



International Federation for Information Processing

<http://www.ifip.org/cal_even.htm> **IFIP Event 02691**

The Seventh International Working Conference



MAKING KNOWLEDGE WORK

HET-NETs 2013



**Performance & Security Modelling and Evaluation of Cooperative
Heterogeneous Networks**

in conjunction with

The First EU PhD Course NET-PEN 2013

Networks and Performance Engineering

Mon 11th – Wed 13th Nov. 2013

Venue: Craiglands Hotel, Ilkley, West Yorkshire, England, UK

HET-NETs 2013

TECHNICAL E-PROCEEDINGS

**Editors: Demetres D. Kouvatsos,
Simonetta Balsamo and Yutaka Takahashi**

**Research Webshelf of River Publishers
Aalborg, Denmark <www.riverpublishers.com>**

ISBN: 978-87-93102-32-3

Table of Contents

Preface

HET-NETs 2013 Chairs and Technical Programme (TPCs) and Local Organising (LOCs) Committees

Lists of HET-NETs 2013 Invited Talks, Technical Papers, Tutorials and Works-in-Progress

PART ONE Performance Modelling and Evaluation

- **Optimal Server Allocation in Service Centres and Clouds**
IT03 Isi Mitrani (Invited Talk - Univ. of Newcastle, England)
- **New Trends in Clinical Trials and Perspectives in Queueing Networks**
IT07 Vladimir Anisimov (Predictive Analytics, Innovation, Quintiles, England)
- **Performance Analysis of a Hybrid Wireless Network with Batch Arrival Process**
P01 Wei Li (ShanDong Univ., China) and Demetres D. Kouvatsos (Univ. of Bradford, England).
- **A Chaos Logistic Map to Model the Delay Tolerant Cyber-Physical Systems**
P03 Shulin Zhao, Hongjun Dai, Wei Li and Lin Lu (Shandong Univ., China)
- **Influence of the Handoff Threshold Hysteresis on Heterogeneous Wireless Network Performance**
P05 Krzysztof Grochla and Konrad Polys (Polish Academy of Sciences, Poland)
- **Performance Evaluation of a Combined System of Random Linear Network Coding and Convolutional Code with Interleaving for Two-Hop Wireless Networks under Rician Fading Channel**
P06 Misfa Susanto, Yim Fun Hu and Prashant Pillai (Univ. of Bradford, England)
- **Performance Analysis of IP Mobile Multicast Mechanisms during Handover in Next Generation Satellite Networks**
P09 Esua Kinyuy Jaff, Prashant Pillai and Yim Fun Hu (Univ. of Bradford, England)
- **Performance Modelling and Analysis of MAC Protocols in Wireless LANs under Multimedia Traffic**
T08 Jia Hu (Liverpool Hope Univ., England), Geyong Min (Univ. of Exeter, England) and Mike E. Woodward (Univ. of Bradford, England)
- **Performance Evaluation of Cognitive Radio Networks under Common Handoff Queue**
WP02 Salah Zahed (The Higher Inst. of Comprehensive Professions-Ghadames, Libya & Univ. of Bradford, England), Irfan Awan and Andrea Cullen (Univ. of Bradford, England)
- **Performance Evaluations for Situational Awareness Based on Scanning Techniques**
WP03 Orabi Shurrab, Evangelos Spatharas, Jules Pagna Disso (EDAS Innovation Work, UK) and Irfan Awan (Univ. of Bradford, England)
- **Performance Evaluation of Cognitive Radio Networks under Reactive-Decision Spectrum Handoff Scheme**
WP04 Salah Zahed (The Higher Inst. of Comprehensive Professions-Ghadames, Libya & Univ. of Bradford, England), Irfan Awan and Andrea Cullen (Univ. of Bradford, England)
- **Performance Analysis of a Pre-emptive Priority Soft Handoff Scheme for Wirelbased Cellular Mobile Networks with Busy Traffic**
WP09 M.I.Bello (Bayero Univ., Nigeria & Univ. of Bradford, England)

- **Performance Modelling and Analysis of Network on Chip under m-port n-tree Bursty Traffic with Virtual Channels**

WP10 Hatem Ibrahim (Univ. of Cairo, Egypt & Univ. of Bradford, England and Irfan Awan (Univ. of Bradford, England)

PART TWO Network Architectures and Protocols

- **Architecture for Automatic Reconfiguration of Network Devices in Response to Events in the Network**

P02 Krzysztof Grochla and Leszek Naruszewicz (Proximity Poland, Poland)

- **A Unified Congestion Control Architecture Design to Improve Heterogeneous Wireless Network Efficiency and Accommodate Traffic by Various Rich Applications**

WP07 Katsuhiko Kusahata, Itsuki Kaneko, Katsuyoshi Iida (Tokyo Inst. of Technology, Japan), Hiroyuki Koga (Univ. of Kitakyushu, Japan) and Masayo Shimamura (Tokyo Inst. of Technology, Japan)

PART THREE Performance vs. Security Trade-offs

- **Performance Modelling of Security Protocols**

P15 Nigel Thomas and Yishi Zhao (Univ. of Newcastle, England)

- **Performance vs. Security Trade-Offs in RANETs based on Hybrid Quantitative Network Models**

T03 Guzman M.A. Miskeen (Univ. of Sebha, Libya & Univ. of Bradford, England), Demetres D. Kouvatso (Univ. of Bradford, England) and Esmail Habibzadeh (IUST - Iran Univ. of Science and Technology & Univ. of Bradford, England)

PART FOUR Network Mobility, Management and QoS

- **Analysing Trade-offs in Mobile Offloading Systems**

IT01 Katinka Wolter (Invited Talk - Univ. of Berlin, Germany)

- **A CIM-based Approach for Managing Computing Servers and Hypervisors Acting as Active Network Elements**

P07 Dimitris Kontoudis and Panayotis Fouliras (Univ. of Macedonia, Greece)

- **Investigating Quality of Service of Wireless Network using Constant Bit Rate Traffic under Random based Mobility Management**

P18 Shahzad Rizwan, Farhan Aadil (COMSATS Inst. of Info. Tech., Pakistan) and Faiz ur Rahman (Univ. of Agriculture, Pakistan)

- **Video Quality Prediction over LTE Using Random Neural Networks**

P20 Tarik Ghalut and Hadi Larijani (Glasgow Caledonian Univ., Scotland)

- **Work in Progress: Environment Aware Optimum Routing Algorithm (EAR) for Resource Efficient Utilization in Delay Tolerant Network (DTN)**

WP01 Shahzad Rizwan (CIIT Attock, Pakistan)

- **Investigating the Pricing Impact on the QoS of a Server Farm Deployed on the Cloud**

WP05 A.M.D Aljohani (Tabuk Univ., Saudi Arabia & Univ. of Bradford, England), D.R.W Holton and Irfan Awan (Univ. of Bradford, England)

PART FIVE Congestion and Admission Control

- **Statistical Tools for Admission Control Decisions in Wireless Networks**

P10 Claudina Rattaro, Pablo Belzarena (Univ. de la Republica, Uruguay), Javier Sim (Univ. Rey Juan Carlos, Espana) and Ivan Hern (Univ. del Cauca, Colombia)

- **Congestion in W-CDMA Networks Supporting Calls of Finite Sources**

P14 I.D. Moscholios (Univ. of Peloponnese, Greece), G.A. Kallos (BT Tech., Service and

Operations, England), V.G. Vassilakis (Univ. of Surrey, England), M.D. Logothetis and M. N. Koukias (Univ. of Patras, Greece)

- **Traffic Control Based Call Admission Control Scheme in Mobile Cellular Networks**

P21 Mohamad Alrowili (King Saud Univ., Saudi Arabia & Univ. of Bradford, England)

PART SIX Research, Standardisation and Networking

- **Research Input to Standardisation: Technology Innovation Through Standardisation**

IT02 Hermann Brand (Invited Talk - European Telecommunication Standards Institute (ETSI), Sophia-Antipolis, France)

PART SEVEN Modelling Techniques and Tools

- **Analysis of Packet Transmission Processes Governed by Immanent Dependences and Heavy-Tailed Distributions**

IT06 Natalia M. Markovich (Inst. of Control Sciences of Russian Academy of Sciences, Moscow, Russia)

- **ITU-R and WINNER II Path Loss Modelling of Femtocells**

P04 Harold O. Kpojime, Ghazanfar A. Safdar, and Mehmet E. Aydin (Univ. of Bedfordshire, England)

- **Transient Analysis of Multi-Priority Queues $M_r/G_r/1$ with Disasters**

P08 Kh. Kerobyan (Society of Computer and Information Systems, Armenia & Los Angeles City College, USA), R. Badalyan (Los Angeles City College, USA) and R. Kerobyan (Soc. of Computer and Inform. Systems, Armenia)

- **A Product-Form Solution for On-Off Components in PEPA**

P12 Nigel Thomas (Univ. of Newcastle, England)

- **Transient MMAP_r(t)/G_r/∞ Model with Disasters**

P13 Kh. Kerobyan (Society of Computer and Inform. Systems, Armenia & Los Angeles City College, USA) and R. Kerobyan (Los Angeles City College, USA)

- **Synchronisation Semantics for Networks with Fork and Join: A Product-Form Approach**

P16 Simonetta Balsamo and Andrea Marin (Univ. of Venice, Italy)

- **The GE/GE/1/N Queue Subject to State-Dependent Arrival Balking**

P23 Neelkamal Shah and Demetres D. Kouvatsos (Univ. of Bradford, England)

- **Analysis of Continuous Time Single Server Queuing Model with Batch Renewal Arrivals: GI^c/M/1/N**

T01 Wei Li (ShanDong Univ., China), Demetres D. Kouvatsos and Rod Fretwell (Univ. of Bradford, England)

- **Extended Entropy Maximisation and Queues with Heavy-Tailed Distributions**

T02 Demetres D. Kouvatsos (Univ. of Bradford, UK)

- **Hybrid Petri Queuing Networks Simulator (HPQNS)**

T05 Esmail Habibzadeh (IUST - Iran Univ. of Science and Technology & Univ. of Bradford, England), Demetres D. Kouvatsos (Univ. of Bradford, England) and Guzman M.A. Miskeen (Univ. of Sebha, Libya & Univ. of Bradford, England)

- **Analytical Modelling of Coordinated Multi-Point (CoMP) based handover for Next Generation Wireless Networks (NGWN)**

WP06 R. R. Ahmed (Quaid_e_Azam University, Pakistan & Univ. of Bradford, England)

➤ **Lightweight Clustering of Cell IDs into Meaningful Neighbourhoods**

WP08 Marios Fanourakis, Katarzyna Wac (Univ. of Geneva, Switzerland)

PART EIGHT Traffic Modelling and Engineering

➤ **Heterogeneous Traffic Scheduling for Multi-Service Communication Networks**

IT05 Geyong Min (Invited Talk - Univ. of Exeter, England)

➤ **Bi-Greedy Cognitive Radio MAC Protocol for Supporting TCP/UDP Traffic**

P24 Humaira Afzal, M. Rafiq Mufti and Irfan Awan (Univ. of Bradford, England)

➤ **Policy Based RAT Selection incorporating IEEE 802.21 Framework for Analysis of Cellular Traffic Offloading through Wi-Fi**

P25 A. Ahmed, M. Ali, Y.F. Hu and P. Pillai (Univ. of Bradford, England)

PART NINE Trust-based Modelling of MANETS

➤ **A Recommendation-Based Trust Model for MANETs to Enhance Dynamic Recommender Selection Using Multiple Rules**

T06 Antesar M.Shabut (Univ. of Benghazi-Libya & Univ. of Bradford), Keshav Dahal (Univ. of Glasgow, Scotland) and Irfan Awan (Univ. of Bradford, England)

➤ **A Trust-Based Monitoring Model to Secure Routing Protocol in MANETs Using Enhanced Trust Metric**

T07 Antesar M.Shabut (Univ. of Benghazi-Libya & Univ. of Bradford, England), Keshav Dahal (Univ. of Glasgow, Scotland) and Irfan Awan (Univ. of Bradford, England)

PART TEN Performance Evaluation Algorithms for VANETs

➤ **Performance Evaluation of VANETs with Multiple Car Crashes in Different Traffic Conditions**

P19 Georgios Charalampopoulos (Hellenic Open Univ. & Univ. of Patras, Greece) and Tasos Dagiuklas (Hellenic Open Univ., Greece)

➤ **Context-aware Rate Adaptation Algorithm for DSRC Vehicular Networks**

P22 Ayoade Ilori, Zuoyin Tang (Aston Univ., England), Yue Li (Research Innovation & Swansea Univ., Wales) and Jianhua He (Univ. of Swansea, Wales)

➤ **Vehicular Ad Hoc Networks (VANETs), Past Present and Future: A Survey**

T04 Farhan Aadil, Shahzad Rizwan (CIIT Attock, Pakistan) and Adeel Akram (Univ. of Agriculture, Pakistan)

PART ELEVEN Performance Issues in Optical Networks

➤ **Sustainable Load in Non-Equidistant Optical Buffers**

P11 W. Rogiest and H. Bruneel (Ghent Univ., Belgium)

➤ **A Novel OLT Based Energy Efficiency Algorithm in TDM Passive Optical Networks**

P17 O.C. Turna (IMT/Telecom SudParis, France, Istanbul Univ., Turkey), M.A. Aydin (Istanbul Univ., Turkey) and Tulin Atmaca (IMT/Telecom A. Sud-Paris, France)

PART TWELVE Acoustic Networks

➤ **Underwater Acoustic Networks**

IT04 Javier Poncea (Invited Talk – Univ. of Malaga, Spain)

Preface

The Seventh Int. HET-NET 2013 Working Conference on the ‘Performance & Security Modelling and Evaluation of Cooperative HETerogeneous NETworks’ was organised under the auspices of IFIP - International Federation of Information Processing (IFIP Event 02691) and in collaboration with the University of Bradford, Bradford, England (U.K.).

This event has its roots to one of the main work-packages of the Research Dissemination Programmes of the Networks of Excellence (NoE) Euro-NGI and Euro-FGI focusing on the ‘Design and Engineering of the Next (NGI) and Future (FGI) Generation Internets’, respectively – towards the ‘Convergence of Multi-service Heterogeneous Networks’. Moreover, it continues to be supported by the earlier EU IST Consortia IASON and VITAL concerning with the ‘Generic Evaluation Platform for Services Interoperability and Networks’ and ‘Mobility Management and Quality-of-Service (QoS) Mechanisms for Convergent IP Multimedia Services (IMS) over Next Generation Networks’, respectively.

The HET-NETs 2013 conference aimed

- i. To motivate fundamental theoretical and applied research into the performance and security modelling, analysis, standardisation and engineering of evolving, cooperative and converging multi-service networks of diverse technology and the FGI;
- ii. To support the staging of the First EU PhD course NET-PEN 2013 on ‘NETworks and Performance ENgineering’.

As in past events, the HET-NETs 2013 Working Conference, along with the associated EU PhD Course NET-PEN 2013, brought together experts as well as research students from both industry and academia worldwide. These events provided the first ever platform for the sharing of knowledge and the exchange of innovative ideas into optimal performance vs. security trade-offs for cooperative heterogeneous networks. They also adopted, as essential main themes, the ‘Mobility management and quality-of service (QoS)’ and ‘Innovative modelling quantitative techniques and tools’.

The HET-NETs 2013 Working Conference adopted original technical submissions in

- *Network Traffic Modelling, Characterisation and Engineering;*
- *Modelling Performance vs. Security Trade-offs in Heterogeneous Networks;*
- *Energy-Aware Security and QoS Routing in Wireless Networks;*
- *Queueing Theoretic-based Models for Intrusion Detection and Elimination Systems;*
- *Performance and Security Issues in Cognitive Radio Networks (CRNs);*
- *Protocols for Vehicular (VANETs) and Robotic (RANETs) Ad Hoc Networks;*
- *Numerical, Simulation and Analytic Methodologies and Quantitative Tools;*
- *Queueing Network Models (QNMs) and Gen. Stochastic Petri Nets (GSPNs) with Blocking;*
- *Network Architectures, Protocols, Clouds and Admission Control Mechanisms*
- *QoS and Mobility Management in Wireless Networks;*
- *Modelling Tools and Quantitative Analysis Techniques;*
- *Management and End-to-End QoS in Heterogeneous Networks;*
- *Networks Standardisation Issues and Activities;*

Following the working conference, submissions of extended HET-NETS 2013 papers and tutorials were invited by the revised date of Monday the 17th of March 2014 for consideration for publication (SCI indexed), subject to additional peer reviews, in Special Issues, including the journals of WIRE, COMCOM and Green Engineering, as appropriate.

The First EU PhD Course NET-PEN 2013 was based on a selective set of invited talks, technical papers and tutorials of the HET-NETs 2013 Working Conference that were associated with the field of Networks and Performance Engineering. The course highlighted important theoretical and applied themes, such as performance vs. security trade-offs in networks, the

impact of bursty and /or correlated traffic flows on network performance, quantitative methodologies and the employment of modelling tools such as arbitrary queueing networks models (QNMs) and generalised stochastic Petri nets (GSPNS) as well as their hybrid integration and related applications. At the end of the PhD course research students were able to appreciate the suitability of alternative quantitative methodologies and tools for the analysis of QNMs and GSPNs for heterogeneous networks and related applications. Consequently, they enhanced further the theoretical and practical contents of research work in the context of their own PhD projects.

The Academic Programme of NET-PEN 2013 PhD Course included the following subset of Invited Talks (ITs), Research Papers (Ps) and Tutorials (Ts) of the HET-NETs 2013 Working Conference, namely IT01 – IT07, P01, P05, P10, P12, P15, P16, T02, T03, T05, T07 and T08 that can be seen in the Contents of HET-NETs E-Proceedings. A special Certificate of Attendance of NET-PEN 2013 was sent progressively as a scanned document to all research students who attended the PhD Course.

We wish to end this Foreword by expressing our deepest thanks to IFIP for the invaluable adoption and support of HET-NETs 2013 and also to the members of TP Committees for their professional services towards the initial assessment and selection of HET-NET 2013 submissions. Our thanks are also due to the invited speakers and the international authors of accepted works for their excellent technical contributions.

Moreover, our thanks are extended for the continuing encouragement, support and collaboration to the staging of HET-NETs 2013 and NET-PEN 2013 of

- The EU Networks of Excellence (NoE) ‘Euro-NGP’ and ‘Euro-FGI’ that involved progressively the services of 57 EU Research Groups and
- The EU IST Projects ‘IASON’ and ‘VITAL’ that drew the participation of several prominent organisations such as Telekom Austria AG (Austria), Teletel SA (Greece), Siemens AG (Germany), Alcatel SEL (Germany), Voiceglobe sprl (Greece), Keletron (Greece), Solinet GmbH (Germany), Telefonica I+D (Spain), Tellas SA (Greece), University of Patras (Greece) and the University of Bradford (U.K.);

Finally, we also wish to thank the Informatics Research Institute (IRI) of the Univ. of Bradford (U.K.) including the staff and student members of the Local Committee for their great support and administrative assistance during the course of the organisation of HET-NETs 2013 and NET-PEN 2013.

Demetres D. Kouvatsos, Simonetta Balsamo and Yutaka Takahashi
Nov. 2013

HET-NET's 2013 General Chair

Demetres D. Kouvatsos (University of Bradford, UK)

HET-NET's 2013 TPC Co-Chairs

Simonetta Balsamo (University of Ca' Foscari of Venetia, Italy)

Yutaka Takahashi (Kyoto University, Japan)

TP Committees

<i>Track I</i> <i>Perform. vs. Security Trade-offs</i>	<i>Track II</i> <i>Mobility Management & QoS</i>	<i>Track III</i> <i>Analytic Methods and Tools</i>
Tulin Atmaca, Paris, France Irfan Ullah Awan, Bradford, U.K. Monique Becker, Paris, France Hermann Brand, ETSI, France Tien Do, Budapest, Hungary Jia Hu, Liverpool, U.K. Karin Anna Hummel, Vienna, Austria Christoforos Kavadias, Teletel SA, Greece Michael Logothetis, Patras, Greece Jim Martin, Clemson, U.S.A. Henrik Schiöler, Aalborg, Denmark Charles Shoniregun, ICITST, U.K. Maciej Stasiak, Poznan, Poland Zhili Sun, Surrey, U.K. Nigel Thomas, Newcastle, U.K. Tereza Vazao, Lisbon, Portugal Sabine Wittevrongel, Ghent, Belgium	Vicente Casares-Giner, Valencia, Spain Der-Jiunn Deng, Nat Univ., Taiwan Ilias Iliadis, IBM Zurich, Switzerland Krzysztof Grochla, Gliwice, Poland Kimon Kontovassilis, Athens, Greece George Kormentzas, Samos, Greece Nicholas Mitrou, Athens, Greece Ioannis Moscholios, Tripolis, Greece Michèle Pagano, Pisa, Italy Rubem Pereira, Liverpool, U.K. Vicent Pla, Valencia, Spain Javier Poncela, Malaga, Spain Adrian Popescu, Karlskrona, Sweden Roberto Sabella, Ericsson, Italy David Remondo-Bueno, Barcelona, Spain Ross Velentzas, AdvTec Ltd., U.K. Vassilis Velentzas, Teletel S.A., Greece	Vladimir Anisimov, Reading, U.K. Pablo Belzarena, Montevideo, Uruguay Chris Blondia, Antwerp, Belgium Richard Boucherie, Amsterdam, Netherlands Herwig Bruneel, Ghent, Belgium Ram Chakka, Meerut, India Tadeusz Czachorski, Gliwice, Poland Hadi Larijani, Glasgow, U.K. Wei Li, JiNan ShanDong, P.R. China Geyong Min, Exeter, U.K. Isi Mitrani, Newcastle, U.K. Sándor Molnár, Budapest, Hungary Natalia M Markovich, Moscow, Russia Michal Pioro, Warsaw, Poland Maria Simon, Montevideo, Uruguay Harry Skianis, Samos, Greece Villen-Altamirano, Madrid, Spain

Local Organising Committee

Guzlan M A Miskeen (Chair), Rana R Ahmed, Irfan Ullah Awan, Mukhtar I Bello, Attila Csenki, Rod J Fretwell, Pauline M L Chan, Esmail Habibzadeh, Neil Saxton, Karim Hamoud Sadik, Neel P Shah, Abdulaziz Aljohani, Hatem Ibrahim.

LISTS OF HET-NETs 2013
INVITED TALKS (ITs), PAPERS (Ps), TUTORIALS (Ts)
& WORKS-IN-PROGRESS (WPs)

INVITED TALKS

- **Analysing Trade-offs in Mobile Offloading Systems**
IT01 Katinka Wolter (Invited Talk - Univ. of Berlin, Germany)

- **Research Input to Standardisation: Technology Innovation Through Standardisation**
IT02 Hermann Brand (Invited Talk - European Telecommunication Standards Institute (ETSI), Sophia-Antipolis, France)

- **Optimal Server Allocation in Service Centres and Clouds**
IT03 Isi Mitrani (Invited Talk - Univ. of Newcastle, England)

- **Underwater Acoustic Networks**
IT04 Javier Poncea (Invited Talk – Univ. of Malaga, Spain)

- **Heterogeneous Traffic Scheduling for Multi-Service Communication Networks**
IT05 Geyong Min (Invited Talk - Univ. of Exeter, England)

- **Analysis of Packet Transmission Processes Governed by Immanent Dependences and Heavy-Tailed Distributions**
IT06 Natalia M. Markovich (Inst. of Control Sciences of Russian Academy of Sciences, Moscow, Russia)

- **New Trends in Clinical Trials and Perspectives in Queueing Networks**
IT07 Vladimir Anisimov (Predictive Analytics, Innovation, Quintiles, England)

PAPERS

- **Performance Analysis of a Hybrid Wireless Network with Batch Arrival Process**
P01 Wei Li (ShanDong Univ., China) and Demetres D. Kouvatsos (Univ. of Bradford, England).

- **Architecture for Automatic Reconfiguration of Network Devices in Response to Events in the Network**
P02 Krzysztof Grochla and Leszek Naruszewicz (Proximity Poland, Poland)

- **A Chaos Logistic Map to Model the Delay Tolerant Cyber-Physical Systems**
P03 Shulin Zhao, Hongjun Dai , Wei Li and Lin Lu (Shandong Univ., China)

- **ITU-R and WINNER II Path Loss Modelling of Femtocells**
P04 Harold O. Kpojime, Ghazanfar A. Safdar, and Mehmet E. Aydin (Univ. of Bedfordshire, England)

- **Influence of the Handoff Threshold Hysteresis on Heterogeneous Wireless Network Performance**
P05 Krzysztof Grochla and Konrad Polys (Polish Academy of Sciences, Poland)

- **Performance Evaluation of a Combined System of Random Linear Network Coding and Convolutional Code with Interleaving for Two-Hop Wireless Networks under Rician Fading Channel**
P06 Misfa Susanto, Yim Fun Hu and Prashant Pillai (Univ. of Bradford, England)

- **A CIM-based Approach for Managing Computing Servers and Hypervisors Acting**

- as Active Network Elements**
- P07** Dimitris Kontoudis and Panayotis Fouliras (Univ. of Macedonia, Greece)
- **Transient Analysis of Multi-Priority Queues $M_r/G_r/1$ with Disasters**
- P08** Kh. Kerobyán (Society of Computer and Information Systems, Armenia & Los Angeles City College, USA), R. Badalyan (Los Angeles City College, USA) and R. Kerobyán (Soc. of Computer and Inform. Systems, Armenia)
- **Performance Analysis of IP Mobile Multicast Mechanisms during Handover in Next Generation Satellite Networks**
- P09** Esua Kinyuy Jaff, Prashant Pillai and Yim Fun Hu (Univ. of Bradford, England)
- **Statistical Tools for Admission Control Decisions in Wireless Networks**
- P10** Claudina Rattaro, Pablo Belzarena (Univ. de la Republica, Uruguay), Javier Sim (Univ. Rey Juan Carlos, Espana) and Ivan Hern (Univ. del Cauca, Colombia)
- **Sustainable Load in Non-Equidistant Optical Buffers**
- P11** W. Rogiest and H. Bruneel (Ghent Univ., Belgium)
- **A Product-Form Solution for On-Off Components in PEPA**
- P12** Nigel Thomas (Univ. of Newcastle, England)
- **Transient $MMAP_r(t)/G_r/\infty$ Model with Disasters**
- P13** Kh. Kerobyán (Society of Computer and Inform. Systems, Armenia & Los Angeles City College, USA) and R. Kerobyán (Los Angeles City College, USA)
- **Congestion in W-CDMA Networks Supporting Calls of Finites Sources**
- P14** I.D. Moscholios (Univ. of Peloponnese, Greece), G.A. Kallos (BT Tech., Service and Operations, England), V.G. Vassilakis (Univ. of Surrey, England), M.D. Logothetis and M. N. Koukias (Univ. of Patras, Greece)
- **Performance Modelling of Security Protocols**
- P15** Nigel Thomas and Yishi Zhao (Univ. of Newcastle, England)
- **Synchronisation Semantics for Networks with Fork and Join: A Product-Form Approach**
- P16** Simonetta Balsamo and Andrea Marin (Univ. of Venice, Italy)
- **A Novel OLT Based Energy Efficiency Algorithm in TDM Passive Optical Networks**
- P17** O.C. Turna (IMT/Telecom SudParis, France, Istanbul Univ., Turkey), M.A. Aydin (Istanbul Univ., Turkey) and Tulin Atmaca (IMT/Telecom A. Sud-Paris, France)
- **Investigating Quality of Service of Wireless Network using Constant Bit Rate Traffic under Random based Mobility Management**
- P18** Shahzad Rizwan, Farhan Aadil (COMSATS Inst. of Info. Tech., Pakistan) and Faiz ur Rahman (Univ. of Agriculture, Pakistan)
- **Performance Evaluation of VANETs with Multiple Car Crashes in Different Traffic Conditions**
- P19** Georgios Charalampopoulos (Hellenic Open Univ. & Univ. of Patras, Greece) and Tasos Dagiuklas (Hellenic Open Univ., Greece)
- **Video Quality Prediction over LTE Using Random Neural Networks**
- P20** Tarik Ghalut and Hadi Larijani (Glasgow Caledonian Univ., Scotland)
- **Traffic Control Based Call Admission Control Scheme in Mobile Cellular Networks**
- P21** Mohamad Alrowili (King Saud Univ., Saudi Arabia & Univ. of Bradford, England)

- **Context-aware Rate Adaptation Algorithm for DSRC Vehicular Networks**
P22 Ayoade Ilori, Zuoyin Tang (Aston Univ., England), Yue Li (Research Innovation & Swansea Univ., Wales) and Jianhua He (Univ. of Swansea, Wales)
- **The GE/GE/1/N Queue Subject to State-Dependent Arrival Balking**
P23 Neelkamal Shah and Demetres D. Kouvatsos (Univ. of Bradford, England)
- **Bi-Greedy Cognitive Radio MAC Protocol for Supporting TCP/UDP Traffic**
P24 Humaira Afzal, M. Rafiq Mufti and Irfan Awan (Univ. of Bradford, England)
- **Policy Based RAT Selection incorporating IEEE 802.21 Framework for Analysis of Cellular Traffic Offloading through Wi-Fi**
P25 A. Ahmed, M. Ali, Y.F. Hu and P. Pillai (Univ. of Bradford, England)

TUTORIALS

- **Analysis of Continuous Time Single Server Queuing Model with Batch Renewal Arrivals: GI^G/M/1/N**
T01 Wei Li (ShanDong Univ., China), Demetres D. Kouvatsos and Rod Fretwell (Univ. of Bradford, England)
- **Extended Entropy Maximisation and Queues with Heavy-Tailed Distributions**
T02 Demetres D. Kouvatsos (Univ. of Bradford, UK)
- **Performance vs. Security Trade-Offs in RANETs based on Hybrid Quantitative Network Models**
T03 Guzman M.A. Miskeen (Univ. of Sebha, Libya & Univ. of Bradford, England), Demetres D. Kouvatsos (Univ. of Bradford, England) and Esmail Habibzadeh (IUST - Iran Univ. of Science and Technology & Univ. of Bradford, England)
- **Vehicular Ad Hoc Networks (VANETs), Past Present and Future: A Survey**
T04 Farhan Aadil, Shahzad Rizwan (CIIT Attock, Pakistan) and Adeel Akram (Univ. of Agriculture, Pakistan)
- **Hybrid Petri Queuing Networks Simulator (HPQNS)**
T05 Esmail Habibzadeh (IUST - Iran Univ. of Science and Technology & Univ. of Bradford, England), Demetres D. Kouvatsos (Univ. of Bradford, England) and Guzman M.A. Miskeen (Univ. of Sebha, Libya & Univ. of Bradford, England)
- **A Recommendation-Based Trust Model for MANETs to Enhance Dynamic Recommender Selection Using Multiple Rules**
T06 Antesar M. Shabut (Univ. of Benghazi-Libya & Univ. of Bradford), Keshav Dahal (Univ. of Glasgow, Scotland) and Irfan Awan (Univ. of Bradford, England)
- **A Trust-Based Monitoring Model to Secure Routing Protocol in MANETs Using Enhanced Trust Metric**
T07 Antesar M. Shabut (Univ. of Benghazi-Libya & Univ. of Bradford, England), Keshav Dahal (Univ. of Glasgow, Scotland) and Irfan Awan (Univ. of Bradford, England)
- **Performance Modelling and Analysis of MAC Protocols in Wireless LANs under Multimedia Traffic**
T08 Jia Hu (Liverpool Hope Univ., England), Geyong Min (Univ. of Exeter, England) and Mike E. Woodward (Univ. of Bradford, England)

WORKS-IN-PROGRESS

- **Work in Progress: Environment Aware Optimum Routing Algorithm (EAR) for Resource Efficient Utilization in Delay Tolerant Network (DTN)**
WP01 Shahzad Rizwan (CIIT Attock, Pakistan)

- **Performance Evaluation of Cognitive Radio Networks under Common Handoff Queue**
WP02 Salah Zahed (The Higher Inst. of Comprehensive Professions-Ghadames, Libya & Univ. of Bradford, England), Irfan Awan and Andrea Cullen (Univ. of Bradford, England)
- **Performance Evaluations for Situational Awareness Based on Scanning Techniques**
WP03 Orabi Shurrab, Evangelos Spatharas, Jules Pagna Disso (EDAS Innovation Work, UK) and Irfan Awan (Univ. of Bradford, England)
- **Performance Evaluation of Cognitive Radio Networks under Reactive-Decision Spectrum Handoff Scheme**
WP04 Salah Zahed (The Higher Inst. of Comprehensive Professions-Ghadames, Libya & Univ. of Bradford, England), Irfan Awan and Andrea Cullen (Univ. of Bradford, England)
- **Investigating the Pricing Impact on the QoS of a Server Farm Deployed on the Cloud**
WP05 A.M.D Aljohani (Tabuk Univ., Saudi Arabia & Univ. of Bradford, England), D.R.W Holton and Irfan Awan (Univ. of Bradford, England)
- **Analytical Modelling of Coordinated Multi-Point (CoMP) based handover for Next Generation Wireless Networks (NGWN)**
WP06 R. R. Ahmed (Quaid_e_Azam University, Pakistan & Univ. of Bradford, England)
- **A Unified Congestion Control Architecture Design to Improve Heterogeneous Wireless Network Efficiency and Accommodate Traffic by Various Rich Applications**
WP07 Katsuhiko Kusahata, Itsuki Kaneko, Katsuyoshi Iida (Tokyo Inst. of Technology, Japan), Hiroyuki Koga (Univ. of Kitakyushu, Japan) and Masayo Shimamura (Tokyo Inst. of Technology, Japan)
- **Lightweight Clustering of Cell IDs into Meaningful Neighbourhoods**
WP08 Marios Fanourakis, Katarzyna Wac (Univ. of Geneva, Switzerland)
- **Performance Analysis of a Pre-emptive Priority Soft Handoff Scheme for Wirebased Cellular Mobile Networks with Busty Traffic**
WP09 M.I.Bello (Bayero Univ., Nigeria & Univ. of Bradford, England)
- **Performance Modelling and Analysis of Network on Chip under m-port n-tree Bursty Traffic with Virtual Channels**
WP10 Hatem Ibrahim (Univ. of Cairo, Egypt & Univ. of Bradford, England) and Irfan Awan (Univ. of Bradford, England)

PART ONE Performance Modelling and Evaluation

Optimal server allocation in service centres and clouds

Isi Mitrani

School of Computing Science, Newcastle University, NE1 7RU, UK

e-mail: `isi.mitrani@ncl.ac.uk`

Abstract. There is an important class of dynamic optimization problems arising in the market for computer services. These concern the number of servers to be kept in an operative state in a service center, or the number of servers to be hired from a cloud. At certain moments in time, one has to decide how many servers to use or hire, depending on current loading conditions. The costs of running servers have to be traded off against the benefits to be derived from them. In other words, the aim is to find a dynamic operating policy that minimizes the long-term average value of a particular cost function. The latter depends on the system under consideration, but typically takes into account both server costs and performance costs (or quality of service costs).

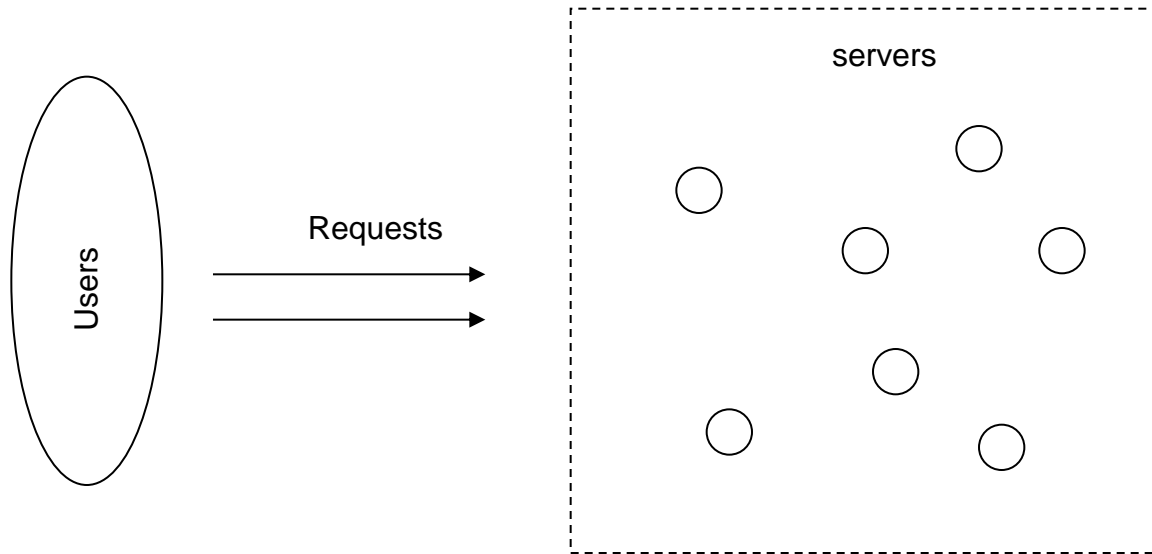
A methodology for tackling such problems is provided by the theory of semi-Markov decision processes. Having made certain assumptions and set the problem in the appropriate framework, it is possible to compute the long-run average cost incurred per unit time by a given operating policy. Moreover, a 'policy improvement' algorithm allows the optimal policy to be determined in a finite number of iterations. That algorithm may be computationally expensive, so one may wish to use heuristic policies that are perhaps sub-optimal, but simple and easily implementable. The quality of a heuristic can then be evaluated by comparing it with the optimal policy.

A brief description of the semi-Markov decision processes and the policy improvement algorithm will be given. Two applications will be presented, one concerning the optimal use of servers in the presence of bursty demands, and the other associated with the optimal brokerage of cloud servers. In both cases, heuristic policies are proposed that turn out to be nearly optimal.

Optimal server allocation in service centres and clouds

Isi Mitrani, Newcastle University

General setup

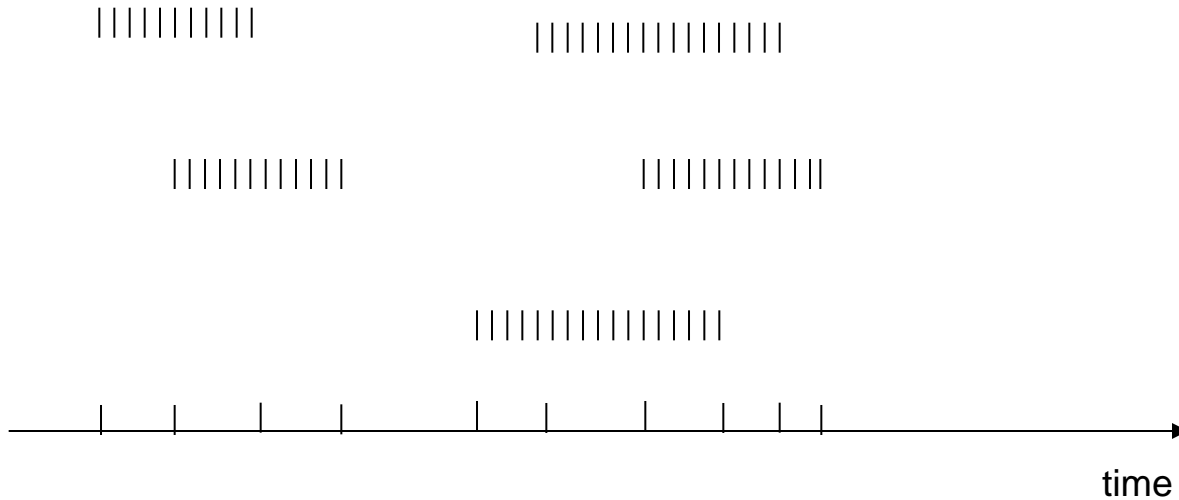


How many servers to allocate?

Dynamic optimization problems:

Given the current loading conditions, together with the costs and benefits of using servers, how many servers should be used at a particular moment in time?

Example 1. Bursty arrivals at a service centre

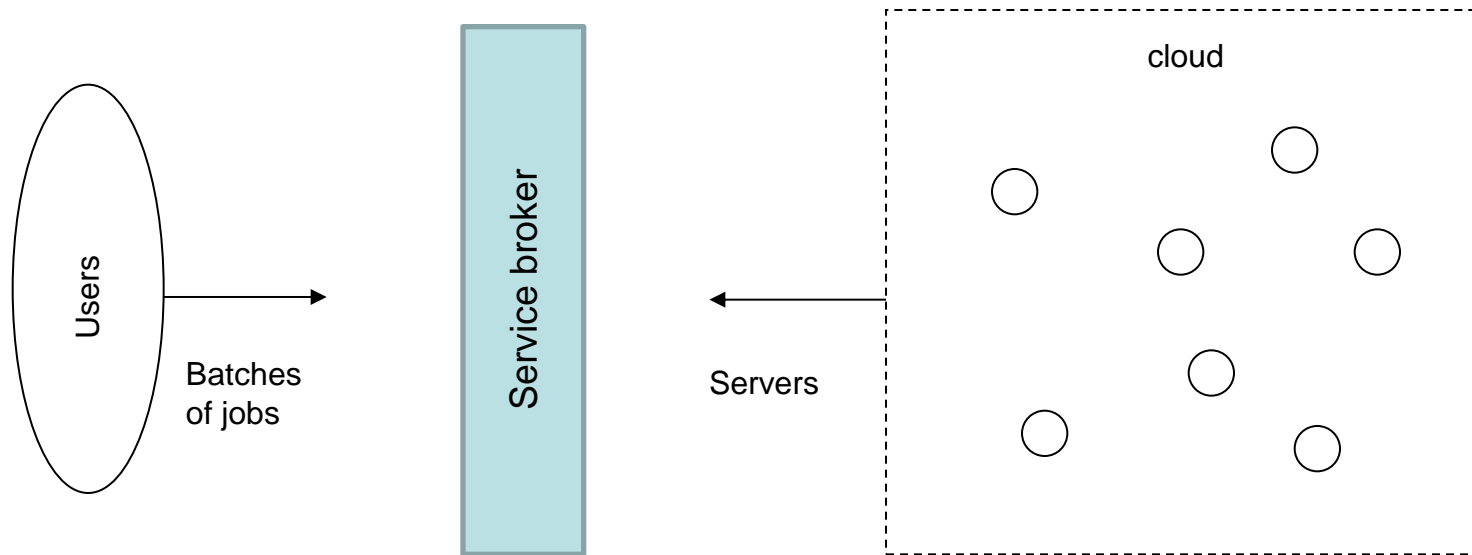


Characteristics of bursts: arrival rate, γ , completion rate, ν ,
job arrival rate, λ , average job service time, b .

Cost of waiting, c_1 , cost of server, c_2 .

How many servers to power on/off when bursts arrive/leave?

Example 2. Hiring servers from a cloud

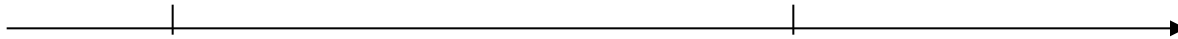


When a new batch of jobs arrives, how many servers should be hired?

General framework (Semi-Markov Decision Process)

State i
action a_i

State j
action a_j



interval until next decision instant: $\tau_i(a)$;

cost incurred until the next decision instant: $c_i(a)$;

transition probability to next state: $p_{i,j}(a)$

Stationary policy: A

Total cost incurred during interval (0,t) under policy A: $Z_A(t)$

Long-term average cost of policy A per unit time: $g(A) = \lim_{t \rightarrow \infty} \frac{1}{t} \mathbf{E}[Z_A(t)]$

Problem: Find policy A that minimizes $g(A)$

Compute the average cost $g(A)$ for a given policy A ,
by solving a set of simultaneous linear equations:

$$v_i = c_i(A) - \tau_i(A)g(A) + \sum_j p_{i,j}(A)v_j$$

$$i = 1, 2, \dots, N$$

Choose an arbitrary state, k , and set $v_k=0$

Find the optimal policy by applying a '*policy improvement*' algorithm:

1. Choose an initial policy, A .
2. Compute v_i and $g(A)$.
3. For every i , find the action a that minimizes v_i .
4. If the new policy is the same as the old one, stop; otherwise repeat from step 2.

The algorithm is guaranteed to terminate after a finite number of iterations:

Greedy heuristic:

For every state i , choose the action that minimizes the cost in the current interval, c_i .

Example 1 (Bursty arrivals):

When i bursts are active, jobs arrive at the rate of $i\lambda$.

New bursts arrive at rate γ .

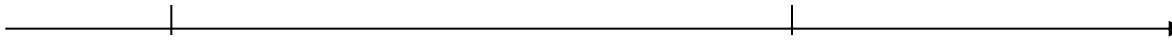
Average length of burst $1/\nu$ (completion rate ν).

Costs incurred for powering servers on,
and also when waiting time exceeds a given limit.

Given the current number of active bursts, how many servers should be powered on?

Decision instants are the moments when new bursts arrive or old ones terminate.

State: i bursts in progress
Action: n servers allocated



M/M/n queue with parameters $(i\lambda, \mu)$;
 $\tau = 1/(\gamma + i\nu)$;
 $p_{i,i+1} = \gamma / (\gamma + i\nu)$;
 $p_{i,i-1} = i\nu / (\gamma + i\nu)$;

$$c(i; n) = c_2 n + i\lambda c_1 P(Q > w) - i\lambda r$$

c_1 = cost of waiting (per job)
 c_2 = cost per server
 r = revenue per job

Since neither the interval until the next decision instant, nor the transition probabilities, depend on the action taken, *the greedy heuristic is optimal.*

$$v_i = c_i(A) - \tau_i(A)g(A) + \sum_j p_{i,j}(A)v_j$$

Example 2 (Batch arrivals):

Batches of jobs arrive at rate λ .

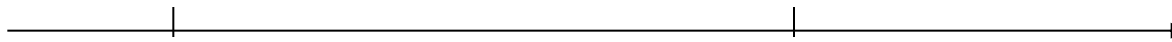
Average number of jobs in batch, b .

Cost incurred for holding jobs in the system, and also for hiring servers.

Given the current number of jobs present, how many servers should be hired from the cloud?

Decision instants are the moments when new batches arrive.

State: i jobs present
Action: n servers hired



$$\tau = 1/\lambda.$$

transition probabilities $p_{i,j}$ depend on n ;
 $j = i - \text{number served} + \text{number in next batch}$

$$c(i,n) = c_1 n/\lambda + c_2 L$$

where $L = \text{average no. of jobs during interval}$

The greedy heuristic is not optimal.

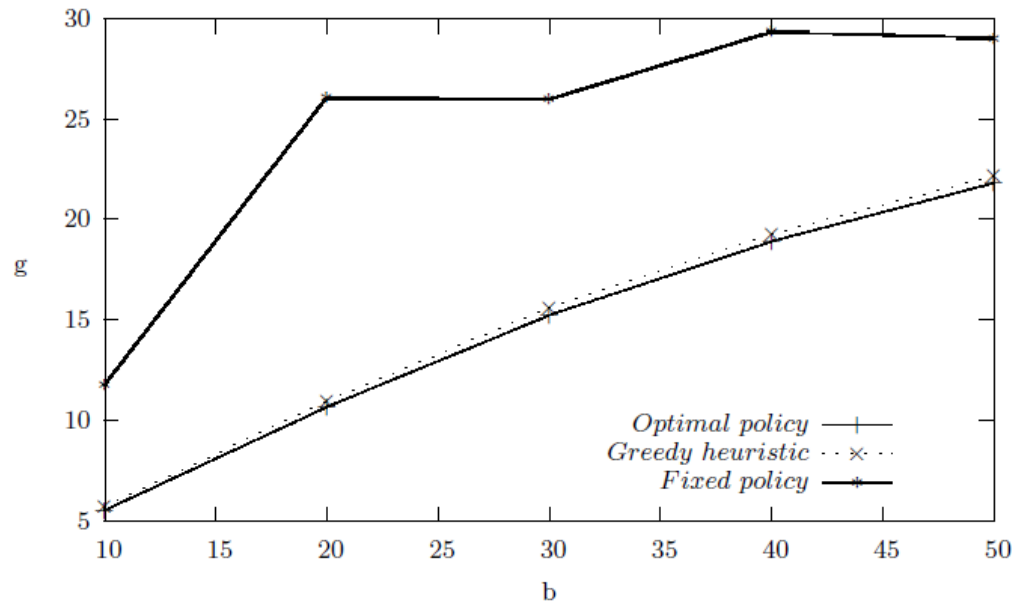


Figure 1: Policy comparison: varying average batch size

New trends in clinical trials and perspectives in queueing networks

Vladimir Anisimov

Innovation, Quintiles, England
Vladimir.Anisimov@quintiles.com

There are many similarities in modelling customer's flows in multichannel queueing networks and patient's flows in multicentre clinical trials. These flows are typically described by multidimensional Poisson processes. However, more complicated models require use of more general type of processes, e.g., rather general queueing models/networks with switching input and service are investigated in book [3] by using switching processes.

Multicentre clinical trials have it's own specific as different centres have different arrival rates and also may have random initiation and closure. To model patient's flows in clinical trials and account for multiple centre's effect, it is natural to introduce several levels of stochastic hierarchy: random variation in enrolment rates of Poisson processes in different clinical centres and random delays of enrolment start-up in each center using some type of distribution.

An analytic methodology for predictive patient's enrolment modelling is developed in [2] and extended further [4]. It accounts for stochasticity in enrolment, variation in enrolment rates described by a gamma distribution, random centres delays and allows predicting with credibility bounds the enrolment over time, time to complete trial and also re-forecasting of enrolment using interim data. It is now accepted as the basic methodology for patient enrolment modelling in clinical trials and implemented on R&D level in some pharmaceutical companies.

The next level of hierarchy is modelling of various operational processes including the number of patients in the trial (similar to customers in service), patient's visits, associated events and related costs. To model these processes constructed on the top of enrolment it is proposed to use hierarchic evolving stochastic processes introduced in [1].

Consider a multichannel network (or multicentre clinical trial) and suppose that each customer at time of arrival generates some evolving process with random life time, in particular, a multi-stage service process or some event process.

To describe the model more precisely, consider a particular node/centre i . Assume that customers (patients) arrive at node i according to a doubly stochastic Poisson process (Cox process) with rate $\lambda_i(t)$ that can be random and time dependent. Denote by $t_{1i} < t_{2i} < \dots$ a sequence of arrival times in centre i . Define a family of stochastic processes $\{\xi_{ki}(t, \theta), 0 \leq t \leq \tau_{ki}\}$, $(\xi_{ki}(t, \theta) = 0, t < 0)$, where τ_{ki} is a random life-time, θ is some unknown parameter, and $\xi_{ki}(\cdot)$ are independent at different i and k with distributions not depending on k . Here

$\xi_{ki}(t, \theta)$ stands for the individual evolving process generated by k -th customer at time of arrival t_{ki} at node i . Consider the sums of evolving processes:

$$Z_i(t) = \sum_{k: t_{ki} \leq t} \xi_{ki}(t - t_{ki}, \theta), \quad Z(t) = \sum_i Z_i(t).$$

In this way we can describe various operational processes associated with patient's/customers flows in multichannel networks and multicentre clinical trials.

Under rather general assumptions it is possible to derive closed-form solutions for the first two moments of the processes $Z_i(t)$ and $Z(t)$. If $\xi_{ki}(\cdot)$ is the process of some events, the predictive distribution of $Z(t)$ can be represented using non-homogeneous doubly-stochastic Poisson process.

As the example, consider modelling the number of patients in the trial (customer's in service). Suppose that in centre i each patient upon arrival can either stay in the trial during follow-up period L or drop-out during this period with a given rate μ_i (possibly random). Define $\xi_{ki}(t) = 1$ as $0 \leq t \leq \tau_{ki}$, and $\xi_{ki}(t) = 0$ as $t > \tau_{ki}$, where τ_{ki} are distributed as $\min(Ex(\mu_i), L)$, where $Ex(\mu)$ has an exponential distribution with parameter μ . Then $Z_i(t)$ represents the number of patients in centre i that stay in trial at time t . Note that $Z_i(t)$ may have a quasi-stationary behaviour during the enrolment period. In a similar way we can represent lost patients and also model the events and related costs.

To describe the evolution of multiple events, e.g., recurrence, death and lost to follow-up (in oncology trials), we use for $\xi_{ki}(t)$ finite Markov or semi-Markov absorbing processes [5]. In this case the number of events at time t can be expressed via Poisson process with time-dependent random cumulative rate which is constructed using arrival rate and transition probabilities of $\xi_{ki}(t)$.

For these models the main predictive characteristics are derived in a closed form, so this technique does not require Monte Carlo simulation. Estimation of the unknown parameter θ is also considered.

The results applied to dynamic modelling of various operational processes and related events in clinical trials including operational costs and are on the stage of implementation. Several case studies are considered.

These models open good perspectives for use in modern queueing networks.

References

1. Anisimov, V.: Limit theorems for evolving accumulation processes. *Theor. Probab. and Math. Statist.*, 43, 5–11 (1991)
2. Anisimov, V., Fedorov V.: Modeling, prediction and adaptive adjustment of recruitment in multicentre trials. *Statistics in Medicine*. 26, 4958-4975 (2007)
3. Anisimov, V.: *Switching Processes in Queueing Models*. John Wiley & Sons, London (2008)
4. Anisimov, V.: Statistical modeling of clinical trials (recruitment and randomization). *Communications in Statistics - Theory and Methods*. v.40, N. 19–20, 3684–3699 (2011)
5. Anisimov, V.: Predictive event modelling in multicentre clinical trials with waiting time to response, *Pharmaceutical Statistics*, v. 10, iss. 6, 517–522 (2011)

New trends in clinical trials and perspectives in queueing models

7th International Conference
HET-NETs 2013

University of Bradford, Ilkley, England



*Prof. Dr. Vladimir Anisimov
Sr Strategic Biostatistics Director*

*Predictive Analytics, Innovation
Vladimir.Anisimov@quintiles.com*

11 Nov 2013

Background - clinical trials



Drug development process includes several stages.

The whole process – 5-10 years, cost £5 -£7 Billions or more.

- **Phases I-II** – testing biologically active compounds, selected the most promising, safety analysis, dose selection,...
- **Phase III** - Late stage: Testing new drug placebo controlled randomized clinical trials
 - large number of patients (hundreds/thousands)
 - patients are enrolled by many clinical centres in different countries
 - after screening patients are randomized and get a prescribed drug
 - it can be a long follow-up period of treatment
- Enrolment is stopped when the total number of recruited patients will reach sample size needed for statistical analysis of patient responses with a given confidence (0.95%)
- **Typical main goal:**
statistically prove that drug A is better than drug B

Design of late-stage clinical trials includes several basic stages:

- Statistical study design
 - models for patient responses and data analysis
- Design of operational characteristics:
 - Patient enrolment
 - duration, number of centres, countries,
 - Events/visits planning/predicting
- Cost/risk and drug supply planning

Design and trial operation are affected by stochasticity in:

- ✓ patient's enrolment; randomisation; event's appearance

Interaction between these stages:

strong impact on the whole process of drug development.

Patient enrolment is a well-recognized bottleneck in clinical trials

- Most of pharma companies and CRO use ad hoc deterministic techniques
- A huge proportion of trials (>70-80%) fail to recruit in time

Natural way:

Model enrolment in multicentre clinical trials using **multiple Poisson processes**

- **similar to modelling flows in multichannel queueing models:**

- Centres – servers or nodes
- Customers – patients
- Input in centre i – Poisson process with rate λ_i

Multicentre clinical trials have it's own specific:

- rates are different in different centres and uncertain (unknown parameters)
- centres may have random delay in initiation and closure.

To model patient's flows it is natural to introduce several levels of stochastic hierarchy:

- random variation in enrolment rates among different centres
- random delays of enrolment start-up in each center

Queueing models vs clinical trials



Similarities & differences

1. Queueing models

- Different types of input flows
Poisson, Markov, renewal, switching,...
- Different types of service
exponential, general, Markov, by stages, retrial,...
- Network models,...

Clinical trials

- Poisson and mixed Poisson
- similar to $(M_R/G_R/\infty)^N$ with random parameters
- No transitions between centres

➤ Basic problems in queueing models

- Analysis of various distributions
queues, customers in service
(waiting, empty and busy period,...)
 - Stationary analysis (queues, times,...)
 - Asymptotic analysis
behaviour of queueing processes at small/large parameters,
low traffic, averaging, diffusion approximation,...
- enrolled patients, pts in trial, random events, visits
 - quasi-stationary, patients in trial
 - Yes averaging, diffusion approx.

Wide opportunities to apply new problems and models in queueing networks

Enrolment models

2. Patient enrolment

- Poisson input flows with random rates with unknown parameters to model variation between centers, regions
- Flows with random delays and stopping times
- Varying number of centres/nodes in time due to random start-up, closure and adding new centres at interim look
- Hierarchic models and event-based trials
“evolving” stochastic processes on the top of enrolment

3. Problems in enrolment modelling

- Statistical estimation of parameters (rates, delays, events) at initial and interim stages
- Analysis of enrolment process – global/in regions, stopping time
- Creating predictive distributions and processes for enrolment, events, centre’s performance, using initial/interim data
- Adaptive adjustment (re-estimating parameters, # of centres at interim look)

Goal - to complete enrolment in time with a given confidence or in optimal way

Analytic statistical methodology is developed *

Poisson-gamma enrolment model:

- Patients arrive according to delayed **Poisson processes**
- Variation in enrolment rates across centres is modelled using **gamma distribution**
- Delays in centre initiation can be random variables

Enrolment process in centre i :

non-homogeneous doubly stochastic Poisson process

$$X_i(t) = \pi_{\lambda_i} (t - v_i) \chi(v_i \leq t \leq b_i)$$

rate $\lambda_i = \gamma(\alpha, \beta)$,

α, β – shape and rate parameters of a gamma distribution,

v_i – delay in centre initiation, b_i – closure time.

Methodology is validated on tens of real trials, world-wide accepted and implemented in some pharma companies

* Anisimov & Fedorov, 2005, 2007; Anisimov, 2009-2011
DIA, JSM (invited sessions), PSI, ...

Case study

Study A – real completed trial:

n=629 pts, N=91 centres

Data -- **special statistic**:

vecv= (v(1), v(2), ...),

v(j) – number of centres recruited j patients.

vecv = (7,11,8,8,9,8,9,7,2,4,1,3,3,4,0,0,2,1,1,2,1,0,0, . .)

7 centres – 1 pt, 11 centres – 2 pts,..., 2 centres – 20 pts, 1 centre – 21 pts.

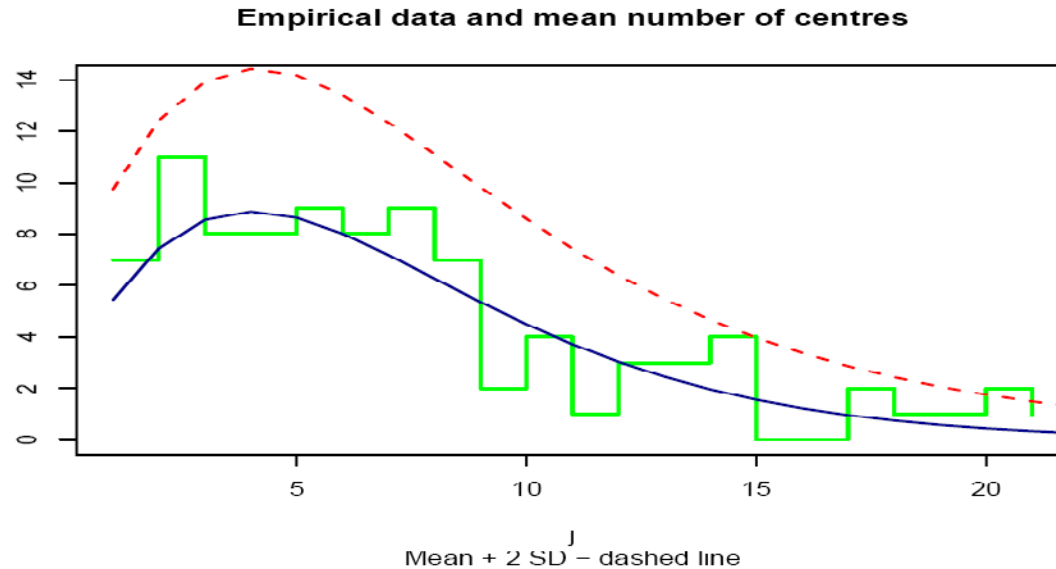
Huge variation in rates,
modelled using a **gamma distribution**.

Theoretical mean number of centres recruited *j* pts –

$$m(n, N, \alpha, j) = N \binom{n}{j} \frac{B(\alpha + j, \alpha(N - 1) + n - j)}{B(\alpha, \alpha(N - 1))}, \quad j = 0, 1, 2, \dots, n.$$

$B(a,b)$ – Beta function.

Verification of enrolment model



Graph of real data (**vector v**) – **step-wise green line**;
mean number of centres $m(n, N, \alpha, j)$ recruited j pts
(theoretical) – **solid blue line**; $\alpha = 2.846$ – ML estimator;
the mean + 2sd – **dashed red line**.

Poisson-gamma model fits real data.

Poisson models with fixed rates behave very differently.

Patient enrolment modelling tool



Basic features:

- Computing predictive distributions, mean and credibility bounds at any stage of the trial for the predictive number of patients recruited over time and for time to stop enrolment

Additional features – different goals compared to queueing networks:

- Evaluating the **minimal number of centres** needed to complete in time with a given confidence
- **Adaptive adjustment (how many centres to add for optimal stopping)**
- Predicting **performance** of centres/countries

The basic version is implemented for all late stage GSK trials

- **significantly improved the efficiency of trials planning/monitoring**

In **Quintiles** it is also developed software packages using this methodology.

Technique:

- Poisson-gamma enrolment model
- Parameters (α, β) of rates are estimated using planned or historical data;
- Delays of centre's initiation have some distribution:
 - fixed, uniform, beta, gamma;
- $\text{mean}[n(t)]$, and $\text{sd}[n(t)]$ are calculated in a closed form ($n(t)$ – total number of recruited patients).
- Predictive bounds for $n(t)$ and for enrolment time are created

If predicted probability to complete in time is less than needed:

- Evaluate # of centres needed to complete with a given confidence

Opportunity to use cost-benefit analysis:

Example: for some trial with 80 centres – average enrolment time 6 months.
to complete in 6 months with prob. 90% - required extra 15 centres
..... with prob. 80% extra 10 centres

Optimal decision can be chosen.

Random centre initiation

n - number of patients, N - number of centers,
vector of rates $\{\lambda_i\}$ in centers,
vector of dates of center's readiness to be initiated $\{\underline{t}_i\}$,
the time of initiation of center i is a gamma distributed variable
with parameters (ψ_i, θ_i) and cumulative probability function $\text{Ga}(x, \psi_i, \theta_i)$.

Denote by $\Lambda_i(t)$ a cumulative rate in center i . Then for $t \geq \underline{t}_i$

$$\mathbf{E}[\Lambda_i(t)] = \mathbf{E}[\lambda_i] \left[(t - \underline{t}_i) \text{Ga}(t - \underline{t}_i, \psi_i, \theta_i) - \frac{\psi_i}{\theta_i} \text{Ga}(t - \underline{t}_i, \psi_i + 1, \theta_i) \right]$$

$$\mathbf{E}[n_i(t)] = \mathbf{E}[\Lambda_i(t)]$$

$$\mathbf{E}[n_{\bullet}(t)] = \sum_{i=1}^N \mathbf{E}[\Lambda_i(t)]$$

where $n_i(t)$ - enrolment process in centre i , $n_{\bullet}(t)$ - global enrolment, rates λ_i are random.
Variance has a similar expression.

For global enrolment process

$$\mathbf{E}[n_{\bullet}(t)] = \sum_i \mathbf{E}[\Lambda_i(t)], \quad \text{Var}[n_{\bullet}(t)] = \sum_i \mathbf{E}[\Lambda_i(t)] + \sum_i \text{Var}[\Lambda_i(t)]$$

(1- δ)-predictive bounds calculated using normal approximation:

$$\mathbf{E}[n_{\bullet}(t)] \pm z(1-\delta/2) \sqrt{\text{Var}[n_{\bullet}(t)]}$$

Trial/centre performance



- Particularities in large trials with many centres:
 - Many low performing sites (unbalanced study)
 - A large proportion of patients is recruited by a small # of sites
 - Increasing enrolment time and study costs
- The question: can it be explained by natural statistical fluctuations?
- The answer: YES

Trial performance for large number of patients n and centres N is mainly determined by only two factors:

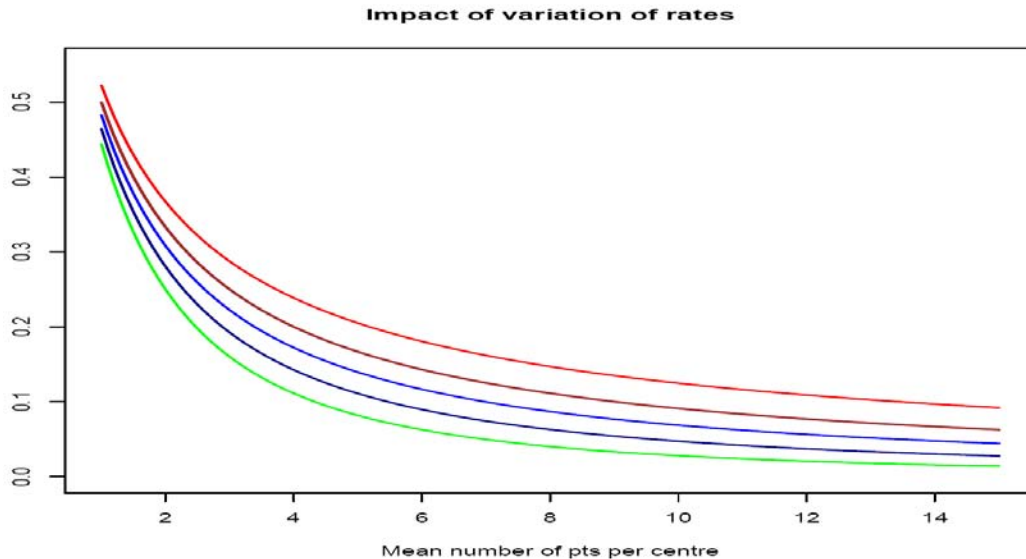
- Mean number of patients in a centre
 $q = n/N$
- Coefficient of variation Cv in enrolment rates across centres:
(shape parameter $\alpha = 1/c_v^2$).

Lemma: When $n \rightarrow \infty$, $n/N \rightarrow q$, for any j ,
(number of centres recruited j patients)/ $N \approx \frac{\Gamma(\alpha + j)}{j! \Gamma(j)} \frac{q^j \alpha^\alpha}{(\alpha + q)^{\alpha+j}}$

Trial/centre performance



Predicting the total fraction of “empty” centres in complete trial



Different curves:

Cv- coef. var.[rates]

Cv = 1.1, 1, 0.9, 0.8, 0.7.

Low curve -- Cv = 0.7.

The variation in rates increases the number of empty centres.

Fraction of the number of empty centres as a function of $q = n/N$.

In trials with small (on average) number of pts per centre the number of “empty” centres is rather substantial.

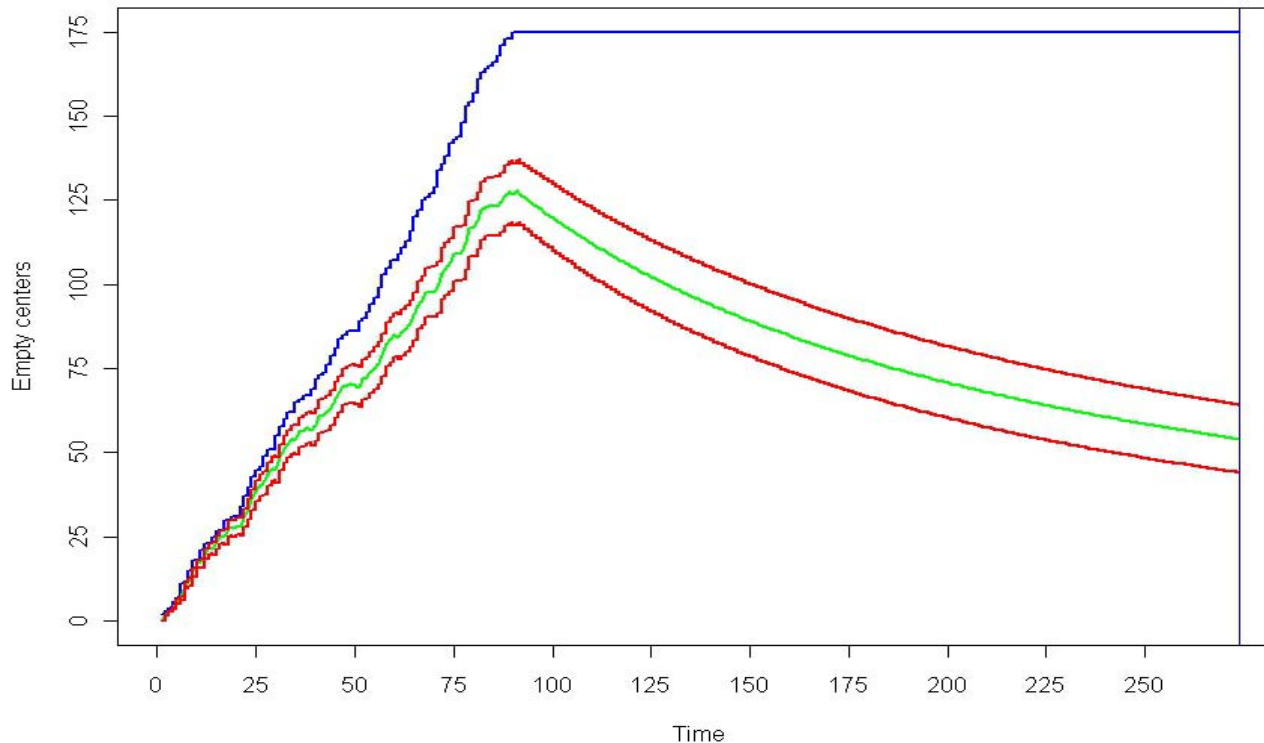
Example: if on average 4 pts/centre - ~15-20% empty centres (oncology trials).

Same consequence for multichannel networks:

number of “empty” nodes on fixed time interval or interval to reach in total n customers.

Predicting dynamics of empty centres

Predicting the number of empty centers



Blue curve:
site's start-up
green curve:
mean # of empty
sites
red curves- 90%
predictive bounds

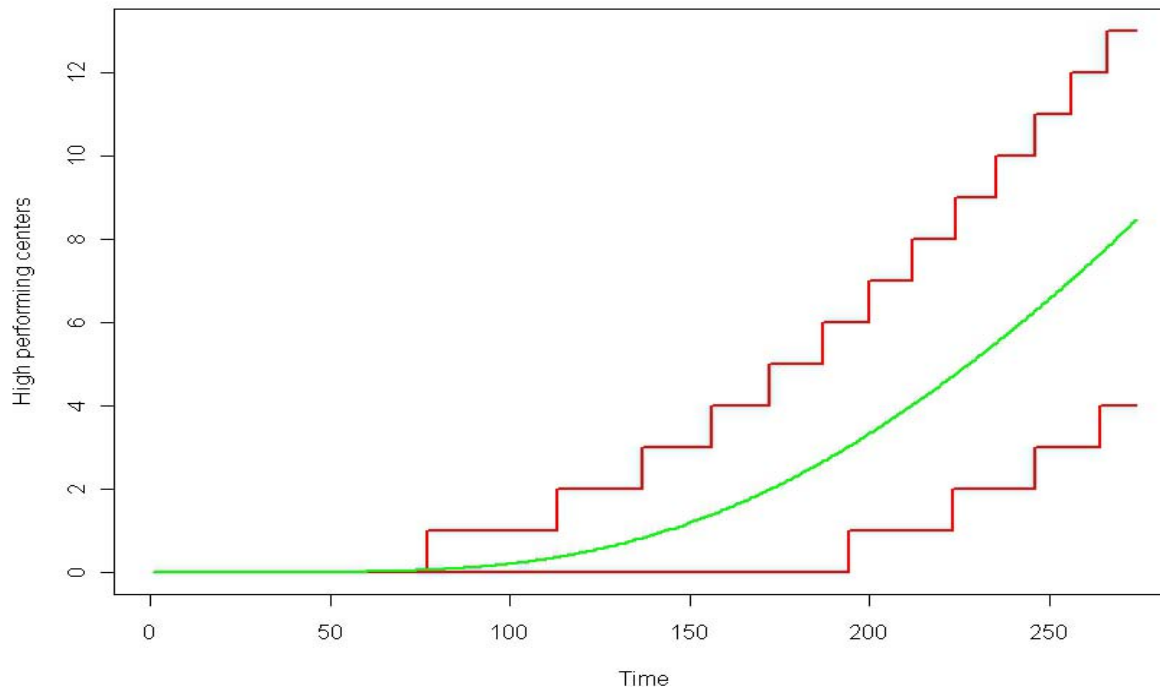
At the end of enrolment:
mean 55,
low bound = 45,
upper bound = 67

Case study: number of patients =450;
number of clinical centres =175;
time to complete enrolment =9 months;
all centres initiated uniformly in 3-month period.

Analogy: networks with random opening of nodes and total performance of nodes.

Dynamics of highly performing centres QUINTILES®

Predicting the number of highly performing centers



Green curve:
mean # of sites
recruited > 8 pts
red curves- 90%
predictive bounds

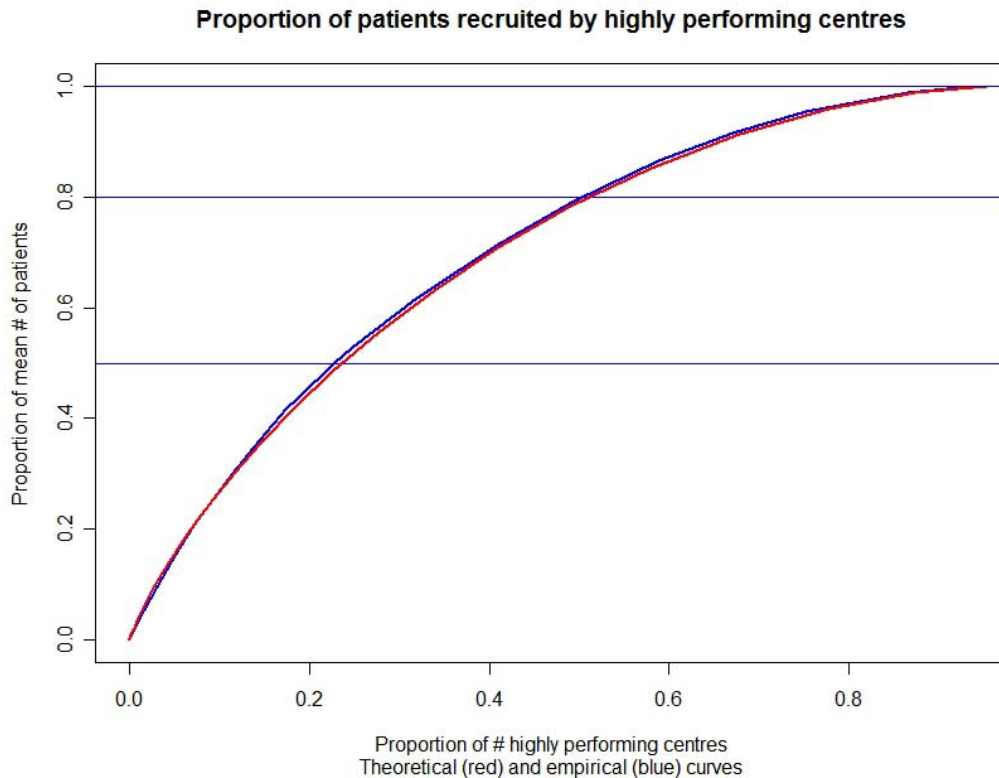
At the end of enrolment:
mean 8,
low bound = 4,
upper bound = 13

Number of centers recruited more than 8 pts each.

Case study: number of patients =450; number of clinical centres =175;
planned time to complete enrolment =9 months; all centres initiated at start up.

Conclusion: during enrolment period with confidence 90% not more than 13 centres can recruit each >8 patients.

Impact of the most productive centres QUINTILES®



Case study A:
Empirical data (blue),
theoretical curve (red):

50% of pts recruited by
~ 22% of centres;

80% of pts recruited by
~ 52% of centres

Typical behaviour:
not a result of a bad planning,
but a Statistical Law

Study A: data: 629 patients, 91 centres;
vecv= (6, 7, 11, 8, 8, 9, 8, 9, 7, 2, 4, 1, 3, 3, 4, 0, 0, 2, 1, 1, 2, 1, 0, 0,...),
Data - number of centres recruited 0, 1, 2, ...pts), $Cv=0.67$

Results confirmed by analysis of large databases of trials.

Methodology:

- Data **at some interim time**:
 - in each centre - duration of active enrolment u_i
 - the number of recruited patients k_i
- Information about future centre initiation (if planned)
- Parameters of enrolment rates (α, β) in **Poisson-gamma** model estimated using ML
- Individual rates in each centre are re-estimated:
Bayesian approach: posterior rate – gamma distributed
$$\lambda_i = \gamma(\alpha + k_i, \beta + u_i)$$
- Predictive process in each center
 - **Poisson-gamma process with rate λ_i**
- Predictive number of patients (over time) and the remaining enrolment time are evaluated (mean and predictive bounds)
closed-form expressions

- **Adaptive adjustment:**
Evaluate **Probability** to recruit in time
If **Probability** is not so high, calculate the number of additional centres required to complete in time with a given probability
- **Predicting low and high enrolling centres:**
given interim data, for any centre evaluate:
 - probability not to recruit any patients during future interval $[0,x]$
(if probability is high (e.g. $\geq 80\%$), alert this centre)
 - probability to recruit not more (or more) than k patients

Example. Predictive probability not to recruit:

(α, β) - estimated parameters of a Poisson-gamma model at interim time,
centre recruited k patients during active enrolment duration τ .

Probability not to recruit any patients in future interval of the length x is

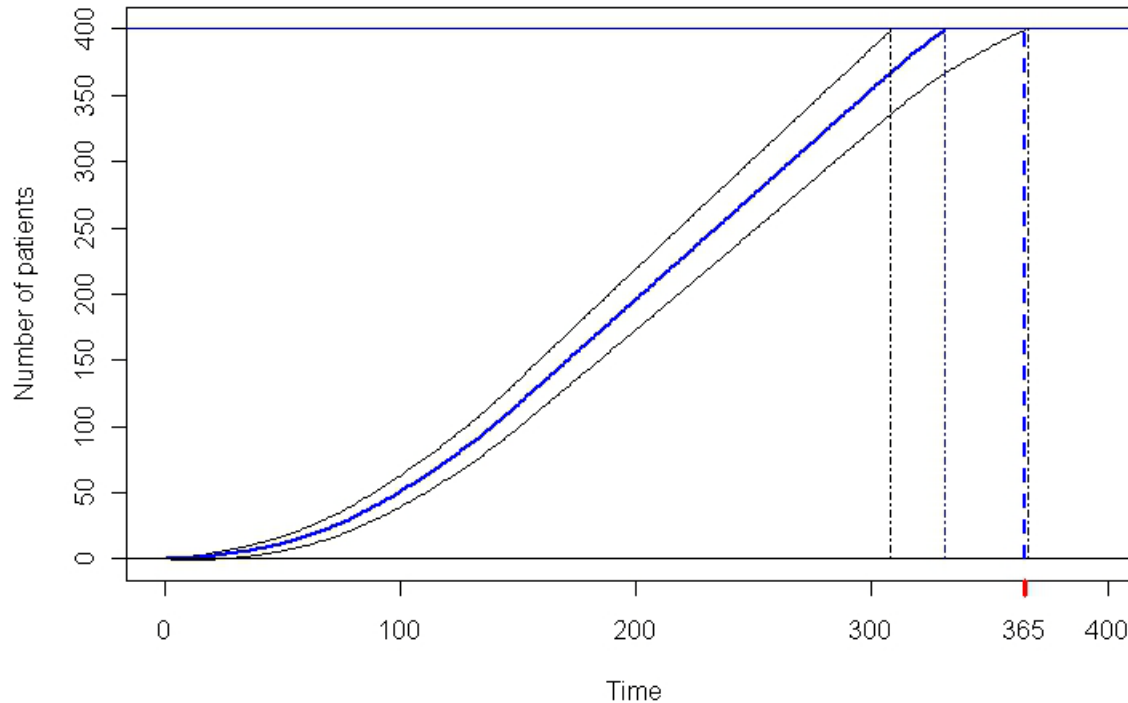
$$\left(\frac{\beta + \tau}{\beta + \tau + x} \right)^{\alpha + k}$$

Similar network problem if unknown input rates.

Optimal enrolment predicting



Adaptive enrollment adjustment



Scenario:

400 pts,
70 centres,
planned time
= 365 days.

70 sites initiated in
5-month period,
half of sites will be
closed during two
months before the
end of enrolment.

New paradigm: “**Plan with confidence**”

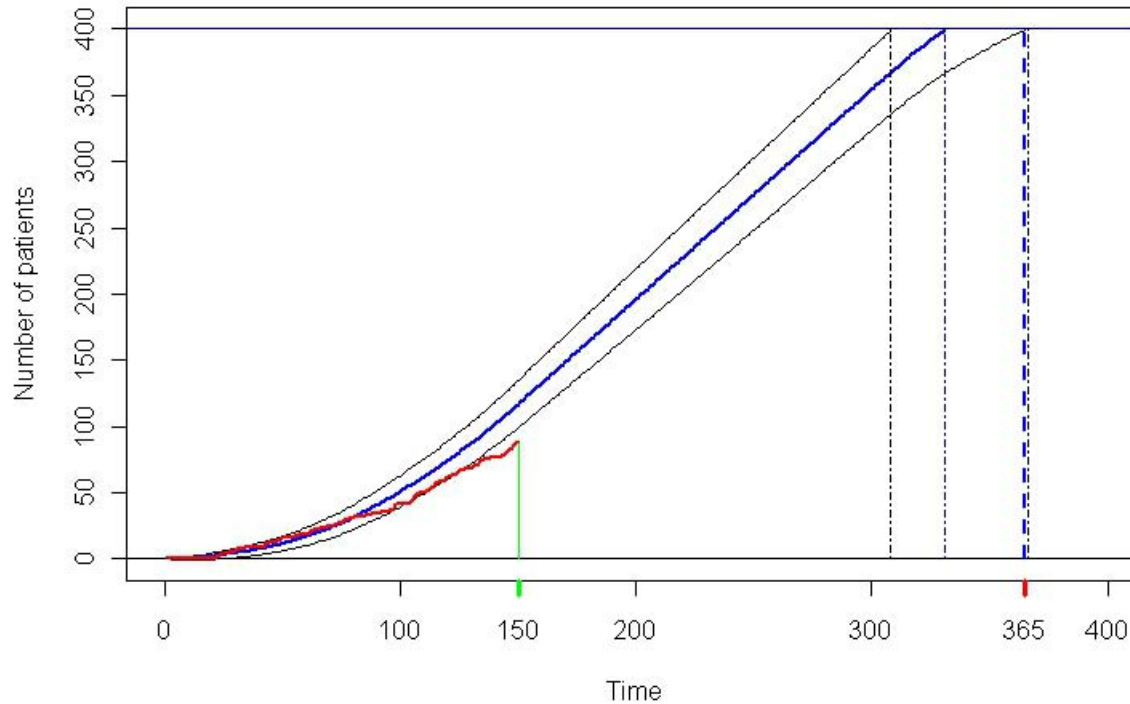
Initial plan: to complete with 90% confidence:

Predictive area: mean and 90% credibility bounds.

Optimal enrolment predicting



Adaptive enrollment adjustment



Scenario:

400 pts,
70 sites,
time = 365 days.
70 sites initiated in
5-month period,
half of sites will be
closed during two
months before the
end of enrollment.

Interim analysis after 150 days:

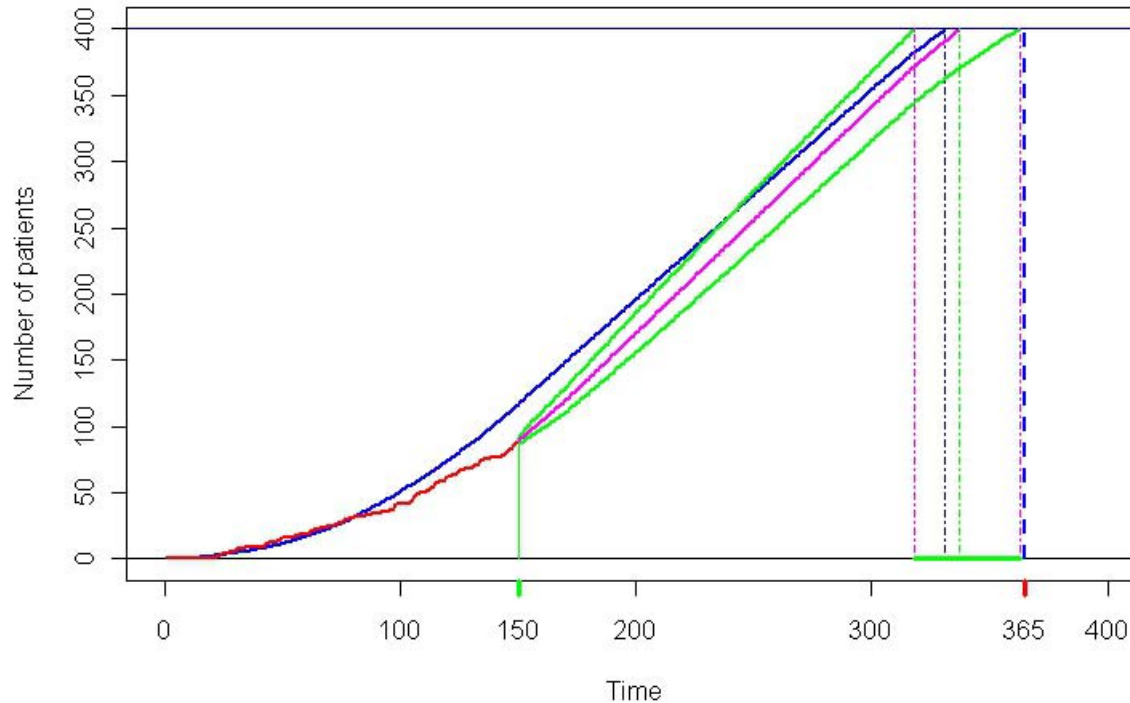
88 pts recruited

Study going slower than predicted

Optimal enrolment predicting



Adaptive enrollment adjustment



Scenario:

400 pts,
70 sites,
time = 365 days.
70 sites initiated in
5-month period,
half of sites will be
closed during two
months before the
end of enrollment.

Interim adjustment:

to complete with 90% confidence: 22 new sites need to be added.
Adjusted prediction is going on time.

Modelling operational characteristics

Methodology for modelling patient enrolment forms the basis for further developments. The next level of hierarchy:

Predictive modelling of operational characteristics.

The technique using **evolving stochastic processes** is developed. It allows analytic dynamic modelling:

- **Patients in trial (follow-up), lost patients**
 - analogy – customers in service, unreliable servers
- **Patient's visits** - analogy – multi-stage service
- **Number of various events** (oncology, CRV)
 - analogy – service interruptions, Markov/semi-Markov type service
- **Operational costs (per visits, maintenance, supply)**

Software tools on the implementation stage

- **Predictive Analytics Team, Quintiles**

Evolving stochastic processes ¹⁾

are constructed on the top of enrolment process.

Consider in a centre i a sequence of arrival times

$$t_{1i} \leq t_{2i} \leq \dots$$

Associate with each t_{ki} a random process (independent in i and k)

$$\xi_{ki}(t), \quad 0 \leq t \leq \tau_{ki}, \quad \xi_{ki}(t) = 0, \quad t < 0, \quad \text{defined on a random interval } [0, \tau_{ki}]$$

Denote

$$Z_i(t) = \sum_k \xi_k(t - t_{ki}) \quad - \text{sum of evolving processes in centre } i,$$
$$Z(t) = \sum_i Z_i(t)$$

$\xi_{ki}(t)$ – can be process of follow-up, visits, event's, lost patients, costs, ...
e.g. some Markov (semi-Markov) absorbing process ²⁾

¹⁾ Anisimov, 1991.

²⁾ Anisimov, 2011

Hierarchical evolving process

Evolving accumulative process in one site:

$$Z(t) = \sum_{k: t_k \leq t} \xi_k(t - t_k),$$

t_k - arrival times with (random) rate $\lambda(x)$.

Denote $A(t) = \mathbf{E}[\xi(t)]$, $B^2(t) = \mathbf{E}[\xi^2(t)]$.

Assume that rate $\lambda(x) = \lambda\psi(x)$, $\psi(x)$ - deterministic function ($\psi(x) = 1, x \in [a, b], = 0, x \notin [a, b]$).

Theorem.

$$\begin{aligned} \mathbf{E}[Z(t)] &= \mathbf{E}[\lambda] \int_0^t A(t-x)\psi(x)dx \\ \mathbf{Var}[Z(t)] &= \mathbf{E}[\lambda] \int_0^t B^2(t-x)\psi(x)dx + \mathbf{Var}[\lambda] \left(\int_0^t A(t-x)\psi(x)dx \right)^2 \end{aligned}$$

Lemma: *Event's modelling.*

Let for some event A , $p_A(t) = \mathbf{P}(\xi_k(t) \in A)$.

The number of events A in $[0, t]$ has a Poisson distribution with random rate

$$\Lambda_A(t) = \int_0^t \lambda(x)p_A(t-x)dx.$$

1. Predicting follow-up patients

- Each patient stays in the trial for time interval L - follow-up period.
- k -th patient in centre i can drop-out with rate μ_{ki} (random)

$$\xi_{ki}(t) = 1, 0 \leq t \leq \tau_{ki}, \xi_{ki}(t) = 0, t > \tau_{ki}, \text{ where } \tau_{ki} = \min(\text{Ex}(\mu_{ki}), L)$$

- Analogy – customers in service, $(M_R/G_R/\infty)^N$ network with bounded service time and random rates

2. Predictive event modelling

- event-driven trials (oncology)

Each patient is followed-up until some event happens

- for one type of events

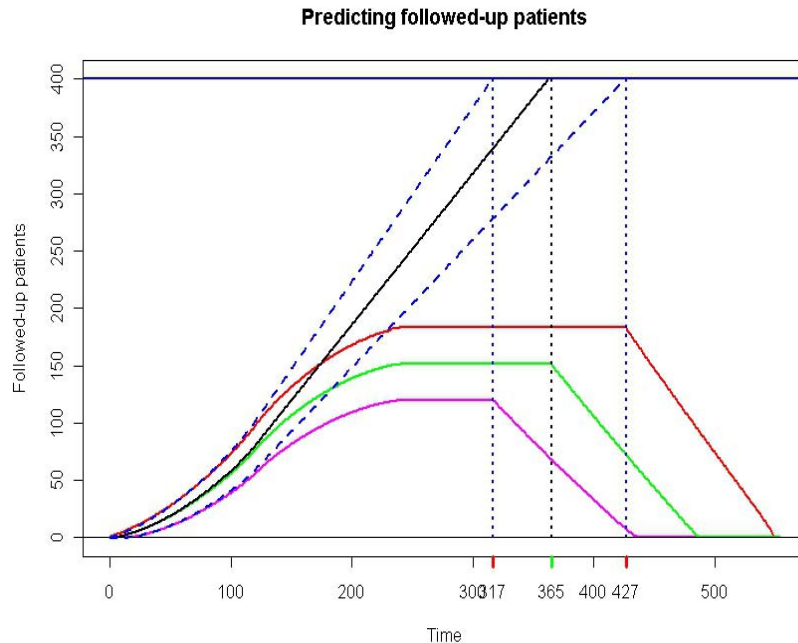
$$\xi_{ki}(t) = 0, 0 \leq t < \tau_{ki}, \xi_{ki}(t) = 1, t \geq \tau_{ki}, \tau_{ki} - \text{time to event}$$

$Z_i(t)$ – total number of events in centre i .

- for multiple events

$\xi_{ki}(t)$ – finite Markov absorbing process

Patients in trial



$L = 4$ months; rate $\mu=0.001$

Scenario: 400 pts, 80 centres, planned enrolment 12 months, site initiation 4 months. Enrolment predictive curves: black and blue lines, drop-out probability ~11%.

90% predictive interval for enrolment time: mean – 365; bounds (317,427)

90% predictive bounds for patients in trial: mean – green, bounds -magenta lines

Quasi-stationary regime in enrolment period: if some centre starts at zero, for $L < t < T$,

$$n(t) \sim n_{\text{stat}} - \text{Poisson distr. with parameter } \lambda(1 - e^{-\mu L}) / \mu$$

Patients arrive according to a Poisson-gamma model. Follow-up period L , drop-out rate μ .

Closed-form expressions for the mean and predictive bounds are derived.

Analogy – customers in service in $(M_R/G_R/\infty)^N$ network (random-effect setting)

Predictive event's modelling



Patients are recruited and follow-up until particular events happen.

- τ time till some event,

$$p(x) = \Pr(\tau \leq x).$$

Consider some centre with enrolment rate $\lambda(t)$

(rate can be random and time depend)

Lemma: The number of events in any interval $[0,t]$, generated by newly recruited patients has a Poisson distribution with cumulative rate

$$\int_0^t \lambda(u) p(t-u) du$$

- Typical enrolment scenario:
recruit in $[0,T]$, then stop and wait until n events will happen.
- If τ is exponential, the predictive mean, sd and bounds for the number of events are calculated analytically.

Event modelling – exponential time



Interim time $t_0=0$. Number of patients at risk n_0 , rate of event μ .
Denote $\hat{\Lambda}$ - global posterior rate (using data at interim time).

The total number of events in interval $[0, t]$ generated by patients at risk and by patients to be recruited:

$$Z(t) = \text{Bin}(n_0, p(t)) + \Pi_{q(t,T)} \hat{\Lambda},$$
$$q(t, T) = \begin{cases} t - \frac{1-e^{-\mu t}}{\mu} & t \leq T, \\ T - \frac{e^{-\mu(t-T)} - e^{-\mu T}}{\mu} & t > T. \end{cases}$$

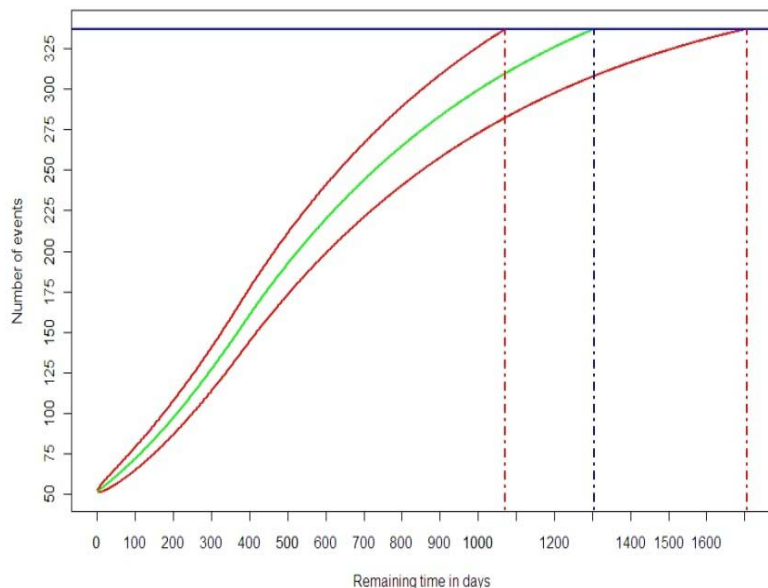
For values $A(t) = \mathbf{E}Z(t)$, $S^2(t) = \mathbf{Var}Z(t)$ – exact formulae.
 $(1 - \varepsilon)$ -predictive bounds for $Y(t)$:

$$(A(t) - z_{1-\varepsilon/2} S(t), A(t) + z_{1-\varepsilon/2} S(t)).$$

Time of study completion (to reach n events) can be numerically calculated (with mean and predictive bounds).

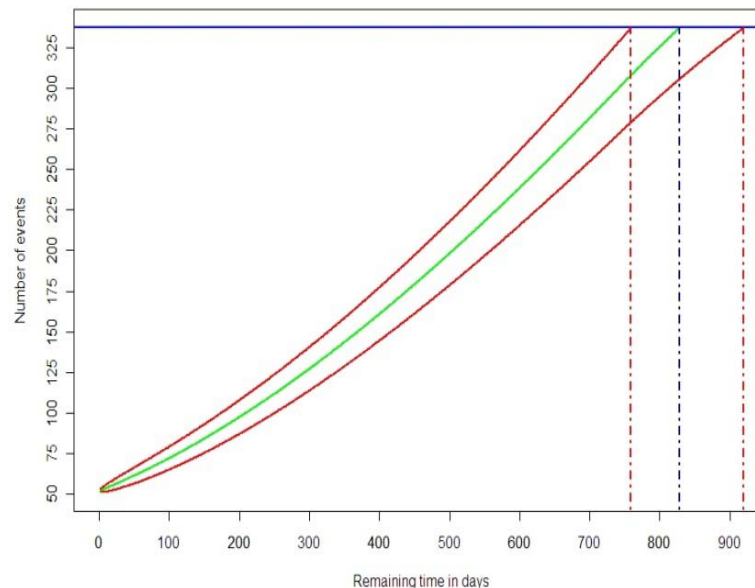
Case study

Predicting the number of events in time with 90% bounds



Sc. 1: Stopping enrolment in one year

Predicting the number of events in time with 90% bounds



Sc. 2: Stopping enrolment in two years

Data: 227 recruited pts; 131 active centres; 32 new centres to be initiated; 52 –events happened; 157 –patients at risk - without events; **Target 336 events.**

Predictive mean time and interval to reach target (in days):

Sc. 1: 1304 (1072, 1705);

Sc. 2: 757 (698, 828) - on average on 550 days earlier.

Multiple events - Markov models



Three types of events:

R - recurrence, no death; *D* - death; *L* - patient is lost to follow-up.

Consider an ongoing trial at some interim time t_0 .

Recruited patients can be divided into 6 groups:

- group *O*: k_O patients recruited and at risk (waiting for any event);
- group *R*: k_R patients with recurrence and followed-up;
- group *D1*: k_{D1} dead patients, the death was the first event;
- group *D2*: k_{D2} dead patients, the death was after recurrence;
- group *L1*: k_{L1} lost to follow-up patients before any event;
- group *L2*: k_{L2} lost to follow-up patients after recurrence.

Evolving process of events - Markov chain in a continuous time with 6 states: $\{O, R, D1, D2, L1, L2\}$.

- transition probabilities are calculated explicitly given transition rates
- transition rates are estimated in a closed form using event data and ML

Anisimov, Pharma Stats, 2011

Multiple events - Markov models

Theorem. Assume that at time $t_0 = 0$ there are N active centres, recruitment data $\{(v_i, k_i)\}$ are given and recruitment will be stopped at time T . Then

$$Z(t, D) \doteq Bin(k_O, P_{O,D}(t)) + Bin(k_R, P_{R,D2}(t)) + \Pi_{Q(t,T,D)\hat{\Lambda}}, t > 0,$$

where $\hat{\Lambda} = \sum_{i=1}^N \hat{\lambda}_i$, $P_{O,D}(t)$, $P_{R,D2}(t)$ are transition probabilities, $\mu_1 = \mu_R + \mu_{D1} + \mu_{L1}$, $\mu_2 = \mu_{D2} + \mu_{L2}$, and

$$Q(t, T, D) = \left(\mu_{D1} + \frac{\mu_R \mu_{D2}}{\mu_2} \right) \frac{q_0(t, T, \mu_1)}{\mu_1} - \frac{\mu_R \mu_{D2}}{\mu_2} \frac{q_0(t, T, \mu_2) - q_0(t, T, \mu_1)}{\mu_2 - \mu_1},$$

$$q_0(t, T, \mu) = \min(t, T) - (e^{-\mu \max(0, t-T)} - e^{-\mu t}) / \mu.$$

Transition probabilities:

$$P_{O,D}(t) = \left(\mu_{D1} + \frac{\mu_R \mu_{D2}}{\mu_2} \right) \frac{1 - e^{-\mu_1 t}}{\mu_1} - \frac{\mu_R \mu_{D2}}{\mu_2} \frac{e^{-\mu_1 t} - e^{-\mu_2 t}}{\mu_2 - \mu_1}$$

For values $E [Z(t,D)]$ and $\text{Var} [Z(t,D)]$ – closed-form expressions, predictive bounds evaluated using normal approximation.

There are some similarities and differences in modelling clinical trials and queueing processes.

- ✓ Some models are similar but the goals may be different.
- ✓ In clinical trials appear special problems
 - using input flows and service with random parameters
 - predicting cumulative flows, random stopping times, event analysis,...
 - using variable number of nodes, estimation and interim adjustment,
 - predicting performance of centres, cost/risk analysis, ...

Predictive analytic techniques for modelling enrolment and associated evolving hierarchic processes in clinical trials are developed.

Implementation of modelling tools led to substantial improvement of the quality of prediction and significant benefits and cost savings.

We are not aware about similar tools used by other companies and CRO.

Some models:

- mixed-Poisson processes, hierarchic evolving processes, optimal problems..

open wide opportunities for use in modern queueing networks and for MS and PhD students.

1. Anisimov V., Limit theorems for evolving accumulation processes, *Theor. Probability and Math. Statist.*, 43, 1991, 5-11.
2. Anisimov V., Fedorov V., Modeling, prediction and adaptive adjustment of recruitment in multicentre trials, *Statistics in Medicine*, 26, 2007, 4958–4975.
3. Anisimov, V., *Switching Processes in Queueing Models*. John Wiley & Sons, London, 2008.
4. Anisimov V., Switching queueing networks, in “Next Generation Internet”, D. Kouvatsos (Ed.), *Lecture Notes in Computer Sciences 5233*, Springer-Verlag, Berlin Heidelberg, 2011, 258-283.
5. Anisimov V., Effects of unstratified and centre-stratified randomization in multicentre clinical trials, *Pharmaceutical Statistics*, 10, iss. 1, 2011, 50-59.
6. Anisimov V., Predictive event modelling in multicentre clinical trials with waiting time to response, *Pharmaceutical Statistics*, 10, iss. 6, 2011, 517-522.
7. Anisimov V., Statistical modelling of clinical trials (recruitment and randomization), *Communications in Statistics – Theory and Methods*, 40: 19-20, 2011, 3684-3699.
8. Anisimov, V, Predictive hierarchic modelling of operational characteristics in clinical trials, in book: 7th Intern. Workshop on Simulation, May 2013, Dept Statistical Sciences, University of Bologna, Italy, 2013, 45-46.

Performance Analysis of a Hybrid Wireless Network with Batch Arrival Process

Wei Li¹ and Demetres Kouvatsos²

¹ School of Computer Science and Technology
ShanDong University
JiNan China
lindaweili@gmail.com,

² NetPEN – Networks and Performance Engineering Research Group
Informatics Research Institute (IRI), University of Bradford
Bradford, BD7 1BA, UK
D.Kouvatsos@bradford.ac.uk

Abstract. A wireless mesh network is adopted to expand the coverage of WLANs. This hybrid wireless network has risen in popularity in various applications in the literature, such as vehicle communication technology, social networks and multimedia transmissions. However, more investigation is required towards the analytic modelling and prediction of fundamental performance measures, such as mean queue length, blocking probability and waiting time distribution. This paper provides a cost effective analytical model for a hybrid wireless mesh network, based on the evaluation of a finite capacity queue with batch arrival process (BAP) for an accurate performance prediction of the proposed hybrid wireless network.

Keywords: Wireless Mesh Networks (WMNs), Batch Arrival Process (BAP), performance prediction, queueing models, Wireless Local Area Networks (WLANs)

1 Introduction

Wireless mesh networks (WMNs) are dynamically self-organized and self-configured, with the nodes in the network automatically establishing an ad hoc network and maintaining the mesh connectivity. WMNs are widely used to expand the coverage of high-speed WLANs and are developed to offer rapid development and easy reconfiguration of wireless broadband communications for mobile users anytime and anywhere at low cost [1–3]. The hybrid WMNs integrating with WLANs have risen in popularity in various applications such as vehicle communication technology, social network and multimedia transmission.

A typical WMN consists of mesh routers (MRs) forming the backbone of the network. MRs have minimal mobility. Each MR can be considered as an

access point serving a number of users or mesh clients (MCs). The MCs could be mobile users, stationary workstations or laptops that exchange data over the Internet. They direct their traffic to their respective MRs, which then forwards it over the backbones, in a multi-hop manner, to reach the gateway that links to the Internet [4]. MRs that facilitated with the gateway functionality are called gateways [5].

Performance evaluation studies on WLANs and WMNs have been widely reported in the literature [6–8]. A queueing model was developed in [6] to investigate the performance metrics of a random access WMN. Garetto etc.[7] developed an analytical model to calculate the throughput of individual traffic flows in multi-hop wireless networks and further identified the dominating and starving flows in the network. Min etc. [8] analysed the performance of integrated WLANs and Internet access mesh networks by using M/M/1/N queueing model. Niyato and Hossain [9] presented a bandwidth management and admission control framework for the integrated WLANs and multi-hop infrastructure mesh networks. However, most of the analytical models assume interarrival time between packets is exponential distributed, i.e., most researchers use single arrival queueing model. Actually, the packets from different resources may arrive at APs or MRs simultaneously. In other words, the batch arrival queueing model describes the arrival process at APs and MRs more precisely.

The main contributions of this paper are:

- A new analytical model is developed to investigate the performance metrics, such as queue length distribution, mean queue length, blocking probability and waiting time distribution for the hybrid WMNs integrating with WLANs.
- The batch arrival queueing model is adopted to analyse the performance at APs and MRs, and the expressions for the performance measures are derived.

This paper is organized as follows. The Hybrid WMN is described and illustrated in section 2. The performance metrics at MCs, APs and MRs are analysed and derived in section 3. Conclusions and future work are presented in section 4.

2 The Hybrid Wireless Mesh Networks

The hybrid WMNs considered in this paper is illustrated by Fig. 1, which includes multiple WLANs and a WMN. The WMN acts as the backhaul network which provides Internet service for MCs through one gateway. The dashed lines indicate the possible transmission path among the entities.

The packets arrive at MCs are generated by the client terminals. Normally, Each AP accepts packets from more than one MCs and forwards them to the gateway in multi-hop manner. Each MR forwards packets relayed by APs and other MRs. It is possible for APs and MRs that packets from different resource arrive to the entities simultaneously. Due to the highly connected feature of WMNs [1], the shortest path routing scheme is used to relay packets in the integrated network.

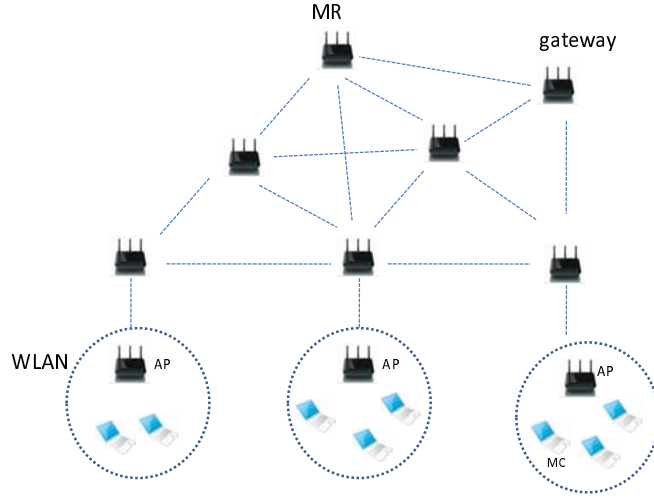


Fig. 1. The Architecture of Hybrid WMNs Integrating with WLANs

3 The analytical model

In the analytical model, the arrival process at MC is considered as Poisson process. To predict the performance metrics at MCs, the queueing model $M/M/1/N$ is adopted. Each AP is connected by a few MCs, the arriving packets from different MCs may arrive the AP at the same time. It is reasonable to treat the arrival process at APs as batch Poisson process, same as that at MRs. The queueing model with batch arrival process $M^{Geo}/M/1/N$ is used to analyse the performance measures at APs and MRs.

3.1 Packets arrival process at MCs

Let λ_c denote the packet generation rate at each MC. Due to the finite capacity of the queue at MC, it is possible the packet is lost with probability π_N^c when the queue is full. The effective arrival rate to the queue is as follows [8]

$$\lambda_c^e = \lambda_c * (1 - \pi_N^c) \quad (1)$$

The blocking probability π_N^c can be obtained by the queueing theory of $M/M/1/N_c$ (N_c is the buffer size of the queue in MC) [10], and can be expressed as

$$\pi_N^c = \frac{(1 - \lambda_c/\mu_c) * (\lambda_c/\mu_c)^{N_c}}{1 - (\lambda_c/\mu_c)^{N_c+1}} \quad (2)$$

where μ_c is the service rate at each MC.

3.2 Packets arrival process at APs and MRs

The packets leaving from MC need to access the associated AP. Each AP accepts packets from multiple MCs in one WLAN, therefore packets from different MCs are possible to arrive AP at the same time. Packets arrive at one AP simultaneously forms a batch. Assume the number of packets in one batch is geometrically distributed. The batch arrival process is described by two component distributions - batch interarrival time and batch size distributions, defined as

$$a_A(t) = \lambda_A e^{-\lambda_A t}, t > 0 \quad (3)$$

$$b_A(n) = (1 - \sigma_A)^{n-1} \sigma_A, n = 1, 2, \dots \quad (4)$$

where λ_A and σ_A are defined as batch arrival rate and batch size geometric parameter at APs respectively.

The packet arrival processes at MRs have similarity as at APs. The arriving batch is comprised of packets from APs and other MRs. The arrival processes at MRs, therefore, is also considered as batch Poisson arrival processes. The batch arrival process at APs can be expressed by equations as follows

$$a_R(t) = \lambda_R e^{-\lambda_R t}, t > 0 \quad (5)$$

$$b_R(n) = (1 - \sigma_R)^{n-1} \sigma_R, n = 1, 2, \dots \quad (6)$$

where λ_R and σ_R are defined as batch arrival rate and batch size geometric parameter at MRs respectively.

The blocking probability at APs can be calculated by the analytical results of queueing model $M^{Geo}/M/1/N$, which is a specific case of the more general queueing model $M^{Geo}/M^{Geo}/1/N$ [11] by setting the service batch generalisation parameter $\tau = 1$. The service utilisation is defined as $\rho_A = \sigma_A \lambda_A / \mu_A$.

$$\pi_N^A = \frac{1}{Z_{N_A}} \cdot \frac{\mu_A - \lambda_A}{\mu_A} \cdot (1 - \sigma_A + \rho_A)^{N_A} \quad (7)$$

where Z_{N_A} is the normalisation factor, and can be determined by the following equation.

$$Z_{N_A} = 1 - \frac{\lambda_A}{\mu_A} \cdot (1 - \sigma_A + \rho_A)^{N_A} \quad (8)$$

The blocking probability at MRs, π_N^R can be computed according to (7) and (8) by substituting the superscript A by R . And μ_A is the service rate at APs.

3.3 Packets service rate at MCs, APs and MRs

The service processes at MCs, APs and MRs are all considered as Poisson process, i.e., the service time at each node is exponential distributed. According to Min [8], service rate at MC can be calculated approximately by the following equation

$$\mu_c \approx \frac{\ln(1 - P_c)^{-1}}{t_l} \quad (9)$$

where P_c denotes the successful channel access probability for MC. The probability of AP's and MR's successive channel access can be obtained by substituting the subscript by A and R respectively, i.e., P_A and P_R . The service rates at APs and MRs can be computed according to equation (9). Each packet is transmitted within time of length t_l in a WLAN.

3.4 Performance metrics at MCs

The average number of packets, L_c , in the queue at MC can be calculated according to the queueing theory of M/M/1/N finite capacity queue [12] as

$$L_c = \frac{\lambda/\mu_c}{1 - \lambda/\mu_c} - \frac{\lambda/\mu_c * [N_c * (\lambda/\mu_c)^{N_c} + 1]}{1 - (\lambda/\mu_c)^{N_c+1}} \quad (10)$$

The total delay D_c that a packet spend at MC is equal to the queueing time plus the service time and given by the following equation.

$$D_c = L_c/\lambda_c^e + 1/\mu_c \quad (11)$$

3.5 Performance metrics at APs and MRs

The queue length distribution $P_A^N(n)$ can be calculated by the queueing theory of $M^{Geo}/M^{Geo}/1/N$ according to the mean queue length expression in [11] by specifying $\tau = 1$ as equation (12).

$$P_A^N(n) = \begin{cases} \frac{1 - \rho_A}{1 - \rho_A(1 - \sigma_A(1 - \rho_A))^N} & n = 0 \\ \sigma_A \rho_A \cdot (1 - \sigma_A(1 - \rho_A))^{n-1} P_A^N(0) & n = 1, 2, \dots, N-1 \\ \sigma_A \rho_A \cdot (1 - \sigma_A(1 - \rho_A))^{N-1} P_A^N(0) & n = N \end{cases} \quad (12)$$

The mean queue length at APs, L_A can be calculated as

$$L_A = \frac{\sigma_A \lambda_A (\mu_A - \lambda_A)}{Z_{N_A} \cdot \mu_A^2} \cdot \frac{1 - (N_A + 1)\rho_A^{N_A-1} + N_A \rho_A^{N_A}}{(1 - \rho_A)^2} + \frac{1}{Z_{N_A}} \cdot \frac{\mu_A - \lambda_A}{\mu_A} \cdot \rho_A \cdot N_A \cdot (\rho_A + 1 - \sigma_A)^{N_A-1} \quad (13)$$

The waiting time distribution of $M^{Geo}/M/1/N$ is given by

$$W_A^N(t) = \frac{\sum_{k=0}^{N-1} \sum_{r=i}^{\infty} \sum_{i=1}^{N-k} P_A^N(k) b_A(r) \int_0^{\infty} \frac{e^{-\lambda_A t} (\lambda_A t)^{k+i-1}}{k!} dt}{(1 - \sigma_A)(1 - \pi_N^A)} \quad (14)$$

4 Conclusions

The hybrid wireless networks integrating WLANs have risen in popularity in various applications such as vehicle communication technology, social network and multimedia transmission. This paper provides a cost effective analytical model for hybrid wireless mesh network with finite capacity queueing model with batch arrival process to predict accurately the performance of the integrated wireless network, such as queue length distribution, mean queue length, blocking probability and waiting time distribution for MCs, APs and MRs respectively. The analytical tool can be used to investigate the performance of hybrid WMNs integrating with WLANs.

The hybrid WMN can be analysed by more complex queueing models, such as batch renewal arrival process which is able to capture the bursty and correlation features in the wireless network. Further work will focus on extending the queueing model to more general ones which will capture the traffic correlation feature.

References

1. I.F. Akyildiz and X. Wang, A Survey on Wireless Mesh Networks, *IEEE Communications Magazine*, Vol. 43, No. 9, pp. S23-S30, 2005.
2. M. Peltomaki, J. Koljonen, O. Tirkkonen and M. Alava, Algorithms for self-organized resource allocation in wireless network, *IEEE Trans. Veh. Technol.*, Vol. 61, No. 1, pp. 346-359, 2012.
3. Yulei Wu, Geyong Min and Laurence T. Yang, Performance Analysis of Hybrid Wireless Networks Under Bursty and Correlated Traffic, *IEEE Trans. on Vehicular Technology*, Vol. 62, No. 1, pp. 449-454, Jan. 2013.
4. K.R. Chowdhury and I.F. Akyildiz, Cognitive Wireless Mesh Networks with Dynamic Spectrum Access, *IEEE Journal on Selected Areas in Communications*, Vol. 26, No. 1, pp. 168-181, 2008.
5. S.M. Faccin, C. Wijting, J. Kenckt and A. Damle, Mesh WLAN Networks: Concept and System Design, *IEEE Wireless Communications*, Vol. 13, No. 2, pp. 10-17, 2006.
6. N. Bisnik and A. Abouzeid, Delay and Throughput in Random Access Wireless Mesh Networks, *Proc. of IEEE International Conference on Communications (ICC'06)*, pp. 403-408, 2006.
7. M. Gatto, Y. Salonidis and E.W. Knightly, Modelling Per-Flow Throughput and Capturing Saturation in CSMA Multi-Hop Wireless Networks, *IEEE/ACM Trans. on Networking*, Vol. 16, No. 4, pp. 864-877, 2008.
8. Geyong Min, Yulei Wu and Laurence T. Yang, Performance Modelling and Analysis of Integrated WLANs and Internet-Access Mesh Networks, 2009 International Conference on Computational Science and Engineering, pp.1-8, 2009.
9. D. Niyato and E. Hossain, Integration of IEEE 802.11 WLANs with IEEE 802.16-based Multihop Infrastructure Mesh/Relay Networks: A Game Theoretic Approach to Radio Resource Management, *IEEE Network*, Vol. 2, No. 3, pp. 6-14, 2007.
10. Leonard Kleinrock, *Queueing Systems, Volume 1: Theory*, A Wiley-Interscience Publication, John Wiley and Sons, Canada, 1975.
11. Wei Li, R.J.Fretwell and D.D.Kouvatsos, Performance Analysis of Queues with Batch Poisson Arrival and service, 2011 IEEE 13th International Conference on Communication Technology (ICCT), pp.1033-1036, Sept. 2011.

12. D. Gross and C. Harris, Fundamentals of Queueing Theory, 3rd ed., Hoboken, NJ, Wiley, 1998.

A Chaos Logistic Map to Model the Delay Tolerant Cyber-Physical Systems

Shulin Zhao, Hongjun Dai ^{*}, Wei Li, and Lin Lu

Department of Computer Science and Technology, Shandong University,
Jinan. 250101 Shandong, China.P.R.

zhaoshulin.cn@gmail.com, dahogn@sdu.edu.cn, lindaweili@gmail.com, llu@sdu.edu.cn

Abstract. Cyber-Physical Systems (CPS) is an integration of the computation processes and physical processes. A chaos logistic map feedback linearization method to model the delay tolerance in CPS has been investigated. The motivation of this generalization is the fact that chaotic CPS is very sensitive to change in plant model parameters and even initial values such as hardware instantaneously and chaos logistic map decision delay. Because of the closed loop stability, simulations results and mathematical equations show that this method is feasible to stabilize a fixed point, tracking and also synchronization of two chaotic CPS. The extended model can not only support hierarchical CPS modeling, but also support delay tolerance modeling. Being simple and easy to implement make this method more valuable.

Keywords: Cyber-Physical systems, Hybrid automata model, Delay modeling, Chaos control, Synchronization, Fuzzy feedback control

1 Introduction

Cyber-Physical System (CPS) is an integration of computation with physical systems and physical process, which emphasizes the interaction and cooperation of information part and physical part. CPS is recognized as a hybrid system that contains continuous physical dynamic systems and discrete computing systems [1]. CPS has widely been adopted in various critical areas such as distributed energy systems, transportation systems, healthcare monitoring systems, transportation systems and etc, which demonstrates that CPS has been considered as an more and more important trend of information technology which will deeply affect our daily life [2]. Considering the fact that the CPS application areas are often extensively related and of large scale, therefore, reasonable modeling for CPS becomes a key challenge during the CPS development. However, CPS is one of the hybrid systems, whose development and evolution not only depends on the response of discrete transient events, but also the response of the dynamic behavior represented by differential and difference equation [3, 4]. The hybrid automata model solves the problem as automatic verification of real-time indicator

^{*} Hongjun Dai is the corresponding author.

of the finite state machine, which implies that the hybrid automata model is a good abstract model for representation of CPS [5]. However, the typical hybrid automata models are not good at modeling the delay tolerance[6].

Many different types of attractors co-exist in the phase space [7, 8]. In such systems, for example, CPS, we often observe the phenomenon of the transient chaos. Chaotic systems are very sensitive to noises and their dynamic evolution may deviate far from the nominal one even though they are perturbed by a small noise[9, 10]. This fact makes it difficult to distinguish chaotic evolutions from stochastic processes due to the seemingly randomness demonstrated by chaotic processes although the chaotic system is indeed deterministic. Resulting from such a super sensitivity of the chaotic system to noises including the modeling error, the most convenient and easy way to control chaos of CPS is feedback linearization methods which utilize the system's feedback to the system; the key problem is in the sense that a detailed knowledge of the system is required. However, for a chaotic CPS, due to its specific features which make the system so sensitive to small changes in system parameters or initial values (in this case, delay), the difficulty of using feedback linearization methods increases [11, 12]. Therefore how to design a simple feedback linearization controller with the little information of chaotic systems is an open problem.

The new characteristics of chaotic CPS raise many challenges for modeling when using the traditional hybrid automata [13]. From the modeling perspective of chaotic CPS, this paper analyzes the insufficient of transitional hybrid automata, and extends the hybrid model with the methods of fuzzy control and synchronization of chaotic CPS. In order to model the new characteristics of chaotic CPS, we extend the traditional hybrid automata model focusing on three core methods: Entity, Mode and Control of chaotic systems using fuzzy feedback linearization. It is noteworthy that extracting the chaotic model of CPS compensate our poor knowledge of mathematical model and changes in system parameters. First, two chaotic CPS models, *Entity* and *Mode*, have been formulated in order to fit the new characteristics of the chaotic CPS. Second, we formulate centers' average defuzzifier equation to demonstrate that designing the chaotic CPS is equivalent to determining the parameters $\bar{y}^l, \bar{x}_i^l, \bar{\sigma}_i^l$. Third, the proposed fuzzy model is updated by back propagation learning algorithm to identify the CPS model. Consequently, this designed fuzzy model and a state feedback controller is used to control chaotic CPS to the unstable fixed point. Afterwards, because of the closed loop stability, simulations results and mathematical equations show that this method is feasible to stabilize a fixed point, tracking and also synchronization of two chaotic CPS, and the extended model can not only support hierarchical CPS modeling, but also support delay tolerance modeling.

2 Related Work

Because the CPS is not a mature theoretical system yet, most of the researches on the delay tolerance modeling for CPS is based on existing hybrid system the-

ory, which are summarized as follows. [14] presents a real-time hybrid structural testing as a grand challenge for CPS, and indicates the motivation and vision of Cyber-physical Instrument for Real-time hybrid Structural Testing (CIRST). [15] proposes an event-driven monitoring framework for CPS based on the hybrid automata formalism. The framework combines model-based design and formal analysis, aiming to form the monitoring capability early in the system development. Note that it is difficult to express the continuous behavior of a CPS using an Ordinary Differential Equation (ODE) because it can have unknown, unpredictable variables. Consequently, it is necessary to predict the next event time of the model by inference to embed the model in a DEVS model. [16] proposes the simulation framework in which a fuzzy inference module is added to each simulation model in order to determine its next event time. The proposed method enables simulation of hybrid system models to decide which can be represented using an ODE.

As mentioned above, how to design a simple feedback linearization controller with a little information of chaotic CPS is an open problem. To overcome this difficulty several methods have been investigated such as adaptive methods like [17] in which the feedback linearization technique is used in an adaptive manner. The global uniform boundedness of parameter estimation errors and the Lyapunov stability of tracking errors are proved by the Lyapunov stability theory and the LaSalle-Yoshizawa theorem. [18] used an adaptation law based on a Lyapunov function that aims to drive feedback linearization control parameters towards their ideal values. [19] also presented the method of using neural networks and other intelligent control systems to extract the model parameters or dynamics to feedback to the system.

Based on the theory of hybrid Petri nets, researchers have conducted some works related to the CPS. Research in chaotic CPS is relatively new, and it does not have a unified theoretical model for delay tolerance. In this paper, from the point of view of chaotic CPS delay tolerance modeling, we extend traditional hybrid automata and presents a detailed analysis of their semantics.

3 Formalization of the Chaotic CPS model

In order to model the new characteristics of chaotic CPS, we extend the traditional hybrid automata model. The core concepts of it are Entity and Mode as well as Control of chaotic systems using fuzzy feedback linearization.

3.1 Formalization definition of Chaotic CPS model

Definition 1 $Entity=(Input, Output, Init, TM)$

Input describes the set of input ports, and output describes the set of output ports. 'I' represents the initial state. TM represents the mode set of the Entity and these modes definite the dynamic behaviors of the entity. According to the characteristics of the chaotic CPS, the entity is divided into: computation entity, physical entity and interactive entity, as Figure 1 shows.

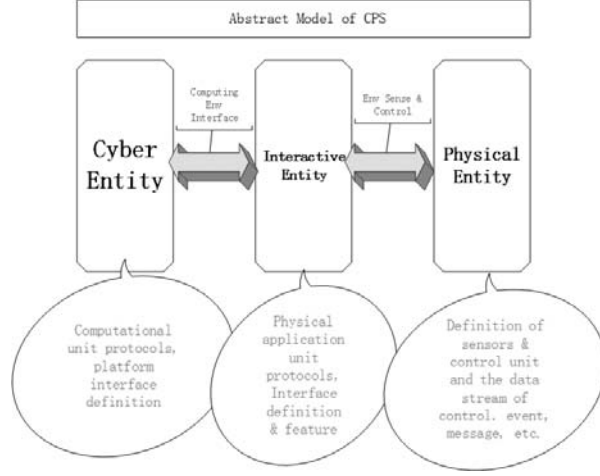


Fig. 1. The simple abstract model of CPS contained C, P and I entities.

Definition 2 $Mode=(Input, Output, S, S0, E, Act)$

Input and Output represent the Input/Output ports, S represents the state set, S0 represents the initial state, E represents the transition edges, and each side has three tuples $(S_i, Guard, S_{i+1})$, Act maps on each state S with differential algebraic equations Fs which represents the relationships of the system continuous state variable x , the state variable differential \dot{x} and continuous variable w :

$$F_s(x, \dot{x}, w) = 0 \quad (1)$$

3.2 The trace and composition of the mode

The mode state is a binary pair $s = (q, x)$, and $q \in Q$ is a discrete state, meets the position invariant $Inv(q)$, x is continuous variable value vector. Given two state $s1 = (q1, x1)$ and $s2 = (q2, x2)$, if there is transition $e \in E : q1 \rightarrow q2$, and the values of $x1, x2$ must meet the guard conditions or reset conditions, we say tuples $(s1, s2)$ for a Jump behavior, and state $s2$ is jump successor of the state $s1$. We call $\sigma = (s0, s1, s2...)$ the trajectory of the mode:

- Arbitrary $s1$ is legal state.
- Arbitrary tuple (s_i, s_{i+1}) is a jump.

Given two modes: $Mode_1 = (I_1, O_1, Init_1, S_1, E_1, Act_1)$
and $Mode_2 = (I_2, O_2, Init_2, S_2, E_2, Act_2)$,
the composition of two modes is $Mode = (I, O, Init, S, E, Act)$:

- $I = I_1 \times I_2$

- $O = O_1 \times O_2$
- $Init = \{(q1, q2), (x1, x2) \mid (q1, x1) \in Init1 \wedge (q2, x2) \in Init2\}$
- $Act = Act_1 \cup Act_2$
- $E : (S_i, Guard, S_{i+1}) \quad S_i, S_{i+1} \in S$

The composition of $Mode_1, Mode_2, \dots, Mode_m, (m > 2)$, synthesis of composite Mode according to the following way definition:

Definition 3 $Mode = Mode_1 \parallel Mode_2 \parallel \dots \parallel Mode_m$

3.3 Chaos control based on fuzzy model design on CPS

Here we assume a chaotic CPS with product inference engine, singleton fuzzifier, centers' average defuzzifier, and Gaussian membership function, given by Equation 2 :

$$f(x) = \frac{\sum_{i=1}^M \bar{y}^l \left[\prod_{i=1}^n \exp\left(-\left(\frac{x_i - \bar{x}_i^l}{\bar{\sigma}_i^l}\right)^2\right) \right]}{\sum_{i=1}^M \left[\prod_{i=1}^n \exp\left(-\left(\frac{x_i - \bar{x}_i^l}{\bar{\sigma}_i^l}\right)^2\right) \right]} \quad (2)$$

Where $\bar{y}^l, \bar{x}_i^l, \bar{\sigma}_i^l$ are free parameters that refer to the mean of output membership functions, mean of input membership function and variance of input membership functions. M is the number of rules that is fixed and not changes during update. N is the number of inputs that should be determined first, equal to our system dimension. By increasing the membership functions and the rules we access more exact model and less error if it is needed, but it makes our calculations more complicated and also more memory is required.

Designing the chaotic CPS is now equivalent to determining the parameters $\bar{y}^l, \bar{x}_i^l, \bar{\sigma}_i^l$. To determine these parameters in some optimal fashion we use the back-propagation learning algorithm and update each parameter against the corresponding error gradient by every input-output pairs.

Assume P is a free parameter, and then the update rule for P would be

$$P(q+1) = P(q) - \alpha \frac{\partial e}{\partial P} \quad (3)$$

While α is a positive constant and e is squared error of chaotic CPS model ($\hat{f}(x, t)$) output and real $f(x, t)$.

$$e = \frac{1}{2} \left[\hat{f}(x, t) - f(x, t) \right]^2 \quad (4)$$

So the update rules for $\bar{y}^l, \bar{x}_i^l, \bar{\sigma}_i^l$ results in:

$$\bar{y}^l(q+1) = \bar{y}^l(q) - \alpha \frac{\hat{f} - f^p}{b} z^l \quad (5)$$

$$\bar{x}_i^l(q+1) = \bar{x}_i^l(q) - \alpha \frac{\hat{f} - f^p}{b} (\bar{y}^l(q) - \hat{f}) z^l \frac{2(x_i^p - \bar{x}_i^l)}{(\sigma_i^l)^2} \quad (6)$$

$$\bar{\sigma}_i^l(q+1) = \bar{\sigma}_i^l(q) - \alpha \frac{\hat{f} - f^p}{b} (\bar{y}^l(q) - \hat{f}) z^l \frac{2(x_i^p - \bar{x}_i^l)^2}{(\sigma_i^l(q))^3} \quad (7)$$

While $q = 0, 1, 2, \dots, L = 1, 2, 3, \dots, M$ is the number of rule that is updated, (x^p, f^p) is the p th input-output pair for learning, and:

$$z^l = \prod_{i=1}^n \exp\left(-\left(\frac{x_i - \bar{x}_i^l}{\sigma_i^l}\right)^2\right) \quad (8)$$

$$b = \sum_{l=1}^M z^l \quad (9)$$

$$a = \sum_{l=1}^M \bar{y}^l z^l \quad (10)$$

As mentioned above, the Mode defines the entity's behavior, and the compositions of many modes define the evolution of dynamic behavior between the entities. Consequently, Therefore, by choosing the enough membership functions for inputs and enough rules defined in the Mode, we can access the desired minimum error.

4 Simulation Results

4.1 Tracking problem

As mentioned above, $\bar{y}^l, \bar{x}_i^l, \bar{\sigma}_i^l$ are free parameters that refer to the mean of output membership functions, mean of input membership function and variance of input membership functions. In this study, \bar{y}^l is defined as synthesis performance of chaotic CPS labeled as x_1 , \bar{x}_i^l is defined as hardware instantaneity labeled as x_2 , $\bar{\sigma}_i^l$ is defined as chaos logistic map decision delay labeled as x_3 . Most of the time Chaos occurrence in CPS is not desired and it is important that systems continue tracking of objective trajectory even if changes in any model parameter or disturbance makes system chaotic. We will extend the proposed method to be able to follow a desired trajectory.

Let x_r be an expectance objective. We choose the $u(t)$ as:

$$u(t) = -\hat{f}(x, t) - k(x - x_r) + \dot{x}_r \quad (11)$$

The closed loop dynamic would be:

$$(\dot{x} - \dot{x}_r) + k(x - x_r) = f(x, t) - \hat{f}(x, t) \quad (12)$$

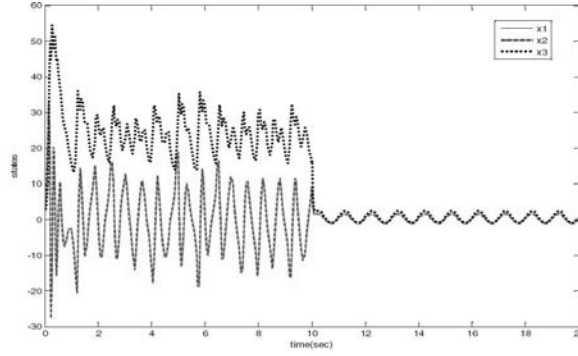


Fig. 2. Tracking a desired trajectory for chaotic CPS, the controller is applied after 10s.

The tracking error is defined as:

$$e = x - x_r \tag{13}$$

Substituting Equation 13 in Equation 12 the dynamics of error is:

$$\dot{e} + ke = f(x, t) - \hat{f}(x, t) \tag{14}$$

Consider $x_r = 0.25 + 2\sin(2\pi t)$ for all states, we use this model to control the chaotic CPS to track this trajectory. The state response curves of CPS is shown in Figure 2. Controller is added after 10s. From simulation results, it is obvious that using this controller can effectively control the chaotic CPS even in tracking mode.

The state control of the delay tolerance and instantaneity by the proposed method is shown in Figure 3. Figure 3 demonstrates that the controller is feasible for different systems and it's not dependent on the system model.

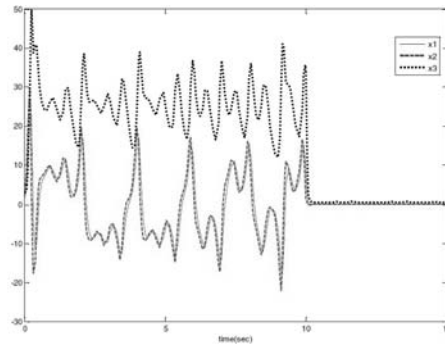


Fig. 3. Control of the delay tolerance and instantaneity, the controller is applied after 10s.

4.2 Synchronization problem

By using the idea of tracking we can extend that for synchronization of two chaotic systems. The states of master system are the desired trajectory for slave one. The drive system is $\dot{x}_m = f_m(x_m, t)$ and the response system is $\dot{x}_s = f_s(x_s, t) + u(t)$, by choosing $x_r = x_m$ and the same control $u(t)$ we can extend the method for synchronization between two chaotic systems. To show the feasibility of the method, the synchronization between two isomerism CPS is simulated. We choose a system with $\alpha = 0.3$ and initial values of $x_{1m}(0) = 3, x_{2m} = 2, x_{3m} = 3$ as a drive system, and a Lorenz system with $\alpha = 0.1$ and initial conditions $x_{1s} = 1, x_{2s} = 0, x_{3s} = 1$ as a response system.

Like before in first seconds of operation a chaotic model of response system is extracted by back propagation algorithm after $10s$, the controller is added to plant to synchronize the response system with master one. Synchronization of two chaotic CPS is shown in Figure 4. Simulation results emphasize that the control method is also effective and feasible for the synchronization.

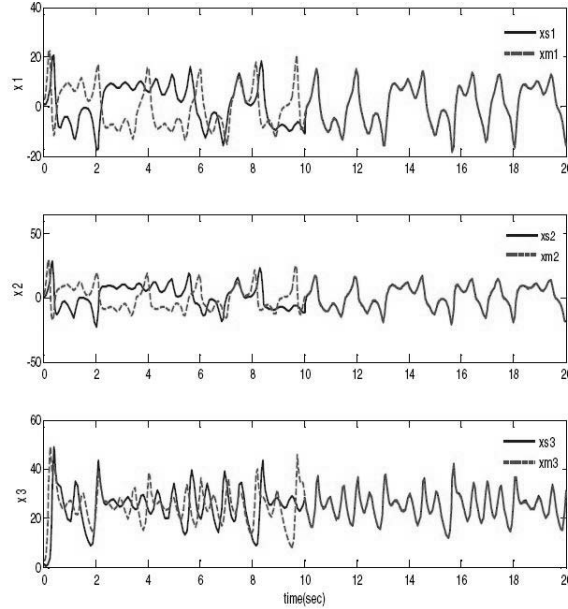


Fig. 4. Synchronization of two chaotic CPS with different parameter and initial conditions, the controller is applied after $10s$.

As demonstrated, our proposed control scheme is highly effective and does not depend on the precise model of the chaotic CPS that makes it very useful for chaotic systems as they are too sensitive to model parameters and other disturbances. From all above we can conclude that presented control can be used for different purposes and different systems.

5 Conclusion

A chaos logistic map feedback linearization method to model the delay tolerance in the CPS has been investigated. The motivation of this generalization is the fact that chaotic CPS is very sensitive to change in plant model parameters and even initial values such as hardware instantaneity and chaos logistic map decision delay. As a result the proposed method that estimates a chaotic model of CPS can be very useful for such systems. Because of the closed loop stability, simulations results and mathematical equations show that this method is feasible to stabilize a fixed point, tracking and also synchronization of two chaotic CPS. Error of system can be reduced as less as it is acceptable by increasing the membership functions and rules of fuzzy model and also by increasing the gain in controller. To sum up, the extended model can not only support hierarchical CPS modeling, but also support delay tolerance modeling. Being simple and easy to implement make this method more valuable.

Acknowledgements

This work has been partially supported by the project "Science and Technology Plan of Shandong Province, China (No.2012GB020108)" and the project "National Science Foundation of China (No. 61202147)".

References

1. Yalei Y, Xingshe Z. Cyber-Physical Systems Modeling Based on Extended Hybrid Automata[C]//Computational and Information Sciences (ICCIS), 2013 Fifth International Conference on. IEEE, 2013: 1871-1874.
2. Chen X, Chen M. Extending the Four-Variable Model for Cyber-Physical Systems[C]//Object/Component/Service-Oriented Real-Time Distributed Computing Workshops (ISORCW), 2012 15th IEEE International Symposium on. IEEE, 2012: 31-36.
3. Alur R, Courcoubetis C, Halbwachs N, et al. The algorithmic analysis of hybrid systems[J]. Theoretical computer science, 1995, 138(1): 3-34.
4. Lee E A. Cyber-physical systems-are computing foundations adequate[C]//Position Paper for NSF Workshop On Cyber-Physical Systems: Research Motivation, Techniques and Roadmap. 2006, 2.
5. Derler P, Lee E A, Sangiovanni-Vincentelli A L. Addressing modeling challenges in cyber-physical systems[R]. CALIFORNIA UNIV BERKELEY DEPT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE, 2011.
6. Kumar P, Goswami D, Chakraborty S, et al. A hybrid approach to cyber-physical systems verification[C]//Proceedings of the 49th Annual Design Automation Conference. ACM, 2012: 688-696.
7. El-Gohary A. Chaos and optimal control of cancer self-remission and tumor system steady states[J]. Chaos, Solitons & Fractals, 2008, 37(5): 1305-1316.
8. El-Gohary A. Chaos and optimal control of equilibrium states of tumor system with drug[J]. Chaos, Solitons & Fractals, 2009, 41(1): 425-435.

9. Zhou J, Wen C, Zhang Y. Adaptive output control of nonlinear systems with uncertain dead-zone nonlinearity[J]. *Automatic Control, IEEE Transactions on*, 2006, 51(3): 504-511.
10. Yau H T. Design of adaptive sliding mode controller for chaos synchronization with uncertainties[J]. *Chaos, Solitons & Fractals*, 2004, 22(2): 341-347.
11. Wang H, Han Z, Xie Q, et al. Sliding mode control for chaotic systems based on LMI[J]. *Communications in Nonlinear Science and Numerical Simulation*, 2009, 14(4): 1410-1417.
12. Salarieh H, Alasty A. Control of stochastic chaos using sliding mode method[J]. *Journal of computational and applied mathematics*, 2009, 225(1): 135-145.
13. Yadmellat P, Nikravesh S. A recursive delayed output-feedback control to stabilize chaotic systems using linear-in-parameter neural networks[J]. *Communications in Nonlinear Science and Numerical Simulation*, 2011, 16(1): 383-394.
14. Derler P, Lee E A, Vincentelli A S. Modeling CyberCPhysical Systems[J]. *Proceedings of the IEEE*, 2012, 100(1): 13-28.
15. Baheti R, Gill H. Cyber-physical systems[J]. *The Impact of Control Technology*, 2011: 161-166.
16. Moon S Y, Park H, Cho T H, et al. Hybrid System Model Simulation Framework for Cyber-Physical Systems[J]. *Applied Mechanics and Materials*, 2012, 110: 4043-4049.
17. Wang C C, Pai N S, Yau H T. Chaos control in AFM system using sliding mode control by backstepping design[J]. *Communications in Nonlinear Science and Numerical Simulation*, 2010, 15(3): 741-751.
18. Lu Z, Shieh L S, Chen G, et al. Adaptive feedback linearization control of chaotic systems via recurrent high-order neural networks[J]. *Information Sciences*, 2006, 176(16): 2337-2354.
19. Sharma B B, Kar I N. Parametric convergence and control of chaotic system using adaptive feedback linearization[J]. *Chaos, Solitons & Fractals*, 2009, 40(3): 1475-1483.

Influence of the Handoff Threshold Hysteresis on Heterogeneous Wireless Network Performance

Krzysztof Grochla¹ and Konrad Polys¹

Institute of Theoretical and Applied Informatics of Polish Academy of Sciences,
Gliwice, Poland,
[kgrochla|kpolys]@iitis.pl,
WWW home page: <http://www.iitis.pl>

Abstract. The paper provides evaluation of the influence of handover threshold hysteresis size on the network performance. The average signal level and signaling overhead is considered. The tradeoffs between the ping-pong effect and the decreased network throughput are presented. The results of the simulation of two different mobility scenarios are given.

Keywords: wireless network, handoff, hysteresis

1 Introduction and motivation

The wireless networks for mobile devices consist of multiple base stations and provide coverage over a large area. The structure of the network may be quite complex, with multiple network technologies used (e.g. WiFi, WiMAX and LTE) or different size of the cells. As the mobile clients move, the communication between them and the network is handled by different wireless point of attachment - base stations or access points. If the data or voice transmission is in progress while the client moves from the range of one access point to another, the handoff procedures are initiated to provide the client with continuous transmission. Handoff is a process of transferring an ongoing call or data session from one access point to another in wireless networks.

The handoff procedure may be initiated by the network (e.g. in LTE) or by the user (e.g. in WiFi). In both cases it is crucial for the network performance and reliability to start the procedure in the right moment. Each handoff procedure causes a short distraction of the communication and requires network signaling to put some overhead to the transmission. Starting the handoff too often would generate unnecessary overhead, but starting it too rarely may result in increasing the probability of losing the connectivity.

In this work we try to evaluate the influence of the threshold and parameters of the procedure used for the initiation of the handoff on the probability of losing the connectivity, average signal level and on the number of wrongly initiated handoff. We consider two cases: the IEEE 802.11 network and heterogeneous network consist of IEEE 802.11 access points and 3G. The handoff procedure is evaluated using extensive simulation, covering two different mobility scenarios.

2 Handoff procedure

There are two types of handoff procedure: horizontal, within the network of a single standard, and horizontal, between two networks using different wireless technology. The vertical handoff requires more signaling and is often slower (causes longer disruption in the transmission) than the horizontal.

The handoff procedure is initiated in response to some events in the network, like the loss of communication with an access point, or when specific criterias are met. In [2] the following criteria are enumerated: received signal strength, network connection time, available bandwidth, power consumption, monetary cost, security and user preferences. Typically the decision whenever the handoff procedure should be initiated is taken by monitoring a metric of the radio signal strength and quality. Metrics such as SNR (Signal to Noise Ratio), RSSI (Received Signal Strength Indication) or CINR (Carrier to Interference plus Noise Ratio). Although a number of more sophisticated algorithms for handover initiation have been proposed in the literature, based e.g. on fuzzy logic [1], the simple threshold on the signal level is easy to measure and is directly related to the service quality. Majority of existing horizontal handover algorithms use RSSI or CINR as the main decision criterion, and RSSI is an important criterion for VHD algorithms as well[2].

The signal level fluctuation may lead to initiation of several handoffs between two access points in a short time period, when both access point are capable of serving the wireless client. These unnecessary handoffs are known as the ping-pong effect[3]. As the number of handoffs increase, forced termination probability and network load also increases. Therefore, handoff techniques should avoid unnecessary handoffs. The hysteresis is a common mechanism used to minimize this effect - the handoff is initiated if the signal level from another access point is higher by more than a specified value. When the signal level drops below some level which indicates the minimum acceptable signal level the hysteresis is ignored, to maintain the network connectivity. The selection of the hysteresis size is an engineering challenge which has not yet been fully addressed. A method of selection of the hysteresis size was described by patent[6], but there is no performance analysis given. The work [7] provides a calculation of optimal threshold using simplified mathematical model. Another method is provided by [8] together with simplified numerical model. But both papers consider generic homogeneous wireless network and lack of taking into account layer 2 WiFi network model.

3 Evaluation of the influence of the threshold hysteresis

We prepared a discrete event simulation model of client mobility to evaluate the handoff efficiency. The two sample paths were generated, representing typical user path in two different environments: a suburban area with low density of buildings (path A) and an industrial zone with large buildings (path B). The paths and location of the APs and buildings are presented on the fig. 1. There are 3 WiFi APs located on path A and 5 APs on path B. The location of the

buildings was taken from the satellite photos, while the path was selected to go through places with different coverage and location with radio signal shadowed by the buildings.

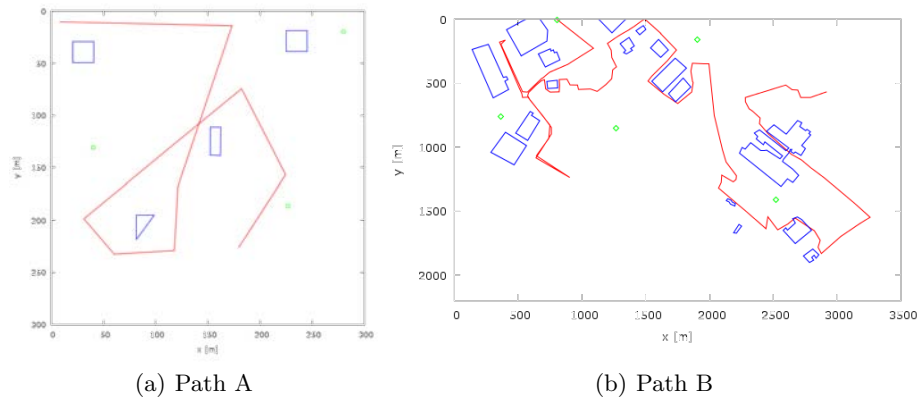


Fig. 1. Paths (red) and environment: buildings (blue) and AP location (green) used in the two simulation scenarios.

The simulation have been implemented using OMNeT++ [4] with INET framework. The lognormal shadowing model was used to represent the radio signal propagation. The default WiFi beacon interval of 100ms was used to broadcast the information about the AP. Simulation for each point was repeated 100 times with different random seeds to gather a credible average values. It was assumed that the client moves along the path with constant speed. The changes of the signal level seen by the client during the simulation are presented on the figures 2 and 3. The path A had practically full WiFi coverage, however often one or two APs were heard with very low signal level. The path B include locations which were not in range of any of the WiFi APs. It was assumed that apart of the WiFi the whole area is covered by the 3G network and the vertical handover is performed when the client goes out of the WiFi coverage.

We have used the simulations to evaluate the influence of the size of the hysteresis. The simulation were executed with hysteresis from 0 (the handoff was initiated directly when any other AP was heard by the client with higher SNE that the current one) to 25 (the handoff was initiated when another AP was heard with SNR more than 25 points higher than the current one).

The simulation showed that the average signal level over the whole simulation becomes lower than the hysteresis threshold is higher, as can be seen on 4. This is caused by the delay in the initiation of the handoff procedure, at which the client still is connected to previous AP with lower signal level. But the increase of the threshold decreases the overhead caused by unnecessary handoffs - the

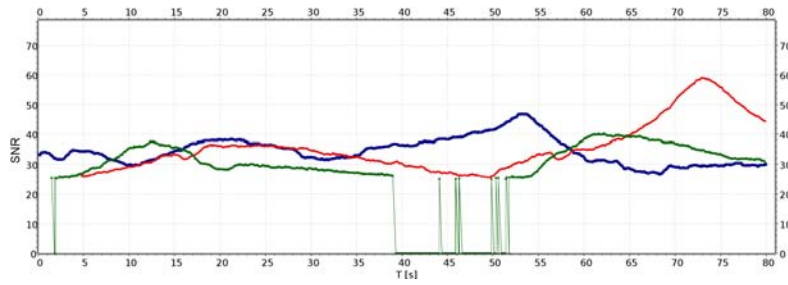


Fig. 2. Path A SNR from different APs measured over time

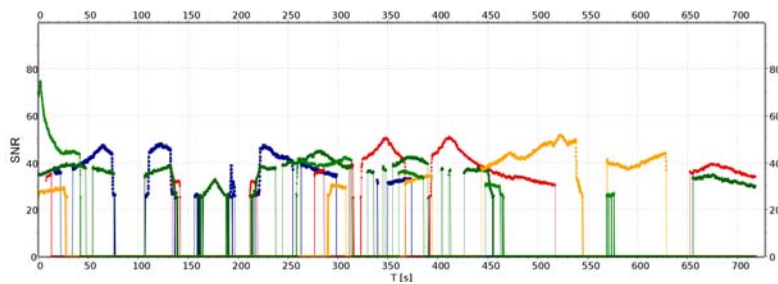


Fig. 3. Path B SNR from different APs measured over time

figure 5 shows that the number of handoffs decreases significantly, especially at the first 5-10 points.

The figure 6 presents how the threshold on signal level allows to mitigate the "ping pong" effect. The number of very fast handoffs within very short period of time (less than 2s) decreases to a constant value after introducing even a small hysteresis of size 2. On both paths we have observed some fast handoffs which were not eliminated even by a very large hysteresis, which were executed because of temporary loss of signal from an AP.

When the signal was too weak to maintain the connectivity the hard handoff had to be executed. The hysteresis increased the number of such events, as it was more probable that the client leaves the range of an AP coverage without performing the handoff to another AP. The fig. 7 shows that on path B the number of such events grows from 9 to 15 when the hysteresis is increased above 20.

4 Conclusions

The analysis performed show that hysteresis is a simple and valuable solution for triggering the handoff procedure. Adding the hysteresis allow to limit the "ping pong" effect and decreases the number of wrong handoffs.

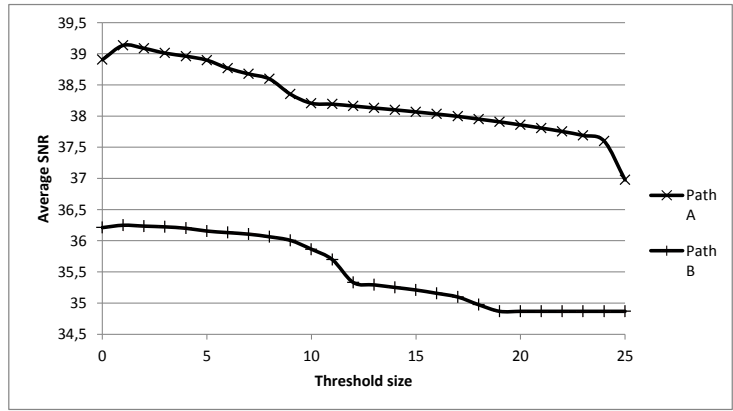


Fig. 4. Average SNR for different hysteresis thresholds

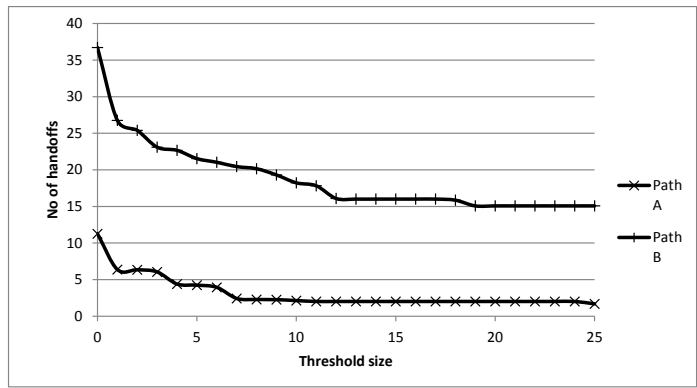


Fig. 5. Avg. no of handoffs for different hysteresis thresholds

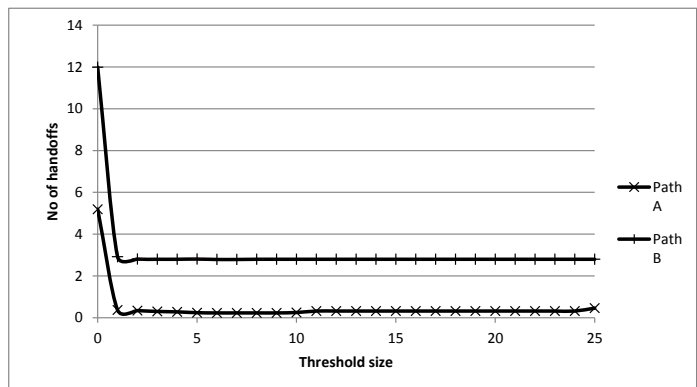


Fig. 6. Avg. no of handoffs within less than 2s after previous handoff for different hysteresis thresholds

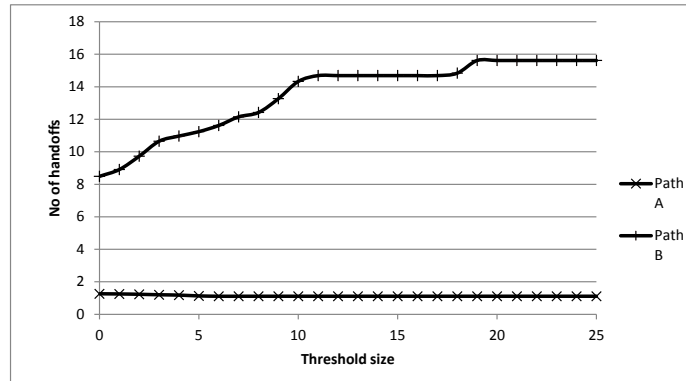


Fig. 7. Avg. no of hard handoffs to 3G for different hysteresis thresholds

5 Acknowledgement

This work has been supported by grant of Polish National Center for Research and Development LIDER/10/194/L-3/11/

References

1. Pooja Dhand and Parwinder Dhillon. Article: Handoff Optimization for Wireless and Mobile Networks using Fuzzy logic. *International Journal of Computer Applications* 63(14):31-35, Feb 2013.
2. Xiaohuan Yan, Y. Ahmet ekerciolu, Sathya Narayanan, A survey of vertical handover decision algorithms in Fourth Generation heterogeneous wireless networks, *Computer Networks*, Volume 54, Issue 11, 2 August 2010, Pages 1848-1863, ISSN 1389-1286
3. Ekiz, Nasif, et al. "An overview of handoff techniques in cellular networks." *International Journal of Information Technology* 2.3 (2005): 132-136.
4. Varga, Andrs. "The OMNeT++ Discrete Event Simulation System (<http://www.omnetpp.org>), European Simulation Multiconference (ESM2001), Prague, Czech Republic." (2001).
5. Coulson, Alan J., Allan G. Williamson, and Rodney G. Vaughan. "A statistical basis for lognormal shadowing effects in multipath fading channels." *Communications, IEEE Transactions on* 46.4 (1998): 494-502.
6. Bringby, Daniel, Per Wijk, and Tomas Lundborg. "Adaptive handoff hysteresis in wireless telecommunication networks." U.S. Patent No. 6,285,883. 4 Sep. 2001.
7. Zonoozi, Mahmood, Prem Dassanayake, and M. Faulkner. "Optimum hysteresis level, signal averaging time and handover delay." *Vehicular Technology Conference, 1997, IEEE 47th. Vol. 1. IEEE, 1997.*
8. Itoh, K-I., et al. "Performance of handoff algorithm based on distance and RSSI measurements." *Vehicular Technology, IEEE Transactions on* 51.6 (2002): 1460-1468.

Performance Evaluation of a Combined System of Random Linear Network Coding and Convolutional Code with Interleaving for Two-hop Wireless Networks under Rician Fading Channel

Misfa Susanto^{1,2}, Yim Fun Hu¹, and Prashant Pillai¹

¹University of Bradford, School of Engineering, Design & Technology, Bradford, UK
{m.susanto, y.f.hu, p.pillai}@bradford.ac.uk

²University of Lampung, Electrical Engineering Department, Bandar Lampung, Indonesia
misfa@unila.ac.id

Abstract. Errors in wireless networks are known fluctuating randomly and bursty, it even becomes worse when the information is sent through multihop scenario. And hence, the design of transmission system for multihop scenario that could promise high performance in term of throughput is crucial. In our previous paper, we have proposed a system of combined random linear network coding and convolutional code with interleaving for multihop wireless network. Our work is based on a system which is called as SRNC (Scattered Random Linear Network Coding) where a series of transmitted block of data information are grouped virtually as “protected” and “vulnerable” blocks for the purpose to achieve a high throughput performance. However, our proposed system has been not evaluated in fading channel. In this paper, we evaluate our proposed system in more realistic scenario in which wireless channel is a series combination of AWGN channel and Rician Fading Channel. AWGN channel represents a random error and Rician Fading channel represents bursty error. A simulation work has been carried out in two-hop scenario for such channel model and simulation results have been determined for the ratio of line of sight and other diffuse components in Rician Fading varied.

Keywords: Error correction code, convolutional code, interleaving, random linear network coding, two-hop transmission, rician fading channel

1 Introduction

Cognitive radio network (CRN) has been emerging as a promising technology to solve the frequency spectrum scarcity problem for wireless communication networks [1]. The ways on how data information is sent in CRN can be through centralised [2] or decentralised/ad hoc manners [2-3]. In any of those both transmission scenarios, it becomes most probably that multihop transmission cannot be avoided and it is possibly that it can be a dominant mode, especially for decentralised/ad hoc mode. For example in a centralised network, a central node/base station/access point always

relays the data information sent by the source node to the destination node, and in an ad-hoc mode an intermediate node functions as a relay node when the destination node is out of range from the source node.

It is also well known that transmission errors in wireless network fluctuating a lot due to propagation loss, terminal mobility, and physical objects surrounding the communicating parties. It faces a fact that errors in wireless channel occur randomly and bursty in nature. Moreover, the errors occurring in wireless networks will be worse when source node need to send its data information through multihop scenario to destination node. It is because of the propagation errors are accumulated hop by hop and hence throughput performance for multihop transmission will be gradually degraded. Therefore, it is important to design a transmission system which can offer high throughput performance for multihop wireless networks. As an illustration in the terms of CRN, consider the scenario in Fig.1 where two secondary mobile users, S and D, want to communicate one to another, but they are out of range each other. Hence, intermediate node which is available in between of those two nodes must be used as a relay node in order to achieve the communication goal. Secondary user node S sends three blocks of data A, B, and C to secondary user node D through the intermediate nodes P and Q which means 3-hop transmissions. Red, blue, and green bars show transmission errors occurred in hop 1, hop 2, and hop 3, respectively. As number of hop increases, more errors occur in the transmitted data blocks. Note that in Fig.1 primary system consists of primary user central node and primary mobile user nodes and only one primary central node is shown in Fig.1. It is also shown that the secondary user only relays the secondary user traffics.

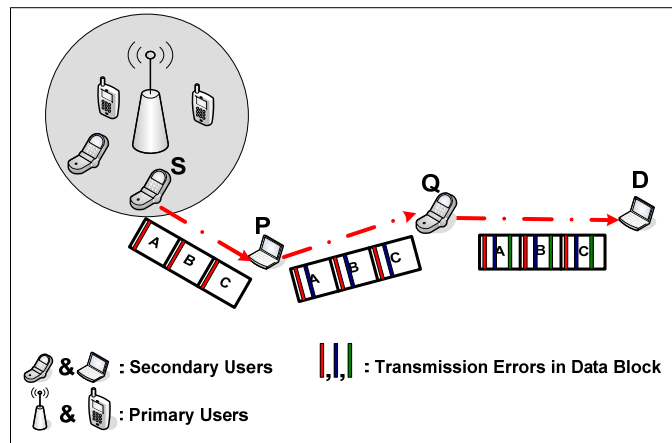


Fig. 1. A Simple example of multihop transmission scenario between secondary users in cognitive radio network. Source node S is transmitting three data blocks A, B, and C through nodes P and Q to destination node D.

On the other hand, network coding has been proposed in 2000 as an elegance technique to improve throughput performance for communication network from information theory point of view [4]. Network coding has received much attention to be

applied in both wired and wireless communication networks as in [5-8] among other references. Linear network coding [9], more specifically Random Linear Network Coding (RLNC) has been a basis to the application in communication network [5]. Seeing errors problem occurring in multihop transmission and network coding as a technique to enhance the throughput performance of wireless network, recently Scattered Random Network Coding (SRNC) has been proposed in [10] to enhance throughput performance for multihop transmission. SRNC is proposed according to the fact that for particular modulation schemes the error pattern occurring in transmission follows a certain rule [11]. SRNC takes advantages of RLNC and error pattern for 16-QAM (Quadrature Amplitude Modulation) and higher order square QAM modulation schemes. SRNC prevents sharing the same error distribution in data information blocks through bit scattering by arranging bits that are in low BER positions to be in the same blocks (protected blocks) and bits that are in higher BER positions in other blocks (vulnerable blocks). Based on SRNC, we have taken one step further to combine Random Linear Network Coding (RLNC) and convolutional code with interleaving for multihop network scenario and we have presented in [12]. We have evaluated that system in [12] under Additive White Gaussian Noise (AWGN) channel. However, our system has not been evaluated in fading channel. In this paper, we evaluate our system in [12] in more realistic that is under Rician fading Channel in series to AWGN channel. AWGN channel characterises randomness of channel errors and rician fading channel exemplifies burstiness of channel errors. We evaluated our system in two-hop transmission scenario. Note that our system can be applied in transmission scenario for as CRN as described in Fig. 1.

This paper is structured as the follows. Following this introduction, section 2 overviews the designed system model and assumptions. Section 3 describes our simulation work and discusses the simulation results. Finally, section 4 concludes this paper and gives the possible future works.

2 System Description Overview and Model

In this section, we describe an overview of our system in [12]. The system takes a step further by applying convolutional code to detect and to correct errors occurred in both high and lower BER blocks. An interleaver has been designed and implemented for the bit scattering process.

Fig. 2 shows the overall block diagram of system, which illustrates a two hop wireless transmission scenario consisting of a transmitter node, a relay node and a receiver node. In CRN, since secondary user devices can be equipped with programmable capabilities, it is possible for secondary users to adapt its modulation scheme parameters during their transmission without changing its hardware. This capability realizes reconfigurability features of cognitive radio terminals [1]. To simplify the explanation and evaluations, it is assumed that each node employs the same modulation and demodulation schemes.

At the transmitter side, the bit stream $d = [d_1, d_2, \dots, d_i]$ is first split into segments. Each segment is then divided into n fixed size blocks denoted as $\mathbf{b} = [b_1, b_2, \dots, b_n]$. A

random linear network encoding module is applied to each segment to generate random linear network coded blocks $\mathbf{c} = [c_1, c_2, \dots, c_m, \dots]$ using predefined encoding coefficients and appended with Cyclic Redundancy Check (CRC) code. The size of the encoding coefficients is also set in advance as a parameter. The random linear network coded blocks are encoded further using channel error correction code to form $\mathbf{c}' = [c'_1, c'_2, \dots, c'_m, \dots]$, which hereafter is referred to as the network-channel coded blocks. Convolution code was adopted due to the random and bursty error nature of the wireless channel. The code rate of the convolutionally encoder is carefully chosen. Moreover, the choices of the generator polynomials together with their constraint lengths to be used are also determined based on the study in [13] in which has given a list of optimal generator polynomials for different values of constraint length for code rate 1/2 and 1/3 of convolutional codes [13-14]. The output of convolutionally encoder which we call it as network-channel coded blocks are interleaved using the interleaving algorithms so as odd and even bits are placed in odd and even numbered blocks to produce the network-channel coded interleaved blocks $\mathbf{c}'' = [c''_1, c''_2, \dots, c''_m, \dots]$. These blocks are then modulated into symbols $\mathbf{s} = [s_1, s_2, \dots, s_m, \dots]$ and transmitted to the relay node over a wireless channel.

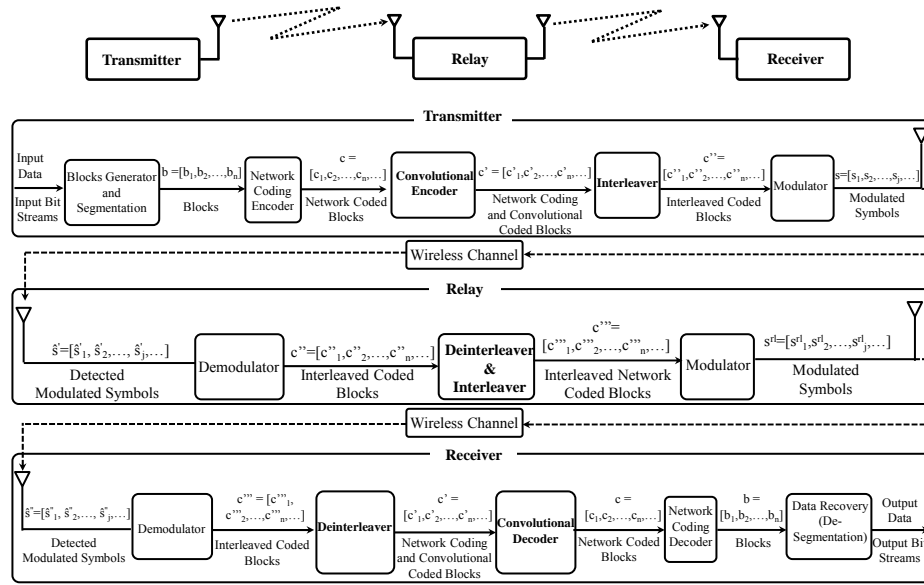


Fig. 2. Block Diagram of a System under the Study of Two-Hop Transmission Processes [12].

Wireless channel model that is applied is series of Rician fading and AWGN channels. AWGN adds noise randomly to the symbols transmitted and Rician fading gives effect to the fluctuations of amplitude and phase components in some modulation schemes such as square QAM modulation scheme. For instance, we will see the effect of Rician fading model in amplitude of transmitted symbols. Rician fading channels comes into a play when line of sight component are dominant than other diffuse (reflected wave components) components in the transmission. It is stated in a ratio of K;

ratio of line of sight and other diffuse components and it will be varied to study the effect of Rayleigh fading in the system.

Upon reception of the transmitted signal $\hat{s}' = [\hat{s}'_1, \hat{s}'_2, \dots, \hat{s}'_n \dots]$, which is now subject to channel errors, the relay node amplifies and forwards \hat{s}' to the destined node [10].

At the receiver side, the receiver detects the arrived signals and demodulates the detected symbols $\hat{s}'' = [\hat{s}''_1, \hat{s}''_2, \dots, \hat{s}''_n \dots]$ to retrieve back the network-channel coded interleaved blocks $c'' = [c''_1, c''_2, \dots, c''_n \dots]$, which are then deinterleaved to obtain $c' = [c'_1, c'_2, \dots, c'_n \dots]$, the network-channel coded blocks. Finally, the convolutional decoder following some decoding algorithms such hard or soft decision of Viterbi algorithm corrects (within its error correction capability) any error occurred to get $c = [c_1, c_2, \dots, c_n \dots]$. From this point, the random network decoder can determine the original data blocks by processing the error free blocks. As a note that any error remaining at this point can be detected using CRC (Cyclic Redundancy Check). Finally, the original data stream $d = [d_1, d_2, \dots, d_i]$ is retrieved.

3 Performance Evaluation and Results

We carried out simulation evaluation for the system described in section 2. Simulations were carried out for 16-QAM modulation with gray mapping rule under series of Rician Fading and AWGN channels in 2-hop transmission scenario using Matlab.

3×10^5 random bits of information data were generated to be input to the system. The size of information data block is set to be 8 bits which is chosen according to the study in [10] and the segment size is chosen to be 10. Size of GF (Galois Field) for RLNC process, m is set equal to the size of information data blocks ($m=8$). The number of rows for RLNC coefficient matrix, H , is set to be (2 x segment size). Elements of first column for H are generated randomly and chosen of $(2^8 - 2)$ of $GF(2^8)$ excluding elements 0 and 1, and those elements are guaranteed have to be having the different values. By setting the size of this random network coding coefficient matrix, a large enough number of coded blocks is generated to study the system behaviour. Each network coded block is appended 4 CRC bits which is generated using generator polynomial $x^4 + x + 1$ known also as CRC-ITU 4 bits [15]. A $\frac{1}{2}$ rate convolutional code with constraint length of 7 and generator polynomials $[g_1, g_2]$ of [67,163] (in octal format) were used based on a study carried out in [13]. Convolutional decoder using Viterbi algorithm with hard-decision is used at the receiver. A square interleaver of size 24 x 24 [bits] was implemented. 16-QAM modulation scheme is used with gray-mapping rule. Rician fading parameters are set with sample frequency of signal 0.1 ms, maximum dopler shift 10 Hertz, and Rician factor K varies from 0 to 10. The energy bit to noise ratio (E_b/N_0) varies from 1dB to 15 dB in steps of 1. The simulation was run for 10 times, and then taking the average value of decoding error probability collected for each simulation run. Fig. 3.(a) depicts the simulation results of decoding error probability for $K = 0 - 10$. It can be noticed that as K increases decoding error probability decreases. It is because of Line of sight component amplitude is larger than other component which causes less faded signals. We also compare the system under study with SRNC system shown in Fig.3.(b). for $K=6$. As expected, our system generally achieves a better performance compared to SRNC.

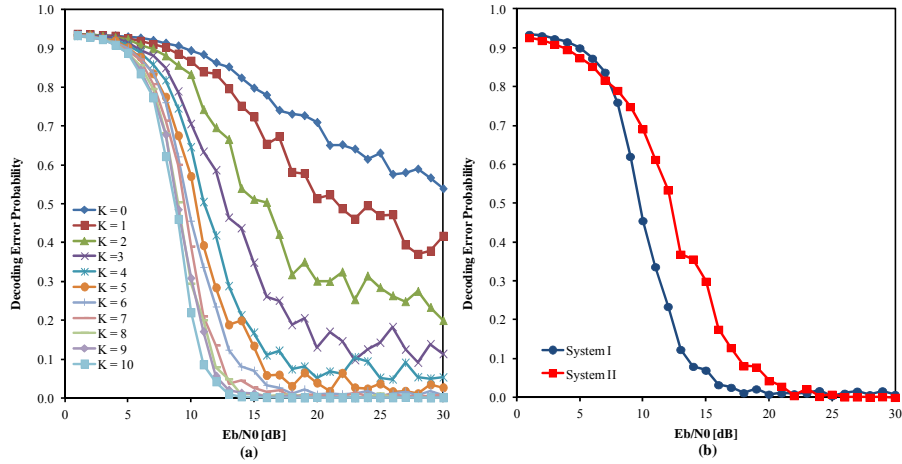


Fig. 3. Simulation Results of Decoding Error Probability for the Combined System of Random Linear Network Coding and Convolutional Code with Interleaving (System I) in Two-Hop Transmission (a) for Various Ratio of Line of Sight and Other Diffuse Components, K , of Rician Fading Channel, (b) in its comparison to SRNC (System II) for $K=6$.

4 Concluding Remark and Further Work

In this paper, we evaluated our system of combined RLNC and convolutional code with interleaving with interleaving for two-hop wireless network under series of Rician Fading and AWGN channels. The performance evaluation is carried out through simulation. Performance parameter term is presented in decoding error probability which directly represents throughput of communication network. The effect of Rician Factor, K , has been studied and simulation results have been as expected. We also compared our system with SRNC system. In general, our system outperforms SRNC. In future, the effect of block error rate on the throughput performance for the overall system under fading channel will be studied. Although we have not applied reconfigurability capability in Relay Node, our study can be extended to support such goal to realises CRN. The system also can be studied to extensions for other error correction schemes as well as other interleaving schemes.

Acknowledgement

First author would like to thank to the Government of Republic of Indonesia through the Ministry of Higher Education of Republic of Indonesia supporting the funding for this study.

References

1. I.F. Akyildiz, W.-Y. Lee, M.C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey", *Computer Networks* 50, pp. 2127 – 2159, 2006.
2. K.-C. Chen, Y.-J. Peng, N. Prasad, Y.-C. Liang, and S. Sun, "Cognitive radio network architecture: part I -- general structure", in *Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication*, pp. 114 – 119, 2008.
3. I. F. Akyildiz, W.-Y. Lee, and K.R. Chowdhury, "CRAHNS: Cognitive radio ad hoc networks", *Ad Hoc Network* 7, pp. 810 – 836, 2009.
4. R. Ahlswede, N. Cai, S.-Y.R. Li, and R.W. Yeung, "Network Information Flow", *IEEE Trans. Inform. Theory*, Vol. 46, No. 4, pp. 1204 – 1216, July 2000.
5. Christos Gkantsidis and Pablo Rodriguez, "Network Coding for Large Scale Content Distribution", *Proceedings of IEEE INFOCOM 2005, 24th Annual Joint Conference of the IEEE Computer and Communications Societies*, pp. 2235 – 2245, 13 – 17 March 2005.
6. S. Katti, H. Rahul, Wenjun Hu, D. Katabi, M. Medard, and J. Crowcroft, "XORS in the Air: Practical Wireless Network Coding", *IEEE/ACM Transactions on Networking*, **Volume** 16 , Issue 3, pp. 497 – 510, June 2008.
7. A. Asterjaldi, E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi, "Toward Network Coding-Based Protocols for Data Broadcasting in Wireless Ad Hoc Networks", *IEEE Transaction on Wireless Communication*, Volume 9, Issue 2, pp. 662 – 673, February 2010.
8. A. Argyriou, "Wireless Network Coding with Improved Opportunistic Listening", *IEEE Transaction on Wireless Communication*, Volume 8, Issue 4, pp. 2014 – 2023, April 2009.
9. S.-Y.R. Li, "Linear Network Coding", *IEEE Transactions on Information Theory*, Volume 49, Issue 2, pp. 371 – 381, February 2003.
10. R.Y. Kim, J. Jin, and Baochun Li, "Scattered Random Network Coding for Efficient Transmission in Multihop Wireless Networks", *IEEE Transactions on Vehicular Technology*, Volume 60, Issue 5, pp. 2383 – 2389, June 2011.
11. M.P. Fitz and J.P. Seymour, "On the Bit Error Probability of QAM Modulation", *International Journal of Wireless Information Networks*, Vol. 1, No. 2, pp. 131 – 139, April 1994.
12. M. Susanto, Y.F. Hu, and P. Pillai, "Joint Random Linear Network Coding and Convolutional Code with Interleaving for Multihop Wireless Network", *27th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pp. 114 – 119, 2013.
13. J. Busgang, "Some Properties of Binary Convolutional Code Generators", *IEEE Transactions on Information Theory*, Volume 11, Issue 1, pp. 90 – 100, January 1965.
14. W.W. Peterson and E.J. Weldon, Jr., "Error Correcting Codes", 2nd Edition, Cambridge, MA, The MIT Press, 1972.
15. International Telecommunication Union (ITU), "ITU-T Recommendation G.704: Synchronous Frame Structures Used at 1544, 6312, 2048, 8448 and 44 736 kbit/s Hierarchical Levels", Geneva, October 1998.

Performance analysis of IP Mobile Multicast Mechanisms during handover in Next Generation Satellite Networks

Esua Kinyuy Jaff, Prashant Pillai, Yim Fun Hu

School of Engineering, Design and Technology, Bradford, UK

ekjaff@student.bradford.ac.uk,{p.pillai,y.f.hu}@bradford.ac.uk

Abstract. Demand for ubiquitous communications as human mobility continues to increase, remains a big challenge for telecommunication industry. Satellite communication stands out as one of the technologies that might ease this demand. IP multicast over satellites utilises the scarce and expensive satellite resources in the most efficient way. Mobility support for multicast receivers as well as sources within a global multi-beam satellite network has witness very little attention from the research community. This paper proposes a comprehensive solution to the challenges of an IP mobile multicast receiver in a global multi-beam satellite network.

Keywords: IP Mobile Multicast, Return Channel Satellite Terminal (RCST), Satellite Beam, Satellite Gateway, Satellite Interactive Network.

1 Introduction

IP multicasting is a technology in which a single copy of IP data is sent to a group of interested recipients and the network replicating the data as required for delivery until a copy reaches all intended group members. This avoids processing overheads and bandwidth overheads associated with sending duplicate packets on the same network link for each interested recipient.

There are lots of proposed IP mobile multicast mechanisms in terrestrial networks as detailed by the authors in [1] but at the moment, there are no such proposed supports for IP multicast on mobile Return Channel Satellite Terminals (RCSTs) [2] during beam, gateway (GWH) or satellite handover (SH) in satellite networks. During gateway/satellite handover (GW/SH) between two satellite beams each served by its own GW having an IP addressing space which is different from every other, a mobile multicast receiver or source i.e., mobile node (MN) will experience some problems which eventually result to multicast packets loss or even link breakage. These problems result from the fact that the MN's IP address changes as the MN changes its point of attachment to the satellite network during a GW/SH from one satellite GW to another. The change in IP address implies that:

- The MN emerges from GW/SH as a completely new device as far as the upper layer identification is concerned.

- The multicast delivery tree (or branch) to the MN becomes invalid and a new delivery tree (or branch) to MN needs to be established (after re-subscription).

The multicast delivery tree discontinuity and the link discontinuity (due to handover latency) at GWH therefore result to multicast packets loss. Due to longer handover and propagation delay in satellite networks compared to terrestrial networks, the effects of these problems in satellite networks is even more severe. This paper proposes Satellite Home Subscription (SHS) and Satellite Remote Subscriptions (SRS) based approaches to mitigate specifically those effects caused by multicast delivery tree discontinuity.

2 IP Multicast on mobile scenarios in satellite networks

The reference network shown in Figure 1 is a global geosynchronous multi-beam satellite network with onboard processing (OBP). This complies with the ETSI standards [2] [3] [4] for DVB-S/S2/RCS.

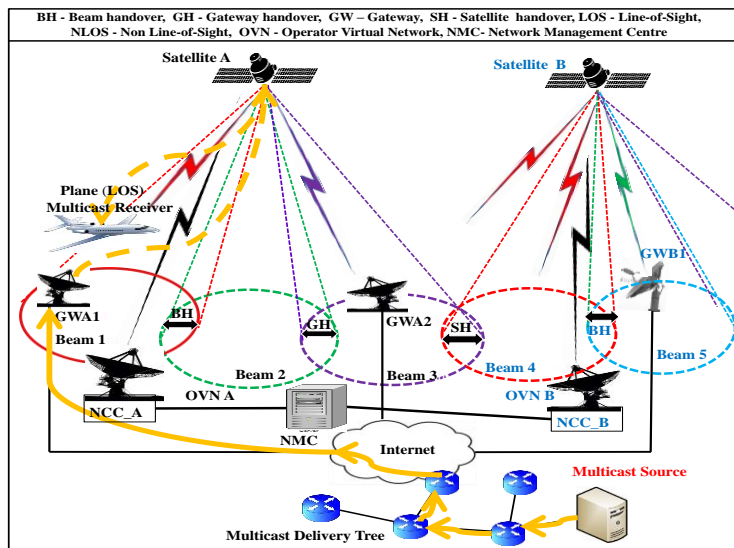


Fig. 1. Mobile multicast receiver at home network in a satellite interactive network

For management purposes and to comprehensively examine all possible handover aspects for Line-Of-Sight Scenarios (LOS) in a satellite network, the network is designed as follows:

- Satellite A has three beams labelled Beam 1, Beam 2 and Beam 3. Operator Virtual Network [4] (OVN) A is made up of two Satellite Virtual Networks [4] (SVNs), SVN 1 and SVN 2. SVN 1 which spans through beams 1 and 2 is served by gateway GWA1. SVN 2 is served by gateway GWA2 and is only limited to beam 3.

GWA1 and GWA2 provide interfaces for interconnection between the satellite system and terrestrial networks (e.g. Internet, ISDN/POTS, Intranet, etc.).

- Satellite B also has three beams although only two are shown on the reference diagram labelled Beam 4 and Beam 5. Part of OVN B shown in our reference diagram in Figure 1 is composed of one SVN (SVN 3) which spans through beams 4 & 5 and is served by gateway GWB1. GWB1 provides interfaces for interconnection between the satellite system and terrestrial networks.

NB: Each SVN forms a Virtual Routing Forwarding (VRF) group as defined in [4]. Gateways GWA1, GWA2 and GWB1 are all Regenerative Satellite Gateways (RSGW). OVN A and OVN B are parts of one global interactive network under the control of one Satellite network Operator (SNO) [4]. The aircraft (RCST) which represents the LOS scenario in beam 1 is in its home network. Figure 1 shows the IP multicast delivery tree from source to User Terminals (UTs) i.e. recipients in the aircraft. It is assumed that the Care-of addresses (CoAs) [5] that the mRCSTs will use in all target GWs are known in advance by all communicating parties. This is because these CoAs are the IP addresses of the FAs i.e., target GWs along the path of the mRCST. This implies there is no need for mobility agents' advertisements.

2.1 Satellite Home Subscription (SHS) based approach

Each of the GWs GWA1 GWA2 and GWB1 is assumed to have the home agent (HA) and foreign agent (FA) functionalities in addition to their normal responsibilities. There is only one HA and one FA under each GW. Due to the advanced knowledge of the mRCST's CoA in the target GW, it is therefore possible for the mRCST to register its CoA in the target GW at its HA before the GWH is completed. Hence, it is proposed that the Synchronization (SYNC) [3] burst which carries handover recommendation to the Network Control Centre (NCC) should carry the mobile IP (MIP) [5] registration message from the mRCST to its HA at GWA1. Because there is only one HA and one FA in each local network of the gateways, only one bi-directional tunnel can be established between the home network and the visited network at any point in time no matter how many mRCSTs from the home network are located at the visited network. This therefore eliminates the tunnel convergence problem experienced in HS based approaches in terrestrial networks.

Beam and Gateway handover

When the aircraft reaches the overlapping area between beams 1 and 2, beam handover will take place. Beam handover is considered as a lower-layer handover in which the NCC coordinates the handover procedure and no higher layer involvement is required in the implementation. Details of beam handover detection/recommendation, decision and execution can be found in [3]. There is little or no change in the multicast delivery tree apart from the fact that if the aircraft is the first member of this group in beam 2, then NCC_A during handover execution will instruct the Onboard Processor (OBP) to forward multicast traffic for this group to beam 2 and also the handover command (with information about resources to be used

in new beam) to the aircraft. Ongoing multicast communications and others inside the aircraft will go on unperturbed in a seamless handover. Upon reception of the handover recommendation from the aircraft as it enters the overlapping area between beams 2 and 3, NCC_A will retrieve the target beam identity from its database and determine whether the beam belongs to a different GW. Note should be taken here that this handover recommendation contains the MIP registration message from a mRCST to its HA at GWA1. NCC_A will realize that the target beam (beam 3) is served by a different gateway, GWA2 and so, a GWH is decided. The NCC_A will then update its service information tables which include Terminal Burst Time Plan (TBTP), SCT Superframe Composition Table (TBTP), Frame Composition Table (FCT) and Time-slot Composition Table (TCT) and signalling between NCC_A, GWA1 and GWA2 is carried out to prepare for GWH. NCC_A will send an SNMP Set-Request message to the GWA2 for events synchronization to ensure that the GWA2 gets ready for connection with the RCST (aircraft). The updated SI tables, together with the routing update information of the aircraft and unicast IP address of the HA at the current serving gateway (GWA1) will be included in this message. The routing update information is generally implemented by sending the location change information to the broadcaster, which is generally handled by the location management scheme.

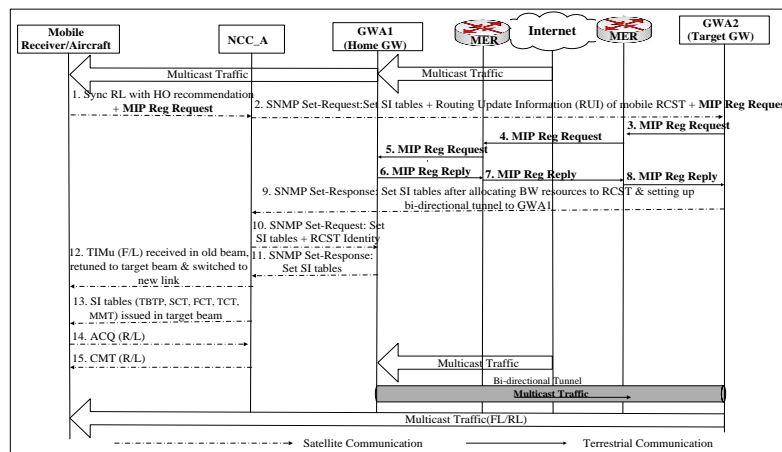


Fig. 2. SHS based approach signalling sequence at gateway handover

Upon reception of the Set-Request signalling, the target GW i.e., GWA2 will allocate bandwidth resources for the aircraft according to the new burst time plan and also forward its IP address to GWA1 as the CoA of the moving aircraft (RCST). The association of the HA (GWA1) and CoA of the aircraft is called binding. After receiving the CoA, the HA creates a binding cache entry that maps the permanent IP address, the multicast group address and the CoA of the aircraft and then sends a binding acknowledgement to indicate that the forwarding to the GWA2 (aircraft) is set. Once the binding process is completed, a bi-directional tunnel [5] is established between the HA at GWA1 and FA at GWA2, and the HA is ready to tunnel all subsequent multicast packets for the aircraft to GWA2 [6]. The acknowledgement Get-Response

message is then sent from the GWA2 to the NCC_A. NCC_A will now send a Set-Request message to GWA1, which includes the aircraft identity and the SI tables. After receiving the Set-Request message from NCC_A, GWA1 will buffer the FL user traffic of the aircraft to be tunnelled to FA at GWA2 during handover. GWA1 will then acknowledge NCC_A by sending it a Get-Response message. GWH always entails beam handover [3]. Upon reception of the Get-Response message from GWA1, a GWH command is issued to the aircraft from NCC_A in a Mobility Control Descriptor carried in a Terminal Information Message Unicast (TIMu) message using old beam. TIMu message also contains new Time Division Multiplex (TDM), Super-frame ID, Group_ID, Logon_ID, PIDs [2] [4] necessary for logging on and functioning in the new beam. GWA1 updates its route mapping table and released resources used by the aircraft. Once the aircraft receives the handover command, it synchronizes with the NCC_A and the GWA2, retunes itself to the new beam and receives traffic from the new beam which comes through the new gateway GWA2. Figure 2 shows the signalling sequence during GW handover. This signalling sequence shows MIP registration message from the mRCST to its HA as proposed, integrated into the standard GW handover signalling sequence given in [3].

The GWH is completed when the aircraft sends an Acquisition (ACQ) message to NCC_A after it finishes the retuning process and receives the Correction Message Table (CMT) message from NCC_A.

Satellite Handover

When the aircraft reaches the overlapping area between beams 3 and 4, it will detect the need for handover [3] [7] and will send a handover request/recommendation (containing MIP registration message) to the NCC_A. Upon reception of the handover recommendation from the aircraft, NCC_A will retrieve the target beam identity from its database and determine whether the beam belongs to a different gateway/satellite. Once NCC_A realized that the target beam belongs to another satellite, then it will start procedures for a satellite handover. NCC_A that controls operator virtual network A (OVN A) will send a message to NCC_B that controls OVN B where the target beam (Beam 4) belongs about the network conditions to see whether it is ready to accept a moving-in RCST (aircraft). It should be noted that OVN A and OVN B, all belong to one global interactive network, so NCC_A and NCC_B have a good depth of information about each other. To maintain a good level of quality of service (QoS) and also Service Level Agreements (SLA) of the ongoing communications in the aircraft, a good estimate of the amount of resources (bandwidth) required by the moving-in aircraft and the type of communication going on will be communicated to NCC_B. If NCC_B confirms that the required resources are available and that it is ready to accept the aircraft, then, NCC_A will send a handover command to the aircraft. Upon reception of the handover command, the aircraft (RCST) synchronizes to the new beam 4 and logs on to the new OVN B under the control of the new NCC_B. During the communication between the two NCCs and the target gateway (GWB1), the MIP registration message from the mRCST (i.e., aircraft) is delivered to GWB1. If the multicast groups with members in the aircraft are new to GWB1, then the FA at GWB1 will forward its IP address as the CoA of the mRCST (aircraft) to

HA at GWA1 for binding update. This will result in a bi-directional tunnel formed between GWA1 and GWB1 as illustrated in the multicast communication signalling sequence in Figure 3.

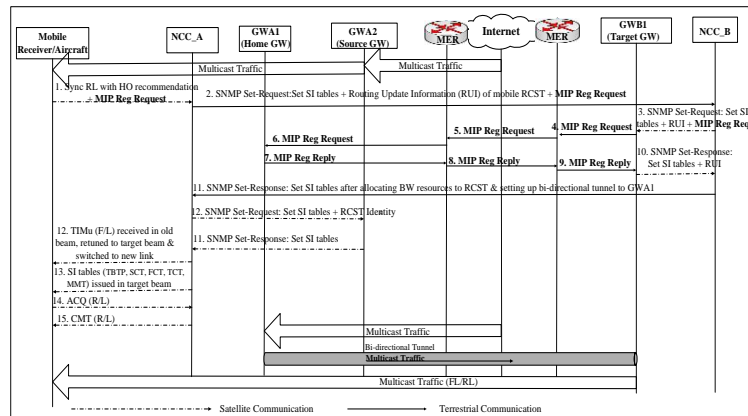


Fig. 3. SHS based approach signalling sequence at satellite handover

The advantage of SHS based approach is its simplicity since the mRCST does not need to rejoin the multicast group as its serving gateway changes. However, this approach suffers from triangular routing through the home network, which increases the join latency which could have a significant negative impact on satellite networks. The fact that SHS relies completely on the HA to forward multicast traffic to the mRCSTs implies a single point of failure, which is very risky. Also, tunnelling through HA incurs overheads in home network.

2.2 Satellite Remote Subscription (SRS) Based approach

Similarly, the advanced knowledge of the mRCST's CoA in the target GW implies that the mRCST can issue the Internet Group Management Protocol (IGMP) [8] join report message to the target GW at the beginning of the GWH procedure. This will make the target GW (upon reception of the IGMP join report) to join the multicast groups of interest to the mRCST before the GWH procedure is completed. It is therefore proposed that the SYNC burst which carrying handover recommendation to the NCC should carry the IGMP join report message from the mRCST target GW.

Gateway handover

When the aircraft enters overlapping area between beams 2 and 3, the handover detection and decision is exactly the same as in SHS based approach described above. The main difference here is that the Synchronization (SYNC) burst which carries handover recommendation to NCC_A contains the IGMP report join message destined for the target gateway GWA2 instead of the MIP registration message. All the multicast groups with members in the aircraft are contained in the IGMP report join message. When the target GW GWA2 which is about to take the responsibility of

servicing the aircraft finally receives the IGMP report message, it will then join the multicast groups that the aircraft is a member of, in order to continue the multicast services to the aircraft when the GWH is completed. Figure 4 illustrates the signalling sequence involved in SRS based approach. Here, there is no binding of the GWA2 IP address to the HA at GWA1 as was the case in SHS based approach.

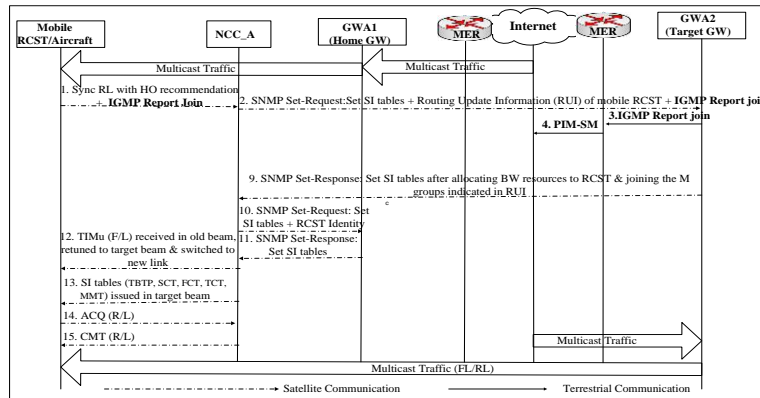


Fig. 4. SRS based approach signalling sequence at gateway handover

Satellite handover

As the aircraft continues towards beam 4, at the overlapping area between beams 3 and 4, a SH will take place. Upon reception of the IGMP report join message which originated from aircraft (mRCST), the GWA1 will join the multicast group(s) that has members in the aircraft as it assumes the responsibility of servicing the aircraft. Figure 5 illustrates the signalling sequence required for SRS based approach at SH. Generally, the SRS based approach enjoys route optimization compared to the SHS based approach since multicast traffic is routed from source to directly to the gateway servicing the aircraft through the shortest possible path.

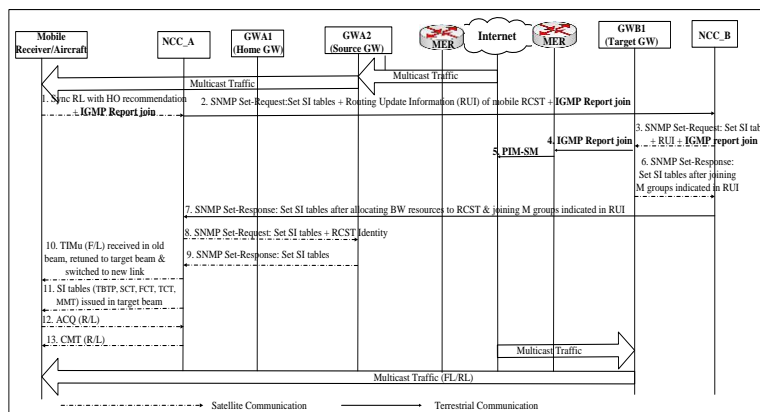


Fig. 5. SRS based approach signalling sequence at satellite handover

3 Comparison of SHS and SRS using Delay Analysis

Delay analysis is used here to compare the signalling times required by the SHS and SRS based approaches. Here, the signalling messages sizes required to handover multicast traffic during GWH and SH are used to perform the delay analysis. From the signalling sequences in Figures 2, 3, 4 and 5, the signalling messages involved in multicast communication during GWH and SH are given in table 1.

MIP registration request and reply messages sizes are assumed to be 90 and 130 bytes respectively [5] [9]. The IGMP and PIM-SM packets each containing 10 multicast groups are assumed to be 64 bytes each. The packet sizes of the SYNC Burst, SNMP Set-Request, SNMP Set-Response, SI Tables, TIM, ACQ, CMT, routing update information table (RUI) containing at least 100 bytes of routing data were obtained from [10].

Table 1. Multicast signalling messages during handover

Messages	Packet Length (Bytes)
Synchronisation (SYNC) Burst + MIP Reg Req/IGMP Report	106/80
SNMP Set-Request: set SI tables + RUI + MIP Reg Req/IGMP Report	826/800
IGMP Join	64
PIM-SM	64
Mobile IP Registration Request	90
Mobile IP Registration Reply	130
SNMP Set-Reply: set SI tables	636
SNMP Set-Request: set SI tables + RCST Identity	640
SNMP Set-Response: set SI tables + RCST Identity	640
TIM (Terminal Information Message)	35
SI Tables (TBTP, SCT, FCT, TCT, MMT)	152
ACQ (Acquisition Burst)	12
CMT (Correction Message Table)	30

The time taken to transmit a single message between two relevant network entities over any given link under ideal conditions i.e., lossless conditions is given by (1) and the time required to send a message during handover to relevant signalling entities under lossy conditions is given by (2) [9].

$$T_{lossless} = T_{trans} + T_{prop} + T_{proc} \quad (1)$$

$$T_{total} = T_{lossless} + (T_{lossless} + T_w) \times \left[\frac{q}{1-q} \right] + T_{INT} \quad (2)$$

Where,

T_{trans} = transmission delay = message size \div link bit rate; T_{Proc} = average processing time at any node. This is assumed to be 5 ms [9] for all nodes; q = probability of a failure transmission over satellite; T_{prop} = propagation delay due to the communication link [9].

Since it is impossible to know the route taken by the packet in the Internet with certainty, TINT is assumed to be 8 ms as suggested in [9]. The data rate in the satellite link is assumed to be 144Kbps [9] and the gateways are assumed to be the Internet using 100 Base-T Ethernet supporting a data rate of 100Mbps [9]. The propagation speed in LAN Ethernet is assumed to be 2/3 the speed of light (2×10^8 m/s) [9]. The distance between the gateway and the MER (multicast edge router) is assumed to be 4m.

Figures 6 compare the time delays at probabilities of failure 0% - 10% for SHS and SRS based approaches during GWH and SH respectively. From these graphs we can deduce that as the probability of failure increases, the time delay also increases and that for any particular probability of failure,

- The time delay for SHS based approach is always slightly higher than that for SRS based approach in both GWH and SH.
- The time delay for SH is always greater than that for GWH.

The last point confirms the understanding that there are more signalling hop by hop messages during SH than GWH.

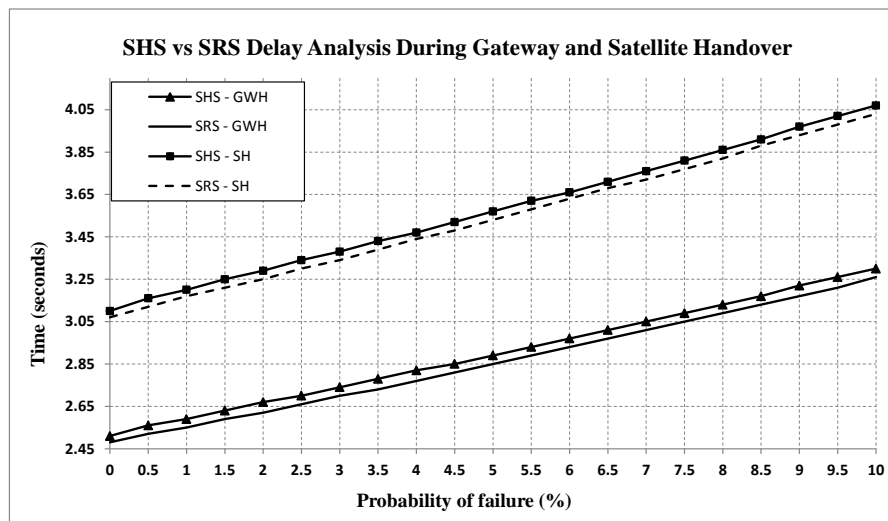


Fig. 6. Comparison of time delay in SHS & SRS at probabilities of failures 0% - 10% during gateway and satellite handovers

4 Conclusion

Due to the fact that IP multicast is a bandwidth conserving technology, it will always be associated to satellite networks. When mobile multicast receivers change their point of attachment to the network, the multicast tree will break down. Solutions proposed to solve this problem in internet are not very suitable to satellite environment. This paper proposes suitable solutions for multicast receiver mobility in a multi-beam satellite network. Additional signalling messages have been integrated into the standard GW handover signalling sequence given in the DVB specification. The proposed solutions will significantly minimise the time the mobile receiver will resume user traffic reception after the multicast delivery tree break down. From the comparison using time delay analysis, there is very little difference between the SHS and SRS – based approaches during GWH and SH. Despite many proposed solutions, IP mobile multicast in general is still difficult to deploy and more so in a satellite network environment.

References

1. I. Romdhani, M. Kellil, L. Hong-Yon, A. Bouabdallah, and H. Bettahar, "IP mobile multicast: Challenges and solutions," *Communications Surveys & Tutorials, IEEE*, vol. 6, pp. 18-41, First Quarter 2004.
2. "Digital Video Broadcasting (DVB); Interaction channel for satellite distribution systems," ETSI EN 301 790, May 2009.
3. "Digital Video Broadcasting (DVB); Interaction channel for Satellite Distribution Systems; Guidelines for the use of EN 301 790 in mobile scenarios," ETSI TR 102 768, April 2009.
4. "Digital Video Broadcasting (DVB); Second Generation DVB Interactive Satellite System (DVB-RCS2); Part 3: Higher Layers Satellite Specification," ETSI TS 101 545-3, May 2012.
5. C. Perkins, "IP Mobility Support for IPv4," IETF RFC 3344, Aug. 2002.
6. A. Z. M. Shahriar, M. Atiqzaman, and W. Ivancic, "Network Mobility in satellite networks: architecture and the protocol," *International Journal of Communication Systems*, vol. 26, pp. 177-197, Feb. 2013.
7. A. Bolea Alamanac, P. M. L. Chan, L. Duquerroy, Y. F. Hu, G. Gallinaro, W. Guo, and D. Mignolo, "DVB-RCS goes mobile: Challenges and technical solutions," *International Journal of Satellite Communications and Networking*, vol. 28, pp. 137-155, Feb. 2010.
8. "Satellite Earth Stations and Systems (SES); Broadband Satellite Multimedia (BSM) services and architectures; IP Interworking over satellite; Multicast group management; IGMP adaptation," ETSI TS 102 293, Feb. 2004.
9. P. M. L. Chan, R. E. Sheriff, Y. F. Hu, P. Conforto, and C. Tocci, "Design and evaluation of signaling protocols for mobility management in an integrated IP environment," *Computer Networks*, vol. 38, pp. 517-530, May 2002.
10. G. Maral and M. Bousquet, *Satellite Communications Systems: Systems, Techniques and Technology*, 5th ed.: John Wiley & Sons Ltd., Dec. 2010.

Performance Modelling and Analysis of MAC Protocols in Wireless LANs under Multimedia Traffic

Jia Hu¹, Geyong Min², Mike E. Woodward³

¹ Department of Mathematics and Computer Science, Liverpool Hope University, Liverpool, L16 9JD, UK

² Department of Mathematics and Computer Science, University of Exeter, Exeter, EX4 4QF, UK

³ Department of Computing, University of Bradford, Bradford, BD7 1DP, UK

Abstract. The wireless Medium Access Control (MAC) protocol, which regulates access and data transmission from wireless nodes to the shared medium, plays a pivotal role in Wireless Local Area Networks (WLANs). Performance modelling and analysis has been and continues to be of great theoretical and practical importance in the design and development of WLANs. This tutorial provides an in-depth overview on MAC protocols modelling while highlighting key challenges and associated solutions, beginning with an overview on Quality-of-Service (QoS) of MAC and advanced analytical techniques, such as Markov Chain and queueing analysis, Heterogeneous Traffic models, and Channel models, suitable for modelling and analyzing MAC protocols. This tutorial also presents details of developing efficient and cost-effective analytical tools for the analysis and enhancement of MAC protocols in WLANs under heterogeneous multimedia traffic. More specifically, comprehensive analytical models are proposed to accommodate the integration of the three QoS schemes of the 802.11e MAC in terms of AIFS, CW, and TXOP under multimedia traffic. Some validation and analytical results are shown to demonstrate the effectiveness and efficiency of the proposed analytical models. The performance results highlight the importance of taking into account the heterogeneous traffic for the accurate evaluation of the TXOP scheme in wireless multimedia networks. Moreover, the results demonstrate that the MAC buffer size has considerable impact on the QoS performance of 802.11e MAC. This tutorial concludes with an in-depth overview on the current and future challenges facing the wireless MAC.

Performance Modelling and Analysis of MAC Protocols in Wireless LANs under Multimedia Traffic

Jia Hu¹, Geyong Min², Mike E. Woodward³

**¹Department of Mathematics and Computer Science, Liverpool Hope University,
Liverpool, L16 9JD, UK**

**²Department of Mathematics and Computer Science, University of Exeter, Exeter,
EX4 4QF, UK**

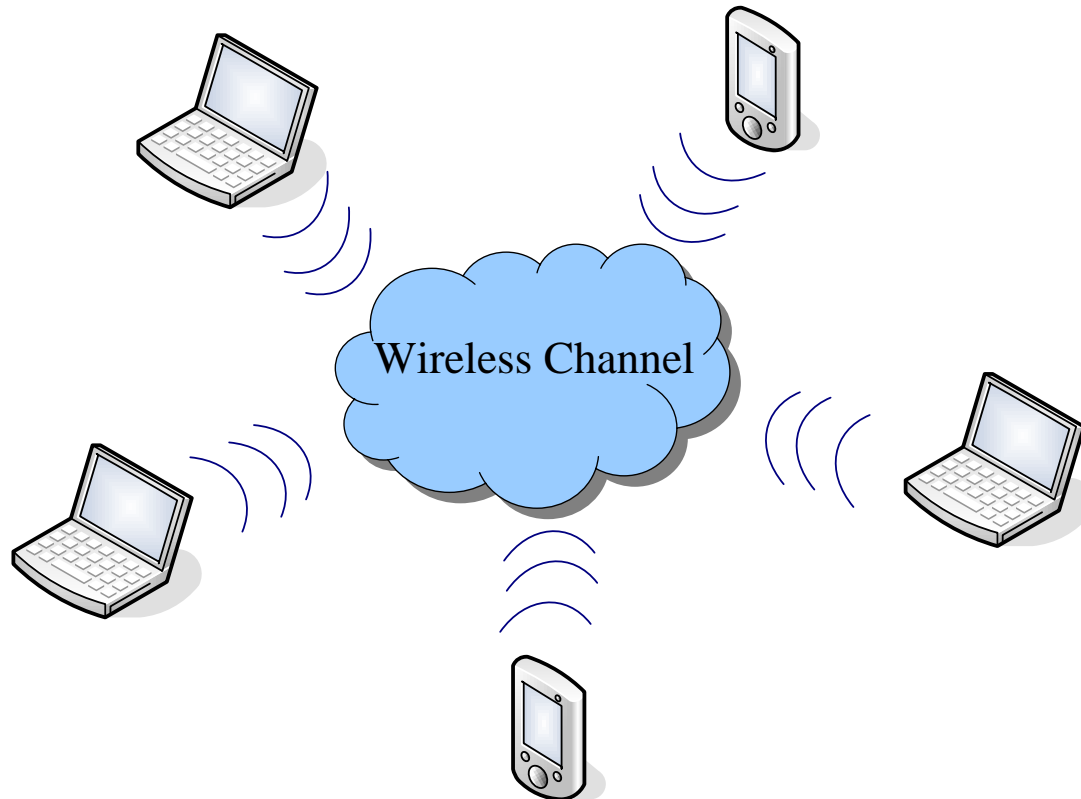
³Department of Computing, University of Bradford, Bradford, BD7 1DP, UK

Outline

- **Backgrounds:** *Markov Chain, Queuing Theory, MAC & Traffic/Channel Models*
- **Models:** *EDCA backoff, A Queuing Framework*
- **Results:** *Validation & Evaluation*
- **Conclusions:** *Summary & Discussion*

Medium Access Control

- ❑ Wireless channel is **shared** among multiple neighboring nodes.
- ❑ **Collision** occurs if more than one node transmit at a time.
- ❑ How to **coordinate** the transmission?



IEEE 802.11 DCF

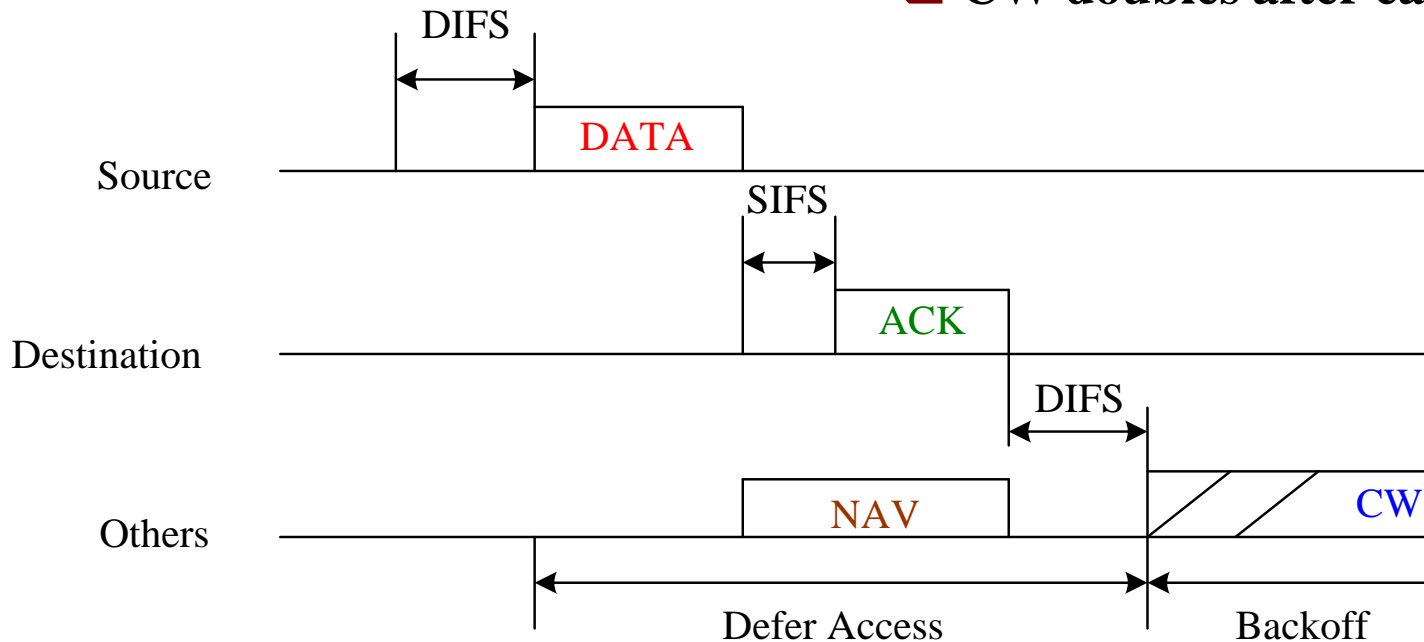
Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA)

Carrier Sensing

- Physical Carrier Sensing
- Virtual Carrier Sensing

Binary Exponential Backoff (BEB)

- Contention window randomly chosen from $[0, CW]$
- CW doubles after each collision



T08-4

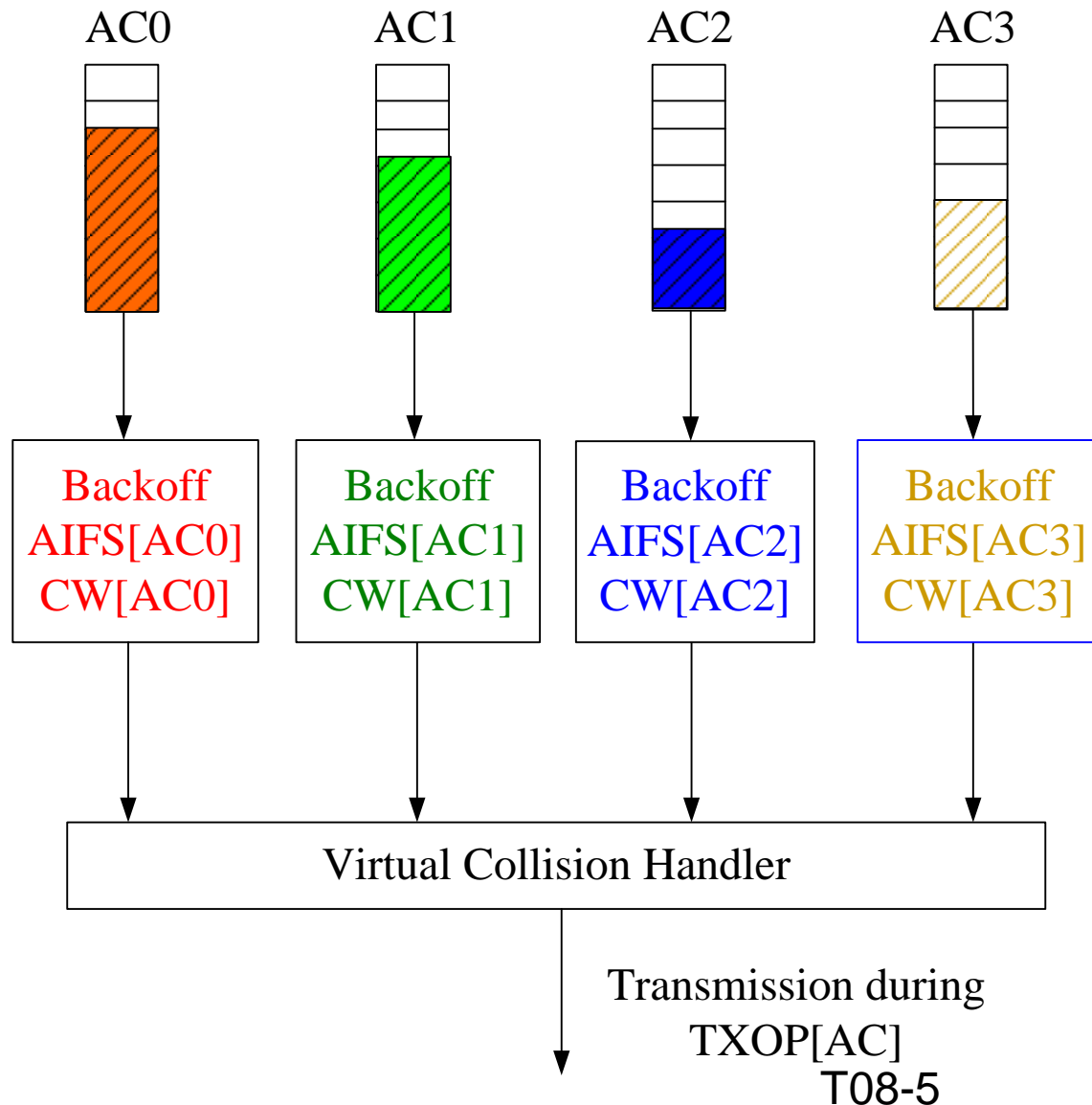
Can it support QoS requirements of various applications?

IEEE 802.11e EDCA

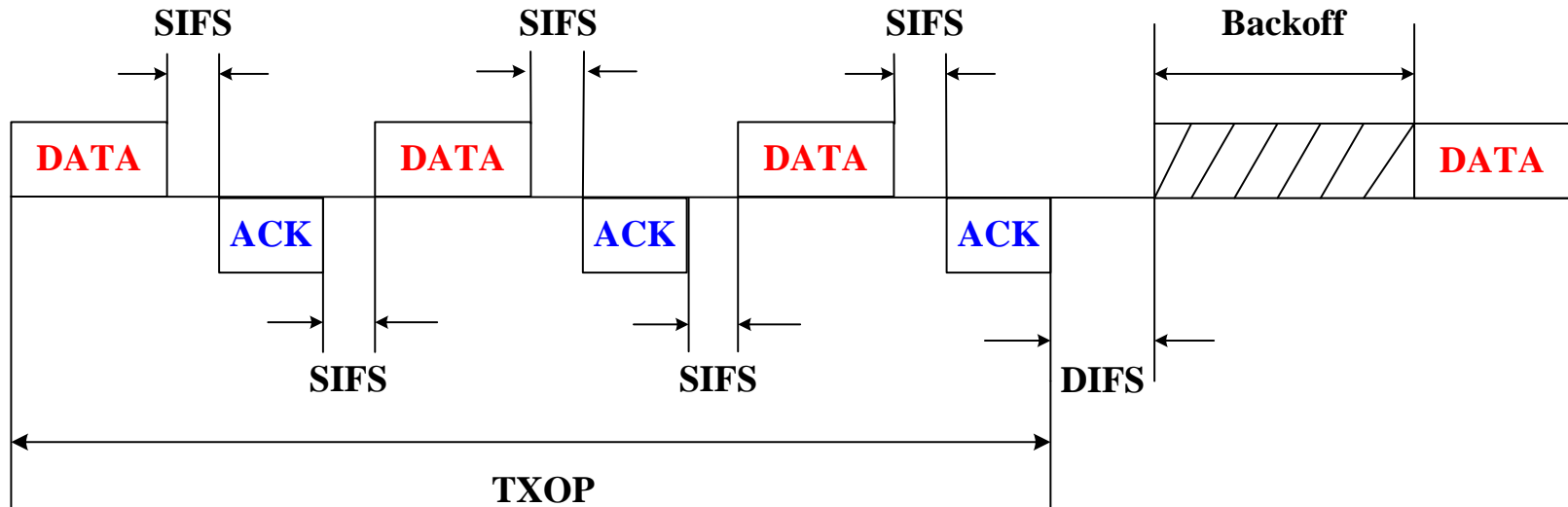
❑ **AIFS:** To control the length of the deferring period

❑ **CW (CW_{min} and CW_{max}):** To control the length of the backoff period

❑ **TXOP:** To control the channel occupation time



TXOP Scheme

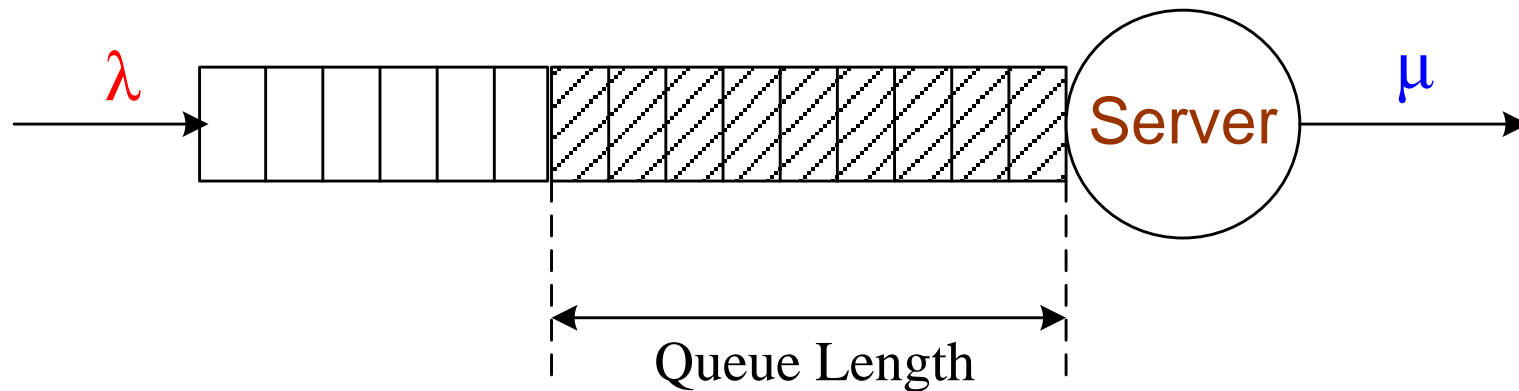


- ❑ Transmit multiple frames consecutively within a TXOP
- ❑ Each frame is acknowledged by an **ACK** after an **SIFS**
- ❑ The next frame is transmitted immediately after it waits for an **SIFS** upon receiving this **ACK**

A Simple Queue

□ distributions of arrival and service process, service discipline

→ Queue length, delay of customers, etc...

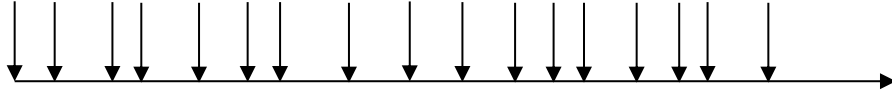


Markov Chain

- ❑ A Markov chain is the **characterization of a system** that transits from one state to another, in a chainlike manner.
 - **Markov property**: i.e., the property, simply said, that the next state depends only on the current state and not on the past.
- ❑ Markov chains are useful as tools for **statistical modelling** in almost all fields of modern applied mathematics.
- ❑ A **discrete-time Markov chain** (discrete set of times) or a **continuous-time Markov chain** ("time" can take continuous values).
- ❑ The changes of state of the system are called **transitions**, and the probabilities (discrete) or rates (continuous) associated with various state-changes are called transition probabilities (discrete) or transition rates (continuous).

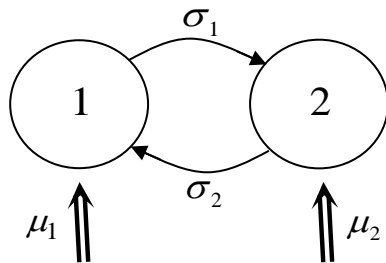
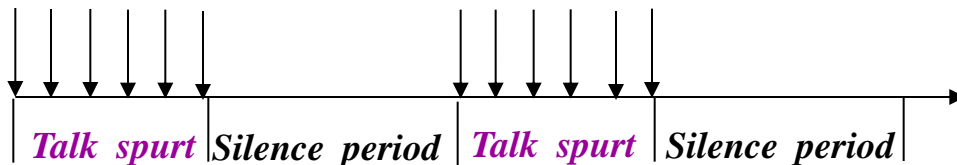
Traffic Model (1)

□ Poisson Arrival Model



- Inter-arrival intervals are **exponentially distributed**.
- Suitable to model “**non-bursty**” text data.

□ Markov-Modulated Poisson Processes (MMPP) Model



Markov model for the MMPP

- A **doubly stochastic process**
- Represented by **two matrices**

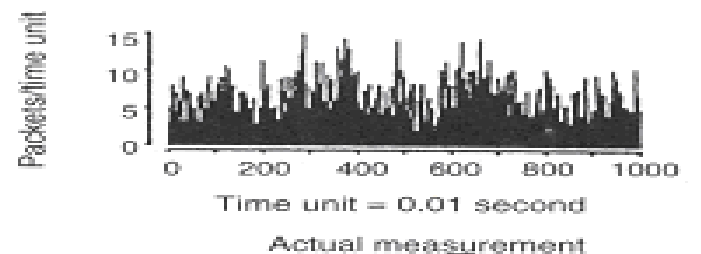
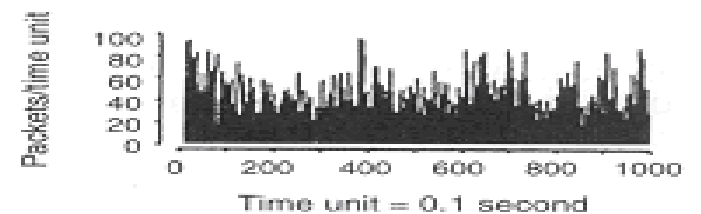
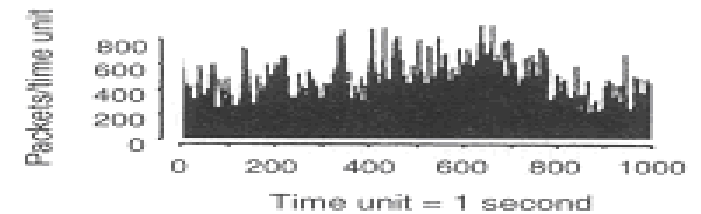
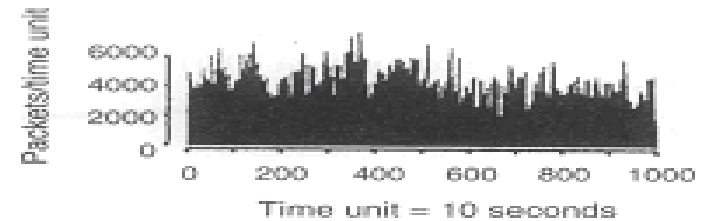
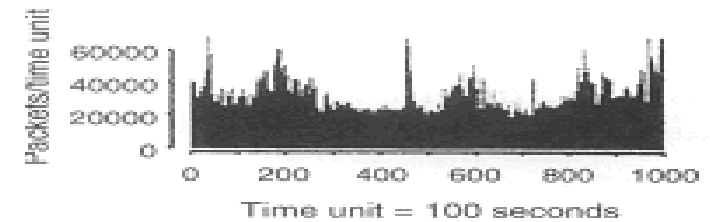
$$Q_s = \begin{bmatrix} -\sigma_1 & \sigma_1 \\ \sigma_2 & -\sigma_2 \end{bmatrix} \quad \Lambda_s = \begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{bmatrix}$$

Traffic Model (2)

□ Self-Similar Model

- Traffic burstiness appears **over a wide range of time scales.**

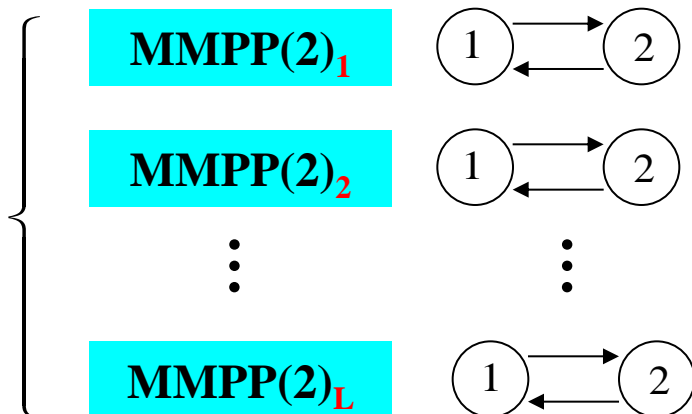
- Ethernet local area traffic
- World Wide Web traffic
- TCP, FTP, and TELNET traffic
- Variable-Bit-Rate (VBR) video
- Wireless and mobile networks traffic



Modelling Methods of Self-Similarity

- **New models** (fractional Gaussian noise, fractional Brownian motion, chaotic maps, ON/OFF sources with heavy-tailed distributions)
→ Tools for queueing analysis of these models are in an embryonic state.

- **Self-similarity is modelled by superposing a number of two-state Markov-Modulated Poisson Processes (MMPPs).**

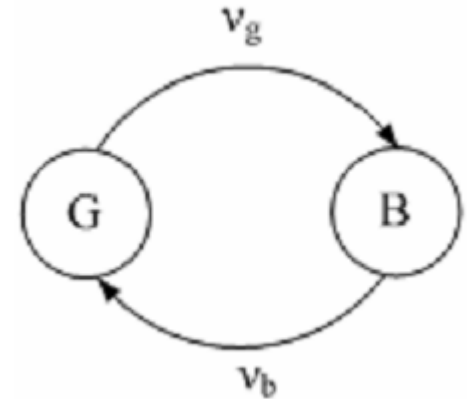


Advantages:

- Modelling of burstiness over multiple time scales.
- Well-developed queueing theory for the MMPP.

Channel Model

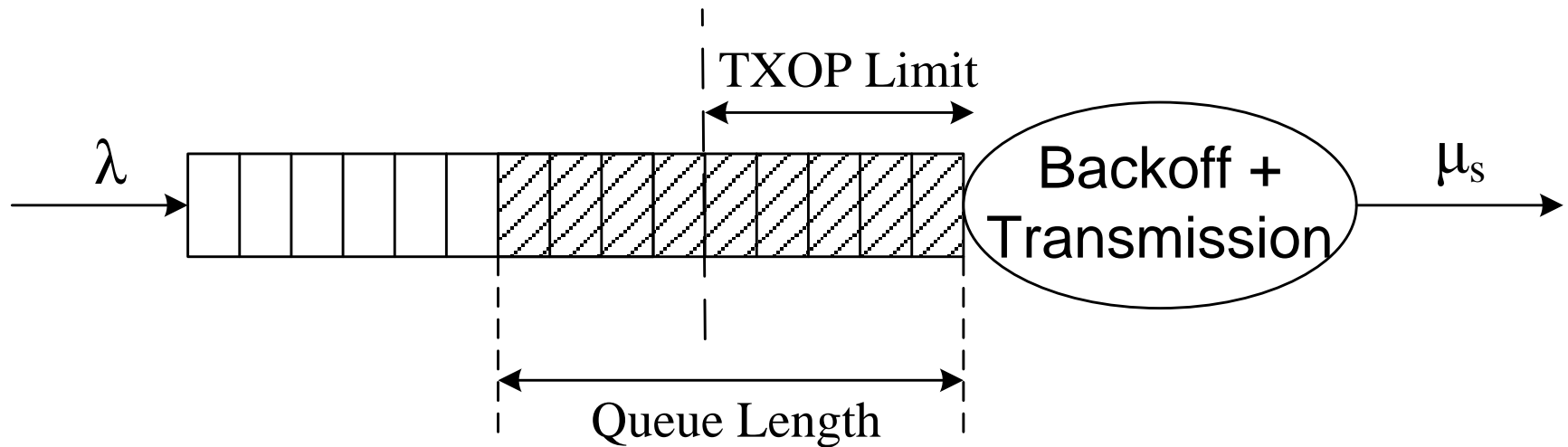
- ▶ The **bursty** characteristics of channel errors can be modelled by a two-state Markov model
- ▶ Channel states
 - ▶ Probability of being in the **good (G)** state
 - ▶ Probability of being in the **bad (B)** state



$$\pi_g = \frac{v_b}{v_g + v_b}$$

$$\pi_b = \frac{v_g}{v_g + v_b}$$

Outline of Model



- ❑ The transmission queue at each AC is modelled as a **bulk service queueing system**.
- ❑ The **service rate**, μ_s , is derived by analyzing the **backoff procedure and burst transmission mechanism**.
- ❑ The **backoff procedure** of each AC is modelled by a **three-dimensional Markov chain**.

Modelling the Backoff Process in EDCA

□ $\{s(t), b(t), c(t)\}$: states of the three-dimensional Markov chain

➤ $s(t)$: backoff stage

➤ $b(t)$: backoff counter

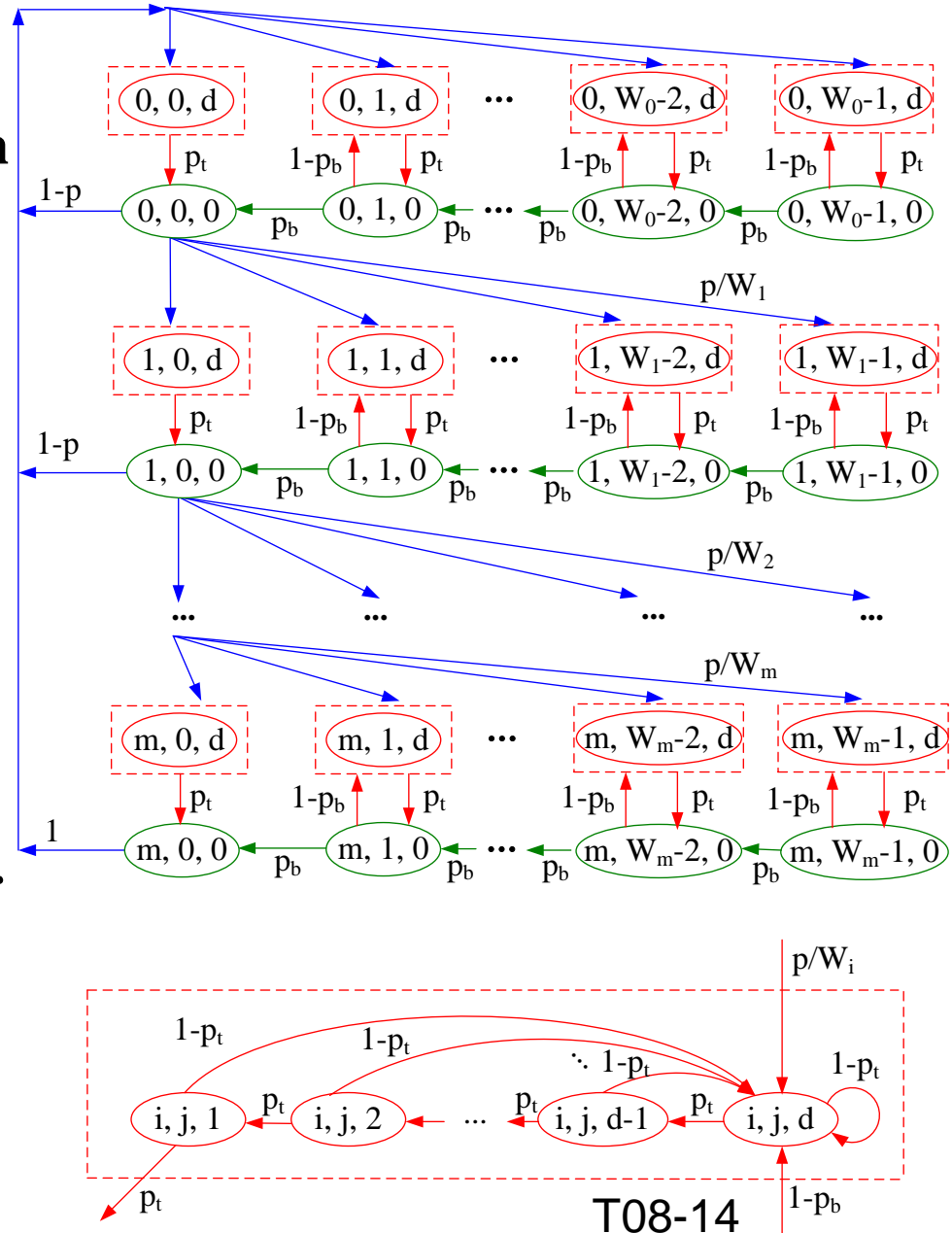
➤ $c(t)$: the number of remaining time slots to complete AIFS[AC]

□ The **sub-Markov chain** denotes the deferring period of the AC.

➤ p : collision probability

➤ p_b : channel idle probability after completing AIFS[AC]

➤ p_t : channel idle probability before completing AIFS[AC].



Service Time Analysis

- ❑ The **service time of a burst with s frames** can be modelled by an **exponential distribution** function with mean $E[S_{sv}]$.

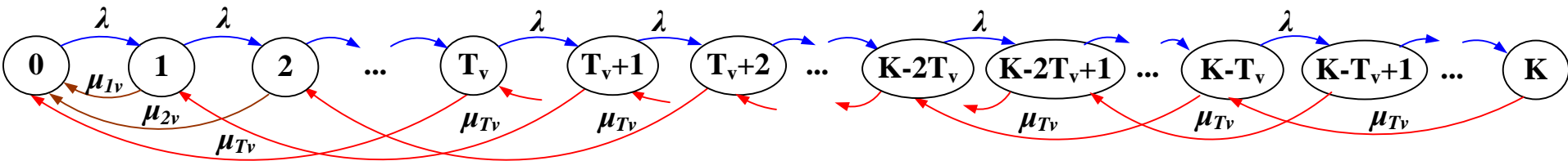
- ❑ $E[S_{sv}] = E[A_v] + E[B_{sv}]$
 - **Channel access delay ($E[A_v]$):** duration from the instant that a frame reaches to the head of the queue, until it wins the contention.

 - **Burst transmission delay ($E[B_{sv}]$):** duration of transmitting a burst.

Queueing Model (Poisson)

□ **M/G^{[1, T_v]/1/K}** queueing system

- [1, T_v] denotes that the **number of frames** transmitted within a burst ranges from 1 to T_v (**TXOP limit**).
- K represents the **buffer size**.



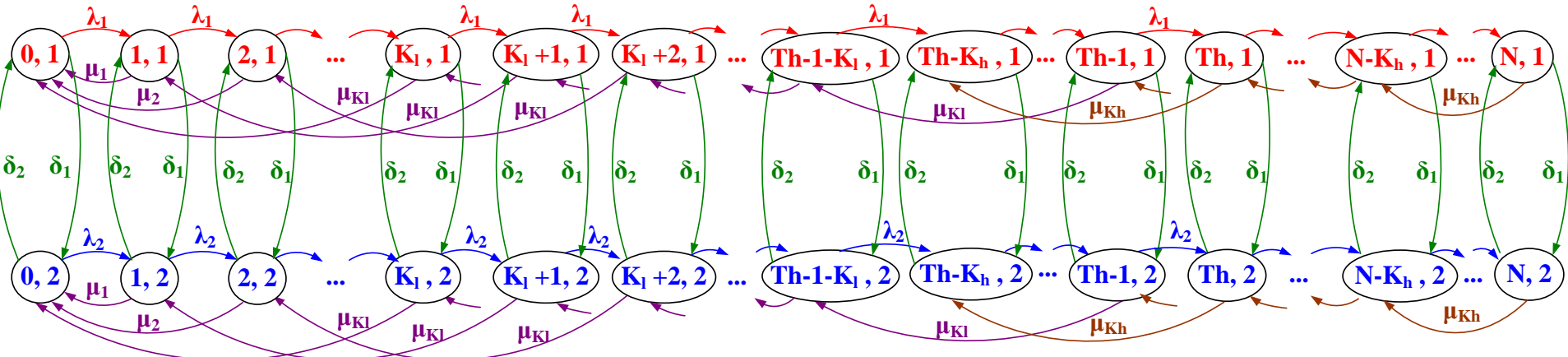
□ The queueing system is modelled by a **Markov chain**

- state i denotes i **frames** in the queueing system

Queueing Model (MMPP)

□ **MMPP/G^{[1, K_h]/1/N}** queueing system

- [1, K_h] denotes that the **number of frames** transmitted within a burst ranges from 1 to K_h.
- N represents the **buffer size**.



□ The queueing system is modelled by a **bi-variate Markov chain**

- state (i, s) , denotes:
 - **i frames** in the queueing system and
 - the multi-state MMPP is at **state s**.

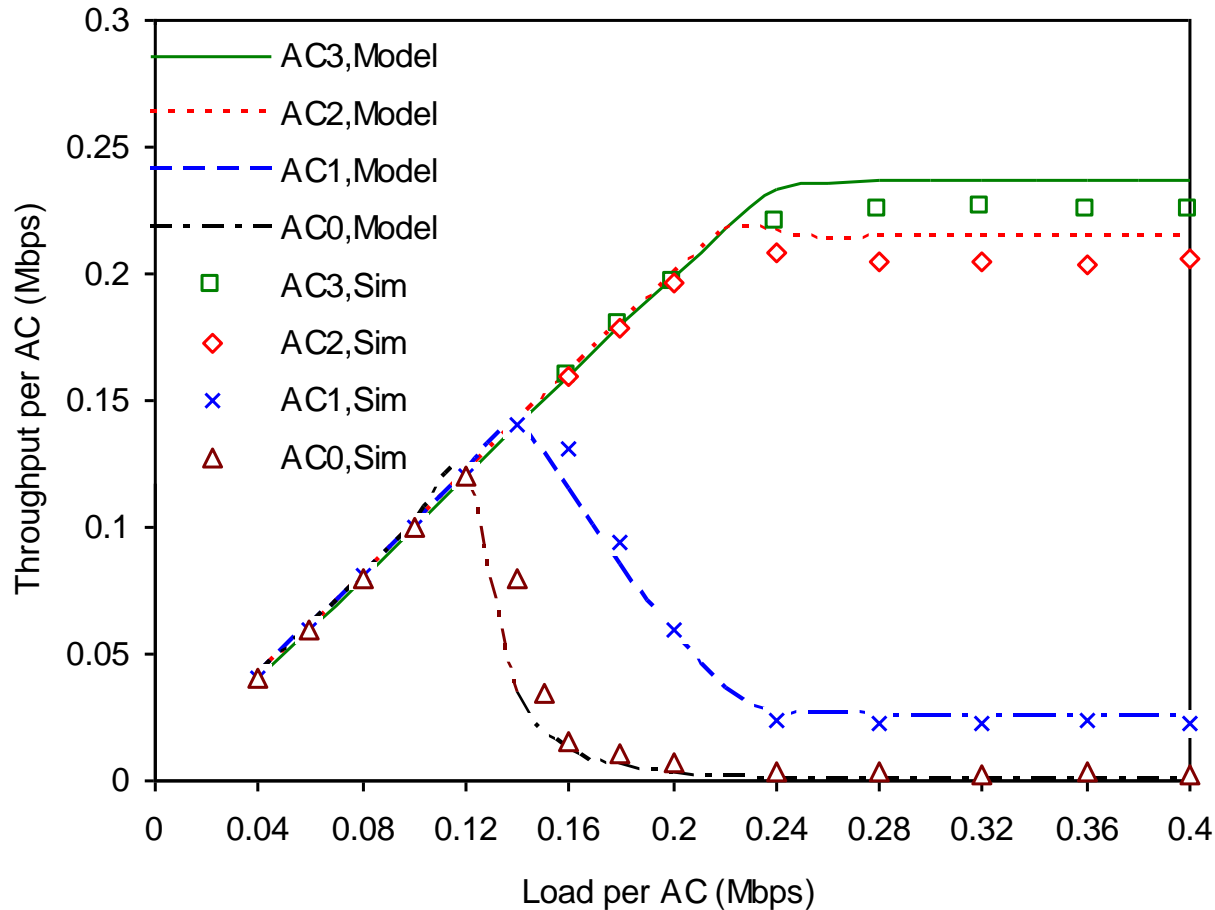
Model Validation

- ❑ **NS-2 simulation**
- ❑ **Basic Service Set (BSS)**
- ❑ **Each station generates four ACs of traffic**
- ❑ **Throughput, delay, delay jitter, and frame loss probability**

System Parameters

Frame payload	8000bits	PHY header	192bits	
MAC header	224bits	ACK	112bits + PHY header	
Channel rate	11Mbit/s	Basic rate	1Mbit/s	
Slot time	20us	Buffer size	50 frames	
SIFS	10us	Retry limit	7	
	AIFSN	CW_{min}	CW_{max}	TXOP
AC₀	6	32	1024	1 Frame
AC₁	2	32	1024	1 Frame
AC₂	2	16	256	4 Frames
AC₃	2	8	128	2 Frames

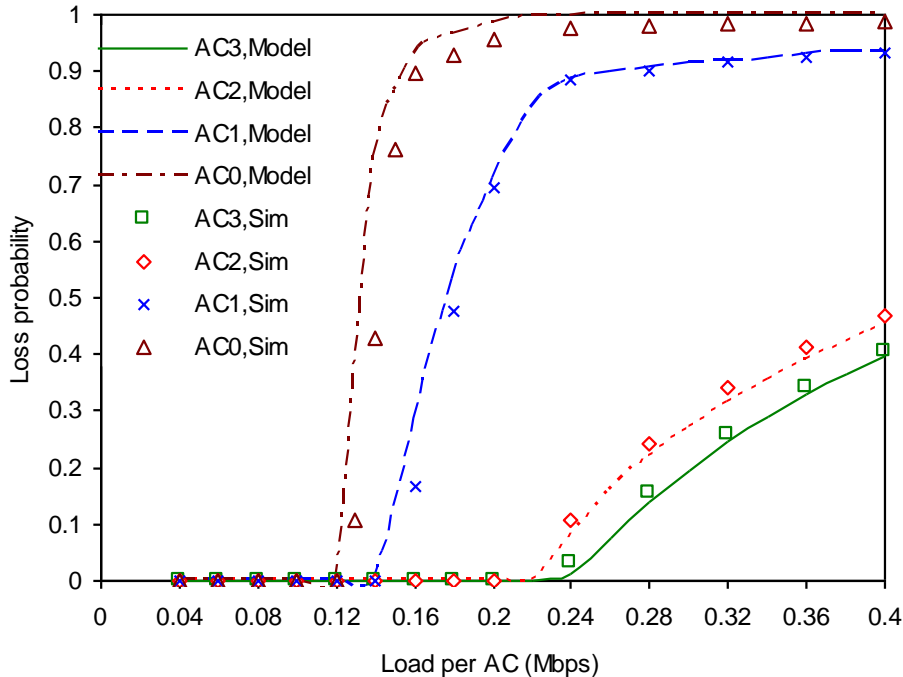
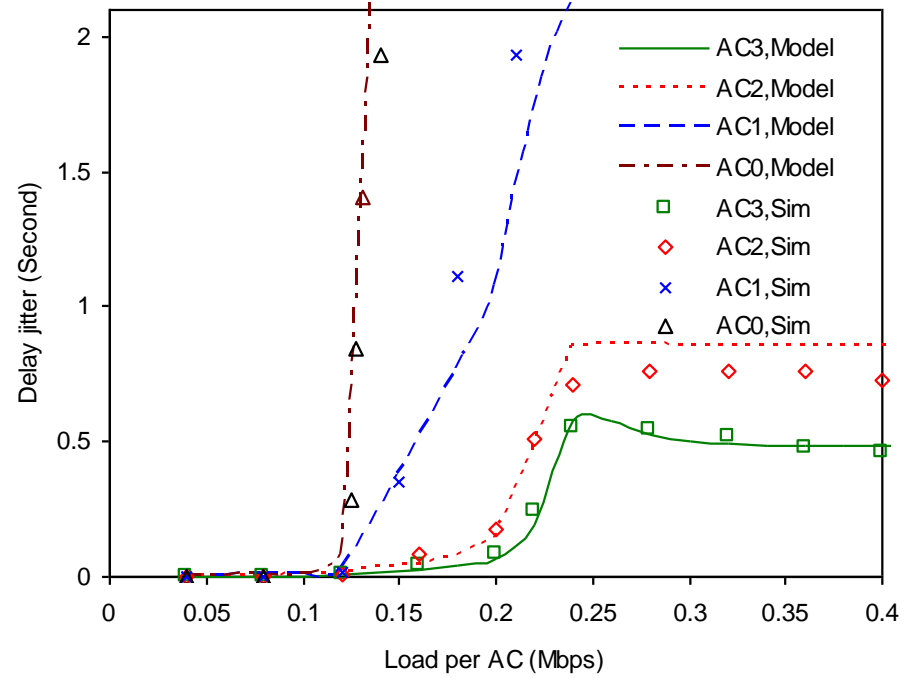
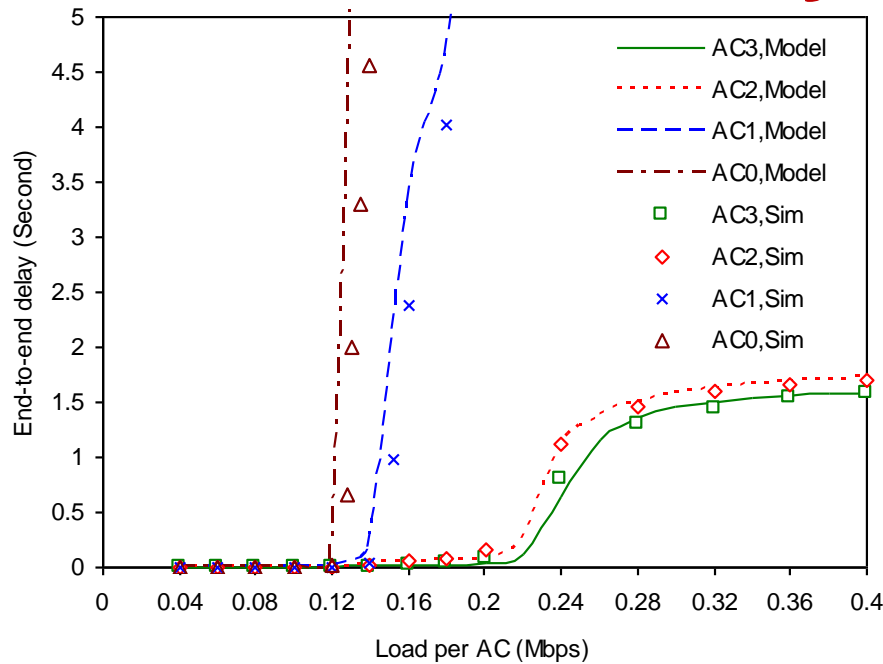
Throughput



❑ The analytical results match the simulations very well.

❑ The throughput differentiation is achieved.

Delay/Jitter/Loss



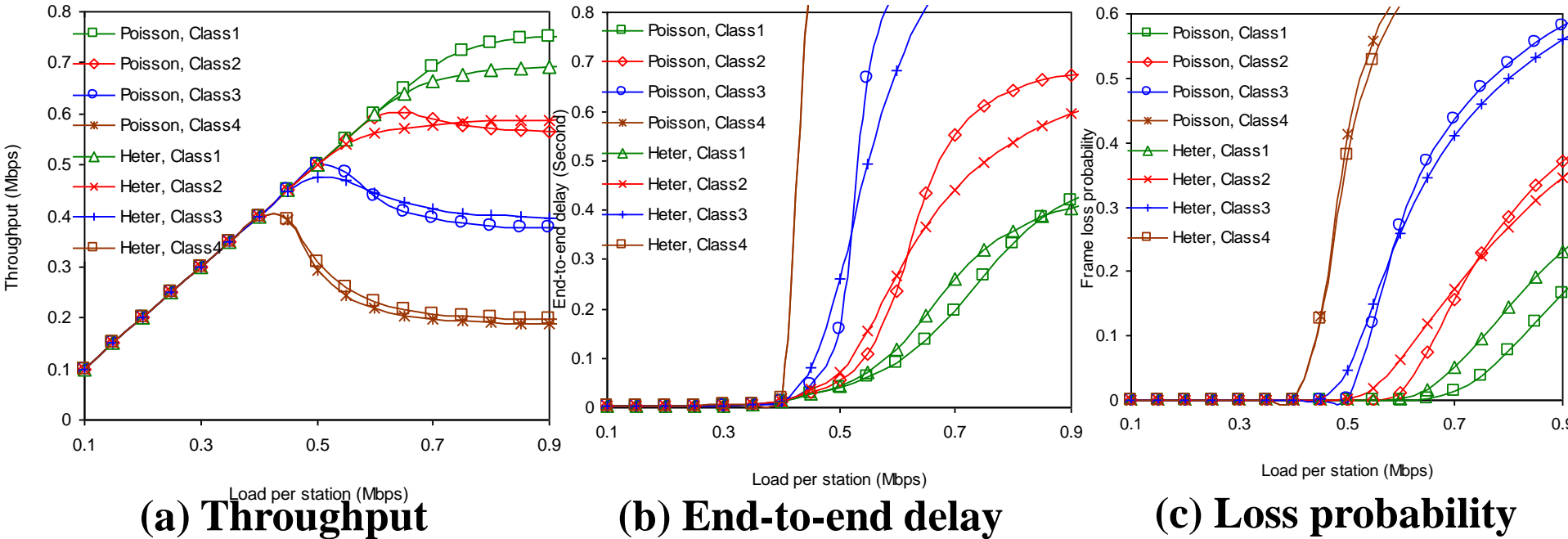
- The analytical results **match the simulations very well.**
- The delay/jitter/loss of the AC_0 and AC_1 is **much larger** compared with the delay of the AC_2 and AC_3 .

Performance Evaluation

- ❑ To investigate the impact of **heterogeneous traffic** on the network performance.
- ❑ We compare the performance measures of heterogeneous classes of stations with that of the homogeneous ones.

	Class 1	Class 2	Class 3	Class 4
Heterogeneous	MMPP	Self-similar	Self-similar	Poisson
Homogeneous	Poisson			

Analytical Results



□ The results show a clear **impact of heterogeneous traffic** on the performance. Under the **heterogeneous traffic**,

- The performance of **C1** station degrades under the medium and heavy loads regions.
- The performance of **C2** and **C3** station deteriorates under the medium loads region.
- The performance of **C4** station is only slightly affected.

Conclusions & Discussions

- ❑ **Heterogeneous traffic models, channel models**
- ❑ **An analytical framework for MAC modelling**
- ❑ **Performance results: throughput, end-to-end delay, jitter, frame loss probability**
- ❑ **Discussion: Model-based admission control, Multi-channel cognitive MAC, 802.11ac/ad/aa very high speed WLANs ...**

References

- ❑ J. Hu, G. Min, and M. E. Woodward, “Performance analysis of a threshold-based dynamic TXOP scheme for intra-AC QoS in wireless LANs”, *Elsevier Future Generation Computer Systems*, DOI: 10.1016/j.future.2013.09.021, 2013.
- ❑ J. Hu, G. Min, W. Jia, and M. E. Woodward, “Comprehensive QoS Analysis of Enhanced Distributed Channel Access in Wireless Local Area Networks”, *Elsevier Information Sciences*, vol. 214, pp. 20-34, 2012.
- ❑ J. Hu, G. Min, and M. E. Woodward, “Performance Analysis and Comparison of Burst Transmission Schemes in Unsaturated 802.11e WLANs”, *Wiley Wireless Communications and Mobile Computing*, vol.12, no.9, pp. 837–848, 2012.
- ❑ G. Min, J. Hu, and M. E. Woodward, “Performance Modelling and Analysis of the TXOP Scheme in Wireless Multimedia Networks with Heterogeneous Stations”, *IEEE Transactions on Wireless Communications*, vo.10, no. 12, pp. 4130 – 4139, 2011.
- ❑ J. Hu, G. Min, W. Jia, and M. E. Woodward, “Admission Control in the IEEE 802.11e WLANs based on Analytical Modelling and Game Theory”, *Proc. IEEE GLOBECOM'09*, 2009.

Performance Evaluation of Cognitive Radio Networks under Common Handoff Queue

Salah Zahed¹, Irfan Awan², Andrea Cullen³

¹The Higher Institute of Comprehensive Professions-Ghadames, Anahda 1, Libya
^{1,2,3}School of Engineering and Informatics, University of Bradford, Bradford, BD7 1DP, UK

{S.M.B.Zahed, I.U.Awan, A.J.Cullen}@bradford.ac.uk

Abstract. Cognitive Radio Networks (CRNs) have been designed to improve spectrum utilization by utilizing temporarily unoccupied spectrum bands in the primary user's licensed spectrum by secondary (unlicensed) users. Even though some spectrum bands in the primary user's licensed spectrum are intensively used, most of the spectrum bands remain underutilized. The introduction of dynamic spectrum access and open spectrum lets the secondary users, supported by cognitive radios capabilities; opportunistically utilize the unoccupied spectrum bands. However, in case of the primary user returning to a band occupied by the secondary user, the occupied spectrum band must be vacated immediately by handing off the secondary user's call to another idle spectrum band. Multiple spectrum handoffs can severely degrade quality of service (QoS) for the secondary users. To reduce the multiple handoffs, an effective spectrum handoff procedure should be initiated to preserve a required level of QoS for secondary users. Moreover, it enables the channel clearing while searching for target vacant channel(s) for completing unfinished transmission. This paper proposes spectrum handoff scheme with shared handoff queue in order to reduce the handoff delay and the total service time. The proposed scheme has been modelled using a preemptive resume priority (PRP) M/M/1 queue. The performance of proposed handoff schemes has been evaluated and compared against the existing spectrum handoff schemes. Experimental results show that the scheme developed here outperforms the existing schemes in terms of average handoff delay and total service time under various traffic arrival rates.



Performance Evaluation of Cognitive Radio Networks under Common Handoff Queue

Salah Zahed¹, Irfan Awan², Andera Cullen³

¹The Higher Inst. of Comprehensive Professions-Ghadames, 1 Anahda, Libya

^{1,2,3}School of Engineering and Informatics, University of Bradford
Bradford, BD7 1DP, UK

Introduction

- Cognitive Radio (CR)
 - Increasing Wireless Applications & Services
 - Most available Radio Spectrum Resources well allocated (Static)
 - Spectrum Underutilization
 - Dynamic Spectrum Access (~CR)
- Primary Users (PUs) → (Licensed Users)
- Secondary Users (SUs) → (Unlicensed Users)
- Spectrum Priority

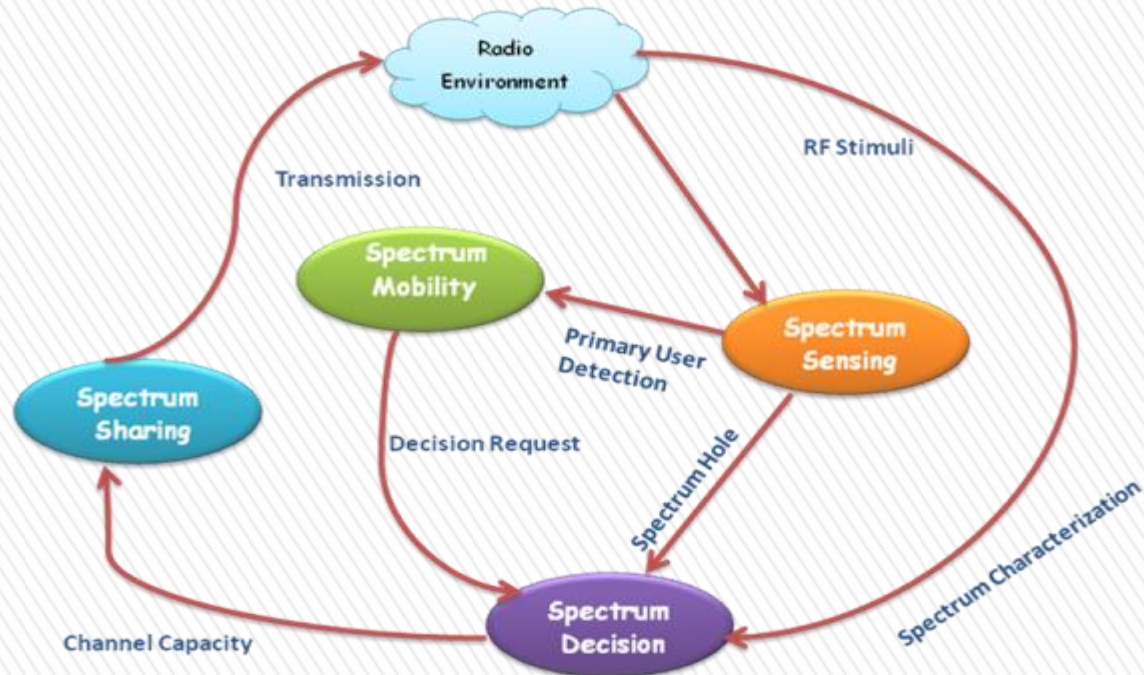


Introduction

- Cognitive Radio Cycle & Functionalities

1. Spectrum Sensing
2. Spectrum Decision

3. Spectrum Sharing
4. Spectrum Mobility



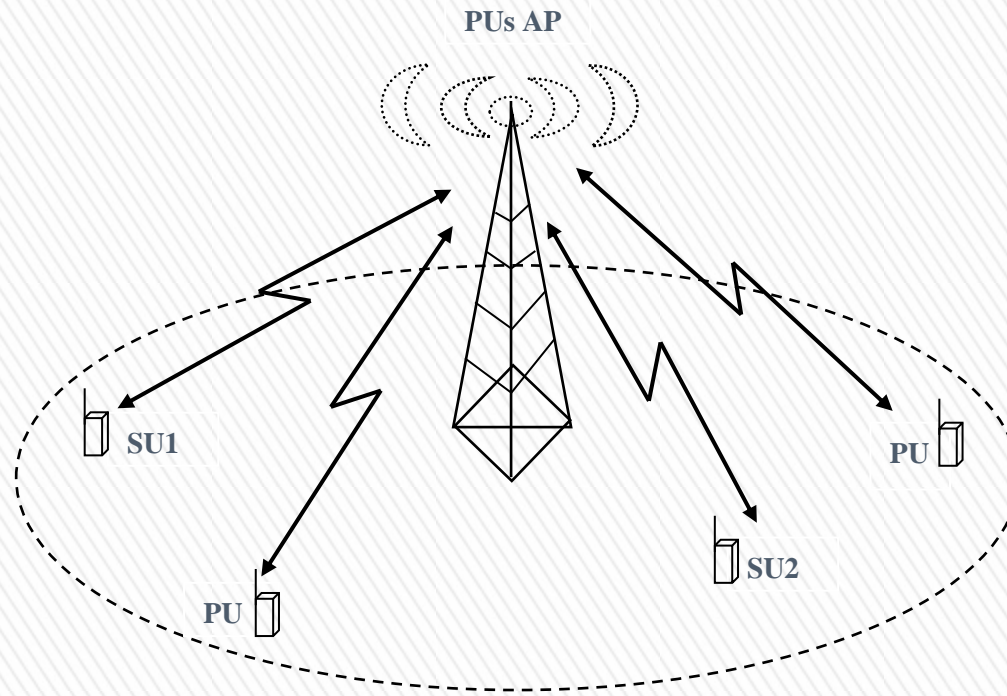
Spectrum Handoff Decision

- CH selection Decision Timing: Target CH for Spectrum Handoff will be selected Based-on
 - Proactive-Decision (Long-term Observations)
 - Reactive-Decision (Mostly, using On-demand Wideband Sensing)

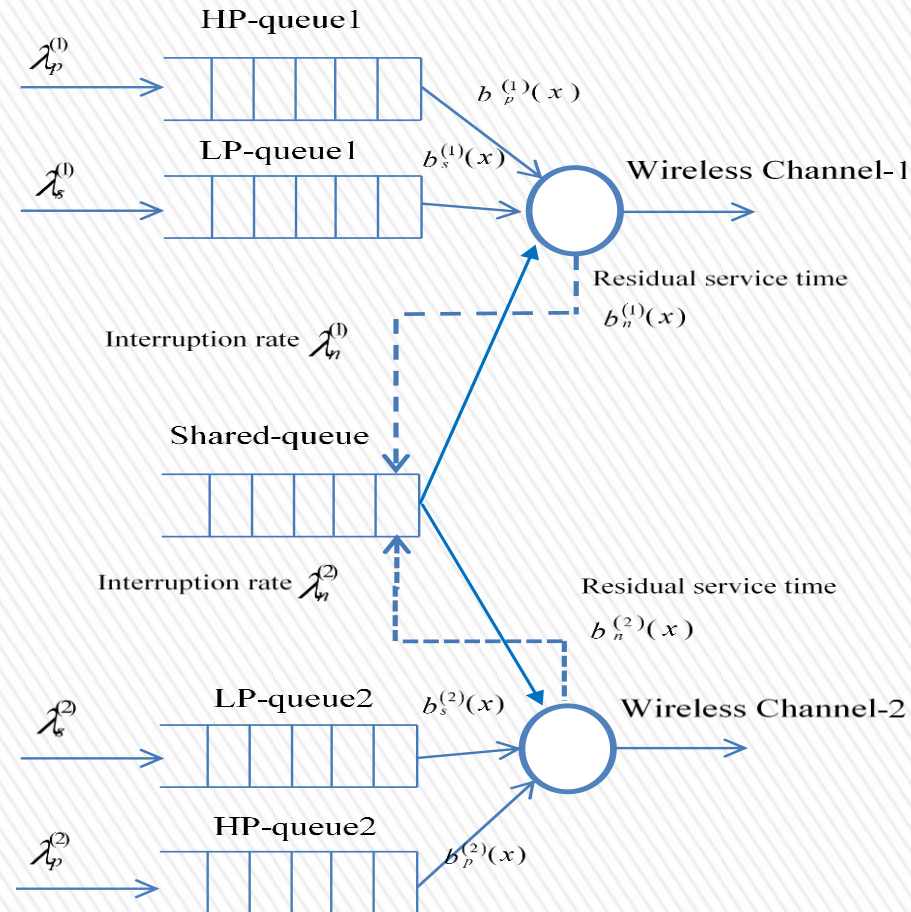


System Model

- Central Entity and SUs & PUs:



Queuing Network Model



Assumptions

- SUs & PUs Arrival Rates (Poisson)
 $\lambda_s^{(k)}$ & $\lambda_p^{(k)}$, respectively
- Exponentially distributed Service Times
with Rates $\mu_s^{(k)}$ & $\mu_p^{(k)}$, respectively
- Pre-emptive Resume Priority (PRP) M/M/1
Queuing Network Model
- Handoff Rate (Poisson) $\lambda_n^{(k)}$



Assumptions

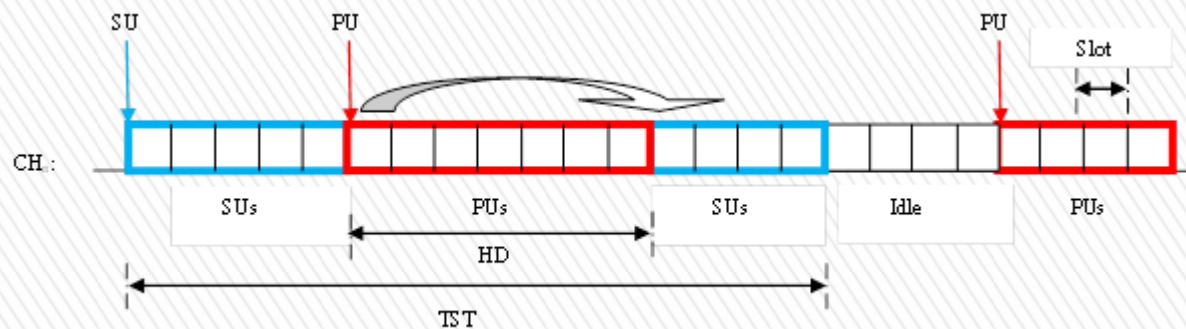
- Interrupted User joins the tail of Shared-Queue
- If an Idle CH exist, then Transmit
- Otherwise, he waits and will be handled by 1st available CH, but before any other SUs in the Low-Priority Queues of both CHs.



Spectrum Handoff

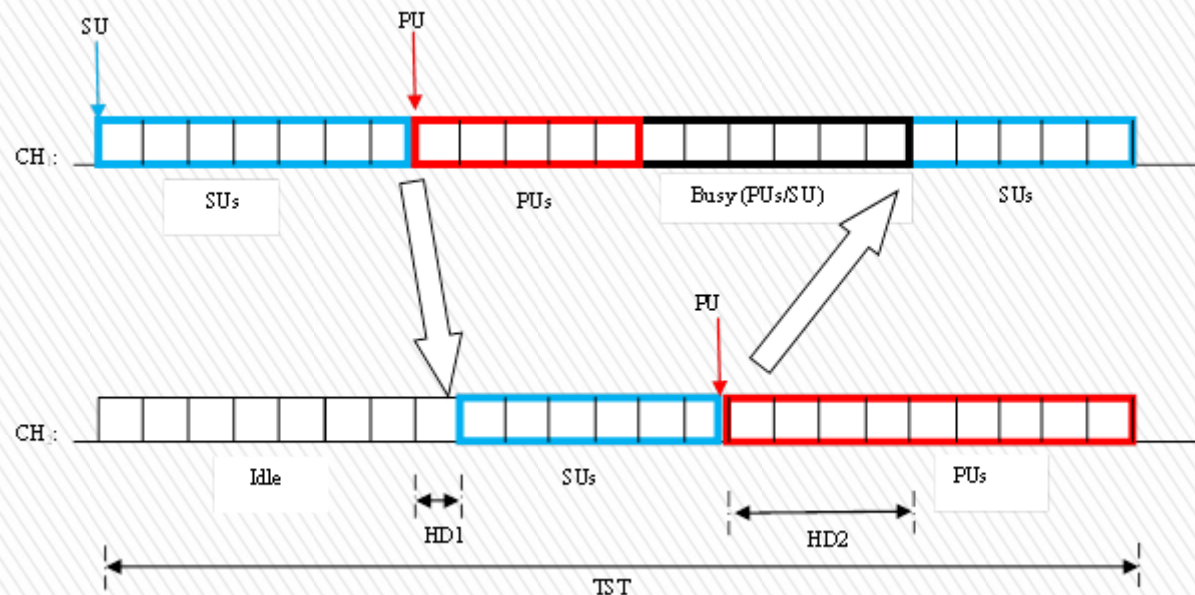
Example 1

- Non-switching-handoff (Non-handoff/non-hopping)



Spectrum Handoff Example 1

- Switching-handoff (Handoff/hopping)



Simulation Experiments

- Discrete Event Simulator (MATLAB)
- Simulation Setup

Simulation parameters	Symbol	Value(s)
PU arrival rate	λ_p	0.05-0.30
SU arrival rate	λ_s	0.15
PU service rate	μ_p	0.60
SU service rate	μ_s	0.40
PU arrival rate Increment	Δ	0.05

- Performance Calculations

$$E[T] = E[X_s] + E[N] * \frac{(\sum \text{Handoff Delays})}{\text{Number of Interruptions}} + T_{sw} + T_{ha}$$

$$E[N] = \frac{\text{Number of Interruptions}}{\text{Number of SUs Arrivals}}$$

$E[T]$: Total Service Time

$E[X]$: Mean Service Time

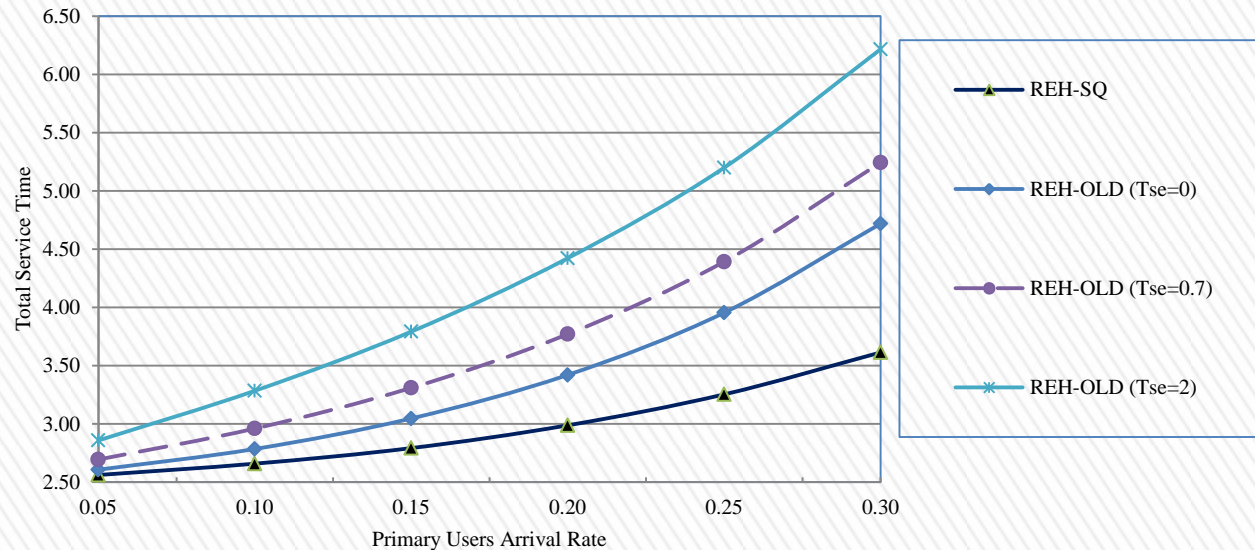
$E[N]$: Avg. No of Handoffs

T_{sw} : Switching Time (Assumed “0”)

T_{ha} : Handshaking Time (Assumed “0”)



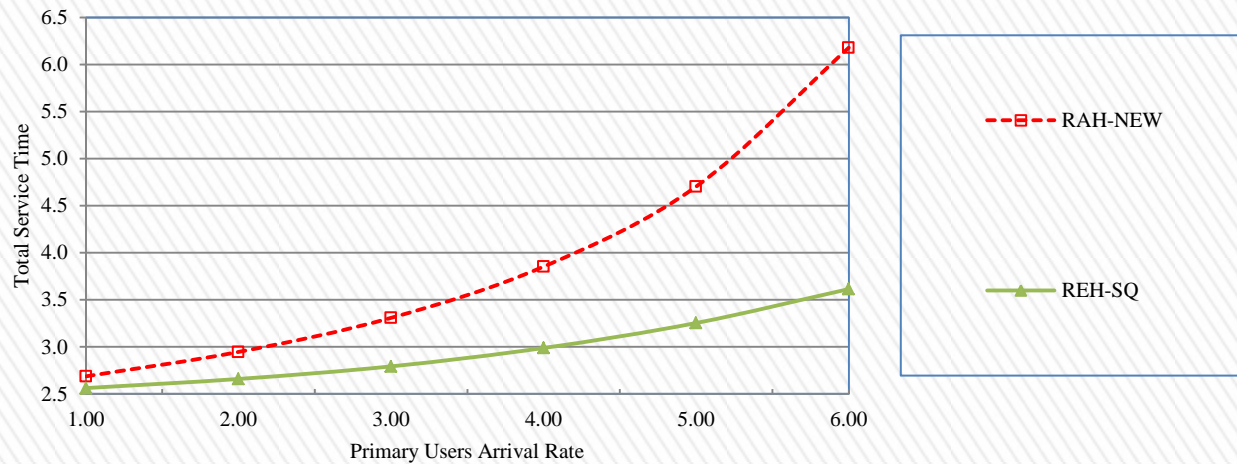
Results 1



- REH-SQ: Reactive with Shared-Queue (Proposed)
- REH-OLD: Reactive Old
- T_{se} : Sensing Time



Results 2



- RAH: Random
- REH-SQ: Reactive with Shared-Queue (Proposed)



Conclusion

- Minimizes the Handoff Delay and hence the Total Service Time of the SUs
- The Proposed Scheme outperforms the Random and existing Reactive Schemes



Any Question ???

Thank you



Performance Evaluations for Situational Awareness Based on Scanning Techniques

Orabi Shurrab, Evangelos Spatharas, Jules Pagna Disso, and Irfan Awan

University of Bradford, Bradford, West Yorkshire, United Kingdom
{omfshurr, E.Spatharas}@student.bradford.ac.uk
i.u.awan@bradford.ac.uk

EADS Innovation Works, Quadrant House, Celtic view, Coedkernew, Newport, UK
julesferdinand.pagna@eads.com

Abstract. Software-based scanning tools are commonly used in experimental research to perform various tasks including penetrator tasks. Here we start a discussion reviewing the problem of using software-based scanning tools, highlighting interesting issues that pose some threats to common beliefs. We started comparing the operator-requested scanning profile against the real behaviour of commonly used port scanning techniques. From this we will find the best performing scanning techniques in terms of speed, stealthiness and accuracy. To analyse the behaviour of the best performing scanning techniques we further aim to perform tests under different size of networks using modified switches (RTT, hostgroup, paralism, and randomness). Preliminary results show notable differences in some cases, opening the way to interesting discussions and further investigations. Furthermore, some of port scanning techniques clearly distinguish between some port status in some experiments and not others. No scanning techniques performed best consistently in all parameters but either it performed well in speed and lacked of accuracy and stealthiness or vice versa.



Performance Evaluation for Situational Awareness Based on Scanning Techniques

Orabi Shurrab, Evangelos Spatharas and Prof. Awan

University of Bradford

Department of Computing Informatics & Media

{omfshurr, e.spatharas} @student.bradford.ac.uk

i.u.awan@bradford.ac.uk

Dr. Jules Pagna Disso

EADS Innovation Works

julesferdinand.pagna@eads.com

- Introduction
- The problem
- Motivation
- Type of port scanning techniques.
- Performance evaluation of port scanning techniques.
- Conclusion
- Future work
- Questions



Introduction

- Scanning tools are commonly used by network administrators and intruders to perform various tasks including penetrations to computer networks..
- They are used to enhance situational awareness for network administrators by point out important information including:
 - Vulnerable services
 - Port status
 - Other valuable information.
- They are used to gain valuable information in order for hackers to launch an attack against computer networks.

The problem

- Port scanning provide quick understanding for network administrators about ports status but we don't know how fast one can scans a network without degrading the quality of information.
- Number of port scanning techniques are available but which one perform best under various contexts including:
 - Different network sizes
 - Speed
 - Round Trip Time (RTT).



Motivation

- To compare different port scanning techniques in term of :
 - Speed : The time it takes to complete the scanning tasks.
 - Quality of information about port status.
 - Stealthiness : Is the number of packet detected by NIDS (Snort)
- To understand the behaviour of port-scanning techniques under various:
 - Network sizes
 - Requested profiles Including the Round Trip Time (RTT).
- To analyse the relations between the speed and accuracy.



Port Scanning Techniques 1

- SYN

Source	Destination	Protocol	Info
192.168.110.129	82.219.166.43	TCP	62317 > http [SYN] Seq=0 Win=2048 Len=0 MSS=1460
82.219.166.43	192.168.110.129	TCP	http > 62317 [SYN, ACK] Seq=0 Ack=1 Win=64240 Len=0 MSS=1460
192.168.110.129	82.219.166.43	TCP	62317 > http [RST] Seq=1 Win=0 Len=0

- Connect

Source	Destination	Protocol	Info
192.168.110.129	82.219.166.43	TCP	35705 > http [SYN] Seq=0 Win=5840 Len=0 MSS=
82.219.166.43	192.168.110.129	TCP	http > 35705 [SYN, ACK] Seq=0 Ack=1 Win=642
192.168.110.129	82.219.166.43	TCP	35705 > http [ACK] Seq=1 Ack=1 Win=5840 Len=
192.168.110.129	82.219.166.43	TCP	35705 > http [RST, ACK] Seq=1 Ack=1 Win=584

- FIN

Source	Destination	Protocol	Info
192.168.110.129	82.219.166.43	TCP	58705 > http [FIN] Seq=1 Win=2048 Len=0
192.168.110.129	82.219.166.43	TCP	58706 > http [FIN] Seq=1 Win=3072 Len=0

- Null

Source	Destination	Protocol	Info
192.168.110.129	82.219.166.43	ICMP	Standard query response, no such name
192.168.110.129	82.219.166.43	TCP	51781 > http [<None>] Seq=1 Win=2048 Len=0
192.168.110.129	82.219.166.43	TCP	51782 > http [<None>] Seq=1 Win=2048 Len=0

- XMASS

Source	Destination	Protocol	Info
192.168.110.129	82.219.166.43	TCP	46509 > http [FIN, PSH, URG] Seq=1 Win=3072
192.168.110.129	82.219.166.43	TCP	46510 > http [FIN, PSH, URG] Seq=1 Win=2048



Port Scanning Techniques 2

- Ack

Destination	Protocol	Info
192.168.110.129	DNS	Standard query response, no such name
192.168.110.130	TCP	35458 > http [ACK] Seq=1 Ack=1 Win=3072 Len=0
192.168.110.129	TCP	http > 35458 [RST] Seq=1 Win=0 Len=0

- Maimon

Destination	Protocol	Info
192.168.110.129	DNS	Standard query response, no such name
192.168.110.130	TCP	38901 > http [FIN, ACK] Seq=1 Ack=1 Win=2048 Len=0
192.168.110.129	TCP	http > 38901 [RST] Seq=1 Win=0 Len=0

- Window

Destination	Protocol	Info
192.168.110.129	DNS	Standard query response, no such name
192.168.110.130	TCP	35458 > http [ACK] Seq=1 Ack=1 Win=3072 Len=0
192.168.110.129	TCP	http > 35458 [RST] Seq=1 Win=0 Len=0



- **Opened**

Actively accepting connections.

- **Closed**

Receives and responds to probe packets, but there is no application listening on it.

- **Filtered**

Cannot determine whether the port is open because packet filtering .

- **Open-filtered**

Unable to determine whether a port is open or filtered.

- **Closed-filtered**

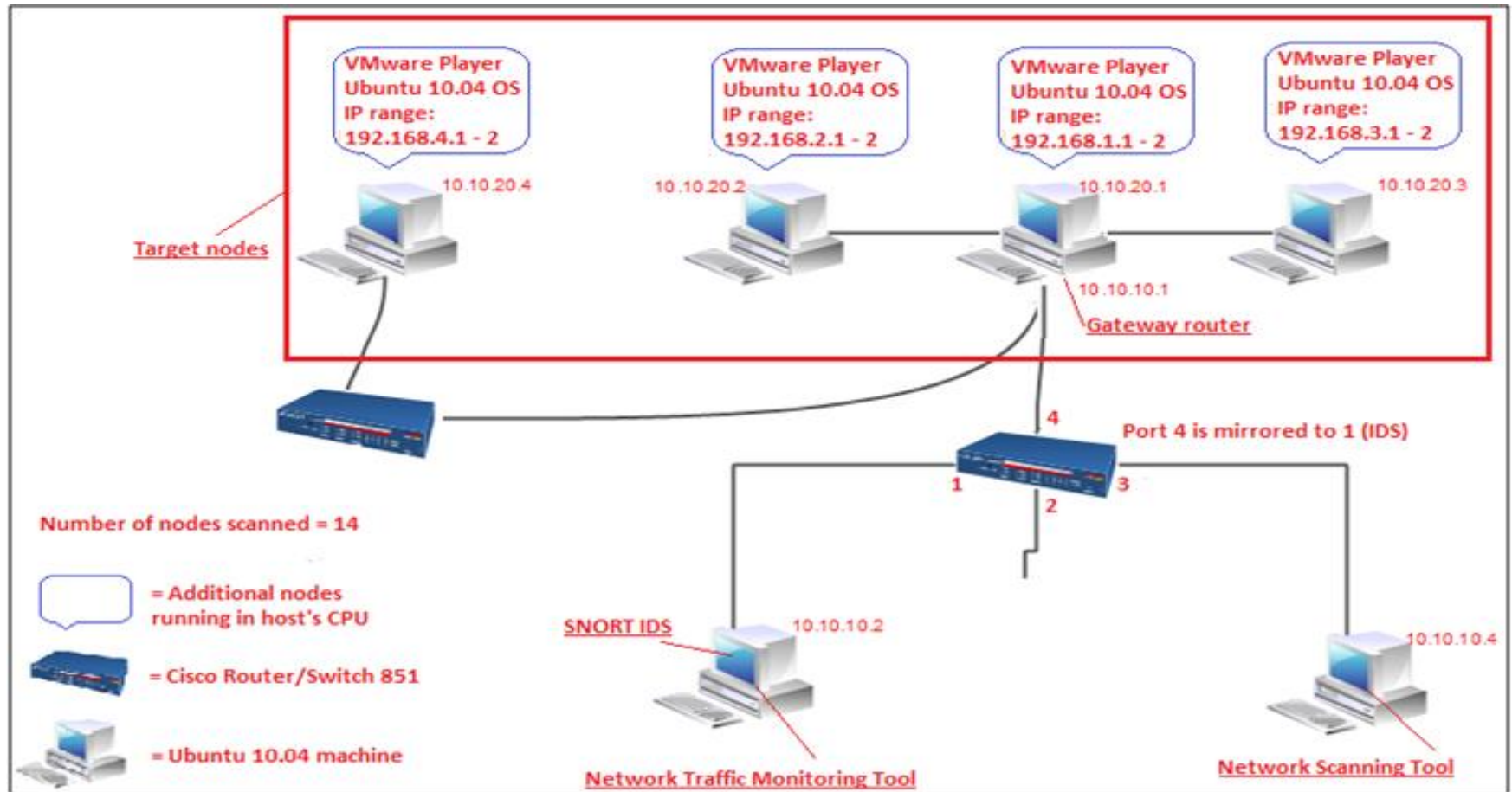
Unable to determine whether a port is closed or filtered.

- **Unfiltered**

Unable to determine whether it is open or closed.



Performance Evaluation (14 Nodes)





Performance Evaluation (Result)

	SYN	Connect	FIN	NULL	XMASS	ACK	Maimon	Windows
Time	16.86	15.89	27.82	29.81	29.81	23.72	24.19	15.99
Snort Alerts	12	12	1	2	2	3	1	1
Open Ports	9/9	9/9	-	-	-	-	-	0/9
Closed Ports	28/28	28/28	4/28	4/28	4/28	-	4/28	37/28
Filtered Ports	13,963/ 13,963	13,963/ 13,963	-	-	-	13,963/ 13,963	0/13,963	13,963/ 13,963
Open-Filtered Ports	-	-	13,996	13,996	13,996	-	13,996	-
Unfiltered Ports	-	-	-	-	-	37/37	-	-

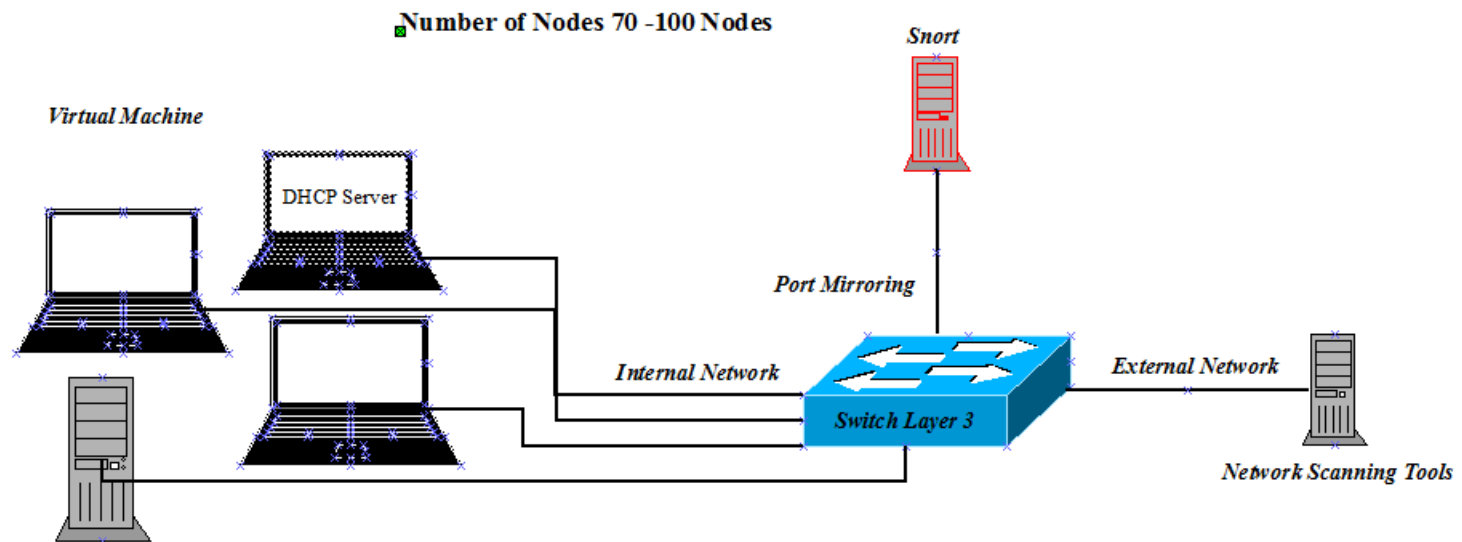
Conclusion

- None of the port-scanning techniques are best in all parameters .
- Some techniques performed well in speed and lacked of accuracy.
- While other showed to be the stealthiest but unfortunately provide misleading information about the actual port status.
- Poor selection of port scanning techniques contribute to poor Situational Awareness (SAW).



Future Work

- We further aim to perform tests under different sizes of computer networks using modified switches of Nmap such as
 - Changing the Round Trip Times (RTT)
 - Randomized and non randomized network scanning.





UNIVERSITY OF
BRADFORD
MAKING KNOWLEDGE WORK

Questions & Answers

Thank you for listening

Performance Evaluation of Cognitive Radio Networks under Reactive-Decision Spectrum Handoff Scheme

Salah Zahed¹, Irfan Awan², Andrea Cullen³

¹The Higher Institute of Comprehensive Professions-Ghadames, Anahda 1, Libya
^{1,2,3}School of Engineering and Informatics, University of Bradford, Bradford, BD7 1DP, UK

{S.M.B.Zahed, I.U.Awan, A.J.Cullen}@bradford.ac.uk

Abstract. In cognitive radio networks (CRNs), spectrum handoff procedure will be initiated and performed whenever the spectrum owner (primary user) returns to its licensed band. Accordingly, operating secondary user (SU) working in that band has to leave the band immediately and transfer to another idle band. This can degrade the quality of on-going communication. In fact, simple spectrum handoff strategy can accomplish reasonable performance for various communication desires, whereas more innovative adaptive strategies are essential to achieve the highest benefit. Based-on on-demand wideband spectrum sensing, the secondary users reactively select the goal channel for spectrum handoff in order to finish the interrupted transmission. However, the cost of spectrum sensing is, longer handoff delay, especially when the number of channels to be sensed is high. Instead of the sensing method, we proposed that secondary users rely on information provided by a central controller regarding the instantaneous secondary users' queues length in order to select the channel with the shortest queue length for resuming unfinished transmission. The proposed scheme is modeled by a Preemptive Resume Priority (PRP) M/M/1 queuing theory. By introducing our novel idea, the simulation study shows that handoff delay could be reduced compared with the methods that use sensing techniques. Moreover, when we compare the new suggested spectrum handoff scheme with the existing schemes in the literature, the new scheme shows an improvement in terms of the cumulative handoff delay at most situations.



Performance Evaluation of Cognitive Radio Networks under Reactive-Decision Spectrum Handoff Scheme

Salah Zahed¹, Irfan Awan², Andera Cullen³

¹The Higher Inst. of Comprehensive Professions-Ghadames, 1 Anahda, Libya

^{1,2,3}School of Engineering and Informatics, University of Bradford
Bradford, BD7 1DP, UK

Introduction

- Cognitive Radio (CR)
 - Increasing Wireless Applications & Services
 - Most available Radio Spectrum Resources well allocated (Static)
 - Spectrum Underutilization
 - Dynamic Spectrum Access (~CR)
- Primary Users (PUs) → (Licensed Users)
- Secondary Users (SUs) → (Unlicensed Users)
- Spectrum Priority

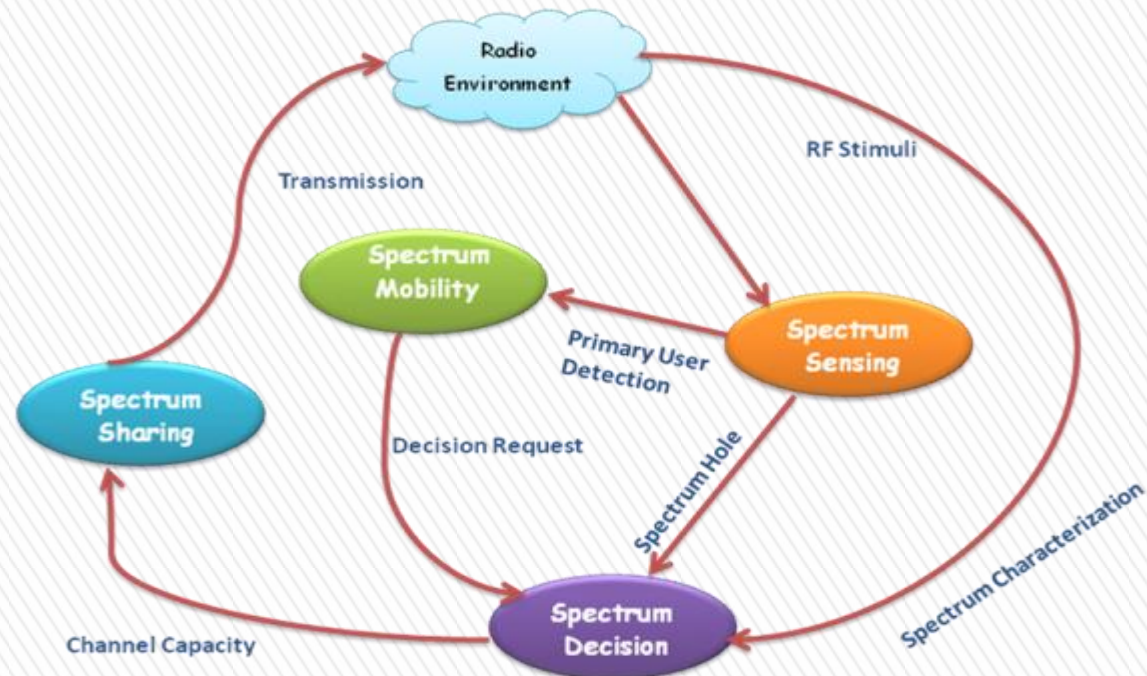


Introduction

- Cognitive Radio Cycle & Functionalities

1. Spectrum Sensing
2. Spectrum Decision

3. Spectrum Sharing
4. Spectrum Mobility



Spectrum Handoff Decision

- CH selection Decision Timing: Target CH for Spectrum Handoff will be selected Based-on:
 - Proactive-Decision (Long-term Observations)
 - Reactive-Decision (Mostly, using On-demand Wideband Sensing)



Sensing-Based Schemes

- Sensing Time (T_{se}): Time Period that Handoff SUs spend until finding an Idle CH
- Sensing Time plays a major Role in Sensing-Based Schemes
- If the No of CHs to be sensed is high, then Sensing Time is high too
- Sensing a few CHs makes it hard to find Idle CHs



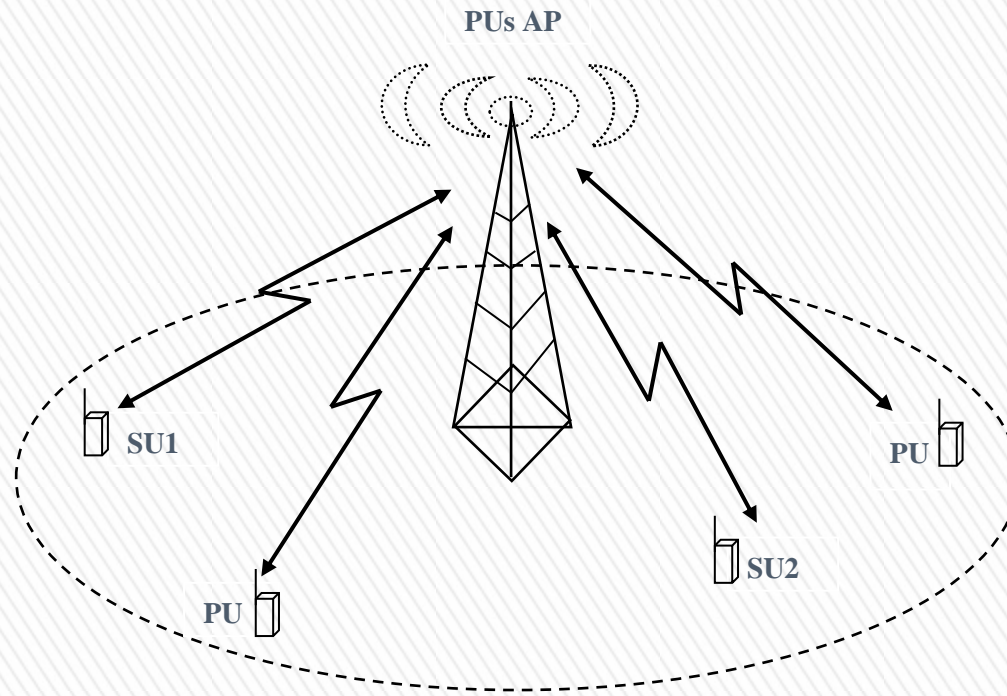
Proposed Handoff Scheme

- Avoid Sensing Process
- Target CH selection is Based-on the shortest instantaneous SUs' Queues Length
- Information regarding Decision Factor (Queues' Length) are provided by a Central Controller

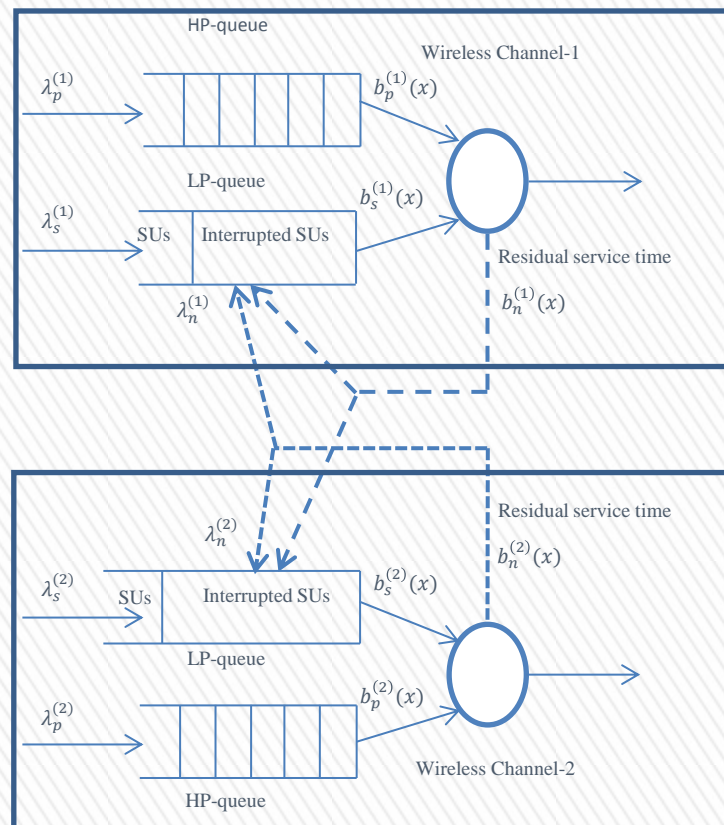


System Model

- Central Entity and SUs & PUs:



Queuing Network Model



Assumptions

- SUs & PUs Arrival Rates (Poisson) $\lambda_s^{(k)}$ & $\lambda_p^{(k)}$ to CH k , respectively
- Exponentially distributed Service Times with Rates $\mu_s^{(k)}$ & $\mu_p^{(k)}$, respectively
- Pre-emptive Resume Priority (PRP) M/M/1 Queuing Network Model
- Handoff Rate (Poisson) $\lambda_n^{(k)}$



Assumptions

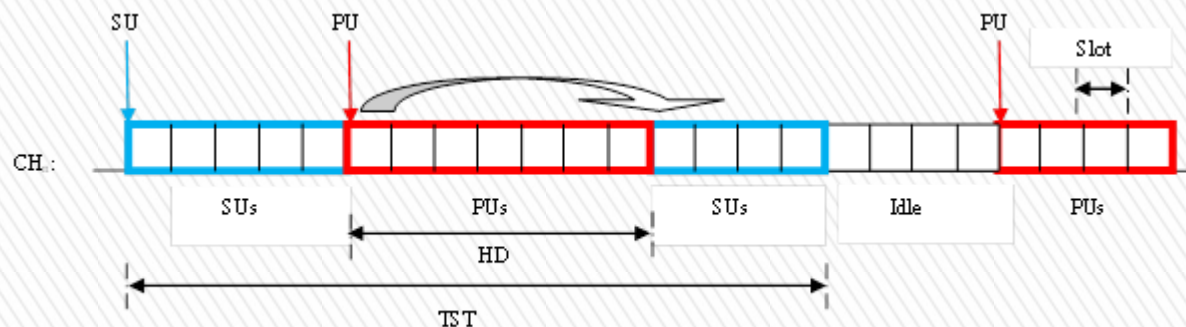
- Interrupted User joins the CH that has the shortest Queue Length
- If Both CHs have same Queue Length, then User waits in Its Operating CH till becomes Idle



Spectrum Handoff

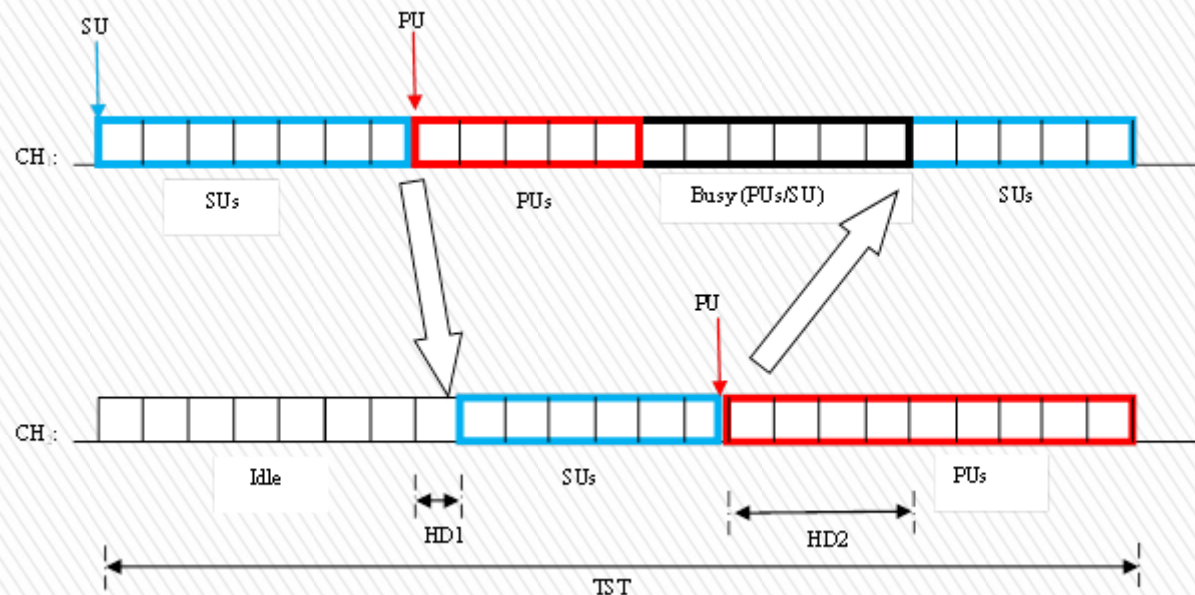
Example 1

- Non-switching-handoff (Non-handoff/non-hopping)



Spectrum Handoff Example 1

- Switching-handoff (Handoff/hopping)



Simulation Experiments

- Discrete Event Simulator (MATLAB)
- Simulation Setup

Simulation parameters	Symbol	Value(s)
PU arrival rate	λ_p	0.05-0.30
SU arrival rate	λ_s	0.15
PU service rate	μ_p	0.60
SU service rate	μ_s	0.40
PU arrival rate Increment	Δ	0.05

- Performance Calculations

$$E[D_{cum}] = E[N] \left(\frac{(\sum \text{Handoff Delays})}{\text{Number of Interruptions}} + T_{sw} \right)$$

$E[D_{cum}]$: Cumulative Handoff Delay

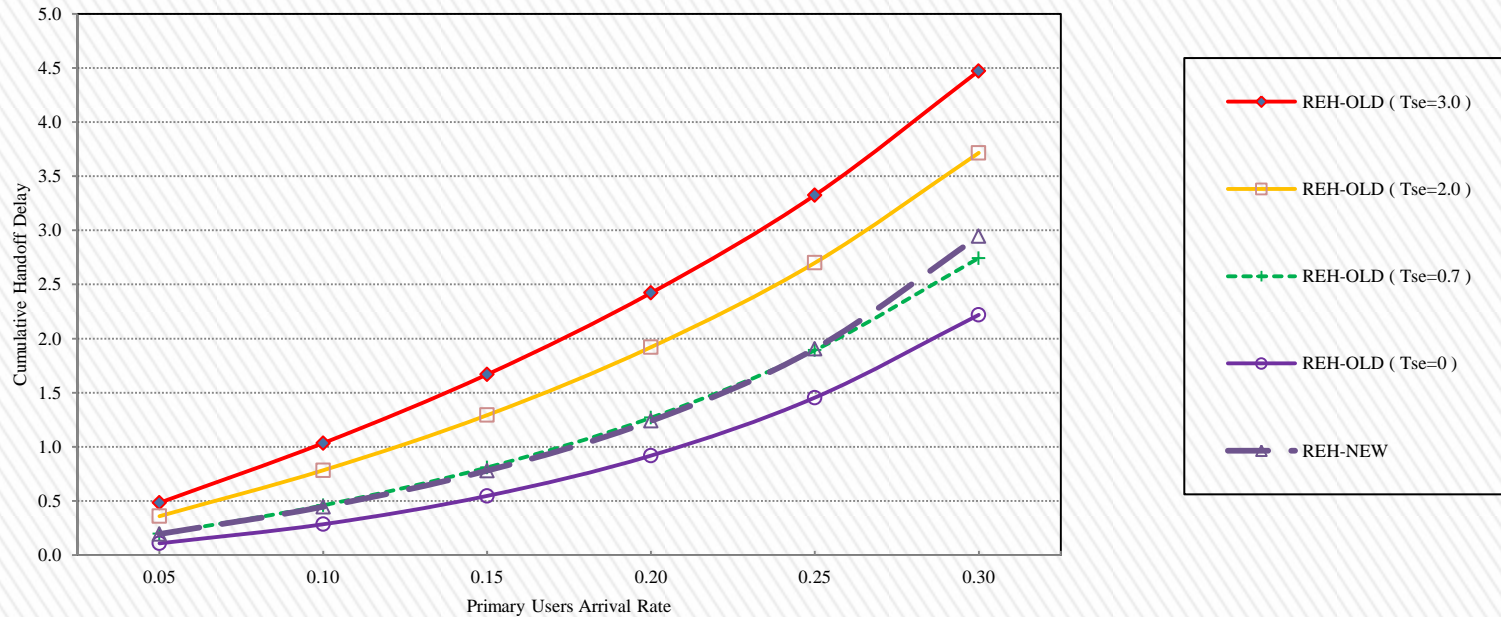
T_{sw} : Switching Time (Assumed “0”)

$E[N]$: Avg. No of Handoffs

$$E[N] = \frac{\text{Number of Interruptions}}{\text{Number of SUs Arrivals}}$$



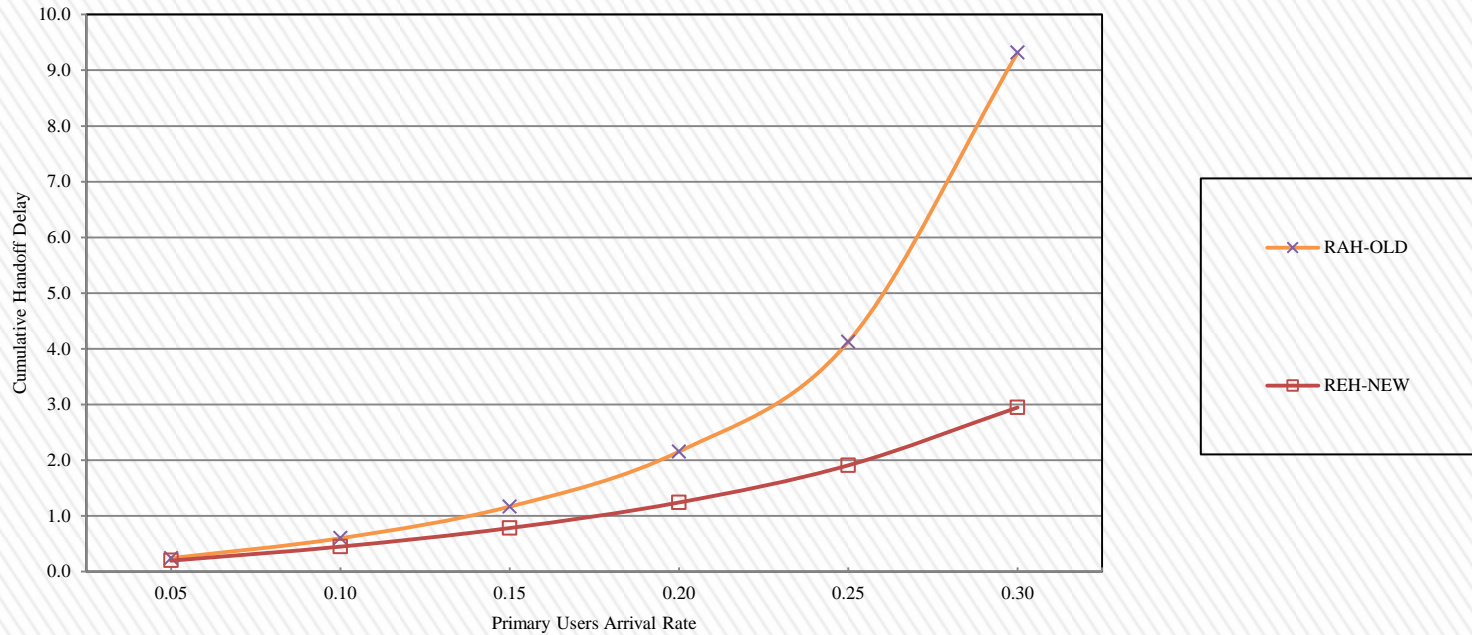
Results 1



- REH-NEW: Reactive New (Proposed)
- REH-OLD: Reactive Old
- T_{se} : Sensing Time



Results 2



- RAH-OLD: Random Old
- REH-NEW: Reactive New (Proposed)



Conclusion

- Minimizes the Cumulative Handoff Delay and hence the Total Service Time of the SUs
- The Proposed Spectrum Handoff Scheme outperforms the Random and Sensing-Based Schemes especially at relatively High Values of Sensing Time



Any Question ???

Thank you



Performance Analysis of a Pre-emptive Priority Soft Handoff Scheme for Wireless based Cellular Mobile Networks with Busty Traffic

M.I. Bello

School of Engineering and Informatics
University of Bradford

Abstract. With current trends in cellular network, where real-time and non-real-time services are accommodated, it is essential to design efficient handoff in order to preserve the quality of service (QoS) by giving priority to delay sensitive services. The IS-95 standard came with the flexibility of allowing mobile users (MU) connecting to multiple cells during soft handoff. Although it improves the performance of user communications, it also results in valuable radio resources that may be required by other users being held unnecessarily, thereby reducing the capacity of the network in accepting handoff request and new incoming calls. Furthermore, the multiple connections increase the interference in those cells and cause more energy to be expended. In this work, we propose an analytical model with busty traffic for handoff scheme of integrated real-time and non-real-time calls in a wireless mobile network that frees some channels for real-time handoff request calls and reduce the interference arising from multiple connections of a user in soft handoff. Real-time handoff calls are giving pre-emptive and priority privileges over other services and when interference level is high, the service given to non-real time handoff calls is suspended and the calls are returned into the queue.

Keywords: Key words: Cellular Network, Handoff

Performance Analysis of a Pre-emptive Priority Soft Handoff Scheme for Wireless based Cellular Mobile Networks with Busty Traffic.

M.I. Bello

NetPEn - Networks and Performance Engineering Research Group
Informatics Research Institute (IRI)
University of Bradford

Outline

□ Introduction

- Cellular networks
- Handoff
- Interference
- Objectives

□ The Proposed Model

□ Further Work

Introduction – Wireless Cellular Network

- ❑ Support multiple types of services simultaneously (real time and non real time).
 - transmission of real-time service are sensitive to delay and interruptions.
 - While non-real-time service are delay insensitive.
- ❑ In wireless cellular networks mobility is the most important feature and continuous service is achieved by supporting handoff from one cell to another.

Introduction - Handoff

- Handoff is the process of transferring an active call or data session from one cell in a cellular network to another or from one channel in a cell to another
- Handoffs are broadly classified into two categories
 - Hard handoff – break before make
 - Soft handoffs – break after make

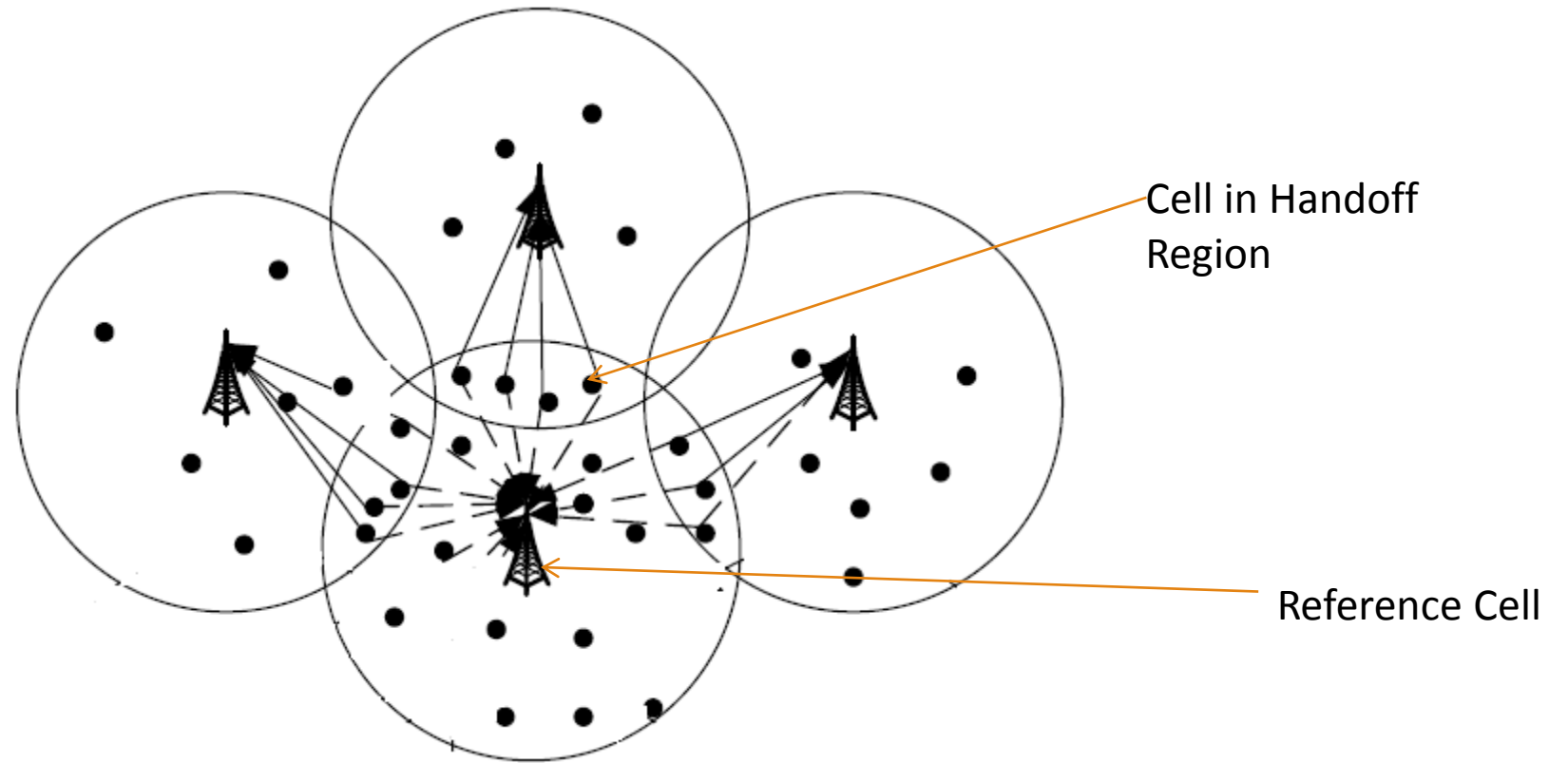
Soft Handoff

- ❑ Mobile units transmit to and receive the same signal from several base stations simultaneously
- ❑ Advantage - Higher performance than hard handoff
 - increases the cell coverage and system capacity
 - Use of same frequency all over the covered area Simplifies system planning.
- ❑ Disadvantages
 - increases interference seen in the system
 - decreases the global availability of radio resources

Introduction - Interference

□ Interference – disrupt signals as it travels along a channel

- Arises due to individual contributions by all operational users in the system. This implies that interference level will increase with increase in number of users in the system.
- Normally, to overcome this interference, the transmitting powers are increased, and thereby more energy is expended.
- When the interference level in a cell become stronger than the maximum transmitted power of at least one ongoing mobile station, the cell is said to be overloaded and cannot admit more mobile users.



Multiple Connections in Soft Handoff

Introduction – Objectives

- ❑ To develop reliable, efficient, and cost-effective analytical tools for soft handoff in wireless cellular networks with bursty traffic.
- ❑ To use the analytical model in investigating the performance characteristics and development of efficient resource allocation schemes.

The proposed Model

- Model consists of;
 - Homogenous cells with fixed number of channels C
 - Four types of calls originating real-time, originating non-real time, Real time handoff and non real time hand off calls
 - Two queues RHQ and NHQ for real-time and non real time handoff calls respectively
- The traffic characterised by bursty generalised exponential (GE) inter-arrival and service times and finite capacity to be used in the modelling
- Some channels (Guard channels) are reserved for handoff calls and the remaining channels are shared equally by handoff calls and the new calls.

-
- ❑ Originating real-time and non-real-time service calls – will be served only if there are free channels available in SC otherwise blocked.
 - ❑ real-time service handoff calls – high priority and can be served even if there is no idle channel by pre-empting one of the current non-real-time handoff calls, if there is at least one ongoing non-real-time handoff service call.
 - The interrupted non-real-time service call returns back to NHQ and waits for an idle channel to be served based on the first-in-first-out rule.
 - Call is queued in RHQ when it cannot get the service. When RHQ is full, handoff request is blocked.

-
- Handoff requests waiting in RHQ is dropped when the mobile user moves out of handoff area before it gets the service.
 - Non real-time service handoff calls – to be served if there is idle channel in the system.
 - Sent to NHQ when all channels are busy
 - To reduce the interference level, it is assumed that disconnecting mobile stations with multiple connections from the cell for a period they maintains multiple connections may not affect their quality of service. Therefore, all non real time communications in the handoff region are stopped and the units sent to NHQ when the interference level is high

Further Work

- The model to be represented and solved by using a multi-dimensional Markov chain
- Performance Metrics – the blocking probabilities of originating calls, the probability of forced termination and the average delays experienced in the system to be calculated numerically.

Performance Modelling and Analysis of Network on Chip under m-port n-tree Bursty Traffic with Virtual Channels

Hatem Ibrahim^{1,2} and Irfan Awan²

¹ University of Cairo, Egypt

² University of Bradford, England

Abstract. In large network system, the Network On Chip (NOC) is a solution to increase the communication requirements instead of bus-based interconnect network. A number of different topologies have been proposed for MPSoC platforms. Almost all these topologies root in the parallel computing science. As we know high throughput and low latency are two of the most important characteristics of MPSoC platforms. Tree and Fat Tree are very common topologies for NoCs. Fat tree is differing from traditional tree in the amount of resource bandwidth available at different levels of the tree. Many recent measurement studies have revealed that the traffic generated by many real-world applications exhibits a high degree of burstiness. Markov Modulated Poission Process (MMPP) has been widely adopted to model the message arrival behaviour of bursty traffic in the temporal domain since it can capture the time-varying arrival rate and the important correlations between inter-arrival times and still remain analytically tractable. The two-state MMPPs (the subscript s denoting the traffic generated by source nodes) used to capture the traffic burstiness in the temporal domain, can be characterised by the infinitesimal generator Q_s of the underlying Markov chain and rate matrix Λ_s . In past, the virtual channels used to enhance the performance of the interconnect networks. This work will presents a new analytical model for fat tree m-port n-tree in the presence of bursty traffic and the effect of using virtual channels. The accuracy of the model will validated through extensive simulation experiments of an actual system.

Performance Modelling and Analysis of Network on Chip under m-port n-tree Bursty Traffic with Virtual Channels

Hatem Ibrahim
Prof. Irfan Awan

Outline

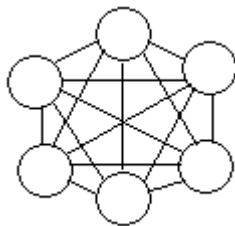
- Introduction
 - Interconnection Network and Network On Chip
 - Traffic Modelling
 - Markov-Modulated Poisson Process (MMPP)
 - Virtual Channels
- Network Topology
- Switching & Routing
- Methodologies
- Future Work

Introduction

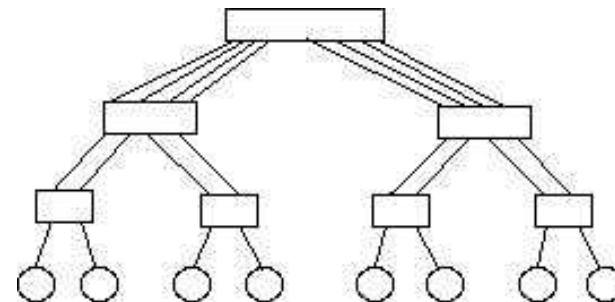
■ Interconnection Network:

- Is a programmable system that transports data between terminals, and it acts as a main role in determining the overall performance of multi-computers system which make the study of interconnection networks very important.
- INs simply classify according to topology:
 - Direct connection, all nodes have direct connection to other nodes (Hypercubes, Mesh, and tours) widely employed in multi-computer systems
 - Indirect connection, nodes connected with each other by intermediate nodes (Fat-tree, butter-fly, and crossbar)

Mesh



Fat-Tree



Introduction

- Network-On-Chip (NoC) have been proposed as a solution to mitigate complex on-chip communication problems.
- The demand for scalable, low latency and power efficient System-On-Chip interconnection, leads the development of NoC.
- On-chip networks share significant similarities with the traditional interconnection networks for parallel computers with multiple processors. Nevertheless, NoCs can be distinguished from the parallel computers by its characteristics.

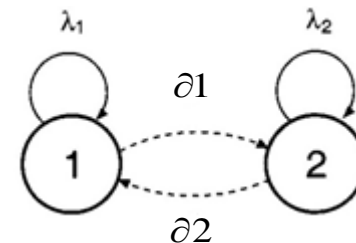
Traffic Modelling

- Traffic Pattern has significant impact on network performance.
- the performance study of wired or wireless networks has to combine with accurate models of the network traffic.
- Costly and time consuming, mathematical modelling of network traffic can be employed instead to evaluate the network performance in early stage.
- Two key parameters of message
 - Arrival process
 - Destination distribution

Markov-Modulated Poisson Process

- MMPP Is the doubly stochastic Poisson process with arrival rate λ_i
- Has widely applied the arrival process of bursty traffic because of the following reasons:
 - MMPP has the ability of capturing the time-varying arrival rate and the important correlations among inter-arrival time.
 - MMPP is closed under the splitting and superposition, and it can be used to model the splitting and superposition in the wireless or wired networks.
 - The studied have been widely reported the queuing-related result of MMPP, these derive the network modelling with the MMPP arrival process analytically soluble.
 - The characterised of MMPP are infinitesimal generator \mathbf{Q}_s of the underlying Markov chain and rate matrix Λ_s .

- $$\mathbf{Q}_s = \begin{bmatrix} -\partial_1 & \partial_1 \\ \partial_2 & \partial_2 \end{bmatrix} \quad \text{and} \quad \Lambda_s = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



Virtual Channels

- In past, the virtual channels used to enhance the performance of the interconnect networks.
- Adding virtual-channels lead to helps in boosting performance and circumventing message-dependent deadlock
- Improvements of Quality-of-Service (QoS) possibly by prioritising the allocation of virtual-channels and switch bandwidth
- Virtual Channels (V) , ($V \geq 2$)
- The average degrees of multiplexing of virtual channels that take place at a physical channel can be estimated by:

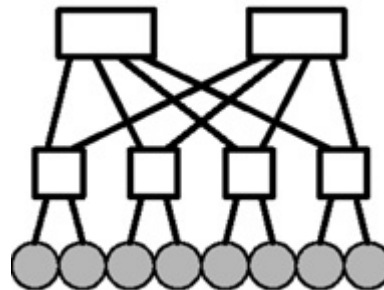
$$V_i = \sum_{v=1}^V \frac{v^2 P_{iv}}{v P_i}$$

Research Methodologies

- The network performance can be accomplished through two ways:
 - Analytical modelling
 - Can gain significant insights into the system performance.
 - Can provide quantitative relations between performance metrics and input parameters.
 - Using Matlab
 - Simulation
 - To understand complex investigations may takes months or years.
 - Using OMNeT++

Network Topology

- Fat-tree is the most popular network topology for fifty years.
- m-port n-trees is a typical example for fat-tree.
- It can rich connectivity among nodes this can lead to maintain paths between all source and destination nodes, and has full bisectional bandwidth.
- This is critical for satisfying the performance requirements of cluster computing.
- Consists of $2\left(\frac{m}{2}\right)^n$ processing nodes, and $(2n - 1)\left(\frac{m}{2}\right)^{n-1}$ communication switches.



4-port 2-tree

Switching and Routing strategies

- Wormhole switching has raised in popularity in cluster based system, because of its low buffering.
- Message is divided into flits, for transmission and flow control.
- Header flit of a message arrives at a node, have to obtain a channel before being forwarded to next node.
- The rest of the flits follow the header in pipelined fashion.
- Deterministic routing has been employed, message crosses a fixed path between source and destination
- Gomez has shown that deterministic routing can achieve a similar, and some scenarios higher lever of performance than adaptive routing in fat-tree.

References

1. A. Bouhraoua, O. Diraneyya, and M. E. Elrabaa, "A simplified router architecture for the modified Fat Tree Network-on-Chip topology," in NORCHIP, 2009, 2009, pp. 1-4.
2. Y. M. Boura and C. R. Das, "Performance analysis of buffering schemes in wormhole routers," *Computers, IEEE Transactions on*, vol. 46, pp. 687-694, 1997.
3. A. Bucur and D. Epema, "The Influence of the Structure and Sizes of Jobs on the Performance of Co-allocation," 2000, pp. 154-173.
4. A. Clematis and A. Corana, "Modeling performance of heterogeneous parallel computing systems," *Parallel computing*, vol. 25, pp. 1131-1145, 1999.
5. W. J. Dally and C. L. Seitz, "Deadlock-free message routing in multiprocessor interconnection networks," *Computers, IEEE Transactions on*, vol. 100, pp. 547-553, 1987.
6. W. J. Dally and B. Towles, *Principles and practices of interconnection networks*: Morgan Kaufmann, 2004.
7. C. Gomez, F. Gilabert, M. E. Gomez, P. Lopez, and J. Duato, "Deterministic versus adaptive routing in fat-trees," 2007, p. 292.
8. F. Kathleen, "The Markov-modulated Poisson process (MMPP) cookbook," *Performance Evaluation*, vol. 18, pp. 149-171, 1993.

Thank you

PART TWO Network Architectures and Protocols

Architecture for automatic reconfiguration of network devices in response to events in the network

Krzysztof Grochla, Leszek Naruszewicz
Proximetry Poland,
40-203 Katowice, Al. Rozdzińskiego 91
[kgrochla|lnaruszewicz]@proximetry.pl

Abstract. The article presents the architecture of the network management system capable of autonomous reconfiguration of the managed devices by preprogrammed rules. The proposed architecture allows implementation of simple logical functions providing self-healing and self-optimization functions in wireless networks. The architecture is based on the centralized network management system and business rule management engine, which processes events in the network and decides whenever a preprogrammed rule should be started.

Keywords: Network management, autonomous operation, self-optimization, SON

1 Introduction

Current wired and wireless networks consist of a very large number of devices. These devices are configured by the network operator. To automate this task, network management systems (NMS) have been developed. They consist of software and hardware which allow performing typical network administrator tasks. An NMS employs various protocols to accomplish these tasks. For example, the SNMP[1] protocol can be used to gather information from devices and send configuration change commands. Other popular network management protocols include NETCONF[2] and CWMP[3] (often referred also as TR-069).

Typically, network management systems do not apply a logic which executes operations other than the ones directly initiated by the operator. The operator uses an NMS to execute configuration changes easier and quicker compared to logging directly into the devices. Some of the NMSes offer group configuration changes to automate reconfiguration of a large number of devices, or allow scheduling of an operation at a predefined time, but the network management software does not trigger configuration changes by itself. This paradigm changes with the introduction of the Self-Optimized Network (SON) concept, which was proposed as the part of the LTE standard by 3GPP [4]. The SON feature allows automation of several tasks to lower operation costs (OPEX). For example, in a SON-enabled LTE network the neighbor relations and Physical Cell IDs are selected automatically [5]. The main functionality of SON includes self-configuration, self-optimization and self-healing. The costs of the

network installation may be lowered by auto-configuration functions, which allow adding new active devices to the network, such as base stations, almost without any manual configuration [6].

Self-optimization and auto-configuration require implementation of new functions in the network management systems, namely algorithms that react to changes in the network state and reconfigure the network when needed as well as calculate the optimal parameters of the network operation. The NMS needs to react to planned changes, such as addition of a new base station, to failures and to variation of radio signal propagation. To perform such an operation correctly, it needs to monitor the statistics representing the current transmission conditions, as well as to react to events representing registration or loss of connectivity with a device.

The automation of network management operations is very specific and related directly to the particular network installation or application. Some of the SON functions can be predefined and are valid for every network installation (eg. the PCI assignment in LTE), but others (e.g. reconfiguration in response to failures) require adaptation to the operator requirements. A vendor of the network management system is incapable of defining all possible operations, therefore it is more effective when such actions are defined or programmed by domain experts or network administrators.

Some of the novel network management systems allow the network operator to predefine operations which are executed upon particular events, e.g. Adrem Net-Crunch [7] allows setting corrective actions to be taken in response to an alert. This functionality is often very limited and uses only a predefined set of triggers. In this paper, we propose an open architecture which enables building automation functions using a rule engine, which is external to the NMS server or central servers. This architecture allows flexible prototyping and validation of the autonomous operational rules, which change operation of the network in response to some events or to changes in the network performance.

2 Proposed architecture of the system

2.1 Generic architecture of the autonomous rule engine

NMSes are implemented as a centralized or a distributed system. A centralized NMS is based on a single server or a set of servers collocated within one location, which communicate over an IP network with managed devices and executes management operations. It stores the device configuration, network state, statistics and other data related to the network operation in a database. A distributed NMS consists of agents distributed among multiple network devices or distributed servers, without a single coordination point. The centralized approach is adopted by most of the network management tools available commercially and was the design goal of the management protocols[1-4]. In this work, we concentrate on the centralized architecture.

We propose to extend the architecture by a new entity – the Autonomous Operational Rule Engine. It interfaces with the NMS using a network connection. The Rule Engine is a software system that executes one or more business rules in a runtime

production environment. It subscribes to the events and data it wants to monitor and executes rules defined by the operator.

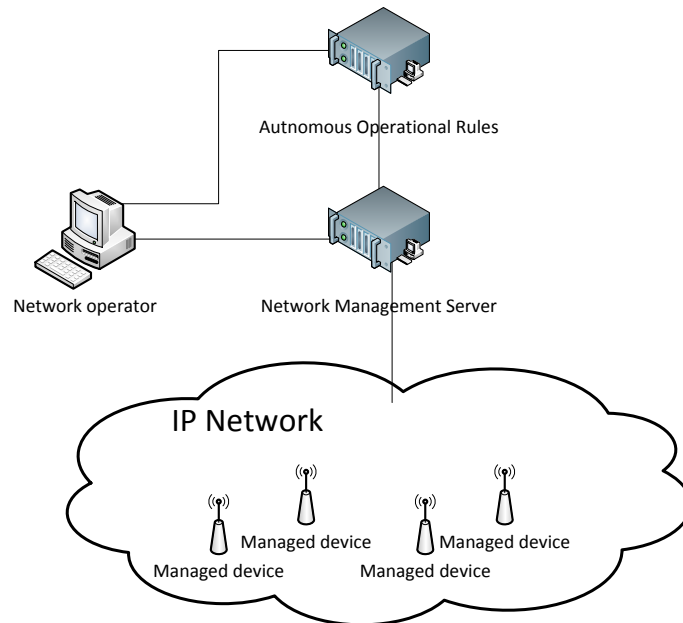


Fig. 1. General view of the architecture for automatic reconfiguration of network devices

2.2 Detailed description of the proposed autonomous rule operation architecture

In the proposed architecture, the NMS server interfaces with the Autonomous Operational Rule Engine (AOR) through two layers of communication: the syslog events and the refreshing and provisioning operation protocol. The refreshing protocol is the standard external interface provided by the NMS for interaction with other systems or for communication with the GUI. It can be based on HTTP and web services, or any other protocol. It provides the following functionality that is used by the AOR:

- transmission of the selected statistics gathered from network devices
- subscription to network events
- subscription to selected statistics
- reconfiguration of network devices
- definition and monitoring of QoS parameters.

To allow triggering of an operation in response to any event logged by the NMS, syslog has been used as a secondary means of communication.

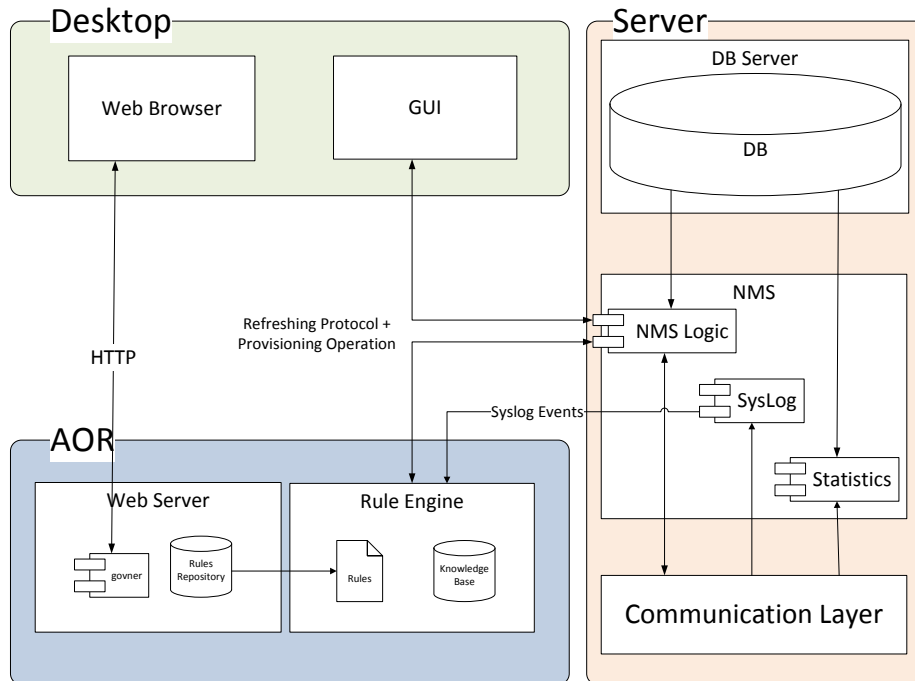


Fig. 2. Interfaces between the NMS and the rule engine

The proposed Autonomous Operational Rule Engine consists of two modules: the Web Server providing the user interface and the Rule Engine which executes configured rules. The rules are stored in a database. The rule engine goes through all the rules either periodically or in response to a received event.

The user interface needs to be extended with the functionality to define and configure the rules. This can be accomplished by using a web interface to simplify the communication.

3 Sample implementation

The proposed architecture of the automatic reconfiguration of network devices in response to events in the network has been implemented using the RedHat Drools business rule engine[8] and Proximetry AirSync NMS[9]. The rule engine has been divided into three elements: a queue manager, a network state builder, and a configuration manager.

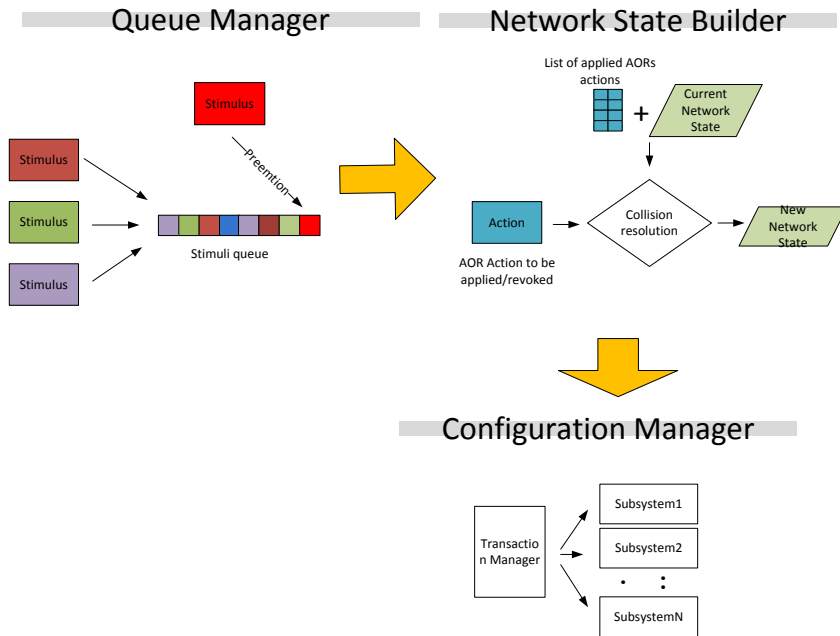


Fig. 3. Structure of the implemented rule engine

The **AOR Queue Manager** is responsible for queuing multiple stimuli which can be received from the network or external systems. By default, stimuli are inserted at the end of the queue and processed in the FIFO manner. The system defines a set of stimuli which are critical and preempts all other stimuli which are queued in the system. The **Network State Builder** takes a single stimulus from the head of the queue and, based on a list of actions defined by a given stimulus, builds a new network state. Each action is analyzed by the collision resolution module against the current network state and applied (or not) based on the predefined policy. The **Configuration Manager** is responsible for applying the new state to the network. It interfaces with all subsystems which need to be reconfigured, creates and manages configuration change transactions.

3.1 Sample functions

In the implemented system, the statistics related to a signal level were processed. As a sample operational rule we implemented reconfiguration of an existing QoS profile in response to change in modulation which is used on the link between a CPE and a base station. It allowed allocation of the same amount of radio resource blocks for this particular CPE, regardless of its distance to the BS.

4 Security and performance considerations

User accounting and access level verification should be maintained in the AOR Engine on the same level as in case of actions executed by the user directly in the NMS. Therefore, the AOR Engine needs to verify the user access rights in the NMS before applying any of the rules.

When multiple operational rules are executed a large number of statistics needs to be forwarded to the AOR Engine. This may thwart the performance of both the NMS and the rule engine. To maximize the performance, the Rule Engine should be collocated with the NMS server within the same physical location and connected using fast network connection. When a particular rule is heavily exploited (e.g. applied on all devices), the AOR can be treated as a rapid prototyping tool and after testing the functionality, the algorithm can be implemented within the NMS. Our performance tests executed with Proximity AirSync and JBoss Drools allowed execution of a single rule on all devices within the network composed of thousands of managed devices, but the performance depends heavily on the number of processed statistics and events.

5 Conclusions

The proposed architecture enables easy implementation and prototyping of the network self-optimization and self-configuration functionalities. The sample implementation showed that it can be used effectively with the JBoss Drools rule engine.

References

1. K. McCloghrie, K. McCloghrie, J. Schoenwaelder, and D. Perkins, "Structure of Management Information Version 2 (SMIv2)". <http://tools.ietf.org/html/rfc2578>.
2. Bierman, R. Enns, M. Bjorklund, and J. Schoenwaelder, "Network Configuration Protocol (NETCONF)". <http://tools.ietf.org/html/rfc6241>.
3. "CPE WAN Management Protocol". TR-069 Amendment 4. Broadband Forum. July 2011. Retrieved February 16, 2012.
4. Hämäläinen, Seppo, Henning Sanneck, and Cinzia Sartori, eds. LTE Self-Organising Networks (SON): Network Management Automation for Operational Efficiency. John Wiley & Sons, 2012.
5. Lukasz Chrost, Krzysztof Grochla: Conservative Graph Coloring: A Robust Method for Automatic PCI Assignment in LTE. *Computer Networks*, 268-276
6. Tragos E., Siris V., Bruno R., Grochla K, Ancillotti E.: Automatically configured, optimised and QoS aware wireless mesh networks, *Proceedings of IEEE PIMRC 2010*.
7. AdRem NetCrunch, <http://www.adremsoft.com/netcrunch/>
8. RedHat JBoss Drools, <http://www.jboss.org/drools/>
9. Proximity AirSync, <http://www.proximity.com/airsync-overview.php>

A Unified Congestion Control Architecture Design to Improve Heterogeneous Wireless Network Efficiency and Accommodate Traffic by Various Rich Applications

Katsuhiko Kusahata¹, Itsuki Kaneko¹, Katsuyoshi Iida¹, Hiroyuki Koga², and
Masayoshi Shimamura¹

¹ Tokyo Institute of Technology,
{katsu,kaneko,shimamura}@netsys.ce.titech.ac.jp, iida@gsic.titech.ac.jp
² University of Kitakyushu, h.koga@kitakyu-u.ac.jp

Abstract. To mitigate the severe congestion of mobile data networks, many researchers are investigating traffic offloading technology. In this paper, we propose a network architecture to improve network efficiency and satisfy the QoS requirements of applications using centralized control at a data center accommodating many servers.

Keywords: Heterogeneous wireless networks, Transport layer protocol, Unified congestion control architecture.

1 Introduction

The use of many rich applications, such as video conferencing and cloud storage applications, is becoming more widespread in the mobile environment. Users of such applications would like to receive a satisfactory quality of service (QoS) from the networks. On the other hand, mobile operators would like to distribute this traffic to supplementary deployed mobile networks, such as Wi-Fi hotspots, since the large amounts of traffic generated by rich applications can cause severe congestion in mobile data networks. Thus, two major requirements need to be met in heterogeneous wireless networks (HWNs): (1) that the QoS of rich applications be maintained, and (2) that mobile data traffic can be distributed into appropriate wireless network interfaces such as 3G, LTE, Wi-Fi, and mobile WiMAX to improve overall network efficiency.

To illustrate these two requirements, we show a simple illustration of flow assignments in HWNs in Fig. 1. In this illustration, there are two devices of user equipment (UE), each of which runs two applications App. 1 (Application 1) and App. 2 via wireless access points AP1 (Access Point 1) and AP2. For example, App. 1 is a delay sensitive application and AP1 provides low latency compared with AP2. App. 2 may be a throughput sensitive application which prefers higher bandwidth wireless networks. In this way, the overall network efficiency will be improved under the constraints related to QoS requirements.

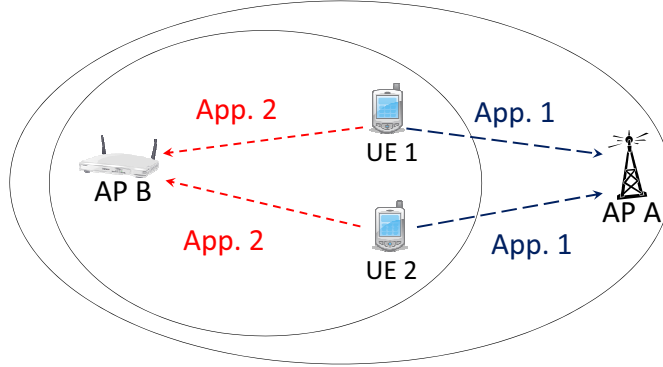


Fig. 1. Flow assignments in heterogeneous wireless networks

Related research in the literature includes the following. In [1], Ma et al. proposed enhancing mobile Stream Control Transmission Protocol (SCTP), which is a transport layer protocol supporting a multi-home technology that will be a key technology to satisfy requirement (2). In [2], C-RAN was presented, which combines centralized based-band pool processing, co-operative radio with distributed antenna equipped with remote radio head (RRH) and real-time cloud infrastructures (RAN). The purpose of C-RAN, which is related to requirement (2), is to enable centralized management of heterogeneous RAN including 3G, LTE, and Wi-Fi. However, neither of these approaches meet both of the requirements set out above; [1] and [2] fail to satisfy requirements (2) and (1), respectively.

In this paper, we propose the Unified Congestion Control Architecture (UCCA) which introduces a centralized congestion controller inspired by C-RAN as illustrated in Fig. 2. Specifically, servers on data centers manage each UE to improve network efficiency. Unlike C-RAN, UCCA servers manage the transport layer flows to maintain the QoS of applications for multiple networks operated by different mobile operators.

2 Unified congestion control architecture

To improve network efficiency while satisfying the QoS requirements of applications, our basic idea is to collect congestion information at a data center because many servers are located at data centers. In UCCA, we deploy a transport connection manager (TCM), which is located at the data center as shown in Fig. 2. TCM collects congestion information in the HWNs. Using the collected information, TCM performs two actions: network interface selection and congestion window (CWND) determination. Network interface selection means interface selection for any UE connected to multiple wireless networks.

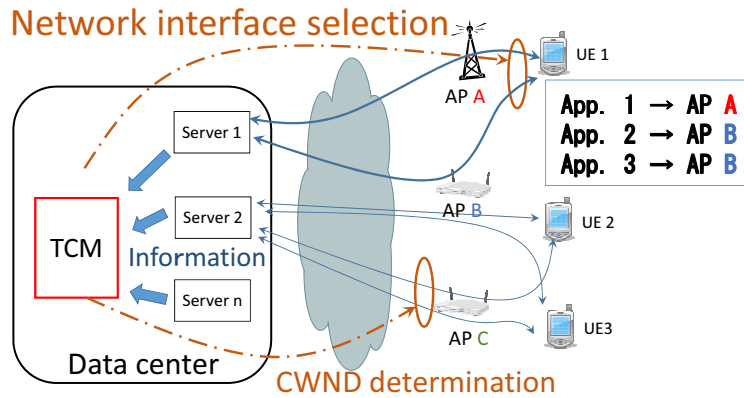


Fig. 2. Unified congestion control architecture (UCCA)

UCCA consists of four components: (A) a congestion information collector in APs, (B) a congestion controller and (C) a path manager in each client and server, and (D) a unified congestion control engine in TCM, as illustrated in Fig. 3. Note that (A), (B), (C), and (D) symbols are shown in the figure. In this paper, we assume that clients and servers employ SCTP [3,4] which supports multi-homing functions to use appropriate wireless network interfaces according to network conditions.

In this diagram, the client establishes SCTP association with the server as a logical connection, which contains two transmission paths through AP1 or AP2. Congestion information such as transmission rate, transmission delay time, and transmission error rates collected by the congestion information collector in APs from their physical and datalink layers is notified to the unified congestion control engine in TCM by storing the information into an SCTP heartbeat chunk transmitted from the client to the server, which is commonly used to confirm the availability of transmission paths to manage the association in original SCTP operation. QoS requirements of applications are also notified to the unified congestion control engine in a similar way.

The unified congestion control engine selects an appropriate network interface satisfying the QoS requirements from applications on the clients and determines the suitable CWND for effective communications based on the notified congestion information and QoS requirements of applications. The information on network interface selection and CWND determination is notified to the server. The path manager on the server informs that on the client of the information on the appropriate network interface to use. For congestion control, two directions of data flows should be considered: a downward direction from the server to the client and an upward direction from the client to the server. Although the congestion controller on the server can control its transmission rate of data flow

based on the notified information on suitable CWND in the downward direction, it has to inform the congestion controller on the client of the information on CWND to control data flow in the upward direction. We note that the above control flows are illustrated in Fig. 4.

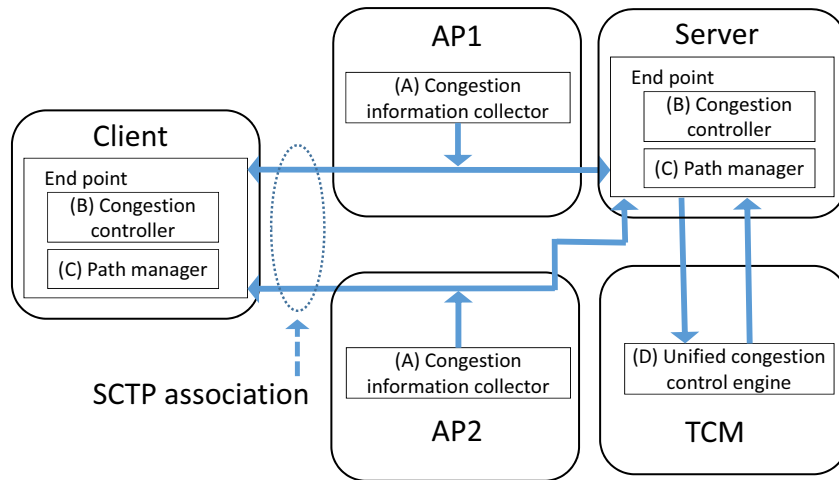


Fig. 3. Function diagram of UCCA

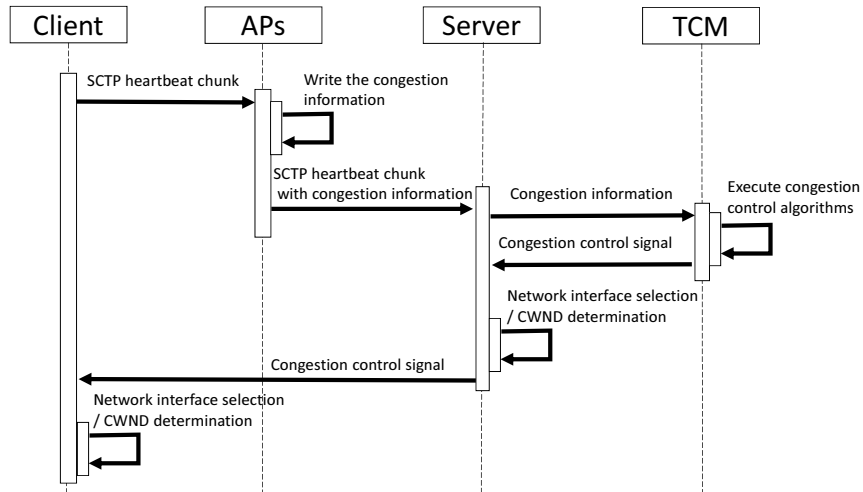


Fig. 4. Sequence diagram of UCCA

Through these operations, UCCA can achieve two objectives in HWNs: (1) that the QoS of rich applications can be maintained, and (2) that mobile data traffic can be distributed into appropriate wireless network interfaces to improve overall network efficiency. Although the effectiveness of UCCA depends on the algorithms in the unified congestion control engine to determine an appropriate network interface and CWND based on the collected information, we first focus on the key idea to collect congestion information at a data center. To evaluate the potential of UCCA, we are now working on designing and implementing detailed mechanisms in UCCA.

3 Concluding remarks and future directions

We have developed UCCA as a means to achieve two goals: maintain application QoS and improve network efficiency. Our basic idea is to introduce TCM at a data center and have it collect congestion information in HWNs. TCM can then select network interfaces in a way that achieves both goals.

Currently, we are designing detailed mechanisms in UCCA. We also plan to evaluate our architecture using flow level simulations.

Acknowledgment

This work was supported in part by JSPS KAKENHI Grant-in-Aid for Scientific Research (B) Number 25280028.

References

1. Ma, L., Yu, R. and Leung, V.C.M.: Performance improvements of mobile SCTP in integrated heterogeneous wireless networks, *IEEE Trans. Wireless Comm.*, 6, 3567–3577, (2007).
2. China Mobile Research Institute: C-RAN, the road towards green RAN: White paper, [http://labs.chinamobile.com/cran/wp-content/uploads/CRAN_white_paper_v2_5_EN\(1\).pdf](http://labs.chinamobile.com/cran/wp-content/uploads/CRAN_white_paper_v2_5_EN(1).pdf), (2011).
3. Stewart, R.: Stream control transmission protocol, IETF RFC4960, (2007).
4. Stewart, R. and Xie, Q.: Stream control transmission protocol (SCTP), Addison-Wesley, (2002).

PART THREE Performance vs. Security Trade-offs

Performance modelling of security protocols

Nigel Thomas and Yishi Zhao

School of Computing Science, Newcastle University, UK.
`nigel.thomas@ncl.ac.uk`

Abstract. In this paper we state the case for the performance evaluation of secure systems. We provide evidence of a wide variation in performance of different secure methods. We further explore the overhead introduced by secure functions in considering a case study in non-repudiation. We present a model of a non-repudiation protocol specified using the Markovian process algebra PEPA and present results derived using mean value analysis and mean field approximation.

1 Introduction

The security of modern computer and communication systems is a major concern for governments, organisations and individuals, resulting in a significant effort to ensure, and prove, that systems remain secure and data remains private. However, it is also essential that security measures do not impose excessive constraints on the user which then encourage subversion of those measures in order to make the system more usable. It should be clear that any security measure that degrades usability is undesirable. However, all security measures will entail some additional work being undertaken which imposes a performance overhead. It is therefore essential that this overhead is understood, measured and minimised. In some practical situations there may be a choice of methods (including varying protocols, algorithms or parameters) which could be employed. Changing the choice of method could have a potentially significant impact on the system performance without degrading the security. In other situations methods can be modified (e.g. by changing a key length or refresh rate) which might improve performance at the cost of some level of security, thus giving a security performance trade-off [1]. However, quantifying this trade-off is not generally possible due to a lack of quantitative methods for evaluating system security. Instead our approach is based purely on evaluating the performance of the system and thus giving the system designer information on competing designs.

Cryptographic protocols are one of the few areas of security that have received attention from both security and performance communities [2–5]. For example consider Figures 1-3, which illustrate the performance of different aspects of secure system behaviour. Figure 1 shows the average execution time for a variety of cryptographic protocols for a specific message length. Figure 2 illustrates the dramatic variation in performance that can be achieved by varying the key length for public key encryption algorithms. Figure 3 shows the measured performance of different functions within a secure stock trading application, with

encryption and without. By considering data such as this, a system designer can modify a secure system to take account of performance.

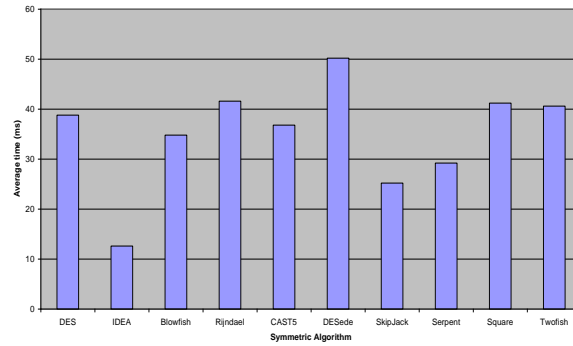


Fig. 1. Average execution time for symmetric key encryption algorithms [4]

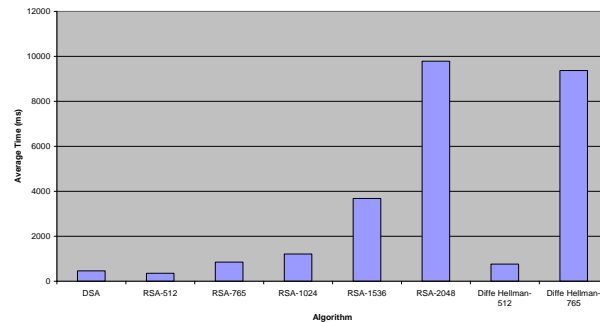


Fig. 2. Execution time varied against key length for RSA and Diffie-Hellman [2]

2 Performance models of secure systems

A greater level of understanding of secure system performance can be gained by specifying and analysing a performance model. A *Key Distribution Centre* (key exchange protocol) has been studied in our previous work, which shows the possibility of modelling by a stochastic process algebra PEPA and analysis by several alternative techniques [7–9]. The advantage of using a formal specification for such models is that it is possible to check specific properties to ensure that the model correctly depicts behaviour which is essential to the security of the

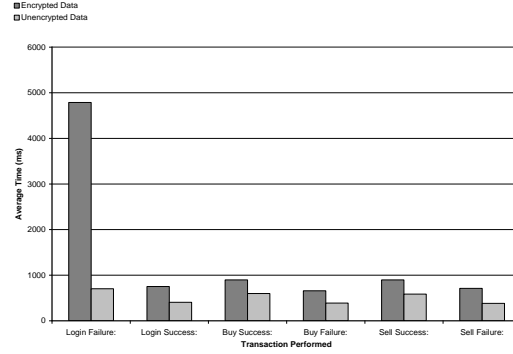


Fig. 3. Average response time for functions within a secure trading application [6]

system. Thus a formal performance model and a formal security model of a given system can be shown to exhibit equivalence, giving the system designed some reassurance that the performance behaviour is valid. A process algebra allows detailed behaviour to be modelled and has the potential to be modified automatically through model transformations to facilitate alternative forms of analysis.

3 A case study in non-repudiation

A non-repudiation service will prevent either of the principals involved from denying the contract after the agreement. The two such protocols were proposed by Zhou and Gollmann [10, 11] and use a non-repudiation server, known as a *Trusted Third Party* (TTP). We will specify the first of these protocols, from now on referred to as ZG1.

- A : originator of the non-repudiation exchange
- B : recipient of the non-repudiation exchange
- TTP : on-line trusted third party provide network services accessible to the public
- M : message sent from A to B
- C : ciphertext for message M
- K : message key defined by A
- $NRO = sS_A(f_{NRO}, B, L, C)$: Non-repudiation of origin for M
- $NRR = sS_B(f_{NRR}, A, L, C)$: Non-repudiation of receipt of M
- $sub_K = sS_A(f_{SUB}, B, L, K)$: proof of submission of K
- $con_K = sS_T(f_{CON}, A, B, L, K)$: confirmation of K issued by TTP

First, A sends the ciphertext (C) and a non-repudiation origin (NRO) for message M to B , and then B replies back with a non-repudiation receipt (NRR) to A . Now B possesses the ciphertext, but cannot read it as he still hasn't got the key to decrypt M . According to the non-repudiation requirement, B is not

a trusted agency to A for sending the key directly to B , they only can resort to a trusted third party (TTP). After receiving the key and proof of submission (sub_K), the TTP will generate a confirmation of K (con_K) and publish in a read only public area. Finally, B can get the key from this public area to decrypt ciphertext (C) and A fetches the confirmation of submission as non-repudiation evidence.

From this protocol specification we can derive the following PEPA model for the complete system when there are N pairs of principals.

$$\begin{aligned}
TTP &\stackrel{def}{=} (publish, r_p).TTP \\
AB_0 &\stackrel{def}{=} (sendB, r_b).AB_1 \\
AB_1 &\stackrel{def}{=} (sendA, r_a).AB_2 \\
AB_2 &\stackrel{def}{=} (sendTTP, r_t).AB_3 \\
AB_3 &\stackrel{def}{=} (publish, r_p).AB_4 \\
AB_4 &\stackrel{def}{=} (getByA, r_{ga}).AB_5 \\
&\quad + (getByB, r_{gb}).AB_6 \\
AB_5 &\stackrel{def}{=} (getByB, r_{gb}).AB_7 \\
AB_6 &\stackrel{def}{=} (getByA, r_{ga}).AB_7 \\
AB_7 &\stackrel{def}{=} (work, r_w).AB_0 \\
SystemZG1 &\stackrel{def}{=} TTP[K] \underset{publish}{\boxtimes} AB_0[N]
\end{aligned}$$

AB_0 to AB_7 in the above ZG1 PEPA model denote the different behaviours of the AB component, and its evolution along the sequence of prescribed actions in the protocol. The choice from AB_4 to AB_5 and AB_6 means step 4 and step 5 in ZG1 can happen in any order. The *work* action is used to define that B can do something with the key and ciphertext after he has obtained these, before returning to the state AB_0 to make a new request again, which forms a working cycle to investigate the steady state.

Figure 4 shows the average queue length varied with number of customer involved in this non-repudiation system solved by an ODE (approximate) solution supported by the PEPA tools [?] and by exact mean value analysis [?].

The ODE approximation does not depend on deriving the state space of the underlying Markov chain, hence it scales very well [?]. Furthermore, the solution converges to the exact solution as the number of components increases, thus it becomes extremely attractive for solving models of extremely large systems, as illustrated in Figure 5.

Finally we show the relative performance between the ZG1 protocol and another proposed by the same authors which is shown to be less scalable. Thus a designer could use such a comparison to choose the appropriate protocol.

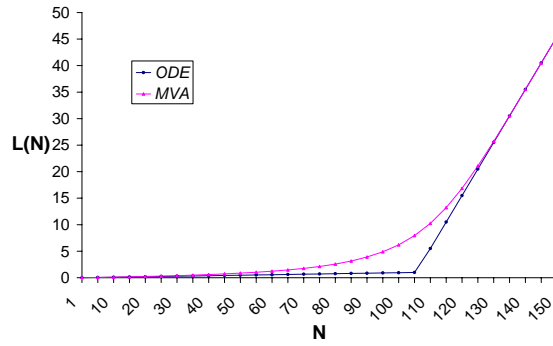


Fig. 4. Average queue length varied with population size calculated by the ODE, $r_b = r_{t1} = r_{ga1} = r_b = r_{t2} = r_{gb} = r_{ga2} = 1, r_w = 0.01, K = 1$

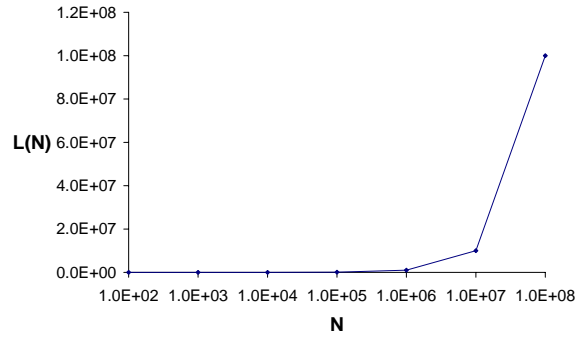


Fig. 5. Average response time varied with population size calculated by the ODE, $r_b = r_{t1} = r_{ga1} = r_b = r_{t2} = r_{gb} = r_{ga2} = 1, r_w = 0.01, K = 1$

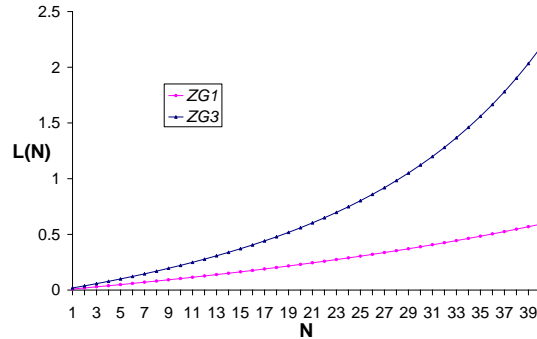


Fig. 6. Average number of waiting jobs with ZG1 and ZG3, $r_p = r_{t1} = r_{ga1} = r_b = r_{t2} = r_{gb} = r_{ga2} = 1, r_w = 0.01$

4 Conclusions

In this paper we have explored the issue of modelling secure systems. Estimating the costs (as well as the benefits) of security is important. We have shown that

not only can there be an appreciable overhead introduced by secure methods, but also that the overhead can vary considerably according to the method employed. Thus there is a clear opportunity for the system analyst to improve, or even optimise, performance by choosing or tuning the various algorithms and protocols.

To date our analysis has focussed on identifying and employing efficient solution methods. There is considerable scope for further work to investigate the relationship between formal security models and formal performance models. The ultimate goal would be to create a system which could automatically produce analysable performance models from security models. However, the choice of security solution, driven by the performance security trade-off should always remain an expert task.

References

1. K. Wolter, and P. Reinecke. Performance and security tradeoff, in: *Formal methods for quantitative aspects of programming languages*, pp. 135-167, Springer, 2010.
2. S. Dick and N. Thomas, Performance analysis of PGP, in: F. Ball (ed.) *Proceedings of 22nd UK Performance Engineering Workshop*, Bournemouth University, 2006.
3. W. Freeman and E. Miller, An Experimental Analysis of Cryptographic Overhead in Performance-critical Systems, *Proceedings of the 7th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, IEEE Computer Society, 1999.
4. C. Lamprecht, A. van Moorsel, P. Tomlinson and N. Thomas, Investigating the efficiency of cryptographic algorithms in online transactions, *International Journal of Simulation: Systems, Science & Technology*, 7(2), pp 63-75, 2006.
5. M. Buchholtz, S. Gilmore, J. Hillston and F. Nielson, Securing statically-verified communications protocols against timing attacks, *Electronic Notes in Theoretical Computer Science*, 128(4), Elsevier, 2005.
6. L. Thorpe, unpublished report, University of Durham.
7. Y. Zhao and N. Thomas, Approximate solution of a PEPA model of a key distribution centre, in: *Performance Evaluation - Metrics, Models and Benchmarks: SPEC International Performance Evaluation Workshop*, pp. 44-57, LNCS 5119, Springer-Verlag, 2008.
8. N. Thomas and Y. Zhao, Fluid flow analysis of a model of a secure key distribution centre, *Proceedings 24th Annual UK Performance Engineering Workshop*, Imperial College, 2008.
9. Y. Zhao, N. Thomas, Efficient solutions of a PEPA model of a key distribution centre, *Performance Evaluation*, **67**(8), pp. 740-756, 2010.
10. J. Zhou and D. Gollmann, A Fair Non-repudiation Protocol, in: *Proceedings of IEEE Symposium on Security and Privacy (SP'96)*, IEEE Computer Society, 1996.
11. J. Zhou and D. Gollmann, Observation on Non-repudiation, in: *Advances in Cryptology-ASIACRYPT'96*, 133-144, LNCS 1163/1996, Springer-Verlag, 1996.

Performance vs. Security Trade-Offs in RANETs based on Hybrid Quantitative Network Models

Guzlan M.A. Miskeen^{1,2}, Demetres D. Kouvatsos²
and Esmail Habibzadeh^{2,3}

1 University of Sebha, Libya

2 University of Bradford, England

3 IUST - Iran University of Science and Technology

{G.M.A.Miskeen, D.Kouvatsos,
E.habibzadeh}@brad.ac.uk

Abstract. Most used security mechanisms in Robotic Ad Hoc Networks (RANETs) have been adopted without adequate quantification of their impact on performance degradation under bursty traffic conditions. This tutorial presents an effective quantitative methodology for RANETs using quantitative hybrid Generalised Stochastic Petri Nets (GSPNs) and Queuing Networks (QNs) models. Performance vs. security trade-offs of RANETs can be evaluated and predicted by employing this hybrid model through discrete-event simulation (DES) by means of Hybrid Petri Queuing Networks Simulator (HPQNS) [2] under bursty traffic conditions, subject to Combined Performance and Security Metrics (CPSMs) proposed in [3]. The Generalised Exponential (GE) distribution is used to model inter-arrival and service times at each node in order to capture the traffic burstiness of the network and predict pessimistic ‘upper bounds’ of network performance. In this context, each robotic node is represented by a hybrid Gated GSPN (G-GSPN) and a QN mode, where G-GSPN incorporates bursty multiple class traffic flows, nodal mobility, security processing and control whilst the QN model has, generally, an arbitrary configuration with finite capacity channel queues reflecting ‘intra’-robot (component-to-component) communication and ‘inter’-robot transmissions. The inclusion of robot’s mobility in the model [1] enables realistic decisions in mitigating the performance of mobile robotic nodes in the presence of security. Furthermore, a theoretical case study from the literature is adapted to illustrate the utility of the QN towards modelling ‘intra’ and ‘inter’ robot communications. Extensions of CPSMs proposed in the literature [3] are suggested to facilitate investigating and optimising RANET’s performance vs. security trade-offs. As a future work, the overall CPSMs for N robotic nodes as well as Marginal and Aggregate CPAMs for R classes can be determined. This framework has a promising potential modelling more meaningfully and explicitly the behaviour of security processing and control mechanisms as well as capturing the robot’s heterogeneity (in terms of the robot architecture and application/task context) in the near future. Moreover, this framework should enable testing robot’s configurations during design and development stages of RANETs as well as modifying and tuning existing configurations of RANETs towards enhanced ‘optimal’ performance and security trade-offs.

References

1. H. Bhatia, R. Lening, S. Srivastava, and V. Sunitha, "Application of QNA to analyse the queueing network mobility model of MANET," Technical Report, Dhirubhai Ambani Institute of Information & Communication Technology (DAICT), Gandhinagar, India, 2007.
2. E. Habibzadeh, D. D. Kouvatsos and G. M.A. Miskeen, Hybrid Petri Queuing Networks Simulator (HPQNS), HET-NETs 2013 Conference, 11th -13th of November 2013, Ilkely, Bradford, UK. <http://www.scim.brad.ac.uk/~gmamiske/Final%20PDF%20Files/T05%20hpqns%20PPT.pdf>
3. K. Wolter and P. Reinecke "Performance and security trade-off," in *Formal Methods for Quantitative Aspects of Programming Languages, Computer science Lecture Notes*, vol. 6154, A. Aldini, M. Bernardo, A. Di Pierro, and H. Wiklicky, Eds. Berlin, Heidelberg: Springer 2010, pp. 135-167.

Performance vs. Security Trade-Offs in RANETs based on Hybrid Quantitative Network Models

Guzlan M.A. Miskeen^{1,2}, Demetres D. Kouvatsos² and
Esmaeil Habibzadeh^{2,3}

¹ University of Sebha, Libya

² University of Bradford, England

³IUST - Iran University of Science and Technology

{G.M.A.Miskeen, D.Kouvatsos, E.habibzadeh}@brad.ac.uk

12th of Nov. 2013

Outline

- Performance and Security Tradeoffs' Concepts
- Security and Tradeoffs Modeling Tools
- Robotic Ad Hoc Networks (RANETs): Related Concepts
- Trade-off Modeling in Literature
- The Proposed Hybrid Framework
- Simulation Experiments
- Final Remarks & Conclusions
- Extensions for Futurework

Performance and Security Tradeoffs' Concepts

- **Performance** is *“the degree to which a system accomplishes its designated functions within given constraints, such as speed, accuracy, or memory usage.”* [IEEE-610.12]
- **Security service** is the collection of mechanisms implemented to reduce the risk associated with threats of stored & exchanged data in communication networks.
- **Performance and Security Trade-off**
Either security is compromised for better overall performance or vice versa & it can be quantified and evaluated using quantitative tools such as QNs and GSPNs thus leading to the trade-off optimisation.

Performance and Security Tradeoffs' Concepts

- The performance-related security can be modelled and evaluated , considering :
 - 1- Modelling performance and security accurately using effective quantitative modelling tools (e.g., QNs and GSPNs which have various features in terms of **simplicity** and **modelling power**);
 - 2- Defining /introducing suitable metrics for both performance and security as well as appropriate **combined** metrics [WOL10].

Security Metrics

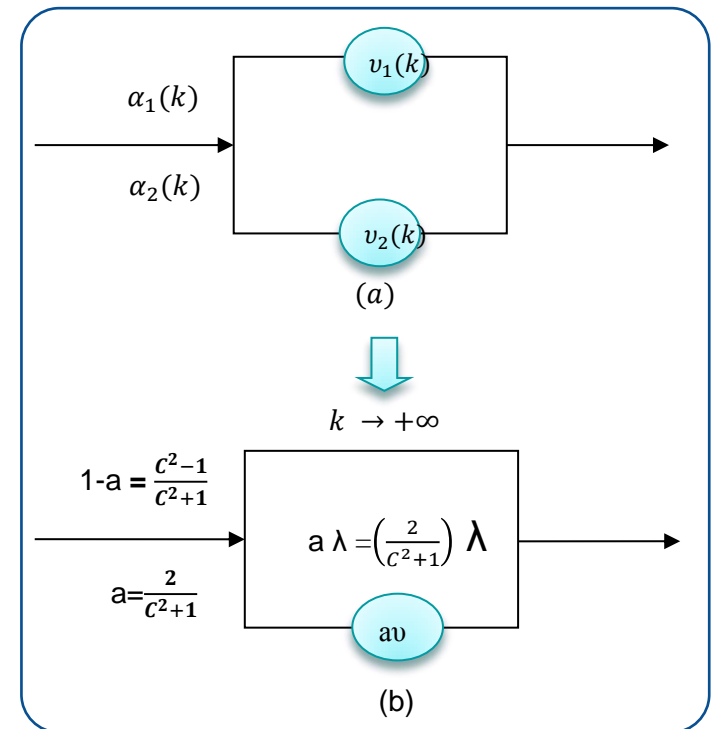
- Since security can be seen as a process, security metrics can be expressed as:
 1. ***Time of security failure occurrence/Time between security failures***;
 2. ***Probability of being in a certain condition***, e.g., probability of the encryption key to be valid”.
- For MANETs, encryption key length (and the corresponding encryption time) can be seen as a metric.

Trade-off Evaluation Under Bursty Traffic Condition

When the security mechanism is activated, how to quantify the corresponding degradation of performance under bursty traffic conditions?

Traffic burstiness can be captured using Generalised Exponential (**GE**) distribution, which can give the upper performance bounds.

- GE distribution is expressed with [KOU88]: λ & C^2
- $F(t) = P(W \leq t) = 1 - a e^{-a\lambda t}$, $t \geq 0$ $a = \frac{2}{(C^2+1)}$,
 $C^2 = \frac{\text{Var}(W)}{E^2(W)}$
- When $C^2=1$, GE r.v. \rightarrow exponential r.v
- GE can be seen as a limiting case of H2, with tuning parameter k with the same λ & C^2 [KOU88];
- When: $k \rightarrow +\infty$, $H_2 \rightarrow GE$.



QNs

- QNs are “concise graphical description” of service centres, queues and their disciplines and routing amongst these nodes;
- Used Scheduling strategies such as First-Come-First-Served (*FCFS*) and Head-of-the-Line (*HoL*) priority rules with either infinite or finite capacity queues

Key QNs Advantages & Limitations

Advantages	Limitations
<ul style="list-style-type: none">• Powerful for modelling hardware contention and scheduling strategies;• Robust exact / approximate algorithms for calculating performance measures for multiple class QNs with or without blocking.	<ul style="list-style-type: none">• Not suitable, in general, for modelling synchronization, simultaneous resource possession, security control and software contention;

GSPNs

- They are credible modelling tools for the qualitative/quantitative analysis of complex system involving software contention, blocking and synchronisation.
- A PN is composed of two types of nodes: Places, p and transitions, t :
 - A **place**: accommodates customers called '**tokens**' which defines GSPN state;
 - A **transition**: represent event that changes the system's state;
 - Arcs: are the links between places and transitions.

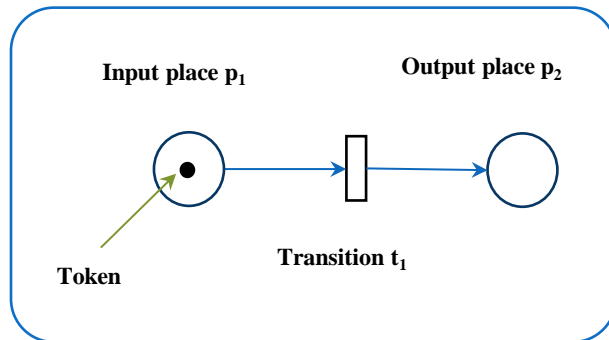


Fig. A SPN Components

Operation Rules

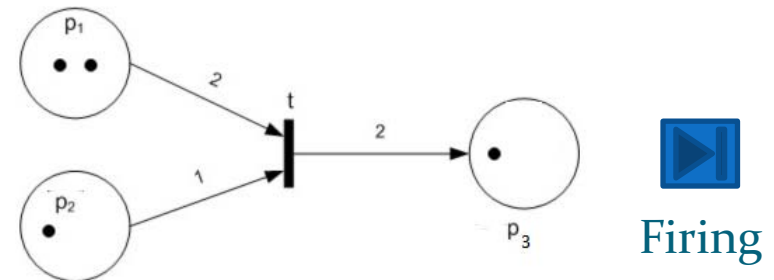


Fig. B SPN Enabling and firing

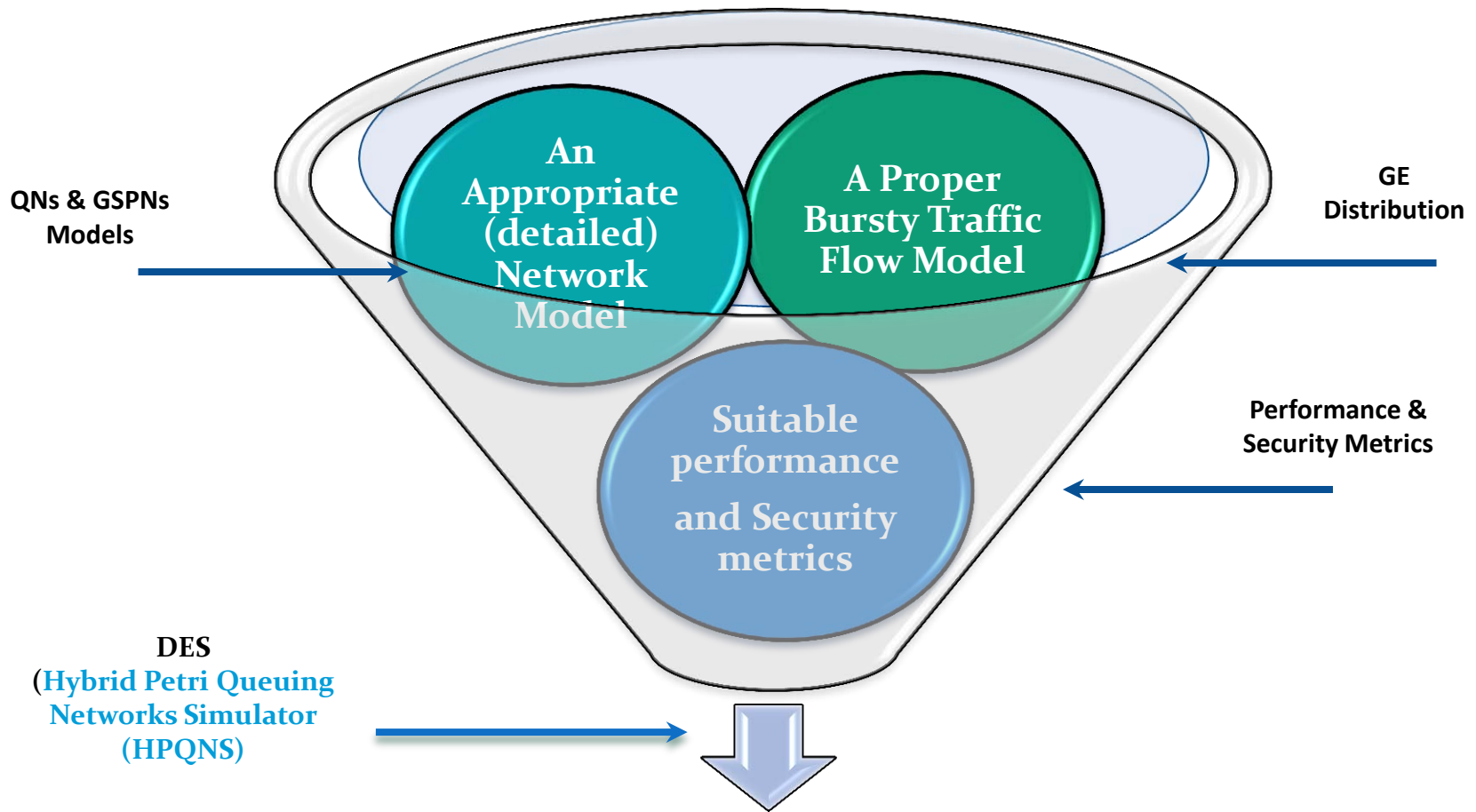
Key GSPNs Advantages & Limitations

Advantages	Limitations
<ul style="list-style-type: none"> • Have great “descriptive power” Allow easy modelling of blocking, synchronization, simultaneous resource possession, security control and software contention. 	<ul style="list-style-type: none"> • The solution of GSPNs for large systems, is computationally very expensive due to exponentially growth of the state space.

RANETs: Related Concepts

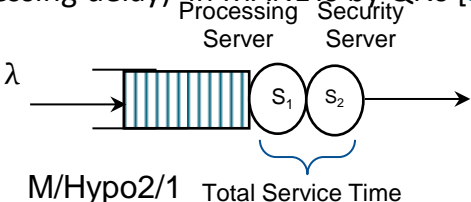
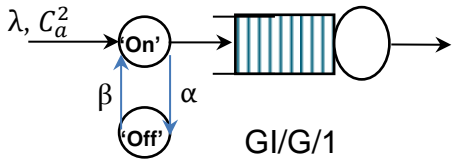
- RANETs is an infrastructure-less networks at which MANETs is used as underlying technology;
- Robots can be '*tele-operated*' (controlled by a human being) or '*autonomous*' (work without human interaction);
- Robots among the RANET might be *heterogeneous* in terms of sensory skills, mobility and robot's architecture.
- The main five components of any robot are: Controller, Sensing Unit, Power Unit, Transmission Unit, Actuators.
- **RANETs Encryption Protocols:** provide confidentiality (via encryption) and/ or integrity to the exchanged data; however, they add extra bits which require more processing time!

The Main Concept



Quantitative Methodology of Performance Degradation and Associated Trade-off in RANETs Under Bursty Traffic Conditions

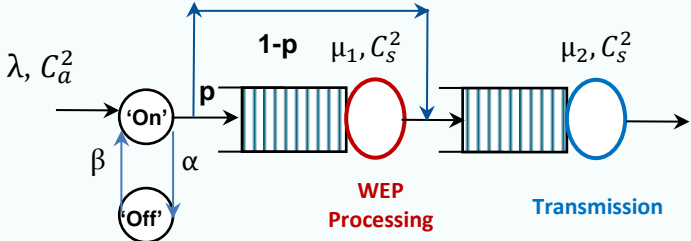
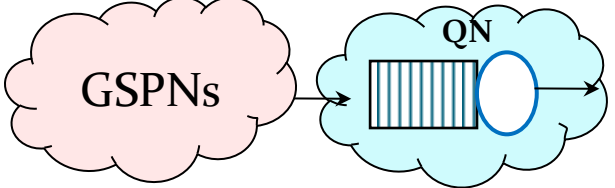
Previous studies– Security & Mobility Modelling

Modelling Approach	Application Context & Modelling Level	Performance and/or Security Metrics
<p>Explicit modelling of Security (Security processing delay) in MANETs by QNs [Saleh05]</p>  <p>M/Hypo2/1 Total Service Time</p>	<ul style="list-style-type: none"> • A MANET's node runs WEP encryption protocol (<i>Node-level</i>) • security and forwarding service times are modelled explicitly by Hypo2 in the queueing model. 	<ul style="list-style-type: none"> ▪ Performance Metrics: mean throughput & packet loss
<p>Mobility Modelling in MANETs by G-QNs [Bhatia07].</p>  <p>GI/G/1</p>	<p>(G-Queues) indicates node's mobility subject to IPP with α & β transition rates (the model is made at <i>Network-level</i>)</p>	<ul style="list-style-type: none"> • Performance Metric: mean marginal end-to-end delay.

Previous studies– Performance & Security Trade-off Modelling Tools

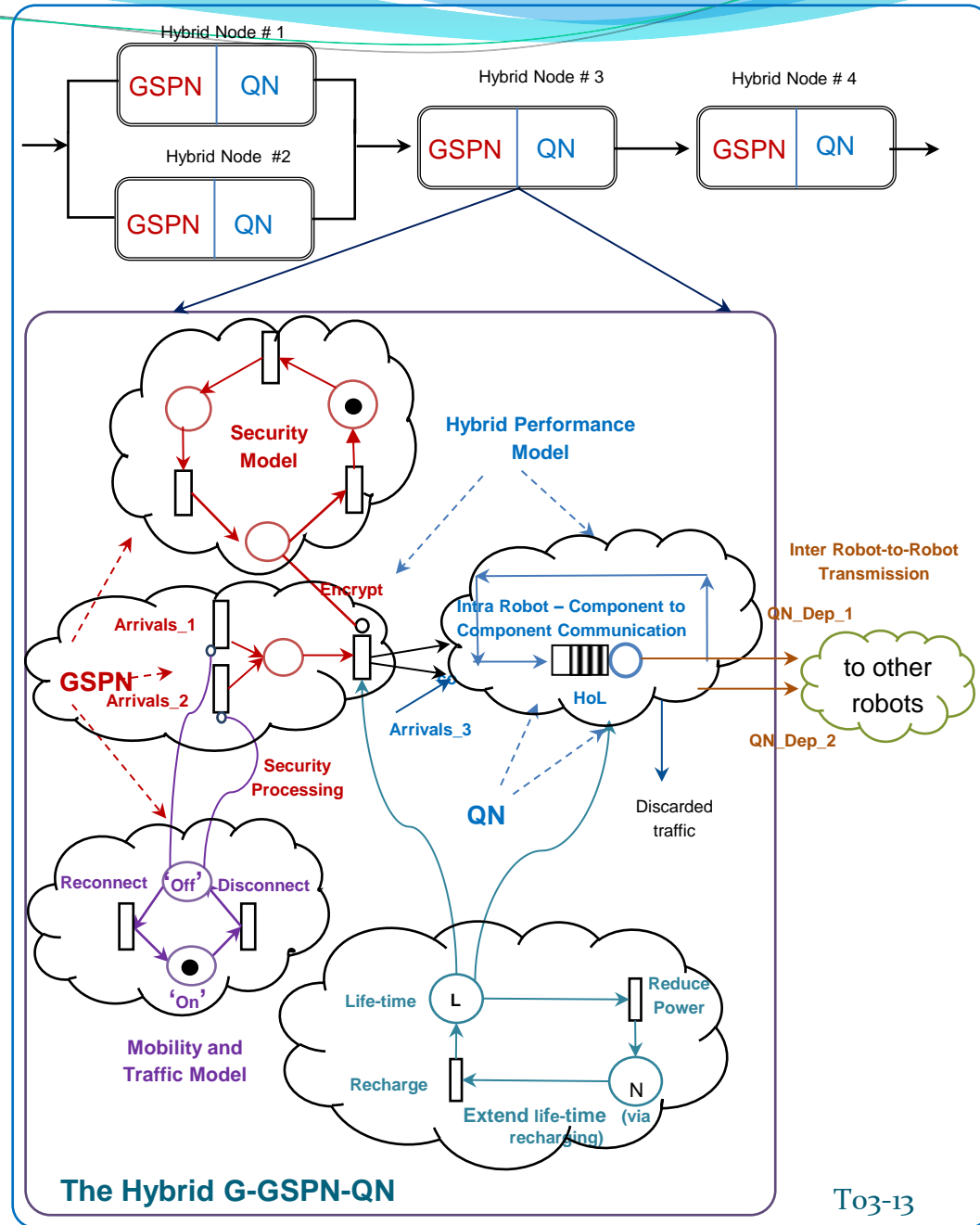
Modelling Approach	Modelling Level & Application Context	Performance and/or Security Metrics
<p><u>Wolter 's model</u> [Wolter10] GSPN to trade-off between the system' throughput and the system being secure. It provides 'Explicitly Detailed-Modelling' of Security Processing & Control.</p>	<p>Combined GSPN performance-security model (at <i>Node-level</i>) for an abstract communication system to evaluate & simultaneously optimise performance and security trade-off using <u>CPSMs</u> [Wolter10].</p>	<ul style="list-style-type: none"> ▪ Performance Metric: Throughput. ▪ Security Metrics: Probability of system being secure, unsecure, or recovering. ▪ Analysis Approach: Free Simulation Package ▪ Limitations: <ul style="list-style-type: none"> ➤ Traffic burstiness was not included. ➤ Security sub-model can be extended further.
<p><u>Szczerbicka's Model</u> [Szczerbicka92] (modelling Performability (i.e., the ability of a system to perform)).</p> <ul style="list-style-type: none"> ▪ <u>The Concept of Combined QNs and GSPNs</u> has been introduced by Balbo [Balbo88]. 	<p>Trade-off between performance & Reliability for a communication channel at (<i>Network-Level</i>). <u>Features of combining QNs and GSPNs</u> This combination was introduced to overcome the limitations in modelling power of QN and the state space explosion in GSPNs when modelling large systems). Combining & exploiting the <i>best features</i> of both QNs and GSPNs to provide an effective . In the context of DES, adopting combined QNs & GSPNs will help in reducing the <i>time</i> required to update transitions status after each firing event in a <i>pure GSPN</i> model.</p>	<ul style="list-style-type: none"> ▪ Combining QNs and GSPNs ▪ Performance Metric: ratio between mean throughput of correctly transmitted jobs and the throughput of all transmitted jobs. ▪ Analysis Approach: mathematical model based on GSPN ▪ Limitations: <ul style="list-style-type: none"> ➤ Traffic burstiness was not included. ➤ Customised model.

Employing the Quantitative Methodology in (RANETs)

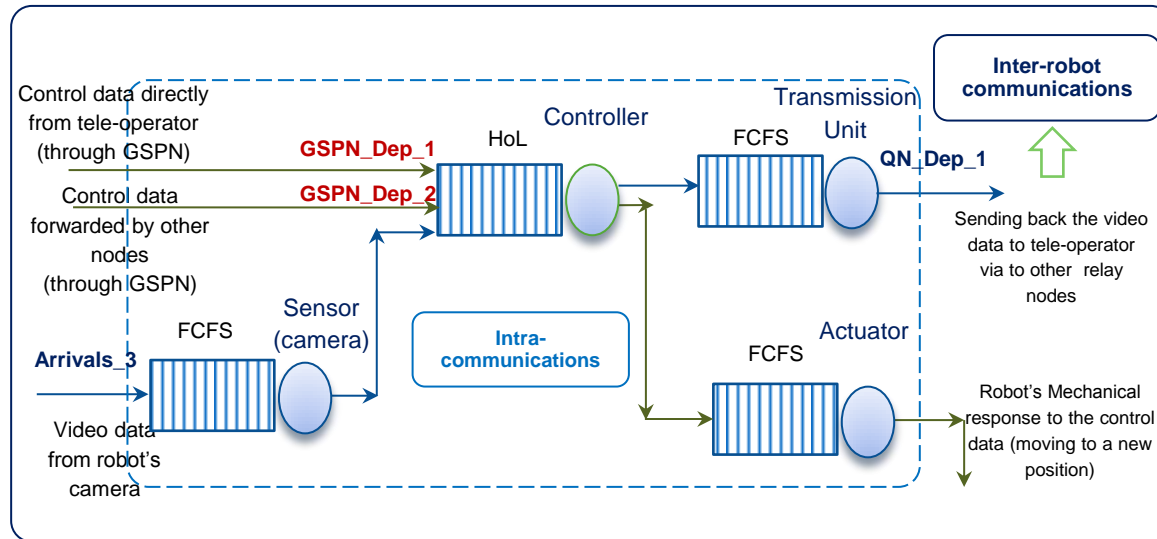
Proposed Models, Application Context, Adopted Modelling Approach & Evaluation Tool)	Modelling Level & Performance and/or Security Metrics
<p>WEP & SS (Security processing is Explicitly Modelled by G-QNs [KOU12])</p> 	<ul style="list-style-type: none"> ▪ Network-Level Incoming traffic is lost when the gate is 'Off', i.e., the node is disconnected from the network. ▪ Performance Metric: mean marginal end-to-end delay. ▪ Security Metrics: N/A • Limitation: Security processing delay in WEP node was only parameterised through the mean rate and SCV (due to the limited modelling power of QNs thus security control was not explicitly reflected!) • This was the motivation to propose framework in which by replacing QN security node is replaced by detailed GSPN one).
<p>The Proposed Hybrid G-GSPNs & QNs Framework (Explicit & Detailed-Modelling of Security processing and Control, mobility, Power consumption & intra/inter-robot communications) with associated two Extended CPSMs).</p> 	<ul style="list-style-type: none"> ▪ Network-Level ▪ Extended CPSM (based on CPSM proposed by Wolter)

The Proposed Hybrid QNs and GSPNs Framework

- It models, at node-level, for each node by explicitly modelling: security processing and control with mobility, power consumption using GSPN;
- QNs reflect 'intra'-robot (**component-to-component**) communication & 'inter'-robot transmissions.
- The **main feature** is to model more effectively the performance & security behaviour of RANET node & subject extended CPSMs.



Case Study



QN model for intra and inter robot communications

- Communication between the robot and tele-operator can be directly or via intermediate mobile nodes;
- Video transmission can be delayed due to the transmission media and encryption protocol's computations which might prevent the '*Timely control*' of the robot;
- Therefore, investigating a balanced trade-off in will reduce transmission delay

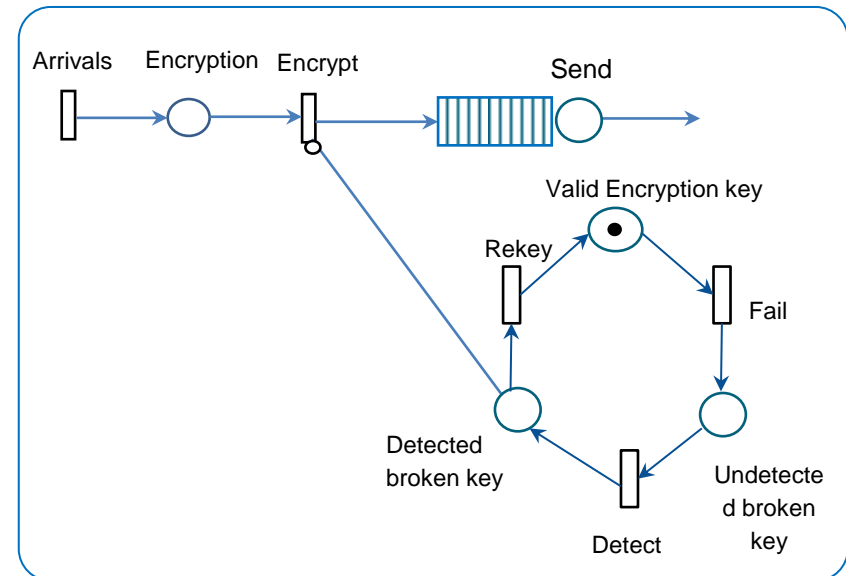
while preserves reasonable security level and video quality.

Extended CPSMs

- Selected performance metrics, which are **compatible** in scale of magnitude e.g., Utilisation **U** and **PLP** are used to resolve the issue in Wolter's CPSM to facilitate investigating and optimising RANET's performance vs. security trade-offs.

Definition of the Extended CPSMs

- These CPSMs are defined in terms of the model in the Figure, as following:
 - CPSM1** = $U + P(\text{Valid Encryption Key})$
 - CPSM2** = PLP in 'encryption' place + $P(\text{Undetected Broken Key})$.



Discrete Event Simulation (DES)

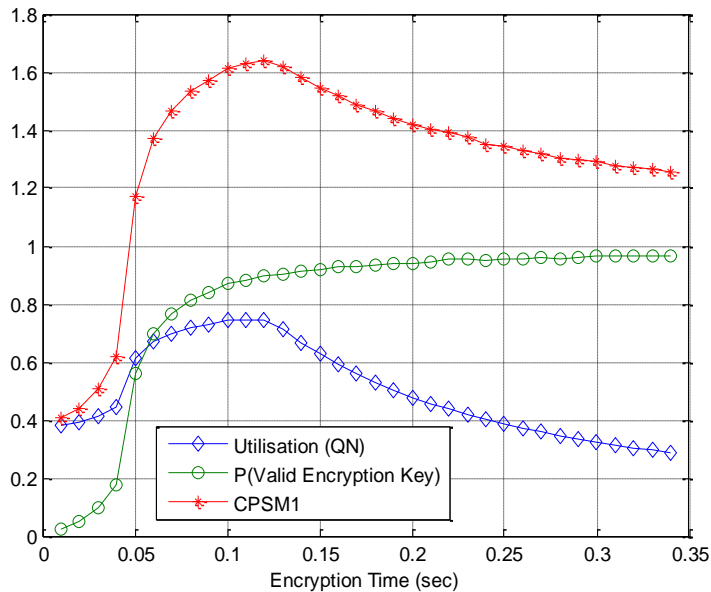
- DES, implemented in java, has been used for its flexibility, simplicity & ability to analyse complex systems since it overcomes queueing theory's limitation and it model reality in greater detail;
- A system's model can be constructed and the required measurements are calculated over time to obtain performance metrics of interest;
- DES is used to generate GE random variable as well as simulating QNs and GSPNs.
- **Hybrid Petri Queuing Networks Simulator (HPQNS)** has been developed to evaluate the trade-off in RANETs

Input parameters of hybrid QN and GSPSN

Parameter	Value
Arrivals	8 messages per sec
Encrypt	0.01 to 0.34 step 0.01 sec
Fail (time to key breaking)	1.25, 25, 50, 100, 600 to 15100 step 500 sec
Undetected Broken key	12 sec
Rekey	36 sec
Send	10 messages per sec
SCV of inter-arrival times	1, 50, 100
Capacity of 'Encryption' place and 'Send' queue buffer	100

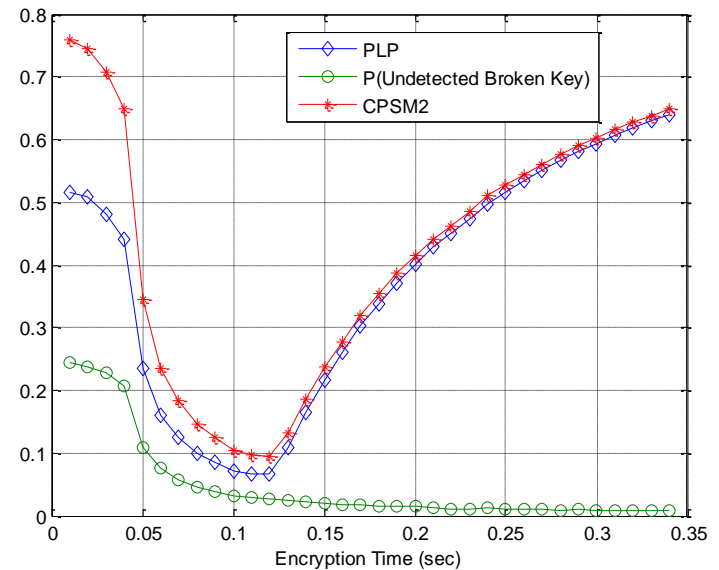
Results

CPSM1 and Corresponding Components



$$\text{CPSM1} = \text{Utilisation} + \text{P(Valid Encryption Key)}$$

CPSM2 and Corresponding Components

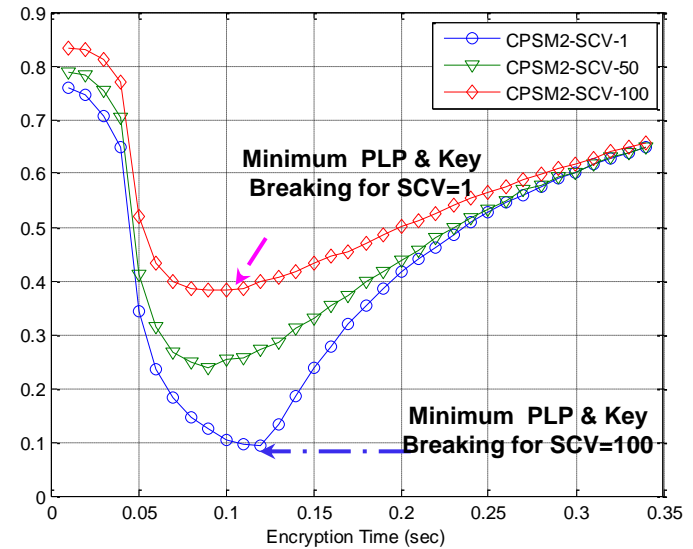
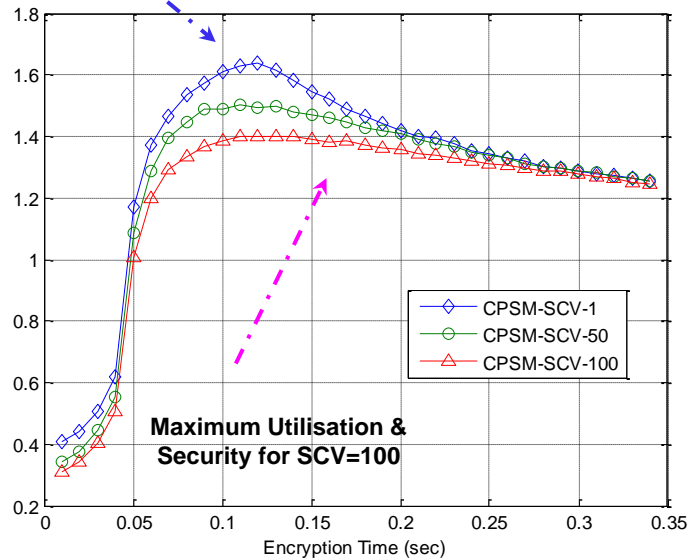


$$\text{CPSM2} = \text{PLP in 'encryption' place} + \text{P(Undetected Broken Key)}$$

SCV Impact on CPSM₁

SCV Impact on CPSM₂

Maximum Utilisation & Security for SCV=1



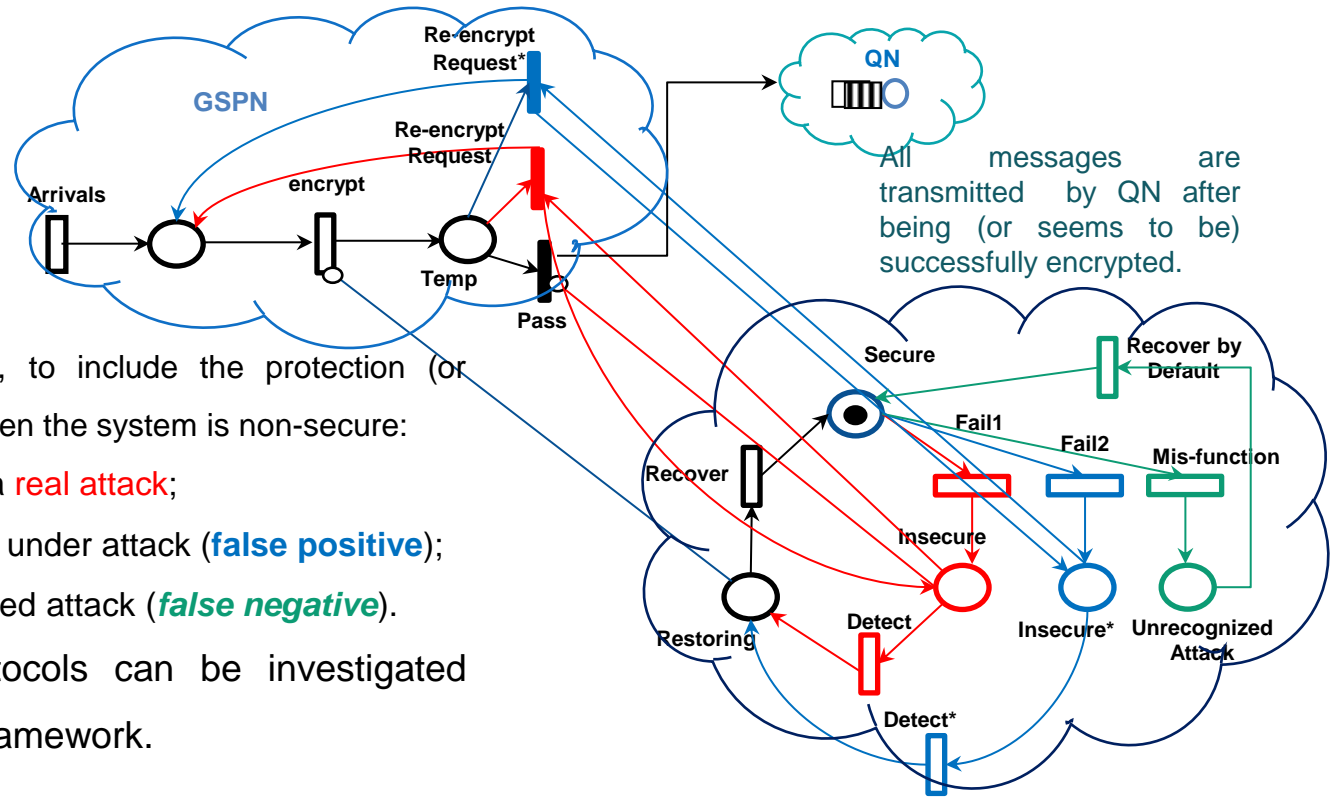
SCV	Max- CMPS1	Optimised Encryption Time (sec)
1	1.64	0.12
50	1.479	0.13
100	1.3998	0.14

SCV	Min- CMPS2	Optimised Encryption Time (sec)
1	0.146	0.11
50	0.311	0.12
100	0.444	0.13

Final Remarks & Conclusions

- This study contributes in **quantifying** and predicting performance **degradation – at node level-** in RANETs caused due to volume incoming traffic in conjunction with security computations by the use of GE and exploiting the **Extended CPSMs** to facilitate investigating and optimising the trade-offs;
- The proposed framework reflects the **behaviour** of **security** mechanisms as well as **intra-and inter- robot** communications which enables capturing the **robot's heterogeneity** (in terms of the robot architecture and application);
- The impact of increasing the **burstiness** of input traffic **increases** the encryption time required for **simultaneous** optimisation of performance and security (Extended CPSM).

Future work: The Extended Model



Three scenarios are considered, to include the protection (or reduction) of information leakage when the system is non-secure:

- (1) system becomes unsecure by a **real attack**;
- (2) system erroneously considered under attack (**false positive**);
- (3) system being under unrecognised attack (**false negative**).

More advanced security protocols can be investigated utilising the proposed hybrid framework.

Net-CPSM₁ & **Net-CPSM₂** (at network-level) of N hybrid nodes can be expressed as:

$$Net_CPSM_1 = \sum_{i=1}^N \frac{U_i}{N} + \sum_{i=1}^N \frac{P_i(\text{system in secure state})}{N}$$

$$Net_CPSM_2 = \sum_{i=1}^N \frac{PLP_i}{N} + \sum_{i=1}^N \frac{P_i(\text{system in detected broken key})}{N}$$

Furthermore, **Aggregate CPSMs** & **Marginal CPSMs** for R arrival classes in the hybrid network can be determined in a similar way.

The conversion from a security state to another is caused by enabling and firing the following transitions:

Transitions

Pass: messages are passed to QN after being encrypted using a valid key.	Fail1: the system fails by a real attack.	Detect: when the integrity check discovers that the encryption key is compromised, it will inform/ alarm the system.	Recover: the system recover from a real attack by performing the rekeying process.	Re-encrypt Request: Any encrypted messages with invalid key has to be re-encrypted.	Fail2: The system fails by a non- real attack
Detect*: The system recover from non- real attack.	Recover: The system recover from non- real attack.	Re-encrypt Request *:Any messages is suspected to be encrypted with invalid key (but in reality it is valid, i.e. false positive detection) has to be re-encrypted.	Mis-function: the possibility of the system being secure whilst it is under a real attack (so fail3 leads to unrecognized attack state)	Recover by Default*: the system recovers from non- real attack, by performing rekeying process after a particular number of messages (or given period of time).	

Places (representing the security status of the model)

Secure: A system is secure (encrypted messages are allowed to pass to the QN to be transmitted).	Temp: Temporary buffer in which all encrypted messages are further checked whether they have been encrypted by a valid key or not .	Insecure: the system becomes non-secure by a real attack (compromised key)	Insecure*: The system is secure but is considered non-secure due to error occurrence (false positive) – (inability to detect a broken key).	Unrecognized Attack: The system is non- secure but was considered secure by error (false negative), i.e., inability to detect the key is compromised
---	--	---	--	---

Key References

- G. Balbo, S. C. Bruell, and S. Ghanta, "Combining queueing networks and generalized stochastic petri nets for the solution of complex models of system behavior," *IEEE Trans. on Computers*, vol. 37, no. 10, pp. 1251-1268, 1988.
- H. Bhatia, R. Lening, S. Srivastava, and V. Sunitha, "Application of QNA to analyze the queueing network mobility model of MANET," Technical Report, Dhirubhai Ambani Institute of Information & Communication Technology (DAIICT), Gandhinagar, India, 2007.
- D. D. Kouvatsos, "A maximum entropy analysis of the G/G/1 queue at equilibrium," *The Journal of the Operational Research Society*, vol. 39, no. 2, pp. 183-200, 1988.
- D. D. Kouvatsos, "Entropy maximisation and queueing network models," *Annals of Operations Research*, vol. 48, no. 1, pp. 63-126, 1994.
- M. Saleh and I. Al Khatib, "Performance of secure ad hoc sensor networks utilizing IEEE802.11b WEP," in *Proc. of Systems Communications*, Montreal, Que., Canada, 2005, pp. 68 - 72.
- H. Szczerbicka, "A combined queueing network and stochastic Petri net approach for evaluating the performability of fault-tolerant computer systems," *Performance Evaluation*, vol. 14, pp. 217-226, 1992.
- H. Szczerbicka and P. Ziegler, "Simulation with active objects: an approach to combined modelling," *Simulation Practice and Theory*, vol. 1, no. 6, pp. 267–281, 1994.
- K. Wolter and P. Reinecke "Performance and security trade-off," in *Formal Methods for Quantitative Aspects of Programming Languages, Computer science Lecture Notes*, vol. 6154, A. Aldini, M. Bernardo, A. Di Pierro, and H. Wiklicky, Eds. Berlin, Heidelberg: Springer 2010, pp. 135-167.

Acronyms

- **CDF** Cumulative Distribution Function
- **CPSM** Combined Performance and Security Measure
- **DES** Discrete-Event Simulation
- **FCFS** First-Come-First-Served
- **G-QNs** Gated QNs
- **G-Queue** Gated-Queue
- **G-GSPN_QN** Gated-Generalised Stochastic Petri Nets_QNs
- **GE** Generalised Exponential
- **GSPNs** Generalised Stochastic Petri Nets
- **HoL** Head-of-the-Line
- **IPP** Interrupted Poisson Process
- **MANETs** Mobile Ad Hoc Networks
- **MTTF** Mean Time To Security Failure
- **PDF** Probability Density Function
- **PNs** Petri Nets
- **PLP** Packet Loss Probability
- **PFQN** Product-Form Queueing Networks
- **RANETs** Robotic Ad Hoc Networks
- **QNs** Queueing Networks
- **QoS** Quality-of-Service
- **SCV** Squared Coefficient of Variation
- **SPNs** Stochastic Petri Nets
- **SS** Selective Security
- **WEP** Wired Equivalent Privacy

The End of the Presentation

Thank
you



Wolter's GSPN Model

- It provides *Explicitly Detailed-Modelling* of Security Processing & Control by GSPNs.
- Combined GSPN performance-security model for an abstract communication system to evaluate & simultaneously optimise the trade-off between performance & security using *CPSMs*.

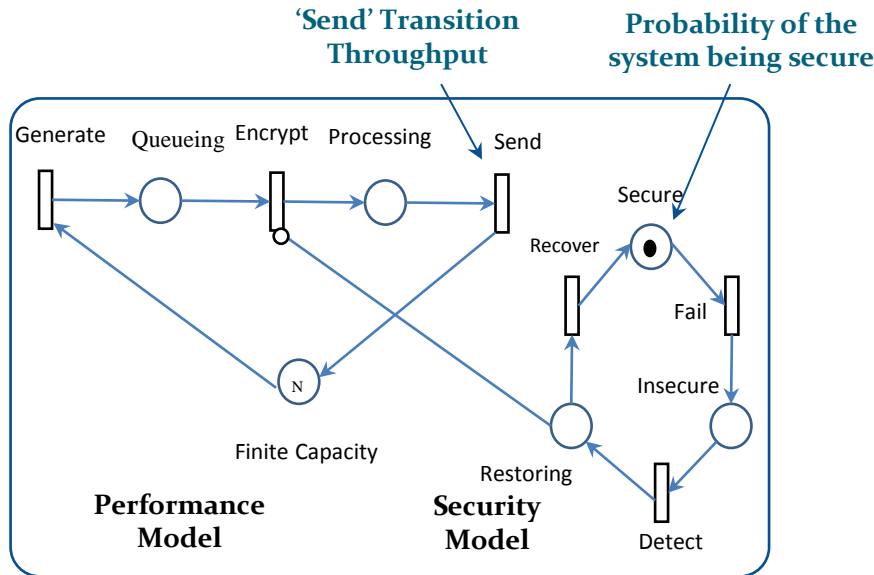


Fig. A GSPN for RANETs [WOL10]

CSPM	Throughput of the transition 'Send'+ Probability of system to be secure state.
------	--

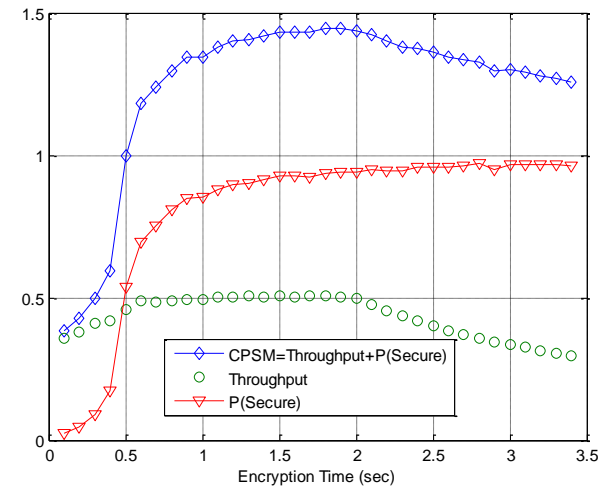


Fig. B CPSM proposed by Wolter [WOL10]

This model investigates how to choose an appropriate encryption key length in terms of its corresponding encryption time (the longer the key size the longer the encryption time!).

- **Remarks:** Throughput is very large compared with probability values! which dominates the overall value of the *CPSM*).



Szczerbicka 's Combined QN and GSPN Model

- Trade-off model between performance & Reliability for a data processing system transmits jobs to a remote facility via a transmission channel that is broken and corrected and thus number of jobs are re-transmitted.
- The higher channel's failure rates \rightarrow the lower the system's throughput:
(i.e., too many jobs to reprocess by the limited capacity system!

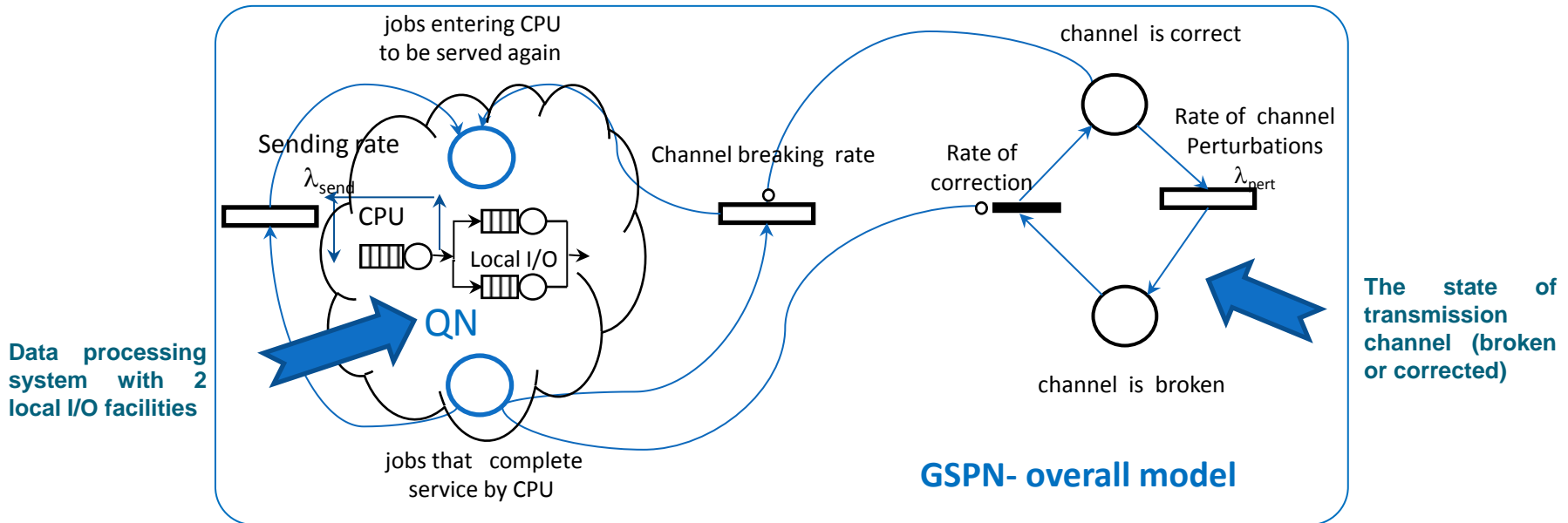


Fig. The Combined QN and GSPN proposed by zczerbicka's



PART FOUR Network Mobility, Management and QoS

Analysing Tradeoffs in Mobile Offloading Systems

Katinka Wolter

Work with Qiushi Wang, Huaming Wu, Martí Giera Jorba,
Joan Martínez Ripoll

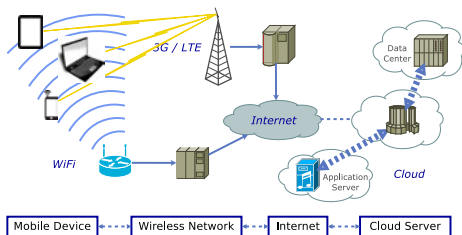
Freie Universität Berlin
Institut für Informatik

12th November 2013

Table of Contents

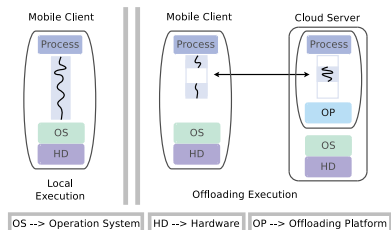
- 1 Mobile offloading on the application level
- 2 Evaluation Approaches
- 3 Reliable offloading system
- 4 Unreliable offloading system architecture
- 5 Dependability and performance models
- 6 Metrics for offloading
- 7 Experiment and simulation
- 8 Conclusions

Mobile Offloading System



- Mobile offloads to a cloud server

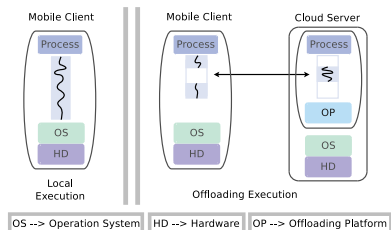
Mobile Offloading System



- Mobile offloads to a cloud server
- Offloading system \implies *Objectives* :

Shorten execution time + Reduce energy consumption

Mobile Offloading System



- Mobile offloads to a cloud server
- Offloading system \implies *Objectives* :
Shorten execution time + Reduce energy consumption
- Tradeoff in the failure-free system:
 - Energy usage versus communication cost

Mobile Offloading System



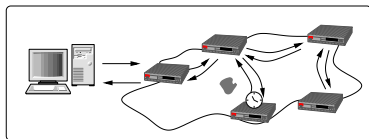
- Mobile offloads to a cloud server
- Offloading system \implies *Objectives* :
Shorten execution time + Reduce energy consumption
- Tradeoff in the failure-free system:
 - Energy usage versus communication cost
- Tradeoff in the unreliable system

Mobile Offloading System



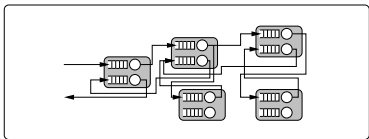
- Mobile offloads to a cloud server
- Offloading system \implies *Objectives* :
Shorten execution time + Reduce energy consumption
- Tradeoff in the failure-free system:
 - Energy usage versus communication cost
- Tradeoff in the unreliable system
 - Restart offloading or do local re-execution
 - How long should one wait and when to launch?

Evaluation Approaches



- Experimentation on test-beds – low abstraction level
 - Set up the system in the lab, emulate faults using stochastic fault-models
 - Implement restart algorithms
 - Measure response-times with and without restart

Evaluation Approaches



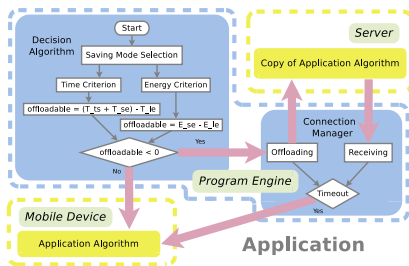
- Experimentation on test-beds – low abstraction level
 - Set up the system in the lab, emulate faults using stochastic fault-models
 - Implement restart algorithms
 - Measure response-times with and without restart
- Simulations – medium abstraction level
 - Build a simulation (e.g. NS-2, OMNeT++, Möbius)
 - Observe response-times with and without restart

Evaluation Approaches

$$F(x) = \int_0^x f(u)du$$

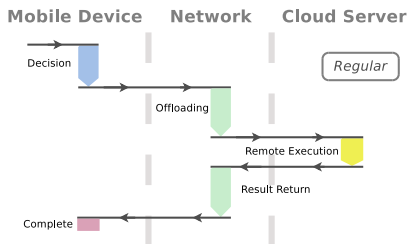
- Experimentation on test-beds – low abstraction level
 - Set up the system in the lab, emulate faults using stochastic fault-models
 - Implement restart algorithms
 - Measure response-times with and without restart
- Simulations – medium abstraction level
 - Build a simulation (e.g. NS-2, OMNeT++, Möbius)
 - Observe response-times with and without restart
- Analytical Approaches – high abstraction level
 - Formalise problem
 - Compute completion-times with and without restart

Reliable offloading system architecture



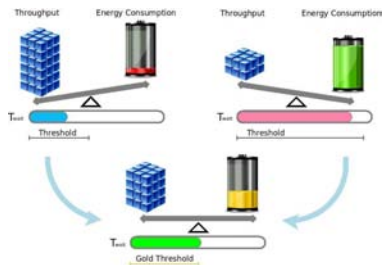
- Program engine and decision algorithm run on mobile device, that interacts with cloud server

Reliable offloading system architecture



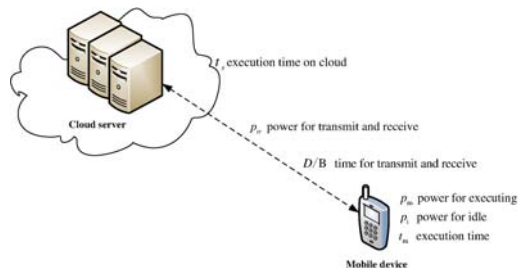
- Program engine and decision algorithm run on mobile device, that interacts with cloud server
- Offloading uses copy of algorithm in the server

Reliable offloading system architecture



- Program engine and decision algorithm run on mobile device, that interacts with cloud server
- Offloading uses copy of algorithm in the server
- When to offload?

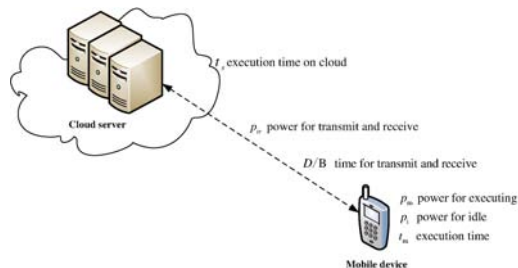
Offloading decisions



$$C = D/B \quad \text{communication time}$$
$$p_{tr} = pC \quad \text{communication cost}$$

- Offloading condition for time: $t_m > t_s + C$

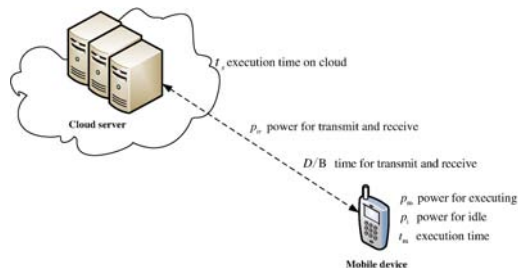
Offloading decisions



$$C = D/B \quad \text{communication time}$$
$$p_{tr} = p_C \quad \text{communication cost}$$

- Offloading condition for time: $t_m > t_s + C$
- Offloading condition for power: $p_m t_m > p_i t_s + p_C C$

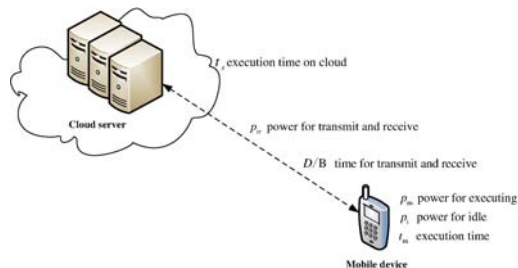
Offloading decisions



$$C = D/B \quad \text{communication time}$$
$$p_{tr} = p_C \quad \text{communication cost}$$

- Offloading condition for time: $t_m > t_s + C$
- Offloading condition for power: $p_m t_m > p_i t_s + p_C C$
- Assume time on server is multiple of mobile time
 $t_s = N t_m, 0 < N < 1$.

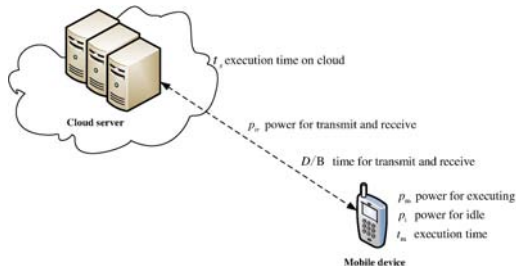
Offloading decisions



$$C = D/B \quad \text{communication time}$$
$$p_{tr} = p_C \quad \text{communication cost}$$

- Offloading condition for time: $t_m > t_s + C$
- Offloading condition for power: $p_m t_m > p_i t_s + p_C C$
- Assume time on server is multiple of mobile time
 $t_s = N t_m, 0 < N < 1$.
- Define ratio: $G_1 = \frac{t_m}{N t_m + C}$ and

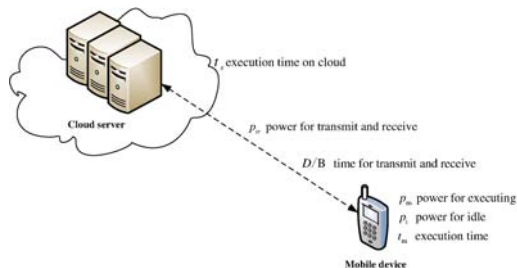
Offloading decisions



$$C = D/B \quad \text{communication time}$$
$$p_{tr} = p_C \quad \text{communication cost}$$

- Offloading condition for time: $t_m > t_s + C$
- Offloading condition for power: $p_m t_m > p_i t_s + p_C C$
- Assume time on server is multiple of mobile time
 $t_s = N t_m, 0 < N < 1$.
- Define ratio: $G_1 = \frac{t_m}{N t_m + C}$ and
- $G_2 = \frac{p_m t_m}{N p_i t_m + p_C C}$

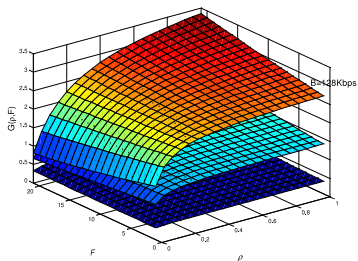
Offloading decisions



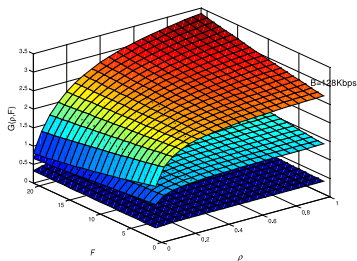
$$C = D/B \quad \text{communication time}$$
$$p_{tr} = p_C \quad \text{communication cost}$$

- Offloading condition for time: $t_m > t_s + C$
- Offloading condition for power: $p_m t_m > p_i t_s + p_C C$
- Assume time on server is multiple of mobile time
 $t_s = N t_m, 0 < N < 1$.
- Define ratio: $G_1 = \frac{t_m}{N t_m + C}$ and
- $G_2 = \frac{p_m t_m}{N p_i t_m + p_C C}$
- Combined metric: $G = G_1^\rho \cdot G_2^{1-\rho}$, where $0 < \rho < 1$

Offloading tradeoff guidelines

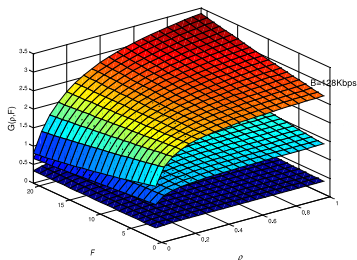


Offloading tradeoff guidelines



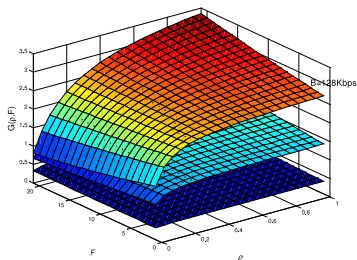
- Network bandwidth B has large impact

Offloading tradeoff guidelines



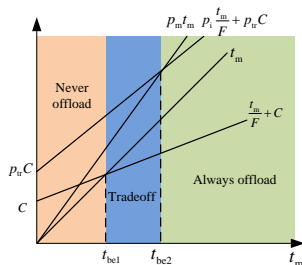
- Network bandwidth B has large impact
- The larger $F = \frac{1}{N}$ the more can be saved

Offloading tradeoff guidelines



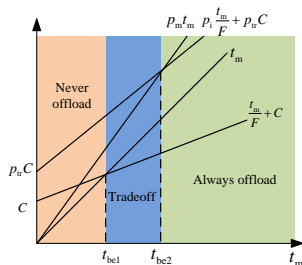
- Network bandwidth B has large impact
- The larger $F = \frac{1}{N}$ the more can be saved
- ρ must be chosen with care

Offloading tradeoff guidelines



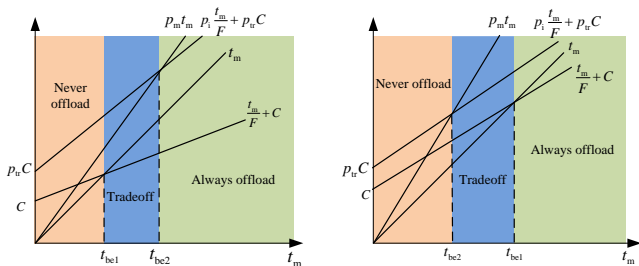
- Network bandwidth B has large impact
- The larger $F = \frac{1}{N}$ the more can be saved
- ρ must be chosen with care
- Break-even for time at $t_{be1} = \frac{C}{1-N}$ and for energy at $t_{be2} = \frac{p_r C}{p_m - N p_i}$

Offloading tradeoff guidelines



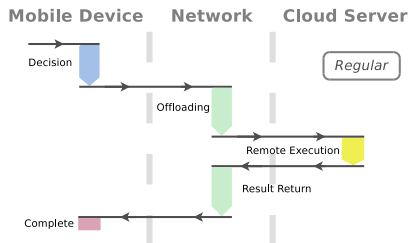
- Network bandwidth B has large impact
- The larger $F = \frac{1}{N}$ the more can be saved
- ρ must be chosen with care
- Break-even for time at $t_{be1} = \frac{C}{1-N}$ and for energy at $t_{be2} = \frac{pC}{p_m - Np_i}$
- If $t_{be1} < t_{be2}$ save time, waste energy

Offloading tradeoff guidelines



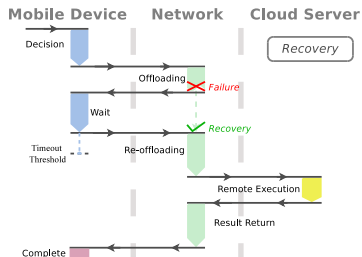
- Network bandwidth B has large impact
- The larger $F = \frac{1}{N}$ the more can be saved
- ρ must be chosen with care
- Break-even for time at $t_{be1} = \frac{C}{1-N}$ and for energy at $t_{be2} = \frac{pC}{p_m - Np_i}$
- If $t_{be1} < t_{be2}$ save time, waste energy
- If $t_{be2} < t_{be1}$ save energy, waste time.

Unreliable offloading system architecture



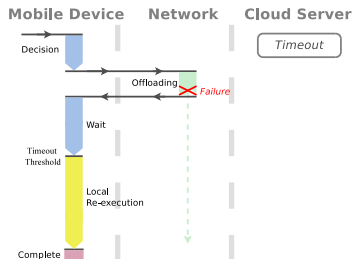
- Offloading uses copy of algorithm in the server

Unreliable offloading system architecture



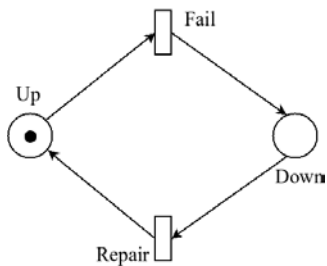
- Offloading uses copy of algorithm in the server
- Recovery process restarts offloading

Unreliable offloading system architecture



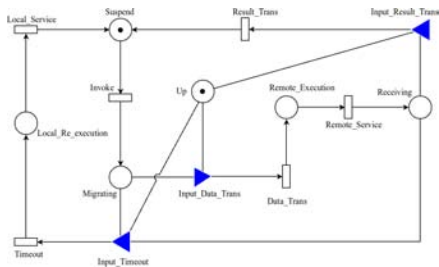
- Offloading uses copy of algorithm in the server
- Recovery process restarts offloading
- Local re-execution process will not offload

Stochastic activity networks



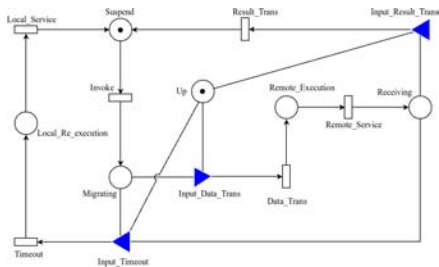
- Wireless Network Model
- Firing time of activities T_{fail} and T_{repair} is exponentially distributed.

Stochastic activity networks



- Wireless Network Model
- Firing time of activities T_{fail} and T_{repair} is exponentially distributed.
- Execution State Model
- $T_{timeout} = P \times (E[T_{local_service}] + 2 \times E[T_{data_trans}])$

Stochastic activity networks



- Wireless Network Model
- Firing time of activities T_{fail} and T_{repair} is exponentially distributed.
- Execution State Model
- $T_{timeout} = P \times (E[T_{local_service}] + 2 \times E[T_{data_trans}])$
- Solve model using simulation

Instability

$$Pr_{instability} = Pr((\#Migrating = 1 \vee \#Receiving = 1) \wedge \#Up = 0)$$

Instability

$$Pr_{instability} = Pr((\#Migrating = 1 \vee \#Receiving = 1) \wedge \#Up = 0)$$

Throughput

H : number of tasks completed in the simulated system lifetime of a given simulation, i.e. number of *Invoke* activations.

Instability

$$Pr_{instability} = Pr((\#Migrating = 1 \vee \#Receiving = 1) \wedge \#Up = 0)$$

Throughput

H : number of tasks completed in the simulated system lifetime of a given simulation, i.e. number of *Invoke* activations.

Energy Consumption

$$E'_{re} = T_{trans} \times p_{wifi} + T_{idle} \times p_{idle} + T_{re} \times p_{cpu}$$
$$E_{re} = E'_{re}/H$$

Combination of metrics

Normalization

- Three metrics have different orders of magnitude and units.
- All data of the same metric $\rightarrow [0, 1]$

$$y = \frac{x - \text{Min}(X)}{\text{Max}(X) - \text{Min}(X)}$$

Combination of metrics

Normalization

- Three metrics have different orders of magnitude and units.
- All data of the same metric $\rightarrow [0, 1]$

$$y = \frac{x - \text{Min}(X)}{\text{Max}(X) - \text{Min}(X)}$$

Geometric Distance

- Three metrics make up a three-dimensional vector $\langle \mu_i, e_i, h_i \rangle$.
- μ_i : Instability, e_i : Energy Consumption, h_i : Throughput
- Performance \sim The distance from the best vector $\langle 0, 0, 1 \rangle$.

$$A = \sqrt{w_1 \times u_i^2 + w_2 \times e_i^2 + w_3 \times (1 - h_i)^2}$$

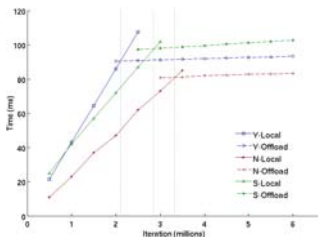
$\langle w_1, w_2, w_3 \rangle$ the weight vector for different scenarios

The testbed



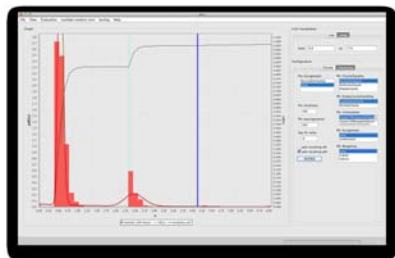
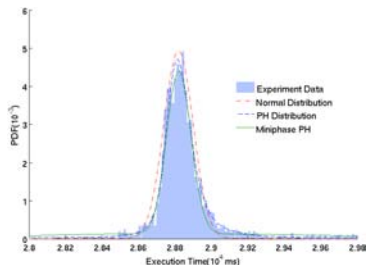
- Mobile Devices:
 - Y: Samsung Galaxy Young, 832 MHz;
 - N: Samsung Galaxy Nexus, 1.2 GHz Dual Core;
 - S: Sony Xperia MT25, 1 GHz;
- Wireless Network: 100M/s WiFi
- Cloud Server:
 - 1 Server with 4 cores (Intel Xeon CPU E5649 2.53 GHz)
 - 2 Google App Engine

The testbed



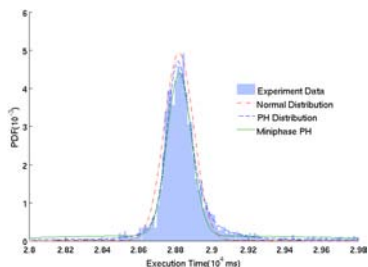
- Mobile Devices:
 - Y: Samsung Galaxy Young, 832 MHz;
 - N: Samsung Galaxy Nexus, 1.2 GHz Dual Core;
 - S: Sony Xperia MT25, 1 GHz;
- Wireless Network: 100M/s WiFi
- Cloud Server:
 - 1 Server with 4 cores (Intel Xeon CPU E5649 2.53 GHz)
 - 2 Google App Engine
- Sample Application: Iteration $\rightarrow i + +$
 - 1 Mobile device: Execution time increases quickly with the number of iteration.
 - 2 Remote server: The number of iterations has nearly no impact.

Parameter Fitting



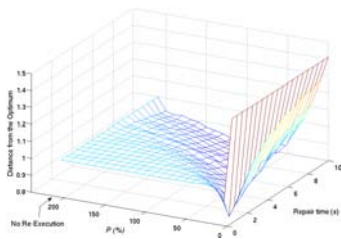
- Tool: HyperStar (a cluster-based algorithm)
- Phase-type (PH) distribution → Accurate but large state space.

Parameter Fitting



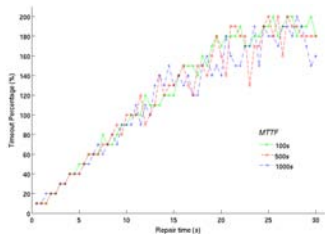
- Tool: HyperStar (a cluster-based algorithm)
- Phase-type (PH) distribution → Accurate but large state space.
- Normal distribution → Low accuracy on both tails but shorter simulation time.

Simulation Results



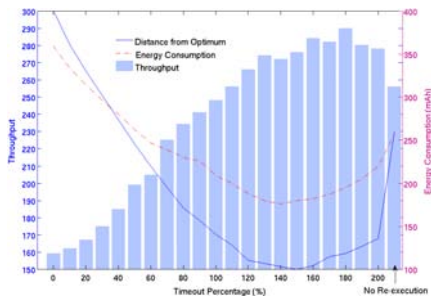
- The bottom is best

Simulation Results



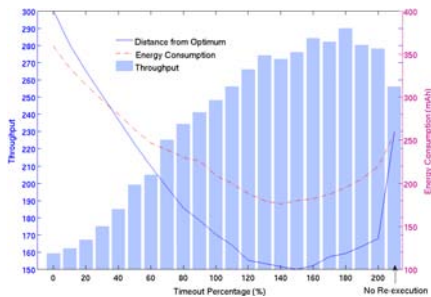
- The bottom is best
- Mean time to failure (MTTF) has no impact
- The optimal timeout increases with the repair time (MTTR).

Validation by Experiment



- Experimental metrics and their difference from simulation results.

Validation by Experiment



- Experimental metrics and their difference from simulation results.
- The distances of the same timeout under different MTTRs are summed up.
- Throughput reaches the upper limit as the timeout is close to the optimum.
- Energy Consumption is minimal at the optimum timeout.

Conclusions

- Mobile offloading can help to improve user experience.

Conclusions

- Mobile offloading can help to improve user experience.
- In order to achieve good performance, different tradeoffs must be considered:

Conclusions

- Mobile offloading can help to improve user experience.
- In order to achieve good performance, different tradeoffs must be considered:
 - Energy versus connectivity

Conclusions

- Mobile offloading can help to improve user experience.
- In order to achieve good performance, different tradeoffs must be considered:
 - Energy versus connectivity
 - Retry versus local re-execution

Conclusions

- Mobile offloading can help to improve user experience.
- In order to achieve good performance, different tradeoffs must be considered:
 - Energy versus connectivity
 - Retry versus local re-execution

Reliable case

- Combination of time and energy criteria must be considered
- Connectivity is important
- Propose decision criteria for offloading

Conclusions

- Mobile offloading can help to improve user experience.
- In order to achieve good performance, different tradeoffs must be considered:
 - Energy versus connectivity
 - Retry versus local re-execution

Reliable case

- Combination of time and energy criteria must be considered
- Connectivity is important
- Propose decision criteria for offloading

Unreliable case

- Decide whether to retry or re-execute locally
- Decision based on SAN model
- Parameters obtained from testbed

Open issues

- Can we build a general offloading engine?
- Take decisions online?

Open issues

- Can we build a general offloading engine?
- Take decisions online?
- We need to consider security!

Open issues

- Can we build a general offloading engine?
- Take decisions online?
- We need to consider security! As a cost?

Thank you.

A CIM-based approach for managing computing servers and hypervisors acting as active network elements

Dimitris Kontoudis, Panayotis Fouliras
University of Macedonia
Department of Applied Informatics
156 Egnatia str., 54006, Greece
{kontoudis, pfoul}@uom.gr

Abstract. Communication network implementations present an ever increasing use of computing servers and hypervisors assuming the role of active network elements. This fact, along with the virtualization concept which has, also, been adopted in the computer networks field, introduces new challenges that need to be accounted for when managing such networks. In this paper we propose a new modular and extensible information model which can be applied to most scenarios of network architectures where hypervisors are involved, facilitating the efficient representation and management of the provisioned computing server resources.

Keywords: Hypervisor, Network Virtualization, Modeling, Management, CIM, SPC, Statistical Process Control

1 Introduction

The convergence of communications and computing to a common design and operational entity is an inevitable reality, introducing the network engineers with new elements and technologies that need to be taken into account when designing and implementing network architectures. The latter are increasingly based on virtualized infrastructures – both from the networking as well as from the computing realms. Some of these new elements can be quite novel and totally outside of the traditional network design scope and expertise.

Communication networks in particular implementations, for example in “testbeds” [1], in network simulation environments [2] and in scenario-based infrastructure management [3], increasingly rely on the use of computing servers to act as active network nodes (Fig.1), as these servers allow for flexible experimentation on new architectures, protocols and services. This has been made possible due to advances in server virtualization (also referenced to as system or machine virtualization, implemented by a specific thin software layer, the hypervisor, also known as the virtual machine monitor/VMM). The network’s “last-hop switch” has, consequently, been shifted from the pure active network elements to become a characteristic of the hypervisor or of the physical server’s hardware [4]. In the Cloud Computing and in the Networking and Computing Infrastructure as a Service (IaaS) approaches [5], for example, the

server loads – thus, now, the active network elements – can and should dynamically shift between physical servers in the same data center or, even, in data centers at different geographical locations. The core networking support in server virtualization environments is based on the IEEE 802.1Q VLAN implementation. Here, virtual network segments are created on top of the server’s hardware features (the hypervisor works as a virtual Ethernet switch and supports queues for each VLAN in the system’s memory). In this way it is possible to establish network communication across different virtual servers without routing network traffic outside the physical system that hosts them, performing the virtualization.

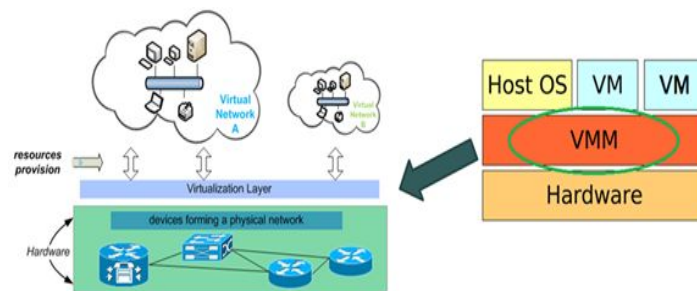


Fig. 1. Communication networks using virtualized computing server resources

New issues are, therefore, raised and need to be considered in network design, operation, monitoring and administration: the involved physical servers’ hardware resources (capacity of processors, memory, virtual switch etc.), the optimization of these resources so as to reach a required system performance and behavior, etc. To tackle these problems, from standardization and management points of view, it is essential that the network’s architectural, technological and operational complexity is semantically represented in a formal way, allowing for virtualization and hypervisor support. Our research has been motivated by the aforementioned issues and we worked towards designing a model that is concise and extensible so as to be able to capture virtualized components and their characteristics in a variety of architectures. We propose the “*KF model*” based on DMTF’s CIM standard [6]. Our proposed work can, conceptually, capture any physical or logical element that can be instrumented by CIM (thus, allowing the object’s management too), be that a networking, a computing system or other resource.

The remainder of this paper is organized as follows: section 2 provides an overview of related research. Section 3 describes the proposed model. We conclude in section 4 by summarizing the findings of our research and presenting future directions.

2 Related work

Many parties, industrial, commercial and academic alike, are actively involved in network virtualization research on a wide variety of topics, ranging from very specific

technical issues (interfacing, signaling and bootstrapping, resource and topology discovery, resource allocation, admission control, virtual nodes and virtual links, naming and addressing, mobility management, monitoring, configuration and failure handling, security and privacy, interoperability) to large scale network implementations, like GENI [1], EmuLab [7], PlanetLab [8], PanLab [9], VINI [10][11], commonly referred to as “testbeds”. A concise survey, from a general perspective, of network virtualization research is provided in [12] and [13].

Virtualization hypervisor concepts in a modeling and management context (whether networking-related or not) are, largely, overlooked. Sporadic support can be found in some proposals but is limited in scope compared to the complexity and details involved in managing a hypervisor. Current information models treat hypervisors as transparent elements of the virtualization layer and begin abstraction from the virtual system or virtual network point. Partial and indirect support can only be found in the Common Information Model (CIM) [6] (proposed by the Distributed Management Task Force [14]), the DEN-ng model [15] as well as in Management Information Bases (MIBs) [16]. In CIM, a hypervisor (not a virtual machine) can be instantiated as a subclass via the `OperatingSystem` class and the built-in hypervisor virtual switch, respectively, via the `UnitaryComputerSystem` class. Although CIM (in the System Virtualization Model [17]) elaborates on modeling and management actions on a virtual machine and on its host computer system, it does not account for the hypervisor layer. DEN-ng, in a similar fashion, could be extended via subclassing from either the `PhysicalResource` and `LogicalResource` or the `VirtualSystem` and `VirtualImage` classes. In the MIBs domain the only relevant references are the VM-MIB [18] and the VMM-MIB [19], both at IETF draft status. These MIB objects can store basic hypervisor information (list of guest virtual machines, virtual CPU information and mappings of logical storage and network interfaces). Current hypervisor technologies are very complex and incorporate several details and operational specifics than what can be abstracted and managed by current models.

3 The proposed information model

To meet the need for semantic representation of virtualized computing server resources provisioned to computer networks we propose the KF model, a novel CIM-based conceptual representation of the different components that constitute a virtual machine-based network. The KF model can cover the physical and logical components supporting the virtual network along with its settings, modes of operation and statistical elements of the hypervisors and virtual machines involved. The KF model is extensible so as to include, nearly, any new element that needs to be introduced, in a hardware-agnostic way. As a result, the model can be applied to a wide variety of scenarios and is not depended on any particular hardware implementation. This model semantically incorporates, at the logical level, a virtual network spanning a number of virtual server hosts (which act as active network elements and provide its core resources) along with the virtualization techniques (physical nodes, hypervisors, virtual machines – VMs) [20] employed in such a design. System provisioned resources

(such as CPUs, memory and I/O) as well as other relevant operational parameters are included in the model. The model’s representation of the virtual machine includes methods for the control of the provisioned resources based on statistical analysis of their performance. Given the agnostic nature of the model all virtualization platforms are supported as long as proper providers are developed adhering to the CIM approach.

3.1 Conceptual Design

The KF model consists of ten classes (Fig.2) representing the virtual machine, network, configuration and operation parts of the NVE architecture¹. These classes, along with the extensibility characteristics of the model, are adequate for representing the basic design of any network employing virtual machines hosted on a hypervisor (this approach is of similar nature regardless the choice of hypervisor). Additional features and facilities can be abstracted by extending the model, thus eliminating the need for initial classes overcommit which would incur difficulties in the model design.

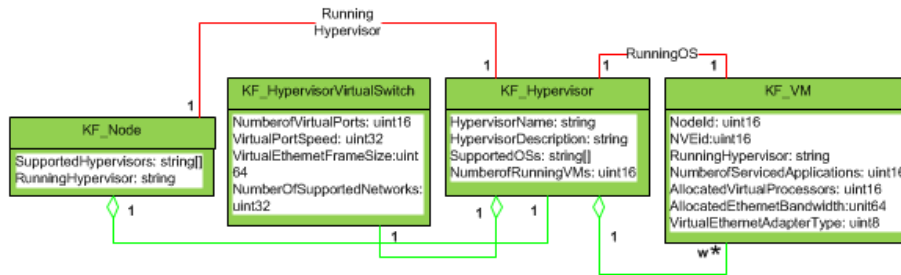


Fig. 2. KF model sample classes and associations¹

An exclusive namespace has been applied to class naming by which each class name is prefixed with a “KF_” convention. In the current version of the model the following elements of a virtual host based network are referenced: computing (systems) nodes, hypervisors, hypervisor virtual switches, virtual machines, processes, applications, virtual networks along with their settings and statistics. All these elements are semantically represented at the logical level and the virtual network environment, as a whole entity, is conceived as a collection of the referenced entities. This design approach allows for the simplification of handling and for the consolidation of global characteristics, such as settings, statistics, naming etc. The virtual machine part uses six classes handling the physical server infrastructure as an hardware node with a running hypervisor, a number of participating VMs, and the processes and applications running on these VMs. Intra-node networking is, in part, represented by a specific “KF_HypervisorVirtualSwitch” class which details the hypervisor’s in-memory

¹ Detailed UML diagrams, complete textual class descriptions as well as MOFs are available at <http://users.uom.gr/~kontoudis/research/>

IEEE802.1Q VLAN compatible virtual switch. Networking information is, also, shared with the “KF_Network” class which, also, includes the properties needed for mapping a virtual host based network notion as a whole entity. A special class is used for handling statistical data resulting in a total of three classes dealing with networking information. A number of associations have been designed which, being double-ended references, return specific operational data depending on the invocation method (i.e. reporting how many virtual Ethernet adapters are allocated per VM, which is the physical node’s running hypervisor, which applications operate per VM and per virtual network, etc.)

The KF model design allows for the inclusion of any manageable entity by implementing proper extensions which can augment the model’s scope and, thus, the managing application’s (that makes use of the model) functionality. For example, suppose that the need arises for the handling of transaction-based performance characteristics or for the management of a virtual router instantiated by a virtual server. The former need can be tackled by a CIM_UnitOfWork derivative subclass [21] whereas the latter need via extending the KF_VM class to include the required management methods. This extensibility of the KF model derives from its design logic and allows for the easy inclusion of new features and elements. CIM schemas are expressed in UML and their syntax description is composed in the Managed Object Format (MOF) [22] (a .mof file is a text file that defines the class name and attributes of a managed resource).

4 Conclusions and Future Work

Recent technological advancements, where virtualized computing play an integral part in computer networks, complicate the network’s architecture, operation and management, introducing new aspects that need to be considered for proper end-to-end service delivery. In this paper we introduced the KF model, a CIM-based approach showing that standardization of the representation of virtual networks where computing servers are involved is, indeed, feasible. The model allows for the conceptual representation of involved components and for the introduction of targeted actions against them. Our ongoing work focuses on enhancing the KF model with semantics for performance patterns and characteristics of network and system performance. Furthermore, we investigate the application of statistical process control, by means of operating system techniques, for creating dynamically adaptive virtual machine performance, adhering to SLA-specified constraints.

References

1. GENI, <http://www.geni.net>, 2013
2. Canonico, R., Emma, D. and Ventre, G., *An XML Description Language for Web-based Network Simulation*, in Proceedings of the 7th IEEE International Symposium on Distributed Simulation and Real-Time Applications, IEEE Press, DOI=<http://dx.doi.org/10.1109/DISRTA.2003.1242989>, Delft, Netherlands, pp. 76-81, 2003

3. Galan, F., Fernandez, D., Fuertes, W., Gomez, M. and Vergara, J., *Scenario-based Virtual Infrastructure Management in Research and Educational Testbeds with VNUML*, Annals of Telecommunications, Vol. 64, No.5&6, pp. 305-323, 2009
4. Creeger, M., *Moving to the Edge: A CTO Roundtable on Network Virtualization*, *Communications of the ACM*, DOI= <http://dx.doi.org/10.1145/1787234.1787251>, Vol. 53, No. 8, , pp. 55-62, 2010
5. Armbrust, M., Fox, A., Griffith, R., Joseph, A., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I. and Zaharia, M., *Above the Clouds: A Berkeley View of Cloud Computing*, Technical Report No. UCB/EECS-2009-28, University of California at Berkeley, <http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html>, 2009.
6. The Common Information Model (CIM), <http://www.dmtf.org/standards/cim>, 2013
7. EmuLab, <http://www.emulab.net>, 2013
8. PlanetLab, <http://www.planet-lab.org>, 2013
9. PanLab, <http://www.panlab.net>, 2012
10. VINI, <http://www.vini-veritas.net>, 2013
11. Boucadair, M., Georgatsos, P., Wang, N., Griffin, D., Pavlou, G., Howarth, M. and Elizondo, A., *The AGAVE Approach for Network Virtualization: Differentiated Services Delivery*, Annals of Telecommunications, DOI= <http://dx.doi.org/10.1007/s12243-009-0103-4>, Vol. 64, No. 5, pp. 277-288, 2009
12. Chowdhury, N. M. M. K. and Boutaba, R., *A Survey of Network Virtualization*, Computer Networks, DOI= <http://dx.doi.org/10.1016/j.comnet.2009.10.017>, Vol. 54, No. 5, pp. 862-876, 2010
13. Chowdhury, N. M. M. K. and Boutaba, R., *Network Virtualization: State of the Art and Research Challenges*, IEEE Communications Magazine, DOI= <http://dx.doi.org/10.1109/MCOM.2009.5183468>, Vol. 47, No. 7, pp. 20-26, 2009
14. Distributed Management Task Force, <http://www.dmtf.org>, 2013
15. J. Strassner, S. Van der Meer, D. O 'Sullivan and S. Dobson, *The Use of Context-Aware Policies and Ontologies to Facilitate Business-Aware Network Management*, Journal of Network and Systems Management, Vol. 17, No. 3, pp. 225–284, 2009
16. B. Fenner, MIB index, <http://www.icir.org/fenner/mibs/mib-index.html>, 2012
17. CIM System Virtualization Profile, http://www.dmtf.org/standards/published_documents/DSP2013_1.0.0.pdf, 2012
18. M. McFaden, J. Schoenwaelder, T. Tsou and C. Zhou, *Definition of Managed Objects for Virtual Machines Controlled by a Hypervisor*, <http://tools.ietf.org/html/draft-schoenw-opsawg-vm-mib-01>, 2012
19. H. Asai, Y. Sekiya, K. Shima and H. Esaki, *Management Information Base for the Virtual Machine Manager*, <http://tools.ietf.org/html/draft-asai-vmm-mib-00>, 2012
20. Fenn, M., Murphy, M., Martin, J. and Goasguen, S., *An Evaluation of KVM for Use in Cloud Computing*, in Proceedings of the 2nd International Conference on Virtual Computing Initiative (ICVCI), IBM Corp. RTP, Raleigh, USA, 2008
21. CIM Metrics Model, <http://www.dmtf.org/sites/default/files/standards/documents/DSP0141.pdf>, 2012
22. Managed Object Format, <http://www.omg.org/mof>, 2013

Investigating the Quality of Service of Wireless Network using Constant Bit Rate Traffic under Random based Mobility Management

Shahzad Rizwan¹, Farhan Aadil² and Faiz ur Rahman³

^{1,2}COMSATS Institute of Information Technology Attock Campus, Attock, Pakistan

³The University of Agriculture, Peshawar Pakistan

Abstract

Ad hoc networks are defined by multi-hop wireless networks, with sudden change in network structure and topology i.e. they are provided with lack of predefined infrastructure. The nodes are mobile in nature and perform function as host and a router which discover routes and forward packets to other nodes. Due to dynamic topology, routing and mobility management is a challenging task. Ensuring the Quality of Service is difficult task due to random mobility pattern of the nodes and limited battery power and bandwidth. In this research paper, the Quality of Service parameters are investigated under two prominent Mobility Models. i.e. Random Waypoint Model (RWP) and Random Direction (RD) Model using Constant Bit Rate (CBR) Traffic Model under on-demand routing protocol for mobile ad hoc networks—We demonstrated that by using Two Ray Ground propagation model. Based on the observations, the conclusion and recommendations are given about how the performance of Quality of Service can be improved in Wireless Networks.

Keywords: Ad hoc; Ad Hoc On-demand Distance Vector Routing (AODV); Random Waypoint (RWP); Two Ray Ground;

1 INTRODUCTION (*MANETS*)

Recently mobile ad hoc networks (MANETs) have established great attention because of their self-organizing property and self-configuring capabilities. At the same time early research effort assumed an open-ended environment which pays an important consideration on different problems like multihop routing, wireless channel access and as well as security. Security has become a major objective to supply protected communication between nodes in a possible aggressive situation [1].

adfa, p. 1, 2011.

© Springer-Verlag Berlin Heidelberg 2011

Mobile Ad-Hoc Network (MANET) is the combination of different wireless nodes in the network, in which every single node uses the multi hops relations without infrastructure or centralized administration. Depend on the prerequisite, different nodes in wireless network can modify its method dynamically and randomly setup routes between source and destination [2]. The topology in network can modify quickly, when nodes move in a wireless environment. So, it is likely that packets have to be sent through different routes every time. They are three categorizes of routing protocols. i.e. proactive, reactive and hybrid [3]. Within proactive routing protocols nodes keep routing information to all the other nodes of the network. When topology changes, it reorganized the routing information after predefined time interval [4]. Reactive protocols as the name suggests, only try to keep up to date routing information to the destinations when needed. i.e. network topology is kept recent as on demand [5].

2 ENTITY BASED MOBILITY MODELS

The following entity based mobility models are described under:

1. RANDOM WAYPOINT MODEL (RWP)

The Random Waypoint Mobility Model causes mobile nodes (MNs) to change direction and/or speed randomly. It introduces speed with pause time. A MN starts by choosing random destination and moves towards it, when reaches it stays for a certain period of time of pause time. This process is repeated again and again until the end of simulation time [6].

2. RANDOM DIRECTION MODEL (RD)

The Random Direction Mobility Model (Royer et al, 2001) [7] is created to overcome node's density produced by the Random Waypoint Mobility Model. In Random Direction Mobility Model a MNs randomly chooses a random direction in similar fashion as in Random Waypoint and Random Walk Mobility Model. Upon reaching the simulation border, a MN pauses for a specified time and chooses another angular direction (between 0 and 180 degrees) and again continues the process [6]. Thus it creates more network topology and structure which is highly dynamic than Random Waypoint and Random Walk Mobility Models.

3 RELATED WORK

Propagation model is a group of mathematical expressions, algorithms and diagrams used to show the radio characteristics of a known environment [8]. Rhattoy and Zatni (2012) [9] compared the performance of different routing protocols i.e. AODV, DSR and DSDV for using fading propagation models like Rayleigh, Ricean, Shadowing and Nakagami using Network Simulator-2 (NS-2) [12]. Tamilarasan and Sivaram (2012) [2] described the comparative study as well as performance analysis of three mobile ad hoc routing protocols i.e. AODV [11], DSR [13], and TORA [14] with the help of Two Ray Ground Propagation Model using the different parameters i.e. packet delivery ratio, end-to-end delay, path optimality, media access delay, and routing overhead. Debnath et al (2011) [10] examined the non-fading and fading propagation models in ad hoc network. The propagation models that are used in the research work are free-space; two ray ground as well as shadowing models.

4 SIMULATION ENVIRONMENT

All the simulations are conducted using Constant Bit Rate (CBR) traffic application over two mobility models on AODV routing protocol [11] under NS-2 simulator [12]. The simulation parameters are shown in the Table 1.

Table 1. Simulation Parameters

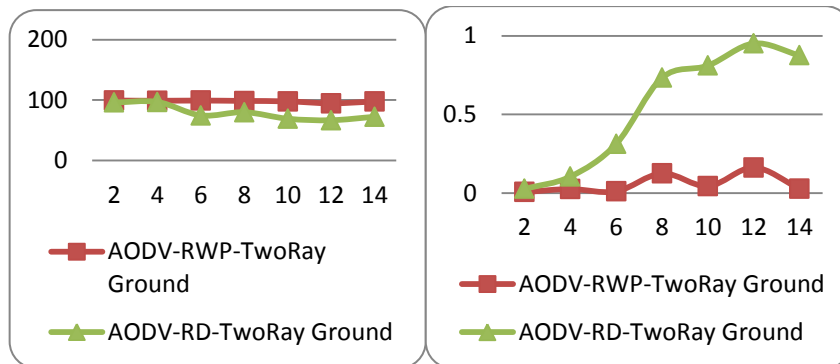
Parameter	Value
Terrain Size	500m x 500m
Number of Nodes	10, 40
Speed	1 to 14 m/sec
Pause time	20, 50, 100, 200, 300, 400, 500 sec
MAC Type	IEEE 802.11
Propagation Model	Two Ray Ground
Transmission Range	150m, 250m
Traffic Model	Constant Bit Rate (CBR)
Simulation Time	600 seconds

4.1 QUALITY OF SERVICE (QoS) METRICS

- Packet Delivery Ratio (PDR): The ratio between the data packets received to data packets send.
- Network Access Delay: The Network Access Delay is the total time experiences during buffering of packets, route discovery phase, queuing delay and MAC packets retransmission delays etc.
- Average Throughput: The throughput is calculated as the total number of data packets received divided by the simulation time.

4.2 SIMULATION RESULTS

1. Scenario 1 with varying max speed for small terrain area (nodes=10): In this scenario, the network size is kept to 10 nodes moving according to RWP and RD mobility patterns in the entire simulation area of 500x500 meters. The CBR traffic sources are kept to 5 connections at sending rate of 4 packets per second for 10 communication nodes. The mobility speed is varied from 2 to 8 m/sec and nodes pause for period of 20 seconds during the entire simulation time of 600 seconds. The graphs are shown below.



(a) Packet Delivery Ratio

Fig 2. Network Access Delay

Fig. 1. QoS metrics vs. Mobility Speed for 10 nodes, Area: 500x500 meters, pause time: 20 seconds, Mobility pattern: RWP and RD, Propagation Model: Two Ray Ground, Traffic Model: CBR, Transmission Range: 250 meters

The Fig. 1. (a) shows that, by changing mobility pattern, the effect of varying mobility speed has more impact on packet delivery ratio of AODV protocol at low CBR traffic load. However, packet delivery ratio of AODV declines when

using RD over Two Ray Ground model, due to fact that in RD nodes chooses angular direction after reaching the boundary of the simulation area, that drastically effect the deviation of convergence of radio signal between sender and receiver. The Fig. 1. (b) shows that the Network Access Delay of AODV across mobility speed for small number of nodes within relatively large area of RD over Two Ray Ground propagation model is surprisingly increases as mobility speed increases beyond 4 m/sec. However, AODV using RWP over Two Ray Ground propagation model also showed irregular increase in Network Access Delay due to more sparseness of nodes in the area. Moreover, Two Ray Ground propagation model suffered a lot in case of Network Access Delay, due to non-direct line of sight propagation of signals and due to other interferences than line of sight propagation.

- Scenario 2 with varying pause time for small terrain area (nodes=40): In this network, 40 nodes are sparsed in area of 500x500 meters with high number of traffic sources. i.e. 20. Also mobility speed is kept to 6 m/sec. The transmission range is decreased to 150 meters for this network.

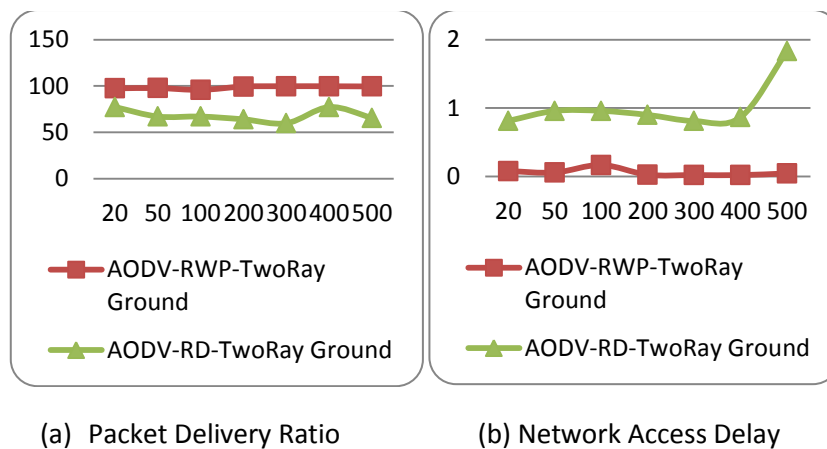


Fig. 2. QoS metrics vs. Pause Time for 40 nodes, Area: 500x500m, Mobility Speed: 6 m/sec, Mobility pattern: RWP and RD, Propagation Model: Two Ray Ground, Traffic Model: CBR, Transmission Range: 150 meters

The graph depicts in Fig. 2. (a) shows that AODV performed excellently in terms of packet delivery ratio under high pause time using RD. i.e. 100% at pause time of 300 seconds than RWP, while AODV showed decline in packet delivery ratio. i.e. 80% in case of RD model. The Fig. 2. (b) shows that de-

crease in transmission range i.e. 150 meters has less impact on AODV using RD and RWP over Two Ray Ground propagation Model. However, RD over Two Ray Ground Model showed highest Network Access Delay of 1.8 seconds at pause time of 500 seconds. i.e. Higher pause time causes less approximation of network convergence and higher hop count.

5 CONCLUSION AND FUTURE WORK

The main conclusion of this research work is that, the random based mobility patterns plays an important role on the quality of service of ad hoc networks under a routing protocol. The Quality of Service parameters fluctuates with the change in propagation model across different traffic and mobility parameters. AODV is a reactive based routing protocol so it is highly robust against sudden increase in mobility speed in dynamic environments. AODV performed excellently considering direct line of sight (LOS) and non direct line of sight (NLOS) paths (i.e. Two Ray Ground) using Random Waypoint and Random Direction Models. When the terrain size and number of nodes are small and also pause time and traffic load set to low, then performance of AODV using Random Waypoint Model is better in terms of Quality of Service parameters with increase in mobility speed at default or large transmission range. Thus it is seen that the performance of AODV in terms of different Quality of Service metrics is highly affected by changing mobility and propagation models. So by proper choosing mobility and traffic models management for specific environment, the Quality of Service of Wireless Networks can be improved.

In future, we like to simulate AODV routing protocol for wireless networks using random based mobility such as Manhattan mobility model and traffic model as constant bit rate (CBR) in Vehicular Ad hoc Networks (VANETs).

6 ACKNOWLEDGMENT

We are highly obliged and thankful for the valuable comments from the anonymous referees that have highly enhanced our research article. We thank Virtual InterNetwork Testbed (VINT) project for developing Network Simulator (NS-2) for providing ns-2 code for our research work.

References

1. Yang, H., Luo, H., Ye, F., Lu, S., Zhang, L.: Security in Mobile Ad Hoc Networks: Challenges and Solutions. *IEEE Wireless Communications*, 11(1), 38-47 (2004)
2. Tamilarasan, S., Sivaram, R.: An Analysis and Comparison of Multi-Hop Ad-Hoc wireless Routing Protocols for Mobile Node. *International Journal of Computer Science and Information Security (IJCSIS)* 10(4), 33-37 (2012)
3. Chatzistavros, E. and Stamatellos, G.: Comparative Performance Evaluation of Routing Algorithms in IEEE 802.11 Ad Hoc Networks. *International Journal of Computer Science Issues (IJCSI)* 7(3), 1-10 (2010)
4. Abolhasan, M. and Wysocki, T.: A Review of routing protocols for mobile ad hoc networks. *Ad Hoc Networks* 2(1), 1-22 (2004)
5. Bisen, D., Suman, P., Shukla, R.S.S.: Effect of Pause Time on DSR, AODV and DYMO Routing Protocols in MANET. 2009.
6. Camp, T., Boleng, J., Davies, V.: A Survey of Mobility Models for Ad Hoc Network Research. *Wireless Communication & Mobile Computing: Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications* 2(5), 483-502 (2002)
7. Royer, E., Melliar-Smith, P.M., Moser, L.: An analysis of the optimum node density for ad hoc mobile networks. In: *ICC 2001: Proceedings of the IEEE International Conference on Communications*, pp. 857-861 (2001)
8. Neskovic, A., Neskovic, N, Paunovic, G.: Modern approaches in modeling of mobile radio systems propagation environment. *IEEE Communications Surveys* 3(3), 2-12 (1999)
9. Rhattoy, A. and Zatni, A.: The Impact of propagation environment and traffic load on the performance of routing protocols in ad hoc networks. *International Journal of Distributed and Parallel Systems (IJDPS)* 3(1), 75-87 (2012)
10. Debnath, D., Hossain, C.A, Islam, R., Tarique, M., Dutta, I. K.: Minimizing Shadowing Effects on Mobile Ad hoc Networks. *Cyber Journals: Multidisciplinary Journals in Science and Technology: Journal of Selected Areas in Telecommunications (JSAT)*, pp. 46-51 (2011)
11. Perkins, C.; Belding-Royer, E., Das, S.: RFC 3561, Ad hoc On-Demand Distance Vector (AODV) Routing. IETF Network Working Group (2003)
12. The Network Simulator - 2. <http://www.isi.edu/nsnam/ns/index.html>
13. Johnson, D. B., Maltz, D. A., Hu, Y.-C.: The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR), IETF Internet Draft, draft-ietf-manet-dsr-09.txt, work in progress (2003)

Park, V., Corson, S.: Temporally-Ordered Routing Alogorithm(TORA) VERSION 1 Internet Draft, draft-ietf-manet-tora-spec- 03.txt, work in progress (2001)

Video Quality Prediction over LTE Using Random Neural Networks

T. Ghalut, H. Larijani and A. Shahrabi

Glasgow Caledonian University
School of Engineering and Built Environment
Glasgow, G4 0BA, Scotland, UK.

Email: {Tarik.Ghalut, H.Larijani, A.Shahrabi}@gcu.ac.uk

Abstract. Long Term Evolution (LTE) is a fourth generation (4G) mobile communication system that supports wireless multimedia applications. The success of video applications over LTE networks very much depends on meeting the Quality of Experience (QoE) requirements of users. The aim of this paper is to present a new QoE prediction model for video over LTE networks. The model uses a combination of parameters in the application and network layer. The application layer parameters considered are the Content Type (CT), Sender Bit Rate (SBR) and Frame Rate (FR) and the network parameters are delay, packet loss rate (PLR). The video quality was predicted in terms of the Mean Opinion Score (MOS). Our method for calculating video MoS is based on statistic learning using Random Neural Network (RNN).

Keywords: Video quality prediction, LTE, QoE, MOS, PSNR, RNN

1 Introduction

QoE framework is an essential element of Fourth Generation (4G) wireless technologies i.e. LTE, its goal to deliver evolving Internet applications to customers, and management of network resources. QoE measures total system performance using subjective and objective measures of customer satisfaction, which assesses the performance of hardware and software services delivered by a vendor. The main aim of subjective measurement is to predict QoE from a given set of Quality of Service (QoS) parameters.

Transmission of video content over wireless communication is increasing exponentially and gaining popularity. In most recent studies conducted by Cisco [1], indicates that global mobile data traffic grew by 70% in 2012, and mobile data traffic will grow at a Compound Annual Growth Rate (CAGR) of 66% between 2012 and 2017. Two-thirds of the world's mobile data traffic will be multimedia by 2017. This increase in multimedia traffic is one of the key drivers of the evolution to new mobile broadband standards i.e. LTE. Recently many mobile Internet applications, such as video call, video streams, and IPTV have different traffic characteristics thus they require a variety of QoS. However, owing to the bandwidth constraints of wireless

networks (low bandwidths and bottlenecks) QoS delivery that is acceptable is difficult to maintain for users. For this reason, the low quality of the multimedia leads to reduced use of applications and services thus leading to lower revenue.

The rest of the paper is organized as follows. Section 2 provides related work on QoE. In section 3 we give briefly describe the fundamentals of RNNs. Section 4 explains our model to evaluate video quality over LTE. Finally, Section 5 draws the conclusion.

2 Related Work

Video quality prediction has many challenges: 1) there are many parameters that influence quality, 2) collection of data for the calculations in WLAN and 3G networks has an overhead. The majority of QoE estimation models use different coefficient methods to map the relationship between the input pattern QoS parameters and the output QoE. For example: the model in [2] used a fixed coefficient, which means the curve fitting algorithm used does not provide the best fit and is only sufficient to prove the proposed concept, while, the models in [3] used an optimization approach for the coefficients based on mapping the fitting functions. These mapping functions are able to find an optimal fit for the given measurement points; these functions derived expressions via mathematical modeling of the dataset to calculate the QoE from the QoS parameters.

In addition to coefficient methods, there are many other artificial intelligence techniques (Machine Learning) used in this field, such as Artificial Neural Networks (ANN), Random Neural Networks (RNN), and Adaptive Neural Inference System (ANFIS). Using the first technique, Ying [4] presented a method that connects the QoE directly to QoS metrics according to the corresponding level of QoE which was estimated by employing a Multilayer Artificial Neural Network (ANN): the network QoS parameters were selected as the input layer, while the MOS, Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) and Video Quality Metric (VQM) as the output layer. The second technique, RNN, was used by [5, 6] to adjust the input network parameters in order to get the ideal output and satisfy the users' need. More specifically, Samir and Rubino [5] proposed a framework for estimating the QoE, based on the RNN and dependent on several parameters: the source Bit Rate (BR), the Frame Rate (FR), the Packet Loss Rate (PLR), the Consecutively Lost Packet (CLP), and the Ratio of the encoded (RA). However, this framework was limited to cabled networks and not applicable to wireless systems which were instead investigated in the work proposed by Piamrat [6] where only three parameters could be applied: 1) LR, 2) packet mean loss burst size (MLBS) and 3) Delay.

Another learning technique can be found in the work of Khan [7] which presented new models to estimate the video quality based on an Adaptive Neural Inference System (ANFIS). The authors predicted the video quality using both application and network layer parameters. However, this work was limited to four parameters: FR, SBR, packet error rate (PER), and bandwidth. These limited numbers of features cannot satisfy the network requirements, especially when network development becomes complex i.e. 4G (LTE).

3 Random Neural Network Models

The Random Neural Network (RNN) is a special type of Neural Network (NN) and is a mathematical representation of neurons or cells which exchange spiked signals. This tool [8] has shown high accuracy and performance more robust. RNN models have many features that make them a more appropriate for modelling the QoE of Video [9]: 1) its standard learning algorithm has low complicity and strong generalization capacity, 2) It represents in a closer manner the signals transmitted in a biological neuronal network than ANNs, 3) It can be easily implemented in both software and hardware. RNNs are composed of a set of interconnected neurons, these neurons exchange signals from one neuron to other, also send and receive signals to and from environment. Each neuron is represented by an integer (+1 or -1) whose value increases when it receives excitation spikes and decreases when an inhibition spike arrives. Thus the excitation spikes are represented as (+1) and inhibition spikes are represented as (-1). The spikes can originate either from outside the network or from another neuron within the network. Neurons which have the excitation state as positive are allowed to send out spikes to either kind of neurons in the network. When a neuron sends out a spike it loses one unit of potential going from state q_i to $q_i - 1$. The probability that the spike signal sent out by neuron i to neuron j as a positive one is represented as $\rho^+_{i,j}$ and a negative one is represented as $\rho^-_{i,j}$.

When a neuron receives a positive signal, either from another neuron or from the environment, its potential is increased by 1; conversely if it is received if a negative one, its potential decreases by 1 if it was strictly positive, and it does not change if its value was 0. Similarly, when a neuron sends a signal, positive or negative, its potential is decreased by one unit it was necessarily strictly positive since only excited neurons send signals [10].

4 Video Quality Prediction Model

In this section we describe our novel model to evaluate video quality over LTE, and with results close to those that can be obtained from objective methods like MOS and PSNR. Initially a database needs to be developed with all the important input and output parameters. This database will be used for both training and experiments. Firstly, the most effective parameters of video quality related to the type of video applications and LTE Networks need to be chosen.

LTE-Sim [11] simulation has been used to generate a video distorted database composed of sequences corresponding to different configurations of the selected parameters. When the simulation uses to send video sequences from the source to destination, every configuration in defined input data must be mapped into the system composed of the network, the source and the receiver. The destination stores the corresponding values of the parameters of the transmitted video sequence. Then, by running the simulation many times, we can generate and store a set of distorted video

sequences with corresponding parameters values.

After completing the distorted database, we choose PSNR [12] as objective quality assessment parameter to compare between original and distorted video sequences, because it is the most common and simple objective video quality assessment used. Then a set of PSNR values were obtained, the corresponding MoS values were extracted. Table 1 shows the correlation of PSNR and MoS values. The PSNR and MOS values with the corresponding parameters' values related to network and application layer were stored in second database called quality database.

In the last step, we must select appropriate RNN architecture and a training algorithm. The quality database is divided randomly into two parts: first one will be used to train the RNN and second one will be used to test its accuracy. The Fig.1 shows feed-forward architecture of 3-layer random neural network model we used.

PSNR [dB]	MOS	Quality
> 37	5	Excellent
31 - 37	4	Good
25 - 31	3	Fair
20 - 25	2	Poor
< 20	1	Bad

Table 1: Possible PSNR to MOS conversion

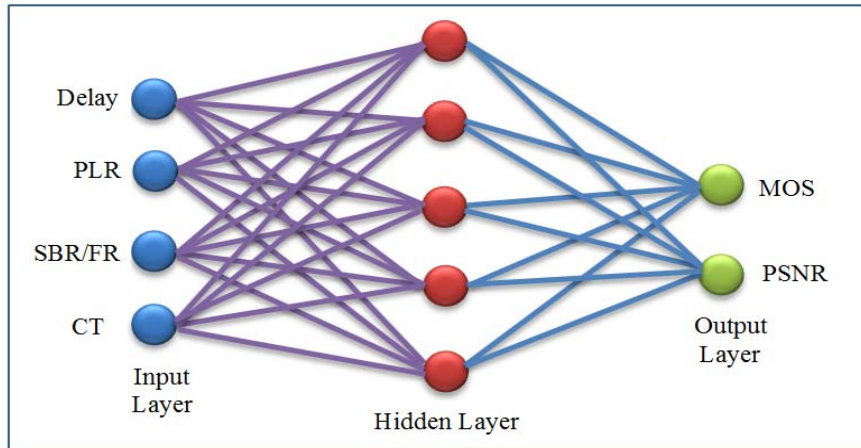


Figure 1: Feed-forward Architecture RNN model

5 Conclusion

This research is currently work in progress and results are preliminary. Further simulations will be carried out and we hope to provide evidence on the suitability of RNN for calculating QoE.

Reference

- [1] Index, Cisco Visual Networking. *Global mobile data traffic forecast update, 2012-2017*, Cisco white paper, 2013.
- [2] T. Wang, et al, *VQM-based QoS/QoE mapping for streaming video*, In Broadband Network and Multimedia Technology (IC-BNMT), 2010 3rd IEEE International Conference on, pp. 807-812, 2010.
- [3] F. Agboma, and A. Liotta, *QoE-aware QoS management*, In Proceedings of the 6th International Conference on Advances in Mobile Computing and Multimedia, pp. 111-116, November 2008.
- [4] P. Calyam, P. Chandrasekaran, G. Trueb, N. Howes, R. Ramnath, D. Yu, L. Ying, L. Xiong, and D. Yang, *Multi-Resolution Multimedia QoE Models for IPTV Applications*, International Journal of Digital Multimedia Broadcasting, 2012.
- [5] S. Mohamed, and G. Rubino, *A study of real-time packet video quality using random neural networks*, Circuits and Systems for Video Technology, IEEE Transactions on, vol. 12, no. 12, pp. 1071-1083, 2002.
- [6] K. Piamrat, C. Viho, J. Bonnin, and A. Ksentini, *Quality of experience measurements for video streaming over wireless networks*, In Information Technology: New Generations (ITNG'09), Sixth International Conference on, pp. 1184-1189, April 2009.
- [7] A. Khan, L. Sun, E. Ifeachor, J. Fajardo, F. Liberal, and H. Koumaras, *Video quality prediction models based on video content dynamics for H. 264 video over UMTS networks*, International Journal of Digital Multimedia Broadcasting, 2010.
- [8] E. Gelenbe, *Random neural networks with negative and positive signals and product form solution*, Neural computation, vol. 1, no. 4, pp. 502-510, 1989.
- [9] S. Timotheou, *The Random Neural Network: A Survey*, The Computer Journal, no. 3, pp. 251-267, 2010.
- [10] E. Gelenbe, *Stability of the Random Neural Network Model*, Neural Computation, vol. 2, pp. 239-247, 1990.
- [11] Piro, Giuseppe, et al, *Simulating LTE cellular systems: an open-source framework*, Vehicular Technology, IEEE Transactions on, vol. 60, no. 2, pp. 498-513, 2011.
- [12] Q. Huynh-Thu, and M. Ghanbari, *Scope of validity of PSNR in image/video quality assessment*, Electronics letters, vol. 44, no. 13, pp. 800-801, 2008.

Work in Progress: Environment Aware Optimum Routing Algorithm (EAR) for resource efficient utilization in Delay Tolerant Network (DTN)

Shahzad Rizwan

Assistant Professor
Department of Computer Science
COMSATS Institute of Information Technology
Attock Campus, Pakistan
email: shehzad.rizwan@ciit-attock.edu.pk

ABSTRACT.

For the last few years there has been enormous growth in intermittently connected networks i.e. ad hoc networks and delay/disruption tolerant networks (DTN). They are usually deployed on temporary basis especially in emergency situation or in disaster environment. This research paper proposes a novel routing scheme called as environment aware optimum routing (EAR) protocol, in order to minimize the computation, operating cost and utilize resources efficiently than existing routing protocols.

Keywords: Delay Tolerant Networks, Entity and Group based Mobility Model, Wireless Network

1 Introduction

Delay-tolerant networking (DTN) is a type of network that is developed in order to cope with technical issues in heterogeneous networks, under any network connectivity condition [1] as shown in the Figure 1. The concept behind the development of DTN is to facilitate the smoothness of routing of data in predefined time interval, to utilize the network resources efficiently, even in lack of end-to-end connectivity. In the later case, the routing protocols uses "store and forward" approach. In this approach, the data is temporary stored in each forwarding hops until it reaches its destination [2][3][4]. A more commonly technique is being used that maximizes the message trans-

ferring probability is to replicate or duplicate many copies of the message in each hop, that eventually results in reaching the destination [5].

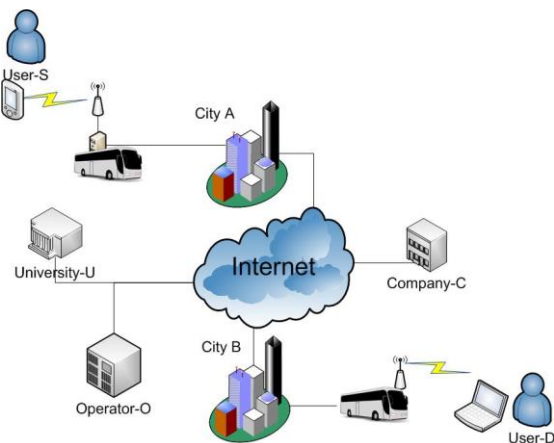


Fig. 1. Example DTN Scenario:
Rural Area Connectivity over Transportation Infrastructures [8]

2 Proposed Work

In recent years, many routing protocols are developed in the field of a Delay-Tolerant Network. The aim of this research work is to devise a routing protocol that considers both contact oblivious and contact based technique. The historic information allows rule of least estimated time for routing bundles. The research work will also be focused on novel environment aware routing protocol (EAR) for efficient resource utilization during non-scheduled contacts and long-duration network partitioning in hostile environments such as disaster and rescue operation etc.

3 Literature Review

Balasubramanian et al. proposed a RAPID Routing protocol for Vehicular intentional DTN testbed. The RAPID “routes” by sending a packet through replication process, until a copy reaches a destination [6]. Jain et al. [7] proposed a forwarding algorithm related to future predicated routing using vehicular network, and tries to predict connection condition that are affected by environmental factors such as extreme weather, radio signal interference, and system failure [6].

4 Conclusion

In this paper we have presented the research project proposal on the development of novel EAR routing protocol for DTNs. Our investigation showed that efficient network resource utilization depends upon several factors such as limited radio range and bandwidth provided by the environment, predication of routes from source to destination, cluster or gateway selection, resilient routing protocol and proper mobility management. We have plan to complete the development of the DTN nodes, cluster and gateway units under the novel EAR for DTN over DTN testbed or simulator.

5 References

1. Wikipedia <http://en.wikipedia.org>
2. Burgess, J., Gallagher, B., Jensen, D., Levine, N. B.: MaxProp: Routing for vehicle-based disruption-tolerant networks.: In: INFOCOM 2006: Proceedings of the 25th IEEE International Conference on Computer Communications, pages 1-11 (2006)
3. Juang, P., Oki, H., Wang, Y., Martonosi, M., Peh, S.L., Rubenstein, D.: Energy-efficient computing for wildlife tracking: design tradeoffs and early experiences with zebranet.: In SIGOPS Oper. Syst. Rev. 36(5), 96–107 (2002)

4. Chaintreau, A., Hui, P., Crowcroft, J., Diot, C., Gass, R., Scott, J.: Impact of human mobility on opportunistic forwarding algorithms. *IEEE Transactions on Mobile Computing* 6(6), 606–620 (2007)
5. Vahdat, A., Becker, D.: Epidemic routing for partially connected ad hoc networks. Tech. Rep. CS-2000-06, Duke University (2000)
6. Balasubramanian, A., Levine, N. B., Venkataramani, A.: DTN Routing as a Resource Allocation Problem. In *SIGCOMM 2007: Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications ACM SIGCOMM*, pages 373–384 (2007)
7. Jain, S., Fall, S., Patra, R.: Routing in a Delay Tolerant Network. In: *SIGCOMM 2004: Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications ACM SIGCOMM*, pages 145–158 (2004)
8. Delay Tolerant Network <http://sprout.ics.uci.edu/projects/dtn/>

Investigating the Pricing Impact on the QoS of a Server Farm Deployed on the Cloud

A.M.D Aljohani ^{1,2}, D.R.W Holton ² and Irfan Awan ²

¹ Tabuk university, Saudi Arabia

² university of Bradford, England

Abstract. One of the attractive characteristics of the cloud computing model is scalability. Web servers deployed in the cloud can scale computing resources upward and downward to match the demand. This achieved because the cloud model is provided in pay-as-you-go manner. However, some of the cloud providers have rules in the pricing manner which makes the decision to scale difficult such as Amazon Elastic Compute Cloud (Amazon EC2) when an instance terminated will be charged for each hour and each partial instance-hour consumed will be billed as a full hour. In this paper a discrete event simulation used to examine the impact of using this method of billing on the cost of the cloud and the QoS. Performance metrics of throughput, delay and utilization are obtained via simulation.

Keywords: Cloud Computing, QoS, Discrete event simulation

Investigating the Pricing Impact on the QoS of a Server Farm Deployed on the Cloud

Authors:

Abdulaziz Aljohani, D.R.W Holton, I Awan

HET-NETs 2013

Outline:

- ▶ Introduction
- ▶ Problem statement
- ▶ Model description
- ▶ Simulation
- ▶ Conclusion
- ▶ Future work

Introduction:

- ▶ Many businesses and organisation have been using the web server farms locally or hosted into web hosting provider.
- ▶ However, With emerge of cloud computing many have moved to the cloud to harvest the cloud benefit.

Introduction:

▶ What is Cloud Computing?

One of the best definitions is presented in (A break in the clouds: towards a cloud definition):

“Clouds are a **large pool of easily usable and accessible virtualized resources** (such as hardware, development platforms and/or services). These resources can be **dynamically reconfigured to adjust to a variable load (scale)**, allowing also for an optimum resource utilization. This pool of resources is typically exploited by a **pay-per-use model** in which guarantees are offered by the Infrastructure Provider by means of customized SLAs”

Introduction

▶ **Cloud benefits:**

- 1- Scaling up and down to meet demands.
- 2- Paying only for what you use.
- 3- eliminating the need for plan far ahead for provisioning

Problem statement:

- ▶ Web server which deployed on the cloud can scale up to a very large number of virtual machine (VM) to improve the QoS and meet the demand.
- ▶ Also, it can release the resource to reduce cost when it no longer required.

Problem statement:

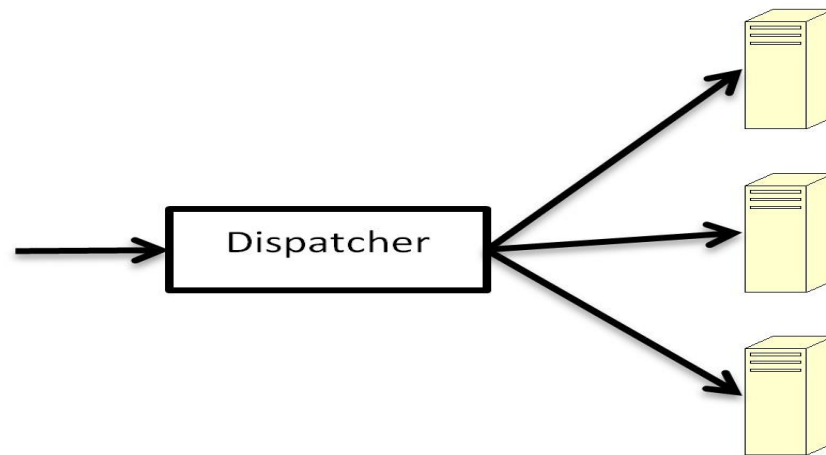
- ▶ The ideal cloud computing should charge the customer for exact time it deployed the server.
- ▶ However, that is not the case in reality. Cloud providers such as Amazon they pricing is per instance-hour consumed for each instance from the time it launched until it is terminated. Also, each partial instance-hour consumed will be billed as a full hour.

Problem statement:

- ▶ In the ideal cloud when an instance is not need it will be terminated to reduce the cost.
- ▶ However, if the user will be charged for the full hour that instance should remained on until the full hour is finished since the user will pay for it.

Model description:

- ▶ In this work a model has been presented to investigate the effect of the pricing method.
- ▶ We present a simple web server farm contains front-end load balancer and a number of back-end servers.

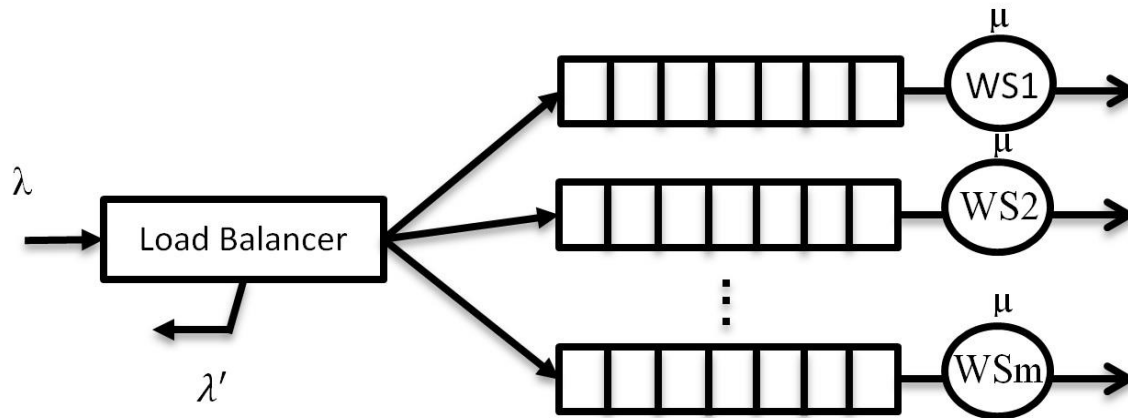


Problem statement:

- ▶ The server farm starts with one server and when the demand increase a server will be added to the pool. The number of request in the pool is used as trigger to add server using the upper threshold.
- ▶ When a server switched ON the time will be recorded in a table. At the end of each hour for each server it will check the number of request in the pool. If the number of request is less than the lower threshold the server will be terminated otherwise it will remain ON.

Simulation:

- ▶ A discrete-event simulation (DES), were built using Java to investigate the model.
- ▶ In this model we assumed the inter-arrival time and the service time is exponentially distributed.
- ▶ Also, each server modelled as a node with finite buffer space (K) and all servers are identical.



Conclusion:

- ▶ This work introducing the cloud pricing method and its impact on cost and QoS.
- ▶ A discrete-event simulation used to examine the model.

Future work:

- ▶ We intend to include extra parameters such as the time between requesting a server until it deployed and fully function.
- ▶ Also, we planning to simulate the same model using a network simulator such as OPNET.

Thank you



PART FIVE Congestion and Admission Control

Statistical Tools for Admission Control Decisions in Wireless Networks

Claudina Rattaro¹, Pablo Belzarena¹, Javier Simó², and Iván Hernández³

¹ Universidad de la República, Montevideo, Uruguay.

² Universidad Rey Juan Carlos, Madrid, España.

³ Universidad del Cauca, Popayán, Colombia.

Abstract. The family of IEEE 802.11 protocols has become the most popular wireless access method. In recent years, other technologies have started to complement 802.11 networks, for example, for Internet access in rural networks. Specifically, point-to-point links based on WiMAX and on TV White Space Dynamic Spectrum Access technologies are used to connect a wired Internet access with a set of 802.11 end users. These heterogeneous and multi-hop networks face many challenges in order to provide QoS guarantees. One of these challenges is the design of admission control algorithms. In this paper we develop a blackbox approach for designing admission control algorithms suitable for these kind of networks. The methodology is based on a combination of active measurements and statistical learning tools. The results obtained during simulation and testing in a laboratory testbed show that the methodology achieves good accuracy.

Keywords: Statistical learning, Support Vector Machines, Wireless Communications, Admission Control, Quality of Service, IEEE 802.11e, WiMax

1 Introduction

Wireless technologies have become increasingly popular as methods to provide flexible broadband connectivity. The main reasons are their low cost, and their fast and easy deployment. In particular, the family of IEEE 802.11 [1] protocols has become the most popular access method. In recent years, other technologies have started to complement 802.11 networks, for example, for Internet access in rural area networks (e.g. [2]). Specifically, point-to-point links based on WiMax (IEEE 802.16 [3]) and on TV White Space Dynamic Spectrum Access technologies. These technologies are typically used to connect with point-to-point links a wired Internet access with a rural community. At the end of these wireless hops, the users are connected using 802.11 links. This architecture is an heterogeneous and multi-hop network with two or more different technologies. This kind of networks must face many challenges to guarantee a certain level of Quality of Service (QoS). One of these challenges is the design of admission control algorithms. There are different proposals to implement admission control mechanisms in wireless networks [4,5], however, in heterogeneous networks the

algorithms cannot be implemented based on one specific technology and so, a blackbox approach is needed.

In this paper we extend the results of our previous work [6] in different directions. In our previous work we presented a methodology to estimate QoS parameters seen by applications in 802.11 networks. In this work we extend this methodology for developing admission control algorithms for QoS sensitive networks, including multi-hop and heterogeneous networks.

Other previous works focus on the problem of estimating QoS parameters by active measurements in the network (see for example [7] and the references therein), but our methodology is based on a combination of active measurements and the application of statistical learning tools. It consists in training the system (as a blackbox) during short periods with application flows and probe packets bursts. After the system has been trained, the QoS parameters of the blackbox can be estimated by sending the probe packets only. Although our proposal is valid for any broadband wireless network, some specific architectures of wireless rural networks were analyzed in the simulations and in the test-bed experiments.

We found that in a wireless multihop heterogeneous network it is possible to estimate the QoS parameters of a future connection by observing the statistical behavior of a light probe packet burst, and hence to decide on the convenience of accepting the connection. In simulations and testbed experiments our methodology proved to be accurate enough for practical use.

1.1 Organization of the paper

The rest of the paper is structured as follows. In section 2 we introduce the methodology. In section 3 we describe the network's state estimator. In section 4 we describe the statistical learning tool used in this work. In section 5 we show the performance of our proposal in different simulation scenarios and in section 6 we discuss experimental results based on a laboratory testbed. Finally, we conclude and discuss future works in section 7.

2 Methodology

Our methodology is based on the estimation of the network state using probe packets and on inferring the QoS seen by applications using statistical learning algorithms. We are interested on inferring the end-to-end QoS seen by an application from the behavior of the probe packets, so the probe packets should generate a traffic similar to the one the application generates (e.g. TCP traffic if it is a file download). Since the volume of this traffic may overload the network, we proposed a trade-off solution in which the system is trained during short periods with application flows and probe packets bursts. After this training period, the system sends only light probe packets that do not overload the network. We used an statistical learning approach to decide if a connection should be accepted or not from its QoS requirements and the behavior of the probe packets.

For example a new connection will be accepted if its delay is less than certain threshold, otherwise it will be blocked.

The main assumption is that the probe packets give an estimation of the state of the wireless path. If the state of the network is such that there are virtually no queues at the links, a low collision probability and low interference, the probe packets will not suffer major changes and their interarrival times will be almost the same as the periods between departures. However, if the network has for example many collisions, the interarrival times of probe packets will suffer strong modifications. This is illustrated in Fig. 1.

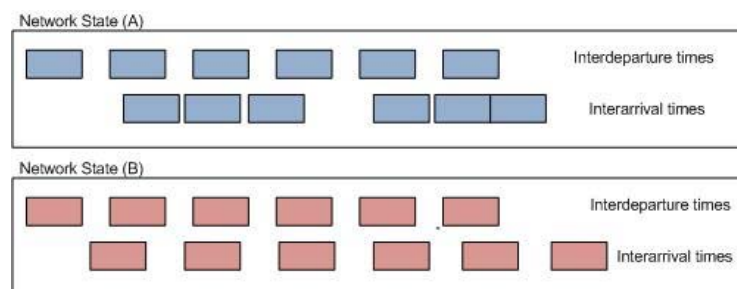


Fig. 1. Interarrival times of probe packets in two different network states. A: High collision probability and/or high interference. B: Low collision probability and low interference.

We proposed a binary classification problem, the idea is to predict the associated label $Z \in \{+1, -1\}$ from any new sample X , where Z is a binary variable that indicates if the connection is accepted ($Z = 1$) or not ($Z = -1$), and X is an estimation of the state of the wireless path. We use Support Vector Machines (SVM) as the statistical learning tool, therefore the objective is to find the maximum margin classification function (Φ). To estimate this function we divide the algorithm into two phases (learning and monitoring).

During the learning phase, when a new connection arrives to the network and starts sending traffic we measure the QoS parameters (e.g. delay) of the new connection. Afterwards, when this transmission finishes, the system sends a burst of probe packets with fixed size and interdeparture times. We build the variable X_i by measuring in each experiment i the interarrival times of the probe packets burst. Therefore, in each experiment of the learning phase we have a pair (X_i, Z_i) . Z_i is a binary sequence obtained by comparing the QoS parameters with their thresholds. After we have collected a set of samples (X_i, Z_i) we use SVM to obtain the estimation of the classification function ($\hat{\Phi}$). After the system is trained in the learning phase, the second one begins. During the monitoring phase, the new connection sends probe packets only and we build the variable X in the same way as in the learning phase. Using $\hat{\Phi}$ we can decide if the connection is accepted or not.

We point out that this procedure does not load the network during the monitoring phase because it does not send the application packets to measure the QoS parameters.

3 The estimator of the state of the wireless link, X

X will be estimated from the probability distribution of the variable component of the delay seen by the probe packets. We first consider a single wireless link, and then a network path.

We consider a probe packet n that arrives to the queue of the wireless link at time t_n^i and leaves the link at the time t_n^o (in practice, these can be obtained with timestamps). If the latency of the link is D , the free capacity is $C_n(p, I)$ (we consider the general case of adaptive multirate where the link capacity varies with p (collision probability) and I (channel interference)), P is the packet's size, tq_n is the waiting time in the queue and $V_n(p, I)$ represents the delay caused by retransmissions, the difference between t_n^o and t_n^i is such that:

$$t_n^o - t_n^i = \frac{P}{C_n(p, I)} + D + V_n(p, I) + tq_n \quad (1)$$

Some factors in (1) are constant and other depend on n , so we can express equation (1) as:

$$t_n^o - t_n^i = K + K_n(p, I) \quad (2)$$

Applying equation (2) recursively for n probe packets we have:

$$K_n = K_0 + \sum_{j=1}^n [(t_j^o - t_{j-1}^o) - (t_j^i - t_{j-1}^i)] \quad (3)$$

Equation (3) allows us to estimate the probability distribution of the variable component of the delay using only the arrival and departure times (K_0 is a constant independent of n).

The previous analysis can be generalized to the case of a multi-hop path. In that case, the difference between the arrival time of packet n to the first queue and the time that this packet leaves the multi-hop path has a constant part and other that depends on n . Therefore, this difference can be written as in equation (2), and equation (3) follows for the multi-hop case.

It is important to note that we didn't use the fact that the interdeparture times of the test packets are constant, so this result is valid for any distribution.

In order to give an insight on how K_n or its statistics depends on the state of the network, in Fig. 3 and 4 we show the empirical distribution of K_n for different states of the wireless link using 802.11. These simulations were done using the topology shows in Fig. 2 in saturated traffic condition. In the first, we can see that when the number of nodes in the network m increases the mean value of K_n increases too.

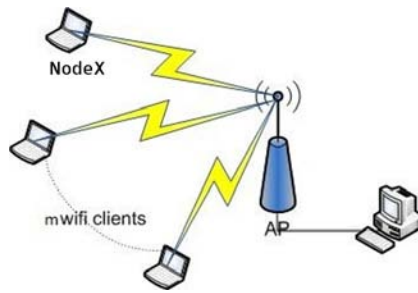


Fig. 2. IEEE 802.11 network. The network consists in a reference node (AP) and $m + 1$ nodes generating traffic to the AP. The distance between AP and NodeX (which generates the probe packets to estimate the state of the link) is the same in all the simulations.

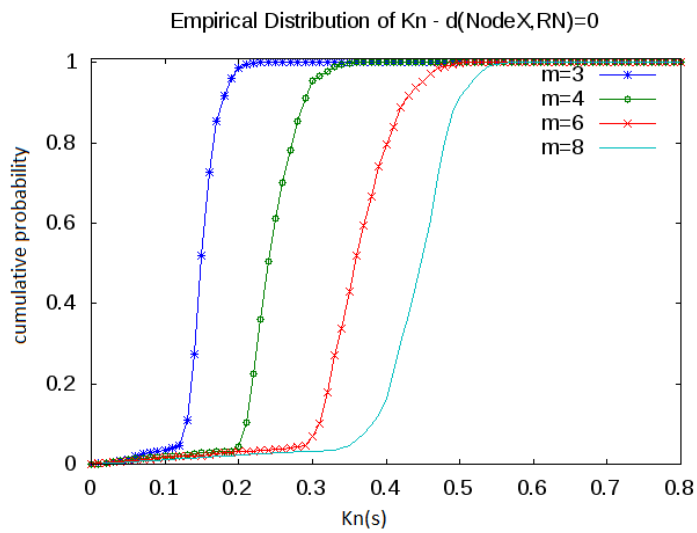


Fig. 3. Empirical distribution of K_n for different states of the wireless link, changing the number of fixed nodes (m) in a Wireless Local Area Network (see Fig. 2).

On the other hand, in Fig. 4 the number of fixed nodes is the same in all the simulations. In this case we vary the distance between the RN and NodeX. Fig. 4 shows, varying d varies the mean and variance of K_n .

Following the previous analysis we propose to use some statistics of K_n (expected value, variance, etc..) as an estimator of the state of the wireless network path (X).

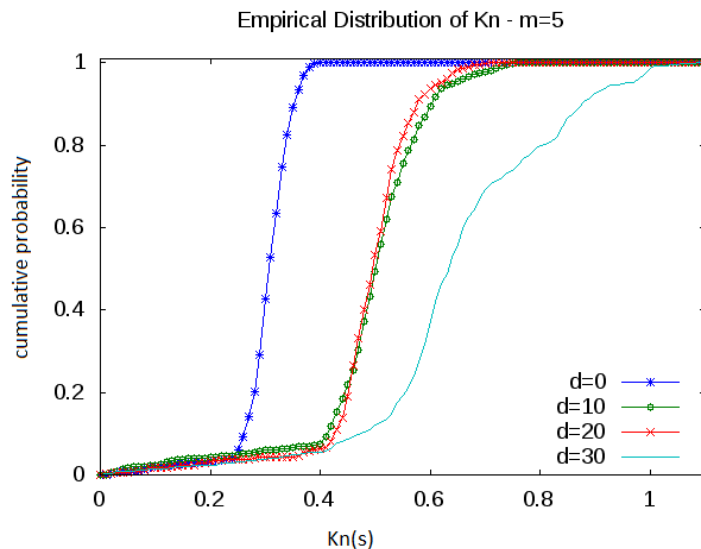


Fig. 4. Empirical distribution of K_n for different states of the wireless link, changing the distance (d) between NodeX and AP.

4 Support Vector Machines

There are several supervised statistical learning tools. We have selected Support Vector Machines (SVM). SVM is a set of classification and regression techniques, introduced in the early nineties by Vladimir Vapnik [8]. A complete description of the method can be found in [9]. SVM is a supervised learning tool well known for its discriminative power and has shown very good performance in different applications. In networking it has been used in several works showing very good performance (see for example [10]).

In this work, training and prediction were done with the LIBSVM library [11]. In particular, we use a radial basis function (rbf) kernel due to the good performance shown in different applications.

5 Simulations

We ran several sets of simulations using ns-2 simulator [12]. This simulator has been used in several works for simulating 802.11 networks (see for example [13,14,15]). In particular we used ns-2.28 simulator⁴.

We considered two topologies. Topology I (Fig. 5) is a basic scenario that consists of two wireless 802.11e nodes (the distance between the nodes is 30

⁴ We had to use that version of ns-2 to use the existing EDCA patch developed for TKN group [16] (it was used in other papers, an example is [17])

meters). Topology II (Fig. 6) represents a multi-hop network. In this case, all wireless links were configured using the same frequency band and the average distance between an AP with its clients is 10 meters. Please note traffic is bidirectional in both cases.

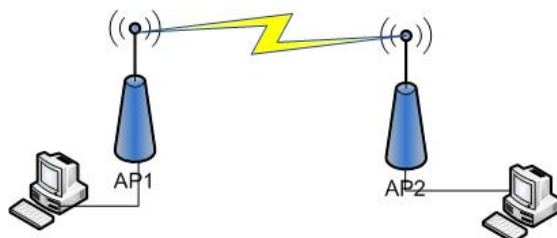


Fig. 5. Topology I: wireless link

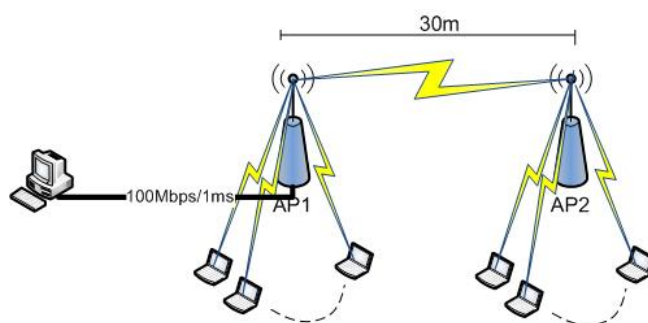


Fig. 6. Topology II: two wireless hop and a wired segment consisting of a duplex-link ($C = 100$ Mbps and $delay = 1$ ms)

We considered EDCA (802.11e), and to simplify, we only included two access categories (ACs) in our networks using their default EDCA parameters in ns-2 (AC_VI and AC_BE).

Network traffic was generated as follows: to model video flows we used an exponential on-off process over UDP, and best effort traffic was modeled as FTP/TCP connections. We considered 1500-byte packet size in both cases.

To estimate the state of the wireless path we injected UDP packets as probing packets. There is a trade-off between the size and interarrival time of these probe packets, since the objective is not to affect the network performance. We did several tests and we concluded that the values: Packet Size (payload)= 10 bytes and Interval Time= 10 ms are sufficient to achieve good results without affecting the system (these may be sent by either AC).

As an example we will show the results where the new connection is of AC_VI and we classified according to the delay. In order to get sufficient statistics for training and to verify the model, in each case we ran several simulations varying the number of flows in the network (the number of TCP and UDP connections).

In both scenarios we built the variable Z that we defined in sec.2 choosing an arbitrary threshold as $maxDelay$. $Z = 1$ if the new connection's delay is less than $maxDelay$, otherwise $Z = -1$. Using the mean value of K_n as the estimator X^5 , in both scenarios we obtained more than 90% of accuracy.

Note: accuracy means the proportion of verification samples that were classified correctly.

6 Testbed for Experimental Validation in Heterogeneous Networks

The proposed methodology was tested experimentally using several topologies typical in rural networks. As a representative example, we present the case of a heterogeneous network that consists of two segments, one with WiMAX and another with WiFi (see Fig. 7).

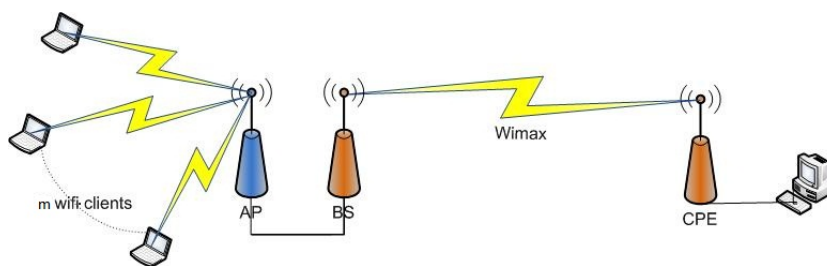


Fig. 7. Hybrid network WiMAX(5 GHz) and WiFi(2.4 GHz). The WiMAX segment consisted of a BS ARBA-556 and a CPE-56-PROA (Albentia Systems [18]). In the WiFi segment we used ALIX boards ([19]) with Voyage Linux v0.6.5 OS [20], one board worked as an Access Point and the other as clients (generating and receiving traffic). In addition, we have used miniPCI 802.11 a/b/g radio interfaces model Mikrotik R52H with 802.11e EDCA support, Atheros chipset and MadWiFi driver.

We have used D-ITGv2.8.0-rc1 [21] to generate and measure traffic. As in previous simulations, we used only two access categories of EDCA (in this case we chose AC_VI and AC_BE) with its default driver parameters. All flows of the class AC_BE were TCP, the other ones, trying to emulate real-time applications, were UDP. We did several tests varying the number of flows of each class in the network (considering bidirectional traffic). The tests were conducted reproducing the steps performed in the simulations.

⁵ We have tested the system using other statistics like the variance or quantiles of K_n but the improve on the accuracy of the system is negligible

In order to build the estimator of the state of the network, a burst of probe packets with fixed size and interdeparture times was injected: packet size (payload)= 10 bytes, interdeparture time= 10 ms, duration= 60s. This causes an overload on the network less than 0.1% which allow us to affirm that the methodology is not invasive.

We used a binary classification based on thresholds. For example for a AC_BE new connection, it will be accepted if its throughput is greater than $minTh$, otherwise it will be blocked. Applying this criteria, we obtained 85% of accuracy using the mean value of $K(n)$ as the estimator of the state of the network.

7 Conclusion

The main contributions of this paper are the proposed estimator of the wireless network state and the methodology presented that uses SVM and probe packets in order to decide whether a connection should be admitted by the network or not. In addition, the proposed methodology is only marginally intrusive. This was verified in the real experiments. The results obtained in the validation process (simulations and experimental testbed) show that the methodology has a good accuracy. It is important to note that the technique is independent of the wireless technology of the network. This allows its use in heterogeneous networks.

This work has many future lines of research like: extending it for more complex topologies and applying the methodology to real networks, in particular, in rural areas (ej: Long-Distance Wireless Links). Also, we are interested in investigating other features of K_n that help raise the accuracy of the estimation and the classification.

Acknowledgments This work was supported by the project CSIC-UdelaR: Algoritmos de Control de Acceso al Medio en Redes Inalámbricas. This research was partially supported by Centro Ceibal para el Apoyo a la Educación de la Niñez y la Adolescencia [22] and MASERATI project funded by the Iberoamerican Programme of Science and Technology for Development CYTED. Also we would like to thank Federico Larroca for helpful discussions.

References

1. IEEE Standard 802.11, 2007 Edition. Available at <http://standards.ieee.org/getieee802/download/802.11-2007.pdf>
2. Simo Reigadas, J., Martinez-Fernandez, A., Ramos-Lopez, J., Seoane-Pascual, J.: Modeling and Optimizing IEEE 802.11 DCF for Long-Distance Links, Mobile Computing, IEEE Transactions on , vol.9, no.6, pp.881,896 (2010)
3. IEEE 802.16. 802.16-2009 IEEE standard for local and metropolitan area networks part 16: Air interface for fixed and mobile wireless access systems., 2009.
4. Yanbing Liu, Man Meng,: Survey of Admission Control Algorithms in IEEE 802.11e Wireless LANs, Future Computer and Communication, 2009. FCC '09. International Conference on , vol., no., pp.230,233, 6-7 (2009)

5. Deyun Gao, Jianfei Cai, King Ngi Ngan.: Admission control in IEEE 802.11e wireless LANs, *Network, IEEE*, vol.19, no.4, pp.6,13, (2005)
6. Rattaro, C., Belzarena, P.: Throughput prediction in wireless networks using statistical learning. *Latin-American Workshop on Dynamic Networks*. Buenos Aires (2010)
7. Mingzhe L., Claypool, M., Kinicki, R.: WBest: A bandwidth estimation tool for IEEE 802.11 wireless networks, *Local Computer Networks*, 2008. LCN 2008. 33rd IEEE Conference on , vol., no., pp.374,381, 14-17 (2008)
8. Vapnik, V.: *The nature of statistical learning theory*, Springer NY (1995).
9. Scholkopf, B., Smola, A.: *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.*, Cambridge, MA, USA: MIT Press (2001).
10. Xiang Yang, Y., Wang, R., Liu, Y., Zhen Li, S., and Yong Zhou, X.: Solving p2p traffic identification problems via optimized support vector machines in AICCSA, 2007, pp. 165-171 (2007).
11. C-C. Chang, C-J. Lin, LIBSVM: a library for support vector machines, (2001), <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
12. McCanne S., Floyd S., ns network simulator, <http://www.isi.edu/nsnam/ns>.
13. Oubaha, J., Habbani, A., ElKoutbi, M.: New approach: Mapping of 802.11e into MPLS domains, *Multimedia Computing and Systems (ICMCS)*, 2011 International Conference on , vol., no., pp.1,6, 7-9 (2011)
14. Xinyu Zhang, Shin, K.G.: Delay-Optimal Broadcast for Multihop Wireless Networks Using Self-Interference Cancellation, *Mobile Computing, IEEE Transactions on*, vol.12, no.1, pp.7,20, (2013)
15. Jain, P., Biswas, G.P.: Design and implementation of an enhanced rate adaptation scheme for wireless LAN IEEE-802.11, *Recent Advances in Information Technology (RAIT)*, 2012 1st International Conference on , vol., no., pp.336,340, 15-17 (2012)
16. Wietholter, S., Hoene, C.: Design and Verification of an IEEE 802.11e EDCA Simulation Model in ns-2.28, TKN Telecommunication Network Group Technische Universität Berlin, (2003)
17. Pompeo Silva Mineiro, E., Muchaluat-Saade, D.C.: Admission Control for QoS Provision in Wireless Mesh Networks, *Latin America Transactions, IEEE*, vol.10, no.1, pp.1135,1142, (2012)
18. Alentia Systems S.A. ARBA550 BS Series User Guide, <http://www.alentia.com>
19. ALIX2C0, <http://www.pcengines.ch/alix2c0.htm>.
20. VOYAGE, <http://linux.voyage.hk/content/voyage-linux-065-released>
21. D-ITG, <http://www.grid.unina.it/software/ITG/>
22. Plan Ceibal, <http://www.ceibal.org.uy>

Congestion probabilities in W-CDMA networks supporting calls of finites sources

I. D. Moscholios¹, G. A. Kallos², V. G. Vassilakis³, M. D. Logothetis⁴ and M. N. Koukias⁴

1. Dept. of Informatics & Telecommunications, University of Peloponnese, 221 00 Tripolis, Greece

2. BT Technology, Service and Operations, IP5 3RE Ipswich, U.K.

3. Dept. of Electronic Engineering, University of Surrey, GU2 7XH Guildford, U.K.

4. WCL, Dept. of Electrical & Computer Engineering, University of Patras, 265 04 Patras, Greece.

E-mails: ids@uop.gr, george.kallos@bt.com, v.vasilakis@surrey.ac.uk, mlogo@upatras.gr, koukias@wcl.ee.upatras.gr

Abstract—We study a multirate teletraffic loss model for the calculation of time and call congestion probabilities in W-CDMA networks that accommodate calls of different service-classes whose arrival follows a quasi-random process. The latter is smoother than a Poisson process in the sense that calls come from a finite source population. The proposed model takes into account multiple access interference, the notion of local (soft) blocking, user's activity and interference cancellation. Although the model does not have a product form solution, we show that the calculation of time and call congestion probabilities is based on approximate but recursive formulas whose accuracy is verified through simulation and found to be quite satisfactory.

Keywords – W-CDMA; time-call congestion; quasi-random; recursive formula; call admission.

1. Introduction

Wideband Code Division Multiple Access (W-CDMA) networks support a wide range of voice and data applications with different QoS requirements. Due to the heterogeneous nature of traffic and the existence of own-cell and other-cell interference, the call-level performance analysis of W-CDMA networks is complicated in both uplink and downlink directions.

Herein, we consider a W-CDMA reference cell that supports calls from K different service-classes and examine the uplink direction (from a mobile user to the base station that controls the cell). The reference cell is modeled as a multirate loss system whose capacity consists of a fixed and integer number of channels. Similarly, the radio resource requirements of each call correspond to an integer number of channels. A new service-class k ($k=1, \dots, K$) call is accepted in the cell if the requested number of channels is available. More precisely, the admission of a new call is based on the estimation of the total interference increase (own-cell and other-cell interference plus thermal noise) caused by the new call's acceptance. An accepted call remains in the system for an exponentially distributed service time. Due to the existence of interferences, soft or local blocking (LB) occurs. That is, a new call can be blocked and lost in any system state if its acceptance increases the noise of all in-service calls above a tolerable level, given that, according to the W-CDMA principle, a call is noise for all other calls. This call admission policy corresponds to the complete sharing (CS) policy in wire networks, whereby all calls compete for all bandwidth resources [1]. The previous model of a W-CDMA cell has been adopted in [2]-[5]. In these papers, the formulas proposed for the calculation of call blocking probabilities (CBP) in the cell, resemble the classical Kaufman-Roberts formula used for the CBP calculation in the Erlang Multirate Loss Model (EMLM). The EMLM refers to a single link of certain capacity that accommodates, under the CS policy, Poisson arriving calls of different service-classes with different bandwidth requirements [6], [7]. More precisely, in [2] and [3] an extension of the EMLM is considered which is based on the Delbrouck's model [8]. The latter generalizes the EMLM since it allows the call-arrival process to have different peakedness factors. In [4], calls arrive in the cell according to a Poisson process. In [5], calls come from a finite number of sources, a realistic case especially in cells that have a limited coverage area. Apart from the different call arrival processes considered in [2]-[5], there is also another difference in the way the notion of LB is modelled. In [2], [3], this modelling is more complicated compared to the one proposed in [4] (also adopted in [5]). In subsection 3.2, we compare these approaches and adopt the model of [4] as it is more realistic for W-CDMA systems.

In this paper, we consider the abovementioned model of a W-CDMA reference cell under the assumption that calls come from a finite number of sources. This call arrival process is known as a quasi-random process

and is smoother than the Poisson process [9]. To this end, we review and make some necessary corrections in [5] while we provide formulas for the calculation of time and call congestion probabilities (TC and CC probabilities, respectively). TC probabilities are determined by the proportion of time the system is congested, while CC probabilities are determined by the proportion of arriving calls that find the system congested. These probabilities coincide in the case of Poisson arrivals due to the Poisson Arrivals See Time Averages (PASTA) property [1]. Based on [5], we also study the effect of interference cancellation (IC) on TC and CC probabilities. IC receivers reduce only the own-cell interference and not the other-cell interference or the thermal noise [10]. This reduction results in the decrease of TC and CC probabilities.

This paper is organized as follows: In section 2, we present the basic relations in the uplink direction of a W-CDMA cell. In subsection 2.1, we review the relations for the total received power of a service-class k call with or without IC. In subsection 2.2, we propose a formula for the determination of the maximum number of service-class k calls in the cell under the IC existence. In subsection 2.3, we calculate the uplink capacity and the bandwidth requirement of service-class k calls. In section 3, we consider Poisson arrivals and calculate CBP in the case of hard blocking (subsection 3.1) and when hard and soft blocking co-exist (subsection 3.2). In section 4, we consider the case of quasi-random arrivals and calculate TC and CC probabilities in the case of hard blocking (subsection 4.1) and when hard and soft blocking co-exist (subsection 4.2). In subsections 3.2 and 4.2 we consider the case of IC. In section 5, we present numerical results and evaluate the proposed formulas based on simulation results. We conclude in section 6.

2. Basic relations in the uplink direction of a W-CDMA reference cell

2.1. Determination of the total received power of a service-class k call

Consider a W-CDMA reference cell controlled by a base station and surrounded by other cells. We examine the uplink direction of the reference cell and model the latter as a multirate loss model. The cell accommodates K different service-classes. A service-class k ($k=1, \dots, K$) call alternates between periods of transmission (active periods) and periods of non-transmission (passive periods). The ratio of “active” over “active + passive” periods is the activity factor of a service-class k call, v_k , where $v_k \leq 1$. Typical values of v_k are: $v_k = 1.0$ if k is a data service-class and $v_k = 0.67$ if k is a voice service-class (see pp. 187, in [11]).

In W-CDMA systems, all users transmit within the same frequency band which means that a single user “sees” the signals generated by all other users as interference. In that case, the base station’s capacity in the W-CDMA reference cell is limited by the so called Multiple Access Interference (MAI) [11]. The latter consists of the own-cell interference, P_{own} , caused by the mobile users of the reference cell and the other-cell interference, P_{other} , which refers to the interference power received from mobile users of the neighbouring cells. Since MAI has a stochastic nature, we speak about the soft or interference limited capacity of the radio interface (see pp. 225 in [12]). We also consider thermal noise, P_{noise} , which corresponds to the interference of an empty W-CDMA system. A typical value of the thermal noise power density is 174 dBm/Hz (see [11]).

The values of P_{own} can be reduced by the application of IC; the latter is not effective towards P_{other} and P_{noise} . The IC efficiency is denoted by β and defined by the ratio [10]:

$$\beta = \left(P_{own}^{NO IC} - P_{own} \right) / P_{own}^{NO IC} \Rightarrow P_{own} = P_{own}^{NO IC} (1 - \beta) \quad (1)$$

where $P_{own}^{NO IC}$ is the own-cell interference without IC.

Herein, β is constant ($0 \leq \beta < 1$) and common to all service-classes while it is independent of the specific receiver implementation and radio link conditions ([10], [13]). Due to IC, and by denoting as p_k the total received power from a service-class k user, we write the power control equation for service-class k as [10]:

$$\left(E_b / N_0 \right)_k = G_k p_k / \left[\left(P_{own} - p_k \right) (1 - \beta) + P_{other} + P_{noise} \right] \quad (2)$$

where $\left(E_b / N_0 \right)_k$ is the signal energy per bit divided by the noise spectral density, required to meet a predefined Block Error Rate, $G_k = W / v_k R_k$ is the processing gain of service-class k in the uplink direction with user activity factor v_k , data rate R_k and W the chip rate of 3840 kcps.

Based on eq. (2), the values of p_k can be obtained by:

$$p_k = \left(E_b / N_0 \right)_k \frac{\left(P_{own} (1 - \beta) + P_{other} + P_{noise} \right)}{G_k + \left(E_b / N_0 \right)_k (1 - \beta)} \Rightarrow p_k = \frac{\left(P_{own} (1 - \beta) + P_{other} + P_{noise} \right)}{1 - \beta + G_k / \left(E_b / N_0 \right)_k} \quad (3)$$

2.2. Determination of the maximum number of service-class k calls in the cell

Let N_k be the maximum number of service-class k calls in the cell. Assuming that $P_{\text{own}}=p_k N_k$, we can calculate P_{own} , via eq. (3), as a function of N_k :

$$P_{\text{own}} = N_k (P_{\text{other}} + P_{\text{noise}}) / \left[(1-\beta) - N_k(1-\beta) + G_k / (E_b/N_0)_k \right] \quad (4)$$

where $P_{\text{total}} = P_{\text{own}} + P_{\text{other}} + P_{\text{noise}}$ is the total received power at the base station.

Consider now the Noise Rise (NR) which is defined as the ratio of the total received power at the base station to the thermal noise (pp. 227 in [12]):

$$NR = P_{\text{total}} / P_{\text{noise}} = (P_{\text{own}} + P_{\text{other}} + P_{\text{noise}}) / P_{\text{noise}} \quad (5)$$

The NR is related to the total uplink cell load, η_{UL} , according to the formula (see also eq. (8.9) in [11]):

$$NR = 1 / (1 - \eta_{UL}) \quad (6)$$

where: $\eta_{UL} = (P_{\text{own}} + P_{\text{other}}) / P_{\text{total}}$.

Based on eqs. (5) and (6) we have:

$$(P_{\text{own}} + P_{\text{other}} + P_{\text{noise}}) / P_{\text{noise}} = 1 / (1 - \eta_{UL}) \quad (7)$$

Substituting eq. (4) in eq. (7) and solving for N_k we have:

$$N_k = \left[(1-\beta) + G_k / (E_b/N_0)_k \right] \frac{[\eta_{UL}(\delta+1) - \delta]}{[1-\beta(\eta_{UL}(\delta+1) - \delta)]}, \quad \delta = P_{\text{other}} / P_{\text{noise}} \quad (8)$$

If $\beta=0$ (IC is not applied) and $\delta=0$ (which means that $P_{\text{other}} = 0$), then eq. (8) takes the form:

$$N_k = \left[1 + G_k / (E_b/N_0)_k \right] \eta_{UL} \quad (9)$$

which is eq. (3) of [3] and shows the maximum number of service-class k calls in an isolated cell.

2.3. Determination of the uplink capacity and the bandwidth requirement of service-class k calls

Having determined N_k according to the proposed eq. (8), we calculate the spread data rate $R_{s,k}$ of service-class k , as the proportion of W which is utilised by a call of service-class k :

$$R_{s,k} = W / N_k \quad (10)$$

Now, we transform W to the uplink capacity C , and $R_{s,k}$ to the corresponding bandwidth requirement per call, b_k , of each service-class k . To achieve this, we define a basic bandwidth unit (*bbu*) which can be determined as the greatest common divisor of the required call resources of all service-classes, or it can take an arbitrarily chosen small value. E.g., if *bbu* = 20 Kcps (arbitrarily chosen), then C and b_k are given by:

$$C = \left\lceil \frac{W}{\text{bbu}} \right\rceil = 192 \text{ channels}, \quad b_k = \left\lceil \frac{R_{s,k}}{\text{bbu}} \right\rceil \text{ channels} \quad (11)$$

Consider now an example of $K=4$ service-classes accommodated in a W-CDMA cell. In Table I, we present the results of N_k and $R_{s,k}$, for $\eta_{UL}=0.5$ and $i=0.55$, when the calculation of N_k is based either on eq. (8) or eq. (9). Equation (9), which does not take into account the other-cell interference, overestimates the maximum number of service-class k calls in the cell, compared to the proposed eq. (8). On the other hand, the results of eq. (8) show that the increase of β (from 0.5 to 0.8) increases N_k as expected.

Table I: N_k and $R_{s,k}$ when $\eta_{UL}=0.5$, $i=0.55$, $\delta=0.5$ (when $\beta > 0$) and various values of the IC efficiency β .

Service-class k	R_k (kbps)	v_k	$(E_b/N_0)_k$ (in dB)	$(E_b/N_0)_k$	$\beta=0, \delta=0$ [2]-[4]		$\beta=0.2$ (proposed)		$\beta=0.5$ (proposed)		$\beta=0.8$ (proposed)	
					N_k (eq. 9)	$R_{s,k}$ (eq. 10)	N_k (eq. 8)	$R_{s,k}$ (eq. 10)	N_k (eq. 8)	$R_{s,k}$ (eq. 10)	N_k (eq. 8)	$R_{s,k}$ (eq. 10)
1	7.95	0.67	4.0	2.51	144	26.67	75.7	50.70	82.1	46.75	89.7	42.78
2	7.95	0.67	7.0	5.01	72.4	53.02	38.1	100.88	41.2	93.11	45.0	85.31
3	32	1.0	3.0	2.00	30.6	125.61	16.0	239.44	17.3	221.63	18.9	203.64
4	64	1.0	2.0	1.58	19.4	197.65	10.2	377.47	10.9	350.39	11.9	322.88

3. Call blocking probabilities assuming Poisson arrivals

3.1. A system with hard blocking only

In connection-oriented systems, every system state j ($j=0, 1, \dots, C$) can be a non-blocking or a blocking state for service-class k calls, depending on the bandwidth requirement b_k . Assuming Poisson arriving calls, exponentially distributed service times and the CS policy, the EMLM results. In the EMLM, the un-normalized values of the system state probabilities, $q(j)$, are given by the following accurate and recursive formula, known as Kaufman-Roberts formula (or recursion) [6], [7]:

$$q(j) = \frac{1}{j} \sum_{k=1}^K a_k b_k q(j-b_k), \text{ for } j=1, \dots, C \quad (12)$$

where: $q(0)=1$, $q(x) = 0$ if $x < 0$, $\alpha_k = \lambda_k/\mu_k$ is the offered traffic-load of service-class k calls (in erl) while λ_k and μ_k are the mean arrival and service rate of service-class k calls, respectively.

The calculation of TC probabilities of service-class k , P_{b_k} , is based on the following formula:

$$P_{b_k} = \sum_{j=C-b_k+1}^C G^{-1} q(j) \quad (13)$$

where $G = \sum_{j=0}^C q(j)$ is the normalization constant.

3.2. A system with both hard and soft blocking

In W-CDMA networks we distinguish two types of states j : a) those that are blocking states for service-class k calls (hard blocking states) and b) those that are blocking states with a probability $0 < L_{j,k} < 1$ (soft or local blocking states) due to the existence of other-cell interference. In what follows, we show how we can incorporate the notion of LB in eq. (12). To consider in the cell the other-cell interference, we approximate it by an independent, lognormally distributed random variable with parameters μ and σ :

$$\mu = \frac{P_{other} + P_{noise}}{P_{own} + P_{other} + P_{noise}} C \xrightarrow{i = \frac{P_{other}}{P_{own}}, \delta = \frac{P_{other}}{P_{noise}}} \mu = \frac{i + i/\delta}{1 + i + i/\delta} C, \quad \sigma = \mu \quad (14)$$

Note that the parameter μ is chosen to be equal to σ as proposed in the literature (e.g., [2]-[5], [14]-[15]). The value of μ expresses the average capacity that is lost from the reference cell due to the other-cell interference.

If the thermal noise P_{noise} is not considered, then eq. (14) takes the form (see also eq. (27) in [3]):

$$\mu = \frac{P_{other}}{P_{own} + P_{other}} C \xrightarrow{i = \frac{P_{other}}{P_{own}}} \mu = \frac{i}{1 + i} C, \quad \sigma = \mu \quad (15)$$

The LB probability (LBP) in state j , L_j , is the probability that the other-cell interference is greater than the available cell's capacity ($C-j$) and is independent of b_k [2]:

$$L_j = P(j' > C - j) = 1 - P(j' < C - j) \Rightarrow L_j = 1 - CDF(C - j) \quad (16)$$

where j' denotes the occupied channels due to the other cell interference and $CDF(x)$ is the cumulative distribution function of the lognormal distribution.

The values of $CDF(x)$ can be determined by:

$$CDF(x) = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{\ln(x) - M}{S\sqrt{2}} \right) \right) \quad (17)$$

where erf is the error function, while M and S refer to the parameters of the normal distribution:

$$M = \ln \left(\mu^2 / \sqrt{\mu^2 + \sigma^2} \right), \quad S = \sqrt{\ln \left(1 + (\sigma^2 / \mu^2) \right)} \quad (18)$$

Consider now a new service-class k call which requires b_k channels in order to be accepted in the cell. We can express the passage factor $1 - L_{j,b_k}$, i.e., the probability that the call will not be blocked due to the other-cell interference, as a function of the number of channels occupied in the cell and b_k [2], [3]:

$$1 - L_{j,b_k} = \prod_{x=j}^{j+b_k-1} (1 - L_x) = (1 - L_j)(1 - L_{j+1})(1 - L_{j+2}) \dots (1 - L_{j+b_k-1}) \quad (19)$$

Note that the right hand side of eq. (19) consists of b_k terms of the form $(1-L_x)$. This means that every time a service-class k call obtains one channel, the value of the LBP changes (e.g., if in state x , the value of $1-L_x$ becomes $1-L_{x+1}$ etc.) and the call is finally accepted in the cell if all b_k channels are assigned to the call.

In [4], a simpler approach is proposed. A new service-class k call will be accepted in the cell if all b_k channels are assigned to the call simultaneously. This means, that the other-cell interference (and consequently the LBP) remains the same during this channel allocation process. In that case, the passage factor $1-L_{j,b_k}$ is equal to the last term of eq. (19), i.e.,:

$$1-L_{j,b_k} = (1-L_{j+b_k-1}) \quad (20)$$

Herein we adopt the approach of [4] as it is more realistic in W-CDMA systems.

Due to the introduction of passage factor according to eq. (20), the transition rate from state $(j-b_k)$ to j , equals $(1-L_{j-b_k,b_k})\lambda_k = (1-L_{j-1})\lambda_k$. Figure 1 presents the system's state transition diagram which is depicted by a one-dimensional Markov chain. Note that $y_k(j)$ is the average number of service-class k calls in state j .

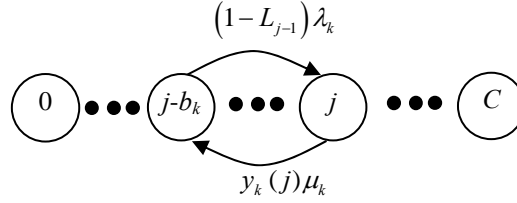


Figure 1: State transition diagram for Poisson service-class k calls with LB between states $j-b_k$ and j .

The un-normalized values of state probabilities, $q(j)$, are given by an approximate formula [4]:

$$q(j) = \frac{1}{j} \sum_{k=1}^K a_k b_k q(j-b_k) (1-L_{j-b_k,b_k}), \text{ for } j=1, \dots, C \quad (21)$$

where: $q(0)=1$, $q(x) = 0$ if $x < 0$, $\alpha_k = \lambda_k/\mu_k$ is the offered traffic-load of service-class k (in erl).

The determination of TC probabilities of service-class k , P_{b_k} , is based on the formula:

$$P_{b_k} = \sum_{j=0}^C G^{-1} L_{j,j+b_k} q(j) \quad (22)$$

where $G = \sum_{j=0}^C q(j)$ is the normalization constant.

To summarize the differences between [2] (or [3]), [4] and the proposed method, we present in Table II the algorithmic steps for the calculation of TC probabilities in each case.

Table II: Algorithms for the calculation of TC probabilities in the case of Poisson traffic.

Algorithm for the calculation of TC probabilities based on [2], [3]	Algorithm for the calculation of TC probabilities based on [4]	Proposed algorithm for the calculation of TC probabilities
1) Determine N_k (eq. (9))	1) Determine N_k (eq. (9))	1) Determine N_k (eq. (8))
2) Determine $R_{s,k}$ (eq. (10))	2) Determine $R_{s,k}$ (eq. (10))	2) Determine $R_{s,k}$ (eq. (10))
3) Determine C, b_k (eq. (11))	3) Determine C, b_k (eq. (11))	3) Determine C, b_k (eq. (11))
4) Determine μ, σ (eq. (15))	4) Determine μ, σ (eq. (15))	4) Determine μ, σ (eq. (14))
5) Determine L_j (eqs. (16-18))	5) Determine L_j (eqs. (16-18))	5) Determine L_j (eqs. (16-18))
6) Determine $1-L_{j,b_k}$ (eq. (19))	6) Determine $1-L_{j,b_k}$ (eq. (20))	6) Determine $1-L_{j,b_k}$ (eq. (20))
7) Determine $q(j)$ (eq. (21))	7) Determine $q(j)$ (eq. (21))	7) Determine $q(j)$ (eq. (21))
8) Determine P_{b_k} (eq. (22))	8) Determine P_{b_k} (eq. (22))	8) Determine P_{b_k} (eq. (22))

Consider again the example of the four service-classes presented in section 2.3. Calls follow a Poisson process and have an exponentially distributed service time in the cell. The offered traffic-loads are: $\alpha_1= 8.0$ erl, $\alpha_2= 4.0$ erl, $\alpha_3= 0.5$ erl and $\alpha_4= 1.0$ erl. If $bbu = 13.5$ Kcps we obtain the values of C and b_k ($k=1, \dots, 4$):

a) Based on the 7th column of Table I ($\beta=\delta=0$): $C=284$, $b_1=2$, $b_2=4$, $b_3=10$ and $b_4=15$ channels. These values will be used for the calculation of TC probabilities in the case of [2] (or [3]) and [4].

b) Based on the 9th column of Table I ($\beta=0.2, \delta=0.5$): $C=284, b_1=4, b_2=8, b_3=18$ and $b_4=28$ channels.

c) Based on the 11th column of Table I ($\beta=0.5, \delta=0.5$): $C=284, b_1=4, b_2=7, b_3=17$ and $b_4=26$ channels.

d) Based on the 13th column of Table I ($\beta=0.8, \delta=0.5$): $C=284, b_1=4, b_2=7, b_3=16$ and $b_4=24$ channels.

The values of (b), (c) and (d) are used to calculate TC probabilities in the case of the proposed algorithm.

In the x-axis of Figs. 2-7, $\alpha_1, \alpha_2, \alpha_3$ and α_4 increase in steps of 2.0, 1.0, 0.5 and 0.25 erl, respectively. So, in Point 1 we have: $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (8.0, 4.0, 0.5, 1.0)$, while in Point 13 $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (32.0, 16.0, 6.5, 4.0)$. In Figs. 2-3, we present the TC probabilities of all service-classes following [2] (or [3]) and [4], respectively. The results show that the approach of [2] gives much higher TC probabilities compared to the more realistic approach of [4]. In Figs. 4-7, we adopt the approach of [4] for the calculation of the passage factors, consider three values of the IC efficiency $\beta=0.2, 0.5, 0.8$ and show the TC probabilities for each service-class, respectively. The TC probabilities obtained by the proposed algorithm cannot be approximated by [2], [3] or [4], since a different approach is followed for the calculation of N_k 's, μ and σ . In addition, the increase of β results in the decrease of TC probabilities as expected.

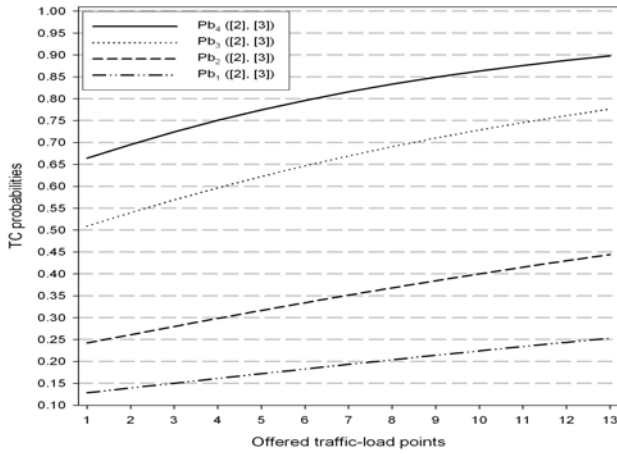


Figure 2: TC probabilities of all service-classes according to [2], [3] (IC is not included).

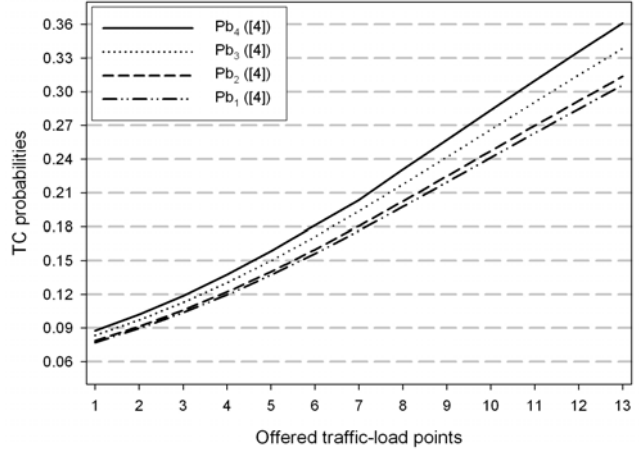


Figure 3: TC probabilities of all service-classes according to [4] (IC is not included).

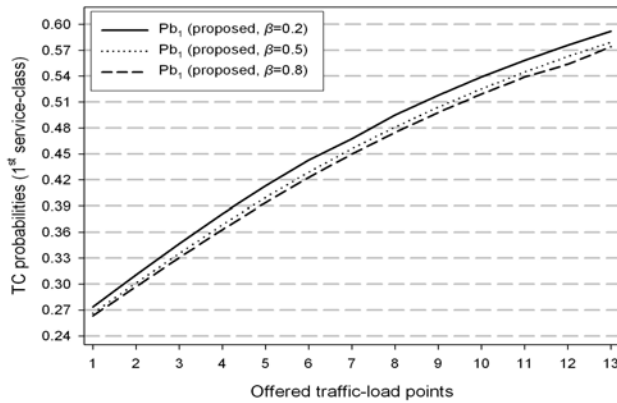


Figure 4: TC probabilities of the 1st service-class for three different values of the IC efficiency.

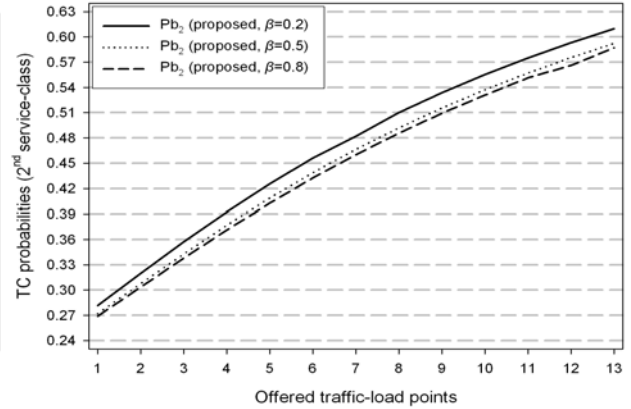


Figure 5: TC probabilities of the 2nd service-class for three different values of the IC efficiency.

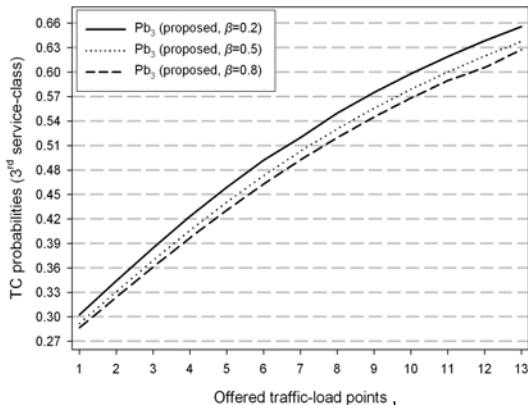


Figure 6: TC probabilities of the 3rd service-class for three different values of the IC efficiency.

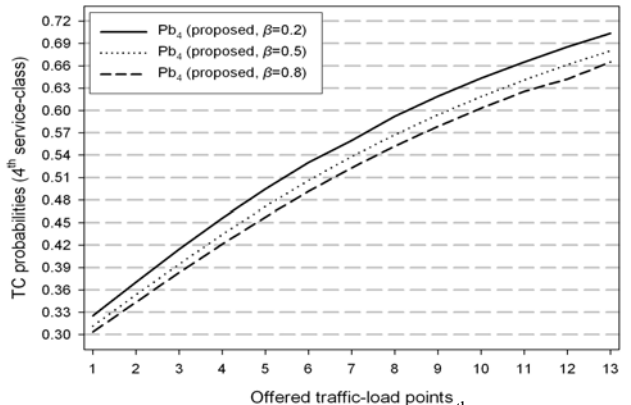


Figure 7: TC probabilities of the 4th service-class for three different values of the IC efficiency.

4. Congestion probabilities assuming quasi-random arrivals

4.1. A system with hard blocking only

In the case of quasi-random traffic, calls of service-class k come from a finite number of traffic sources, S_k . In that case, the calculation of $q(j)$'s can be based on the following accurate and recursive formula [16]:

$$q(j) = \frac{1}{j} \sum_{k=1}^K (S_k - n_k(j) + 1) a_{k,fin} b_k q(j - b_k), \text{ for } j = 1, \dots, C \quad (23)$$

where: $q(0)=1$, $q(x) = 0$ if $x < 0$, $\alpha_{k,fin} = \gamma_k / \mu_k$ is the offered traffic-load per idle source of service-class k and γ_k is the arrival rate from an idle source of service-class k .

We name the model of [16], Engset Multirate Loss Model (EnMLM) because when $K=1$, eq. (23) and eq. (13) give the same TC probabilities with the classical Engset formula [9]. The determination of $q(j)$'s in eq. (23) requires the value of the in-service calls of service-class k in state j , $n_k(j)$, which is unknown. In other finite multirate loss models (e.g., [16]-[20]) there exist calculation methods for the determination of n_k in each state j through the use of an equivalent stochastic system, with the same traffic description parameters and exactly the same set of states. However, the state space determination of the equivalent system is complex, especially for large capacity systems that serve many service-classes. Thus, we avoid such methods and approximate $n_k(j)$, as the mean number of service-class k calls in state j , $y_k(j)$, when Poisson arrivals are considered, i.e., $n_k(j) \approx y_k(j)$ and $n_k(j) - 1 \approx y_k(j - b_k)$. Such approximations are common in the literature and induce little error (e.g., [21]-[26]). Based on this approximation, eq. (23) takes the form:

$$q(j) = \frac{1}{j} \sum_{k=1}^K (S_k - y_k(j - b_k)) a_{k,fin} b_k q(j - b_k), \text{ for } j = 1, \dots, C \quad (24)$$

where: $q(0)=1$, $q(x) = 0$ if $x < 0$ and the values of $y_k(j)$ are given by eq. (25) in the case of Poisson traffic:

$$y_k(j) = a_k q(j - b_k) / q(j) \quad (25)$$

Note that in eq. (25), the values of $q(j)$'s are determined by eq. (12).

Having determined $q(j)$'s we can calculate TC and CC probabilities based on eq. (13). However, to calculate CC probabilities of service-class k , $q(j)$'s should be determined assuming $N_k - 1$ traffic sources.

4.2. A system with both hard and soft blocking

To introduce LB in the EnMLM we adopt the approach of [4]. Based on eq. (20), the transition rate from state $(j - b_k)$ to j , equals $(1 - L_{j-b_k, b_k})(S_k - \bar{n}_k(j - b_k))\gamma_k = (1 - L_{j-1})(S_k - \bar{n}_k(j - b_k))\gamma_k$. Figure 8 shows the system's state transition diagram. Note that $\bar{n}_k(j)$ is the average number of service-class k calls in state j .

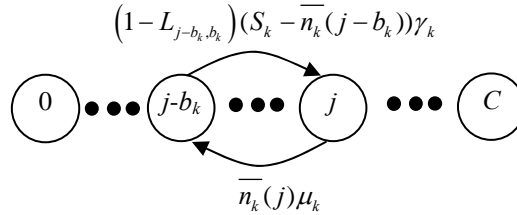


Figure 8: State transition diagram for quasi-random service-class k calls with LB between states $j - b_k$ and j .

To derive an approximate but recursive formula for the calculation of the (un-normalized) link occupancy distribution, $q(j)$, we assume that the following local balance equation does exist:

$$(S_k - \bar{n}_k(j - b_k))(1 - L_{j-b_k, b_k})\gamma_k q(j - b_k) = \bar{n}_k(j)\mu_k q(j) \quad (26)$$

Multiplying both sides of eq. (26) by b_k and summing up over all service-classes $k = 1, \dots, K$, we obtain:

$$q(j) = \frac{1}{j} \sum_{k=1}^K (S_k - n_k(j) + 1) a_{k,fin} b_k q(j - b_k) (1 - L_{j-b_k, b_k}), \text{ for } j = 1, \dots, C \quad (27)$$

since $j = \sum_{k=1}^K \bar{n}_k(j) b_k$ and $\bar{n}_k(j - b_k) = n_k(j) - 1$ assuming that $\bar{n}_k(j) = n_k(j)$.

The determination of $q(j)$'s in eq. (27) requires the value of the in-service calls of service-class k in state j , $n_k(j)$, which is unknown. To circumvent this problem we approximate $n_k(j)$, as the mean number of service-

class k calls in state j , $y_k(j)$, when Poisson arrivals are considered, i.e., $n_k(j) \approx y_k(j)$ and $n_k(j)-1 \approx y_k(j-b_k)$. Based on this approximation, eq. (27) takes the form:

$$q(j) = \frac{1}{j} \sum_{k=1}^K (S_k - y_k(j-b_k)) a_{k,\text{fin}} b_k q(j-b_k) (1-L_{j-b_k, b_k}), \text{ for } j=1, \dots, C \quad (28)$$

where the values of $y_k(j)$ are given by eq. (29) in the case of Poisson traffic:

$$y_k(j) = a_k (1-L_{j-b_k, b_k}) q(j-b_k) / q(j) \quad (29)$$

Note that the values of $q(j)$'s in eq. (29) are determined by eq. (21) which refers to Poisson arrivals.

Having determined LBP by eq. (20) and link occupancy distribution by eq. (28) we calculate TC probabilities of service-class k calls based on eq. (22). Equation (22) can also be used for CC probabilities of service-class k , but $q(j)$'s should be determined by eq. (28) with $S_k - 1$ traffic sources.

5. Numerical results

In this section, we compare the analytical and simulation results of TC probabilities, obtained by the proposed model for the quasi-random process, for different values of the IC efficiency. For comparison, we also present the corresponding analytical results obtained in the case of the Poisson process. Simulations are carried via the SIMSCRIPT III language [27] and are mean values of 7 runs.

Consider a W-CDMA cell that accommodates calls of $K=3$ different service-classes. Calls arrive in the cell according to a quasi-random process. Accepted calls remain in the link for an exponentially distributed service time with mean value $\mu_1^{-1} = \mu_2^{-1} = \mu_3^{-1} = 1$. In Table III, we present the traffic characteristics of all service-classes. The last column of Table III presents the corresponding offered traffic-load in the case of Poisson traffic. In addition, we assume that: $\eta_{UL}=0.75$, $i=0.35$, $\delta=2.0$, $bbu=13.5$ kcps while the IC efficiency β takes the values 0.0 and 0.8. In the x-axis of Figs 9-11 the offered traffic load of the 1st, 2nd and 3rd service-class increase in steps of 0.05, 0.10 and 0.002 erl, respectively. So, point 1 refers to: $(\alpha_{1,\text{fin}}, \alpha_{2,\text{fin}}, \alpha_{3,\text{fin}}) = (0.15, 0.20, 0.01)$ while point 6 to: $(\alpha_{1,\text{fin}}, \alpha_{2,\text{fin}}, \alpha_{3,\text{fin}}) = (0.40, 0.70, 0.02)$.

Figures 9a-9b, present the analytical and simulation TC probabilities of the 1st service-class for $\beta=0.0$ and 0.8, respectively. Figures 10a-10b and 11a-11b, present the corresponding results of the 2nd and 3rd service-class, respectively. Based on these results, we conclude that: 1) The proposed formulas give quite accurate results compared to simulation. 2) The increase of β results in the decrease of TC probabilities. This is anticipated, since the IC reduces the own-cell interference. 3) The TC probabilities obtained by assuming Poisson arrivals fail to approximate the corresponding TC probabilities in the case of the quasi-random process. This result shows the necessity of the proposed model.

Table III: Traffic parameters of all service-classes.

Service-class k	R_k (kbps)	v_k	$\left(\frac{E_b}{N_0}\right)_k$ (in dB)	$\left(\frac{E_b}{N_0}\right)_k$	S_k	$a_{k,\text{fin}}$ (in erl)	$a_k = S_k a_{k,\text{fin}}$ (in erl)
1	7.95	0.67	4.0	2.51	20	0.15	3.0
2	12.20	0.67	4.0	2.51	10	0.20	2.0
3	144.00	1.0	2.0	1.58	5	0.01	0.05

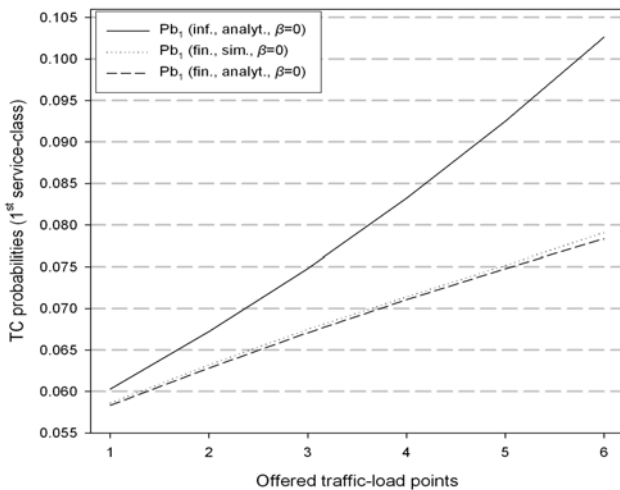


Figure 9a: TC probabilities of the 1st service-class ($\beta=0.0$).

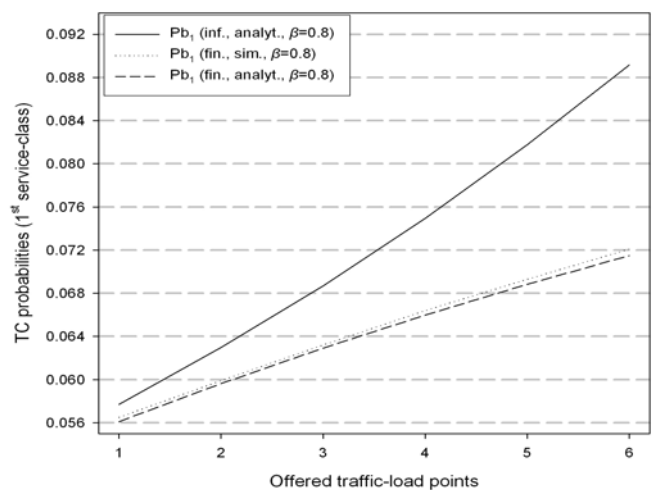


Figure 9b: TC probabilities of the 1st service-class ($\beta=0.8$).

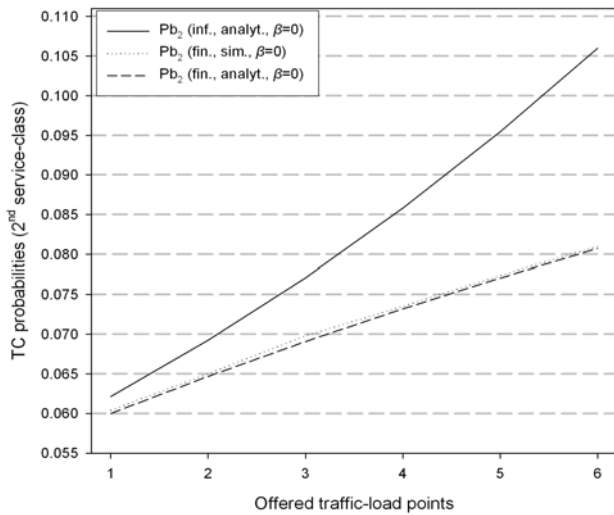


Figure 10a: TC probabilities of the 2nd service-class ($\beta=0$).

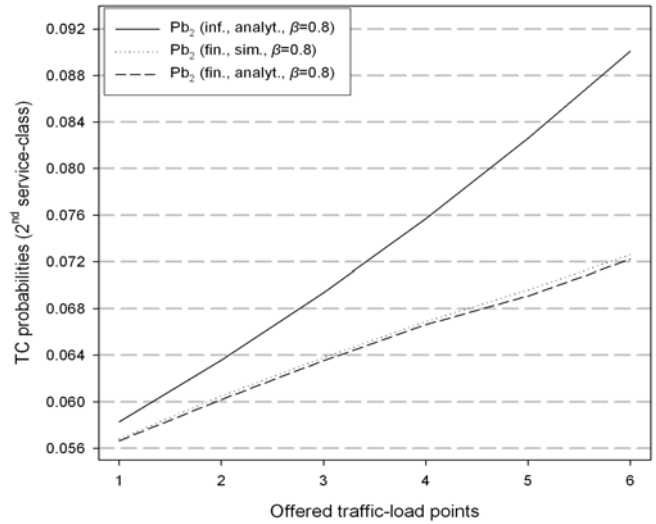


Figure 10b: TC probabilities of the 2nd service-class ($\beta=0.8$).

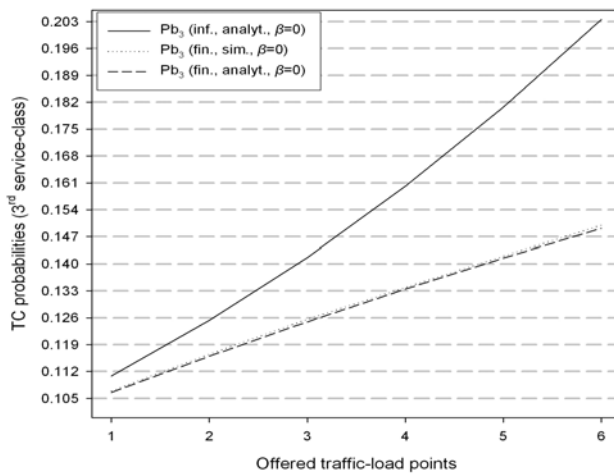


Figure 11a: TC probabilities of the 3rd service-class ($\beta=0$).

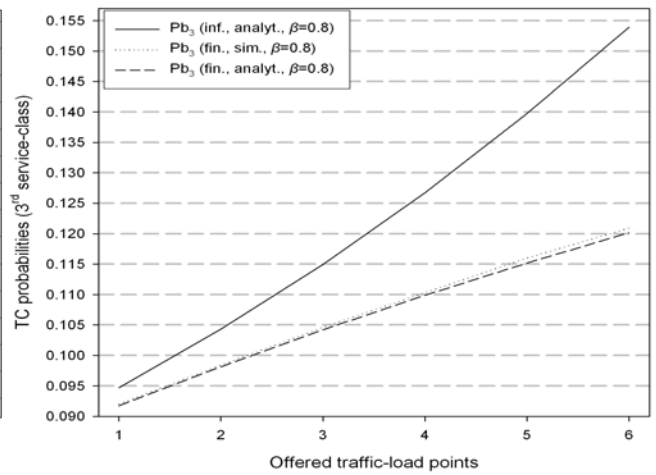


Figure 11b: TC probabilities of the 3rd service-class ($\beta=0.8$).

6. Conclusion

We propose a multirate loss model for the call-level analysis of W-CDMA networks that support calls from different service-classes whose arrival follows a quasi-random process. The new model takes into account multiple access interference, the notion of local blocking, user's activity and interference cancellation. Due to the existence of local blocking, the proposed model does not have a product form solution. However, we show an approximate but recursive formula for the calculation of occupancy distribution and consequently the determination of time and call congestion probabilities. Simulation results verify the accuracy of the proposed model.

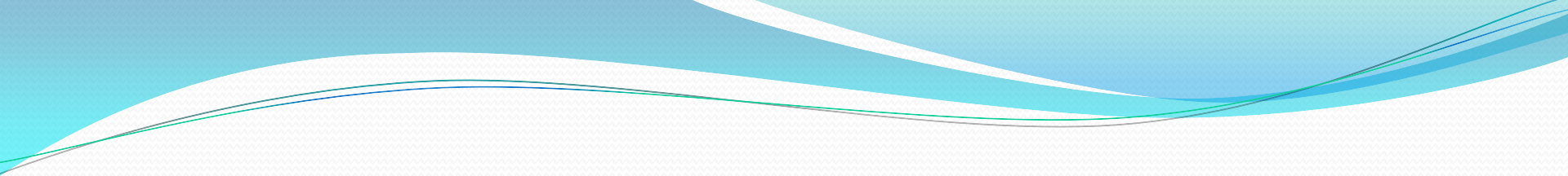
References

- [1] K. Ross, *Multiservice Loss Models for Broadband Telecommunication Networks*, Springer, London, 1995.
- [2] V. Iversen, V. Benetis, N. Ha, and S. Stepanov, "Evaluation of Multi-service CDMA Networks with Soft Blocking", Proc. ITC Specialist Seminar, pp. 223-227, Antwerp, Aug./Sept. 2004.
- [3] V. Iversen, "Evaluation of Multi-service CDMA Networks with Soft Blocking", Proc. of 3rd Conference on Smart Spaces, ruSMART 2010, and 10th Int. Conference, NEW2AN 2010, St. Petersburg, Russia, pp. 160-171, 23-25 Aug. 2010.
- [4] L. Popova and W. Koch, "Analytical performance evaluation of mixed services with variable data rates for the uplink of UMTS", Proc. ISWCS'06, Valencia, Spain, Sept. 2006.
- [5] G. Kallos, V. Vassilakis and M. Logothetis, "Call-level performance analysis of a W-CDMA cell with finite population and interference cancellation", European Transactions on Telecommunications, vol. 22, issue 1, pp. 25-30, Jan. 2011.

- [6] J. Kaufman, "Blocking in a shared resource environment", *IEEE Trans. Commun.* vol. 29, no. 10, pp. 1474-1481, Oct. 1981.
- [7] J. Roberts, "A service system with heterogeneous user requirements", in: G. Pujolle (Ed.), *Performance of Data Communications systems and their applications*, North Holland, Amsterdam, pp.423-431, 1981.
- [8] L. Delbrouck, "On the steady state distribution in a service facility with different peakedness factors and capacity requirements", *IEEE Trans. Commun.*, 31 (11), pp. 1209-1211, Nov. 1983.
- [9] H. Akimaru and K. Kawashima, *Teletraffic – Theory and Applications*, 2nd edition, Springer-Verlag, Berlin, 1999.
- [10] C. Rosa, T. Sorensen, J. Wigard and R. Mogensen, "Interference cancellation and 4-branch antenna diversity for W-CDMA uplink packet access", *Proc. IEEE Vehicular Technology Conference 2005*, pp. 1758-1762, 30 May – 1 June 2005.
- [11] H. Holma and A. Toskala, eds., *W-CDMA for UMTS – HSPA Evolution and LTE*, 4th edition, Wiley, 2007.
- [12] M. Stasiak, M. Glabowski, A. Wisniewski, and P. Zwierzykowski, *Modeling and Dimensioning of Mobile Networks*, Wiley, 2011.
- [13] S. Hamalainen, H. Holma and A. Toskala "Capacity evaluation of a cellular CDMA uplink with multiuser detection", *Proc. of 4th Int. Symposium on Spread Spectrum Techniques and Applications*, Sept. 1996, vol. 1, pp. 339-343.
- [14] D. Staehle, K. Leibnitz, K. Heck, B. Schröder, A. Weller, and P. Tran-Gia, "Approximating the othercell interference distribution inhomogeneous UMTS networks", *Proc. IEEE VTC 2002*, pp. 1640-1644, 6-9 May 2002.
- [15] D. Staehle and A. Mäder, "An analytic approximation of the uplink capacity in a UMTS network with heterogeneous traffic," *Proc. 18th International Teletraffic Congress (ITC18)*, Berlin, pp. 81-91, 31 Aug./Sept. 2003.
- [16] G. Stamatielos and J. Hayes, "Admission control techniques with application to broadband networks", *Comput. Commun.*, vol. 17, no. 9, pp. 663-673, 1994.
- [17] I. Moscholios, M. Logothetis and P. Nikolaropoulos, "Engset Multi-Rate State-Dependent Loss Models", *Performance Evaluation*, Vol. 59, Issues 2-3, pp. 247-277, Feb. 2005.
- [18] I. Moscholios and M. Logothetis, "Engset Multirate State-Dependent Loss Models with QoS Guarantee", *International Journal of Communications Systems*, Vol. 19, Issue 1, Feb. 2006, pp. 67-93.
- [19] I. Moscholios, M. Logothetis and M. Koukias, "An ON-OFF Multirate Loss Model of Finite Sources", *IEICE Trans. Commun.*, Vol. E90-B, No.7, pp. 1608-1619, July 2007.
- [20] I. Moscholios, M. Logothetis and G. Kokkinakis, "On the calculation of Blocking Probabilities in the Multirate State-Dependent Loss Models for Finite Sources", *Mediterranean Journal of Computers and Networks*, Vol. 3, No. 3, July 2007, pp. 100-109.
- [21] M. Glabowski and M. Stasiak, "An approximate model of the full-availability group with multi-rate traffic and a finite source population", in *Proc. of 12th MMB&PGTS*, Dresden, Germany, pp. 195-204, Sept. 2004.
- [22] V. Vassilakis, G. Kallos, I. Moscholios and M. Logothetis, "Call-Level Analysis of W-CDMA Networks Supporting Elastic Services of Finite Population", *IEEE ICC 2008*, Beijing, China, 19-23 May 2008.
- [23] M. Glabowski, K. Kubasik, and M. Stasiak, "Modelling of Systems with Overflow Multi-rate Traffic and Finite Number of Traffic Sources," in *Proc. CNSDSP 2008*, pp. 196–199, July 2008.
- [24] M. Glabowski, A. Kaliszan and M. Stasiak, "Modeling product-form state-dependent systems with BPP traffic", *Performance Evaluation*, vol. 67, issue 3, pp. 174-197, March 2010.
- [25] J. Vardakas, I. Moscholios, M. Logothetis, V. Stylianakis and F. Röck, "Evaluation of Dynamic Wavelength Allocation Scenarios in WDM-TDMA PONs Servicing ON-OFF Traffic of Finite Sources", *Proc. of 11th Int. Conf. on Telecommun., ConTEL 2011*, Graz, Austria, 15-17 June 2011.
- [26] I. Moscholios, J. Vardakas, M. Logothetis and M. Koukias, "A Quasi-random Multirate loss model supporting elastic and adaptive traffic", *Proc. of 4th Int. Conference on Emerging Network Intelligence*, Barcelona, Spain, 23-28 Sept. 2012.
- [27] Simscript III <http://www.simscrip.com/>. (Accessed on September 2013).

Traffic Control Based Call Admission Control Scheme in Mobile Cellular Networks

Mohammed Alrowili

- 
- In this presentation, we present :
 - Introduction
 - System model
 - Assumptions
 - Results and discussion
 - Future work

Call Admission Control schemes(CAC):

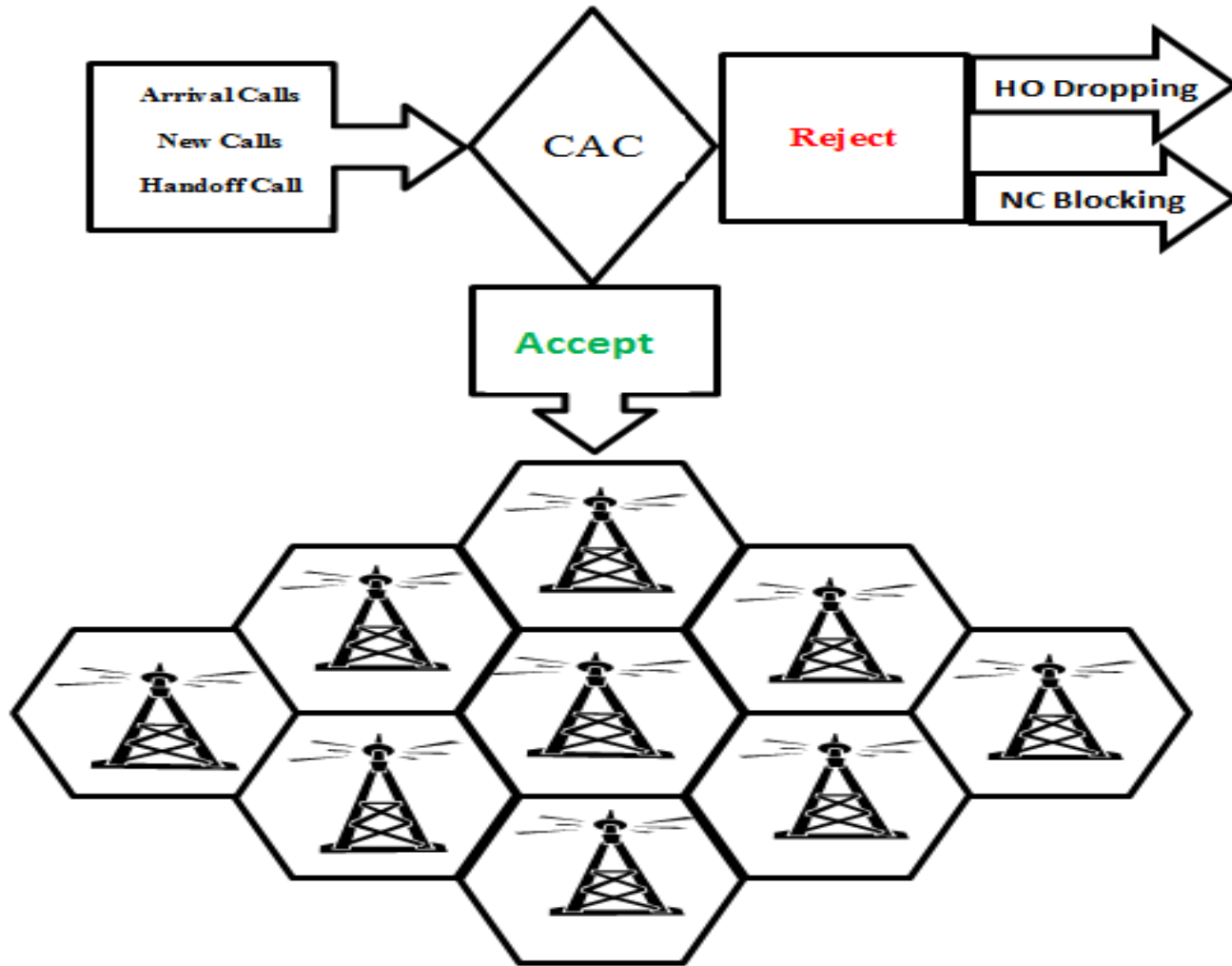
Definition of CAC:

CAC is one of the radio resource management policies, which is used as a QoS provisioning strategy [1, 2]. CAC can determine the total number of calls into the network in order to meet the QoS requirements and minimize the probabilities of dropping and blocking of handoff calls and new calls respectively by accepting or rejecting the connection requests.

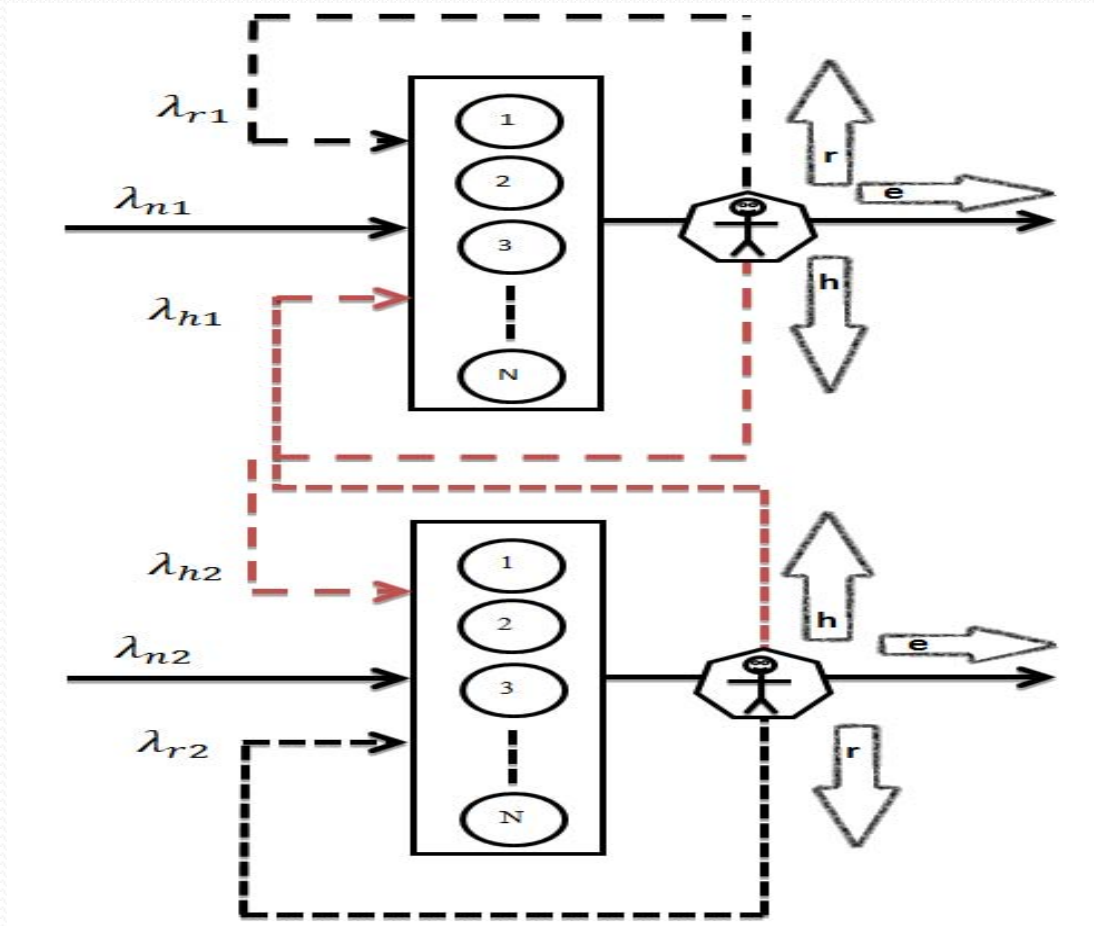
[1] L. Barolli, G. Mino, T. Ikebata, A. Barolli, F. Xhafa, and M. Takizawa, "Performance Evaluation and Comparison of Fuzzy-Based Intelligent CAC Systems for Wireless Cellular Networks," in *Intelligent Networking and Collaborative Systems (INCOS), 2010 2nd International Conference on*, 2010, pp. 177-184.

[2] N. Naghavi, V. Friderikos, and H. Aghvami, "Joint Revenue-Based Call Admission Control and Routing in Wireless Mesh Networks," in *Communications (ICC), 2011 IEEE International Conference on*, 2011, pp. 1-6.

Call Admission Control system:



System Model:



P21-5

System model description and assumptions:

- Two identical cells each of which has 100 channels.
- Active calls divide into two branches 50% Handover calls(H) and 50% end calls(E) .
- The threshold of triggering messaging system will be 25%, 50%, 75% and 100% of a cell capacity.
- In case of exceeding target cell threshold, H will divide into three categories depending on the user behaviour(h, r and e):
 - (r) When user stays at the same cell with active call.
 - (h) when user moves to the neighbouring cell.
 - (e) when user ends his call before handover.
- Poisson distribution is used for arrival traffic with rate between 50 and 500 user/minute, and 5 minutes as a mean for holding time.

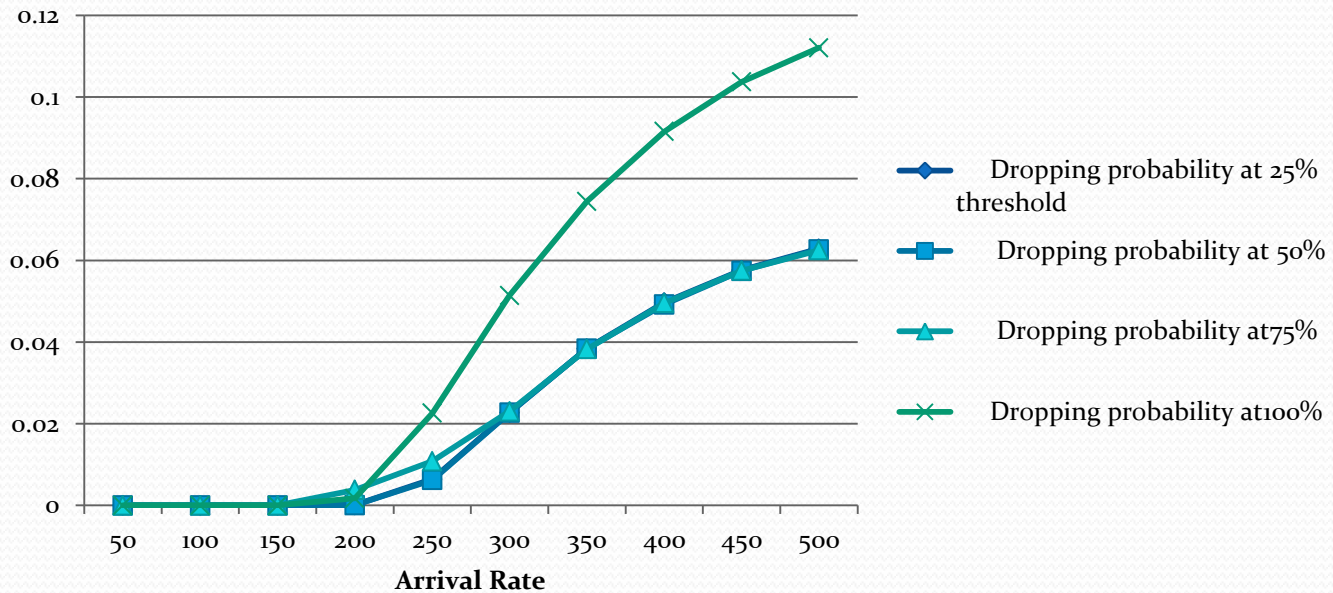
Five different scenarios were examined as shown in the following table:

Scenario #	1	2	3	4	5
h	10%	100%	34%	60%	20%
r	40%	0%	33%	20%	60%
e	50%	0%	33%	20%	20%

Scenario # 1

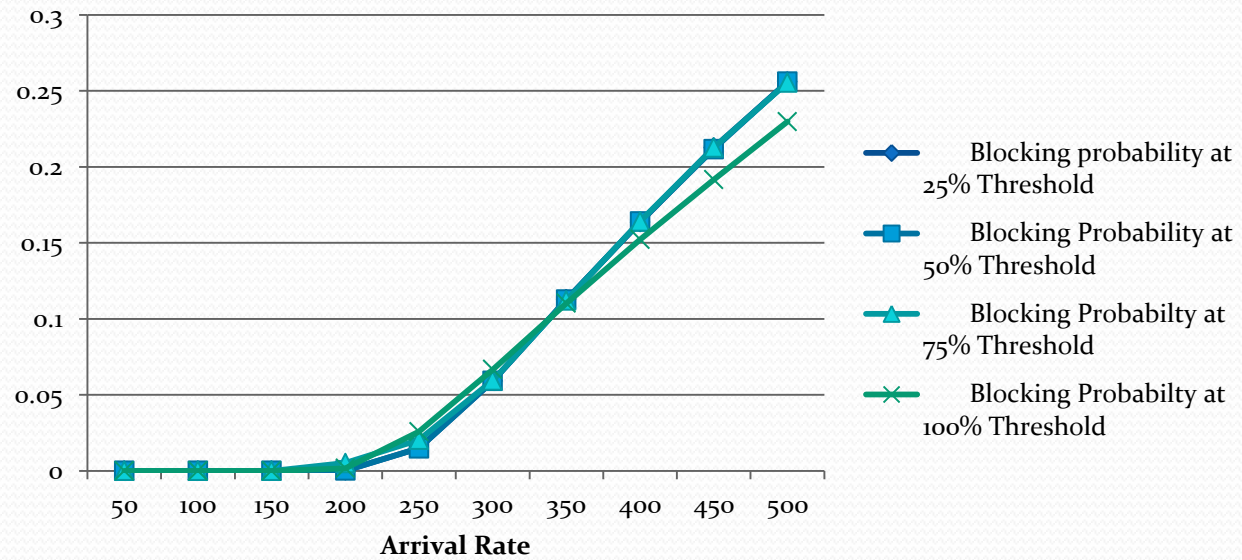
E	H	h(%H)	r(%H)	e(%H)
50%	50%	10%	40%	50%

Handoff Call Dropping Probability



Scenario # 1

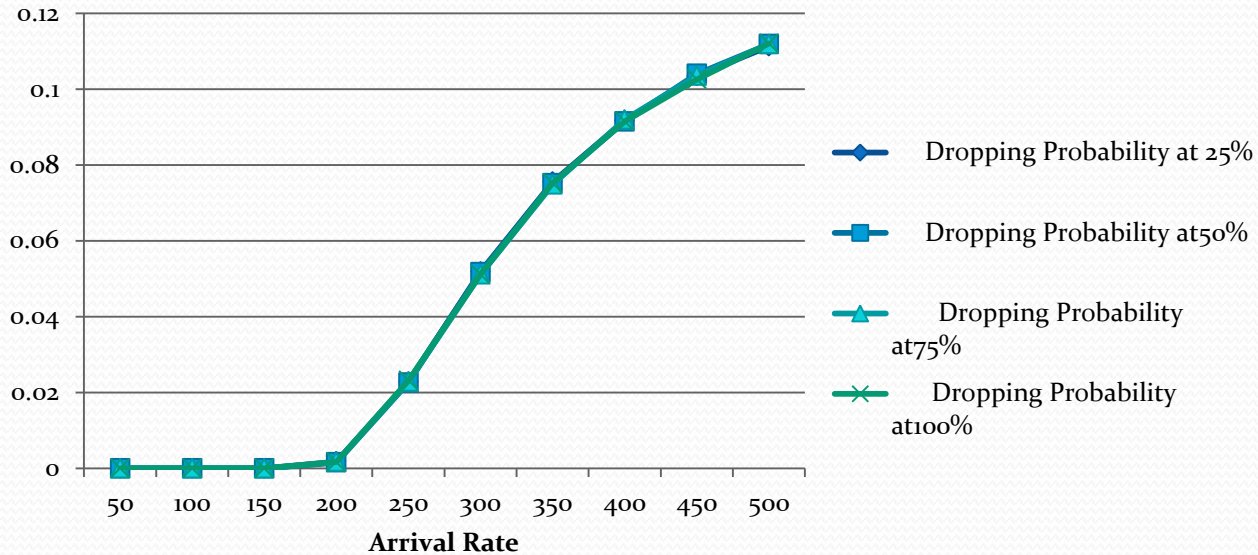
New Call Blocking Probability



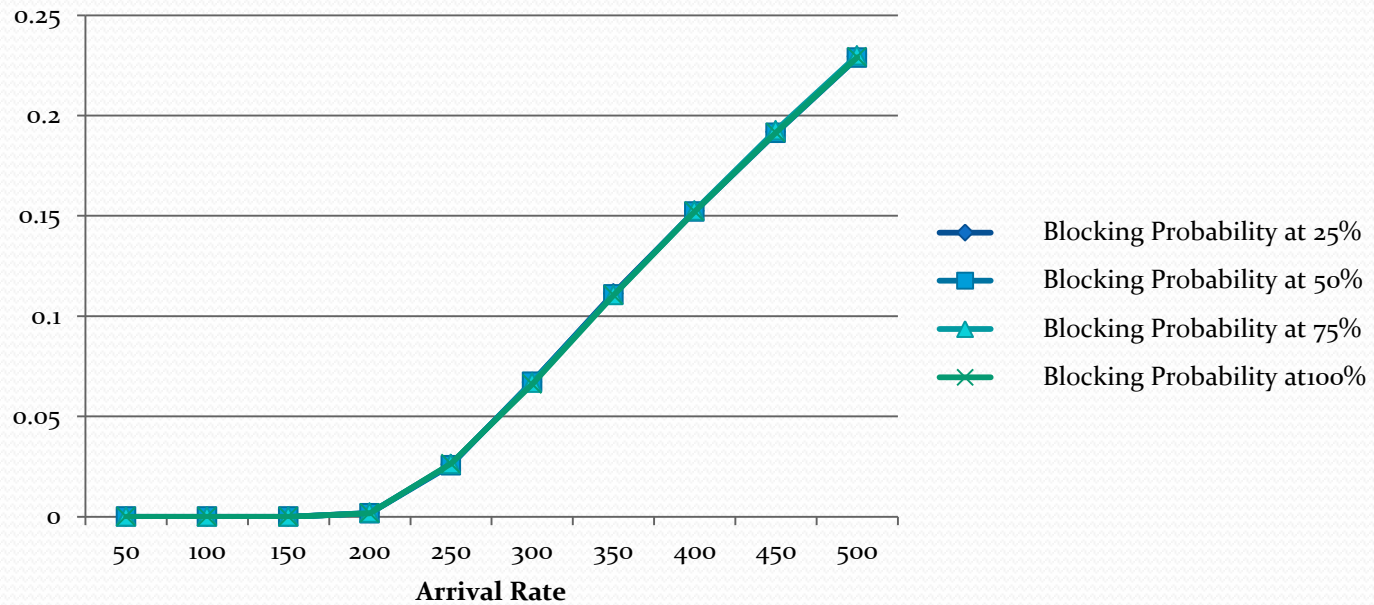
Scenario# 2:

E	H	h(%H)	r(%H)	e(%H)
50%	50%	100%	0%	0%

Handoff Call Dropping Probability



New Calls Blocking Probability

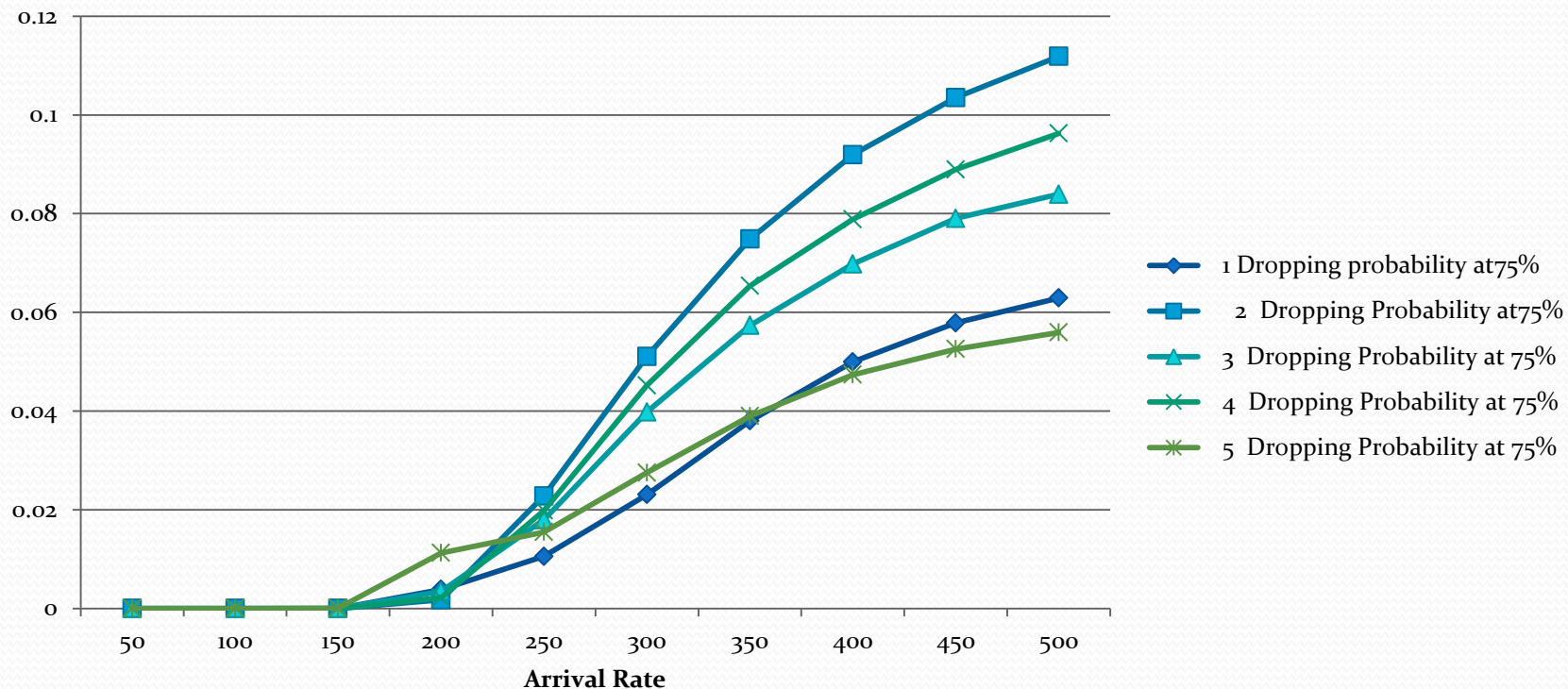


As predicted from theory we can see from the figures above that even though the messaging system has been utilized but in the condition where the 100% of the users decided to make the handoff no increase in the blocking probability was found and no improvement in the dropping probability. But users will be aware of the network status.

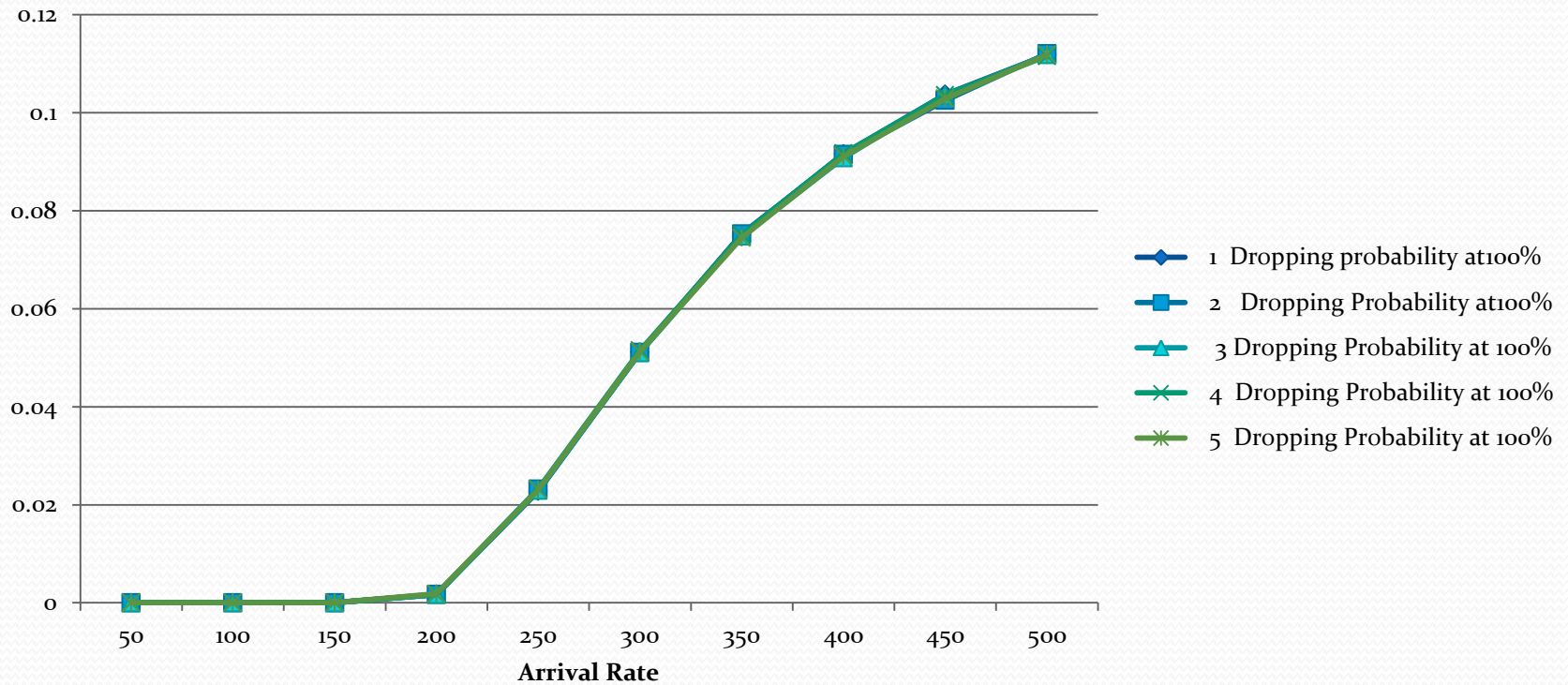
Now from looking at the figure below it is evident that if the same conditions are imposed on the network the new call blocking probability increases when compared with the no messaging which can help to cater for more handoff calls and thus reducing the total number of annoying forced terminations. Call blocking and call dropping both have conflicting requirements that is why in order to achieve optimization between the two is extremely difficult.

Comparison between the five scenarios at different thresholds:

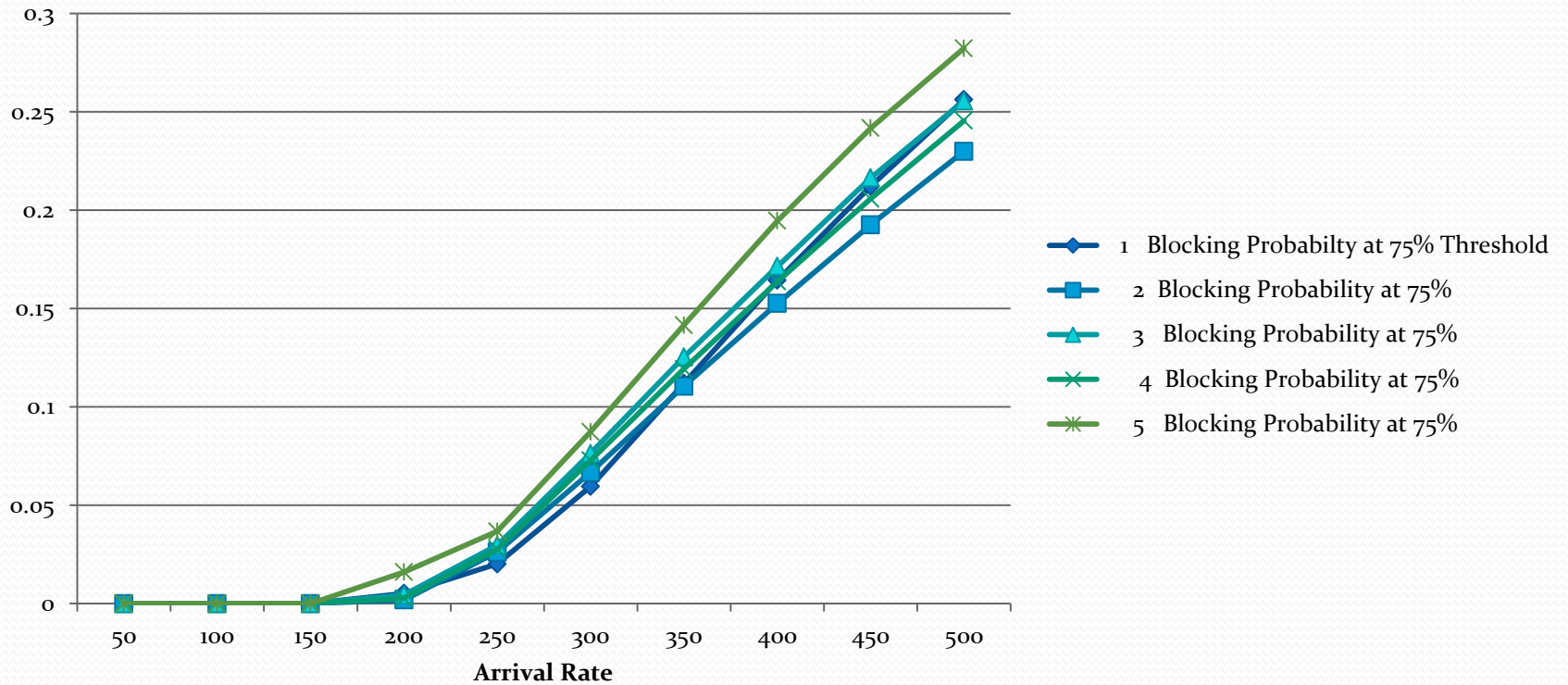
Handoff Call Dropping Probability for 75% threshold in the five scenarios



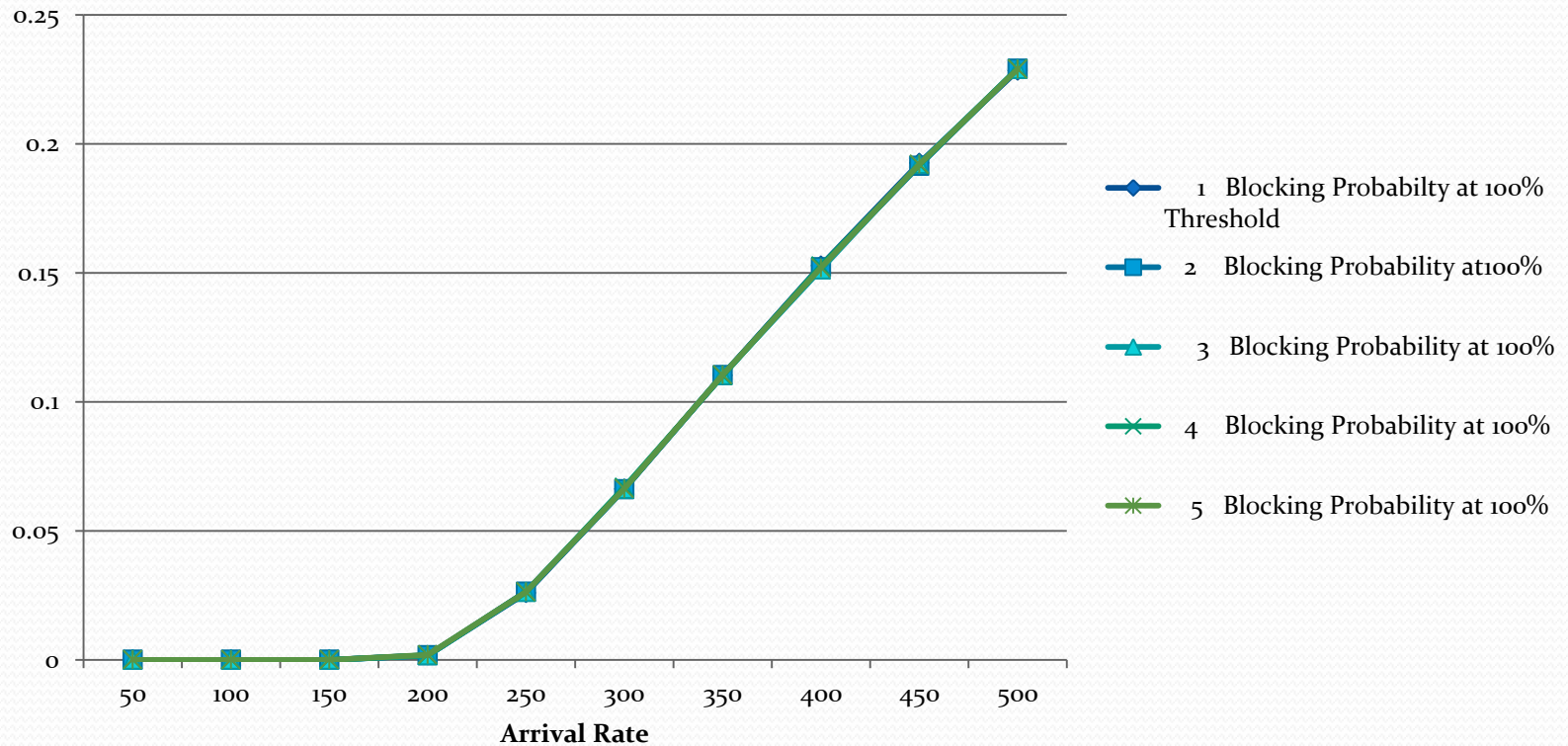
Handoff Call Dropping Probability for 100% threshold in the five scenarios



New call Blocking Probability For 75% Threshold



New call Blocking Probability For 100% Threshold



Conclusion:

Wireless cellular networks performance can be improved by controlling the users' behaviour when there is a sophisticated messaging system which updates users about the network status.

Future Work:

- Using multi class traffic(Data, Video).
- Using machine learning techniques to find the optimal performance.
- Comparing the findings with existing algorithms.

PART SIX Research, Standardisation and Networking



Welcome to the World of Standards



RESEARCH INPUT TO STANDARDISATION **TECHNOLOGY INNOVATION THROUGH STANDARDISATION**



Hermann Brand
Director Innovation

IFIP HET NETs 2013
Ilkley, UK, 11 November 2013

- I will not talk about latest research results
- BUT
- about how standardisation may help to exploit them

- Technical standards are full of scientific discoveries and technologies
- BUT
- There is no point to standardise all technologies

- ***Why standardisation of ICT is important***
- Why standardisation is tricky
- How standardisation/ETSI works
- What ETSI is/does
- Opportunities to get involved in standardisation

-  Technical specifications and standards enable **interoperable solutions** allowing for **network effects**
 - Classic example: telephone
-  Standardisation aims at making **markets as large and homogeneous as possible** to allow for **economy of scale**

What is in standardisation for a researcher?



- Meet the industry and understand their technical problems
- Meet the industry to understand the relevance of your research (reality check)
- Disseminate and exploit research results
 - Present and discuss your research results
 - Improve existing standardised technologies
 - Propose alternative technologies as solution elements
 - Anchor a patent as essential IPR in a standard
 - Reports and specifications included in ETSI deliverables will be maintained and made available to the public forever

- 🌐 Why standardisation of ICT
- 🌐 ***Why standardisation is tricky***
- 🌐 How standardisation/ETSI works
- 🌐 What ETSI is/does
- 🌐 How ETSI helps you to get involved

Standardisation is tricky



- Its a timing problem (life cycles)
- Its cooperation in order to compete
- Its consensus based

- technology convergence results in accelerated market segmentation and alternative solutions (e.g. POT, smartphone, Skype, RTCWeb, ...)

- ICT drives innovation in more and more sectors (e.g. Smart transport, smart grid, smart cities, ...), cross-sector standardisation

Meeting the window of opportunity



Pre-Standardization	Standard Setting	Post-Standardization
Embryonic stage: <input type="checkbox"/> Technology breeding <input type="checkbox"/> Business nurturing	'The making of a standard'	Maintenance: <input type="checkbox"/> Corrections <input type="checkbox"/> Incremental feature enhancements
Not yet a best solution	There is a best solution (at least consensus on a solution) Overall system architecture frozen	Enhancements and enrichments of the solution
Studies <input type="checkbox"/> Gap analysis <input type="checkbox"/> Assessments of competing solutions <input type="checkbox"/> Roadmapping	Technical standards	CR and new releases

Technology maturity matrix



High

DIFFERENTIATION

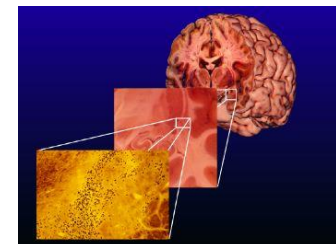
Low

2. Pacemaker	3. Key
1. Emerging	4. Basic

Low

High

INTEGRATION

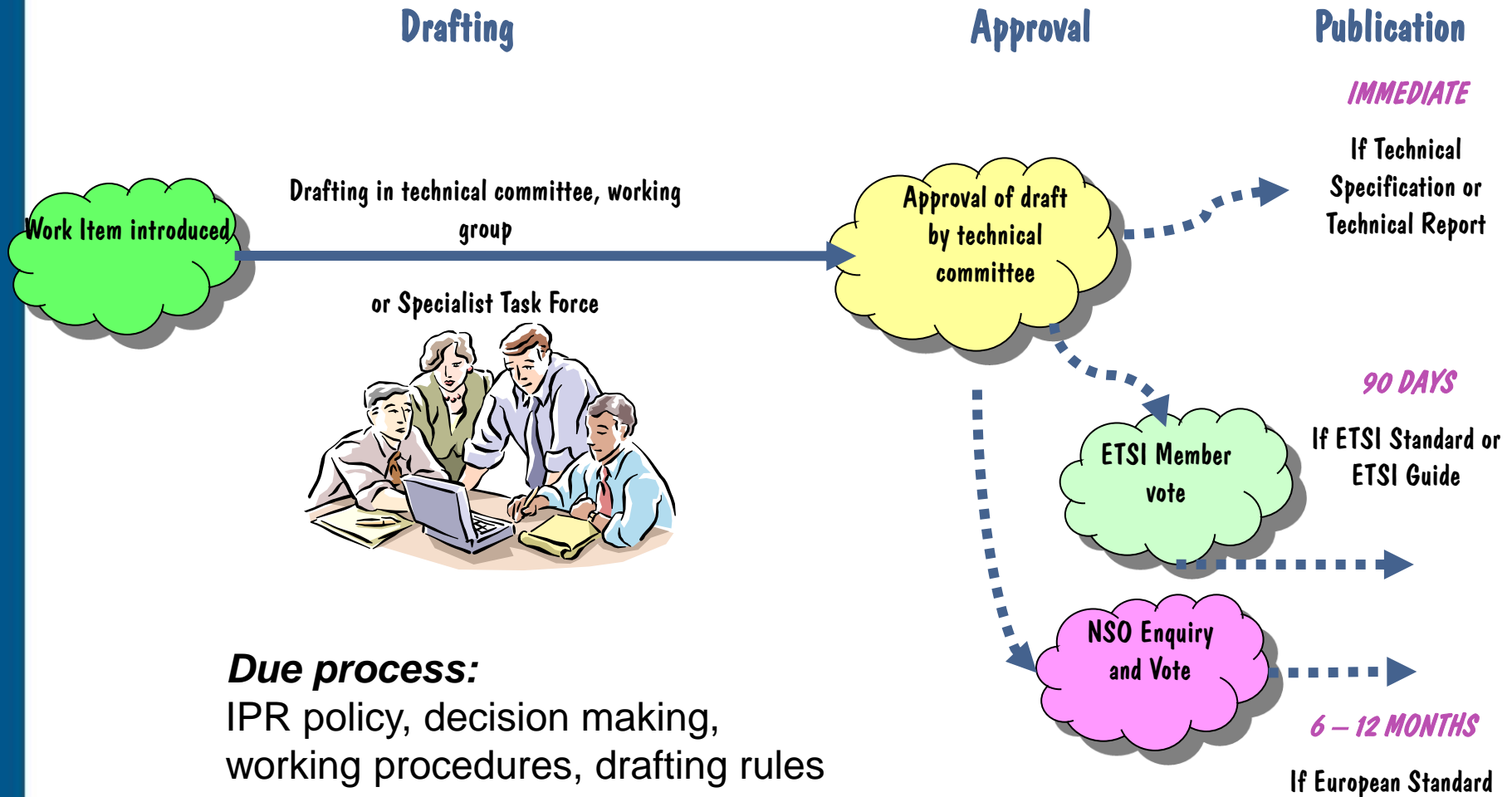


- Research: transforms money into knowhow. Innovation: transforms knowhow into money.
- Technologies do not have a value per se (even if they generate technology push). The value of a technology results from a business model.
- Standardisation is collaborative business development to coordinate technology and market development .
- Standards makers are those stakeholders who need the standards to develop their products and services in order to drive their business. They are also users of the standard.
- ICT markets are global and standards are pervasive in ICT, therefore ICT innovation must result in standardisation

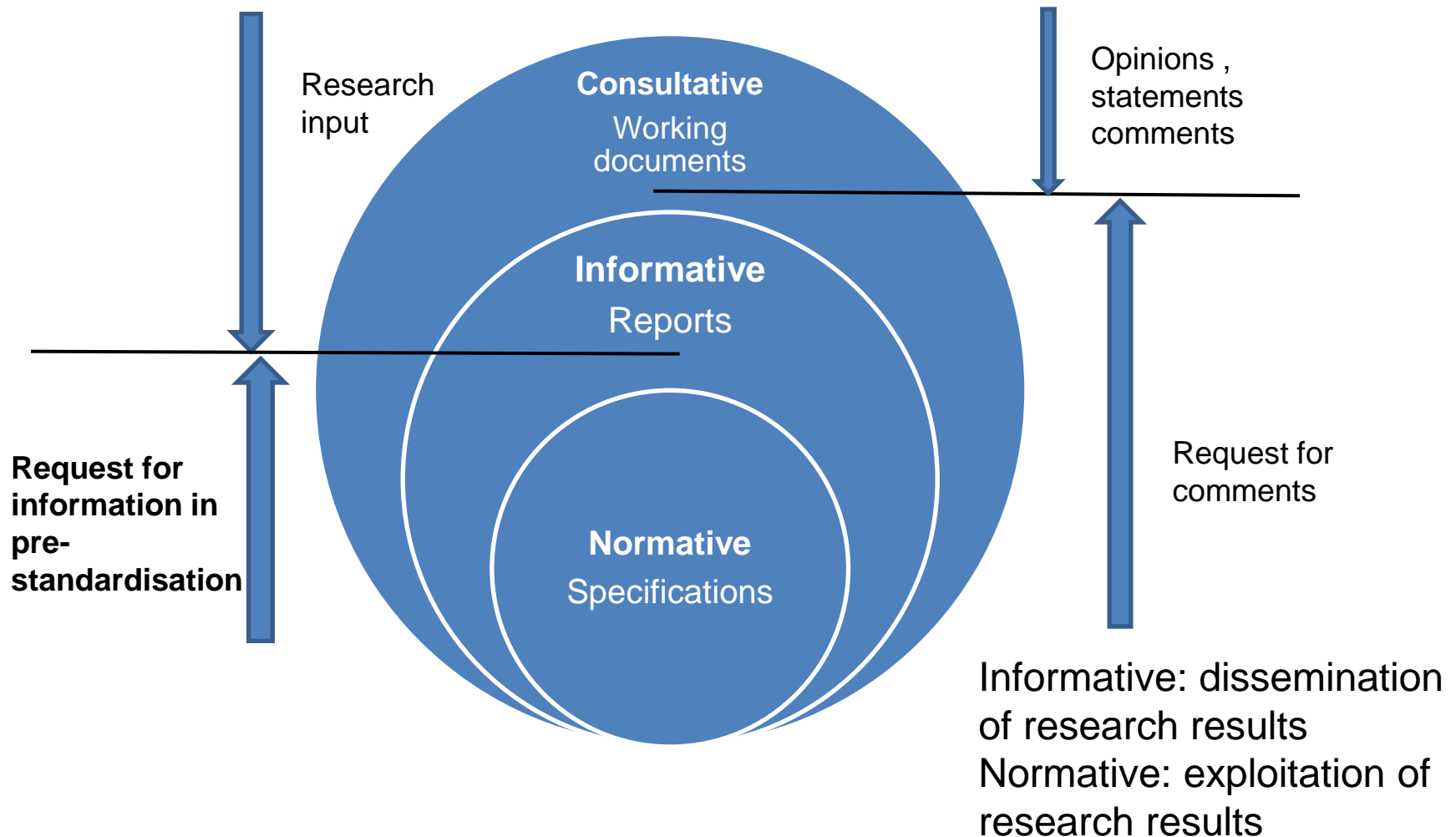
- 🌐 Why standardisation of ICT
- 🌐 Why standardisation is tricky
- 🌐 ***How standardisation/ETSI works***
- 🌐 What ETSI is/does
- 🌐 Opportunities to get involved in standardisation

- ETSI follows a **life-cycle based approach to standardisation** throughout all phases of commercial products/services, from conception to market introduction
- 'Standards engineering' in the ETSI sense is a pragmatic and results-oriented process. It covers:
 - Linking to **research** in order to anticipate and identify standardisation needs
 - Support of **early consensus** and community building by converting research communities into **pre-standardization** communities
 - **Setting standards** in a proper type of organizational setup (based on an e2e system view)
 - And last but not least, hands on **verification of interoperable implementations** of standards (not only from ETSI)

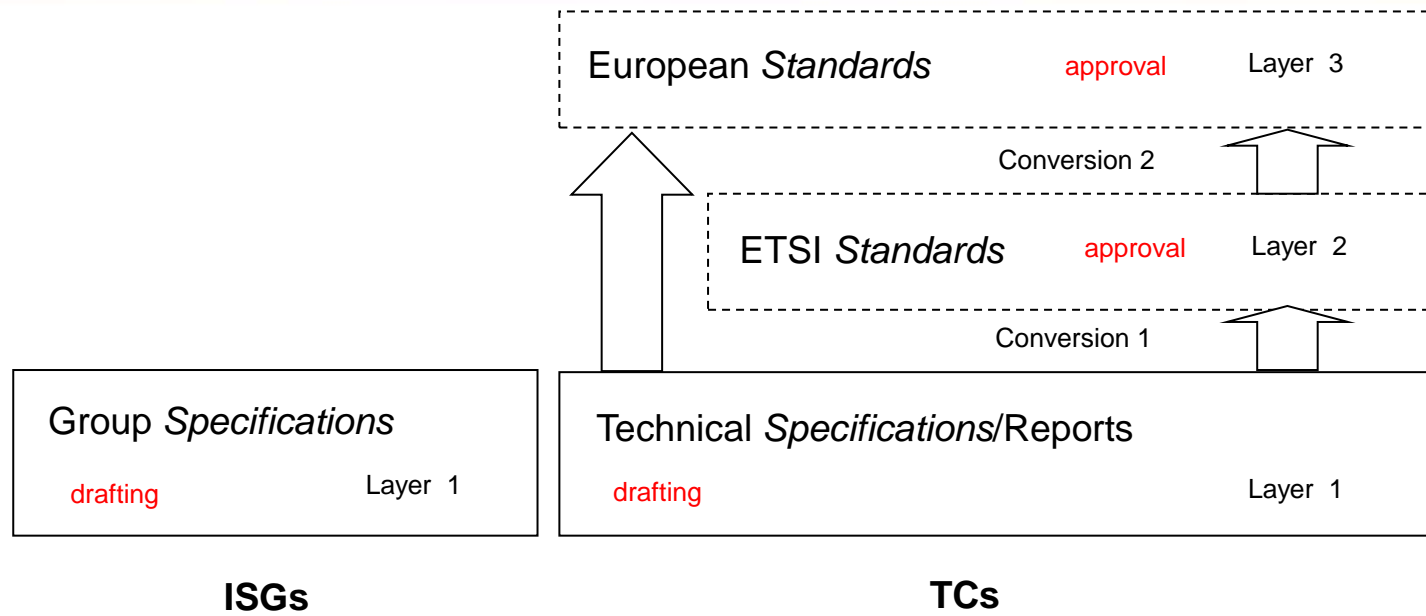
The ETSI standards process



How ETSI works - drafting



How ETSI works – approving



Member delegates develop deliverables by making written contributions

- Separation of working level, i.e. **drafting of deliverables** from **adoption of standards**
- Through approval procedures **consensus is formally established** within different communities of stakeholders (members, NSOs/National Delegations)

How ETSI works - different lines of production



ETSI branded – TCs/EPs and ISGs

	<i>ISGs</i>	<i>TCs/EPs</i>
Decision power	DG	Board
Participation	Members have to sign an agreement, contracted non-ETSI member participants (bound to IPR Policy)	Members only (Guests on a temporary and exceptional basis without the right to make technical contributions because they are not bound to IPR policy)
Voting	Project specific (usually one member one vote)	Weighted voting
Fees	Participation fee per meeting for non-member participants	Annual membership fee
IPR Policy	<i>FRAND</i>	<i>FRAND</i>

Non-ETSI branded (own identity)

- Partnership Projects (PPs)
 - Joint membership of organizational partners
 - Joint technical work (drafting and adoption of specifications)
 - Coordinated adoption by organizational partners in EU, US, Japan, South Korea, China
- Forapolis services for fora/consortia (ETSI inside), e.g.



- 🌐 Why standardisation of ICT
- 🌐 Why standardisation is tricky
- 🌐 How standardisation/ETSI works
- 🌐 ***What ETSI is/does***
- 🌐 Opportunities to get involved in standardisation

ETSI Deliverables

different level of consensus and formal status



3GPP TR 33.812 V9.2.0 (2010-06)

Technical Report

3rd Generation Partnership Project;
Technical Specification Group Services and System Aspects;
Feasibility study on the security aspects of remote
provisioning and change of subscription for
Machine to Machine (M2M) equipment
(Release 9)



ETSI GS NFV 002 V1.1.1 (2013-10)



ISG

Network Functions Virtualisation (NFV);
Architectural Framework

The document has been produced and approved by the Network Functions Virtualisation (NFV) ETSI Industry Specification Group (ISG) and represents the views of those members who participated in this ISG. It does not necessarily represent the views of the entire ETSI membership.

ETSI EN 302 625 V1.1.1 (2009-07)

Harmonized European Standard (Telecommunications series)
Disclaimer

Electromagnetic compatibility
and Radio spectrum Matters (ERM);
5 GHz BroadBand Disaster Relief applications (BBDR);
Harmonized EN covering the essential requirements
of article 3.2 of the R&TTE Directive

Group Specification
Harmonized Standard
European Standard

National Delegations, through the
National Standards Organizations (NSOs)

ETSI Guide
ETSI Standard

ETSI Membership

Direct membership of organizations



Over 750 companies, big and small, from more than 60 countries on 5 continents!

In ICT standardization direct membership and free availability of deliverables is the rule!



11% Research institutes & Universities

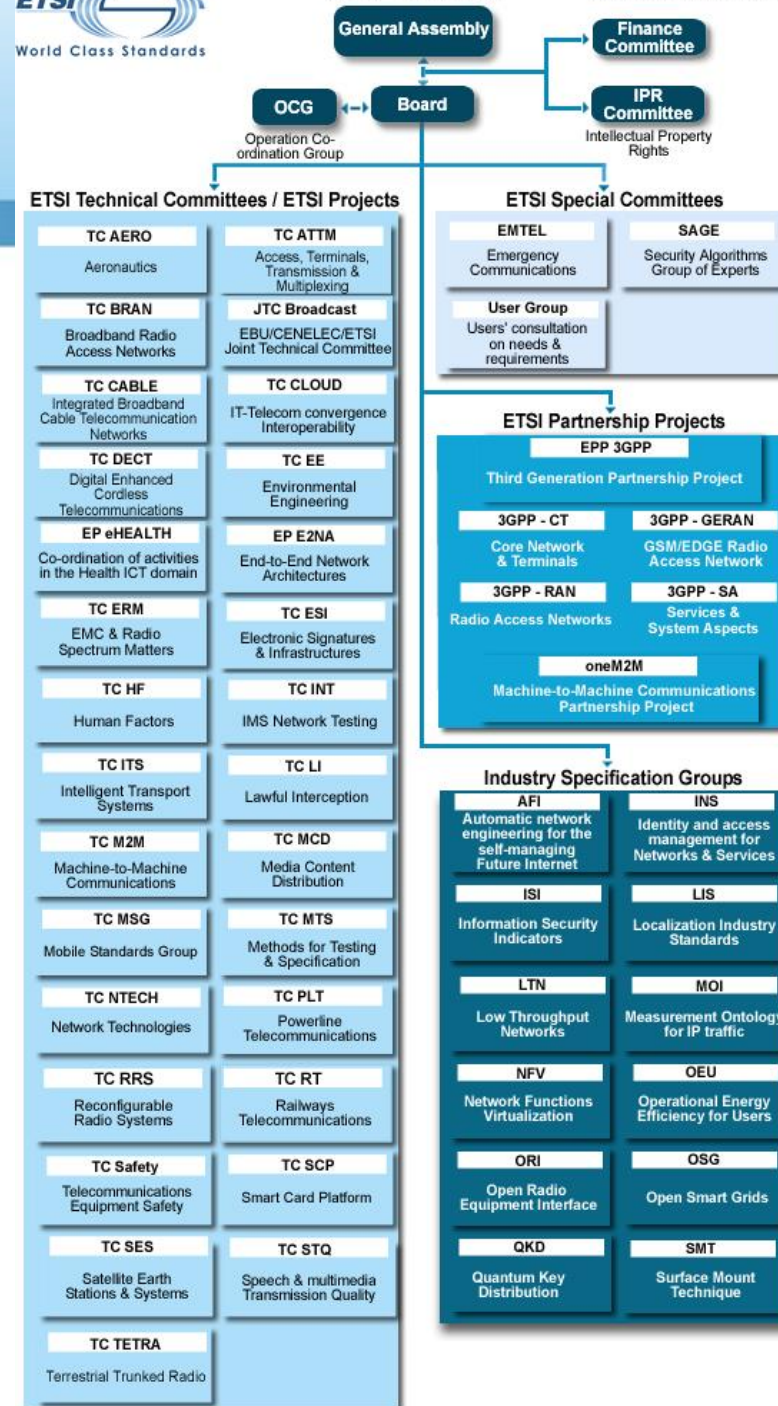
- Manufacturers
- Network operators
- National Administrations
- Ministries
- Service/content providers
- Research bodies
- Universities
- User groups
- Consultancies

Work areas



Work programme at: <http://portal.etsi.org/portal/server.pt/community/home/312>

ETSI structure



General Assembly

Board

- More than **30000** deliverables
- Circa **7000** experts active in technical work program
 - Delegates
 - Experts in STFs

Special Committees

ETSI Technical Body

- Technical Committees
- ETSI Projects
- ETSI Partnership Projects

ETSI Industry Specification Groups

Secretariat

(Experts in Specialist Task Forces)

Partnership Projects

- 3GPP
- oneM2M

Technical Committees

- TC RRS (Reconfigurable Radio Systems)
- TC ITS (Intelligent Transport Systems)
- TC Cloud

ETSI Project E2NA (E2E Network Architectures)

Industry Specification Groups

- ISG QKD (Quantum Key Distribution)
- ISG NFV (Network Function Virtualisation)

ETSI's work includes...



- Telecoms networks
- Content distribution
- DSL, Broadband, Next Generation Networks
- Cable distribution systems
- Exchange equipment
- Protocols
- Power line telecommunications
- Safety
- Speech quality and speech recognition
- Testing
- Grid and cloud networks
- Security
- Electronic signatures, lawful interception, smart cards
- ...



And ETSI's work includes...



- Mobile
 - GSM, UMTS, Railway communications...
 - private/professional land mobile radio systems
- Cordless (DECT)
- Spectrum matters & electromagnetic compatibility
- Radio technologies
- Wireless Local Area Networks
- TETRA & other public safety systems
- Broadcast
- Satellite communications
- Short range devices
- Aeronautical and marine radio
- eHealth and Wireless medical devices
- Intelligent transport
- Internet of the Future



ETSI: one organization – three roles



GSP: Global Standards Producer:
ETSI creates standards intended to meet global needs in ICT (global applicability of technical specifications)



ESO: European Standards Organization:
ETSI produces Harmonized Standards in all areas of telecommunications & ICT used to access European market



SPO: Service Providing Organization:
Interopolis: interoperability engineering
Forapolis: forum management
Plugtests: events, home of TTCN-3



ETSI inside !



- 🌐 Why standardisation of ICT
- 🌐 Why standardisation is tricky
- 🌐 How standardisation/ETSI works
- 🌐 What ETSI is/does
- 🌐 ***Opportunities to get involved in standardisation***

Opportunities to get involved (1)



- Check the work programme for ongoing work and use ETSI deliverables for state of the art analysis (<http://www.etsi.org/standards>)
- Check the ETSI IPR database for patented technical solutions as part of state of the art analysis (<http://ipr.etsi.org/>)
- Consult with us in funded projects (preparation, execution, exploitation phases)

Opportunities to get involved (2)



- Attend and contribute to ETSI workshops/conferences and present your research results
 - Published as eProceedings with ISBN
- Contribute to ongoing standardisation work
 - Expert opinion as guest
 - Contribute to feasibility studies to prepare for standardising the best solution (pre-standardisation)
 - Contributions to normative documents (specifications of implementable technology for interoperable solutions)
- Create a new standardisation task/group/committee (4 ETSI members needed)
 - WI, WG and/or ISG or TC
- Bring product prototypes to plugtests, e.g cloud, engage in trials, e.g. car2car in cities

ETSI work for H2020 societal challenges

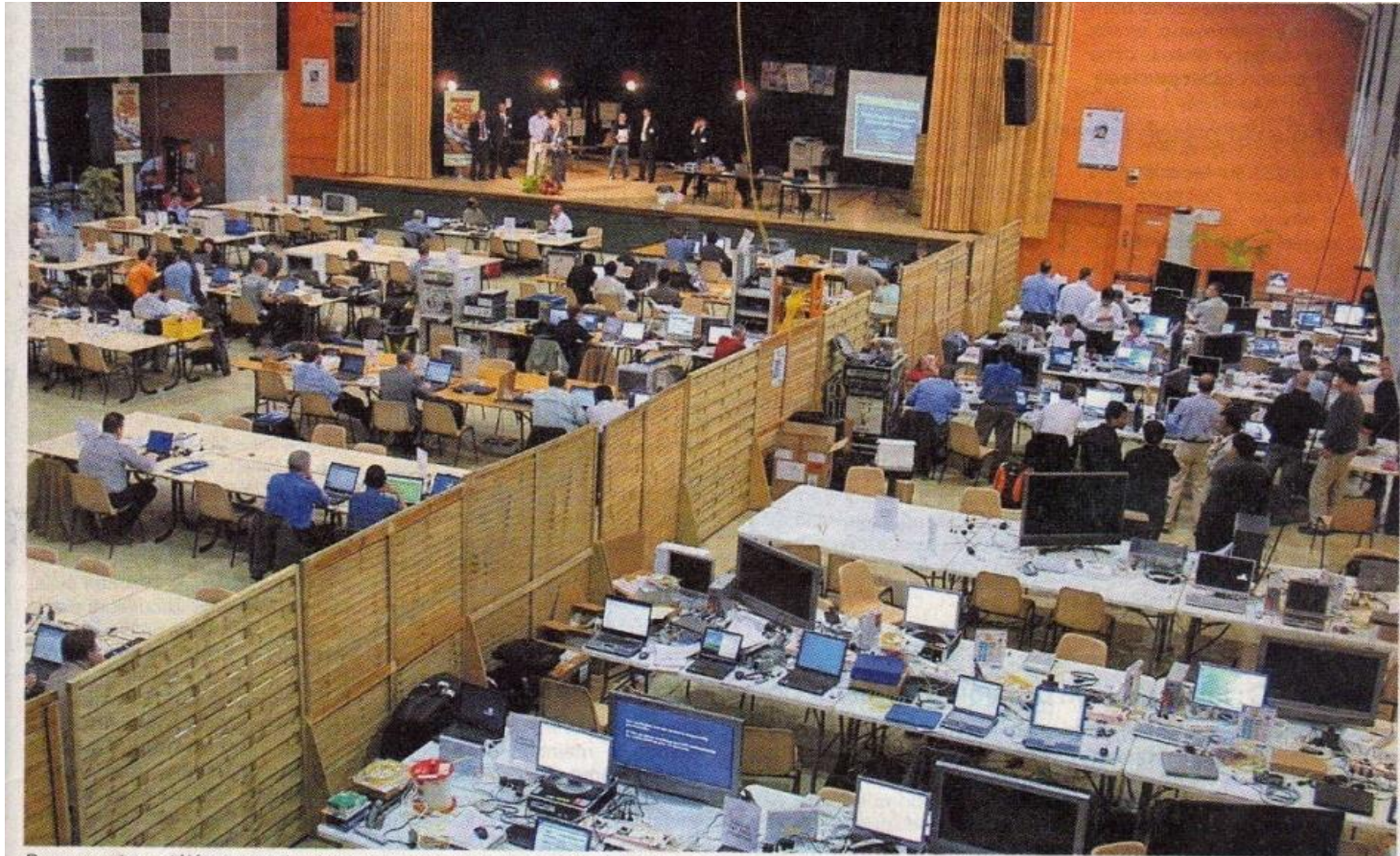


HORIZON 2020 Societal challenge	Current ETSI Standardization projects
Health, demographic change and wellbeing	EP eHEALTH, TC HF (Human Factors), TC SAFETY, TC USER (User group), TC EMTEL (Emergency Telecommunications)
Food security, sustainable agriculture, marine and maritime research and the bio-economy	
Secure, clean and efficient energy	TC EE (Environmental Engineering), TC M2M (Machine to Machine Communications), ISG OSG (Open Smart Grid), ISG OEU (Operational energy Efficiency for Users)
Smart, green and integrated transport	TC ITS (Intelligent Transport Systems), TC AERO (Aeronautic), TC SES (Satellite Earth stations and Systems), TC RT (Railway Telecommunications)
Climate action, resource efficiency and raw materials	TC EE (Environmental Engineering)
Inclusive, innovative and secure societies	TC SCP (Smart Card Platform), TC ESI (Electronic Signatures and Infrastructures), TC LI (lawful Interception), ISG QKD (Quantum Key Distribution), ISG ISI (Information Security Indicators), ISG INS (Identity Management for Network Services), TC SAFETY, TC EMTEL (Emergency Telecommunications), SAGE (Security Algorithms Group of Experts)

Work programme at: <http://portal.etsi.org/portal/server.pt/community/home/312>

EP: ETSI Project, TC: Technical Committee, ISG: Industry Specification Group

Plugtests™ can look like this...



Des experts en télécommunications venus de toute la planète testent entre eux aux Ursulines les produits qui seront demain sur le marché.

... or this (Car2Car Interop)



- All ICT products implement standardised technologies
- Exploitation of research results (not only dissemination) must result in contributions to standardisation
- Research input is needed especially
 - to go beyond the limitations of existing standardised technologies
 - to assess the technical feasibility of alternative technical solutions
- Research input in pre-standardisation phase is crucial to standardise the ‘best’ solution
- ETSI supports funded projects to disseminate and exploit results through standardisation
- ETSI has all tools to meet the standardisation needs of diverse stakeholder groups

Thank you for your attention! Contact us!



ETSI website (<http://www.etsi.org>)

- General public information
- Free standards download

ETSI portal (<http://portal.etsi.org>)

- Easy access to data for each technical body
- Working documents
- ETSI applications and databases

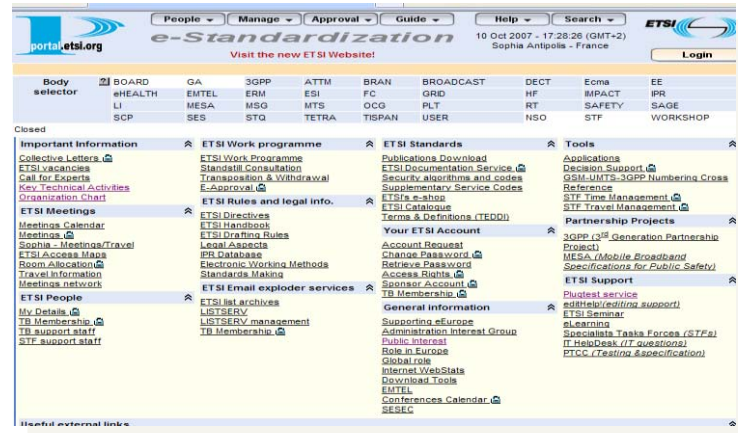
Work Programme (<http://www.etsi.org/images/files/WorkProgramme/etsi-work-programme-2013-2014.pdf>)

hermann.brand@etsi.org

ino@etsi.org

sni@etsi.org

info@etsi.org



PART SEVEN Modelling Techniques and Tools

Analysis of Packet Transmission Processes Governed by Immanent Dependences and Heavy-Tailed Distributions

Natalia M. Markovich

Institute of Control Sciences of Russian Academy of Sciences Moscow, Russia

Abstract: Packet video and voice are the biggest traffic sources in today's Internet. Numerous papers have been written about the influence of packet loss on video and voice quality. This tutorial is motivated by a series of papers [1]-[4]. In these papers, we try to understand causes of packet loss during packets transmission through Internet and how that affects the quality of transmitted video and voice data. To support an adequate teletraffic analysis of the packet traces by the evaluation of relevant performance indices, for instance, of captured packet flows arising from new real-time services in a multimedia Internet, advanced statistical methods are required. They allow us to cope with immanent dependencies and underlying heavy-tailed distributions of interesting features of the traffic such as the bitrates, the inter-arrival times, bit loss and delay distributions of the packet streams, the distribution of path length of the packet, the probability to loss packets or the equivalent bandwidth as a control parameter. In the tutorial we shall discuss useful statistical and probabilistic techniques to handle the arising strongly correlated or long-range dependent time series and heavy-tailed marginal distributions that determine the underlying random variables of the observed data features. New procedures to compute the demanded bandwidth of observed streams or the delay-loss profiles of packet flows during a session will be stated. The inferences are partly based on the last achievements in the theory of extremes in time series. The analysis concepts will be illustrated by real traces arising from some popular Internet applications. The tutorial shall stimulate the participants to incorporate adaptations of the sketched procedures into the practice according to their personal needs. **Key Words:** Packet traffic in Internet, cluster of exceedances over a threshold, quantile, equivalent capacity, quality of transmission, bit loss, missing probability of packet, extremal index, geometric distribution.

References

1. Markovich N.M., Undheim U., Emstad P., Classification of Slice-Based VBR Video Traffic and Estimation of Link Loss by Exceedance. *Computer Networks*. 2009, Vol.53, 1137-1153.
2. Markovich N.: Quality Assessment of the Packet Transport of Peer-to-Peer Video Traffic in High-Speed Networks. *Performance Evaluation*. 70, 28-44 (2013)

3. Markovich N.M.: Analysis of Packet Transmission in Peer-to-Peer Overlay Networks. Preprints of the International IFAC conference on Manufacturing Modelling, Management, and Control, MIM' 2013, Saint-Petersburg, Russia, June 19-21, 956-961 (2013).
4. Markovich N.: Modeling Clusters of Extreme Values. Extremes, Accepted DOI 10.1007/s10687-013-0176-3 (2013)

Analysis of Packet Transmission Processes Governed by Immanent Dependences and Heavy-Tailed Distributions

Natalia Markovich

Institute of Control Sciences
of Russian Academy of Sciences
Moscow, Russia

HET-NET 2013, Ilkley
November 11-13, 2013

Packet video and voice traffic in Internet: extremes in time series

The tutorial is motivated by recent papers:

- Markovich N.M., Undheim A., Emstad P., Classification of Slice-Based VBR Video Traffic and Estimation of Link Loss by Exceedance. *Computer Networks*, Vol.53, 1137-1153 (2009)
- Markovich N.: Quality Assessment of the Packet Transport of Peer-to-Peer Video Traffic in High-Speed Networks. *Performance Evaluation*. 70, 28-44 (2013)
- Markovich N.: Modeling Clusters of Extreme Values. *Extremes*, Accepted (2013) DOI/10.1007/s10687 – 013 – 0176 – 3

Main objectives are to find causes of packet loss during packets transmission through Internet and how that affects the quality of transmitted video and voice data.

- Common subjects and objectives
- M. Perf.Eval.:
Estimating the Quality of Packet Transmission Processes by a Statistical Teletraffic Model
- M., Undheim, Emstad (2009):
Slice-Based VBR Video Traffic Classification
- M. Extremes (2013):
A bit of strong theory and its application
- References

Common subjects and objectives

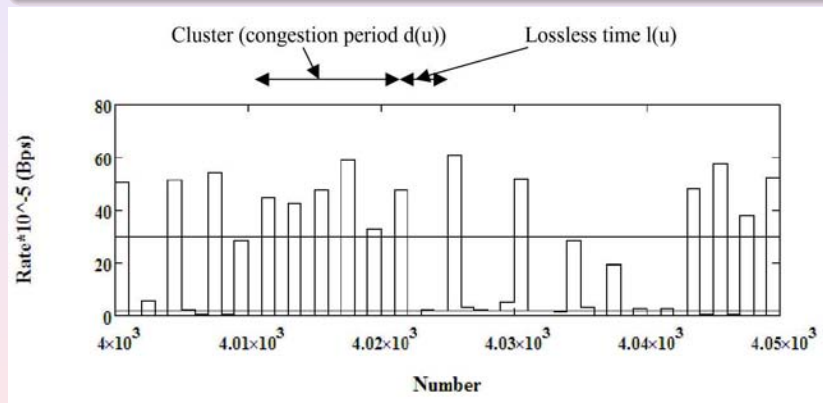
We investigate clusters of exceedances of the underlying process over a threshold and their impact on the quality of packet (frame, chunk) flow transmission

What are these clusters of exceedances and why they are important?

- Clusters of extremal events are conglomerates of exceedances of the process over a threshold.
- Clusters impact the risk for hazardous events like climate catastrophes, huge insurance claims, the loss and delay in telecommunication networks due to overloading
- Definitions of a cluster:
 - a cluster is a block of data with at least one exceedance over a threshold or,
 - clusters are data blocks separated by a fixed number of non-exceedances over the threshold (Beirlant et al. 2004)

Examples of clusters of exceedances

Transmission rates of a P2P SopCast video packet stream



Horizontal lines show the $\approx 82\%$ quantile (upper line) and the median of the rates as thresholds u .

What are causes of clusters?

Clusters are caused

by dependence (long-range dependence) and heavy tails of an underlying process.

Long-range dependence and heavy tails are typical features

of Internet traffic, e.g., of inter-arrival times between packets and the rates of transmission of Skype and IPTV video packets, Markovich & Krieger, ComCom (2010).

Example

A Markov chain generates clusters of exceedances for any threshold u .

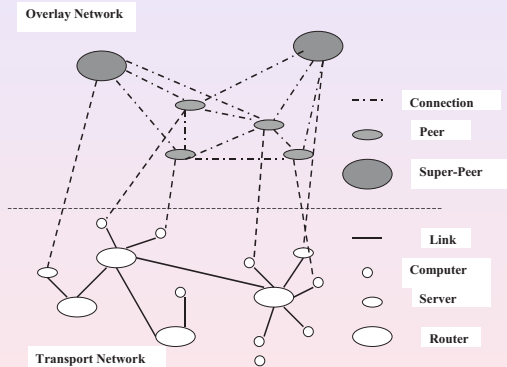
Our approach of cluster investigation is general

and irrespective of technologies, architectures of the networks, schedules, coding etc.

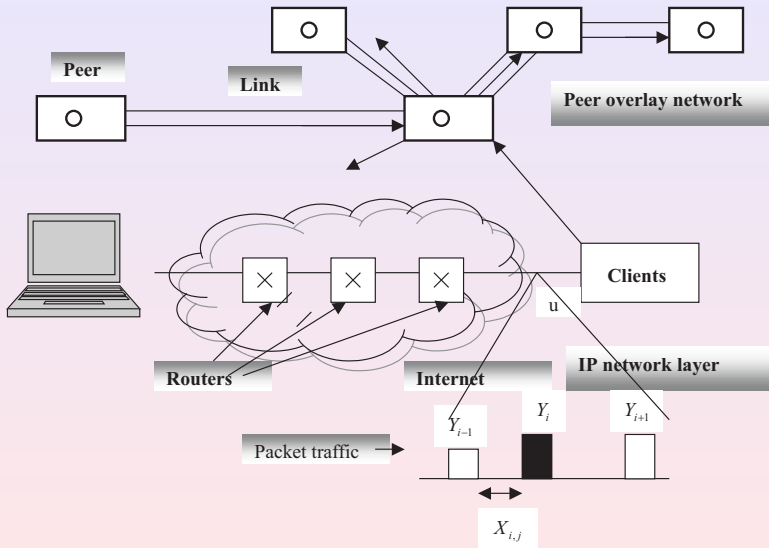
Part 1:

Estimating the Quality of Packet Transmission Processes by a Statistical Teletraffic Model

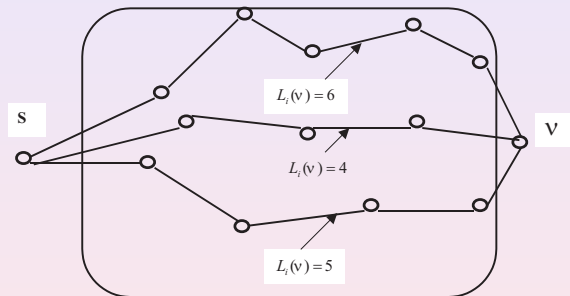
Peer-to-Peer Overlay Network Structure



Packet stream transmission by P2P network



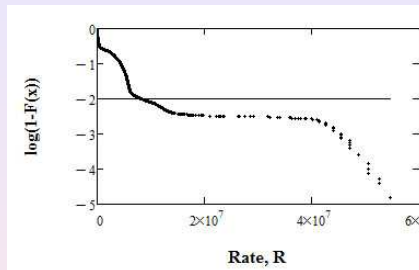
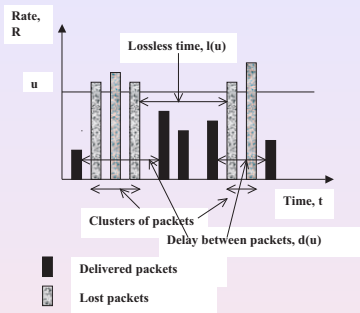
Possible overlay paths of the packets between source and destination nodes



$L_i(\nu)$ is the random length of the overlay path. What is its distribution?

b is the playback delay (it must be \leq playout deadline) at the destination node ν .

Packet traffic



Rates $\{R_i = Y_i/X_i\}$ against the time:

clusters of packets corresponding to the equivalent capacity u , Y_i and X_i are packet lengths (PLs) to the inter-arrival times (IATs), (left); $\log(1 - F(x))$, where $F(x)$ is empirical distribution function of $\{R_i\}$ of a P2P video stream during a SopCast session, and the horizontal level corresponding to 99% quantile (right)

Subjects and Objectives

Subjects

- the statistical traffic characterization,
- the quality control of the packet transmission process arising from advanced streaming applications

Objectives

- The relation between the missing probability P_m of transmitted packets at the receiver and distributions of
 - the length of the overlay path $L(\nu)$,
 - of the upload capacity of the nodes (or equivalent bandwidth) u^* ,
 - of the playback delay (i.e. playout deadline at destination – generation of a packet at the source) at a destination node b in the overlay networks;
- We aim to propose the equivalent capacity u^* that minimize P_m for given playback delay b

Related work

Dán and Fodor, 2009:

- The probability that the i th packet delay $D_i(\nu, L_i) = \sum_{h=1}^{L_i(\nu)} X_h(\nu)$ exceeds the playback delay b at the destination node is considered as the missing probability of the packet $P_m = \mathbf{P}\{D_i(\nu, L_i) > b\}$
- $X_h(\nu)$ are the per-hop (one-way) delays of the overlay path from the source node s to the destination node ν

Embrechts, Klüppelberg and Mikosch, 1997:

If $L_i(\nu)$ and $\{X_h(\nu)\}$ are mutually independent, delays $\{X_h(\nu)\}_{h \geq 1}$ are heavy-tailed subexponentially distributed and independent, the lengths of the overlay paths $\{L_i(\nu)\}$ are light-tailed, then

$$\mathbf{P}\{D_i(\nu, L_i) > b\} \sim \mathbf{E}(L_i(\nu))\mathbf{P}\{X_h(\nu) > b\}$$

Related work. Cont.

Massoulié et al., 2007, Lemma 1:

The relation between

- the playback delay T ,
- the expected time for a packet to travel k ,
- the time D when the node starts playing and
- an expected number of skipped packets α

is obtained

$$k(1 - 1/k)^D + T(1 - 1/k)^{D+k} \leq \alpha.$$

This implies that α increases linearly with increasing T for given D and $k \geq 1$.

Assumptions: IATs between packets at the source and packet transfer times are iid r.v.s and follow exponential distributions; all nodes have the same capacity.

Estimating the Quality of Packet Transmission

Our approach: There are two sources of packet loss:

- the lack of upload (equivalent) capacity of nodes,
- the limited playback delay assigned at the destination node.

Other sources of the packet loss, e.g., due to the node departure, are not considered since they anyway cause a packet delay which exceeds the playback delay.

Main statistical assumptions:

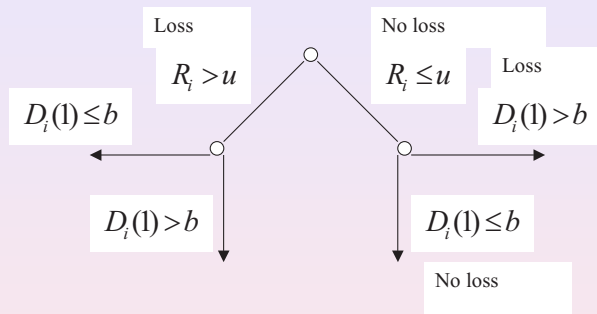
- i each node may have an own capacity;
- ii the overlay network is dynamic;¹
- iii the IATs between packets may be independent or weakly dependent and heavy-tailed distributed with a finite mean;
- iv the packet per-hop delays $D_i(\nu, 1) = X_1(\nu) \triangleq D_i(1)$ of the i th packet at different links are iid;
- v the i th packet delay $D_i(\nu, L_i)$ at a path of the length L_i and its rate $R_{i,j}$ at the j th node are stationary distributed and mutually stochastically independent.

¹These are networks with node arrivals and departures and connectivity changes.

Other assumptions:

- 1 Packet layer (not the chunk layer)
- 2 Bufferless transport channel
 - In real-time scenarios with video and voice streams a bufferless model may be applied since the buffers used in routers along a transport path of the packet-switched network must be short.
 - If a buffer model of a transport channel is used to reduce the loss of information it requires the identification of a service rate and a buffer size.

One-Way Delay

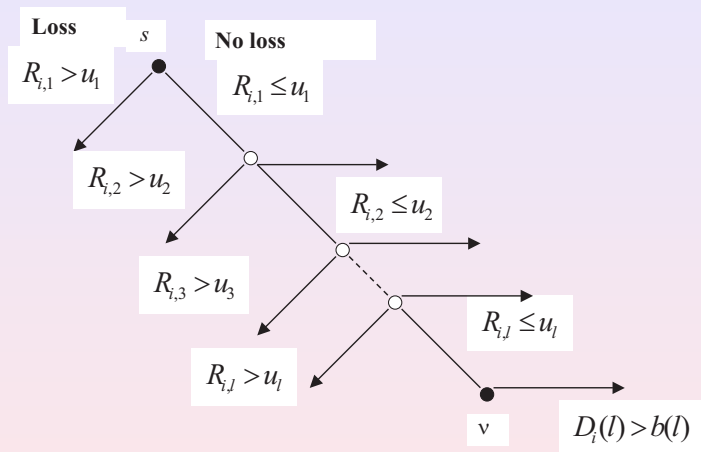


The overlay path contains one link: $L_i(\nu) = 1$

Missing probability:

$$P_m(u, b, \nu) = \mathbf{P}\{L_i(\nu) = 1\} (\mathbf{P}\{R_i > u\} + \mathbf{P}\{R_i \leq u\} \mathbf{P}\{D_i(1) > b\})$$

End-to-End Delay



The overlay path contains more than one link: $L_i(\nu) > 1$
Distributions of the length $L_i(\nu)$ and the missing probability P_m
are given in theorems

The distribution of the length of the overlay path $L_i(\nu)$

Theorem 1: Let the assumption (v) be valid, and

$$\mathbf{P}\{R_{i,j} > u_j\} = \rho_j, \quad \mathbf{P}\{D_i(\nu, l) > b\} = \eta$$

for any $j = 1, 2, \dots, l$ hold. Then it holds

$$\mathbf{P}\{L_i(\nu) = l\} = (\eta + \rho_{l+1}) \prod_{k=1}^l (1 - \rho_k)$$

if each node has an own capacity. If

$$u_1 = \dots = u_l = u, \quad \mathbf{P}\{R_{i,j} > u\} = \rho$$

hold then it follows

$$C_{\rho,\eta} \mathbf{P}\{L_i(\nu) = l\} = \rho(1 - \rho)^{l-1}, \quad \mathbf{E}L_i(\nu) = 1/(\rho C_{\rho,\eta}), \quad (1)$$

where

$$C_{\rho,\eta} = \rho / ((\eta + \rho)(1 - \rho)).$$

Conclusion from (1): the distribution of $L_i(\nu)$ becomes **geometric** after the normalization.

The missing probability of the packet at an overlay path

Theorem 2: Let the assumptions of Theorem 1 hold. Then the probability to miss the packet at the overlay path of the length $l \geq 1$ is equal to

$$P_m(\rho_1, \dots, \rho_l, \eta, l, \nu) = (\eta + \rho_{l+1}) \prod_{k=1}^l (1 - \rho_k) \left(\rho_1 + \sum_{j=2}^l \rho_j \prod_{k=1}^{j-1} (1 - \rho_k) + \eta \prod_{k=1}^l (1 - \rho_k) \right)$$

if each node has an own capacity and to

$$P_m(\rho, \eta, l) = (1 - \rho)^l (\eta + \rho) (1 - (1 - \rho)^l (1 - \eta)) \quad (2)$$

if all nodes have the same capacity.

Concl. from (2): for a sufficiently small ρ the missing probability tends to zero as $\eta \rightarrow 0$ and the corresponding playback delay b increases.

One-Way Delay in Action

We propose to use

- an $(1 - \rho)$ th quantile of the rate as u : $\mathbf{P}\{R_1 > x_\rho\} = \rho$;
- an $(1 - \eta)$ th quantile of the packet delay as b : $\mathbf{P}\{D_i(1) > b\} = \eta$.

Then from Theorem 1 it follows

$$P_m(x_\rho, \eta) = (1 - \rho)(\eta + \rho)(\rho + (1 - \rho)\eta), \quad (3)$$

since $\mathbf{P}\{L_i(\nu) = 1\} = (1 - \rho)(\eta + \rho)$.

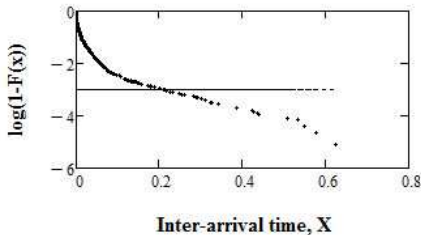
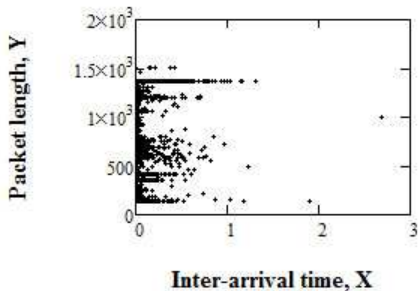
To minimize $P_m(x_\rho, \eta)$ for a given ρ , we take η sufficiently small.

Example

If $\rho = \eta = 0.03$ holds, i.e. u and b are 97% quantiles of R_i and $D_i(1)$, respectively, then by (3) we get $P_m(x_\rho, \eta) = 3.44 \cdot 10^{-3}$.

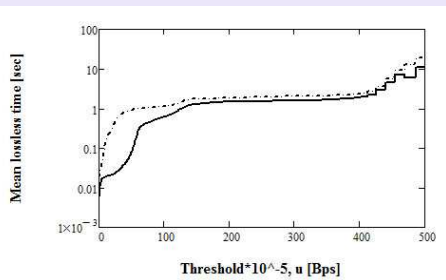
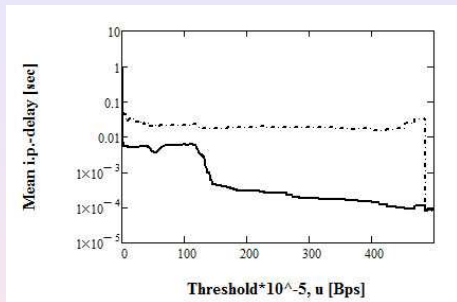
Illustration by a P2P SopCast TV Application

We study a packet trace of the superimposed flows exchanged with an observed peer in the overlay network during a SopCast session.



Scatter plot of the PLs in bytes against the IATs of SopCast video traffic in seconds (left). $\log(1 - F(x))$, where $F(x)$ is empirical distribution function of the IATs, and the horizontal level corresponding to 99.9% quantile (right).

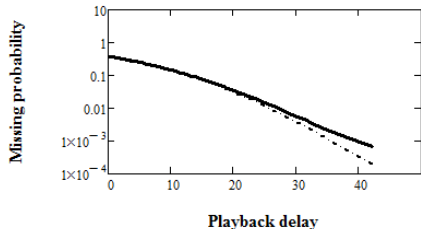
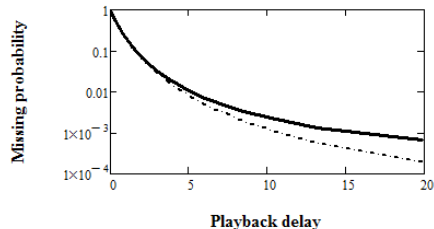
Equivalent Capacity u^* of P2PTV Traffic



The mean delay between successfully delivered packets $\bar{d}(u)$ (left) and the mean lossless time $\bar{l}(u)$ (right) of the SopCast overall traffic (solid line) and pure video traffic without signaling messages (dashed line) against the assigned capacity $u \cdot 10^{-5}$ bps in a logarithmic scale.

u^* is selected as a trade-off between $\bar{d}(u)$ and $\bar{l}(u)$.

Illustration by a P2P SopCast TV Application



Packet missing probability at one link path (dashed line) and multi-link way with $l = 10$ (solid line) calculated by (2) against the playback delay in ms for Pareto and normal distributed one-way delay $D_i(1)$ in logarithmic scale.

Advantages of methodology

- applicability to heavy-tailed, dependent flow data of any streaming application using a bufferless channel;
- applicability to any scheduling algorithm used in the overlay;
- pure nonparametric background;
- the missing probability of the packet transmission is driven by the playback delay and equivalent node capacity;
- the equivalent capacity is much less than the peak rate;
- the possible improved comparison of video encoding/decoding processes using the proposed missing probability;
- the estimation of the playback delay and the equivalent uploading node capacity are important to provide a good streaming quality and the network planning.

Disadvantages

- the playback delay is restricted to be the same for all nodes of the overlay network because of the stability of the control;
- since the playback delay corresponding to heavy-tailed one-way delays may be large, one has to decrease it and get larger missing probability than theoretically possible;
- to estimate the playback delay one needs statistics regarding the one-way (or end-to-end) packet delays that are not always available.

Brief Conclusions

- the packet missing probability as a QoS index is driven by the equivalent capacity, the playback delay and the length of the overlay path;
- the equivalent capacity is driven by bursts of the variable bitrates of packets and is determined as a high quantile of the rate distribution corresponding to the appropriate transmission indices, i.e. the lossless time and the i.p.-delay;
- the playback delay is determined as a quantile of the per-hop (or end-to-end) delay distribution to minimize the missing probability;
- the normalized length of the overlay path is derived to be geometric distributed if all nodes have equal capacities.

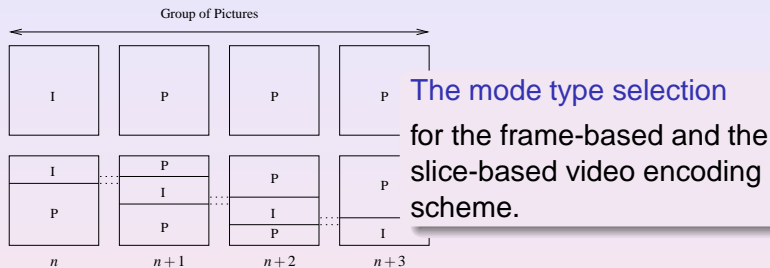
References

- G. Dán and V. Fodor, Delay Asymptotics and Scalability for Peer-to-peer Live Streaming, IEEE Trans. Parallel Distrib., 2009, Vol. 20, 10, p. 1499-1511
- P. Embrechts and C. Klüppelberg and T. Mikosch, Modeling Extremal Events for Finance and Insurance, Springer, Berlin, 1997
- L. Massoulié and A. Twigg and C. Gkantsidis and P. Rodrigues, Randomized decentralized broadcasting algorithms, Proc. of IEEE INFOCOM, p. 1073 - 1081, 2007
- N.M. Markovich, U.R. Krieger, Statistical Analysis and Modeling of Skype VoIP Flows, Computer Communications (2010),

Part 2:

Slice-Based VBR Video Traffic Classification

Description of the slice-based encoded video data



- For the frame-based video encoding the first frame in GoP is intracoded;
- For the slice-based video encoding a slice of each frame in GoP is intracoded.
- The burstiness due to the large intracoded frames is hence reduced.

Description of the model and analyzing data

We consider

- a bufferless fluid model of a communication system;
- non-aggregated flow.

We deal with frame sizes of a single flow

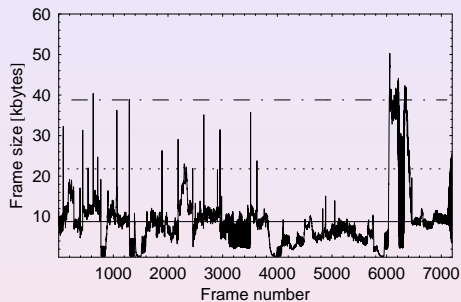
Size of frame = number of packets in the frame \times packet size

- One frame is transmitted every $1/30$ second.
- The packet size is constant apart from the last packet in the frame.
- The frame size is a random variable since the number of packets inside each frame is random.

Our aims are

- to classify the frame sizes to the domains of homogeneity and stationarity;
- to estimate the mean loss due to the exceedances over the capacity of the channel and,
- to estimate high quantiles (e.g., 99%, 99.9%) of the exceedances that determine the quantiles of losses, the amount of losses may exceed these quantiles with a very small probability.

Description of the slice-based encoded video data



The frame-sizes of the slice-based encoded video stream together with the mean frame size $\mu = 8.764$ Kbytes (solid horizontal line) and the 80%, 95% and 99% empirical quantiles.

Classification of the size-frame data

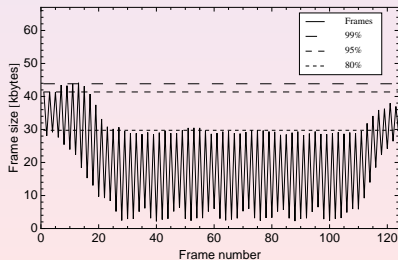
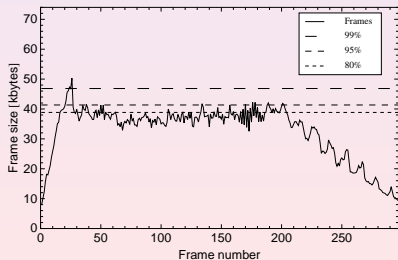
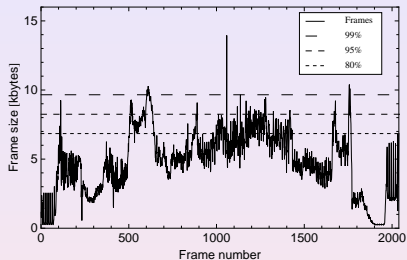
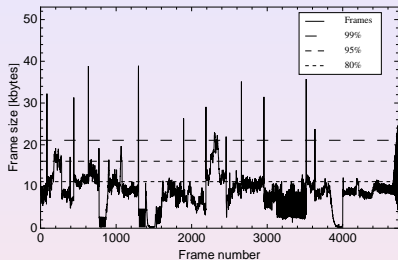
Classification may be done by

- average frame size;

Four classes are determined: for $i \in [0, 4000] \cup [6461, 7198]$,
 $i \in [4001, 6039]$, $i \in [6040, 6210] \cup [6335, 6460]$ and $i \in [6211, 6334]$,
 i is a number of frame.

- scene changes frequency (Undheim, Lin & Emstad 2007);
- the change of marginal distribution;
- the tail index (it shows the heaviness of tails of the class);
- the extremal index (it shows the dependence of the class).

The 80%, 95% and 99% empirical quantiles for four classes separated by average frame size.



Scene change detection

Motivation to divide the video traffic traces into scenes

- to have approximately constant bit rate (Undheim, Lin, Emstad, 2007);
- to have independent blocks (scenes) required for estimation of the extremal index and the mean loss per cluster.

Scene change detection

Let X_j be the j th frame size and the minimum scene length is 12 frames, e.g. one GoP.

Methods of scene detection

- Heyman, Lakshman 1996: the scene change is detected if the inequality is fulfilled for a fixed parameter λ

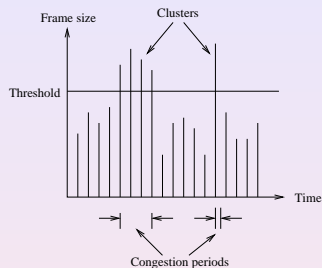
$$\frac{(X_{n+1} - X_n) - (X_n - X_{n-1})}{\left(\frac{1}{6}\right) \sum_{j=n-5}^n X_j} < -\lambda$$

Undheim, Lin, Emstad 2007: $\lambda = 0.4$ based on [visual inspection](#)

- Markovich, Undheim, Emstad 2008: [quantile method of scene detection](#)

The first frame which size exceeds a fixed quantile of frame sizes (we take 80% to have enough data) is detected as a scene change frame.

Estimation of the mean loss in the bufferless model



The cluster exceedance structure showing that the clusters correspond to the congestion periods.

A cluster or a congestion period is defined as

a period in which one or more frame sizes are larger than the threshold.

The mean loss is determined by means of exceedances.

Estimation of the extremal index θ

- Let X_i , $i = 1, 2, \dots$ be a **stationary** process with a marginal distribution function $F(x)$. We have n measurements from this process.
- According to the theory of extremal values, for large n and u_n :

$$\mathbf{P}\{\max(X_1, \dots, X_n) \leq u_n\} \approx F^{n\theta}(u_n),$$

- where $\theta \in [0, 1]$ is the extremal index that reflects the dependence in the sequence.
- For independent, identically distributed sequences $\theta = 1$.
- $1/\theta$ is approximately equal to the **mean cluster size**, i.e. the mean number of exceedances per cluster.
Interpretation: $1/\theta$ means the mean number of lost frames per cluster.

Estimates of the extremal index θ

The blocks estimator: a cluster is defined as a block of the data with at least one exceedance over the threshold u .

$$\bar{\theta}^B(u) = \frac{k^{-1} \sum_{j=1}^k \mathbf{1}(M_{(j-1)r, jr} > u)}{rn^{-1} \sum_{i=1}^n \mathbf{1}(X_i > u)},$$

$M_{i,j} = \max(X_{i+1}, \dots, X_j)$, k is the number of blocks, $r = [n/k]$ is the number of observations in each block.

The runs estimator: a cluster is defined as a block of data with some number of exceedances over the threshold and the following r observations are all below the threshold u .

$$\bar{\theta}^R(u) = \frac{(n-r)^{-1} \sum_{i=1}^{n-r} \mathbf{1}(X_i > u, M_{i+1, i+r} \leq u)}{n^{-1} \sum_{i=1}^n \mathbf{1}(X_i > u)}$$

Interpretation of the extremal index estimators

The $1/\bar{\theta}^B(u)$ ($1/\bar{\theta}^R(u)$) shows the mean number of exceedances in the cluster.

$$\begin{aligned} & \text{The inversion of the blocks (runs) estimator} \\ = & \frac{\text{the number of observations that exceed the thresholds } u}{\text{the number of clusters}} \end{aligned}$$

The overall bit loss over all clusters for a fixed threshold u

$$\begin{aligned} & = \text{cumulative exceedance over } u \text{ of the entire trace} \\ & = \text{the mean excess} \times \text{the mean number of exceedances in the cluster} \\ & \quad \times \text{the number of clusters} \end{aligned}$$

Runs estimate has a better asymptotic bias than the blocks estimate
Smith and Weissman 1994.

Scene blocks estimator of the extremal index θ

New idea

We shall use scenes as blocks.

In this case, we deal with blocks of unequal size.

New scene blocks estimator

$$\bar{\theta}_S^B(u) = \frac{\sum_{j=1}^k \mathbf{1}(M_{\sum_{m=0}^{j-1} r_m, \sum_{m=1}^j r_m} > u)}{\sum_{i=1}^n \mathbf{1}(X_i > u)},$$

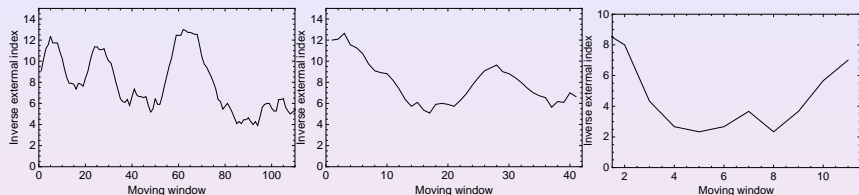
where r_j is the number of frames in j th scene, $\sum_{j=1}^k r_j = n$, $r_0 = 0$ and k is number of scenes.

Classification of dependence of frames by the extremal index

The extremal index θ plays a double role:

- It implies classification of the video traffic according to the dependence of frames in the scenes of each class.
- It detects a non-stationarity of the frame size data $\{X_t\}$, since the point process of exceedances of X_t over u is asymptotically Poissonian with intensity θ .
Then the distribution of the inter-exceedance times is asymptotically exponential with intensity θ .

Classification of dependence of frames by the extremal index



The inverse scene blocks estimate $1/\overline{\theta_S^B}$

is calculated sequentially for moving window containing m scenes for Classes 1-3 (from left to right).

Threshold u equal to the 80% quantiles of frame sizes of the corresponding class. Moving window contains m scenes: $m = 10$ for Classes 1 and 2, and $m = 2$ for Class 3.

Classes 2 and 3 are more stationary than Class 1 with respect to the more homogeneous inter-exceedance times distribution.

Classification by the extremal index

The channel capacity required for each class to satisfy a given bit loss, equal to 3%, is shown.

Table: The 3 % overall bit loss, mean loss per cluster, inter-cluster length, and the corresponding threshold for each class, in Kbytes.

Class	Threshold	Overall bit loss	Mean loss per cluster	Inter-cluster length in number of frames
1	15	1033.8	24.2341	110.14
2	8	109.8	3.3936	63.00
3	37	258.1	21.4438	24.51
4	35	74.68	25.7223	42.41

The occurrence of a few "heavy" clusters or more frequent "light" clusters may influence the QoS and required capacity of channel.

Conclusions

- We classify a video stream by average frame sizes.
- The distributions of the selected classes can only be mixtures of classical heavy- and light-tailed distributions.
- The dependence structure within the classes is variable because of the video encoding.
- A new quantile method for scene change detection is employed. The scenes selected by this method reflects the classification by the average frame sizes.
- The mean bit loss per cluster and the overall bit losses in the bufferless model are estimated.
- High quantiles of bit losses which determine the upper limits of losses for a fixed capacity are evaluated.
- We found the capacity of channel required to give the maximum allowed loss rate for each class.
- This methodology is applied for a test flow of slice-based video data. It can be extended to an aggregated flow.

Reference

- Natalia Markovich, Astrid Undheim and Peder J. Emstad, Slice-Based VBR Video Traffic-Estimation of Link Loss by Exceedance, Proceedings of the 4th International Telecommunication NEtworking WorkShop on QoS in Multiservice IP Networks (QoS-IP 2008), p. 112-117, 2008.
- Astrid Undheim and Yuan Lin and Peder J. Emstad, Characterization of Slice-based H.264/AVC Encoded Video Traffic, Proceedings of the ECUMN 2007, No.4, p. 263-272, 2007.
- S.Mohamed, G.Rubino, A Study of Real-Time Packet Video Quality Using Random Neural Networks, IEEE Transactions on Circuits and Systems for Video Technology 12(2), 2002, p.1071-1083.

Part 3:

A bit of strong theory and its application

Problem

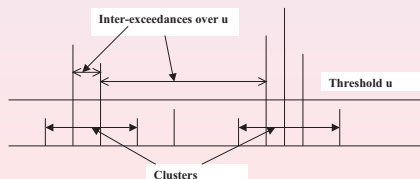
Let $\{R_n : n \geq 1\}$ be a stationary sequence of r.v.s with marginal cdf $F(x)$ and the extremal index θ , $M_n = \max\{R_1, \dots, R_n\}$.

The number of inter-arrival times (IATs) between events arising between two consequent exceedances of the process $\{R_n\}_{n \geq 1}$ over u

$$T_1(u) = \min\{j \geq 1 : M_{1,j} \leq u, R_{j+1} > u \mid R_1 > u\}$$

The number of IATs between two consecutive non-exceedances

$$T_2(u) = \min\{j \geq 1 : L_{1,j} > u, R_{j+1} \leq u \mid R_1 \leq u\}$$



$$M_{1,j} = \max\{R_2, \dots, R_j\},$$
$$M_{1,1} = -\infty$$

$$L_{1,j} = \min\{R_2, \dots, R_j\},$$
$$L_{1,1} = +\infty$$

Main results

We derived limit distributions of

- the inter-cluster size $T_1(u)$ and the cluster size $T_2(u)$

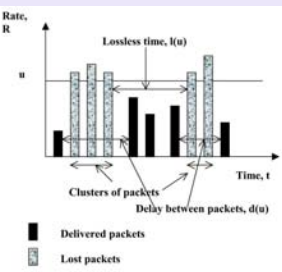
Main question: Selecting the $(1 - \rho)$ th quantile x_ρ of R_t as u may we estimate the means $E(T_1(x_\rho))$ and $E(T_2(x_\rho))$ simply by geometric means $1/\rho$ and $1/q$, $q = 1 - \rho$?

Asymptotically equal distributions of T_1 and T_2 are geometric-like

$$\lim_{n \rightarrow \infty} P\{T_1(x_{\rho_n}) = j\} / (\rho_n (1 - \rho_n)^{(j-1)\theta}) = 1,$$

$$\lim_{n \rightarrow \infty} P\{T_2(x_{\rho_n^*}) = j\} / (q_n^* (1 - q_n^*)^{(j-1)\theta}) = 1.$$

The quality control of a packet flow transmission



Transmission rates $\{R_i = Y_i/X_i\}$,
packet lengths $\{Y_i\}$,
IATs between the packets $\{X_i\}$,
physical capacity u of a channel

- Lossless time of the packet transmission $S_{T_1}(u) = \sum_{i=1}^{T_1(u)} X_i$;
- Delay between successfully transmitted packets $S_{T_2}(u) = \sum_{i=1}^{T_2(u)} X_i$
- Bit loss $B_{T_2}(u) = \sum_{i=1}^{T_2(u)} Y_i$
- Bit number transmitted without loss $B_{T_1}(u) = \sum_{i=1}^{T_1(u)} Y_i$

Skype and SopCast IPTV traffic data

Geometric mean (solid line) and $E(T_1(u))$ fitted theory with $\theta(u) = 0.23$ (dashed line) and by the sample mean $\overline{T_1(u)}$ (thin solid line) obtained by SopCast IPTV data with a sample size equal to $6.553 \cdot 10^4$, against the threshold u (left), and similar curves shown for Skype data with a sample size equal to 4605 and with $\theta(u) = 0.388$ (right)

ITU-R and WINNER II Path Loss Modeling of Femtocells

Harold O. Kpojime, Ghazanfar A. Safdar, Mehmet E. Aydin

Department of Computer Science and Technology,
University of Bedfordshire, Park Square, Luton, LU1 3JU, UK
{harold.kpojime, ghazanfar.safdar, mehmet.aydin}@beds.ac.uk

Abstract. As cellular network users continue to grow; there is a need for increased user capacity, higher throughput and improved system performance. An important issue affecting cellular networks is to make services available to regions of bad or no reception. Various standards and techniques such as, antennas, relays and hotspots are used in achieving this and one of the most reliable is a technique that involves bringing the transmitter and receiver closer to each other. This enhances the performance of the receiver by ensuring a high quality link. In this context, Femtocells are another promising technology to improve indoor coverage; these are short range low-powered base stations and usually installed by home users. However, the statistical analysis retrieved by deploying femtocells on a large scale does not reflect the behavior of a single femtocell. Moreover, the interference mitigation techniques employed take into consideration densely deployed FAPs which is not applicable in a scenario where there are only a handful of femtocells. This paper examines the influence of LTE-femtocells on a small scale in an indoor to indoor environment by comparing two well known path loss models *ITU-R* and *WINNER II*. To achieve this, a site specific scenario is modeled to investigate pathloss, received power, and SINR to show the behavior of a single femtocell as it experiences the abrupt fluctuations under the influence of co-tier interference.

Introduction

Indoor cellular usage which accounts for 50% of all voice calls and 70% of data traffic is mostly faced with poor reception due to low signal penetration through walls and attenuation which may lead to total loss of signal [1]. The use of Macrocell Base Stations (MBS) which motivates existing wireless strategies has proven to be inefficient to tackle the indoor coverage problem as the distance between transmitter and receiver has to be kept relatively close to enhance a high quality link. Femtocells were introduced as a solution to poor indoor network coverage. Femtocells are home base stations which connect standard mobile devices to the network operator using residential Digital Subscriber Line (DSL), optical fibres, cable broadband or wireless last-mile technologies [2], [3]. However, due to the two-tier architecture of femtocells, interference is imminent. The cell sites covered by a number of Femtocells Access Points (FAPs), in some cases overlapping FAPs, is overlaid in the larger cell site of

the MBS, resulting in co-tier/cross interference. The close proximity greatly lowers transmission power and increases the battery life of mobile devices and with a dedicated FAP it offers subscribers a single billing address for mobile phone, broadband and land line as they are all channeled through the same backhaul [2], [4]. Additionally, it acts as a solution to the convergence of landline and mobile systems [5].

Femtocells can be deployed in an urban or suburban scenario where FAPs are randomly located in a MBS with low or high density respectively with emphasis on the general performance and analysis [3]. To mitigate the interference in femtocells, measurement based approaches have been incorporated in literature by measuring scenarios in real life or creating models and mapping out parameters such as the path loss, received power and SINR. These parameters are subsequently used to analyze the performance and implement interference mitigation techniques. ITU-R [6] and WINNERII [7], which is incorporated in this paper are some of the path loss models and can be used to calculate the mentioned parameters in Long Term Evolution (LTE) femtocells between a FAP and Femtocell User Equipment (FUE).

A femtocell model is developed in [8] by randomly generating realistic floor plans, designs and analyzing the signal penetration through walls, doors and windows. The analysis of co-tier interference in femtocells is presented in [9] where a semi analytical approach is implemented on a femtocell model and the interference between FAPs (co-tier Interference) is mitigated using a path loss model, shadowing, wall penetration loss, location of femtocells and user distribution parameters. However, the statistical analysis retrieved by deploying femtocells on a large scale does not reflect the behavior of a single femtocell. Moreover, the interference mitigation techniques employed in most of the literature take into consideration densely deployed FAPs which is not applicable in a scenario where there are only a handful of femtocells. Thus the main motivation of this research work is to examine the influence of LTE-femtocells on a small scale in an indoor to indoor environment by comparing two well known path loss models ITU-R and WINNERII

The paper investigates the path loss and received power by FUEs for an indoor site specific femtocell scenario using the ITU-R and WINNER II models. The rest of the paper is organised as follows. Section II briefly describes ITU-R and WINNER II models before Section III presents scenario details and simulation parameters. Section IV looks into results and analysis whereas paper is finally concluded in section V with some details about future work.

ITU-R and WINNER II Pathloss Models

Path loss propagation models can be either empirical models or semi-empirical [2]. While empirical models are based on statistics and site measurements, semi-empirical models employ theory and real case measurements and also introduce deterministic parts which can be added to a model. This section briefly introduces the ITU-R and WINNER II path loss models.

2.1 ITU-R

The International Telecommunication Union Radio communication (ITU-R) is a semi empirical path loss model that can be employed in different scenarios comprising indoor offices, vehicular and pedestrian. The applicable operational frequency is 900MHz – 6GHz and the equations for the line of sight (LOS) and non-LOS (NLOS) path loss P_l are;

$$P_{l(LOS)} = 18 \log_{10}(f_c) + N \log_{10}(d) - 28 \quad (1)$$

$$P_{l(NLOS)} = 20 \log_{10}(f_c) + N \log_{10}(d) + Lf(n_w) - 28 \quad (2)$$

Where;

f_c = 2000 MHz, carrier frequency

N = 28 (Power Loss Coefficient)

d , Tx-Rx distance in meters (Table II)

Lf = 4 dB (wall penetration loss)

n_w , the number of walls

2.2 WINNER II

The Wireless World Initiative New Radio (WINNER) which developed the WINNER II path loss model is a consortium of partners with the goal of enhancing mobile communication performance. The WINNER II path loss is also semi empirical and allows for a flexible modeling of propagation channels in a wide range of scenarios including indoor offices, indoor-outdoor and urban micro-cell. The indoor path loss incorporated in this paper is the WINNER II A1 LOS and NLOS, specified in [7] as given below;

$$P_{l(LOS)} = 46.8 + 18.7 \log_{10}(d) + 20 \log_{10}\left(\frac{f_c}{5}\right) \quad (3)$$

$$P_{l(NLOS)} = 46.4 + 20 \log_{10}(d) + 20 \log_{10}\left(\frac{f_c}{5}\right) + L_w \quad (4)$$

Where;

f_c = 2000 MHz

d , Tx-Rx distance in meters (Table II)

L_w , the wall penetration losses which is $5(n_w - 1)$ for light walls

Simulation parameters

The femtocell scenario considered in this work is illustrated in Figure 1. It comprises an apartment block with four flats (1-4) each with an FAP and few FUEs. The position of the FAPs and FUEs is assumed as shown and not based on the probability of being in a particular position. The femtocells are circular shaped with a fixed transmission power and operates in the Closed Subscriber Group (CSG) access mechanism.

Table 1. Simulation parameters

<i>Parameter</i>	<i>Value</i>
Number of FAPS/Size(m)	4/10x10
Number of FUES	9
Frequency	2 GHz
Carrier Bandwidth	180 KHz
Tx Power	20 dBm
Cell Layout	Circular Cell
Thermal Noise Level	-121.42 dBm / Hz

The femtocells are assumed to be coexisting with the MBS therefore sharing resources such as centre frequency as defined by the 3rd Generation partnership project (3GPP) LTE standard. The interference due on each FAP is the over-lapping signal strength of neighbouring FAPs. Cross tier interference which is the interference between a femtocell and the MBS and Macrocell User Equipment (MUE) is not considered in this work. The propagation channel of an indoor scenario is more complex compared to an outdoor scenario due to the inherent obstacles with dimensions equivalent to the propagated wavelength. The attenuation is also increased when floors are considered adding to its complexity. For sake of simplicity, multipath fading is ignored as well as multiple floors and mobility of FUEs. Table I shows the simulation parameters whereas the actual distance in meters between the FAPs and FUEs is highlighted in Table II.

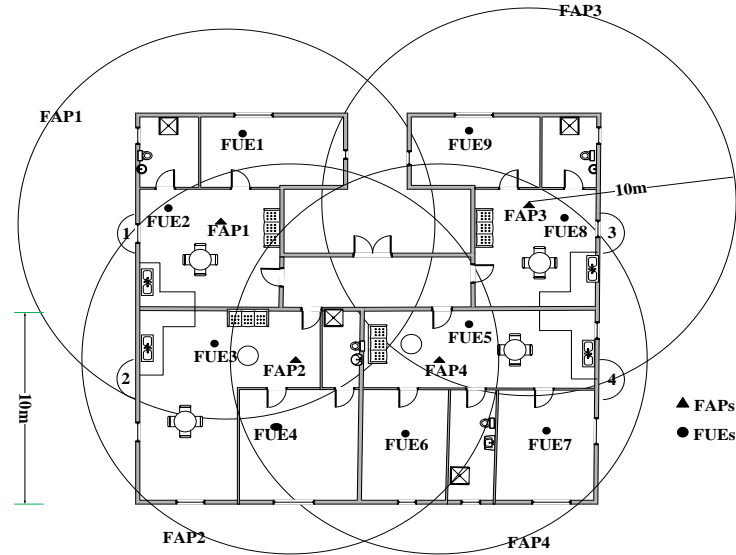


Fig. 1. Femtocell Scenario

Table 2. Distance between FAPs and FUEs in metres (m)

<i>FUE</i>	<i>FAP1</i>	<i>FAP2</i>	<i>FAP3</i>	<i>FAP4</i>	<i>Distance between FAPS</i>		
1	4.3	11.6	13.6	14.6			
2	2.3	9.6	16.5	14.7		<i>FAP2</i>	<i>FAP3</i>
3	6.3	3.7	16.0	10.4			<i>FAP4</i>
4	10.8	3.8	16.4	8.7	<i>FAP1</i>	7.9	14.2
5	12.7	8.1	6.3	2.0			
6	13.9	6.3	12.7	4.1	<i>FAP2</i>		13.2
7	18.7	12.2	11.4	6.2			
8	16.0	14.4	1.8	9.1	<i>FAP3</i>		
9	12.3	14.1	4.5	11.5			8.6

Table 3. Path loss calculation

<i>FUE</i>	<i>Nw (No. of walls)</i>		<i>ITU-R (dB)</i>		<i>WINNER II (dB)</i>	
	<i>w.r.t FAP1</i>	<i>w.r.t FAP2</i>	<i>w.r.t FAP1</i>	<i>w.r.t FAP2</i>	<i>w.r.t FAP1</i>	<i>w.r.t FAP2</i>
1	1	2	59.76	75.83	56.11	69.73
2	0	1	48.15	69.52	45.67	63.08
3	1	0	64.4	53.93	59.42	49.80
4	2	1	74.96	58.25	69.10	55.03
5	3	2	80.93	71.46	75.51	66.61
6	5	3	90.03	72.40	86.30	69.42
7	5	5	93.63	88.44	88.87	85.16
8	2	4	79.74	86.45	72.52	81.42
9	4	4	84.54	86.20	80.23	84.99

Table 4. Received Power calculations

<i>FUE</i>	<i>Nw (No. of walls)</i>		<i>ITU-R (dBm)</i>		<i>WINNER II (dBm)</i>	
	<i>w.r.t FAP1</i>	<i>w.r.t FAP2</i>	<i>w.r.t FAP1</i>	<i>w.r.t FAP2</i>	<i>w.r.t FAP1</i>	<i>w.r.t FAP2</i>
1	1	2	-39.76	-55.83	-36.11	-49.73
2	0	1	-28.15	-49.52	-25.67	-43.08
3	1	0	-44.40	-33.93	-39.42	-29.80
4	2	1	-54.95	-38.25	-49.10	-35.03
5	3	2	-60.92	-51.46	-55.51	-46.61
6	5	3	-70.03	-52.40	-66.30	-49.42
7	5	5	-73.63	-68.44	-68.87	-65.16
8	2	4	-59.74	-66.45	-52.52	-61.60
9	4	4	-64.54	-66.20	-60.23	-61.42

Results and Analysis

Using ITU-R, WINNER II and distance values from Table II, the results for NLOS path loss (Table III) and received power (Table IV) are obtained. Even though our considered scenario (Figure 1) includes four FAPs, due to space constraint only the results for FAP1 and FAP2 are included in this work.

4.1 Path Loss

Individual FUEs experience different path loss due to their varying distances from FAPs (Table II, shows distances of FUEs from different FAPs). Clearly, the path loss will be higher for an FUE at a distance further away from an FAP compared to path loss value for an FUE which is situated closer to the FAP. Figure 2 and Figure 3 plots path loss for all FUEs (1-9) for FAP1 and FAP2 respectively. When compared to ITU-R, WINNER II averaged over all FUEs results into slightly lower (5 dB less), path loss due to differences in the power loss coefficient and wall penetration loss components.

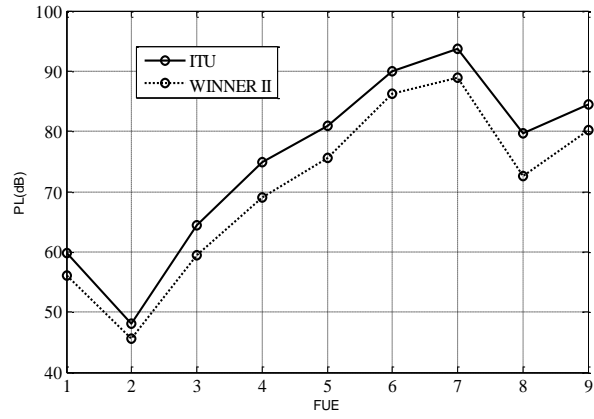


Fig. 2. Path loss w.r.t FAP1

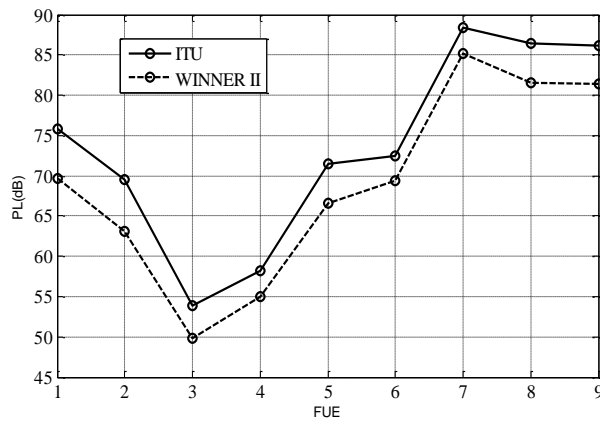


Fig. 3. Path loss w.r.t FAP2

4.2 Received Power

Equation 5 shows received power calculations by employing the values of Tx power and the path loss P_l . The results for received power are in accordance with the path loss values where FUEs which suffered higher losses due to increased distance and wall penetration has lower received power compared to other FUEs in the chosen femtocell scenario. Figure 4 and 5 shows received power plots for all FUEs for the same FAPs (FAP1 and FAP2 respectively).

$$P_r = 20 - P_l \quad (5)$$

Since ITU-R has slightly higher path loss compared to WINNER II, accordingly it resulted into lesser received power values for ITU-R (on average 5 dBm lesser) compared to WINNER II. This further validates the results of path loss and received power both for ITU-R and WINNER II.

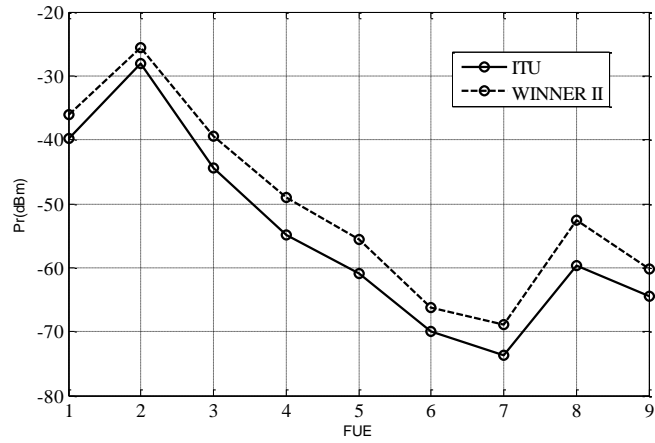


Fig. 4. Received Power w.r.t FAP1

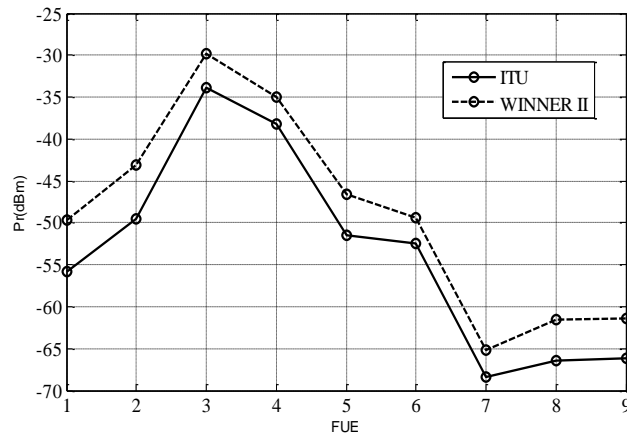


Fig. 5. Received Power w.r.t FAP2

4.3 SINR

In our work, the sources of downlink interference on each FUE are considered to be the received power due to neighbouring FAPs. Thus the signal to Interference plus Noise Ratio (SINR) can be calculated by using Equation 6.

$$\text{SINR} = \frac{P_r}{\sum_i I_{s_i} + N_t} \quad (6)$$

- Where P_r is the received power of a particular FUE (i.e. received power from serving FAP)
- $\sum_i I_{s_i}$ is the summation of the interference power at particular FUE (i.e. received power from all other FAPs)
- N_t is the Thermal Noise value at 180 KHz [10]

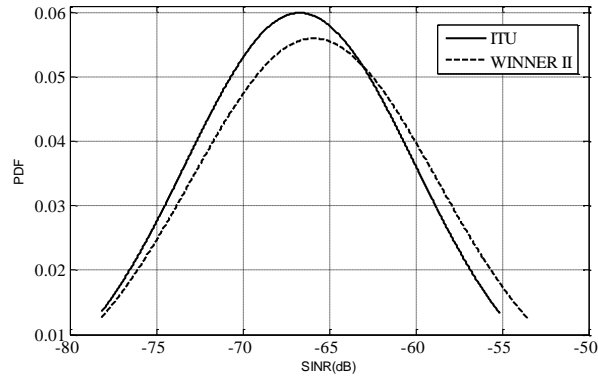


Fig. 6. SINR (PDF plot)

The calculated SINR values are subsequently used to compute the statistics such as probability density function (PDF) and cumulative distribution function (CDF). In accordance with the path loss and received power results, WINNER II while compared to ITU-R, performs better in terms of SINR value both for PDF and CDF plots as shown in Figure 6 and 7 respectively.

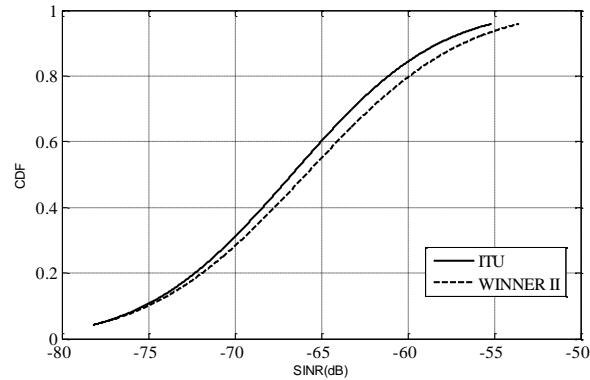


Fig. 7. SINR (CDF plot)

Conclusion

This paper examines the influence of LTE-femtocells on a small scale in an indoor to indoor environment by comparing two well known path loss models ITU-R and WINNER II. A site specific scenario is modeled to investigate pathloss, received power, and SINR to show the behavior of a single femtocell as it encounters the abrupt fluctuations under the influence of co-tier interference. The results have shown that WINNER II performs better over ITU-R for all the investigated parameters due to differences in the power loss coefficient and wall penetration loss components. Future work will introduce indoor – outdoor propagation at different frequencies to introduce more interference sources such as the MBS and Macrocell user equipments (MUEs). Accordingly some conventional and subsequently cognitive radio enabled interference mitigation schemes will be employed to see how they affect to achieve better performance in the presence of increased interference.

References

1. V. Chandrasekhar and J. Andrews and A. Gatherer, "Femtocell networks: a survey," *Communications Magazine, IEEE*, vol.46, no.9, pp.59-67, September 2008
2. Zhang, J. and Roche, G. (2010) 'Femtocells: Technologies and Deployment' Hoboken, NJ: Wiley
3. Andrews, J.G.; Claussen, H.; Dohler, M.; Rangan, S.; Reed, M.C. , "Femtocells: Past, Present, and Future," *IEEE Journal on Selected Areas in Communications*, vol.30, no. no.3, pp.497,508, April 2012
4. H.A. Mahmoud and I. Guvenc, "A comparative study of different deployment modes for femtocell networks," *IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp.1-5, 13-16 Sept. 2009

5. N. Saquib, E. Hossain, Long Bao Le and Dong In Kim, "Interference management in OFDMA femtocell networks: issues and approaches," *Wireless Communications, IEEE*, vol. 19, no. 3, pp. 86-95, 2012
6. "Guidelines for evaluation of radio transmission technologies for IMT-2000", Rec. ITU-R M.1225
7. IST-WINNER II Deliverable D1.1.1 V1.1, WINNER II interim channel models, September 2007
8. M. Mirahmadi, A. Shami and A. Al-Dweik, "A building architecture model for predicting femtocell interference in next-generation networks," *Communications (ICC), 2012 IEEE International Conference on*, pp. 5059-5063.
9. E. Pateromichelakis, M. Shariat, A. ul Quddus and R. Tafazolli, "On the analysis of co-tier interference in femtocells," *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, pp. 122-126.
10. M. Barkat, *Signal Detection and Estimation*. 2nd ed. Norwood: Artech House, 2005, pp. 205-207

Transient Analysis of Multi-Priority Queues $M_r | G_r | 1$ with Disasters

Kh. Kerobyan^{+,*}, R. Badalyan^{*}, R. Kerobyan⁺

⁺Armenian Society of Computer and Information Systems, Armenia

^{*}Los Angeles City College, USA

khankerobyan@ymail.com; rbadalian@gmail.com

Abstract. In this paper, the model of single server priority queues $M_r | G_r | 1$ with r classes of customers and disasters has been considered. Transient and steady-state joint distribution of queues length and Laplace Stieltjes Transformation of busy period for non-preemptive and preemptive priority disciplines have been analyzed using the supplementary variables method.

1. Introduction

It is well known that Next Generation Network and Future Network are based on integration of multiple types of networks and networking technologies, traffic and services, heterogeneous resources and applications and are aimed to create unite, World wide information infrastructure which provides guaranteed high QoS, ubiquitous, unlimited, effective and reliable access to its numerous customers and services. The ways to solve these ambitious and challenging problems are closely related to the networks' modeling, to analysis, to prediction of performance measures of their different components, clusters and entire network, as well as to the research and design of new effective, robust, algorithms and mechanisms that are tolerant to networks' treats. In this article, we consider network server queue model with the priority discipline of service for heterogeneous customers. The "negative arrivals" models, called also G-queues, were introduced by Gelenbe [1], have been used to describe an impact of different types of network threats on the server performance measures. We consider two types of customers: "regular" customers have following behavior - they form a queue and serve according to some service discipline; "negative arrivals" have following behavior - they do not form a queue, they can destroy a single customer, a group of customers or all of them. The last type of "negative arrivals" is called disaster or catastrophe. We suppose that after disasters the model immediately becomes free of regular customers and continues its work from that state.

Queuing models with catastrophe have been studied intensively for a broad variety of service disciplines of regular customers, e.g. first-come-first-served (FCFS), preemptive last-come-first-served (LCFS), processor sharing (PS), queue with retrials, queue with multiple steps of service [1-9,16,17] etc. The first model of G-queue with catastrophes was introduced by Chao [2]. An exponential network of queues with catastrophes is considered under the following strategy of removal: when a catastrophe occurs at one service station, all the regular customers in this station are destroyed immediately, and then the server is sent for the repair. It is shown that the steady-state probability of the queues' sizes has a product form. Jain and Sigman [3] have considered single-server $M | G | 1$ queues with disasters for FIFO and LIFO service disciplines for regular customers. Using the level crossing arguments they obtained a generalization of the Pollaczek-Khintchine formula for workload. Another single server $M | G | 1$ queue model with "negative" arrivals and a special mechanism of removal is examined by Boucherie and Boxma [4]. In this article, it is supposed that "negative" arrivals can destroy random amount of work. By using Wiener-Hopf factorization method and level crossing arguments, the models workload distribution has been investigated. The transient and steady-state solutions for queue length measures, distributions of the busy period and waiting time are obtained by applying the supplementary variables method. The $M | G | 1$ model with PS service discipline for regular customers and disasters is considered by Li and Lin [5]. Using supplementary variable method, the transient and steady-state distributions of queue length, and Laplace-Stieltjes Transformation (LST) for

waiting time and busy period are investigated. For a comprehensive survey of queue with “negative arrivals” and its applications, the reader may refer to Gelenbe [1], Artalejo [6], Bocharov [7], and Do [8].

It is well known that priority mechanisms are effective scheduling tools which allow many concurrent classes of customers to receive a different QoS by allocating limited system resources among them. Priority queues have wide application in communication and computer systems and networks, manufacturing systems, etc. For the priority queues see, for example, [9,10] and the references.

Priority queues have been an interesting subject of intensive investigation since 1960. Many articles have been published on these queues and their applications, e.g. Jaiswal [9], Kleinrock [10], Gnedenko and all [11], Klimov [12], Takagi [13], Matveev and Ushakov [14].

In priority queues, with regard to occupation of the server, two major types of priority disciplines are considered: pre-emptive, when the service of lower priority customer can be interrupted by the high level priority customers, and non-pre-emptive, or head-of-the-line, priority when the service of lower priority customers cannot be interrupted by any priority customers. For the pre-emptive priority discipline, there are many different policies in regard to the customer whose service was interrupted: a) pre-emptive resume – the service of the interrupted customer continues from the interrupted level; b) pre-emptive repeat-service – the service of the interrupted customer is repeated and c) preemptive abort-priority (or pre-emptive loss-priority) – the interrupted customer is lost; for the details see Jaiswal [9], Klimov [12].

The first model of single-server priority queues with disasters was considered by Towsley and Tripathi [15]. They analyzed a preemptive resume priority queuing model $M_2^* | M_2 | 1$ with two classes of customers and an unreliable server. The failures of the server have a disaster effect, i.e. when the server’s failure occurs, it flushes out (removes) all the customers in the system. The model is analyzed under the following assumptions: Poisson arrivals of disaster and correlated groups of priority customers, exponential distribution of each priority class customer service time, and general distribution of server repair time. By means of Semi-Markov Process (SMP) methods, expressions for the Probability Generating Function (PGF) of high and low priority queue lengths, the throughput of system for each priority classes, the distribution of the number of customers which were removed due to the disaster arrival are obtained.

To analyze the single server $M_r | G_r | 1$ priority queues with r classes pre-emptive priority customers and disasters, Kerobyan [16] offered to use the basic queuing model $M | G | 1$ with disasters and an unreliable server. Two types of preemptive priority disciplines are considered: pre-emptive repeat-service, and pre-emptive loss-priority. Using the supplementary variable method the transitive PGF of queue length of each priority classes and the distributions of the model busy period and regeneration period are obtained.

The steady-state behavior of the $M_r | G_r | 1$ model with r priority classes of preemptive and non preemptive customers and disasters has been analyzed by Kerobyan [17]. Two types of preemptive priority disciplines are considered: pre-emptive repeat-service, and pre-emptive loss-priority. To obtain the steady-state joint distribution of queues length of different priority classes and the LST of busy period, the supplementary variable and supplementary event methods are used.

The rest of this paper is organized as follows. In Section 2, we describe the base model of a single server priority queue and notations’ usage. In Section 3, we derive the probability distribution functions (PDF) of a busy period and regeneration period, the emptiness probability of the $M | G | 1$ model with disasters. In Section 4, we derive the transient and steady-state distributions of the queue length for the non-preemptive and preemptive priority models $M_r | G_r | 1$ with disasters.

2. Description of the model

The model $M_r | G_r | 1 | \infty$ with independent Poisson arrivals of r classes of customers and disasters with arrival rates a_1, a_2, \dots, a_r and ν , respectively, is considered. The service of customers of various classes is carried out in accordance with priorities. Further, we assume that the customers of i^{th} class (i^{th} priority

customers) have a higher priority than the customers of j^{th} class, if $i < j$, $i, j = \overline{1, r}$. Within one priority class the customers get served by one according to the discipline FCFS. The service time of i^{th} priority customers is mutually independent, identically distributed (i.i.d.) random variables B_i (r.v.), with distribution function (DF) $B_i(x) = P(B_i < x)$, and density $b_i(x)$, $i = \overline{1, r}$, and has finite first two moments $\beta_{1i} < \infty$, $\beta_{2i} < \infty$. The disasters are not stored and served. We assume that when the disaster arrives to a busy model, it instantaneously destroys all the customers in the model, including the customer in the service. Otherwise, if the model is empty at the moment of disaster's arrival, it disappears without any influence on the model. Then the model continues its work from an empty state.

Basic notations

Let's introduce following notations and definitions: $\mathbf{L}(t) = (L_1(t), L_2(t), \dots, L_r(t))$ - is a vector which describes numbers of customers in the model at the time t , where $L_i(t)$ - is the number of i^{th} priority customers in the model at the time t , $L_i(t) = \overline{0, \infty}$, $i = \overline{1, r}$. Set:

$$\sigma = a_1 + a_2 + \dots + a_r, \quad (\mathbf{a}, \mathbf{z}) = a_1 z_1 + a_2 z_2 + \dots + a_r z_r, \quad \mathbf{z} = (z_1, z_2, \dots, z_r)$$

$$(\mathbf{a}, \mathbf{z})^i = a_{i+1} z_{i+1} + a_{i+2} z_{i+2} + \dots + a_r z_r, \quad i = \overline{0, r-1}, \quad (\mathbf{a}, \mathbf{z})_i = a_1 z_1 + a_2 z_2 + \dots + a_i z_i, \quad i = \overline{1, r}.$$

$P(\mathbf{n}, t) = P(\mathbf{L}(t) = \mathbf{n}) = P(L_1(t) = n_1, L_2(t) = n_2, \dots, L_r(t) = n_r)$ - is the probability that at the time t there are n_1 customers of first priority, n_2 customers of second priority, and n_r customers of r priority, respectively, in the model where $\mathbf{n} = (n_1, n_2, \dots, n_r)$.

Define the PGF of $P(\mathbf{n}, t)$

$$P(\mathbf{z}, t) = \sum_{\mathbf{n}=0}^{\infty} \mathbf{z}^{\mathbf{n}} P(\mathbf{n}, t) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \dots \sum_{n_r=0}^{\infty} z_1^{n_1} z_2^{n_2} \dots z_r^{n_r} P(n_1, n_2, \dots, n_r, t), \quad \eta_i(x) = \frac{b_i(x)}{1 - B_i(x)}, \quad B_i(x) = 1 - e^{-\int_0^x \eta_i(t) dt},$$

$\eta_i(x)$ - is the failure rate (hazard) function of a service time PDF $B_i(x)$ of i^{th} priority customers.

We assume that all random variables defined above are mutually independent.

For an $F(x)$ function, we write $\bar{F}(x) = 1 - F(x)$, and denote its LST as $\tilde{F}(s) = \int_0^{\infty} e^{-sx} dF(x)$.

3. Busy period of the model

Let α_v , α , and χ are i.i.d. random variables which describe the length of busy period of the model with disasters, the length of busy period of the standard model without disasters, and the length of time between two consecutive arrivals of disasters, respectively. Let also $\pi_v(t)$, $\pi_v(t) = P(\alpha_v \leq t)$ be a PDF of the busy period of a model with disasters, and $\tilde{\pi}_v(s)$ and $\bar{\pi}_v$ are its LST and mean value respectively. Also, $\tilde{\pi}(s)$ and $\bar{\pi}$ are the LST of the PDF of the busy period of the standard model without disaster and its mean value, respectively. It is easy to see that for a r.v. α_v the following stochastic equality takes place: $\alpha_v = \min(\alpha, \chi)$. Using this equality for the PDF of a busy period $\pi_v(t)$ and its LST $\tilde{\pi}_v(s)$ we have

$$\pi_v(t) = P(\min(\alpha, \chi) \leq t) = \int_0^t e^{-vx} d\pi(x) + \int_0^t [1 - \pi(x)] d(1 - e^{-vx}). \quad (1)$$

Lemma 1. The LST of the PDF of the length of the model's busy period $\tilde{\pi}_v(s)$, its mean value $\bar{\pi}_v$, and second moment are defined by

$$\tilde{\pi}_v(s) = \frac{v + s\tilde{\pi}(s+v)}{v+s}, \quad \bar{\pi}_v = \frac{1 - \tilde{\pi}(v)}{v}, \quad \bar{\pi}_v^{(2)} = \frac{\bar{\pi}_v + \tilde{\pi}'(v)}{v}, \quad (2)$$

where $\tilde{\pi}(s)$ is the unique solution of functional equation

$$\tilde{\pi}(s) = \tilde{B}(s + \sigma - \sigma\tilde{\pi}(s)). \quad (3)$$

Also, for the standard $M_r | G_r | 1 | \infty$ model, the LST of busy period of i^{th} and higher priority customers $\tilde{\pi}_i(s)$, $i = \overline{1, r}$ are defined from the following system of functional equations

$$\sigma_n \tilde{\pi}_n(s) = \sum_{i=1}^n a_i \tilde{\pi}_{ni}(s), \quad n = \overline{1, r}.$$

For the non-preemptive priority model, the functions $\tilde{\pi}_{ni}(s)$ are determined by the following expressions

$$\tilde{\pi}_{ni}(s) = \tilde{B}_i(s + \sigma_n - \sigma_n \tilde{\pi}_n(s)), \quad i = \overline{1, n}, \quad n = \overline{1, r}. \quad (4)$$

For the preemptive priority model, we will consider two modifications – b) and c).

For the first modification, the functions $\tilde{\pi}_{ni}(s)$ are determined [30] by the system of functional equations

$$\tilde{\pi}_{ni}(s) = \tilde{B}_i(s + \sigma_n - \sum_{j=i}^n a_j \tilde{\pi}_{nj}(s)) - \sum_{j=1}^{i-1} a_j \tilde{\pi}_{nj}(s) \frac{1 - \tilde{B}_i(s + \sigma_n - \sum_{j=i}^n a_j \tilde{\pi}_{nj}(s))}{s + \sigma_n - \sum_{j=i}^n a_j \tilde{\pi}_{nj}(s)}, \quad i = \overline{1, n}, \quad n = \overline{1, r}. \quad (5)$$

For the second modification, the functions $\tilde{\pi}_{ni}(s)$ can be found from the system of functional equations

$$\tilde{\pi}_{ni}(s) = \tilde{B}_i(s + \sigma_n - \sum_{j=i}^n a_j \tilde{\pi}_{nj}(s)) \left[1 - \sum_{j=1}^{i-1} a_j \tilde{\pi}_{nj}(s) \frac{\tilde{B}_i(s + \sigma_n - \sum_{j=i}^n a_j \tilde{\pi}_{nj}(s))}{s + \sigma_n - \sum_{j=i}^n a_j \tilde{\pi}_{nj}(s)} \right]^{-1}, \quad i = \overline{1, n}, \quad n = \overline{1, r}. \quad (6)$$

Lemma 2. The idle state probabilities $p_0(s)$, P_0 of $M | G | 1$ queue with work conservative discipline of service and disasters is defined by

$$p_0(s) = \frac{1}{\lambda + s - \lambda \tilde{\pi}_v(s)} = \frac{v + s}{s[\lambda + v + s - \lambda \tilde{\pi}(s + v)]}, \quad P_0 = \lim_{s \rightarrow 0} s p_0(s) = \frac{v}{v + \sigma(1 - \tilde{\pi}(v))} \quad (7)$$

4. The model $M_r | G_r | 1 | \infty$ with non-preemptive and preemptive priorities

To investigate the queue length of the model, we use the supplementary variable method. Let consider a random process $(\mathbf{L}(t), x(t), i(t))$, where $\mathbf{L}(t) = (L_1(t), \dots, L_r(t))$, $L_i(t)$ is the number of i^{th} priority customers in the model; $x(t)$ denotes the amount of the elapsed service time of the customer in service (otherwise $x(t) = 0$); $i(t) = 0$ if the server is empty and $i(t) = i$ if i^{th} priority customer is served at the moment t .

Obviously, the triple $(\mathbf{L}(t), x(t), i(t))$ describes the Markov process with the set of states

$$\Omega = \{i(t) = 0, x(t) = 0, L_j(t) = 0, 1 \leq j \leq r\} \cup \{i(t) = i, x(t) \geq 0, 0 \leq L_j(t), 1 \leq j, i \leq r\}, \quad t \geq 0.$$

Considering the transitions of $(\mathbf{L}(t), x(t), i(t))$ in the infinitesimal interval of time $(t, t + \Delta)$ we can set up the equations for the following probability density functions and their corresponding PGF.

$$P_i(\mathbf{n}, x, t) = \frac{\partial}{\partial t} P_i(\mathbf{L}(t) = \mathbf{n}, x(t) < x, i(t) = i) \cdot p_i(\mathbf{z}, x, s) = \sum_{\mathbf{n}=\mathbf{0}}^{\infty} \mathbf{z}^{\mathbf{n}} p_i(\mathbf{n}, x, s)$$

Theorem 1. For the transient $p(\mathbf{z}, s)$ and steady $P(\mathbf{z})$ PGF of queue length, we have

$$p(\mathbf{z}, s) = p_0(s) + \sum_{i=1}^r \frac{1 - \tilde{B}_i(s + \sigma + v - (\mathbf{a}, \mathbf{z}))}{s + \sigma + v - (\mathbf{a}, \mathbf{z})} p_i(\mathbf{z}, 0, s),$$

$$P(\mathbf{z}) = P_0 + \sum_{i=1}^r \frac{1 - \tilde{B}_i(\sigma + \nu - (\mathbf{a}, \mathbf{z}))}{\sigma + \nu - (\mathbf{a}, \mathbf{z})} P_i(\mathbf{z}, 0), \quad |z_i| \leq 1, i = \overline{1, r}. \quad (8)$$

Where $p_i(\mathbf{z}, 0, s)$ and $P_i(\mathbf{z}, 0)$ are determined from the recurrent relations

$$\begin{aligned} & \sum_{i=j+1}^r [1 - z_i^{-1} \tilde{B}_i(s + \sigma + \nu - \sigma_j \tilde{\pi}_j(s + \sigma^j + \nu - (\mathbf{a}, \mathbf{z})^j) - (\mathbf{a}, \mathbf{z})^j)] p_i(\mathbf{z}, +0, s) \\ &= 1 + \frac{\nu}{s} - (s + \sigma + \nu - \sigma_j \tilde{\pi}_j(s + \sigma^j + \nu - (\mathbf{a}, \mathbf{z})^j) - (\mathbf{a}, \mathbf{z})^j) p_0(s) \\ & \sum_{i=j+1}^r [1 - z_i^{-1} \tilde{B}_i(\sigma + \nu - \sigma_j \tilde{\pi}_j(\sigma^j + \nu - (\mathbf{a}, \mathbf{z})^j) - (\mathbf{a}, \mathbf{z})^j)] P_i(\mathbf{z}, +0) = \nu - (\sigma + \nu - \sigma_j \tilde{\pi}_j(\sigma^j + \nu - (\mathbf{a}, \mathbf{z})^j) - (\mathbf{a}, \mathbf{z})^j) P_0. \end{aligned} \quad (9)$$

The mean queue size of i^{th} priority customer is given

$$E(n_i) = \left. \frac{\partial}{\partial z_i} P(\mathbf{z}) \right|_{z_1 = z_2 = \dots = z_r = 1}, \quad i = \overline{1, r}.$$

Let consider joint distribution of queues lengths and the remaining service time of a customer in service

$$\Pi_i(\mathbf{n}, y, t) = \frac{\partial}{\partial} P_i(\mathbf{L}(t) = \mathbf{n}, x_+(t) < y, i(t) = i), \quad i = \overline{1, r},$$

where $x_+(t)$ is remaining service time of i^{th} priority customer which is in service at time t .

The z and Laplace transformations of $\Pi_i(\mathbf{n}, y, t)$ are given by

$$\begin{aligned} \Pi_i(\mathbf{z}, y, t) &= \sum_{\mathbf{n}=0}^{\infty} \mathbf{z}^{\mathbf{n}} \Pi_i(\mathbf{n}, y, t) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \dots \sum_{n_r=0}^{\infty} z_1^{n_1} z_2^{n_2} \dots z_r^{n_r} \Pi_i(n_1, n_2, \dots, n_r, t), \\ \Pi_i(\mathbf{z}, \omega_i, s) &= \int_0^{\infty} e^{-\omega_i y} dy \int_0^{\infty} e^{-st} \Pi_i(\mathbf{z}, y, t) dt = p_i(\mathbf{z}, +0, s) \frac{\{\tilde{B}_i(\omega_i) - \tilde{B}_i(s + \sigma + \nu - (\mathbf{a}, \mathbf{z}))\}}{s - \omega_i + \sigma + \nu - (\mathbf{a}, \mathbf{z})}. \end{aligned}$$

Thus, for the joint distribution of queues lengths we have

$$p(\mathbf{z}, s) = p_0(s) + \sum_{i=1}^r \int_0^{\infty} p_i(\mathbf{z}, x, s) dx = p_0(s) + \sum_{i=1}^r \Pi_i(\mathbf{z}, 0, s).$$

Let $U_i(t)$ be the unfinished work of customers of priority i at time t , where $i = \overline{1, r}$. The joint distribution of $U_1(t), U_2(t), \dots, U_r(t)$ at time t is defined by

$$\tilde{U}(\omega_1, \omega_2, \dots, \omega_r, t) = E[e^{-\omega_1 U_1(t) - \omega_2 U_2(t) - \dots - \omega_r U_r(t)}].$$

The Laplace transformation of $\tilde{U}(\omega_1, \omega_2, \dots, \omega_r, t)$ is then given by

$$\int_0^{\infty} e^{-st} \tilde{U}(\omega_1, \omega_2, \dots, \omega_r, t) dt = p_0(s) + \sum_{i=1}^r \frac{\Pi_i(\tilde{B}_1(\omega_1), \tilde{B}_2(\omega_2), \dots, \tilde{B}_r(\omega_r), \omega_i, s)}{\tilde{B}_i(\omega_i)}.$$

Theorem 2. For the transient PGF of queue length $p(\mathbf{z}, s)$ of the model we have

$$p(\mathbf{z}, s) = p_0(s) + \sum_{i=1}^r \frac{1 - \tilde{B}_i(s + \sigma + \nu - (\mathbf{a}, \mathbf{z})^{i-1})}{s + \sigma + \nu - (\mathbf{a}, \mathbf{z})^{i-1}} p_i(\mathbf{z}, +0, s), \quad (10)$$

where $p_0(s)$ is determined from (67), and $p_i(\mathbf{z}, +0, s)$ satisfy the following recurrent relations:

$$\begin{aligned} & \sum_{i=j+1}^r \left\{ 1 - z_i^{-1} \tilde{B}_i(s + \sigma + \nu - (\mathbf{a}, \mathbf{z})^{i-1}) - \frac{1 - \tilde{B}_i(s + \sigma + \nu - (\mathbf{a}, \mathbf{z})^{i-1})}{s + \sigma + \nu - (\mathbf{a}, \mathbf{z})^{i-1}} [\sigma_j \tilde{\pi}_j(s + \sigma^j + \nu - (\mathbf{a}, \mathbf{z})^j) + (\mathbf{a}, \mathbf{z})^{i-1}] z_i^{-\delta} \right\} p_i(\mathbf{z}, +0, s) = \\ &= 1 + \frac{\nu}{s} - (s + \sigma + \nu + \sigma_j \tilde{\pi}_j(s + \sigma^j + \nu - (\mathbf{a}, \mathbf{z})^j) - (\mathbf{a}, \mathbf{z})^j) p_0(s). \end{aligned} \quad (11)$$

Here $p_0(s)$ can be determined from (11), when $j = r$, or from (7).

5. Concluding Remarks

In this paper, the $M_r | G_r | 1 | \infty$ queuing model with priorities and disasters has been considered. Based on the supplementary variable method, the expressions for many performance measure of the model have been derived.

The approach developed in this paper can be used for studying the queue models $M_r | G_r | 1 | \infty$ with disasters, with mixed priorities, server vacations, and server breakdowns.

REFERENCES

1. E. Gelenbe, (1989) Random neural networks with negative and positive signals and product form solution. *Neural Computation*, 1(4), pp. 502–510.
2. X. Chao, (1995) A queueing network model with catastrophes and product form solution. *Oper.Res. Lett.* 18, pp.75-79.
3. G. Jain, and K. Sigman, (1996) A Polleczech-Khintchine Formulation for M/G/1 Queues with Disasters. *Journal of Applied Probability*, 33, 1191-1200.
4. R.J. Boucherie, O.J. Boxma, (1994) The workload in the M/G/1 queue with work removal. *Probability in the Engineering and informational sciences* 10, 261-278.
5. Q.L. Li, and C. Lin, (2006). The M/G/1 processor-sharing queue with disasters. *Computers & Mathematics with Applications*, Vol. 51, pp. 987-998.
6. J.R. Artalejo, (2000) G-networks: A versatile approach for work removal in queue networks. // *European J. Oper. Res.* 126, pp.233-249.
7. P.P. Bocharov, V.M. Vishnevskii, (2003) G-Networks: Development of the Theory of Multiplicative Networks, *Automation and Remote Control*, Vol. 64, No. 5, pp. 714-739.
8. T.V. Do, (2011) Bibliography on G-networks, negative customers and applications. *Mathematical and Computer Modelling*, Volume 53, Issues 1-2, January, pp. 205-212.
9. N. K. Jaiswal, (1968) *Priority Queues*, New York, Academic Press.
10. L. Kleinrock, (1976) *Queueing Systems, Vol. II: Computer Applications*, John Wiley & Sons, Inc.
11. B.V.Gnedenko, E.A. Danielyan, et all, (1973) *Priority Queueing Systems*. Moscow, MSU Press, Moscow, 447.
12. G.P. Kliov, (1966) *Stochastic service systems*. Moscow, Nauka, 245.
13. H. Takagi, (1991) *Queueing Analysis: A Foundation of Performance evaluation, vacation and priority systems, Part 1*, North-Holland, Amsterdam.
14. V.F. Matveev, V.G. Ushakov, (1984) *Queueing systems*. Moscow, MSU Press, 239 .
15. D. Towsley, S.K. Tripathi, (1989) *A Single Server Priority Queue with Server Failures and Queue Flushing*, Technical Report: UM-CS-1989-016.
16. Kh.V. Kerobyan, (2007) The $M_r | G_r | 1$ queuing model with the priorities and "negative" customers, *Bulletin of RAU, Series: Physics-Mathematics and Natural Sciences*, #2, pp. 6-18.
17. K.V. Kerobyan, (2007) The Model M/G/1 with Unreliable Server and «Negative» Customers / *Proc. of YSU, Series: Physics-Mathematics Sciences*, #3, pp. 11-20.

A product-form solution for on-off components in PEPA

Nigel Thomas

School of Computing Science, Newcastle University, UK.
`nigel.thomas@ncl.ac.uk`

Abstract. This paper addresses the issue of how the behaviour of the model may be used to directly show product form results in general models without relying on additional insight from the modeller. To do this properties of the model are defined which allow the identification of model decompositions which are then shown to exhibit product by employing the reversed process.

1 Introduction

Throughout the history of analysing the performance of computer and communication networks, researchers have endeavoured to find efficient mechanisms for tackling models with extremely large state spaces. Within stochastic process algebra there are a number of approaches that have been investigated, including decomposition [1, 2], mean value analysis [3], and fluid approximations [4, 5].

One of the most powerful and mathematically attractive decomposition techniques are so-called product form solutions. Such solutions are derived on the basis that the components in the model are statistically independent in their steady state behaviour and so the steady state solutions for components may be found in isolation without the need to generate the entire state space of the model. Two general approaches have been used to identify product form solutions in stochastic process algebra, structural decomposition [6, 7] relying on identifying patterns in the model specification and reverse processes [9–11].

The aim of this paper is to address the issue of how the behaviour of the model may be used to directly show product form results in general models without relying on additional insight from the modeller. To do this properties of the model are defined which allow the components to be decoupled so that their marginal steady state distributions can be derived. Under certain conditions, these decomposed models give rise to a product form solution.

The paper begins by re-introducing Hillston's Markovian process algebra, PEPA [12], together with the set of concepts required to describe features of a model and then briefly discusses the notion of behavioural independence. In Section 3 the exploitation of behavioural independence is made in relation to simple product form decomposition. Finally some conclusions and future work directions are presented.

2 PEPA

A formal presentation of PEPA is given in [12], in this section a brief informal summary is presented. PEPA, being a Markovian Process Algebra, only supports actions that are negative exponentially distributed at given rates. Specifications written in PEPA represent Markov processes and can be mapped to a continuous time Markov chain (CTMC). Systems are specified in PEPA in terms of *activities* and *components*. An activity (α, r) is described by the *type* of the activity, α , and the *rate* of the associated negative exponential distribution, r . This rate may be any positive real number, or given as *unspecified* using the symbol \top . The syntax for describing components is given as:

$$P ::= (\alpha, r).P \mid P + Q \mid P/L \mid P \bowtie_L Q \mid A$$

- The component $(\alpha, r).P$ performs the activity of type α at rate r and then behaves like P . The component $P + Q$ behaves either like P or like Q , the resultant behaviour being given by the first activity to complete.
- The component P/L behaves exactly like P except that the activities in the set L are concealed, their type is not visible and instead appears as the unknown type τ .
- Concurrent components can be synchronised, $P \bowtie_L Q$, such that activities in the *cooperation set* L involve the participation of both components. In PEPA the shared activity occurs at the slowest of the rates of the participants and if a rate is unspecified in a component, the component is passive with respect to the activities of that type. The parallel combinator, \parallel , is used as shorthand to denote synchronisation with no shared activities, i.e. $P \parallel Q \equiv P \bowtie_{\emptyset} Q$.
- $A \stackrel{\text{def}}{=} P$ gives the constant A the behaviour of the component P .

In the following sections a number of properties of PEPA models will be referred to. Since the definitions of these properties are presented in detail elsewhere [12], only a general description of them is included here. Informally, it can be said that, two PEPA expressions are *isomorphic* if they give rise to Markov chains which are equivalent such that for every state in each Markov chain there is a corresponding state in the other with the same one-step transition rates to states which are similarly equivalent. A *derivative* is the “state” of a component defining its current behaviour. For example, if $P \stackrel{\text{def}}{=} (\alpha, r).Q$ then Q is a derivative of P and if $Q \stackrel{\text{def}}{=} (\beta, r_2).R$ then R is a derivative of both Q and P , and so on. The *derivative set*, $ds(P)$, is the set of all the possible derivatives of a component, P . The *current action type set* of a component P , $\mathcal{A}(P)$, contains all the action types (but not the rates) that are enabled in the current derivative of the component P . The *complete action type set* of a component P , $\mathcal{A}(P)$, contains all the action types (but not the rates) that are enabled in any of the derivatives $P' \in ds(P)$, hence $\mathcal{A}(P) = \bigcup_{P' \in ds(P)} \mathcal{A}(P')$. The *current action set* of a component P , $\mathcal{Act}(P)$, contains all the actions (type and rate pairs) that are enabled in the current derivative of the component P . In addition it is necessary to construct a number of additional definitions.

Fertile action An action γ is said to be fertile in derivative P_i if $P_i \xrightarrow{\gamma} P_j$ and $i \neq j$.

Current fertile action type set The current fertile action type set of a component P , denoted $\mathcal{A}_f(P)$, is the set of all action types of actions that are fertile in the current derivative of P .

Complete fertile action type set The complete fertile action type set of a component P , denoted $\mathbf{A}_f(P)$, is the set of all action types of actions that are fertile in at least one derivative of P .

3 Behavioural Independence

Put simply the notion of *behavioural independence* is simply that a component in a model behaves identically regardless of the current behaviour of the other components in the model. This property can be defined more formally thus:

The component P is said to be **behaviourally independent** in the model $P \bowtie_L Q$ if for every $P_i \in ds(P)$

$$\begin{aligned} & Act \left((P_i \bowtie_L Q_j) / \{ \mathcal{A}(P_i \bowtie_L Q_j) / \{ \mathcal{A}_f(P_i) \cap L \} \} \right) = \\ & Act \left((P_i \bowtie_L Q_k) / \{ \mathcal{A}(P_i \bowtie_L Q_k) / \{ \mathcal{A}_f(P_i) \cap L \} \} \right) \end{aligned}$$

$\forall Q_j, Q_k \in ds(Q)$ s.t. $(P_i \bowtie_L Q_j), (P_i \bowtie_L Q_k) \in ds(P \bowtie_L Q)$

Obviously the trivial case for behavioural independence is where there are no shared actions, i.e. $P \parallel Q$, however this is not the only case where components may be considered to be behaviourally independent. Furthermore, the fact that no actions are shared between two components does not mean they will always be behaviourally independent in the presence of other components. For example, in $(P \parallel Q) \bowtie_L R$ the interaction between P and R may influence the interaction between Q and R , causing P and Q to be behaviourally *dependent*.

If a component is not behaviourally independent then it must be dependent on some other component to perform one of more actions during its evolution. This dependence is referred to by saying that component P *controls* component Q over action $K \subset L$ in $P \bowtie_L Q$ if the rate at which an action of type $k \in K$ can happen in $Q_i \in dsQ$ depends on the current derivative of P . Clearly, if P controls Q over K then Q cannot be behaviourally independent, but the independence, or otherwise, of P is not known by this statement.

The definition for behavioural independence given above relies on knowing the combination of each component's derivatives in every state of the underlying Markov chain; in essence, the entire state space would have to be explored. If the potential benefit of compositionality is to be realised then it is necessary that behavioural independence can be identified at the component level, without knowledge of the underlying CTMC. A number of simple specific cases in which P can be observed to be behaviourally independent in $P \bowtie_L Q$ can be stated.

- $L \cap \mathbf{A}_f(P) = \emptyset$.

- Q is redundant in $P \boxtimes Q$ and P is sequential.
- If an action $l \in (L \cap \mathcal{A}_f(P))$ is enabled in $Q_i \in ds(Q)$ at rate α then it is also enabled in every $Q_j \in ds(Q)$ at rate α (α may be unspecified).
- If an action $l \in (L \cap \mathcal{A}_f(P))$ is enabled in $Q_i \in ds(Q)$ at a rate $\alpha \geq Mr(l, P)$ then it is also enabled in every $Q_j \in ds(Q)$ at rate $\alpha_j \geq Mr(l, P)$ (note: α or α_j may be unspecified).

Note that the final condition is a numerical one whereas the others are structural only. In general numerical values are not considered at the model specification stage and so this last condition is of only limited use.

More formally, if one of the following conditions holds true $\forall \gamma \in \mathcal{A}(Q_i)$ and $\forall Q_i \in ds(Q)$ then P is behaviourally independent in $P \boxtimes Q$.

1. if $\gamma \in \mathcal{A}_f(Q_i)$ such that $Q_i \xrightarrow{\gamma} Q_j$, ($i \neq j$) then $Act(Q_i/\{\{\mathcal{A}(Q_i)/\{L \cap \mathcal{A}_f(P)\}\}\}) = Act(Q_j/\{\{\mathcal{A}(Q_j)/\{L \cap \mathcal{A}_f(P)\}\}\})$
2. if $\gamma \in \{\mathcal{A}_f(P_i) \cap L\}$ such that $P_i \parallel Q_k \xrightarrow{\gamma} P_j \parallel Q_l$, ($i \neq j$) then

$$Act(Q_k/\{\{\mathcal{A}(Q_k)/\{L \cap \mathcal{A}_f(P_j) \cap \mathcal{A}_f(P_i)\}\}\}) = Act(Q_l/\{\{\mathcal{A}(Q_l)/\{L \cap \mathcal{A}_f(P_j) \cap \mathcal{A}_f(P_i)\}\}\})$$

The conditions set out here can be used to test a model for behavioural independence at the component level, however they are unnecessarily strong, thus it is not possible to say that a component failing to meet these criteria is dependent in some way on another component.

3.1 Product form over components

As well as being used to identify independent behaviour leading to decomposition, behavioural independence and control can also be used to identify cases where product form solutions exist. Such a case is the queueing model with breakdowns illustrated below.

$$Queue_0 \stackrel{def}{=} (arrival, \top).Queue_1$$

$$Queue_i \stackrel{def}{=} (arrival, \top).Queue_{i+1} + (service, \top).Queue_{i-1} \quad 1 \leq j \leq N-1$$

$$Queue_N \stackrel{def}{=} (service, \top).Queue_{N-1}$$

$$Server_{on} \stackrel{def}{=} (fail, \xi).Server_{off} + (arrival, \lambda).Server_{on} + (service, \mu).Server_{on}$$

$$Server_{off} \stackrel{def}{=} (repair, \eta).Server_{on}$$

$$Queue_0 \boxtimes_{\{service, arrival\}} Server_{on}$$

It is clear that the *Server* component is behaviourally independent in this model as neither of the shared actions affects its evolution. Similarly it is clear that the *Server* component controls the *Queue* component over the actions *service* and *arrival*. A number of other important factors are also apparent:

- All the actions of *Queue* are shared actions.
- All shared actions are enabled in *Server_{on}*.
- No shared actions are enabled in *Server_{off}*.
- The activity *fail* does not alter the current derivative of *Queue*.

These six factors mean that a product form solution exists over the *Server* and *Queue* components such that the joint steady state probabilities are given as:

$$p(\text{Server}_j | \text{Queue}_i) = p(\text{Server}_j) \cdot p(\text{Queue}_i)$$

where $j \in \{on, off\}$ and $0 \leq i \leq N$.

The model illustrated in here is reversible and it is possible to derive a product form solution using the characterisation derived in [8]. That approach required the identification of reversible components and the application of restrictions on the cooperation between them. This requires a detailed study of both the components and the interface, whereas the approach described here only requires a simple inspection of the components, only adherence to the five criteria.

1. One component, A , of a pair $A \underset{L}{\bowtie} B$ is behaviourally independent.
2. The other component, B , is controlled by A over all the actions in the cooperation set, L .
3. The complete action type set of A , $\mathcal{A}(B)$ is contained within its interface, $\mathcal{A}(B) = L$.
4. All actions in the cooperation set, L , are enabled in exactly one derivative of A .
5. No actions in the cooperation set, L , are enabled in any other derivative of A .

As long as these 5 stated conditions are not broken then the model illustrated above can be easily adapted to incorporate additional (non-reversible) features, such as batch service, without compromising the product form solution. Such a model is illustrated below.

$$\text{Queue}_0 \stackrel{\text{def}}{=} (\text{arrival}, \top) \cdot \text{Queue}_1$$

$$\text{Queue}_i \stackrel{\text{def}}{=} (\text{arrival}, \top) \cdot \text{Queue}_{i+1} + (\text{service}, \top) \cdot \text{Queue}_0 \quad 1 \leq j \leq N-1$$

$$\text{Queue}_N \stackrel{\text{def}}{=} (\text{service}, \top) \cdot \text{Queue}_0$$

$$\begin{aligned} \text{Server}_{on} &\stackrel{\text{def}}{=} (\text{fail}, \xi) \cdot \text{Server}_{off} + (\text{arrival}, \lambda) \cdot \text{Server}_{on} \\ &\quad + (\text{service}, \mu) \cdot \text{Server}_{on} \end{aligned}$$

$$\text{Server}_{off} \stackrel{\text{def}}{=} (\text{repair1}, \eta_1) \cdot \text{Server}_{standby}$$

$$\text{Server}_{standby} \stackrel{\text{def}}{=} (\text{repair2}, \eta_2) \cdot \text{Server}_{on}$$

$$\text{Queue}_0 \underset{\{\text{service}, \text{arrival}\}}{\bowtie} \text{Server}_{on}$$

3.2 Excluded shared actions

In [8] Hillston and Thomas identified the limitations of their approach as being that they could not handle cases where the components were not reversible, but the resultant model was.

$$\begin{aligned}
P_0 &\stackrel{\text{def}}{=} (in_P, r).P_1 + (a, r).P'_1 \\
P_i &\stackrel{\text{def}}{=} (in_P, r).P_{i+1} + (out_P, s).P_{i-1} \\
&\quad 1 \leq i \leq N - 1 \\
P_N &\stackrel{\text{def}}{=} (out_P, s).P_{N-1} \\
P'_i &\stackrel{\text{def}}{=} (a, r).P'_{i+1} \\
&\quad 1 \leq i \leq M - 1 \\
P'_M &\stackrel{\text{def}}{=} (a, r).P_0 \\
\\
Q_0 &\stackrel{\text{def}}{=} (in_Q, r).Q_1 + (b, r).Q'_1 \\
Q_i &\stackrel{\text{def}}{=} (in_Q, r).Q_{i+1} + (out_Q, s).Q_{i-1} \\
&\quad 1 \leq i \leq N - 1 \\
Q_N &\stackrel{\text{def}}{=} (out_Q, s).Q_{N-1} \\
Q'_i &\stackrel{\text{def}}{=} (b, r).Q'_{i+1} \\
&\quad 1 \leq i \leq M - 1 \\
Q'_M &\stackrel{\text{def}}{=} (b, r).Q_0 \\
\\
Sys &\stackrel{\text{def}}{=} P \boxtimes_{\{a, b\}} Q
\end{aligned}$$

In this model it might appear that the P component controls the Q component over the action b and the Q component controls the P component over the action a . However, $b \notin \mathcal{A}(P_0)$ and $a \notin \mathcal{A}(Q_0)$, hence both these shared actions are permanently blocked. Therefore it is clear that no control is exerted between these two components, both are behaviourally independent and the model has a trivial product form solution when the excluded behaviours are removed from the components (the rates of the actions a and b are always zero). However, this is not universally true for excluded behaviours. For example, consider the case where P'_M is redefined as follows.

$$P'_M \stackrel{\text{def}}{=} (b, r).P_0$$

Now $b \in \mathcal{A}(P_0)$, but in this model b will never happen since action a is required to reach P'_M and action a is always blocked. Both components are still behaviourally independent by definition 3.1, however the tests specified in Section 3.2 will not identify this. In general it is not a simple matter to know that an action will not occur unless the states of the underlying CTMC are explored, therefore cases of excluded actions such as this are, in general, extremely problematical when working at the component level.

3.3 Extended example

In this simple example there are a pair of resources. The only stipulation is that at least one must be online at any time. Thus, any resource may choose to go offline only if another remains online. The correctness is held by the fact that in order to make the transition from $xOnline$ to $xOffline$, each resource must have the (passive) cooperation of its partner. Once offline however, the resources are incapable of communicating, and so the resource which is online must remain so.

$$\begin{aligned}
AOnline &\stackrel{def}{=} (Aoff, r_2).AOffline + \\
&\quad (Boff, \top).AOnline \\
AOffline &\stackrel{def}{=} (on, r_4).AOnline \\
BOnline &\stackrel{def}{=} (Boff, r_2).BOffline \\
&\quad +(Aoff, \top).BOnline \\
BOffline &\stackrel{def}{=} (on, r_6).BOnline
\end{aligned}$$

$$AOnline \boxtimes_{\{Aoff, Boff\}} BOnline$$

Thus the model has the excluded state of $AOffline|BOffline$, and a trivial product form over the remaining states, given by,

$$\pi_{(AY, BZ)} = \frac{1}{X} \pi_{AY} \cdot \pi_{BZ}$$

where $Y, Z = \{Online, Offline\}$ and $X = \pi_{AOnline} \cdot \pi_{BOnline} + \pi_{AOnline} \cdot \pi_{BOffline} + \pi_{AOffline} \cdot \pi_{BOnline}$. The model can be made slightly more interesting if resource perform some computation. There are a number of possibilities in this regard.

– *Both resources compute together at all times.*

$$\begin{aligned}
AOnline &\stackrel{def}{=} (compute, r_7).ABusy \\
&\quad +(Aoff, r_2).AOffline \\
&\quad +(Boff, \top).AOnline \\
ABusy &\stackrel{def}{=} (complete, r_8).AOnline \\
BOnline &\stackrel{def}{=} (compute, r_7).BBusy \\
&\quad +(Boff, r_2).BOffline \\
&\quad +(Aoff, \top).AOnline \\
BBusy &\stackrel{def}{=} (complete, r_8).BOnline
\end{aligned}$$

$$AOnline \boxtimes_{\substack{\{compute \\ Aoff, Boff\}}} BOnline$$

– One of the resources is busy, the other must be online.

$$\begin{aligned}
AOnline &\stackrel{\text{def}}{=} (compute, r_7).ABusy \\
&\quad + (compute, \top).AOnline \\
&\quad + (Aoff, r_2).AOffline \\
&\quad + (Boff, \top).AOnline \\
ABusy &\stackrel{\text{def}}{=} (complete, r_8).AOnline \\
BOnline &\stackrel{\text{def}}{=} (compute, r_7).BBusy \\
&\quad + (compute, \top).BOnline \\
&\quad + (Boff, r_2).BOffline \\
&\quad + (Aoff, \top).BOnline \\
BBusy &\stackrel{\text{def}}{=} (complete, r_8).BOnline
\end{aligned}$$

$$AOnline \overset{\boxtimes}{\underset{\{compute, Aoff, Boff\}}{}} BOnline$$

This case only has a product form if $r_7 = r_9$.

4 Conclusions and Further Work

In this paper a discussion of the notions of behavioural independence and control has been presented in relation to product form solution. It is probable that many additional classes of product form solution will have subclasses that can be defined using behavioural independence and control, although this remains to be proved. The simple product form here can be extended by considering parts of components as behaviourally independent and parts which exert control. This ongoing work aims to build simple relations for models which may be more generally characterised by the definitions in [7]. By exploring these subclasses of solutions it will be possible to gain greater understanding of the links between different product form solutions (and non-product form solutions) and where they may overlap. In addition, practical approaches to the identification of these properties allow automation of model decomposition and hence greater support for the performance modeller.

References

1. Hillston J. Exploiting Structure in Solution: Decomposing Composed Models, in: *FMPA Lecture Notes*, Springer-Verlag, 2001.
2. Thomas N. and Bradley J. and Thornley D. Approximate solution of PEPA models using component substitution *IEE Proceedings - Computers and Digital Techniques*, **150**(2), pp. 67-74, 2003.

3. Thomas N. and Zhao Y. Mean value analysis for a class of PEPA models *The Computer Journal*, **54**(5), 643-652, 2011.
4. J. Hillston, Fluid flow approximation of PEPA models, in: *Proceedings of QEST'05*, pp. 33-43, IEEE Computer Society, 2005.
5. Zhao Y. and Thomas N. Efficient solutions of a PEPA model of a key distribution centre *Performance Evaluation*, **67**(8), pp. 740-756, 2010.
6. Harrison P.G. and Hillston J. Exploiting Quasi-reversible Structures in Markovian Process Algebra Models, *The Computer Journal*, **38**(7), pp. 510-520, 1995.
7. Hillston J. and Thomas N. Product Form Solution for a Class of PEPA Models, *Performance Evaluation*, **35**(3-4), pp. 171-192, June 1999.
8. Hillston J. and Thomas N. A Syntactic Analysis of Reversible PEPA Models, in: *Proceedings of the Sixth International Workshop on Process Algebra and Performance Modelling*, 1998.
9. Harrison, P.G., Turning back time in Markovian process algebra, *Theoretical Computer Science*, **290**(3): pp. 1947-1986, 2003.
10. Harrison PG. and Thomas N. Product-form solution in PEPA via the reversed process in: *Network performance engineering*, pp. 343-356, 2011
11. Harrison PG. and Thomas N. Semi-Product-Form Solution for PEPA Models with Functional Rates in: *Analytical and Stochastic Modeling Techniques and Applications*, pp. 416-430, 2013.
12. Hillston J. *A Compositional Approach to Performance Modelling*, Cambridge University Press, 1996.
13. Thomas N. Exploiting behavioural independence and control in Markovian process algebra, in: *Proceedings of the 1st Workshop on Process Algebra with Stochastic Timed Activities*, University of Edinburgh, 2002.
14. Harrison PG. and Thomas N. State-dependent rates and semi-product-form via the reversed process in: *Computer Performance Engineering*, pp. 207-218, 2010.
15. Thomas N. Using ODEs from PEPA models to derive asymptotic solutions for a class of closed queueing networks Technical Report CS-TR-1129, School of Computing Science, Newcastle University, 2008

TRANSIENT $MMAP_r(t)|G_r|\infty$ MODEL WITH DISASTERS

Kh. Kerobyan, R. Kerobyan

Armenian Society of Computer and Information Systems
Los Angeles City College
khankerobyan@gmail.com

In the present paper the $MMAP_r(t)|G_r|\infty$ queue model with r types of customers, non-stationary, Marked Markov Arrival Process of traffic and non-stationary Poisson arrival of disasters is considered. The transient distribution and two moments of number of each type of customers are obtained.

Keywords: queuing model, $MMAP$ arrival, disaster, infinite number of servers.

Introduction

The distinguishing characteristic of New Generation Networks (NGN) is the integration of heterogeneous recourses, applications, technologies, customers and data into one united information infrastructure which is ubiquitous and accessible anytime and anywhere. This integration process includes all layers of NGN and makes its QoS metrics guaranteeing more challenging Kouvatsos [1]. To solve the NGN optimal design and performance providing problems the methods of statistical simulation, modeling, and prediction are widely used. However, the application of statistical modeling methods and tools even for several element of NGN, for example a protocol, a server, or a canal, is complicated because network statistical (traffic and service) processes usually have heterogeneous, non-stable, and correlated character Tripathi, Sharma, Raghavan [2], Ma [3]. To evaluate the network canals' performance main parameters, such as capacity, delay and packet loss probability, the infinity server queue models are widely used. As shown by the large number of measurements, the traffic of modern IP networks can be characterized by the heterogeneousness, the non-stationarity, the burstiness, the short-range (SRD) and the long-range dependence (LRD). These factors make the process of modeling and performance evaluation of modern networks canals more challenging Park, Kim, Crovella [4], Paxson, Floyd [5]. To describe the network traffic, the models based on finite Markovian Processes such as Markov Arrival Process (MAP), Batch Markov Arrival Process (BMAP), Marked Arrival Process (MMAP) and their generalizations are widely used. The $MMAP$ and MAP arrivals properties and their applications are presented in Artalejo [6].

The $BMAP |G|\infty$ model and its some generations have been studied by Breuer [7], Baum and Kalashnikov [8]. By using semi-Markov processes (SMP) and matrix analytic methods, the probability generating function (PGF) of the number of busy servers and its moments are considered. The models with phase type $PH |G|\infty$ and Markov modulated arrival are considered in Ramaswami and Neuts [9] and Blom, Mandjes, and Thorsdottir [10]. The model $M_k |M_k|\infty$ queue with correlated k heterogeneous customers in a batch and exponential service time is studied by Choi [11]. In steady state, the joint PGF of the number of customers of type i being served in the system is derived explicitly by solving a partial differential equation. The generalization of this model for general service time $M_k |G_k|\infty$ is considered by Falin [12]. The model $MAP_k |G_k|\infty$ with the structured batch arrival of k types of customers is considered by Masuyama [13]. In steady state, the differential equations for PGF of the number of busy server and its solution are obtained. Papers of Keilson, Servi [14], Eick, Massey, Whitt [15] are devoted to the study of $M(t)|G|\infty$ model with non-stationary Poisson input stream. The distribution of number of busy servers for the inhomogeneous model $BMAP(t) |G|\infty$ is obtained in Breuer [7].

To evaluate the impact of networks threats one considers the infinity server model with the random environment and disasters. The models $M_k |M_k|\infty$ with semi Markov environment are studied by Linton and Purdue [16] and [17]. The general clearing model has been introduced by Serfozo [18]. The generalization of this model for infinite-server queue subject to an extraneous phase process is considered

by Templeton and Purdue [16]. The clearing phase process is modeled by an M-state irreducible SMP. The model $M|M|\infty$ with disasters in steady-state is studied by Bohn [19] and Erromou and Fakinos [20]. In the present paper the PGF of number of busy servers and its mean value for the $MMAP_k(t)|G_k|\infty$ model with disasters are considered.

Model Description. The model consists of infinity number of identical servers; K type of customers arriving according to non-homogeneous MMAP. The service of arriving batches of customers starts immediately. Service time of i type customers is a random variable (r.v.) γ_i , not depending on input process, on model state, having arbitrary distribution $G_i(t) = P(\gamma_i \leq t)$ and finite mean value $\bar{\gamma}_i$. The disasters arrive to the model according to Poisson process with the rate ν . Arriving disaster destroys (flashes out) instantly all customers in the model. Then the model continues working from the empty state. Assume that the arrival MMAP is given by underlying Markov process (MP) $\{J(t); t \geq 0\}$ (phase process (PP)) with finite set $E = \{1, 2, \dots, m\}$ of states and sequence of characteristic matrices $\{D_0(t), D_h(t); \mathbf{h} \in C^0\}$ of $m \times m$ size. C^0 is a finite or counting set of nonnegative integers (size of arriving batches), $\mathbf{h} = (h_1, h_2, \dots, h_K)$, $\mathbf{h} \in C^0$, and h_r is a number of type r customers, $1 \leq r \leq K$ in a batch. $D_0(t)$ is a non-singular matrix with negative diagonal and non-negative extra-diagonal elements, the column-wise sum of which is less or equal to zero. $D_0(t)$ manages $PP J(t)$ transitions, that are not accompanied with customer generation. Non-negative matrices $D_h(t)$ govern the transitions of the phase MP $J_i(t)$ with a mark $\mathbf{h} = (h_1, h_2, \dots, h_K)$ which are accompanied with generation of batches of customers. Further we assume that PP $J(t)$ is an irreducible Markov process with generating matrix $D(t)$, with a set E of states, and with vector $\boldsymbol{\pi}(t) = (\pi_1(t), \pi_2(t), \dots, \pi_m(t))$. Here $D(t)$ is the matrix of size $m \times m$:

$$D(t) = D_0(t) + \sum_{\mathbf{h} \in C^0} D_h(t), \quad (D(t) \neq D_0(t)),$$

$$\boldsymbol{\pi}(t)D(t) = 0, \quad D(t)\mathbf{e} = 0, \quad \boldsymbol{\pi}(t)\mathbf{e} = 1,$$

where \mathbf{e} is a unit vector column. For MMAP the arrival rate of customers is

$$\lambda(t) = \boldsymbol{\pi}(t) \sum_{n=1}^{\infty} n \sum_{\mathbf{h} \in C^0, h_1+h_2+\dots+h_K=n} D_h(t)\mathbf{e}.$$

Let consider a counting process of MMAP $\{N(t), J(t); t \geq 0\}$, $N(t) = (N_1(t), N_2(t), \dots, N_K(t))$, where $N_r(t)$ represents the number of r type customers arriving on time interval $[0, t)$, $1 \leq r \leq K$. Denote by $P(\mathbf{n}, t)$ the matrix of transition probabilities of MMAP counting process with elements

$$p_{lj}(\mathbf{n}, t) = P(N(t) = \mathbf{n}, J(t) = j | J(0) = l), \quad 1 \leq l, j \leq m,$$

where $\mathbf{n} = (n_1, n_2, \dots, n_K)$ and n_1, n_2, \dots, n_K are nonnegative integers.

We define the matrix of generating functions

$$\tilde{D}(\mathbf{z}, t) = D_0(t) + \sum_{\mathbf{h} \in C^0} \mathbf{z}^{\mathbf{h}} D_h(t), \quad 0 \leq z_r < 1, \quad 0 \leq r \leq K$$

where $\mathbf{z}^{\mathbf{h}} = (z_1^{h_1}, z_2^{h_2}, \dots, z_K^{h_K})$.

$\{P(\mathbf{n}, t), \mathbf{n} \geq 0\}$ satisfy the Kolmogorov differential equations

$$\frac{d}{dt} P(\mathbf{n}, t) = P(\mathbf{n}, t) D_0(t) + \sum_{\mathbf{h} \leq \mathbf{n}, \mathbf{h} \in C^0} P(\mathbf{n} - \mathbf{h}, t) D_h(t), \quad \mathbf{n} \geq 0. \quad (1)$$

the solution for which can be presented in matrix-exponential form $P(\mathbf{n}, t) = e^{\int_0^t D(x) dx}$.

Using the equation (1), it can be shown that generating function of MMAP counting process $N(t)$, $\tilde{P}(z, t) = E z^{N(t)} = \sum_{n \geq 0} z^n P(n, t)$ is given by

$$\tilde{P}_i(z, t) = e^{\int_0^t \tilde{D}(z, x) dx}.$$

Here E means expectation. Let define the following random variables:

If we consider only the point process formed by customers of type r , we obtain a BMAP with characteristic matrices

$$D_0(r, t) = D_0(t) + \sum_{h \in C^0, h_r=0} D_h(t), \quad D_n(r, t) = \sum_{h \in C^0, h_r=n} D_h(t), \quad n \geq 1,$$

and the counting process $N_r(t)$, $1 \leq r \leq K$.

If we do not distinguish the type of customers, the point process consisting of all customers is a BMAP with characteristic matrices

$$D_n(t) = \sum_{h \in C^0, h_1+h_2+\dots+h_K=n} D_h(t), \quad n \geq 1,$$

and the counting process $N_g(t) = \sum_{r=1}^K N_r(t)$.

The expectation matrix of the counting process $N_r(t)$ over the time interval $[0, t)$ is given by

$$m_{1,r}(t) = \int_0^t P^\Phi(0, u) \sum_{n=0}^{\infty} n D(r, u) P^\Phi(u, t) du = \int_0^t e^{\int_0^u D(x) dx} \sum_{n=0}^{\infty} n D(r, u) e^{-\int_0^u D(x) dx} du, \quad 1 \leq r \leq K,$$

where $P^\Phi(u, t)$ is the transition probability matrix of the PP on interval $[u, t)$, $P^\Phi(u, t) = e^{\int_u^t D(x) dx}$.

Let $N^s(t)$ be the number of customers present in the system at the time t , or equivalently the number of busy servers at time t ;

$$N^s(t) = \sum_{i \in S} \sum_{h \in C^0} N_{hi}^s(t),$$

where $N_{hi}^s(t)$ is a number of type r customers present in the system at time t .

The main feature of MMAP|G| ∞ queue is independence of arrival and service processes of different marks $\mathbf{h} = (h_1, h_2, \dots, h_K)$. This allows consider sequences of marks separately.

Model Analysis. First of all, let suppose the states of the model are observed during the time interval $[u, t)$, $0 \leq u \leq t$. Let denote by $N(u, t)$ the number of customers arrived at moment u , $0 \leq u \leq t$, and still in service at the moment t , $N(t) = N(0, t)$, $N(0) = \mathbf{0}$, where $\mathbf{0}$ is a vector with 0 elements. Let $R_{jk}(\mathbf{n}, u, t)$ be the probability that $\mathbf{n} = (n_1, n_2, \dots, n_K)$ customers are in service at moment t , and PP $J(u)$ is in the phase $j \in E$ under condition that at initial moment $t=0$ the system was empty, and PP $J(u)$ was in phase $k \in E$, $R_{jk}(\mathbf{n}, u, t) = P(N(u, t) = \mathbf{n}, J(u) = j | N(0) = \mathbf{0}, J(0) = k)$. $R(\mathbf{n}, u, t)$ is a $m \times m$ matrix with elements $R_{jk}(\mathbf{n}, u, t)$. Assume also $\mathbf{R}(u, t) = (R(\mathbf{n}, u, t), \mathbf{n} \geq \mathbf{0})$ and $\mathbf{R}(t) = \mathbf{R}(0, t)$.

Let suppose that for fixed time t an arrival of mark $\mathbf{h} = (h_1, h_2, \dots, h_K)$ occurred during infinitesimal time interval du . It is well known that for the given mark $\mathbf{h} = (h_1, h_2, \dots, h_K)$ the probability of $n_r = (0, 1, \dots, h_r)$ customers of type r still being served at time t is distributed binomially by

$$b_{h_r}(n_r, t-u) = \binom{h_r}{n_r} (1-G_r(t-u))^{n_r} G_r(t-u)^{h_r-n_r}, \quad 0 \leq n_r \leq h_r.$$

Conditioning upon the size of the marks arrivals, the total rate of $\mathbf{n} = (n_1, n_2, \dots, n_K)$ customers arrivals during the time interval du which are still in service at time t is

$$K_n(u, t) = \sum_{h=n}^{\infty} D_h(u) \prod_{r=1}^K b_{h_r}(n_r, t-u) = \sum_{h=n}^{\infty} D_h(u) \prod_{r=1}^K \binom{h_r}{n_r} (1-G_r(t-u))^{n_r} G_r(t-u)^{h_r-n_r}. \quad (2)$$

The probabilities $R(\mathbf{n}, u, t)$ satisfy the system of differential equations

$$\begin{aligned} \frac{d}{du} R(\mathbf{0}, u, t) &= R(\mathbf{0}, u, t) K_0(u, t) + v \sum_{k \geq 1} R(\mathbf{k}, u, t), \\ \frac{d}{du} R(\mathbf{n}, u, t) &= R(\mathbf{n}, u, t) (K_0(u, t) - v(u)I) + \sum_{k \geq 1} R(\mathbf{k}, u, t) K_{\mathbf{n}-\mathbf{k}}(u, t) \end{aligned} \quad (3)$$

with initial conditions $R(\mathbf{0}, u, 0) = I$ and $R(\mathbf{n}, u, 0) = \mathbf{0}$ where I and $\mathbf{0}$ are unite and null matrices.

Let $\tilde{R}(\mathbf{z}, u, t)$ be the PGF of the number of customers being in service at the moment t .

$$\tilde{R}(\mathbf{z}, u, t) = \sum_{\mathbf{n}=\mathbf{0}}^{\infty} \mathbf{z}^{\mathbf{n}} R(\mathbf{n}, u, t), \quad |z_1| \leq 1, |z_2| \leq 1, \dots, |z_K| \leq 1.$$

Then from (3) for $\tilde{R}(\mathbf{z}, u, t)$ we derive

$$\frac{d}{du} \tilde{R}(\mathbf{z}, u, t) = \tilde{R}(\mathbf{z}, u, t) (\tilde{K}(\mathbf{z}, u, t) - v(u)I) + v(u) \tilde{R}(\mathbf{I}, u, t) \quad (4)$$

with initial conditions $\tilde{R}(\mathbf{z}, u, 0) = I$.

Here

$$\tilde{K}(\mathbf{z}, u, t) = \sum_{\mathbf{n}=\mathbf{0}}^{\infty} \mathbf{z}^{\mathbf{n}} \sum_{h=\mathbf{n}}^{\infty} D_h(u) \prod_{r=1}^K b_{h_r}(n_r, t-u) = \sum_{\mathbf{n}=\mathbf{0}}^{\infty} D_h(u) \prod_{r=1}^K [z_r(1-G_r(t-u)) + G_r(t-u)]^{n_r}$$

Hence, the solution of equation (4) can be presented in matrix-exponential form

$$\tilde{R}(\mathbf{z}, t) = e^{\int_0^t [\tilde{K}(\mathbf{z}, u, t) - v(u)I] du} I + \int_0^t v(u) e^{\int_0^u [\tilde{K}(\mathbf{z}, u, t) - v(u)I] du} \tilde{R}(\mathbf{I}, u) du, \quad |\mathbf{z}| \leq 1.$$

where the matrix $\tilde{R}(\mathbf{I}, t)$ defines from the differential equation

$$\frac{d}{du} \tilde{R}(\mathbf{I}, u, t) = D(u) \tilde{R}(\mathbf{I}, u, t)$$

and has a solution

$$\tilde{R}(\mathbf{I}, u, t) = e^{\int_0^u D(x) dx}.$$

Let consider some particular cases. The homogeneous model $MAP|G|\infty$ with MAP arrival of customer and non-stationary Poisson arrival of disasters with the rate $v(t)$. Suppose that MAP is given by characteristic matrices C, D size of $m \times m$. Then for the PGF of this model $\tilde{R}(\mathbf{z}, t)$ we have

$$\tilde{R}(\mathbf{z}, t) = e^{\int_0^t \{C + D[G(x) + z(1-G(x))] - v(x)I\} dx} \left(I + \int_0^t v(y) e^{\int_0^y \{D(1-z)(1-G(x)) + v(x)I\} dx} dy \right).$$

Thus for probabilities $R(\mathbf{n}, t)$ we obtain

$$R(n, t) = e^{\int_0^t \{C+DG(x)-v(x)I\}dx} \frac{\left(\int_0^t D(1-G(x))dx \right)^n}{n!} + \int_0^t v(y)e^{\int_y^t [(C+D)t-(D(1-G(x))+v(x)I)dx]} \frac{\left(\int_y^t D(1-G(x))dx \right)^n}{n!} dy .$$

For the first and second moments $m_1(t)$, $m_2(t)$ of the number of busy servers we obtain

$$\begin{aligned} \frac{d}{dt} m_1(t) &= [C + D - v(t)]m_1(t) + D(1 - G(t))e^{[C+D]t} \\ \frac{d}{dt} m_2(t) &= [C + D - v(t)]m_2(t) + 2D(1 - G(t))m_1(t) \end{aligned}$$

with the initial conditions $m_1(0) = 0$, $m_2(0) = 0$.

$$\begin{aligned} m_1(t) &= \int_0^t e^{\int_0^y [C+D-v(x)I]dx} D e^{[C+D]y} (1-G(y)) dy, \\ m_2(t) &= 2 \int_0^t e^{\int_0^y [C+D-v(x)I]dx} D m_1(t) (1-G(y)) dy. \end{aligned}$$

The obtained results may be applied for estimating characteristics, as well as searching optimal strategies for managing resources of wide class of subsystems of NGN, whereas the model $MMA P_k(t) | G_k | \infty$ may be used as a model of these subsystems.

REFERENCES

1. Kouvatsos, D. D. Network Performance Engineering: A Handbook on Convergent Multi-Service Networks and Next Generation Internet. Published by Springer-Verlag NY Inc, 2011
2. Tripathi, S.K., Sharma P., Raghavan S.V. (2006) Challenges in Design of Next Generation Networks. Modeling and Simulation Tools for Emerging Telecommunication Networks, pp 19-42
3. Ma Sheng, and Ji Chuanyi, Modeling Heterogeneous Network Traffic in Wavelet Domain. IEEE/ACM Transactions On Networking, Vol. 9, No. 5, October 2001
4. Park K., Kim G., Crovella M. (1997) On the Effect of Traffic Self-Similarity on Network Performance, Proc. SPIE Int'l. Conf. Perf. and Control., p. 296–310.
5. Paxson V., Floyd S. (1994) Wide-area Traffic: The Failure of Poisson Modeling. Proc. Of the ACM94, p. 257–268.
6. Artalejo J.R., Gomez-Corral A., He Q.M. Markovian arrivals in stochastic modelling: a survey and some new results. SORT 34 (2) July-December 2010, 101-144
7. Breuer L. Spatial Queues. Ph.D. Thesis, University of Trier, Germany, 2000.
8. Baum D., Kalashnikov V. (2004) Spatial No-Waiting Stations with Moving Customers Queueing Systems, v. 46, p. 231–247.
9. Ramaswami V., Neuts M.F. (1980) Some Explicit Formulas and Computational Methods for Infinite-Server Queues with Phase-Type Arrival. J. Appl. Prob., № 17, 498–514.
10. Blom J., M. Mandjes, And H. Thorsdottir (2013). Time-scaling limits for Markov-modulated infinite-server queues. Stochastic Models, 29, 112–127.
11. Choi B.D. The $M_k/M/\infty$ Queue With Heterogeneous Customers In A Batch, J. Appl. Prob. 29, 477-481 (1992)
12. Falin G. (1994) The $M_k/G/\infty$ Batch Arrival Queue with Heterogeneous Dependent Demands. Journal of Applied Probability, Vol. 31, No. 3, pp. 841-846

13. Masuyama H. Studies on Algorithmic Analysis of Queues with Batch Markovian Arrival Streams. PhD Thesis, Kyoto University, 2003, 146 p.
14. Keilson J., Servi L.D. Networks of Non-homogenous $M|G|\infty$ Systems. GTE Laboratories. Waltham, MA, 1989.
15. Eick S., Massey W., Whitt W. (1993) The Physics of the $M(t)/G/\infty$ Queue, Operations Research 41(4), pp.731–742
16. Stidham, S. (1974) Stochastic clearing systems. Stoch. Proc. Appl. 2, 85-115.
17. Purdue, P. And Linton, D. (1981) An infinite-server queue subject to an extraneous phase process and related models. J. Appl. Prob. 18, 236-244.
18. Linton, D. And Purdue, P . (1979) An $M/G/\infty$ queue with m customer types subject to periodic clearing. Opsearch 16, 80-88.
19. Böhm, Walter (2008) A Note on Queueing Systems Exposed to Disasters. Research Report Series/ Department of Statistics and Mathematics, 79. Department of Statistics and Mathematics, WUVienna University of Economics and Business, Vienna.
20. Economou A. and Fakinos D. (2008) Alternative approaches for the transient analysis of Markov chains with catastrophes. Journal of Statistical Theory and Practice, Vol. 2, Issue 2.

Synchronisation semantics for networks with fork and join: a product-form approach

Simonetta Balsamo and Andrea Marin

DAIS
Università Ca' Foscari Venezia
via Torino, 155, 30170
Venice, Italy
{balsamo,marin}@dais.unive.it

Abstract. In this paper we survey and compare different semantics for models' synchronisations that may lead to separable solutions for the stationary state probabilities. Indeed, although in the literature the synchronisation semantics for timed stochastic models have been widely studied in terms of expressive power, fewer efforts have been devoted to investigate of whether the modularity that is achieved by the formalism thanks to its synchronisation semantics is then reflected to a possible modular analysis. We discuss the basilar semantics for queue interactions (arrival and departure events) and more complex behaviours like the fork/join constructs inspired by stochastic Petri nets and the synchronisations inspired by the G-networks.

1 Introduction

The complexity of modern computer networks requires complicated models for the performance evaluation possibly consisting of several components that interact in some way. The definition of a general purpose formalism with a strong semantics that allows for the specification of stochastic models with underlying Continuous Time Markov Chains (CTMCs) has been object of many research efforts (see ,e.g., [11, 12, 8, 7]). However, in general the compositionality which is gained by the formalism is lost when one tries to derive the models' performance indices. For the steady-state analysis, when a model admits a stationary distribution that can be expressed in terms of products of the stationary distributions of its isolated elementary components we say that it is in *product-form*. It is worth of notice that although independent models are clearly always in product-form, we are interested in the cases where they are stochastically dependent and hence the exact result on the decomposition holds only in steady-state. The main advantage of product-form models is a drastic reduction of the computational state and space complexity of the algorithms for the stationary distribution (and consequently for the performance indices) computation. The drawback is clearly that not all the models admit a product-form solution. For this reason many papers have been devoted to the characterisation of product-form models at the stochastic process level of abstraction (e.g., [9, 5]) or at the formalism level of

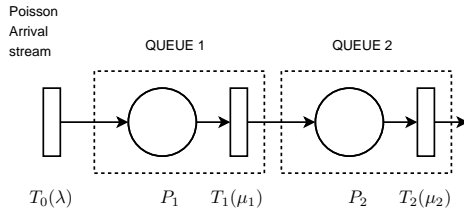


Fig. 1. Tandem of two exponential queues represented in SPNs.

abstraction (e.g., [3, 1, 10]). In this paper we focus on an important aspect of product-form model specification: the synchronisation semantics. Indeed, while it is well-known that Jackson’s networks admit separable equilibrium distributions [3], more recently more flexible synchronisations with product-forms have been studied. The recent development of a lower level and general theorem for the product-form analysis, the Reversed Compound Agent Theorem (RCAT) [5], allows us to compare the more recently proposed synchronisations, e.g. modelling networks with fork and join, under the view of obtaining a product-form stationary distribution. We show that the importance of understanding the nature of synchronisation among different model’s components (especially for what concerns the fork and join semantics) has a strong impact on the steady-state distribution of the stochastic models and hence their performance indices.

2 Basic synchronisations

The basic synchronisation is the pairwise. Let M be a model consisting of a set of cooperating components C_1, \dots, C_N , and let $\mathbf{m} = (m_1, \dots, m_N)$ be a state of M with m_i the state of component i . We say that M has pairwise synchronisations if in its underlying CTMC the existence of transition from state \mathbf{m}' to \mathbf{m}'' implies that \mathbf{m}' and \mathbf{m}'' differs for at most two components. Such a synchronisation can be modelled in PEPA process algebra [8] and using Stochastic Automata Network [12] with the idea of active and passive components. Figure 1 uses a stochastic Petri net formalism to represent a tandem of two exponential queues. Notice that this simplification is valid only for the sake of computing the distribution of the population in the two stations in steady-state. T_0 models the external arrivals that occur according to a Poisson process with intensity λ . P_1 and P_2 are associated with non-negative integer numbers representing the population of the queue and are changed by the firing on one of the three transitions. Customer service occurs at the first station with rate μ_1 and this causes an instantaneous migration of a customer to the second queue, and at the second station with rate μ_2 . The tandem model of Figure 1 has been known to be in product-form since the pioneering work of Burke, i.e., in stability the stationary probability of observing n_1 and n_2 customers in P_1 and P_2 , respectively, is $\pi(n_1, n_2) = (1 - \rho_1)(1 - \rho_2)\rho_1^{n_1}\rho_2^{n_2}$, with $\rho_1 = \lambda_1/\mu_1$ and $\rho_2 = \lambda_2/\mu_2$. Figure 2

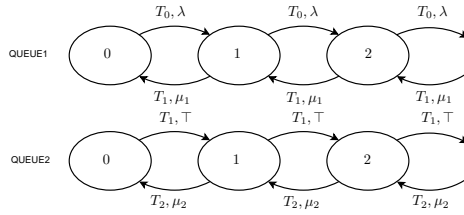


Fig. 2. Processes underlying the model of Figure 1.

shows the processes underlying the model of Figure 1. We label the transitions with some types that are in the model the names of the transitions. Transitions with the same type either occur jointly in the two models or cannot occur. Observe that transition T_1 causes a reduction of the number of customers in the first queue with rate μ_1 and increases the number of customers in the second queue with unknown rate denoted, as in PEPA, by the symbol \top . We say that the first queue is active with respect to T_1 , while the second is passive. We use the product-form theorem presented in [5] to obtain again Burke's product-form:

Theorem 1 (Reversed Compound Agent Theorem [5]). *The synchronisation of two models is in product-form if:*

1. *For every passive type, each state must have one outgoing transition with that type.*
2. *For every active type, each state must have one incoming transition with that type.*
3. *For every active type a , in the isolated model, the reversed rate associated with every active transition with type a must be the same, say $x_a \in \mathbb{R}^+$.*

The *isolated model* is the model where every occurrence of a passive type a with undefined rate \top is replaced with the corresponding reversed rate x_a . As regards the theory of the reversible and reversed Markov chain we refer to [9, 5]. Burke's product-form can be readily derived observing that the structural conditions of RCAT (cond. 1 and 2) are trivially satisfied. Observe that the first queue has an underlying birth and death process which is reversible and hence we easily obtain $x_{T_1} = \lambda$.

3 Modeling the join construct

Join constructs are used to model the simultaneous departures from two or more stations of some customers. In computer networks, given a set of transmitting stations on a shared channel, we may model each station state with a queue containing the packets waiting to be sent, and the collisions of two or more packets by means of join synchronisation. Ideally, the aim is modelling the SPN construct shown in Figure 3-(A) whose underlying process is depicted in Figure 3-(B). For

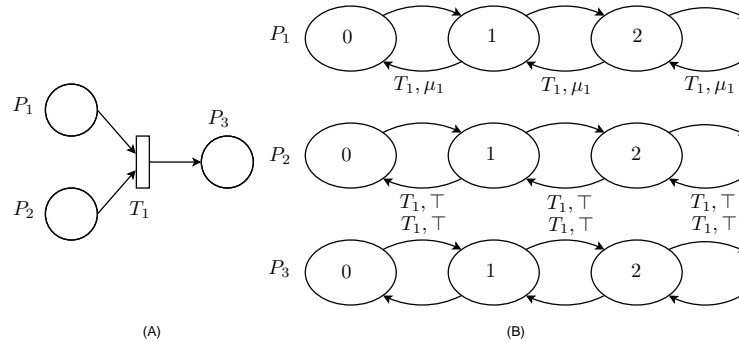


Fig. 3. The join construct in stochastic networks. (A)- SPN representation. (B)- Underlying processes

clarity, we just label the transitions involved in the synchronisations. Notice that the SPN approach has a simple process algebraic specification, where transitions labelled T_1 in P_1 and P_2 synchronise, and the joint transition maintain the same label to synchronise with P_3 . This can be specified also with the Kronecker's algebra in a straightforward way. Despite of the simple modular specification, this type of synchronisations have a product-form with strict conditions (see [2, 10]) that may be rate-dependent. A similar synchronisation has been defined by Gelenbe in the G-networks [4]. An inspection of Figure 3-(B) immediately shows that the product-form of the net cannot be proved by RCAT since the second structural condition is violated by state 0 in P_2 . Therefore, we may define another semantics for the synchronisation, inspired by [4] in which the process underlying P_2 and P_3 are those shown by Figure 4. Notice that in this case every state of P'_2 has an outgoing passive transition labelled T_1 . This allows use to conclude that the synchronisation between P'_1 and P'_2 are in product-form. However, observe that transitions labelled T_1 in P_2 do not satisfy RCAT conditions once they synchronise as active with P'_3 : indeed, state 0 of P'_2 has two incoming transitions labelled T_1 thus violating condition 2 of Theorem 1. The workaround consists in not allowing the self-loop transition in P'_2 to synchronise with those in P_3 by a renaming (which is shown in the picture using the double name). Now, the joint model is in product-form (a sound theory of this approach can be found in [6]).

Observe that the two semantics of synchronisation proposed in Figures 3 and 4 are different. In particular, if we consider a state (n_1, n_2, n_3) in which $n_1 \geq 0$, $n_2 > 0$ and $n_3 \geq 0$ the outgoing transitions in the CTMC underlying the joint model are the same. Nevertheless, when $n_2 = 0$ the two models have different behaviours. Indeed, while in the model of Figure 3 transition T_1 is disabled, in the model of Figure 4 we observe a departure from P_1 which synchronises with the self-loop of P'_2 and hence does not cause an arrival at P_3 . If we assume Poisson arrival streams at P_1 , P_2 and P_3 with intensities λ_1 , λ_2 , λ_3 , respectively,

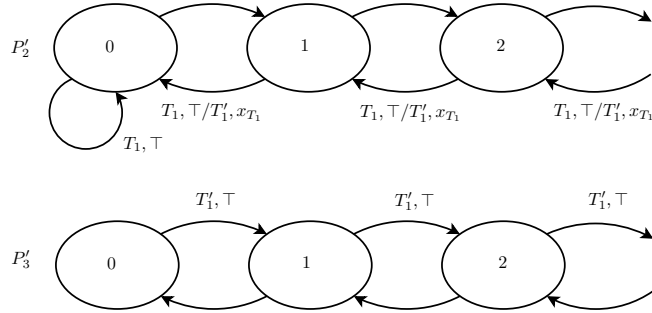


Fig. 4. Alternative semantics for join construct.

independent exponential distributed service times with rates μ_1, μ_2, μ_3 and a rate for the join transition T_1 of χ , then the product-form expression for the model of Figure 4 is $\pi(n_1, n_2, n_3) = (1 - \rho_1)(1 - \rho_2)(1 - \rho_3)\rho_1^{n_1}\rho_2^{n_2}\rho_3^{n_3}$, where:

$$\rho_1 = \frac{\lambda_1}{\mu_1 + \chi}, \quad \rho_2 = \frac{\lambda_2(\mu_1 + \chi)}{\mu_2(\mu_1 + \chi) + \lambda_1\chi}, \quad \rho_3 = \frac{\lambda_2(\mu_1 + \chi) + \lambda_3(\mu_2(\mu_1 + \chi) + \lambda_1\chi)}{\mu_3(\mu_2(\mu_1 + \chi) + \lambda_1\chi)}.$$

Under the same assumptions, the model of Figure 3 is not in product-form. For the modelling point of view, there are some other important differences. First, observe that while the SPN synchronisation may have a conservation law for the customers and hence can be applied to study closed models, the G-network must be used for an open model. Indeed, the departure of a customer from P_1 when P'_2 is empty causes a customer loss that must be compensated in some way, otherwise in steady-state the population of the network will tend to 0. A second aspect is that the SPN model is symmetric with respect to P_1 and P_2 , i.e., they can be interchanged without affecting the CTMC underlying the SPN. Conversely, the role of P_1 and P'_2 are different. In particular if one desires to approximate the SPN model by means of a G-network, it is reasonable to choose the node that is generating the join signal as that with lower load factor with the aim of reducing the probability of observing state 0 in P'_2 . Clearly, this choice may be driven by some simulations or prior knowledge of the model.

4 Modelling to fork construct

Fork construct has many applications at the system-level especially in the field of parallel and concurrent programming. For instance, many dynamic web portals generate the user content according to some static information stored in the disk and some dynamic which is generated by a query to a database.

In our context, modelling the fork construct can be seen as a dual problem of the join.¹

¹ Further details will be given in the extended version.

5 Conclusion

In this paper we investigated different synchronisation semantics in the view of obtaining a model with a product-form solution. Particular attention has been devoted to the modelling of fork and join constructs that are known to have strict conditions for the product-form. The work has shown that there is a tension between the accuracy of the model and its analytical tractability. Specifically, the G-network based synchronisation admits very general conditions for the product-form that are surely rate-independent, but on the other hand do not model explicitly the blocking behaviour which is expected in the join synchronisation. Stochastic Petri nets synchronisation type properly models the blocking of customers waiting for the join, but the product-form conditions on the net structure can be strict and (even worse) may be rate-dependent. This discussion aims at helping to modeller in the choice of the types of synchronisations that are suitable for the analysis purposes taking into account both the accuracy and the complexity of the performance indices derivation.

References

1. Balsamo, S., Harrison, P.G., Marin, A.: A unifying approach to product-forms in networks with finite capacity constraints. In: Proc. of the 2010 ACM SIGMETRICS Int. Conf. on Measurement and Modeling of Computer Systems. pp. 25–36. ACM, New York, NY, USA (2010)
2. Balsamo, S., Harrison, P.G., Marin, A.: Methodological Construction of Product-form Stochastic Petri-Nets for Performance Evaluation. *J. of System and Software* 85(7), 1520–1539 (2012)
3. Baskett, F., Chandy, K.M., Muntz, R.R., Palacios, F.G.: Open, closed, and mixed networks of queues with different classes of customers. *J. ACM* 22(2), 248–260 (1975)
4. Gelenbe, E.: G-networks with triggered customer movement. *J. of Appl. Prob.* 30, 742–748 (1993)
5. Harrison, P.G.: Turning back time in Markovian process algebra. *Theoretical Computer Science* 290(3), 1947–1986 (2003)
6. Harrison, P.G., Marin, A.: Product-forms in multi-way synchronizations. *The Computer Journal* p. To appear (2013)
7. Hermanns, H.: *Interactive Markov Chains: The Quest for Quantified Quality*. LNCS 2428, Springer (2002)
8. Hillston, J.: *A Compositional Approach to Performance Modelling*. Cambridge Press (1996)
9. Kelly, F.: *Reversibility and stochastic networks*. Wiley, New York (1979)
10. Marin, A., Balsamo, S., Harrison, P.: Analysis of stochastic Petri nets with signals. *Perf. Eval.* 85(7), 1520–1539 (2012)
11. Molloy, M.K.: Performance analysis using stochastic Petri nets. *IEEE Trans. on Comput.* 31(9), 913–917 (1982)
12. Plateau, B.: On the stochastic structure of parallelism and synchronization models for distributed algorithms. *SIGMETRICS Perf. Eval. Rev.* 13(2), 147–154 (1985)

The GE/GE/1/N Queue Subject to State-Dependent Arrival Balking

Neelkamal Shah¹, Demetres Kouvatsos²³

NetPEN – Networks and Performance Engineering Research Unit
Informatics Research Institute (IRI)
University of Bradford, UK

Abstract. New links between discrete maximum entropy (ME) distributions and Markov chains are conjectured towards the stochastic analysis of general single server queues with or without arrival balking. An ME solution, the truncated generalised discrete Half Normal ($GdHN_T$) distribution is characterised, subject to appropriate mean value constraints, particularly for inferences of the stationary queue length distributions (QLD's) of stable ordinary G/G/1/N queues or stable G/G/1/N queues subject to extended Morse-type balking. The GE/GE/1/N queue, characterised by bursty generalised exponential (GE) inter-arrival and service times and finite capacity, N is analysed subject to state-dependent arrival balking under three different balking-blocking policies and their QLD's are derived via the global balance (GB) solution technique. In this context, conjectures are proposed, based on comprehensive numerical experimentation, that the latter GB solutions, subject to extended Morse balking are special cases of the $GdHN_T$ ME distribution. Remarks on open research problems emerging from this work are included.

Keywords: Queues, Balking, Maximum Entropy (ME), Global Balance (GB), Queue Length Distribution (QLD), Truncated Generalised discrete Half Normal ($GdHN_T$), Truncated Generalised Geometric ($GGeo_T$), Generalised Exponential (GE).

1 Introduction

Queueing models with or without arrival balking constitute fundamental investigative tools for the performance modelling and analysis of discrete flow systems such as business, transport and production systems and communication networks.

Balking is the immediate rejection of service by arriving customers to a queue anticipating an unacceptably long queueing time and/or their service is not required urgently [1]. Queues with balking have been successfully applied to evaluate multiple real life systems such as telephone call centres and business service systems where

1 npsah@student.bradford.ac.uk

2 D.Kouvatsos@bradford.ac.uk

3 Corresponding Author

customers gain information about anticipated delay, transportation traffic-flow control [2-7], admission control mechanisms [8] and threshold-type customer acceptance/rejection policies dependent on system workload [2, 9, 10] among others. Nevertheless open problems posing analytic challenges in their modelling and evaluation still exist.

The technical material of this paper is founded largely on the works of Morse [11], Kemp [12-14], El-Affendi and Kouvatso [15] and Kouvatso [16-18], which analyse single server finite/infinite - capacity queues in equilibrium, with or without balking, via the Markov chain approach, global balance (GB) solution technique and/or the information theoretic principle of maximum entropy (ME).

The ME principle, attributable to Jaynes [19], provides a method of inference to characterise a unique probability distribution, free from bias, of a random quantity of any general system based on prior partial knowledge of the quantity in terms of its moments. The information theoretic entropy functional, representing the measure of uncertainty associated with the random quantity, which is maximised in this paper is attributable to Shannon [20] and the maximisation is carried out using the Lagrangian multiplier technique.

ME has been employed in the modelling of systems in a vast range of areas from economics and finance to transportation, linear and nonlinear programming, queueing and operational research [21]. In particular, ME solutions have been derived for various types of queueing systems including among others single server queues at equilibrium [15-18, 22], multiserver queues [23, 24], multiple class queues with priorities [25], queues with vacation [26-28], queueing network models (QNM's) [29-32] and queues subject to arrival balking [12, 14, 33].

In the context of this work, the ME principle is employed in the discrete domain to derive a generalised ME solution, namely the truncated generalised discrete half normal (GdHN_T) distribution when the prior information known about the quantity is the first moment, variance, boundary state probabilities p_0^N and p_N^N and the normalising condition. Subsequently, the GE/GE/1/N queue, characterised by bursty generalised exponential (GE) inter-arrival and service times and finite capacity, N is analysed subject to arrival balking under three different balking-blocking policies. The GE model provides a justifiable, cost-effective means to analytically model inter-event durations marked by square coefficient of variation, C^2 greater than one, a property which has been observed in numerous telecommunication systems. A conjecture is proposed, via extensive numerical experimentation, that the queue length distributions (QLD's) of the GE/GE/1/N queue subject to extended Morse arrival balking under the three different policies are special cases of the GdHN_T ME distribution.

The GE/GE/1/N subject to extended Morse balking serves as an ME-significant model from which queue metrics such as the waiting, queueing or busy times, number served in a busy period or balking rates can be justifiably inferred for queues with balking (or appropriate metrics for equivalent queues with state-dependent arrival rates) when the aforementioned prior information of the queue is known. On the other hand, this queueing model comprises a building block component in the approximate product form ME analysis of arbitrarily-configured non-exponential QNM's [32] comprising queues with balking. Furthermore these least biased ME QLD's can

prove useful in applications where the full QLD is required for example in optimal queueing network control [34].

The paper is organised as follows: In Section 2 existing ME solutions of queueing systems pertinent to this work are reviewed and subsequently in Section 3, the GdHN_T ME distribution is derived. In Section 4, the QLD's of the GE/GE/1/N queue subject to extended Morse balking under the three policies are derived and their equivalence to the GdHN_T is conjectured followed by the conclusions and remarks on future work in Section 5.

2 ME Solutions of Queueing Systems

This section presents a review of ME solutions of queueing systems in the literature which are relevant to the results in this paper.

An ME solution has been proposed for the QLD of the M/G/1 queue in [22, 35] assuming prior information of the mean queue length (MQL), which is taken as the celebrated Pollaczek-Khinchin MQL formula for an M/G/1 queue, and the normalisation condition. This effectively models the QLD of the M/G/1 queue by that of an M/M/1 with the same MQL. A more general ME solution for the QLD of a stable M/G/1 queue, the generalised geometric (GGeo) ME QLD, was published in [15] by explicitly incorporating the queue stability (or flow balance) condition (2.1) as the 'p₀' prior information constraint in addition to the MQL and normalisation condition.

$$\lambda = \mu(1 - p_0) \quad (2.1)$$

Where λ and μ are the mean arrival and service rates respectively and p_0 is the fraction of time that the queue is empty. Analysis, based on the ME principle, has also been carried out for the QLD's of the finite capacity M/G/1/N and G/G/1/N queues [16, 17]. For the G/G/1/N queue, the stability condition is expressed as follows (2.2):

$$\lambda(1 - p_N^N) = \mu(1 - p_0^N) \quad (2.2)$$

Where p_N^N and p_0^N are the proportions of time that the finite capacity queue is full and empty respectively, λ is the prospective arrival rate (a fraction of which is blocked from the queue when full) and μ is the service rate. Therefore in order to incorporate the stability condition (2.2), both boundary state probabilities, p_0^N and p_N^N are required in addition to λ and μ .

The discrete distribution having maximum entropy subject to the prior information constraints of the first moment, $E[N_N]$ and boundary state probabilities p_0^N and p_N^N normalised over finite non-negative integer support $[0, N]$ is the truncated GGeo (GGeo_T) (2.3) i.e. the GGeo right-truncated above N and constrained additionally by the upper boundary state probability [16, 17]:

$$p_n^N = \begin{cases} p_0^N, n = 0 \\ p_0^N y_N x_N^n, n = 1, 2, 3 \dots N - 1 \\ p_0^N z_N y_N x_N^n, n = N \end{cases} \quad (2.3)$$

Where the parameters x_N , y_N and z_N correspond to the prior information constraints $E[N_N]$, p_0^N and p_N^N respectively. Unlike the GGeo [18] however, the parameters of the GGeo_T cannot be explicitly expressed in terms of the prior moment information however they can be determined numerically using parameter estimation techniques.

Analogous to their infinite capacity counterparts, the M/G/1/N, G/M/1/N and G/G/1/N queues bearing the GGeo_T ME QLD are satisfied exactly when the general distribution 'G' is specified as the GE, a two parameter mixed distribution comprising a continuous exponential component and a discrete component at the origin and completely defined in terms of its first two moments [15]. The GE CDF can be interpreted as the inter-event time distribution of a compound Poisson process (CPP) with geometrically distributed batch sizes [32].

In the ME analysis of QLD's (marginal or joint) of different queueing systems, in addition to the normalising condition, the MQL (marginal or joint) is almost always included as part of the prior information. Other prior information assumed known include, among others, further queue length (ordinary or central) moments [12, 22, 33, 36], server utilisation (marginal or joint) [16, 18, 31, 36], full buffer probability (marginal or joint) [16, 37] and the set of state probabilities $p_n, n = 0, 1, \dots, c$ in an ME solution of the QLD of a multiserver queue [23].

3 A Generalised Maximum Entropy Solution

Motivated by queueing system modelling, in this section, the ME principle is used to derive generalised least biased probability assignments for discrete values realised by a random quantity of a general system when the prior information known about the quantity is its first moment, variance, the boundary state probabilities p_0^N and p_N^N and the normalising condition.

The discrete Half Normal (dHN) has been characterised as that discrete distribution having maximum entropy constrained by the first moment and variance and normalised over non-negative integer support [13, 14] and it has been defined as in (3.1) below:

$$p_n = p_0 \theta^n q^{\frac{n(n-1)}{2}}, n = 0, 1, 2 \dots, \theta > 0, 0 < q < 1 \quad (3.1)$$

In the interest of consistency in nomenclature, the generalisation of the dHN ME distribution by inclusion of the additional prior information constraints of p_0^N and p_N^N in the ME derivation is referred to here as the truncated generalised dHN (GdHN_T) ME distribution analogous to the case of the GGeo_T.

Maximising Shannon's entropy functional subject to the first moment, variance and boundary state probabilities, p_0^N and p_N^N prior information constraints and normalised over finite non-negative integer support $[0, N]$ yields the GdHN_T (3.2):

$$p_n^N = \begin{cases} p_0^N, n = 0 \\ p_0^N \gamma_N \phi_N^n r_N^{\frac{n(n-1)}{2}}, n = 1, 2, 3 \dots N-1, \zeta_N, \gamma_N, \phi_N, r_N > 0 \\ p_0^N \zeta_N \gamma_N \phi_N^N r_N^{\frac{N(N-1)}{2}}, n = N \end{cases} \quad (3.2)$$

Where again the respective parameters of the GdHN_T cannot be explicitly represented in terms of the prior information constraints however they can be determined numerically from the available information.

The GdHN arises as a special case of the GdHN_T when $N \rightarrow \infty$. Furthermore, setting $r_N = 1$ retrieves the GGeo_T and setting both $\gamma_N = 1$ and $\zeta_N = 1$ in (3.2) yields the truncated dHN (dHN_T) ME distribution i.e. the dHN right-truncated above N.

4 A Maximum Entropy Significant Queueing System

In this section, the queueing system(s) whose QLD's are exact special cases of the GdHN_T ME distribution are sought where in this case the prior information constraints are the MQL, variance of queue length (VQL), QLD probabilities p_0^N and p_N^N and the normalising condition over finite non-negative integer support $[0, N]$.

The irrational probability generating function of the GdHN invalidates its use in the celebrated Pollaczek-Khinchin transform formula to determine, by inversion of its Laplace-Stieltjes transform, the unique service time distribution of the M/G/1 queue with ME QLD the GdHN (and by extension the service-time distribution of the M/G/1/N queue with ME QLD the GdHN_T and by further analogies the inter-arrival and service time distributions of the ordinary G/G/1 and G/G/1/N queues). In light of this limitation an alternative approach had to be used.

Notably, Kemp discovered specific state-transition probabilities of a Markov chain model which result in the dHN steady state probability distribution (3.1) and observed that this equilibrium distribution has as a special case, the QLD of the M/M/1 queue subject to Morse arrival balking [11, 12, 14]. At its heart, the Morse balking model comprises the following instantaneous exact-workload dependent joining function (4.1):

$$q(t) = e^{-\alpha t}, t \geq 0 \quad (4.1)$$

Where α was interpreted by Morse as a measure of the average impatience (unwillingness to wait in line) of customers [11] and t is the instantaneous exact-workload. For most practical purposes though, the instantaneous exact-workload is not known however the instantaneous average-workload may be known resulting in the following population-dependent function (4.2) when service is exponential:

$$q(n) = e^{-\frac{\alpha n}{\mu}}, n = 0, 1, 2 \dots \quad (4.2)$$

Where n is the queue population at the instant of arrival and μ is the service rate. The QLD of the M/M/1 queue subject to Morse balking was derived by analysing a Birth-Death (B-D) chain, via the GB technique, of an equivalent M(n)/M/1/N queue whose state-dependent arrival rates were adjusted to account for the balking resulting in the QLD (4.3) below [11]:

$$p_n = p_0 \left(\frac{\lambda}{\mu}\right)^n e^{-\frac{\alpha n(n-1)}{\mu}}, n = 0, 1, 2 \dots \quad (4.3)$$

Kemp observed that the QLD of the M/M/1 queue subject to Morse balking (4.3) is a special case of the dHN (3.1) when $\theta = \lambda/\mu$ and $q = \exp(-\alpha/\mu)$ [11, 12, 14]. This result implies that incorporating the VQL prior information constraint is equivalent to maximising the additional uncertainty in the ME solution arising from Morse arrival balking, since maximising entropy subject to solely the MQL and normalising condition yields the QLD of the ordinary M/M/1 queue.

Given the value of a suitable performance metric including among others the average balking rate (BR), MQL or mean waiting time and assuming that both λ and μ are known, α can be determined quantitatively by solving a nonlinear equation analogous to that for the case when BR is known as in (4.4) below.

$$\lambda \sum_{n=0}^{\infty} q(n)p_n - (\lambda - BR) = 0 \quad (4.4)$$

Where $q(n)$ and p_n are given by (4.2) and (4.3) respectively.

It is seen that two subclass distributions of the GdHN_T, namely the GGeo_T and dHN satisfy exactly QLD's of queueing systems characterised by GE inter-event times and Morse arrival balking respectively. Therefore, the queueing system bearing the GdHN_T ME QLD was sought by amalgamating these two properties, hence the GE/GE/1/N queue subject to extended Morse balking is analysed below.

4.1 The GE/GE/1/N Queue Subject to Arrival Balking

Within the context of batch arrivals subject to balking and finite capacity queues, three different combinations of balking and blocking policies are identified and analysed in this paper drawn from the batch arrival MG/M/1/N queue subject to state dependent batch arrival rates (i.e. the MG(n)/M/1/N queue) studied in [38]. In that work, the celebrated complete and partial batch blocking policies (cf. [35]) are modelled and the independent and group (or batch) behaviour of members of batches with respect to arrival rates are distinguished with solely the batch behaviour modelled. In this paper three (composite) policies are analysed subject to the general population-dependent joining function, $q(n)$ which is conditional on the queue population, n at

the instant of arrival of a batch and all the members of a batch are considered to ‘see’ the same instantaneous queue size, n . The policies are described below:

1. Complete batch balking and complete batch blocking.

The balking and blocking behaviour of each of the members of a batch is identical resulting in the batch behaving as a single entity. This policy can be interpreted as unity between members of a batch being preserved throughout the system operation i.e. whether or not there is sufficient capacity.

2. Complete batch balking and partial batch blocking.

The balking behaviour of each of the members of a batch is identical however this unity is upheld solely when there is sufficient capacity for the entire batch and is surrendered when there is limited capacity to accommodate the entire batch. Under the latter restrictive condition, following the (positive) decision of the arriving batch to join the queue, as many customers as can be accommodated join the queue successively from the head of the batch.

3. Independent batch balking and partial batch blocking.

Members of a batch behave independently with respect to balking whereby each successive member of a batch makes an autonomous decision to join or balk. Following a positive join decision, batch members proceed to occupy successive positions in the queue until it becomes full, after which subsequent members are blocked. This policy can be interpreted as the customer autonomy upheld irrespective of the level of resource availability.

Following the solution technique employed for the exponential case in the literature for example in [11], the GE/GE/1/N subject to balking under the three balking-blocking policies is solved by balancing the global flow between states of the Markov chain models of the equivalent GE(n)/GE/1/N queue where the operations of the different policies are accounted for in the different upward state transition rates. The following recursive- (policies I and III) and closed- (policy II) form QLD’s result, where λ and $\sigma\lambda$ are the single or batch prospective arrival rates respectively, μ and $\tau\mu$ are the individual and batch service rates respectively and $q(n)$ is the joining function:

1. The QLD of the GE/GE/1/N subject to balking under policies I and III

$$p_n^N = \begin{cases} p_0^N, n = 0 \\ p_0^N \frac{(1-\tau)R_{01} + \tau \sum_{j=1}^N R_{0j}}{\tau\mu + (1-\tau) \sum_{j=2}^N R_{1j}}, n = 1 \\ \frac{(1-\tau) \sum_{i=0}^{n-1} R_{in} p_i^N + p_{n-1}^N (\tau\mu + \sum_{j=n}^N R_{n-1,j}) - \sum_{i=0}^{n-2} R_{i,n-1} p_i^N}{\tau\mu + (1-\tau) \sum_{j=n+1}^N R_{nj}}, 2 \leq n \leq N-1 \\ \frac{\sum_{i=0}^{N-1} R_{iN} p_i^N}{\tau\mu}, n = N \end{cases} \quad (4.5)$$

Where the upward state transition rates, the R_{ij} 's are omitted due to space limitations.

2. The QLD of the GE/GE/1/N subject to balking under policy II

$$p_n^N = \begin{cases} p_0^N \frac{\sigma\lambda q(0)\tau}{(\sigma\lambda q(1)(1-\tau) + \tau\mu)} \prod_{i=2}^n \left(\frac{(1-\sigma)\tau\mu + \sigma\lambda q(i-1)}{\sigma\lambda q(i)(1-\tau) + \tau\mu} \right), & 1 \leq n \leq N-1 \\ p_0^N \frac{1}{\tau\mu} \left(\frac{(1-\sigma)\tau\mu + \sigma\lambda q(N-1)}{\sigma + \tau(1-\sigma)} \right) \left(\frac{\sigma\lambda q(0)\tau}{(\sigma\lambda q(1)(1-\tau) + \tau\mu)} \right) \prod_{i=2}^{N-1} \left(\frac{(1-\sigma)\tau\mu + \sigma\lambda q(i-1)}{\sigma\lambda q(i)(1-\tau) + \tau\mu} \right), & n = N \end{cases} \quad (4.6)$$

When the service durations satisfy any general distribution including the GE CDF, then the extended Morse joining function over finite support $[0, N-1]$ as defined in (4.7) below arises:

$$q(n) = \begin{cases} 1.0, & n = 0 \\ q \left(\frac{C_s^2 - 1}{2} \right) q^n, & n = 1, 2, 3 \dots N-1 \\ 0.0, & n = N \end{cases} \quad (4.7)$$

Where C_s^2 is the square coefficient of variation of the service time distribution and $q = \exp(-\alpha/\mu)$. In this case, the parameter α can be calculated in an analogous manner to that described earlier in this section for the M/M/1 queue subject to Morse balking.

4.2 Discussion of Results

Substitution of the extended Morse joining function (4.7) into (4.5) and (4.6) yields the QLD's of the GE/GE/1/N subject to extended Morse balking under the three policies. Failing algebraic proof of equivalence between the resulting QLD's and their corresponding GdHN_T ME inferences (3.2), numerical validation of this equivalence was sought by computing the absolute differences between the individual QLD probabilities and the corresponding state probabilities of the inferred GdHN_T solutions for each experiment. The GdHN_T solutions were derived numerically by maximising Shannon's entropy functional subject to the normalising condition and the following prior information constraints obtained from the QLD's: MQL, VQL, p_0^N and p_N^N .

Extensive numerical experimentation was carried out in MATLAB version 7.10.0.499 (R2010a) for a wide range of queue input-parameter values which are presented in Table 1 below.

Table 1. Parameter values used in the ME-significance validation of the QLD of the GE/GE/1/N subject to extended Morse balking under Policies I - III.

Parameter	Value(s)
-----------	----------

λ	[10 20 30 40]
μ	20
C_s^2	[1 5 10 20 50 100 200 500]
N	[5 10 15 20] (Policies I & II)
	[5 10] (Policy III)
q	[0.3 0.6 0.9]

The range of queue capacities used in the case of Policy III was smaller due to the memory and speed limitations of the PC used for the experiments.

Equivalence between the QLD of the GE/GE/1/N subject to extended Morse balking under policy I and the GdHN_T ME distribution was evidenced by a maximum absolute difference (error) between any two corresponding probabilities, over all experiments, of 0.018 with the overwhelming majority of maximum absolute differences per experiment less than 0.005. For policy II, the maximum absolute difference observed was 0.019 with the overwhelming majority of differences below 0.007. Under the third policy, the maximum absolute difference observed was 0.0277 with the overwhelming majority below 0.008. Moreover, earlier research showed that an approximate result would have been evidenced by mounting errors as C_s^2 increased [37], which did not occur on this occasion thus further supporting the case for the conjecture.

The cause of these errors was further investigated by conducting analogous experiments and computing the maximum absolute differences between the QLD probabilities and GdHN_T ME inferences of the following two cases whose equivalence has already been proven in the literature, namely the M/M/1/N queue subject to Morse balking [33] and the ordinary GE/GE/1/N queue [16, 17], whose respective QLD's are special cases of the GdHN_T. Absolute differences of up to 0.0066 and 0.000848 were observed respectively implying that the errors can be attributed to numerical limitations of the software package thus further supporting the conjecture of the above equivalence. Larger errors in the policy I-III experiments are attributed to the greater number of parameters involved.

When p_N^N was less than 1.0×10^{-10} , comparison had to be made against the GdHN ME QLD due to the limitation of the software package. Both the constraints and change in objective function were satisfied to within the default tolerance of 1.0×10^{-6} .

Supported by the above experimental evidence, the following conjecture is proposed:

Conjecture: *The QLD's of the GE/GE/1/N queue subject to extended Morse balking under Policies I, II or III (4.5) and (4.6) are special cases of the GdHN_T ME distribution constrained by the prior information of the queues' MQL, VQL (or equiva-*

lently second moment of queue length), empty state probability (or equivalently server utilisation), and full buffer state probability, normalised over finite integer support $[0, N]$.

5 Conclusions

In this paper, a generalised ME solution, namely the truncated generalised discrete Half Normal ($GdHN_T$), has been derived for least-biased probability assignments to the non-negative finite integer values realised by a discrete random quantity when the prior information known about the quantity is its first moment, variance, boundary state probabilities p_0^N and p_N^N and the normalising condition. Subsequently the GE/GE/1/N queue subject to arrival balking under three different balking-blocking policies was analysed and QLD's derived via GB analysis of their Markov chain models. When the balking was of extended Morse type, the QLD's were conjectured, based on extensive numerical experimentation, to be special cases of the $GdHN_T$ ME distribution.

In light of the contributions gained from the research work reported in this paper, potential avenues of future related work include investigating ME significance when a reward-cost structure is expressly modelled in the queue [39], analysing multiple class priority queues subject to arrival balking or generalising the model by including correlated service, multiple-servers and/or server vacation. Furthermore, the balking model readily lends itself to incorporation in the more general retrial queueing system.

Nevertheless, the novel queueing systems studied in this paper are envisaged to play the role, following analysis of the GE-type departing, splitting and merging traffic streams under balking, of cost-effective analytic building block models in the approximate ME analysis of queueing network models with arbitrary topology and balking [16, 17, 32, 37]. Moreover it provides a means for the justifiable inference of performance metrics such as queueing, waiting or busy times or loss rates of queueing systems featuring state-dependent arrival rates, given the appropriate prior information of the system(s) in concern within the broad range of contexts from current communication networks to manufacturing, transportation and business systems among others.

6 References

1. Haight, F.A., *Queueing with Balking*. Biometrika, 1957. **44**(3/4): p. 360-369.
2. Liu, L., *Service Systems with Balking Based on Queueing Time*, in Statistics and Operations Research. 2007, University of North Carolina at Chapel Hill: Chapel Hill, USA.
3. Whitt, W., Improving Service by Informing Customers About Anticipated Delays. Management Science, 1999. **45**(2): p. 192-207.
4. Jouini, O., Z. Akşin, and Y. Dallery, Call Centers with Delay Information: Models and Insights. Manufacturing & Service Operations Management, 2011. **13**(4): p. 534-548.

5. Whitt, W., Engineering Solution of a Basic Call-Center Model. *Management Science*, 2005. **51**(2): p. 221-235.
6. Guo, P. and P. Zipkin, Analysis and Comparison of Queues with Different Levels of Delay Information. *Management Science*, 2007. **53**(6): p. 962-970.
7. Mendelson, H., R.R. Pillai, and L.K. Jones, Inferring Balking Behavior From Transactional Data. *Oper. Res.*, 1999. **47**(5): p. 778-784.
8. Boxma, O.J. and B.J. Prabhu, Analysis of an M/G/1 Queue with Customer Impatience and an Adaptive Arrival Process. 2009, Eindhoven University of Technology: Eindhoven, The Netherlands. p. 17.
9. Combé, M.B., Impatient customers in the MAP/G/I queue. 1994, CWI, The National Research Institute for Mathematics and Computer Science, The Netherlands: Amsterdam. p. 1 - 11.
10. Lu, X. and B.L. Mark. Analytical modeling of optical burst switching with fiber delay lines. in *Modeling, Analysis and Simulation of Computer and Telecommunications Systems*, 2002. MASCOTS 2002. Proceedings. 10th IEEE International Symposium on. 2002.
11. Morse, P.M., *Queues, Inventories and Maintenance, The Analysis of Operational Systems with Variable Demand and Supply*. First ed. Publications in Operations Research, ed. D.B. Hertz. Vol. 1. 1958: John Wiley & Sons, Inc. 202.
12. Kemp, A.W., Steady-state Markov chain models for certain q-confluent hypergeometric distributions. *Journal of Statistical Planning and Inference*, 2005. **135**(1): p. 107.
13. Kemp, A.W., Characterizations of a discrete normal distribution. *Journal of Statistical Planning and Inference*, 1997. **63**(2): p. 223.
14. Kemp, A., The Discrete Half-Normal Distribution, in *Advances in Mathematical and Statistical Modeling*. 2008, Birkhäuser Boston. p. 353-360.
15. El-Affendi, M.A. and D.D. Kouvasos, A maximum entropy analysis of the M/G/1 and G/M/1 queueing systems at equilibrium. *Acta Informatica*, 1983. **19**(4): p. 339.
16. Kouvasos, D.D., Maximum entropy and the G/G/1/N queue. *Acta Inf.*, 1986. **23**(5): p. 545-565.
17. Kouvasos, D.D., A maximum entropy queue length distribution for the G/G/1 finite capacity queue, in *Proceedings of the 1986 ACM SIGMETRICS joint international conference on Computer performance modelling, measurement and evaluation*. 1986, ACM: Raleigh, North Carolina, USA. p. 224-236.
18. Kouvasos, D.D., A Maximum Entropy Analysis of the G/G/1 Queue at Equilibrium. *The Journal of the Operational Research Society*, 1988. **39**(2): p. 183-200.
19. Jaynes, E.T., Information Theory and Statistical Mechanics. *Physical Review*, 1957. **106**(4): p. 620.
20. Shannon, C.E., A Mathematical Theory of Communication. *The Bell System Technical Journal*, 1948. **27**: p. 379-423 and 623-656.
21. Fang, S.-C., J.R. Rajasekera, and H.-S.J. Tsao, Entropy optimization and mathematical programming. *International series in operations research & management science 1997*: Kluwer Academic Publishers, Boston.
22. Shore, J.E., Information theoretic approximations for M/G/1 and G/G/1 queueing systems. *Acta Informatica*, 1982. **17**(1): p. 43.
23. Wu, J.-S. and W.C. Chan, Maximum Entropy Analysis of Multiple-server Queueing Systems. *Journal of the Operational Research Society*, 1989. **40**(9): p. 815-825.
24. Kouvasos, D.D. and J. Almond, Maximum entropy two-station cyclic queues with multiple general servers. *Acta Inf.*, 1988. **26**(3): p. 241-267.

25. Kouvatso, D.D. and N. Tabet-Aouel, A Maximum Entropy Priority Approximation for a Stable G/G/1 Queue. *Acta Inf.*, 1989. **27**(3): p. 247-286.
26. Wang, K.-H., S.-L. Chuang, and W.-L. Pearn, Maximum entropy analysis to the N policy M/G/1 queueing system with a removable server. *Applied Mathematical Modelling*, 2002. **26**(12): p. 1151.
27. Ke, J.-C. and C.-H. Lin, Maximum entropy approach for batch-arrival queue under N policy with an un-reliable server and single vacation. *J. Comput. Appl. Math.*, 2008. **221**(1): p. 1-15.
28. Skianis, C., Arbitrary queueing network models with blocking and server vacations: approximate analysis of queueing network models of manufacturing and computer communication systems with finite capacities, server vacation periods and different types of building block queues using the principle of minimum relative entropy and the generalised exponential distribution. , in Department of Computing. 1997, University of Bradford: Bradford, UK.
29. Walstra, R.J., Nonexponential networks of queues: a maximum entropy analysis. *SIGMETRICS Perform. Eval. Rev.*, 1985. **13**(2): p. 27-37.
30. Wu, J.-S., Maximum Entropy Analysis of Open Queueing Networks with Group Arrivals. *The Journal of the Operational Research Society*, 1992. **43**(11): p. 1063.
31. Kouvatso, D.D. and I. Awan, Entropy maximisation and open queueing networks with priorities and blocking. *Perform. Eval.*, 2003. **51**(2-4): p. 191-227.
32. Kouvatso, D.D., Entropy maximisation and queueing network models. *Annals of Operations Research*, 1994. **48**(1): p. 63-126.
33. Shah, N., D.D. Kouvatso, and R.J. Fretwell. An Analytic Generalisation of a Maximum Entropy Customer Impatience Queueing Solution and its Nonbalking G/M/1/N Equivalence. in *The 11th Annual Postgraduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting*. 2010. Liverpool, UK: The School of Computing and Mathematical Sciences, Liverpool John Moore's University.
34. Cantor, J., A. Ephremides, and D. Horton, Information theoretic analysis for a general queueing system at equilibrium with application to queues in tandem. *Acta Informatica*, 1986. **23**(6): p. 657-678.
35. Guiasu, S., Maximum entropy condition in queueing theory. *The Journal of the operational research society*, 1986. **37**(3): p. 293-301.
36. Yang, D.Y., K.H. Wang, and W.L. Pearn, First two moment entropy maximisation approach for M/G/1 queues with second optional service and server breakdowns. *International Journal of Services Operations and Informatics*, 2011. **6**(4): p. 310-331.
37. Kouvatso, D. and I. Awan, Entropy maximisation and open queueing networks with priorities and blocking. *Performance Evaluation*, 2003. **51**(2-4): p. 191.
38. Choudhury, G.L., K.K. Leung, and W. Whitt, *Resource-Sharing Models with State-Dependent Arrivals of Batches*. 1994, AT&T Bell Laboratories: New Jersey.
39. Hassin, R. and M. Haviv, *To Queue or Not to Queue: Equilibrium Behaviour in Queueing Systems*. *International Series in Operations Research and Management Science*. Vol. 59. 2003: Kluwer Academic Publishers. 191.

Analysis of Continuous Time Single Server Queueing Model with Batch Renewal Arrivals: $GI^G/M/1/N$

Wei Li, Demetres D. Kouvatsos and Rod J. Fretwell

ShanDong University, China
University of Bradford, UK

`lindaweili@gmail.com,`
`D.Kouvatsos@bradford.ac.uk`

Abstract. It has been proven that the batch renewal process (BRP) is the least biased process given the infinite sets of measures of the traffic correlation, (i.e. indices of dispersion, covariance or correlation function) in the discrete time domain. The similar argument is expected for the continuous time BRP but not shown in here.

The correlation induced by BRP is analysed and a queueing model fed by BRP is studied. In this context, a continuous time $GI^G/M/1/N$ queueing model is analysed having a single server, general batch renewal arrival process, exponential service time and finite buffer size N . Closed form expressions for performance distribution, such as, queue length, blocking probability and waiting time distributions are derived. As a consequence, these analytic tools can be used to efficiently assess the adverse effect of traffic correlation induced by the BRP on the queue.

Keywords: batch renewal process (BRP); continuous time queueing model; performance distributions; correlation;

Analysis of Continuous Time Single Server Queueing Model with Batch Renewal Arrivals: $GI^G/M/1/N$

Wei Li

School of Computer Science and Technology
ShanDong University
JiNan China
lindaweili@gmail.com

Demetres D. Kouvatsos and Rod J. Fretwell

NetPEN – Networks and Performance Engineering Research Group
Informatics Research Institute (IRI), University of Bradford
Bradford, BD7 1BA, UK
D.Kouvatsos@bradford.ac.uk

Abstract—It has been proven that the batch renewal process (BRP) is the least biased process given the infinite sets of measures of the traffic correlation, (i.e. indices of dispersion, covariance or correlation function) in the discrete time domain. The similar argument is expected for the continuous time BRP but not shown in here.

The correlation induced by BRP is analysed and a queueing model fed by BRP is studied. In this context, a continuous time $GI^G/M/1/N$ queueing model is analysed having a single server, general batch renewal arrival process, exponential service time and finite buffer size N . Closed form expressions for performance distribution, such as, queue length, blocking probability and waiting time distributions are derived. As a consequence, these analytic tools can be used to efficiently assess the adverse effect of traffic correlation induced by the BRP on the queue.

Index Terms—batch renewal process (BRP); continuous time queueing model; performance distributions; correlation;

I. INTRODUCTION

A persistent problem in the field of performance modelling and evaluation is to choose the most appropriate queueing model. Unfortunately, due to the need for analytical tractability most works assume a Poisson arrival process [9], which is not realistic as shown, for example in [12]. On the other hand, in actual networks, such as IP or ATM networks, it is most possible that packets arrive at the queue in batches rather than individually.

Recently, queueing models fed by batch arrival process have received significant attention [6-11]. Most of the research work falls in the discrete time domain, whilst batch arrival process could be also applied into continuous time domain. The latter removes the restriction that the packets, in discrete time domain, must arrive at the end of a time slot.

BRP is a general case of batch arrival process which involves both count and interval correlations. Note that indices of dispersion have long been known as powerful tools in the analysis of the second-order properties of point processes [13] [14] and computer traffic measurements [1-5].

Sriram and Whitt [1] analysed superposition of bursty correlated arrival processes in terms of indices of dispersion for intervals (IDI). Heffes and Lucantoni [2] modelled the superposition of bursty arrival processes approximated by a correlated Markov modulated Poisson process (MMPP) matching three features of indices of dispersion for counts (IDC) and mean arrival rate. Gusella [3] characterized the variability of

measured packet arrival process with indices of dispersion and discussed the merits of these indices as well as the pitfalls of their indiscriminate use. Moreover he estimated the indices of dispersion for measured LAN traffic which was approximately modelled by 2-phase MMPP which matched on the three IDCs and the squared coefficient of variation (SCV) of the inter-arrival times.

Fowler and Leland [14] reported LAN traffic with unbounded IDC. However, it is expected that performance of a restricted buffer system with deterministic service would not be affected by the magnitude of IDC for long intervals. Andrade and Martinez-Pascua showed that the queue length distribution and other statistics are affected by IDC only up to a certain size of interval (determined by the buffer size) whilst the value of the IDC at infinity is of little importance.

Only a limited number of papers studying on finite-buffer queues with batch renewal process have appeared so far in the literature. In this context, the batch renewal process in a discrete time domain was introduced and applied to the analysis of queueing models by Fretwell and Kouvatsos [5-7], where the generating functions of IDI and IDC were expressed in terms of corresponding generating functions of batch inter-arrival time and batch size distributions. Moreover, Frey [15] analysed a finite-capacity vacation queue with batch arrivals but considered batch Poisson arrivals only. Niu [9] studied a vacation queue with setup and close-down times and batch Markovian arrival processes in discrete time domain. A queueing model with batch arrivals and batch-dedicated servers was analysed by Gullu [11], which nevertheless, has an infinite number of servers and buffer sizes. Finally, Takaki and Wu [10] studied queue fed by a semi-Markovian batch arrival process but with infinite waiting rooms.

The BRP is employed in the continuous time domain and the following analysis of a $GI^G/M/1/N$ queue with a general batch renewal arrival process GI^G , as defined in Section 2. Performance distributions of queue length, blocking probability and waiting time distributions are determined.

Section 2 introduces the continuous time batch renewal process and associated properties. Section 3 gives the transforms of IDI and IDC of continuous time batch renewal process. The performance distributions of a single server queue with batch renewal arrival process and finite-capacity are presented in Section 4. Conclusions are drawn in Section 5.

II. BATCH RENEWAL PROCESSES

A batch renewal process as a traffic process allows simultaneous arrivals such that

- the number of arrivals in different batches are independent and identically distributed
- the intervals between batches are independent and identically distributed
- the batch sizes (number of arrivals in one batch) are independent of intervals between batches

In the continuous time domain, a batch renewal process has continuous inter-arrival times and discrete counts. In the discrete time domain a batch renewal process has been found to exhibit both interarrival correlation between individual arrivals and count correlation between successive epochs. In this paper the corresponding correlations are investigated in the context of a continuous time batch renewal process.

III. CORRELATION IN BATCH RENEWAL PROCESS

Consider a continuous time batch renewal process in which

- the distribution of batch size is given by the probability mass function (pmf) $b(n)$, $n = 1, 2, \dots$, with mean b , squared coefficient of variation (SCV) C_b^2 and probability generating function (pgf) $B(z) = \sum_{n=1}^{\infty} b(n)z^n$.
- the distribution of interval between batches is given by the probability density function (pdf) $a(t)$, $t > 0$, with mean a , SCV C_a^2 and Laplace transform $A(\theta) = \int_0^{\infty} a(t)e^{-\theta t} dt$.

Let $\nu(n, t)$ be the probability that exactly n customers arrive during the interval of length t , $n = 0, 1, \dots$, $t > 0$. Then the transform function of $\nu(n, t)$ is defined as

$$\begin{aligned} N(z, \theta) &= \sum_{n=0}^{\infty} \int_{t=0}^{\infty} \nu(n, t) z^n e^{-\theta t} dt \\ &= \frac{1}{\theta} - \frac{1}{a \cdot \theta^2} \cdot \frac{(1 - A(\theta))(1 - B(z))}{1 - A(\theta)B(z)} \end{aligned} \quad (1)$$

The IDC, I_t , is defined to be the variance of the number of arrivals during an interval of length t divided by the mean number $E(N_t)$ of arrivals in t , namely

$$I_t = \frac{Var(N_t)}{E(N_t)} \quad (2)$$

Let $I(\theta)$ to be the generating function of I_t . From the well-known connection between the indices of dispersion and correlation functions (covariances), the transform of count correlation function $K(\theta)$ can be determined by (c.f.[16]):

$$K(\theta) = b \left(C_b^2 + \frac{1 + A(\theta)}{1 - A(\theta)} - \frac{2}{a} \cdot \frac{1}{\theta} \right) \quad (3)$$

Let $\tau(n, t)$ be the probability that there are n intervals between $n + 1$ successive individual arrivals during interval of length t . Then the transform function of $\tau(n, t)$ is defined as

$$\begin{aligned} T(z, \theta) &= \sum_{n=0}^{\infty} \int_{t=0}^{\infty} \tau(n, t) z^n e^{-\theta t} dt \\ &= \frac{1}{1 - z} - \frac{z}{b \cdot (1 - z)^2} \cdot \frac{(1 - A(\theta))(1 - B(z))}{1 - A(\theta)B(z)} \end{aligned} \quad (4)$$

The IDI, J_n , is defined to be the n times SCV of intervals X between individual arrivals.

$$J_n = \frac{n \cdot Var(X)}{E(X)^2} \quad (5)$$

Defining $J(z)$ to be the generating function of J_n . The transform of interval correlation function $L(z)$ can be obtained by

$$L(z) = b \left(C_a^2 + \frac{1 + B(z)}{1 - B(z)} - \frac{1}{b} \cdot \frac{1 + z}{1 - z} \right) \quad (6)$$

Note that $K(\theta)$ is expressed in terms of batch interarrival time distribution and $L(z)$ in terms of batch size distribution.

From equations (3) and (6), $A(\theta)$ and $B(z)$ can be expressed in terms of $K(\theta)$ and $L(z)$, respectively, by:

$$1 - B(z) = \frac{2b}{L(z) - b(C_a^2 - 1) + \frac{1+z}{1-z}} \quad (7)$$

$$1 - A(\theta) = \frac{2b}{K(\theta) - b(C_b^2 - 1) + 2\lambda \cdot \frac{1}{\theta}} \quad (8)$$

It may be shown that equations (7) and (8) define $A(\theta)$ and $B(z)$ uniquely given $K(\theta)$ and $L(z)$ (c.f.[15]). It may be also shown that counting process $\nu(n, t)$ and timing process $\tau(n, t)$, taken together, define $A(\theta)$ and $B(z)$ uniquely.

IV. CENSORED $GI^G/M/1/N$ QUEUE

This section specifies the $GI^G/M/1/N$ queue and derives the performance distributions.

A. Model Description and Specification

Consider a $GI^G/M/1/N$ queue with a batch renewal arrival process where customers within an arriving batch are turned away and simply lost (i.e. censored arrivals) when the number of occupied buffers reaches capacity N . The service times are assumed to be exponentially distributed and the first customer arriving to an empty system receives service immediately and departs after service completion (immediate service policy). In the following sections, $a(t)$ denotes the probability density function of inter-arrival time between batches, $b(n)$ the probability mass function of batch sizes and $N(0 < N < \infty)$ the buffer size of the queue.

B. Two Embedded Processes

Consider within a continuous time domain two processes embedded at points immediately before and after each batch arrivals. Each process may be described independently by a Markov chain but the processes are mutually dependent.

- For the first chain (chain 'A'), the state is the number of customers in the queue after allowing for any departure at that instant but discounting the new arrivals at that instant. Let $P_N^A(n)$ be the steady state probability that the state be $n = 0, 1, \dots, N - 1$ (where N is the capacity of the system)
- For the second chain (chain 'D'), the state is the number of customers in the queue after allowing for any departure at that instant but including the new arrivals. Let $P_N^D(n)$ be the steady state probability that state be $n = 1, 2, \dots, N$
- Let $P_N(n)$ be the steady state probability that there are $n = 0, 1, \dots, N$ customers in the system (including queueing and receiving service) at any time

To see the relation between the two Markov chains, first consider the state of each chain at an arrival instant. Chain 'D' may be in state $n = 1, 2, \dots, N - 1$ when chain 'A' is in state $k = 0, 1, \dots, n - 1$ and there are $n - k$ arrivals in the batch. Alternatively, chain 'D' may be in state N when chain 'A' is in state $k = 0, 1, \dots, N - 1$ and there are at least $N - k$ arrivals in the batch, therefore

$$P_N^D(n) = \begin{cases} \sum_{k=0}^{n-1} P_N^A(k) b(n-k) & n=1, 2, \dots, N-1 \\ \sum_{k=0}^{N-1} P_N^A(k) \sum_{r=N-k}^{\infty} b(r) & n=N \end{cases} \quad (9)$$

Now consider the states of each chain at successive arrival instants. At the later instant, the chain 'A' may be in state $n = 1, 2, \dots, N - 1$ when the chain 'D' may be in state $k = n + 1, \dots, N$ at earlier arrival instant and there are $k - n$ departures in the interval between two arrivals of batches. Since the service times are exponentially distributed, the probability that $k - n$ customers depart during interval of length t is given by

$$\int_0^{\infty} e^{\mu t} \cdot \frac{(\mu t)^{k-n}}{(k-n)!} dt \quad (10)$$

The relationship between P_N^A and P_N^D at two successive arrival instants

$$P_N^A(n) = \begin{cases} \sum_{k=0}^{N-n} P_N^D(n+k) \int_0^{\infty} e^{-\mu t} \cdot \frac{(\mu t)^k}{k!} \cdot a(t) dt & n = 1, 2, \dots, N \\ \sum_{k=0}^N P_N^D(k) \int_0^{\infty} \int_0^t e^{-\mu s} \cdot \frac{(\mu s)^k}{k!} ds \cdot a(t) dt & n = 0 \end{cases} \quad (11)$$

C. Queue Length Distribution

If immediately after the arrival of a batch, there are $n+k = 1, 2, \dots, N$ customers in the system and the interval between that and the next batch be the length of t , then

$$P_N(n) = \begin{cases} \sum_{k=0}^{N-n} P_N^D(n+k) \int_0^{\infty} \int_0^t e^{-\mu s} \cdot \frac{(\mu s)^k}{k!} \cdot \frac{1}{a} \cdot ds \cdot a(t) dt & n = 1, 2, \dots, N \\ \sum_{k=1}^N P_N^D(k) \int_0^{\infty} \int_0^t e^{-\mu s} \cdot \frac{(\mu s)^k}{k!} \cdot \frac{1}{a} \cdot (t-s) \cdot ds \cdot a(t) dt & n = 0 \end{cases} \quad (12)$$

D. Blocking probability

If an arriving batch of size $N - k + r$ finds k customers in the finite buffer system, then only the first $N - k$ customers of the arriving batch enter the system and r customers will be blocked and lost.

The probability that the arriving batch finds k customers is $P_N^A(k)$, the probability that the batch be of size $N - k + r$ is $(N - k + r) \cdot \frac{1}{b} \cdot b(N - k + r)$ and the probability of a customer being in one of the r positions, given the batch being of size $N - k + r$, is $\frac{r}{N - k + r}$. Therefore, the blocking probability Π_N^B is given by

$$\Pi_N^B = \sum_{k=0}^{N-1} P_N^A(k) \sum_{r=1}^{\infty} \frac{r}{b} b(N - k + r) \quad (13)$$

E. Waiting time distribution

The waiting time of a customer is given by the interval from the instant at which it arrives in the queue to that it is receiving service.

If the new arrival batch finds k customers in the system, then the waiting time be the service time of k customers plus the service time of $i - 1$ customers before the tagged customer given it is in the i^{th} position in the batch.

$$W_N(t) = \frac{\sum_{k=0}^{N-1} \sum_{r=i}^{\infty} \sum_{r=1}^{N-k} P_N^A(k) b(r) \int_0^{\infty} \frac{(\mu t)^{k+i-2}}{(k+i-2)!} \cdot \mu e^{-\mu t} dt}{b(1 - \Pi_N^B)} \quad (14)$$

V. CONCLUSIONS

A continuous time $GI^G/M/1/N$ queue with single server, general batch renewal arrivals process, exponentially distributed service time and finite capacity N is analysed. Closed form expressions for basic performance distributions, such as queue length, waiting time and blocking probability distributions are derived. As a consequence, these analytic tools can be used to efficiently assess the adverse effect of traffic correlation induced by the BRP on the queue.

ACKNOWLEDGMENT

The authors would like to thank the Scientific Research Foundation for Returned Overseas Chinese Scholars, State Education Ministry, Shandong Province Natural Science Foundation (2009ZRB02550) and Shandong Province Science Foundation for Post-doctoral Scientists of China (200803053)....

REFERENCES

- [1] Kotikalapudi Sriram, Ward Whitt, *Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data*, IEEE Journal on Selected Areas in Communication, Vol. SAC-4, No. 6, pp. 833-846, September 1986
- [2] Harry Heffes, David M. Lucantoni, *A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance*, IEEE Journal on Selected Areas in Communication, Vol. SAC-4, No. 6, pp. 856-868, September 1986
- [3] Riccardo Gusella, *Characterizing the Variability of Arrival Processes with Indexes of Dispersion*, IEEE Journal on Selected Areas in Communication, Vol. 9, No. 2, pp. 203-211, February 1991
- [4] Leonard Kleinrock, *Queueing Systems Volume1: Theory*, A Wiley-Interscience Publication, John Wiley and Sons, Canada, 1975.
- [5] Demetres Kouvatso, Rod Fretwell, *Discrete Time Batch Renewal Processes with Application to ATM Performance*, Proceeding of the 10th UK Performance Engineering Workshop, Hilston, J.Etal(eds.). Edinburgh University Press, pp. 187-192, Sep.1994
- [6] Demetres Kouvatso, Rod Fretwell, *Closed Form Performance Distribution of a Discrete Time $GI^G/D/1/N$ Queue with Correlated Traffic*, In Data Communications and their Performance, Fdida S. and Onvural.(eds.), Chapman and Hall, pp. 141-163, 1995
- [7] R.J.Fretwell, *An Investigation into traffic correlation in high speed communication networks by means of discrete time models*, Ph.D thesis, 2002
- [8] Sebastia Galmes, Ramon Ruigjaner, *Correlation analysis of a discrete-time flexible arrival process*, Computer Networks, Vol. 41 (2002), pp. 795-814
- [9] ZhiSheng Niu, Tao Shu, Yoshitaka Takahashi, *A vacation queue with setup and close-down times and batch Markovian arrival processes*, Performance Evaluation, Vol. 54 (2003), pp 225-248
- [10] Hideaki Takagi, De-An Wu, *Multiserver queue with semi-Markovian batch arrivals*, Computer Communications, Vol. 27 (2004), pp 549-556
- [11] Refik Gullu, *Analysis of an $M/G/\infty$ queue with batch arrivals and batch-dedicated servers*, Operations Research Letters, Vol. 32 (2004), pp 431-438
- [12] Vern Paxson, Sally Floyd, *Wide-Area Traffic: The Failure of Poisson Modeling*, ATM/IEEE Transactions on Networking, 3 (3), pp. 226-244, June 1995
- [13] D.D.Brillinger, *Comparative aspects of the study of ordinary time series and of point processes*, in Development in Statistics, P.R.Krishnaiah, Ed. New York: Academic, 1978, Vol.1, pp33-133
- [14] D.R.Cox and P.A.W.Lewis, *The Statistical Analysis of Series of Events*. London: Chapman and Hull, 1966.
- [15] A.Frey, Y.Takahashi, *An $M^{[x]}/GI/1/N$ queue with close-down and vacation time*, J.Appl.Math.Stochast.Anal.12 (1999) 63-83
- [16] W.Li, R.Fretwell, D.D.Kouvatso, *Continuous Time Batch Renewal Processes and the Application to Analysis of Queueing Models*, Research Report, May 2005

Extended Entropy Maximization and Queues with Heavy Length Tails

Demetres D. Kouvatsos

Networks and Performance Engineering Research Group (NetPEn),
Informatics Research Institute (IRI), University of Bradford, Bradford, BD7 1DP, UK

D.Kouvatsos@brad.ac.uk

Abstract. Empirical traffic characterization studies for heterogeneous networks and the Internet have shown that traffic streams often exhibit burstiness and self-similarity causing performance degradation and the formation of queues with bursty and/or heavy length tails. This tutorial undertakes an exposition of the ‘extensive’ (EME) and ‘non-extensive’ (NME) maximum entropy formalisms, whose roots may be found in diverse disciplines such as Information Theory and Statistical Mechanics. These formalisms are based, respectively, on the maximization of the classical Gibbs-Boltzmann-Shannon and generalised Havrda-Charvat-Tsallis entropy functions, subject to prior information in the form of suitable mean value constraints. The credibility of these formalisms, as methods of inductive inference for the study of physical systems with short-range and long range interactions, respectively, will be explored in terms of four potential consistency axioms, namely uniqueness, invariance, system independence and subset independence. Focusing on stable single server queues, it will be shown that the analytic EME and NME state probabilities are characterized by generalised modified geometric (GGeo) and Zipf-Mandelbrot (G-Z-M) distributions depicting, respectively, bursty generalized exponential (GE) and/or heavy length tails with asymptotic power law behaviour. Typical numerical experiments will be used to highlight the validity and robustness of these analytic EME and NME solutions and evaluate the adverse combined impact of traffic burstiness and self-similarity on the performance of the queues. Moreover, the suitability of these entropy-based solutions, as simple and cost-effective analytic building blocks, will be assessed towards the establishment of product-form approximations and queue-by-queue decomposition algorithms for complex open QNMs, subject to bursty and self-similar traffic flows. Finally, performance modelling applications for telecommunication networks will be highlighted in conjunction with associated open problems.

IFIP HET-NETs 2013

TUTORIAL T02

**Extended Entropy Maximization and
Queues with Heavy Length Tails**

Demetres D. Kouvatsos

NetPEn

**Networks and Performance Engineering
Research Group, University of Bradford, UK**



Outline

- Motivation: *Info. Theory, Statistical Mechanics & Quantification Theory* → *Queues with Bursty & Heavy Tails*
- *On the Interpretation of Classical and Generalised Entropies;*
- *The Non-Extensive Max Entropy (NME) Formalism;*
- *The NME Formalism as a Method of Inductive Inference;*
- *An Application to a Finite capacity Queue with Heavy Tails;*
- *Numerical Experiments;*
- *Conclusions and Suggestions for Future Work.*

Motivation: Info. Theory, Stat. Mechanics & Quantification Theory → Queues with Heavy Tails

- The optimisation of **Havrda-Charvat-Tsallis entropy function**, a generalisation of the classical **Boltzmann-Gibbs-Shannon entropy function**, subject to suitable mean value constraints for the analysis of queues and queueing network models (QNMs) of high speed networks **with long range interactions** and **bursty** and/or **heavy queue length tails**.

Motivation: Info. Theory, Stat. Mechanics & Quantification Theory → Queues with Heavy Tails

- *Empirical traffic characterisation studies concerning high speed networks of diverse technology and the Internet have shown that traffic flows often exhibit **burstiness**, **self-similarity (SS)** and/or **long-range dependence (LRD)** causing performance degradation and the formation of queues with heavy length tails and asymptotic power law behaviour (e.g., Sahinoglu et al (1999)).*

Traffic Modelling and Characterisation in High Speed Networks

- *Traffic distributions with such properties are often employed to generate network workloads in simulation modelling for performance evaluation studies which, however, tend to be rather inflexible, computationally display unusual characteristics (e.g., Crovella (1997), Rezaul & Crout (2007)).*

*Motivation: Info. Theory, Stat. Mechanics &
Quantification Theory → Queues with Heavy Tails*

- *Analytic mechanisms for estimating the tail index of Internet traffic with heavy tails, based on the Pareto distribution have been reported in Rezaul & Grout (2007).*
- *Analytic exploration can be found in Karmeshu & Sharma (2005) and Kouvatsos & Assi (2002, 2007, 20011) , based on the optimisation of extensions of Boltzmann-Gibbs-Shannon entropy, such as those proposed in the fields of Quantification Theory (c.f., Havrda & Charvat (1967)), Statistical Mechanics (c.f., Tsallis (1988)) and Info. Theory (c.f., Shannon (1947)).*

The Classical Boltzmann-Gibbs-Shannon Entropy

For a general physical system \mathbf{Q} with an integer number of possible (microscopic) configurations or states $N(N>0)$, the **Boltzmann-Gibbs-Shannon** classical **entropy function**, $H^*(p_N)$ is defined by

$$H^*(p_N) = -c \sum_{n=1}^N p_N(n) \log p_N(n)$$

where $\mathbf{c}(c > 0)$ is a positive constant and $\{p_N(n), n = 1, 2, \dots, N\}$ are the associated event or state probabilities of system \mathbf{Q} . It can be **interpreted** as the measure of uncertainty or information content that is implied by $p_N(n)$ about a physical system **with short range interactions**.

Physical Systems with Short-Range Interactions

- For physical systems with **short-range interactions**, such as *the formation of chemical bonds, holding the matter together etc. in Statistical Mechanics*, **energy** and **entropy** are **extensive** variables, in the sense that

'Total Energy → ~ System Size' .

- For **queueing systems** with **short-range dependent (SRD)** traffic flows, such as Poisson and Compound Poisson (bursty) types, the **state** and **entropy** variables are **extensive**
→ **exponential/generalised exponential (GE) queue length tails**

A Generalised Entropy Function

For a general system Q with an integer number of possible (microscopic) configurations or states $N(N>0)$, the **Havrda-Charvat-Tsallis non-extensive entropy function**, $H^*(p_{q,N})$ can be defined by

$$H^*(p_{q,N}) = c(1 - \sum_{n=1}^N p_{q,N}(n)^q) / (q - 1)$$

where q is a real number known as the **non-extensivity** parameter and $\{p_{q,N}(n), n = 1, 2, \dots, N\}$ are the state probabilities of a Q at equilibrium. As $q \rightarrow 1$, $H^*(p_{q,N})$ reduces to the classical entropy function $H^*(p_N)$.

ME Formalism, Statistical Mechanics & Short/Long-Range Interactions

- **Energy** and **entropy** are **non-extensive** variables, in the sense that

‘Total Energy → No longer ~ System Size’

(c.f., due to **long-range interactions**, such as **gravity**).

For queueing systems with **long-range dependent (LRD)** traffic flows, the **state** and **entropy** variables are **non-extensive** → **heavy queue length tails**.

ME Formalisms, Statistical Mechanics Queues & Short Range Interactions

□ *Classical Maximum Entropy (ME) Principle*

{**max** Boltzmann-Gibbs-Shannon **Extensive** Entropy,
subject to **mean value constraints** of a quantity
(e.g., system energy, # of molecules, volume)}

→ Analogies with Jaynes (1957) in Stat. Physics and Shore (1982), Kouvatsos (1984, 1985 , ...), Walstra (1984) etc.
in the analysis of queues and QNMs with regular or bursty traffic

Applying the **Method of Lagrange Undetermined Multipliers**

→ **Geometric and Gen. Geo Type Prob. Distributions**

ME Formalism, Statistical Mechanics & Long-Range Interactions

- **Generalised Maximum Entropy Principle**
{**max** Havrda-Charvat-Tsallis **Non-Extensive** Entropy}
subject to appropriate **mean value constraints**

Applying the **Method of Lagrange Undetermined Multipliers**

- *Zipf-Mandelbrot Type State Prob. Distribution*
with **power-law** (heavy) tails and
non-extensivity real-valued parameter q
- *Analogies with Statistical Mechanics applications [Tsallis 1988] and the analysis of queues with bursty traffic & heavy tails [Kouvatsos & Assi 2002, Karmeshu and Sharma 2005]*

The Zipf-Mandelbrot Distribution

The **Zipf-Mandelbrot distribution** is a discrete probability distribution. It is a power-law distribution on ranked data.

The probability mass function (pmf) is of the form

$$p(n, u, s) = \frac{(n + u)^{-s}}{\sum_{n=1}^N (n + u)^{-s}}$$

N - the number of elements

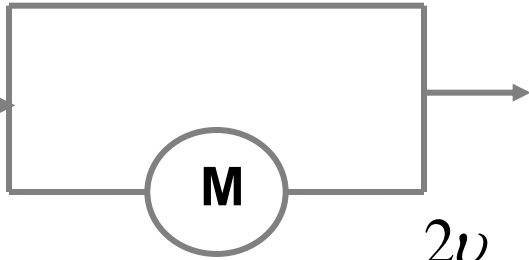
n, u - real numbers

s - the value of the exponent characterizing the distribution

The GE-type Distribution

- The **ME** solution of a stable **M/G/1** queue is exact if **G** \equiv **GE**. Similarly for a stable **GE/G/1** queue
- **GE-type** distribution with parameters α and β ($0 \leq \alpha \leq 1$)

$$F(t) = 1 - \alpha e^{-\beta t}, \quad t \geq 0,$$

$$1 - \alpha = \frac{C^2 - 1}{C^2 + 1}$$

$$\alpha = \frac{2}{C^2 + 1}$$
$$\beta = \frac{2\nu}{C^2 + 1}$$

- The underlying counting process of the GE-type distr. is a compound Poisson process with **Geo-type** distributed batch sizes and mean batch size $1/\alpha = (C^2 + 1)/2$.

Interpretation of GE-type distribution

- **GE** is an extremal case of the family of two-phase exponential distributions having the same $\{v, C^2(>1)\}$
- **GE** is a bulk type distribution with an underlying counting process equivalent to a **Compound Poisson Process (CPP)** with parameter $2v/(C^2 + 1)$ and a geometrically distributed bulk sizes with mean $(1 + C^2)/2$ and SCV $(C^2 - 1)/(C^2 + 1)$ i.e.,

$$P(N_{cp} = n) = \begin{cases} \sum_{i=1}^n \frac{\sigma^i}{i!} e^{-\sigma} \binom{n-1}{i-1} \tau^i (1-\tau)^{n-i}, & n \geq 1 \\ e^{-\sigma}, & n = 0 \end{cases}$$

where N_{cp} is the random variable of the number of events per unit time.

Non- Extensive Maximum Entropy Formalism

■ Consider a general system \mathcal{Q} with a set \mathcal{S} of possible discrete states $\{\mathcal{S}_1, \mathcal{S}_2, \dots\}$, which may be finite or countable infinite and $\mathcal{S}_i, i = 0, 1, \dots$ may be specified arbitrarily.

■ A **NME** framework can be established to determine the form of $p_q(\mathcal{S}_n)$ that maximises the generalised **Havrda-Charvat-Tsallis** entropy, $H^*(p_q)$

$$H^*(p_q) = c \frac{1 - \sum_{\mathcal{S}_n \in \mathcal{S}} p_q(\mathcal{S}_n)^q}{q - 1}$$

subject to the constraints

$$\sum_{\mathcal{S}_n \in \mathcal{S}} p_q(\mathcal{S}_n) = 1$$



$$\sum_{\mathcal{S}_n \in \mathcal{S}} f_{k,q}(\mathcal{S}_n) p_q(\mathcal{S}_n) = F_{k,q}$$

Non- Extensive Maximum Entropy Formalism

where $c(c > 0)$ is a positive constant, $\{F_{k,q}, k = 1, 2, \dots, m\}$ are the prescribed mean values defined on the set of functions $\{f_{k,q}(S_n), k = 1, 2, \dots, m\}$.

By employing the **Method of Lagrange's of Undetermined Multipliers**, the **non-extensive** solution for the state probabilities is determined by

$$p_q(S_n) = \frac{1}{Z_q} \left[1 + \sum_{k=1}^m \beta_k (1 - q) f_{k,q}(S_n) \right]^{\frac{1}{q-1}}$$

Non- Extensive Maximum Entropy Formalism

where $\{\beta_k\}$, $k = 1, 2, \dots, m\}$ are the Lagrangian multipliers corresponding to the constraints (4), $Z_q = \exp\{\beta_0\}$ is the normalizing constant and β_0 is the Lagrangian multiplier determined by the normalisation constraint.

- The **NME** solution (6) can be interpreted as a **Generalised Z-M (G-Z-M)** type distribution depicting heavy queue tails and asymptotic power law behaviour.

The NME Formalism

as a Method of Inductive Inference

- The principle of EME is a uniquely correct method of inductive inference satisfying four information theoretic consistency axioms [Shore & Johnson (1980)], namely
 - *Uniqueness*
 - *Invariance*
 - *System Independence*
 - *Subset Independence*

- The credibility of the NME principle, as a method of inductive inference, may be assessed in terms of these four consistency axioms.

The NME Formalism

as a Method of Inductive Inference

- **Uniqueness:** “If the same problem is solved twice in exactly the same way, the same answer is expected in both cases i.e., the solution should be unique”;
- **Invariance:** “The same solution should be obtained if the same inference problem is solved twice in two different coordinate systems”;
- **System Independence:** “It shouldn’t matter if one accounts for ind. information about ind. systems separately in terms of different marginal probabilities **or** together in terms of a joint probability”;

The NME Formalism

as a Method of Inductive Inference

- **Subset Independence:** “It does not matter whether one treats an independent subset of system states in terms of a separate conditional density or in terms of the full system density”
- It can be shown (c.f., Kouvatsos and Assi (2011a, b) that the **NME formalism** satisfies the three axioms of **uniqueness, invariance and subset independence** but **not the one on system independence!**

System Independence

Let Q and R are two separate systems each of which has a set of N ($N > 0$) possible discrete states $\{x_n, n = 1, 2, \dots, N\}$ and $\{y_n, n = 1, 2, \dots, N\}$, respectively.

Let X and Y be the r. vs describing the state of the systems Q and R with state probabilities $f_{q,N}(x_n) = Pr\{X = x_n\}$ and $g_{q,N}(y_n) = Pr\{Y = y_n\}$, respectively.

- The joint probability, $h_{q,N}(x_k, y_n) = Pr(x_k, y_n)$, $k, n = 1, 2, \dots, N$ finition
of the $H^* [h_{q,N}] \neq H^* (f_{q,N}) + H^* (g_{q,N})$

System Independence

- The inequality implies **correlation/interdependence** between the **non-extensive** systems Q and R and hence, the **NME formalism** does not satisfy the axiom of system independence.
- Thus, it is implied that the **NME formalism** is a most suitable method of inductive inference to be used in quantitative studies of **'dependable' non-extensive dynamic systems with long range interactions'** in order to determine credible NME state probabilities at equilibrium with heavy tails and power law behaviour.

System Independence

- In the case of $q \rightarrow 1$, it follows that

$$H^* [h_{q,N}] = H^* (f_{q,N}) + H^* (g_{q,N})$$

- This result verifies that the Classical-Boltzmann-Gibbs-Shannon EE function, as expected, satisfies the axiom of system independence and thus, *the ME formalism is a most appropriate method of inductive inference **for the study of extensive systems with short range interactions.***

G/G/1 Queue: An EME Framework

■ Maximise Generalised Entropy Functional

$$\max_P \left\{ S_q = \frac{C \left(1 - \sum_{i=0}^{\infty} p(n)^q \right)}{q-1} \right\}$$

subject to

- The normalization, $\sum_{n=0}^{\infty} p(n) = 1$
- The mean queue length, $\sum_{n=0}^{\infty} np(n) = L$
- The utilisation, $\sum_{n=0}^{\infty} h(n)p(n) = 1 - p(0) = \rho$, $\rho = \frac{\lambda}{\mu}$, $0 < \rho < 1$
where $h(n) = 1$, if $n = 0$ or otherwise.

Applying the Method of Lagrange's Undetermined Multipliers

[Kouvatsos & Assi 2002, 2011]

G/G/1 Queue: An EME Framework

A **G** Zipf-Mandelbrot **EME** power-type distribution

$$p(n) = \frac{[1 + \alpha(1 - q)n + \beta(1 - q)h(n)]^{\frac{1}{q-1}}}{\sum_{n=0}^{\infty} [1 + \alpha(1 - q)n + \beta(1 - q)h(n)]^{\frac{1}{q-1}}}, \quad n = 0, 1, \dots$$

- At the **q** → **1** limit,

$$p(n) = \frac{e^{-\lambda n - \beta h(n)}}{Z} = \frac{x^n g^{h(n)}}{Z}, \quad \text{with } Z = \sum_{n=0}^{\infty} x^n g^{h(n)}, \quad x = e^{-\lambda}, \quad g = e^{-\beta}$$

→ ME state probability distribution of a stable **G/G/1** queue

- **x** and **g** are the Lagrangian coefficients corresponding to **MQ**L and server utilisation constraints. Moreover, $\frac{1}{2} < q < 1$.

G/G/1/N Queue: An EME Framework

■ Maximise Generalised Entropy Functional

$$\max_P \left\{ S_q = \frac{C \left(1 - \sum_{i=0}^{\infty} p_N(n)^q \right)}{q-1} \right\},$$

subject to:

- The normalization, $\sum_{n=0}^N p_N(n) = 1$
- The mean queue length, $\sum_{n=0}^N n p_N(n) = L_N$
- The utilisation, $\sum_{n=0}^N h(n) p(n) = 1 - p_N(0) = U$
- Full buffer state probability, $\sum_{n=0}^N s(n) p_N(n) = \varphi = p_N(n)$, $0 < \varphi < 1$

where $h(n) = 1$, if $n = 0$ or 0 ow. and $s(n) = 1$, if $n = N$ or 0 ow,
satisfying the flow balance condition: $\lambda (1 - \pi) = \mu (1 - P_N(0))$
c.f., [Assi 2000, Kouvatso and Assi 2002, 2007].

G/G/1/N Queue : An EME Framework

A **Truncated Generalised Zipf-Mandelbrot EME** power-type distribution

$$p_N(n) = \frac{[1 + \alpha(1-q)n + \beta(1-q)h(n) + \gamma(1-q)s(n)]^{\frac{1}{q-1}}}{\sum_{n=0}^N [1 + \alpha(1-q)n + \beta(1-q)h(n) + \gamma(1-q)s(n)]^{\frac{1}{q-1}}}$$

- At the $q \rightarrow 1$ limit,

$$p_N(n) = \frac{e^{-\alpha n - \beta h(n) - \gamma s(n)}}{\sum_{n=0}^N e^{-\alpha n - \beta h(n) - \gamma s(n)}} = \frac{x^n g^{h(n)} y^{s(n)}}{\sum_{n=0}^N x^n g^{h(n)} y^{s(n)}}, \quad n = 0, 1, \dots, N$$

$$x = e^{-\alpha}, \quad g = e^{-\beta}, \quad y = e^{-\gamma}$$

This is the corresponding known solution of a **GE/GE/1/K** queue. For $q < 1$ and for large number of jobs n , the **EME solution follows the power law:** $p_N(n) \sim n^{\frac{1}{q-1}}, \quad 1/2 < q < 1$

Boundary Conditions and a Heuristic Relationship Between q and H

A heuristic relation between the non-extensivity parameter, q and the Hurst parameter, H can be achieved by using the boundary conditions.

- The boundary conditions of the non-extensivity parameter q of Tsallis entropy solutions are $\frac{1}{2} < q < 1$; The boundary conditions of Hurst parameter H of the **fractional Brownian Motion (fBm)** is $\frac{1}{2} < H < 1$. It is implied that for $q \rightarrow 1$ (Shannon's Entropy) $H \rightarrow 0.5$ (exponential distribution);
- For $q \rightarrow 1/2$ (max value of non extensivity parameter) $\rightarrow H \rightarrow 1$ (e.g., **Pareto** distribution with **power law tails** corresponding to max value of H);
- The following simple heuristic relationship may be defined

$$H = 1.5 - q$$

An EME Mean Queue Length (MQL)

- A MQL constraint for a **fBM/M/1** queue is motivated by a reinterpretation of a formula in Norros [1994] for calculating buffer capacity of a simple storage model with self-similar input traffic process modelled by a fBm as an input process with **Hurst parameter, H**, and exponential service time:

$$\langle n \rangle = \frac{\rho^{1/[2(1-H)]}}{(1-\rho)^{H/(1-H)}}, \quad 0.5 < H < 1, \quad \rho = \lambda/\mu$$

where λ and μ are the mean arrival and service rates.

A Heuristic Generalisation for the EME MQL

- A **heuristic extension of Norros formula** was conjectured in [Kouvatsos & Assi 2007, 2011] for the buffer capacity of a simple storage model with generalised fractional Brownian motion (gfBm) process as an input traffic and GE-type service time distribution, namely

$$\langle n \rangle = \frac{\rho^{1/[2(1-H)]}}{2^{1/[2(1-H)]}} \left(\frac{(1-\rho + C_a^2 + \rho C_s^2)^{1/[2(1-H)]}}{(1-\rho)^{H/(1-H)}} \right), \quad 0.5 < H < 1, \quad \rho = \frac{\lambda}{\mu}$$

where **C^2a** and **C^2s** are the interarrival time and service time SCV and H is the Hurst parameter taking values in the interval **$[1/2, 1]$** .

- This extended formula was adopted as the MQL, **$\langle n \rangle$** , of a stable infinite capacity **gfBm/GE/1** queue. For **$H = 1/2$** it yields the result for MQL of a stable **GE/GE/1** queue (c.f., **$q \rightarrow 1$** in the proposed framework).

Server Utilisation and Blocking Probability

- The probability that server is busy (i.e., the **server utilization, U**)

$$\begin{aligned} U &= 1 - p_K(0) \\ &= 1 - \frac{[\alpha(1-q) + \beta(1-q)h(n) + \gamma(1-q)s(n)]^{\frac{1}{q-1}}}{\zeta \left[\frac{1}{1-q}, \frac{1 + \beta(1-q)h(n) + \gamma(1-q)s(n)}{\alpha(1-q)} \right]} \end{aligned}$$

- Using the flow balance condition, the **blocking probability** can be obtained by

$$\pi = 1 - U/\rho$$

EME Analytic Algorithms

EME ALGORITHM 1: The gfBm /GE/1 Queue

□ **Input Data** $\{q, \lambda, C_a^2, \mu, C_s^2\}$

Begin

Step 1 Calculate $H=1.5-q$ and mean queue length, $\{n\}$;

Step 2 Set initial approximations s of Lagrangian multipliers $\{\alpha, \beta\}$;

Step 3 Solve constraints (2) and (3) via Newton-Raphson Method wrt $\{\alpha, \beta\}$;

Step 4 Obtain new values for $\{\alpha, \beta\}$;

Step 5 Return to Step 3 until convergence of $\{\alpha, \beta\}$;

End

□ **Output Statistics:** The Lagrange's multipliers $\{\alpha, \beta\}$ and state probabilities, $\{p(n)\}$.

EME Analytic Algorithms

EME ALGORITHM 2: The Censored gfBm /GE/1/N Queue

■ **Input Data** $\{N, \alpha, \beta, q, \lambda, C_a^2, \mu, C_s^2\}$

Begin

Step 1 Initial approximation of Lagrangian multiplier γ ;

Step 2 Solve constraints (1) and (4) using the Newton-Raphson method wrt γ ;

Step 3 Obtain new values for γ ;

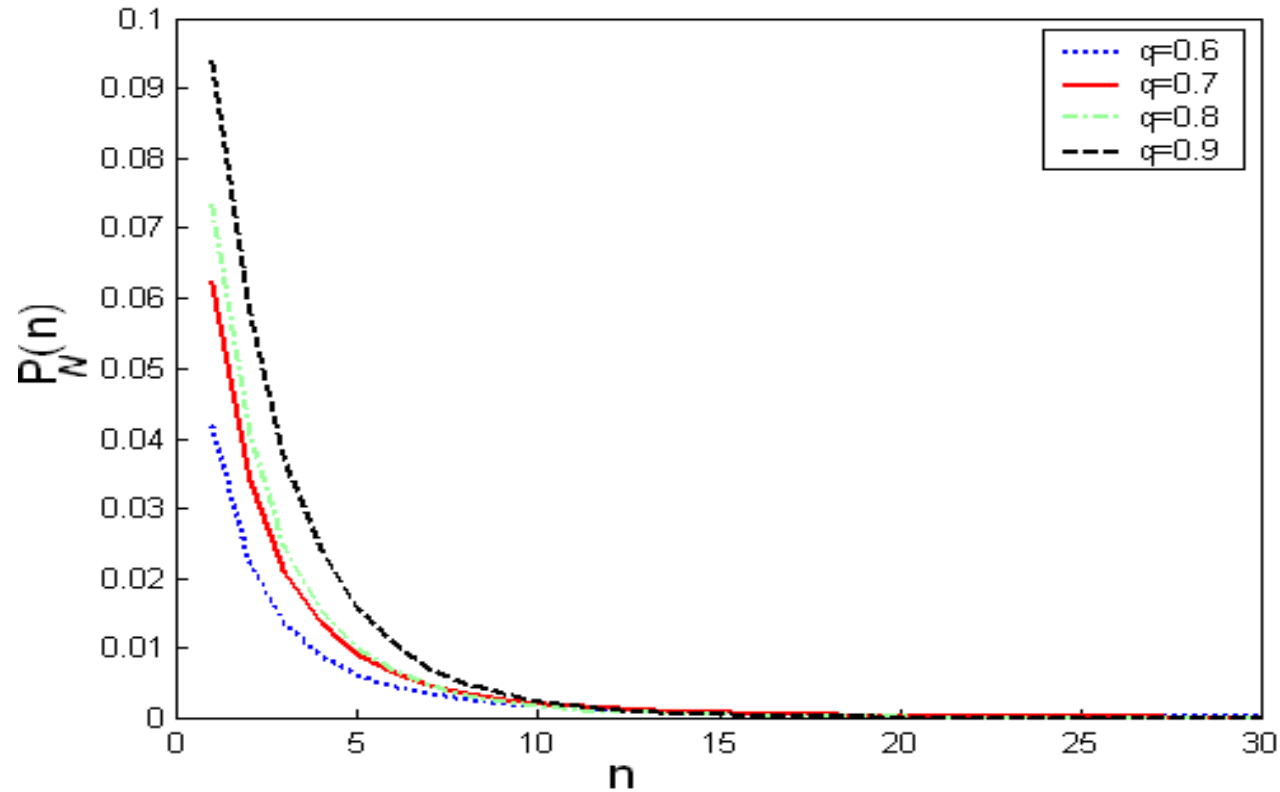
Step 4 Return to Step 2 until convergence of γ ;

Step 5 Using flow balance condition to compute blocking probability.

End

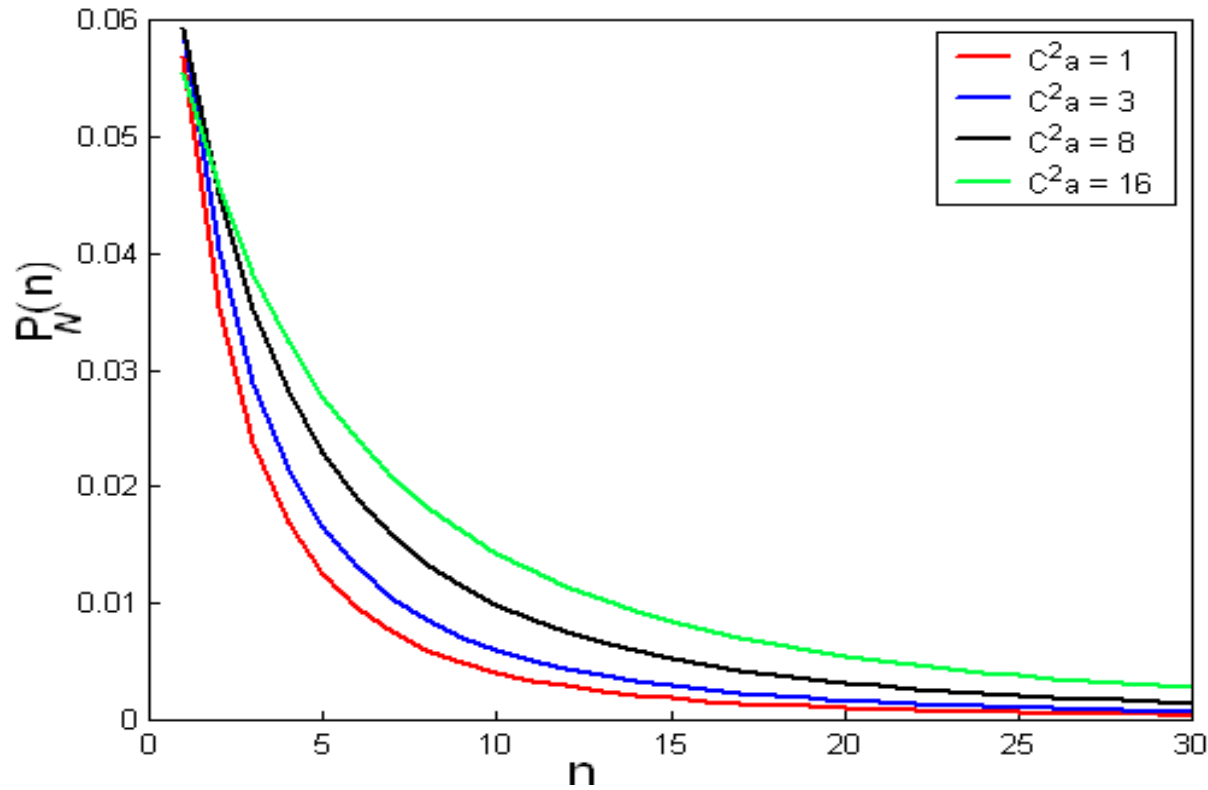
■ **Output Statistics:** The Lagrangian multipliers, γ , state probability $\{P_N(n)\}$ and the blocking probability, π .

Numerical Results



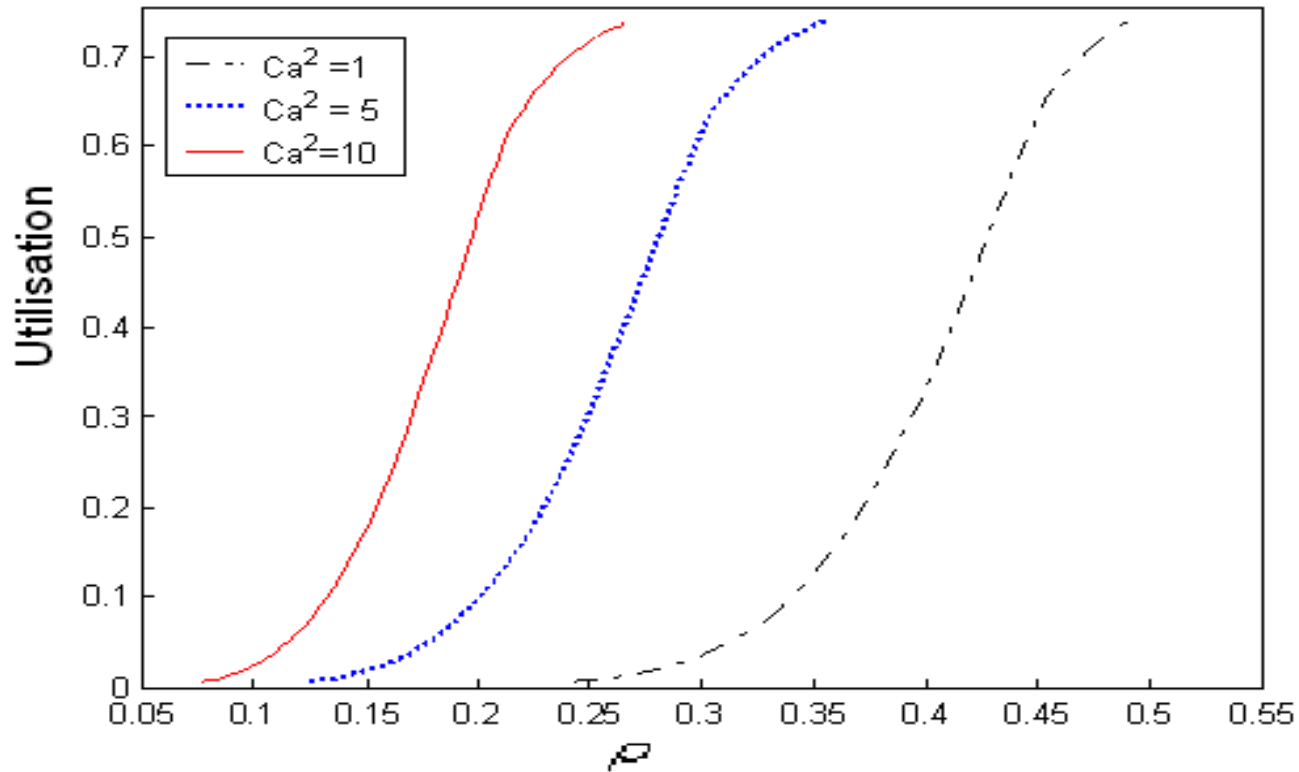
The relation between $p_N(n)$ and n for a finite capacity queue with $(q = 0.5, 0.6, 0.7, 0.9)$, $C^2_a = 8$, $C^2_s = 4$, $\lambda=0.03$, $\mu=0.2$ and $N=30$.

Numerical Results



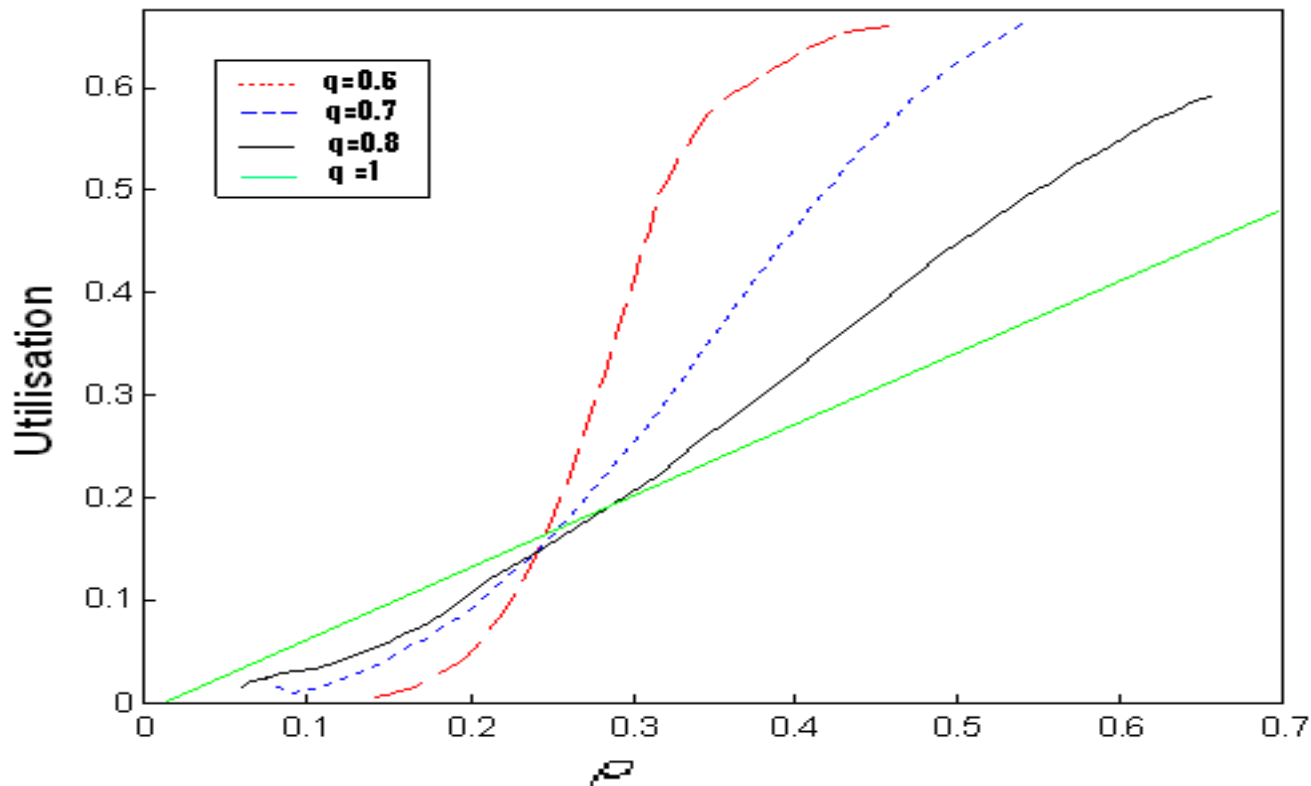
The relation between the queue length distribution and n for a finite capacity queue with $\lambda=0.02$, $\mu=0.4$, $q = 0.6$, $C^2_s = 4$ and $N = 30$

Numerical Results



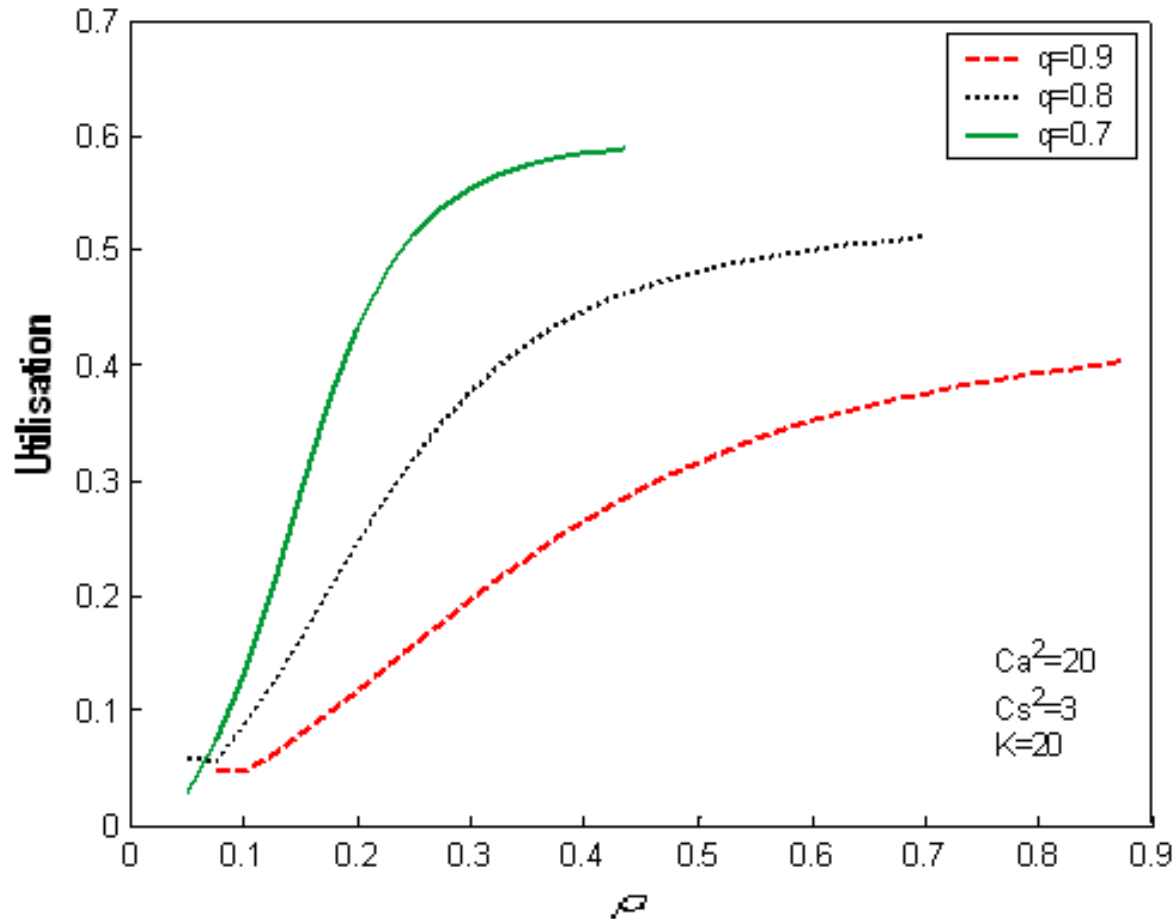
The relation between ρ and U (Utilisation) with $\{C^2a = 1, 5 \text{ \& } 10\}$, $C^2s = 3$, $q = 0.6$ and $N = 20$.

Numerical Results



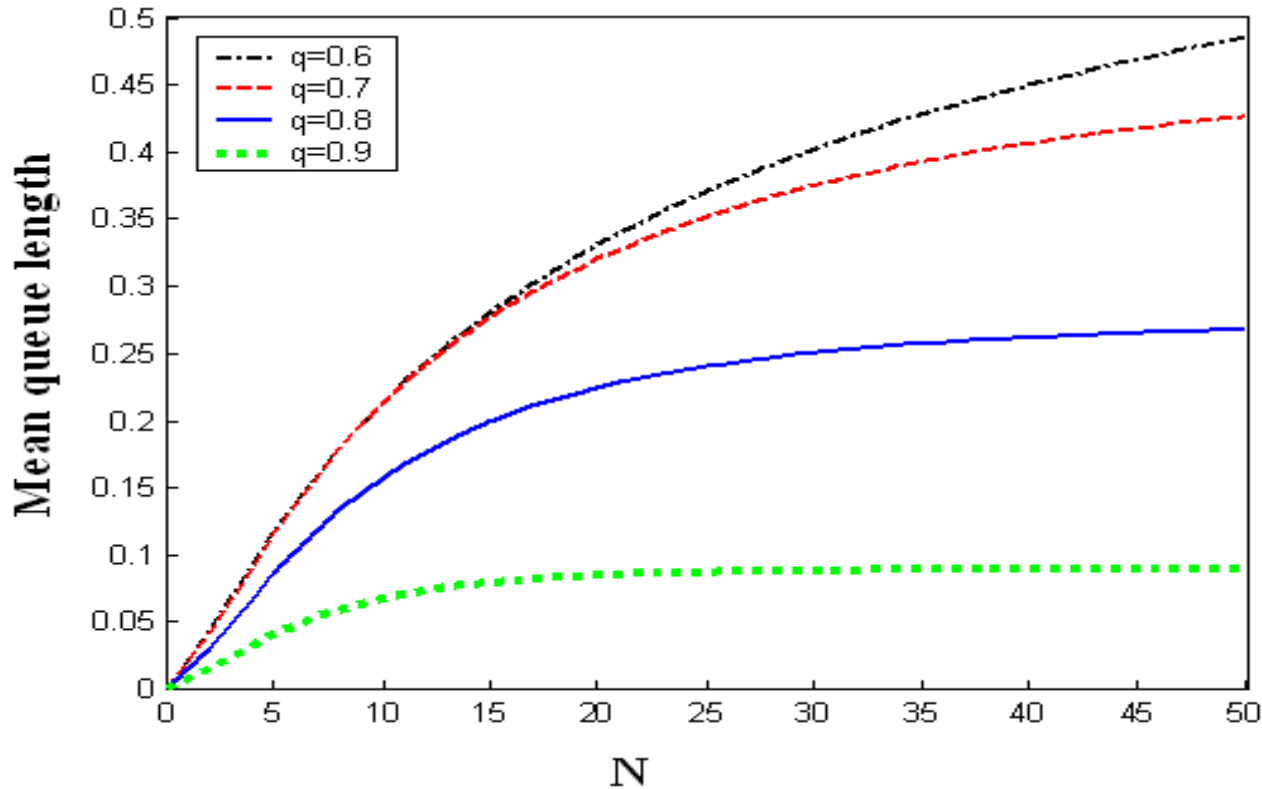
The relation between ρ and U (Utilisation) with $C^2a = 3$,
 $C^2s = 4$, $q = \{0.6, 0.7, 0.8, 1\}$

Numerical Results



The relation between traffic intensity, ρ and server utilisation, $U = 1 - P(0)$ for a finite capacity GfBm/GE/1/K queue with $Ca^2 = 20$, $Cs^2 = 3$ and $q = 0.7, 0.8, 0.9$.

Numerical Results



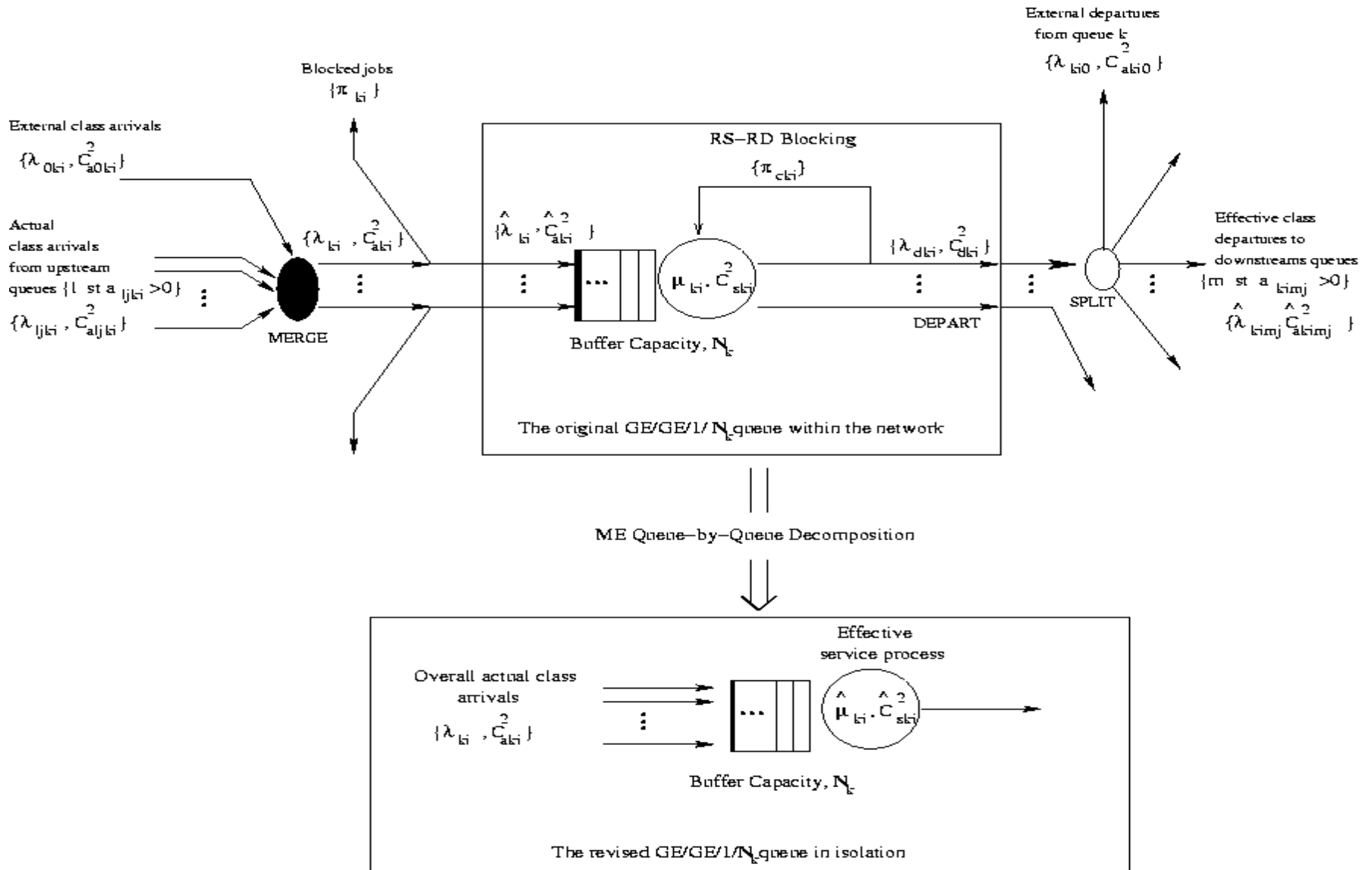
The relation between the mean queue length and N (queue capacity) with $C^2_a = 4$, $C^2_s = 9$, $\lambda = 0.1$, $\mu = 0.5$ and $(q = 0.6, 0.7, 0.8, 0.9)$

*Conclusions & Extensions to
Arbitrary Open Queueing Network Models
(QNM_s)*

- **Product-Form Approximations** and **Queue-by-Queue Decomposition** of Arbitrary Open Queueing Network Models (QNM_s) with **Blocking**

[Kouvatsos & Awan 2003, Kouvatsos et al 2011]

Queue-by-Queue *Decomposition of Open QNMs*



Selected References

- [1] I. Norros, 'A Storage Model with Self-similar Input', *Queueing Systems*, Vol. 16, pp. 387-396, 1994.
- [2] J.E. Shore and R.W. Johnson, 'Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy', *IEEE Trans. Inf. Theory*, Vol. IT-26, pp. 26-37, 1980.
- [3] J.E. Shore and R. W. Johnson, Properties of Cross-Entropy Minimisation, *IEEE Tran, Information Theory*, Vol. IT-27, pp. 472-482, 1981.
- [4] J.H. Havrda and F. Charvat, 'Quantification Methods of Classificatory Processes: Concept of Structural Alpha Entropy', *Kybernetika*, Vol. 3, pp. 30-35, 1967.

Selected References

- [5] E.T. Jaynes, '*Information Theory and Statistical Mechanics*', Phys. Rev. Vol. 106, pp. 620-630, 1957a.
- [6] E.T. Jaynes, '*Information Theory and Statistical Mechanics II*', Phys. Rev. Vol. 108, pp. 171-190, 1957b.
- [7] C. Tsallis, '*Possible Generalisation of Boltzmann-Gibbs Statistics*', Journal of Statistical Physics, Vol. 52, Nos. 1-2, pp. 479-487, 1988.
- [8] Karmeshu and S. Sharma, 'Long Tail Behaviour of Queue Lengths in Broadband Networks: Entropy Framework', Technical Report Private Communication, School of Computing and System Sciences, J. Nehru University, New Delhi, India, August 2005.

Selected References

[9] D.D. Kouvatsos and Salam Adli Assi, '*Generalised Entropy Maximisation and Queues with Bursty and / or Heavy Tails*', Chapter in '*Network-²Network Performance Engineering - A Handbook on Convergent Multi-Service Networks and Next Generation Internet*', Springer, Series: Lecture Notes in Computer Science (LNCS), Vol. LNCS 5233, pp. 357- 392, May 2011.

[10] Demetres D. Kouvatsos and Salam Adli Assi, *On the Analysis of Queues with Heavy Tails: A Non-Extensive Maximum Entropy Formalism and a Generalisation of the Zipf-Mandelbrot Distribution*, in '*Performance Evaluation of Computer and Communication Systems*', K.A. Hummel et al (Eds.), Lecture Notes in Computer Science (LNCS), Vol. LNCS 6821, ISSN 0302-9743, DOI 10.1007/978-3-642-25575-5, pp. 99 – 111, Dec. 2011.

Selected References

[11] Demetres D. Kouvatsos (Ed.), *Network Performance Engineering - A Handbook on Convergent Multi-Service Networks and Next Generation Internet*, Springer, Series: Lecture Notes in Computer Science, Vol. LNCS 5233, Subseries: Computer Communication Networks and Telecommunications, Demetres D. Kouvatsos, (Ed.), 1st Edition, 2011, XVIII, Softcover, ISBN 978-3-642-02741-3, 43 Chapters, 1080 pages, May 2011.

[12] Demetres D. Kouvatsos, Jumma S. Alanazi and Kevin Smith, *A Unified ME Algorithm for Arbitrary Open QNMs with Mixed Blocking Mechanisms*, Numerical Algebra, Control and Optimization (NACO), The American Institute of Mathematical Sciences (AIMS), ISSN2155-3289, doi:10.3934/naco.2011.1.781, Vol. 1, No. 4, pp. 781 – 816, Dec. 2011.

Thanks



END - ΤΕΛΟΣ

Hybrid Petri Queuing Networks Simulator (HPQNS)

Esmail Habibzadeh^{1,2}, Demetres D. Kouvatsos²
and Guzman M.A. Miskeen^{2,3}

1 IUST - Iran University of Science and Technology
2 University of Bradford, England
3 University of Sebha, Libya

{E.habibzadeh, D.Kouvatsos,
G.M.A.Miskeen}@brad.ac.uk

Abstract. Among others, Petri Nets (PNs) and Queuing Networks (QNs) are the most widely used modelling paradigms that can be employed to capture and represent the dynamic behaviour and relevant properties of discrete flow systems. While the PN modelling formalism enables us to examine the system both qualitatively and quantitatively, QN models may merely be exploited for quantitative studies. Although so flexible and powerful, there are some inherent limitations to these modelling approaches. PNs, in particular, suffer from the “State Space Explosion” problem when the size of the model increases, and QNs are unable to effectively capture such system dynamics as “concurrency”, “synchronization”, “mutual exclusion” and “conflict”. To overcome these constraints and exploit the best features of both modelling techniques, a hybrid simulation modelling environment-Hybrid Petri Queuing Networks Simulator (HPQNS)—has been developed where sub-models of both PNs and QNs may coexist. The main clear advantage of such a combined simulation modelling tool is to improve the expressive power of system models which, in turn, can ultimately lead to more efficient and accurate performance analysis systems under study.

Hybrid Petri Queuing Networks Simulator (HPQNS)

Esmaeil Habibzadeh

Demetres Kouvatsos

Guzlan Miskeen



University of Bradford

2013

Hybrid Petri Queuing Networks Simulator

Contents

- ✓ Introduction
 - ✓ Class Hierarchy
 - ✓ Service Facility
 - ✓ Queue
 - ✓ Arrival/Service Time Distributions
 - ✓ Queuing System
 - ✓ Place
 - ✓ Transition
 - ✓ Building Block
 - ✓ Model Builder
 - ✓ Future Developments
 - ✓ References
-

Hybrid Petri Queuing Networks Simulator

Introduction

- ✓ So many applications already available in community
 - ✓ Petri Nets
 - ✓ <http://www.informatik.uni-hamburg.de/TGI/PetriNets/tools/quick.html>
 - ✓ Queuing theory
 - ✓ <http://web2.uwindsor.ca/math/hlynka/qsoft.html>
-

Hybrid Petri Queuing Networks Simulator

Introduction

- ✓ Issues with the existing applications
-

Hybrid Petri Queuing Networks Simulator

Introduction

- ✓ Issues with the existing applications
 - ✓ No such simulator which would satisfy all our needs
-

Hybrid Petri Queuing Networks Simulator

Introduction

- ✓ Issues with the existing applications
 - ✓ No such simulator which would satisfy all our needs
 - ✓ Usually customization is not straightforward
-

Hybrid Petri Queuing Networks Simulator

Introduction

- ✓ Issues with the existing applications
 - ✓ No such simulator which would satisfy all our needs
 - ✓ Usually customization is not straightforward
 - ✓ Verification requires broad, in-depth coding knowledge
-

Hybrid Petri Queuing Networks Simulator

Introduction

- ✓ Issues with the existing applications
 - ✓ No such simulator which would satisfy all our needs
 - ✓ Usually customization is not straightforward
 - ✓ Verification requires broad, in-depth coding knowledge
 - ✓ Technical support is fairly poor, if any
-

Hybrid Petri Queuing Networks Simulator

Introduction

- ✓ Issues with the existing applications
 - ✓ No such simulator which would satisfy all our needs
 - ✓ Usually customization is not straightforward
 - ✓ Verification requires broad, in-depth coding knowledge
 - ✓ Technical support is fairly poor, if any
 - ✓ Sometimes just available for limited platforms
-

Hybrid Petri Queuing Networks Simulator

Introduction

- ✓ Issues with the existing applications
 - ✓ No such simulator which would satisfy all our needs
 - ✓ Usually customization is not straightforward
 - ✓ Verification requires broad, in-depth coding knowledge
 - ✓ Technical support is fairly poor, if any
 - ✓ Sometimes just available for limited platforms
 - ✓ Your favorite one may not be free of charge
-

Hybrid Petri Queuing Networks Simulator

Introduction

- ✓ Customized solutions might be more beneficial
-

Hybrid Petri Queuing Networks Simulator

Introduction

- ✓ Customized solutions might be more beneficial
 - ✓ The nature of research requires suitable special tools
-

Hybrid Petri Queuing Networks Simulator

Introduction

- ✓ Customized solutions might be more beneficial
 - ✓ The nature of research requires suitable special tools
 - ✓ Reasonably flexible to be modified as per requirements
-

Hybrid Petri Queuing Networks Simulator

Introduction

- ✓ Customized solutions might be more beneficial
 - ✓ The nature of research requires suitable special tools
 - ✓ Reasonably flexible to be modified as per requirements
 - ✓ Efficient coding wrt running times and memory usages
-

Hybrid Petri Queuing Networks Simulator

Introduction

- ✓ Customized solutions might be more beneficial
 - ✓ The nature of research requires suitable special tools
 - ✓ Reasonably flexible to be modified as per requirements
 - ✓ Efficient coding wrt running times and memory usages
 - ✓ Better and instant support
-

Hybrid Petri Queuing Networks Simulator

Introduction

- ✓ Customized solutions might be more beneficial
 - ✓ The nature of research requires suitable special tools
 - ✓ Reasonably flexible to be modified as per requirements
 - ✓ Efficient coding wrt running times and memory usages
 - ✓ Better and instant support
 - ✓ Will enable us to study models of combined PN & QN
-

Hybrid Petri Queuing Networks Simulator

Introduction

- ✓ Customized solutions might be more beneficial
 - ✓ The nature of research requires suitable special tools
 - ✓ Reasonably flexible to be modified as per requirements
 - ✓ Efficient coding wrt running times and memory usages
 - ✓ Better and instant support
 - ✓ Will enable us to study models of combined PN & QN
 - ✓ Implemented by Java
-

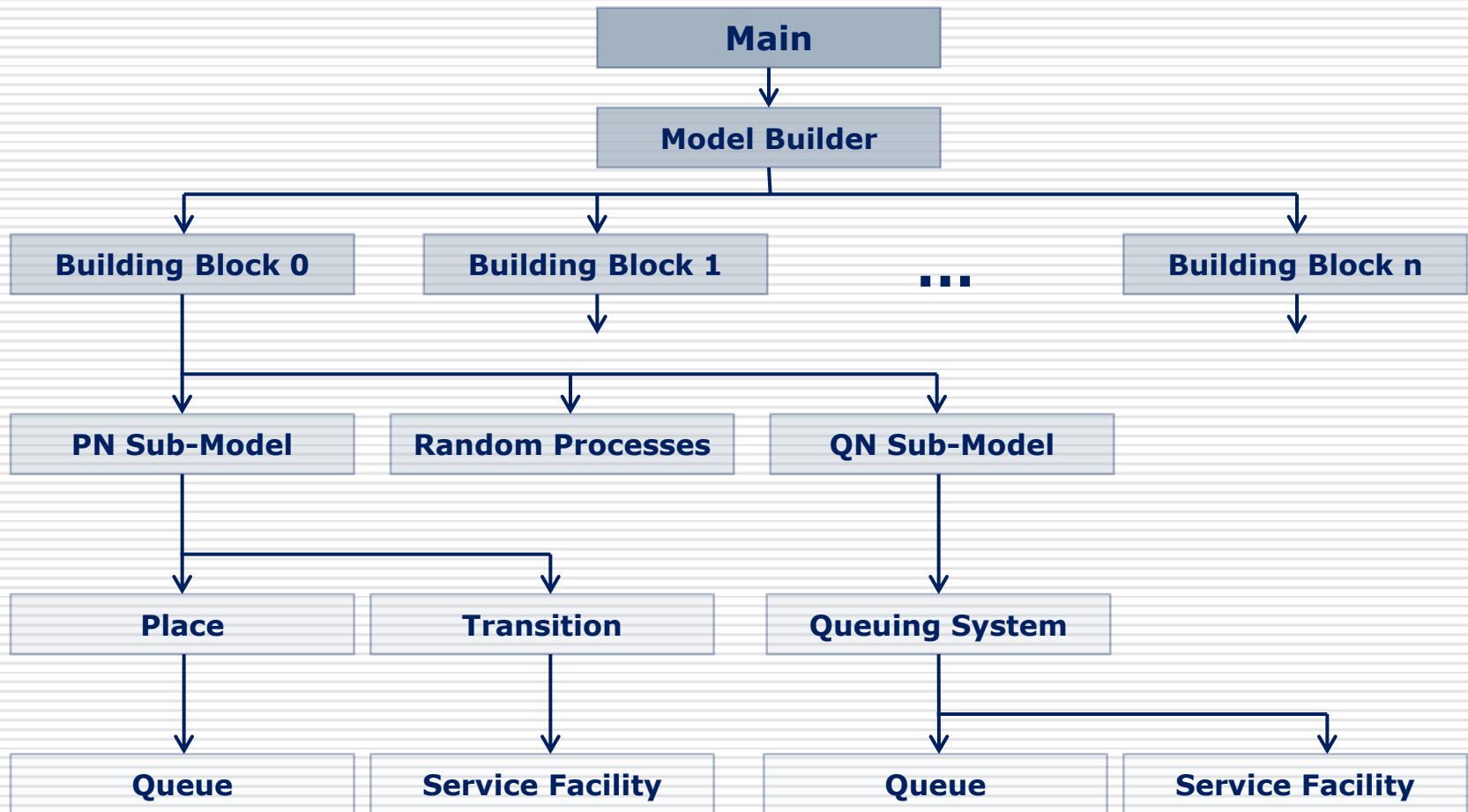
Hybrid Petri Queuing Networks Simulator

Introduction

- ✓ Customized solutions might be more beneficial
 - ✓ The nature of research requires suitable special tools
 - ✓ Reasonably flexible to be modified as per requirements
 - ✓ Efficient coding wrt running times and memory usages
 - ✓ Better and instant support
 - ✓ Will enable us to study models of combined PN & QN
 - ✓ Implemented by Java
 - ✓ Easy to learn
 - ✓ Platform-independent
-

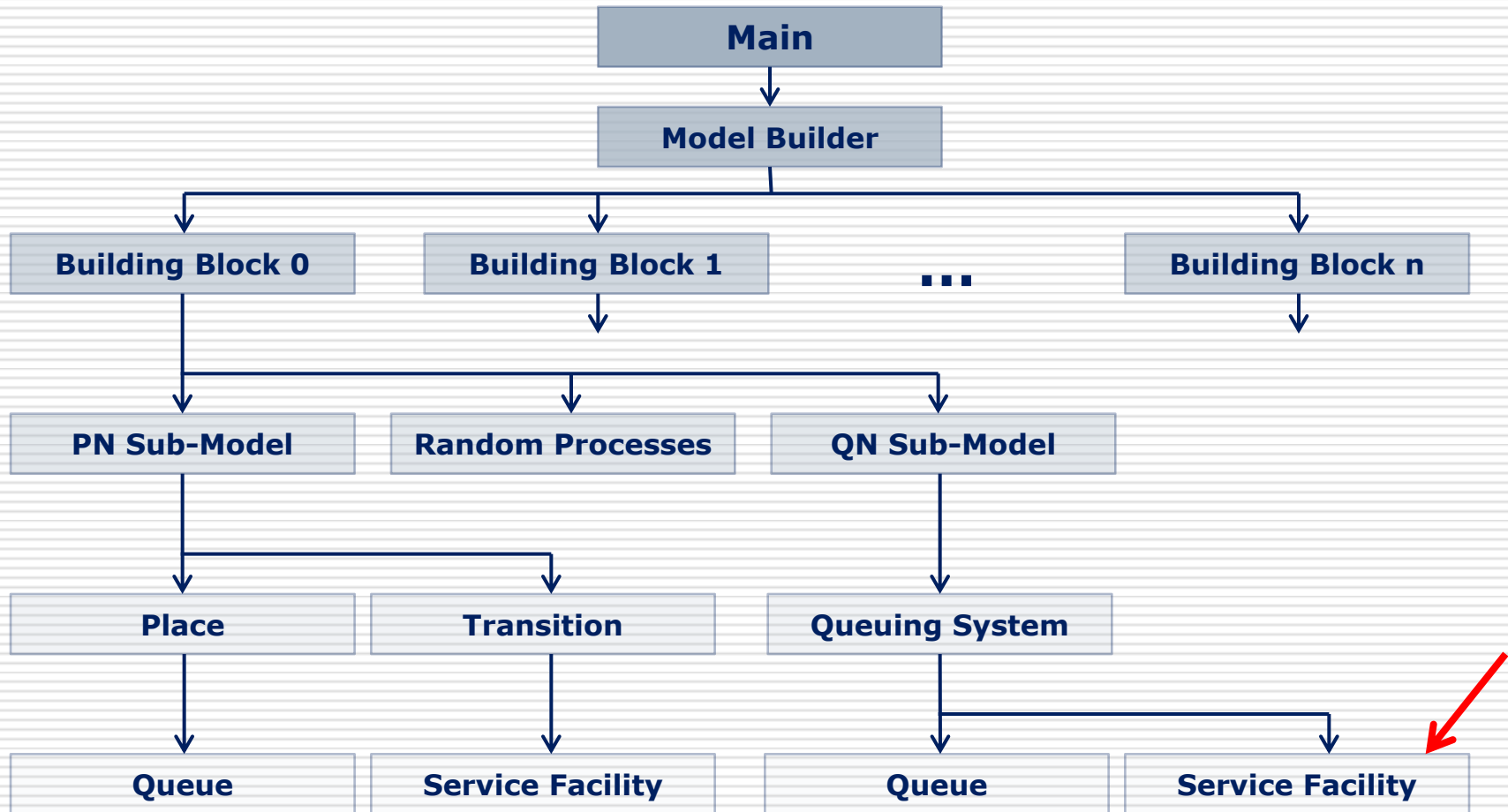
Hybrid Petri Queuing Networks Simulator

Class Hierarchy



Hybrid Petri Queuing Networks Simulator

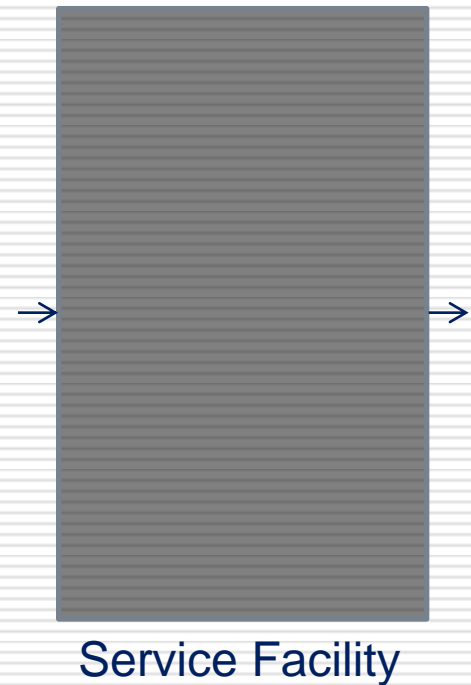
Class Hierarchy



Hybrid Petri Queuing Networks Simulator

Service Facility (SF)

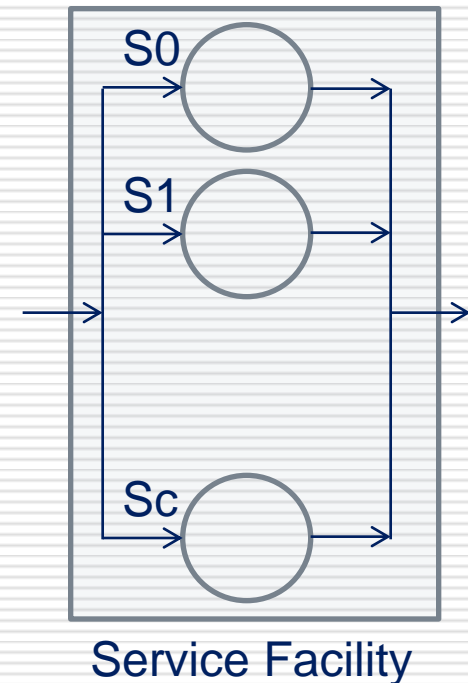
- ✓ Provides service for customers



Hybrid Petri Queuing Networks Simulator

Service Facility (SF)

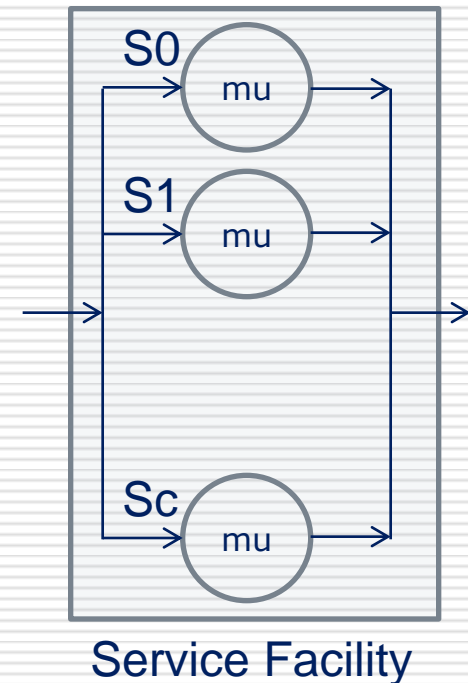
- ✓ Provides service for customers
- ✓ May have one or more servers



Hybrid Petri Queuing Networks Simulator

Service Facility (SF)

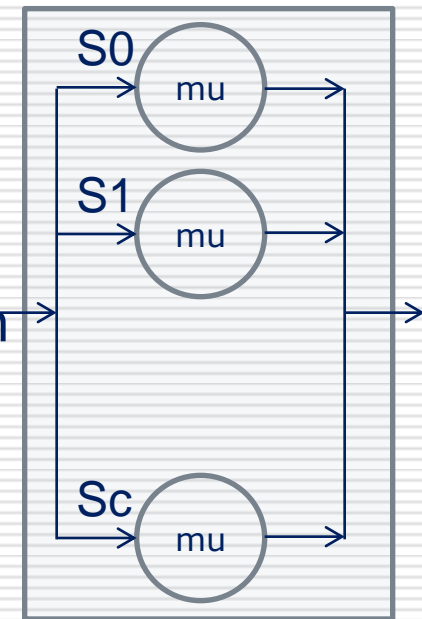
- ✓ Provides service for customers
- ✓ May have one or more servers
- ✓ Servers are homogeneous



Hybrid Petri Queuing Networks Simulator

Service Facility (SF)

- ✓ Provides service for customers
- ✓ May have one or more servers
- ✓ Servers are homogeneous
- ✓ Sharing Service Time (ST) distribution

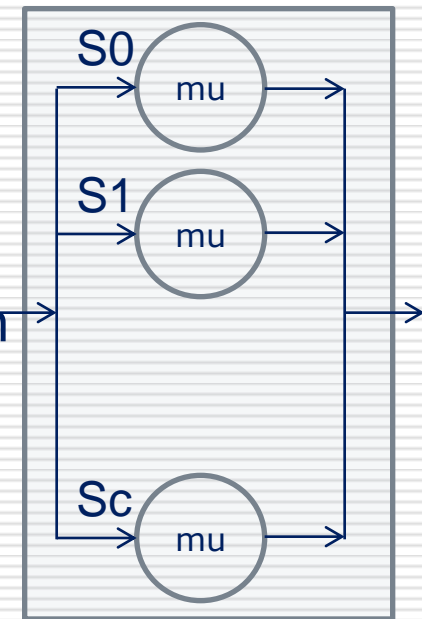


Service Facility

Hybrid Petri Queuing Networks Simulator

Service Facility (SF)

- ✓ Provides service for customers
- ✓ May have one or more servers
- ✓ Servers are homogeneous
 - ✓ Sharing Service Time (ST) distribution
 - ✓ Multi-STs for Multi-Class jobs

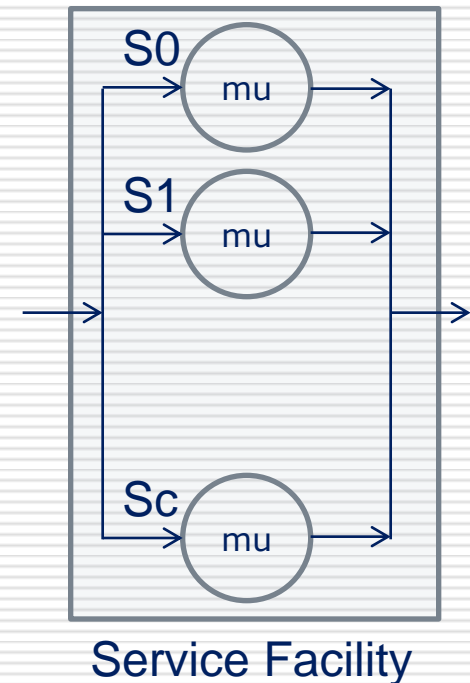


Service Facility

Hybrid Petri Queuing Networks Simulator

Service Facility (SF)

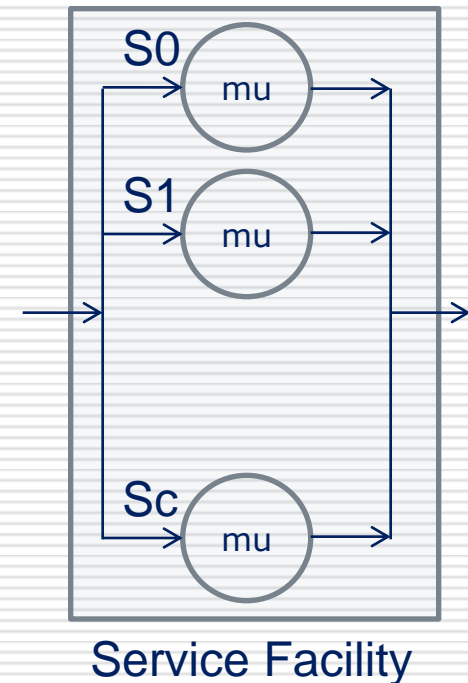
- ✓ Server Selection Disciplines (SSDs)



Hybrid Petri Queuing Networks Simulator

Service Facility (SF)

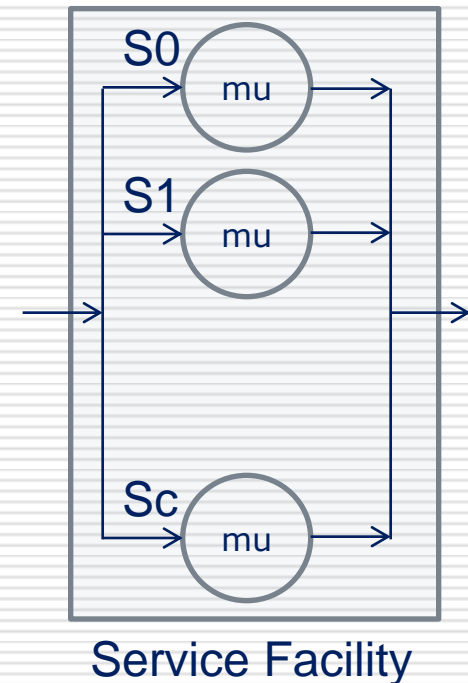
- ✓ Server Selection Disciplines (SSDs)
 - ✓ In Order



Hybrid Petri Queuing Networks Simulator

Service Facility (SF)

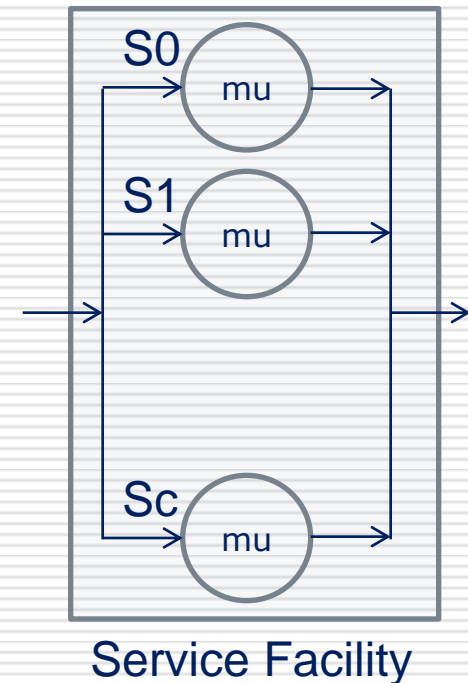
- ✓ Server Selection Disciplines (SSDs)
 - ✓ In Order
 - ✓ Random



Hybrid Petri Queuing Networks Simulator

Service Facility (SF)

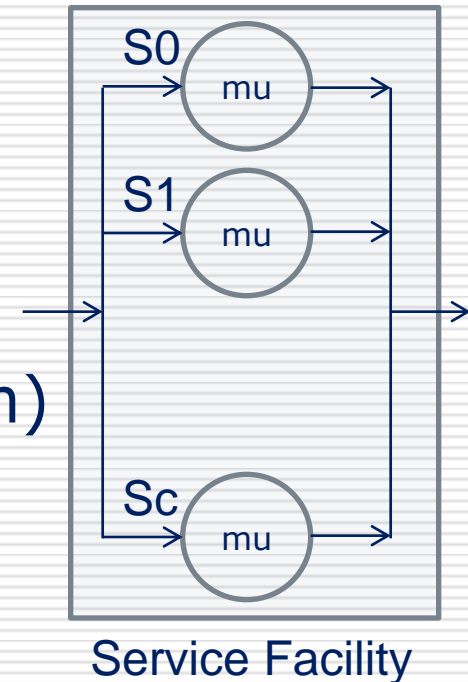
- ✓ Server Selection Disciplines (SSDs)
 - ✓ In Order
 - ✓ Random
 - ✓ Equity



Hybrid Petri Queuing Networks Simulator

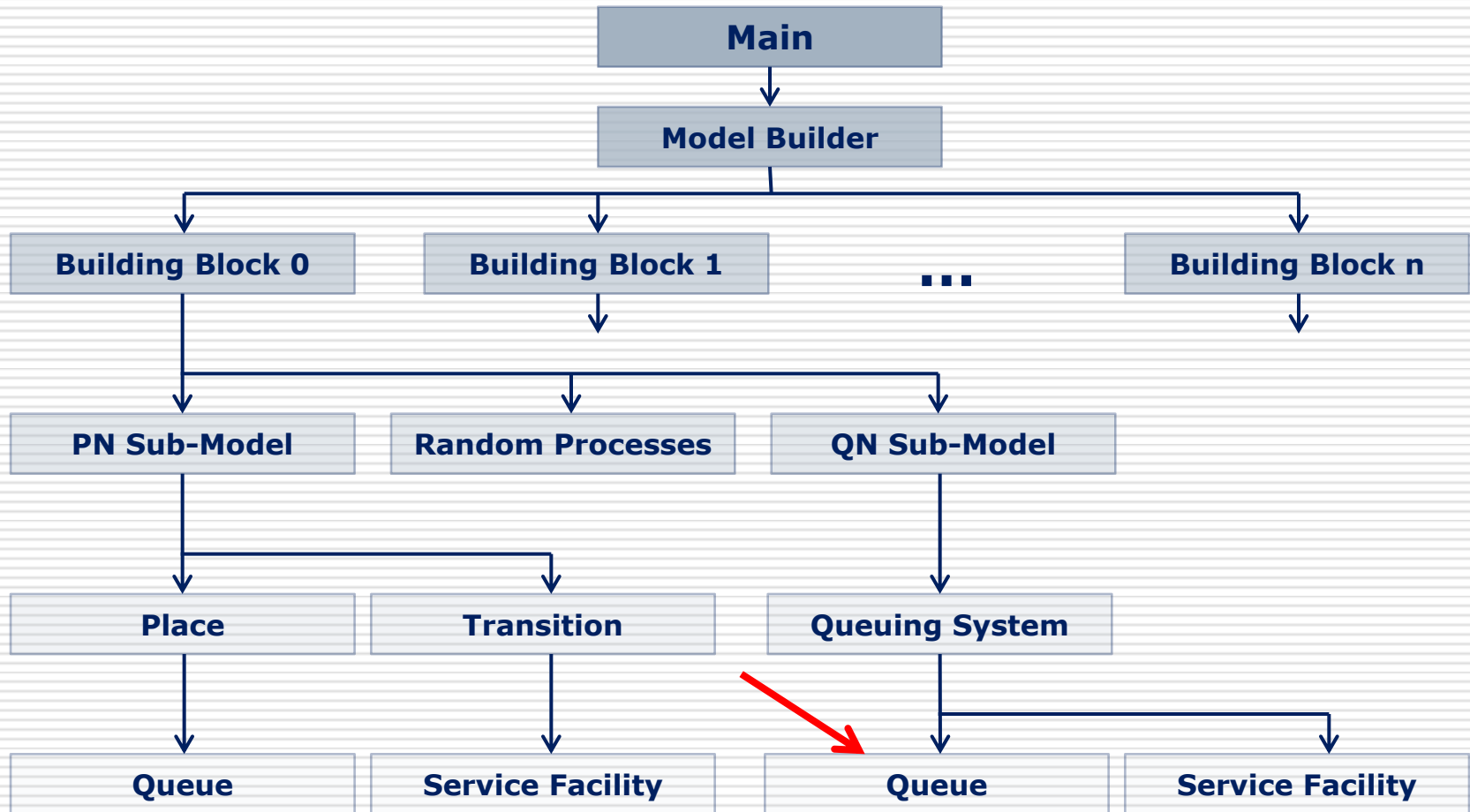
Service Facility (SF)

- ✓ Server Selection Disciplines (SSDs)
 - ✓ In Order
 - ✓ Random
 - ✓ Equity
- ✓ Infinite Servers (aka Delay System)
if c assumes large enough values



Hybrid Petri Queuing Networks Simulator

Class Hierarchy



Hybrid Petri Queuing Networks Simulator

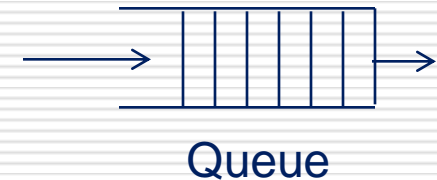
Queue (Q)

- ✓ Customers will join a queue (waiting line) until a server is available
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

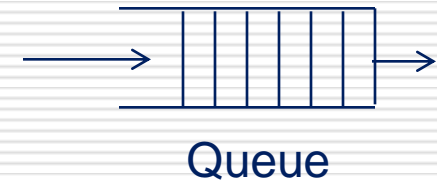
- ✓ Customers will join a queue (waiting line) until a server is available



Hybrid Petri Queuing Networks Simulator

Queue (Q)

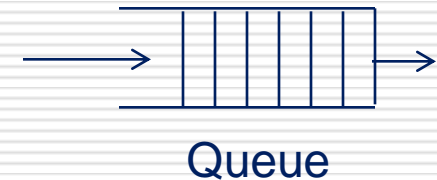
- ✓ Customers will join a queue (waiting line) until a server is available
- ✓ Can be finite or infinite (≥ 0)



Hybrid Petri Queuing Networks Simulator

Queue (Q)

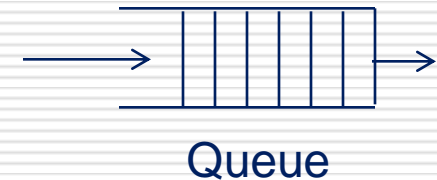
- ✓ Customers will join a queue (waiting line) until a server is available
- ✓ Can be finite or infinite (≥ 0)
- ✓ Queuing Disciplines (QDs)



Hybrid Petri Queuing Networks Simulator

Queue (Q)

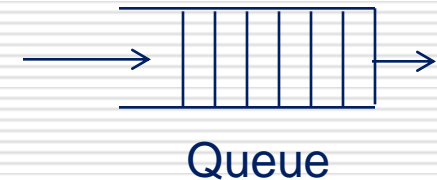
- ✓ Customers will join a queue (waiting line) until a server is available
- ✓ Can be finite or infinite (≥ 0)
- ✓ Queuing Disciplines (QDs)
 - ✓ Non-Preemptive
 - ✓ Preemptive



Hybrid Petri Queuing Networks Simulator

Queue (Q)

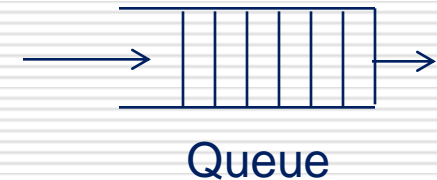
- ✓ Customers will join a queue (waiting line) until a server is available
- ✓ Can be finite or infinite (≥ 0)
- ✓ Queuing Disciplines (QDs)
 - ✓ Non-Preemptive
 - ✓ Preemptive
 - ✓ Resume (PR)



Hybrid Petri Queuing Networks Simulator

Queue (Q)

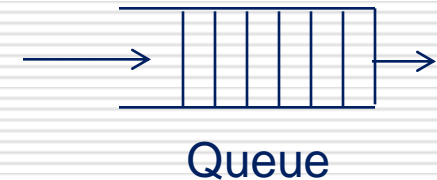
- ✓ Customers will join a queue (waiting line) until a server is available
- ✓ Can be finite or infinite (≥ 0)
- ✓ Queuing Disciplines (QDs)
 - ✓ Non-Preemptive
 - ✓ Preemptive
 - ✓ Resume (PR)
 - ✓ Repeat Identical (PRI)



Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ Customers will join a queue (waiting line) until a server is available
- ✓ Can be finite or infinite (≥ 0)
- ✓ Queuing Disciplines (QDs)
 - ✓ Non-Preemptive
 - ✓ Resume (PR)
 - ✓ Repeat Identical (PRI)
 - ✓ Repeat Different (PRD)



Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ First-Come, First-Served (FCFS)
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ First-Come, First-Served (FCFS)
 - ✓ Last-Come, First-Served (LCFS)
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ First-Come, First-Served (FCFS)
 - ✓ Last-Come, First-Served (LCFS)
 - ✓ Non-Preemptive -> Head-Of-the-Line (HOL)
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ First-Come, First-Served (FCFS)
 - ✓ Last-Come, First-Served (LCFS)
 - ✓ Non-Preemptive -> Head-Of-the-Line (HOL)
 - ✓ Preemptive-Resume (LCFS-PR)
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ First-Come, First-Served (FCFS)
 - ✓ Last-Come, First-Served (LCFS)
 - ✓ Non-Preemptive -> Head-Of-the-Line (HOL)
 - ✓ Preemptive-Resume (LCFS-PR)
 - ✓ Random-Selection-for-Service (RSS)
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ First-Come, First-Served (FCFS)
 - ✓ Last-Come, First-Served (LCFS)
 - ✓ Non-Preemptive -> Head-Of-the-Line (HOL)
 - ✓ Preemptive-Resume (LCFS-PR)
 - ✓ Random-Selection-for-Service (RSS)
 - ✓ Priority Service (PRI)
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ Processor-Sharing
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ Processor-Sharing
 - ✓ Jobs enter a single queue and wait in FCFS fashion
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ Processor-Sharing
 - ✓ Jobs enter a single queue and wait in FCFS fashion
 - ✓ Jobs receive service according to the Round-Robin scheduling
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ Processor-Sharing
 - ✓ Jobs enter a single queue and wait in FCFS fashion
 - ✓ Jobs receive service according to the Round-Robin scheduling
 - ✓ Quantum Δt shrinks to zero
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ Processor-Sharing
 - ✓ Jobs enter a single queue and wait in FCFS fashion
 - ✓ Jobs receive service according to the Round-Robin scheduling
 - ✓ Quantum Δt shrinks to zero
 - ✓ Pros and cons of PS
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ Balking: Refusing to join a queue
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ Balking: Refusing to join a queue
 - ✓ Reneging: Leaving a queue before service
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ Balking: Refusing to join a queue
 - ✓ Reneging: Leaving a queue before service
 - ✓ Blocking: Turned away
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ Balking: Refusing to join a queue
 - ✓ Reneging: Leaving a queue before service
 - ✓ Blocking: Turned away
 - ✓ Various combinations might be of interest
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ Balking: Refusing to join a queue
 - ✓ Reneging: Leaving a queue before service
 - ✓ Blocking: Turned away
 - ✓ Various combinations might be of interest
 - ✓ Complete Balking, Complete Blocking
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ Balking: Refusing to join a queue
 - ✓ Reneging: Leaving a queue before service
 - ✓ Blocking: Turned away
 - ✓ Various combinations might be of interest
 - ✓ Complete Balking, Complete Blocking
 - ✓ Complete Balking, Partial (Independent) Blocking
-

Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ Balking: Refusing to join a queue
 - ✓ Reneging: Leaving a queue before service
 - ✓ Blocking: Turned away
 - ✓ Various combinations might be of interest
 - ✓ Complete Balking, Complete Blocking
 - ✓ Complete Balking, Partial (Independent) Blocking
 - ✓ Partial Balking, Complete Blocking
-

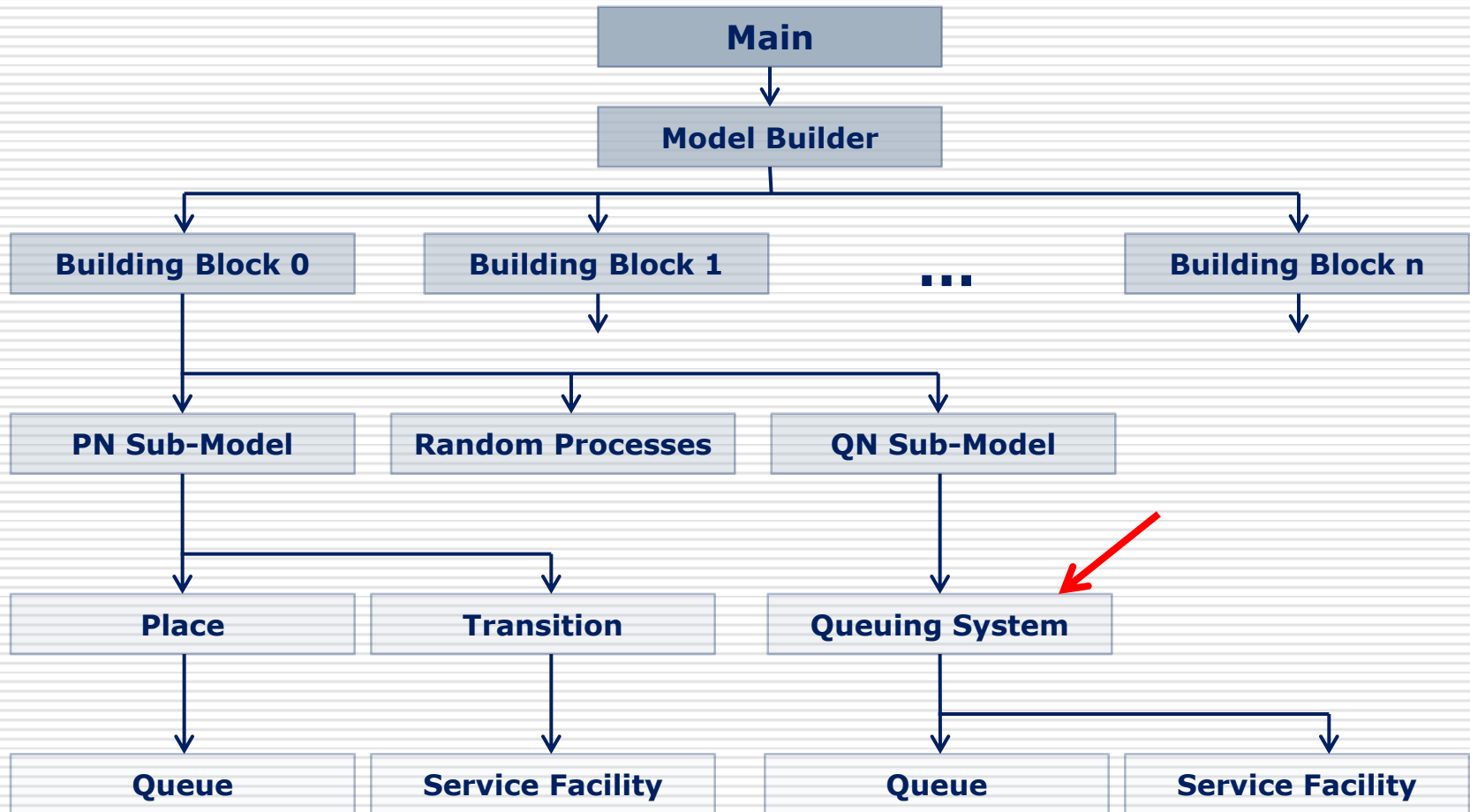
Hybrid Petri Queuing Networks Simulator

Queue (Q)

- ✓ Balking: Refusing to join a queue
 - ✓ Reneging: Leaving a queue before service
 - ✓ Blocking: Turned away
 - ✓ Various combinations might be of interest
 - ✓ Complete Balking, Complete Blocking
 - ✓ Complete Balking, Partial (Independent) Blocking
 - ✓ Partial Balking, Complete Blocking
 - ✓ Partial Balking, Partial Blocking
-

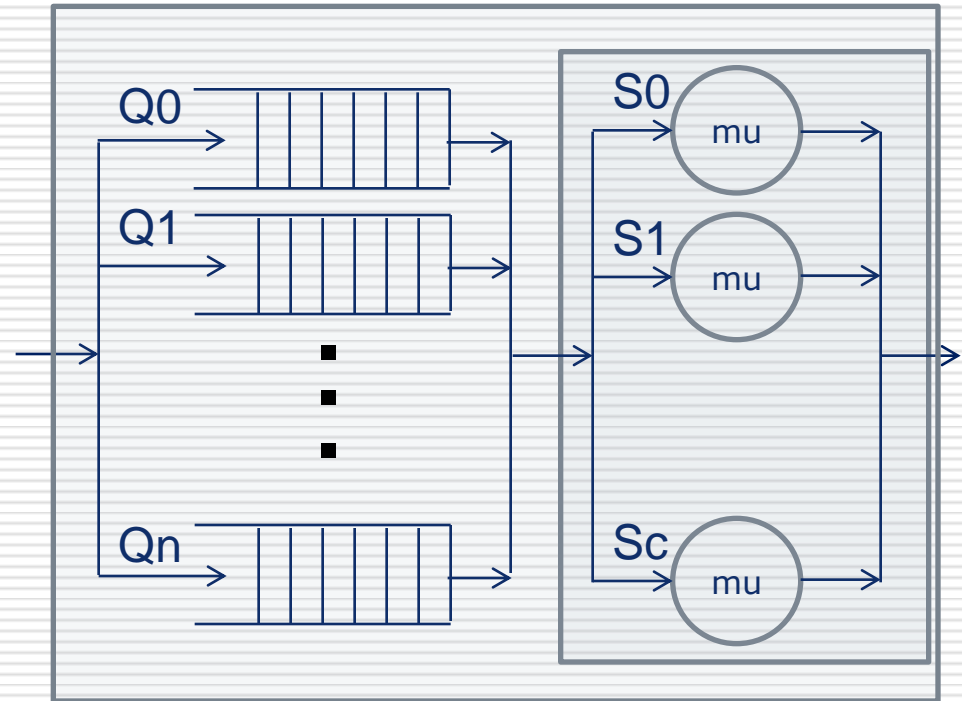
Hybrid Petri Queuing Networks Simulator

Class Hierarchy



Hybrid Petri Queuing Networks Simulator

Queuing System (QS)

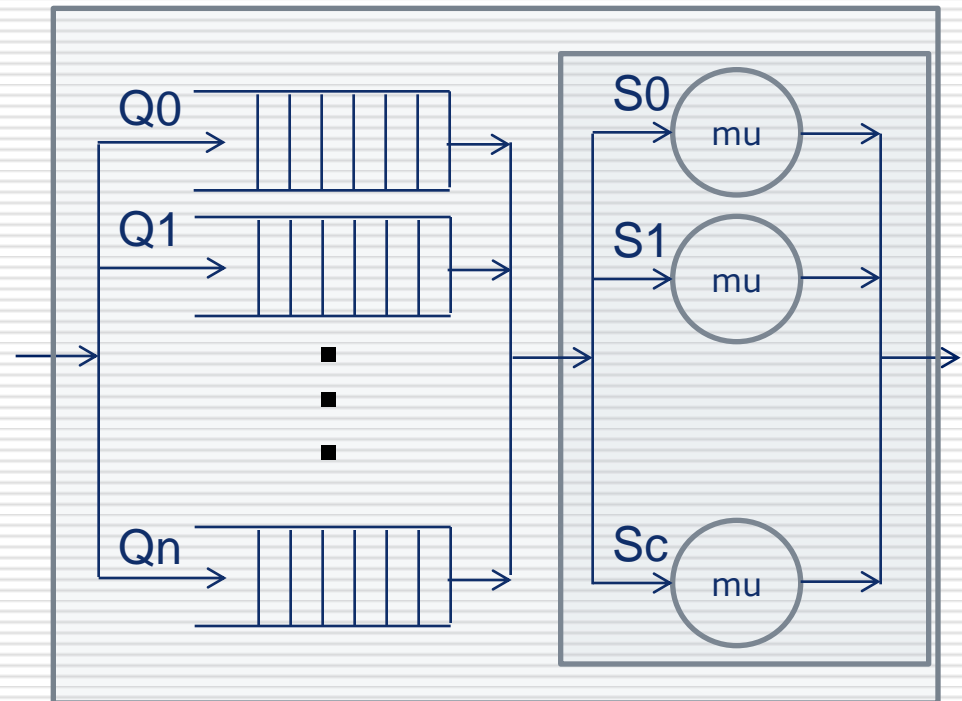


Queuing System

Hybrid Petri Queuing Networks Simulator

Queuing System (QS)

- ✓ Employed in QNs

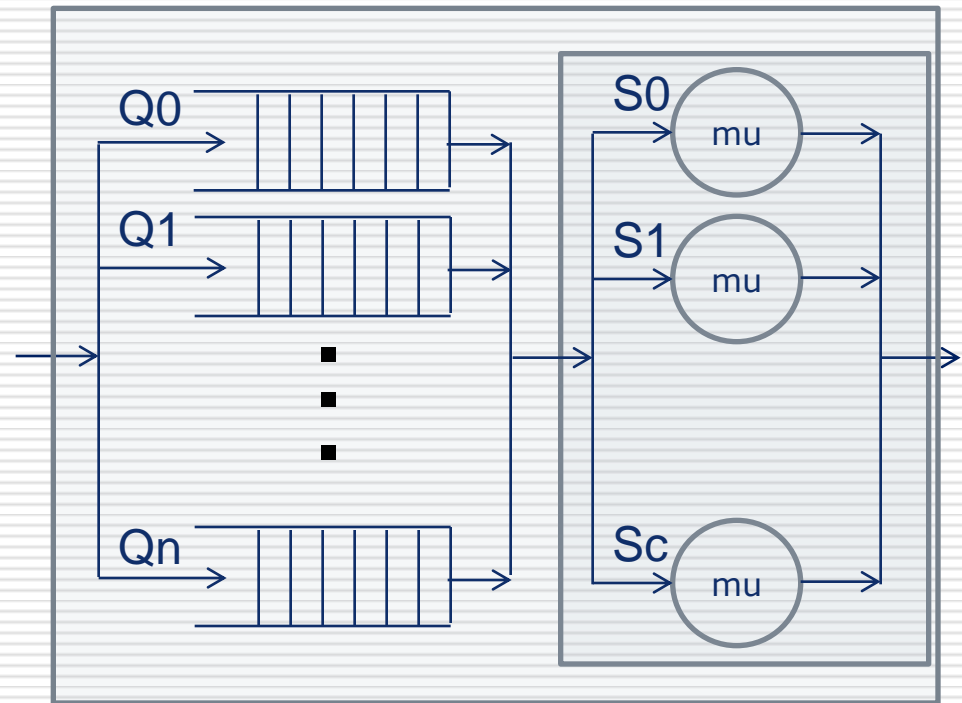


Queuing System

Hybrid Petri Queuing Networks Simulator

Queuing System (QS)

- ✓ Employed in QNs
- ✓ Multiple Servers

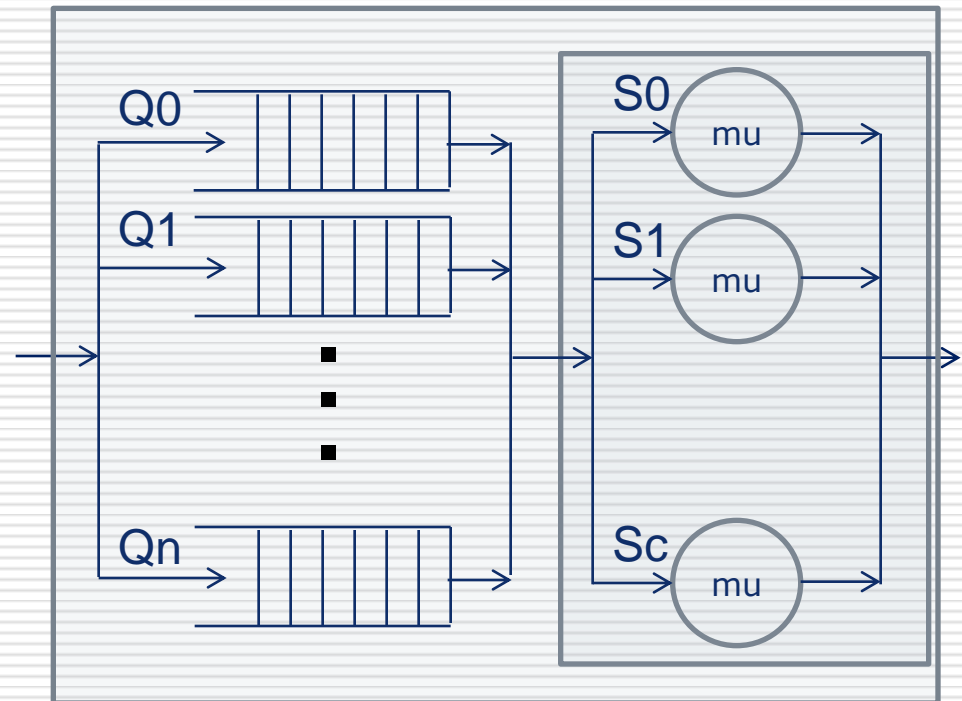


Queuing System

Hybrid Petri Queuing Networks Simulator

Queuing System (QS)

- ✓ Employed in QNs
- ✓ Multiple Servers
- ✓ Multiple Queues

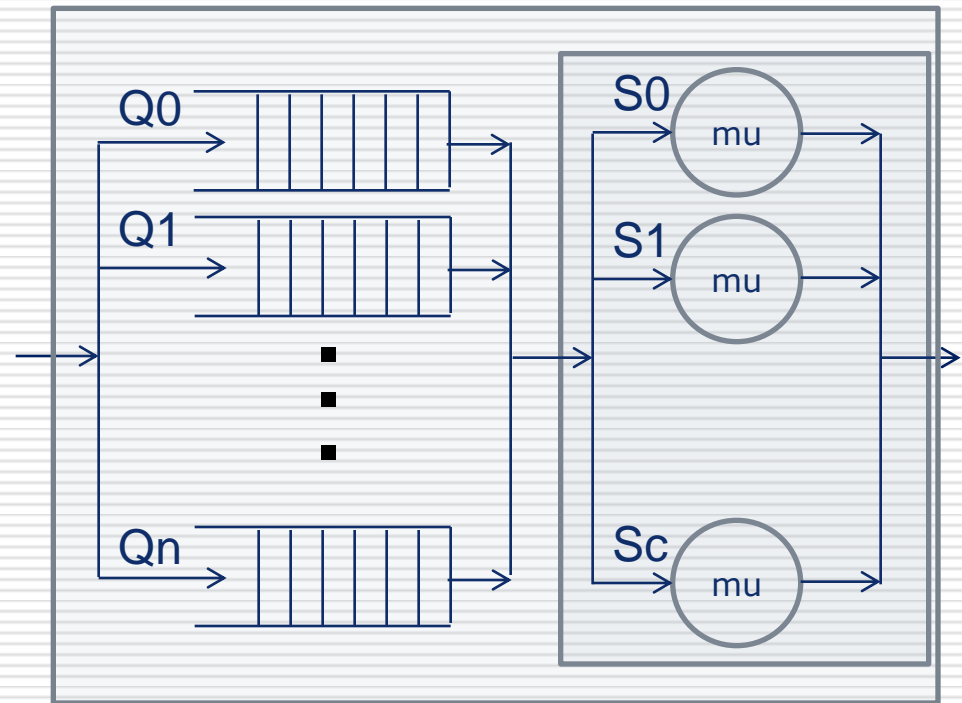


Queuing System

Hybrid Petri Queuing Networks Simulator

Queuing System (QS)

- ✓ Employed in QNs
- ✓ Multiple Servers
- ✓ Multiple Queues
- ✓ Finite or Infinite

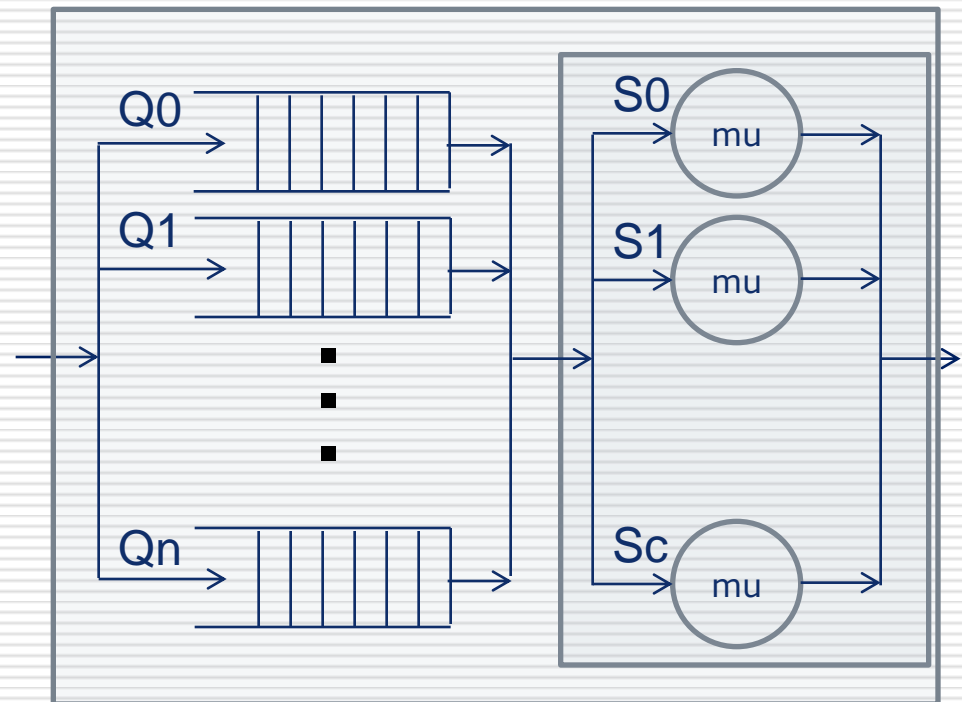


Queuing System

Hybrid Petri Queuing Networks Simulator

Queuing System (QS)

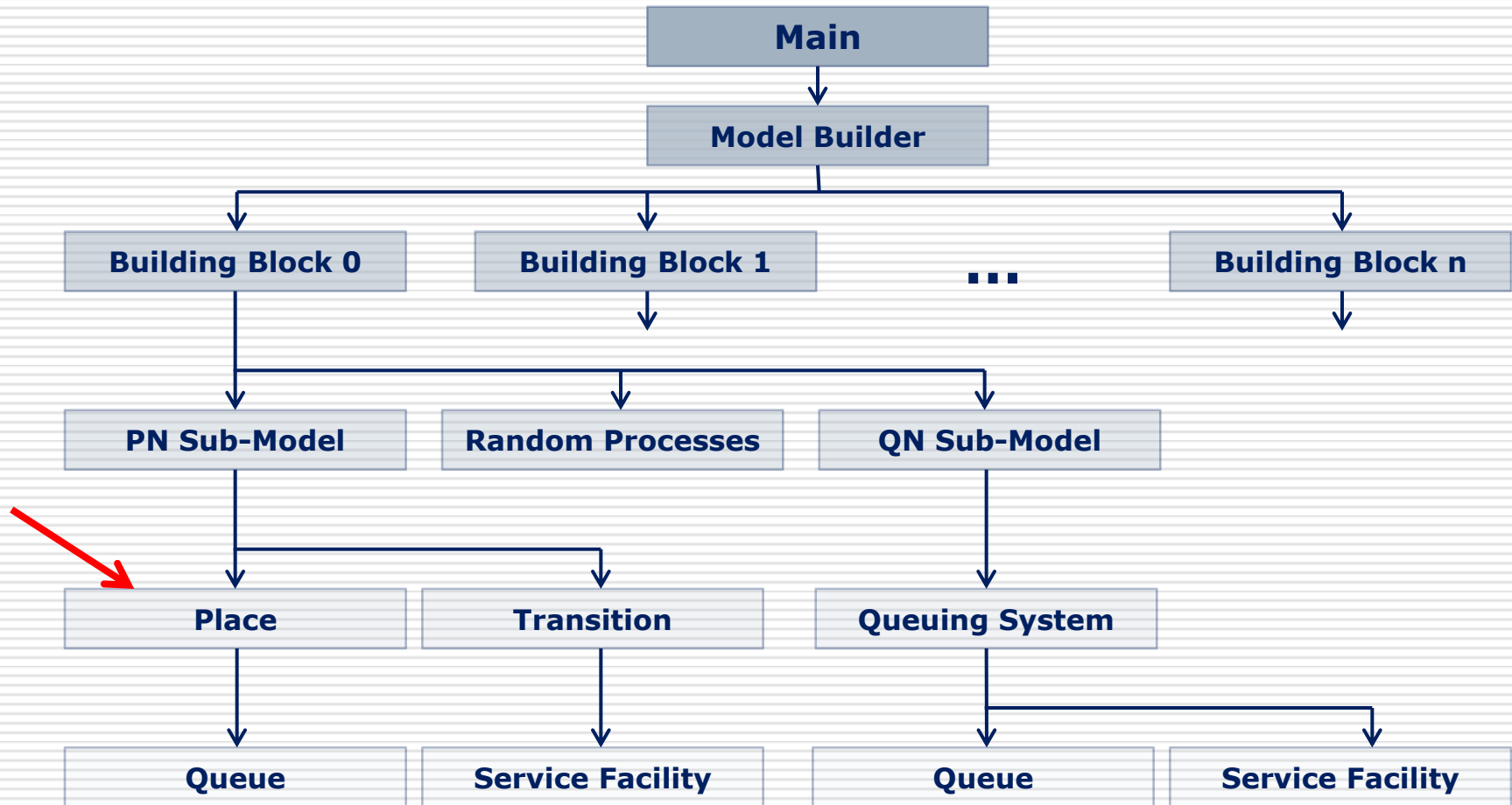
- ✓ Employed in QNs
- ✓ Multiple Servers
- ✓ Multiple Queues
- ✓ Finite or Infinite
- ✓ Various SSDs/QDs



Queuing System

Hybrid Petri Queuing Networks Simulator

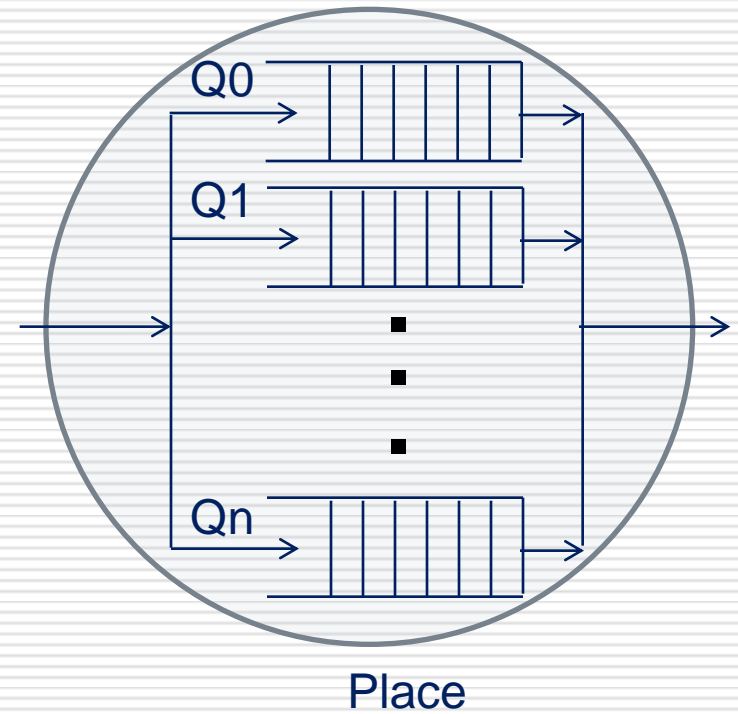
Class Hierarchy



Hybrid Petri Queuing Networks Simulator

Place (P)

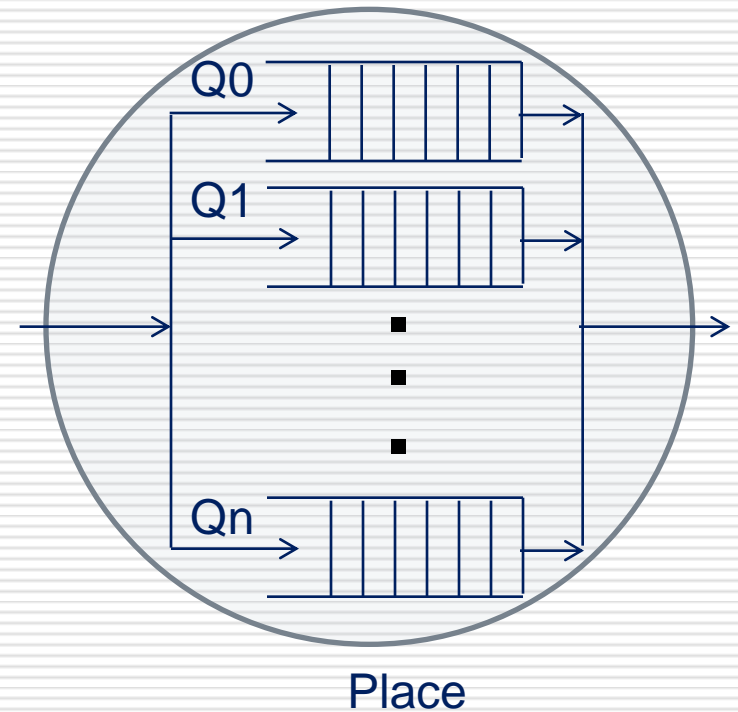
- ✓ Employed in PNs



Hybrid Petri Queuing Networks Simulator

Place (P)

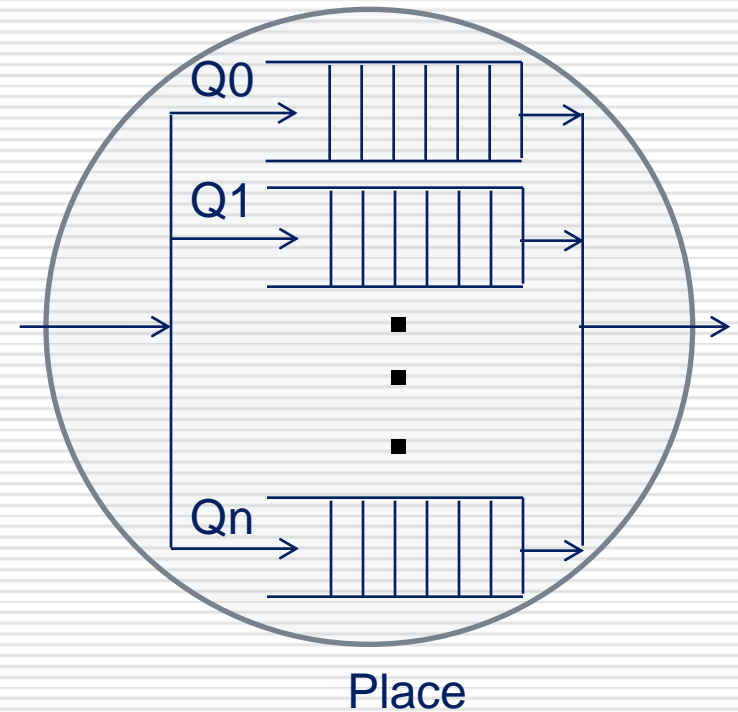
- ✓ Employed in PNs
- ✓ Multiple Queues



Hybrid Petri Queuing Networks Simulator

Place (P)

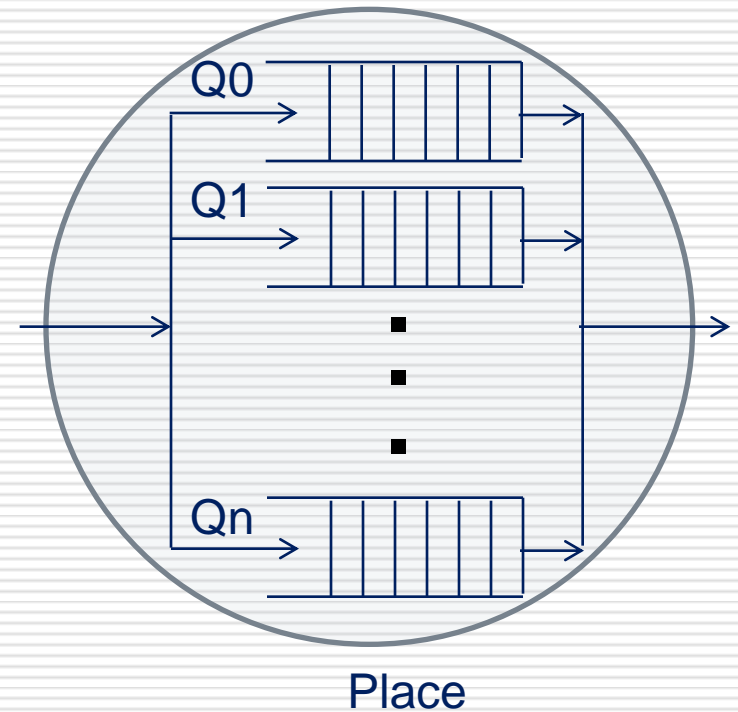
- ✓ Employed in PNs
- ✓ Multiple Queues
- ✓ Finite or Infinite



Hybrid Petri Queuing Networks Simulator

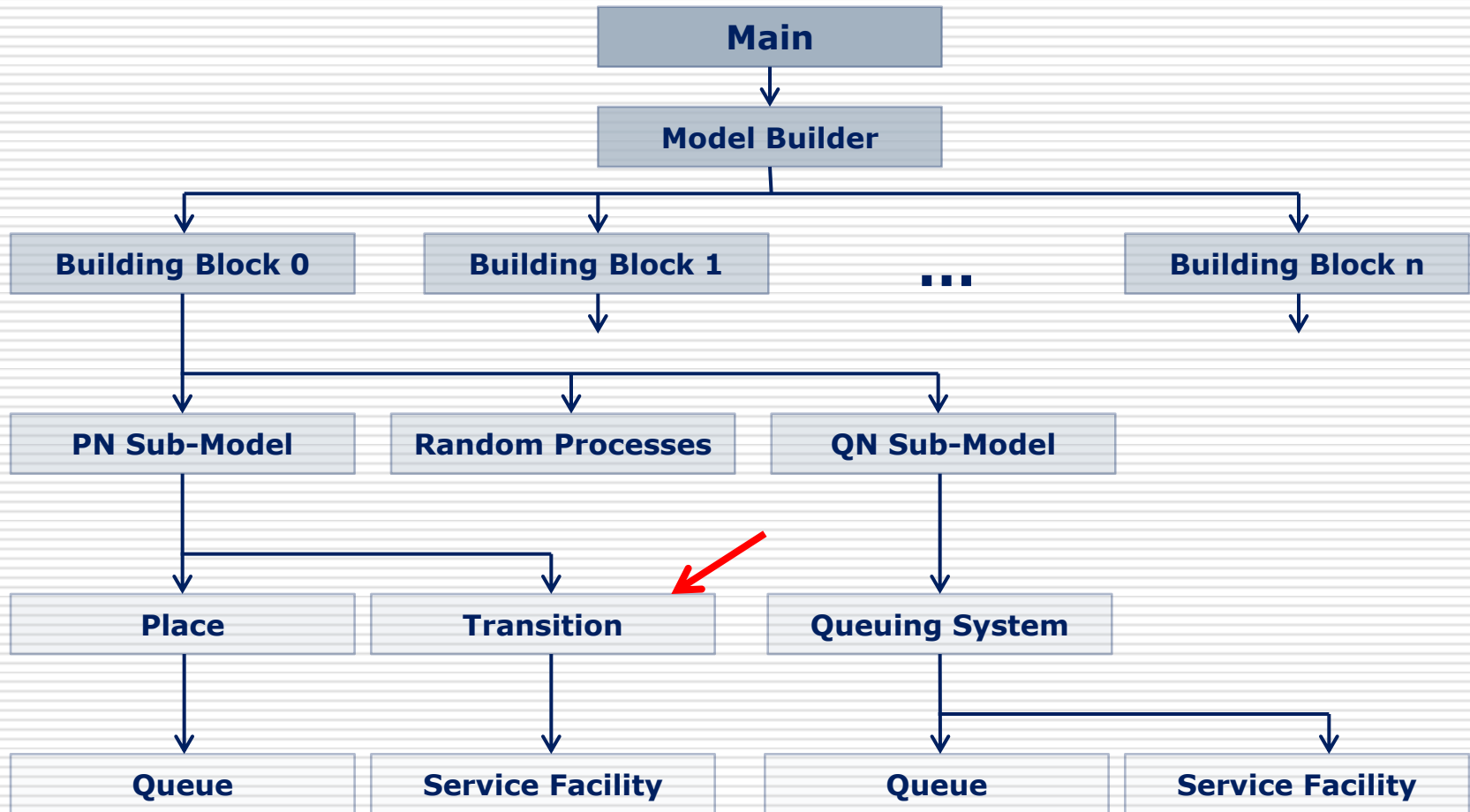
Place (P)

- ✓ Employed in PNs
- ✓ Multiple Queues
- ✓ Finite or Infinite
- ✓ Various QDs



Hybrid Petri Queuing Networks Simulator

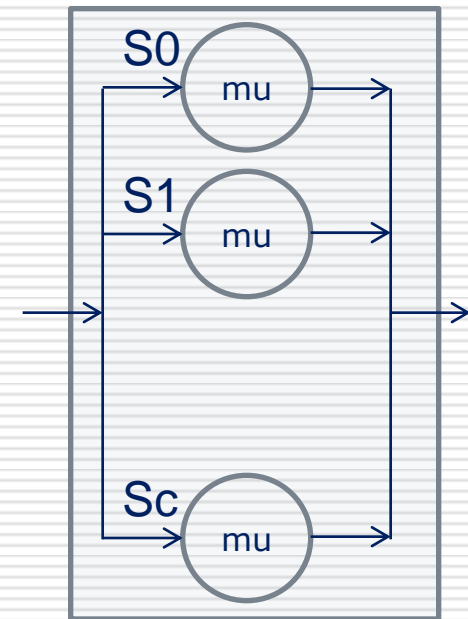
Class Hierarchy



Hybrid Petri Queuing Networks Simulator

Transition (T)

- ✓ Employed in PNs

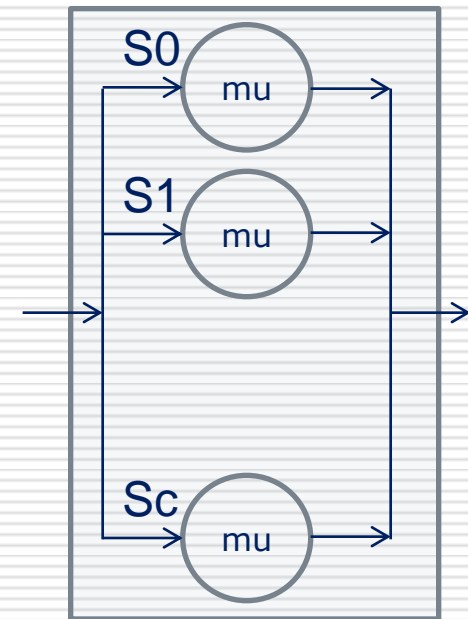


Transition

Hybrid Petri Queuing Networks Simulator

Transition (T)

- ✓ Employed in PNs
- ✓ Multiple Servers

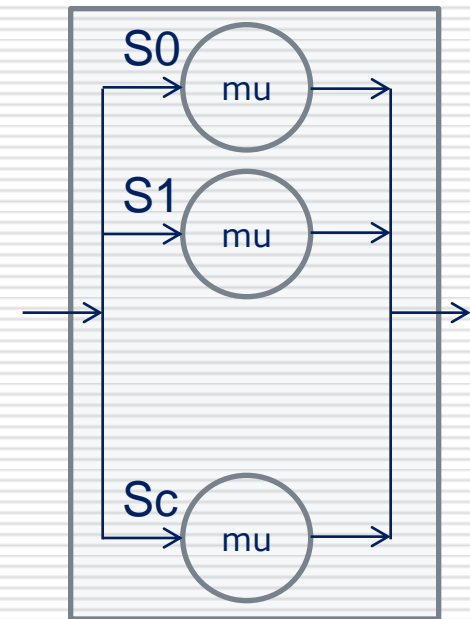


Transition

Hybrid Petri Queuing Networks Simulator

Transition (T)

- ✓ Employed in PNs
- ✓ Multiple Servers
- ✓ Finite or Infinite

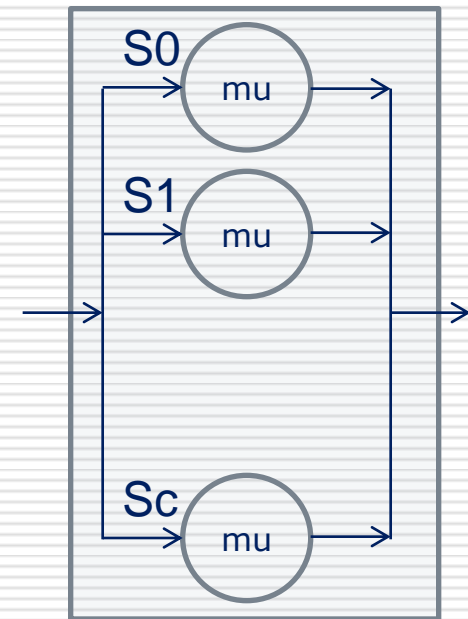


Transition

Hybrid Petri Queuing Networks Simulator

Transition (T)

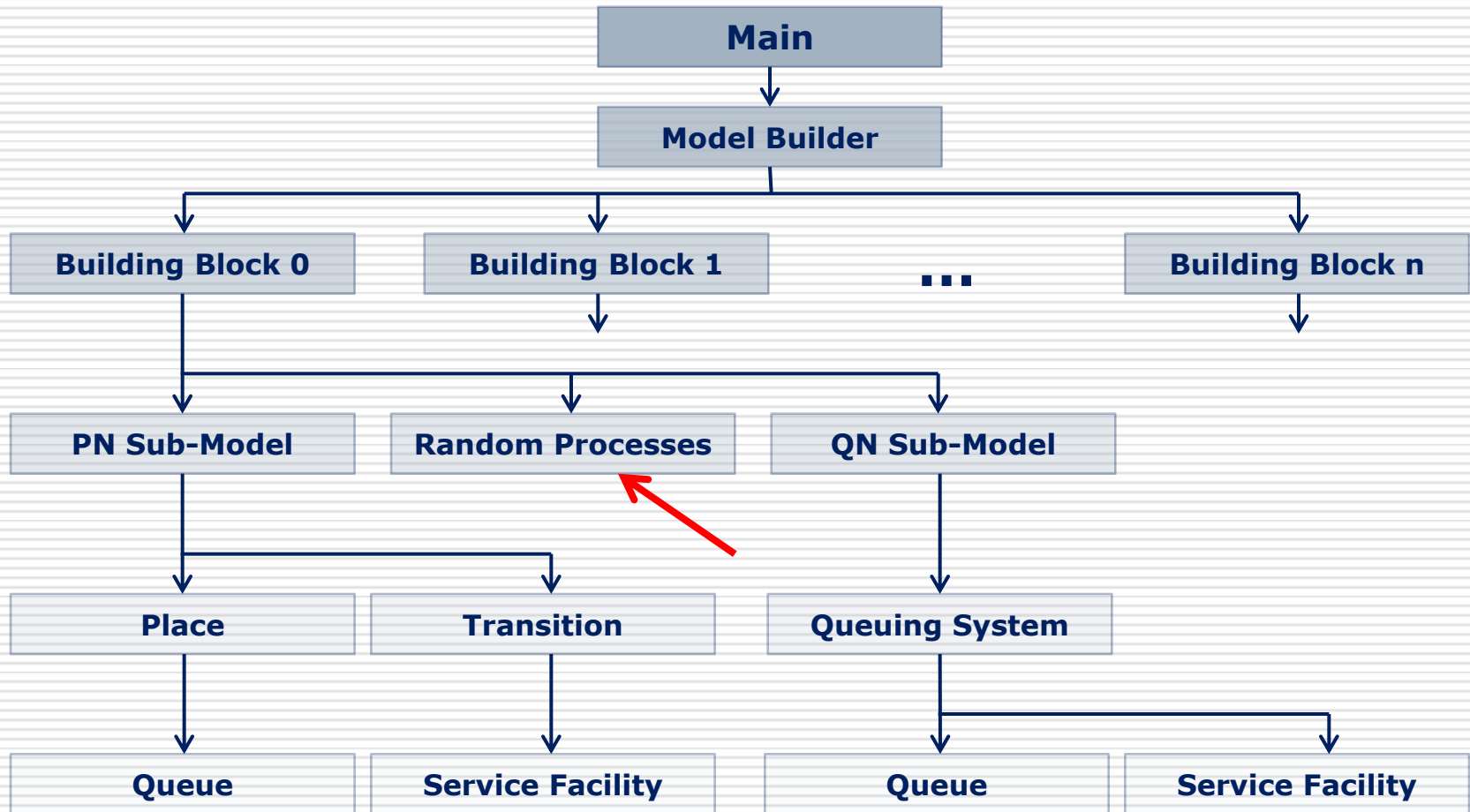
- ✓ Employed in PNs
- ✓ Multiple Servers
- ✓ Finite or Infinite
- ✓ Various SSDs



Transition

Hybrid Petri Queuing Networks Simulator

Class Hierarchy



Hybrid Petri Queuing Networks Simulator

Random Processes

- ✓ Support for some standard Probability Density Functions (PDFs)
-

Hybrid Petri Queuing Networks Simulator

Random Processes

- ✓ Support for some standard Probability Density Functions (PDFs)
 - ✓ Uniform (U)
 - ✓ Exponential (Exp)
 - ✓ Hyper-exponential (HE)
 - ✓ Generalized Exponential (GE)
-

Hybrid Petri Queuing Networks Simulator

Random Processes

- ✓ Support for some standard Probability Density Functions (PDFs)
 - ✓ Uniform (U)
 - ✓ Exponential (Exp)
 - ✓ Hyper-exponential (HE)
 - ✓ Generalized Exponential (GE)
 - ✓ Inclusion of arbitrary PDFs is straightforward
-

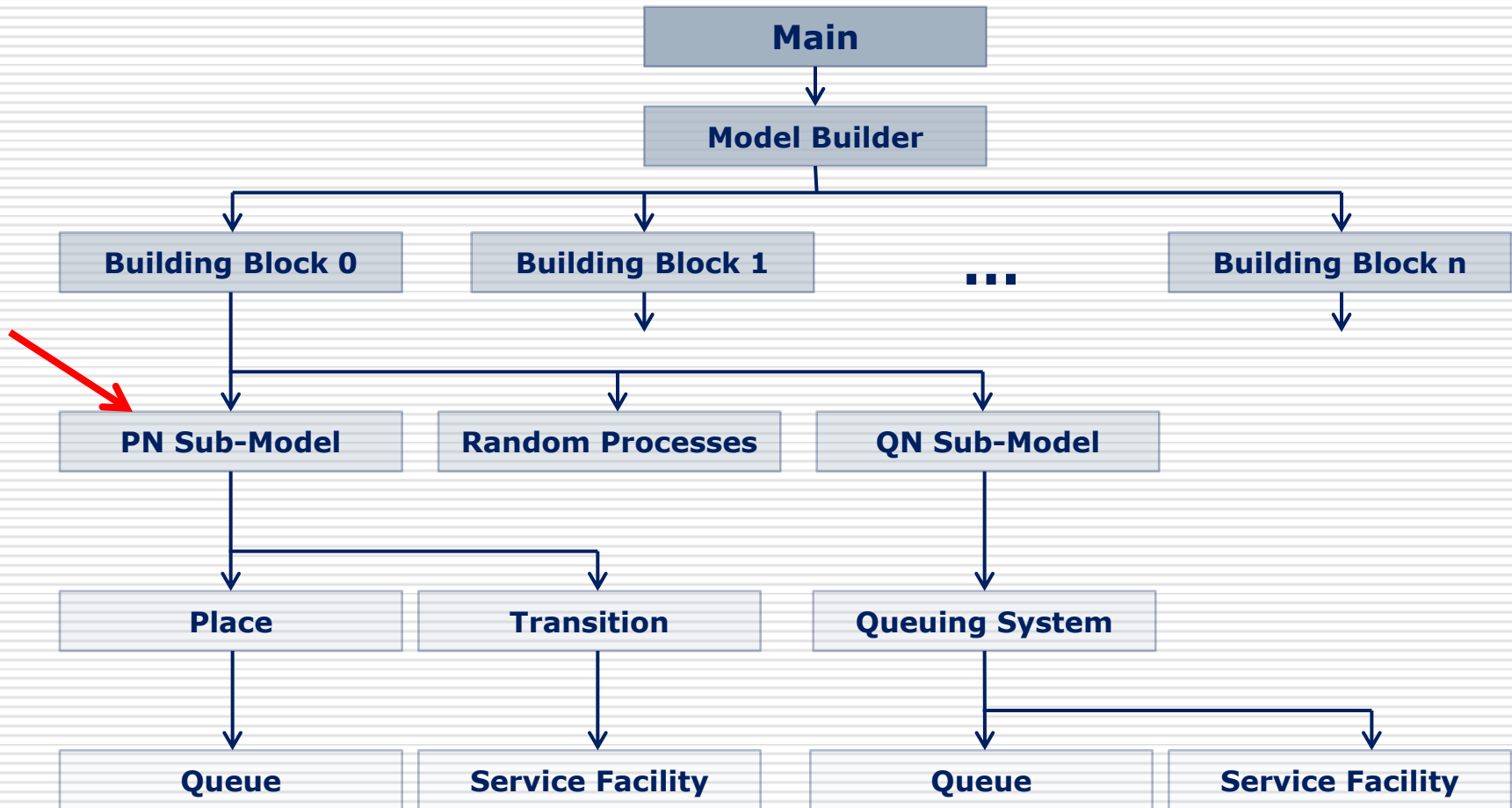
Hybrid Petri Queuing Networks Simulator

Random Processes

- ✓ Support for some standard Probability Density Functions (PDFs)
 - ✓ Uniform (U)
 - ✓ Exponential (Exp)
 - ✓ Hyper-exponential (HE)
 - ✓ Generalized Exponential (GE)
 - ✓ Inclusion of arbitrary PDFs is straightforward
 - ✓ Employed in modeling Arrival/Service Times
-

Hybrid Petri Queuing Networks Simulator

Class Hierarchy



Hybrid Petri Queuing Networks Simulator

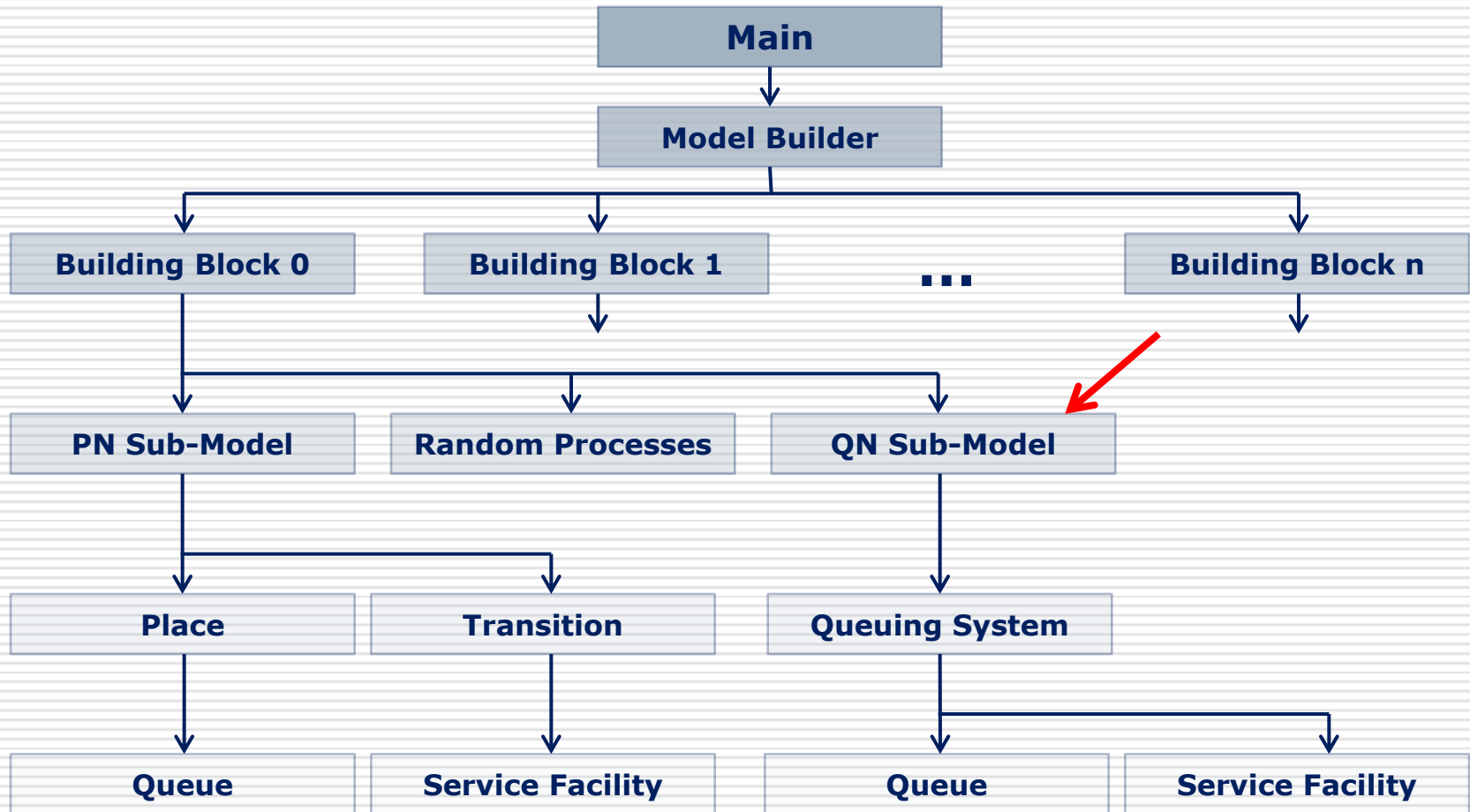
PN sub-models

- ✓ Composed of Places, Transitions, Arcs
- ✓ Arbitrary topologies
- ✓ Heterogeneous
- ✓ Support of Immediate Transitions and Inhibitors



Hybrid Petri Queuing Networks Simulator

Class Hierarchy



Hybrid Petri Queuing Networks Simulator

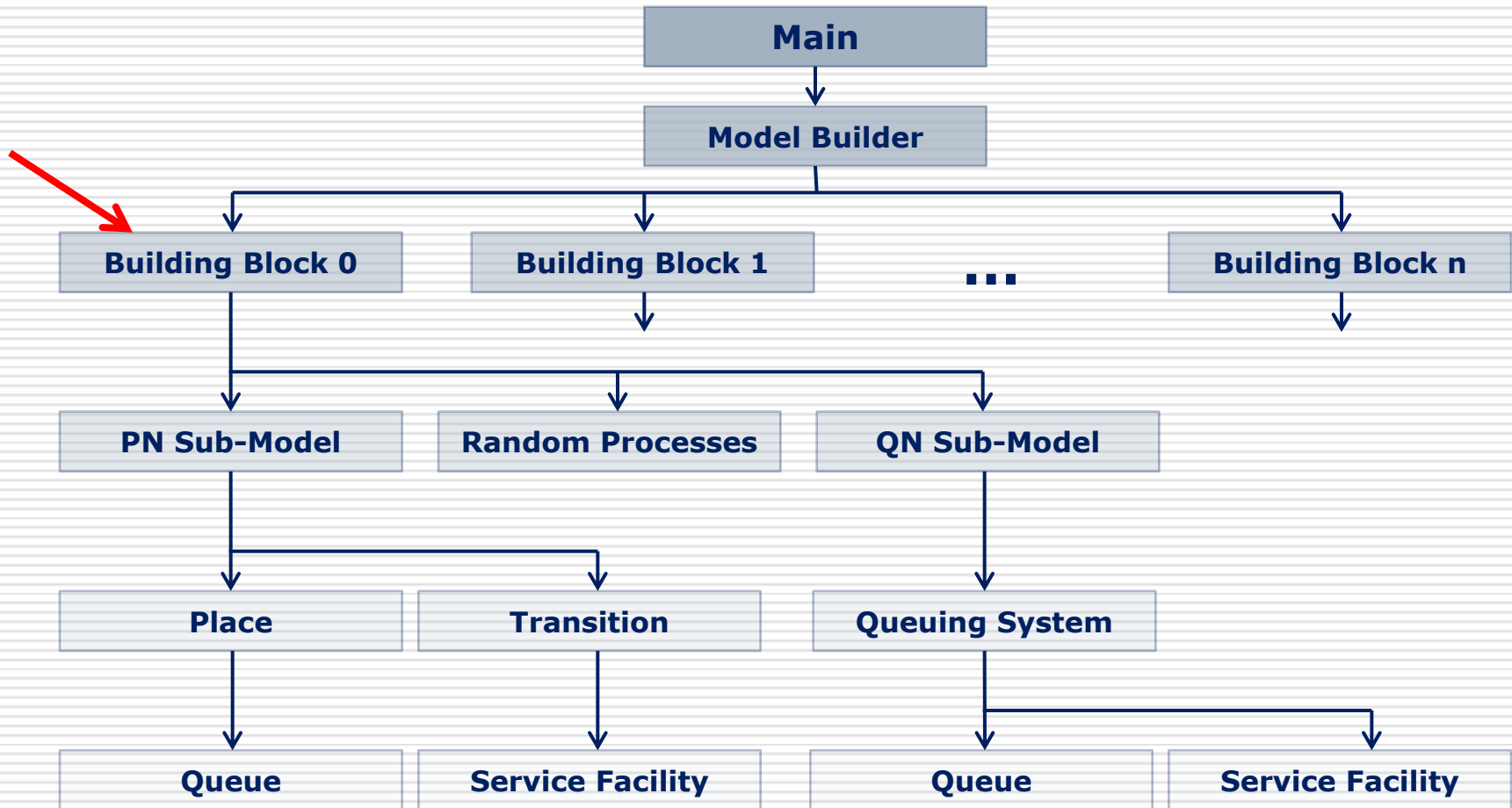
QN sub-models

- ✓ Composed of Queues and Service Facilities
- ✓ Arbitrary topologies
- ✓ Heterogeneous



Hybrid Petri Queuing Networks Simulator

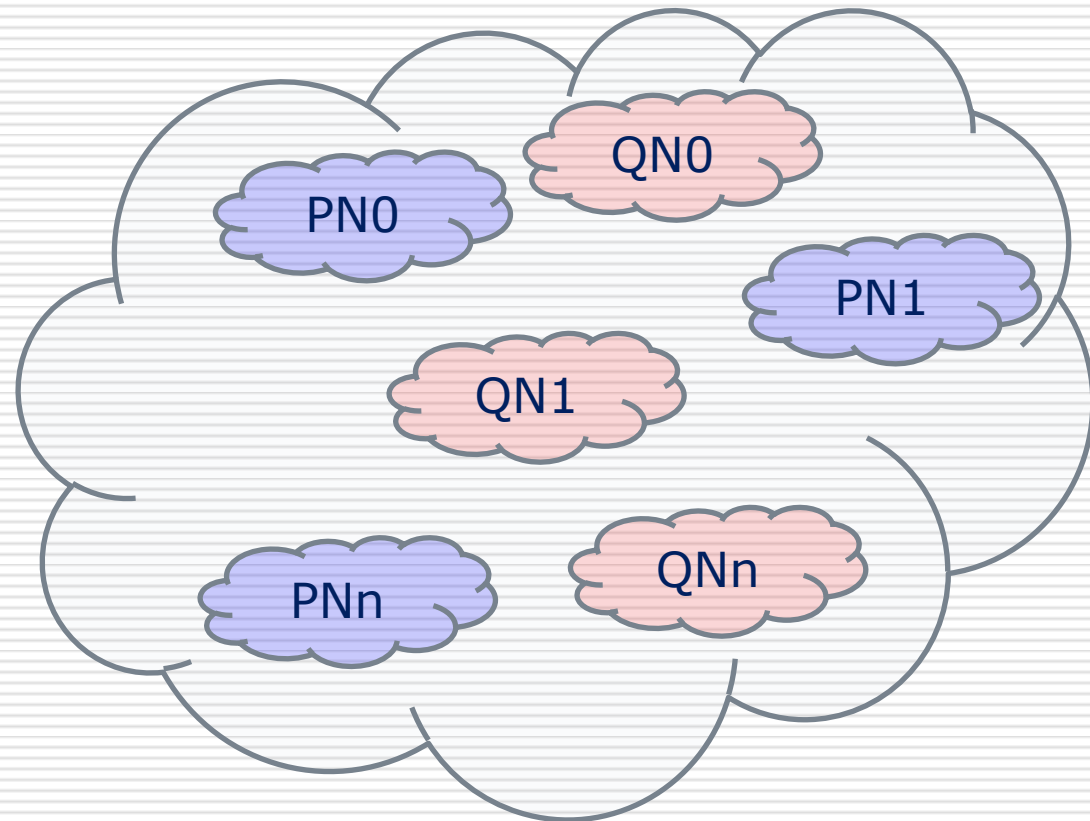
Class Hierarchy



Hybrid Petri Queuing Networks Simulator

Building Block ^(BB)

- ✓ Combined PN and QN sub-models

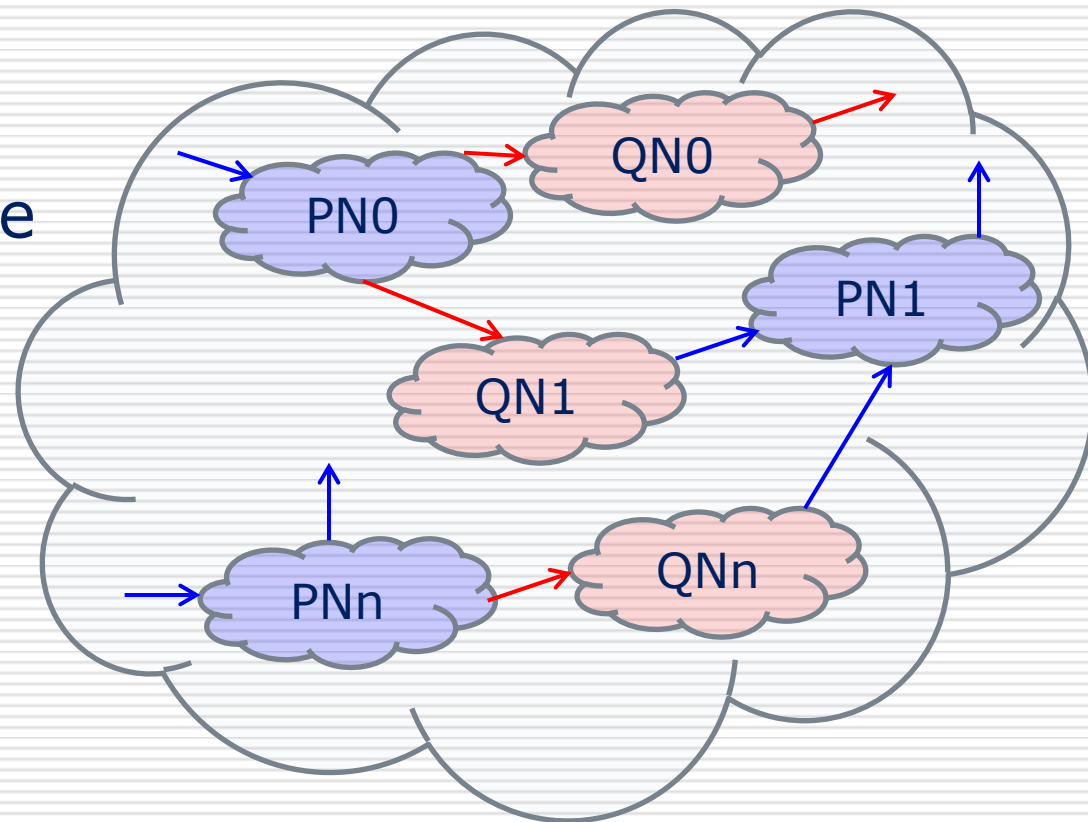


Building Block

Hybrid Petri Queuing Networks Simulator

Building Block (BB)

- ✓ Combined PN and QN sub-models
- ✓ Sub-models may inter-communicate

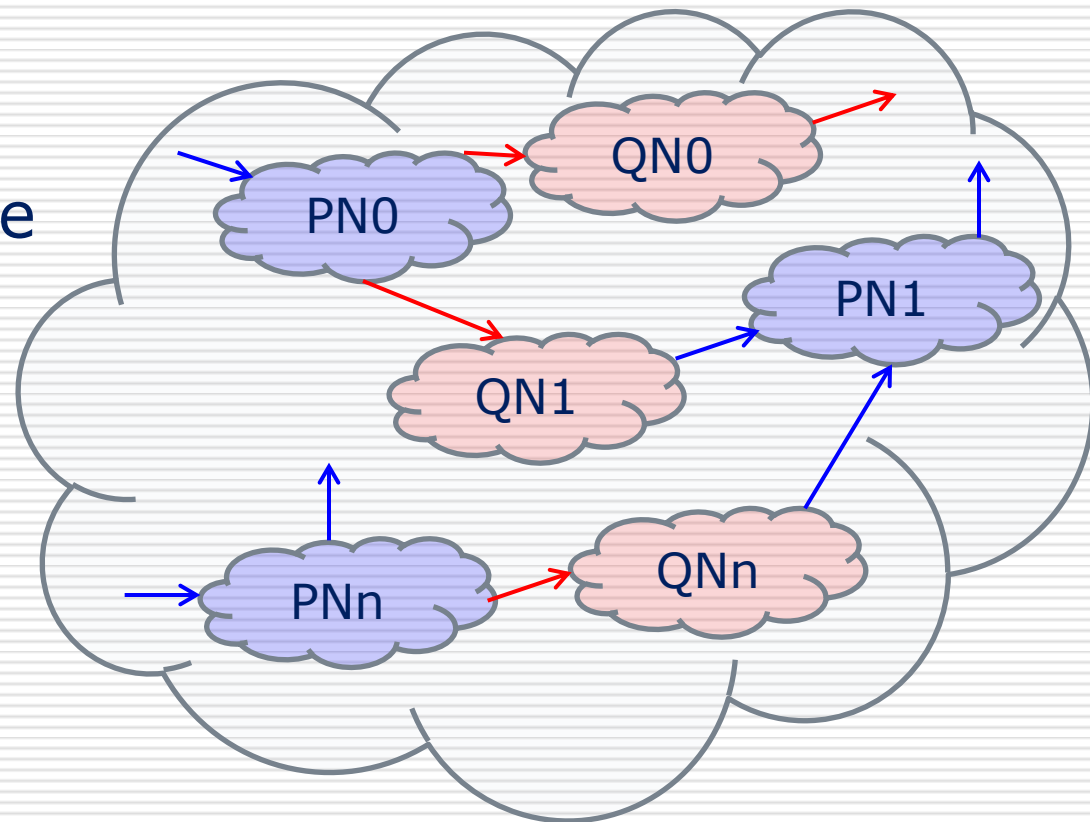


Building Block

Hybrid Petri Queuing Networks Simulator

Building Block (BB)

- ✓ Combined PN and QN sub-models
- ✓ Sub-models may inter-communicate
- ✓ Represents intra-communications

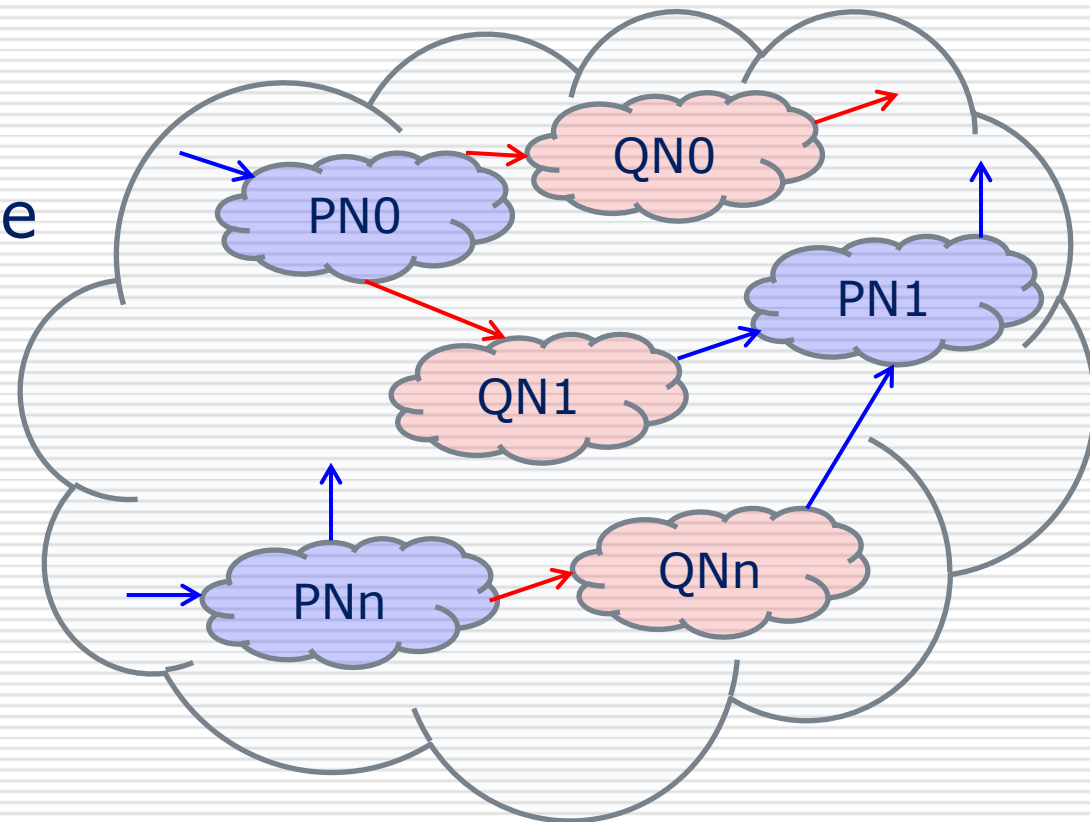


Building Block

Hybrid Petri Queuing Networks Simulator

Building Block (BB)

- ✓ Combined PN and QN sub-models
- ✓ Sub-models may inter-communicate
- ✓ Represents intra-communications
- ✓ Cross connections
 - ✓ $T \rightarrow Q$ or SF

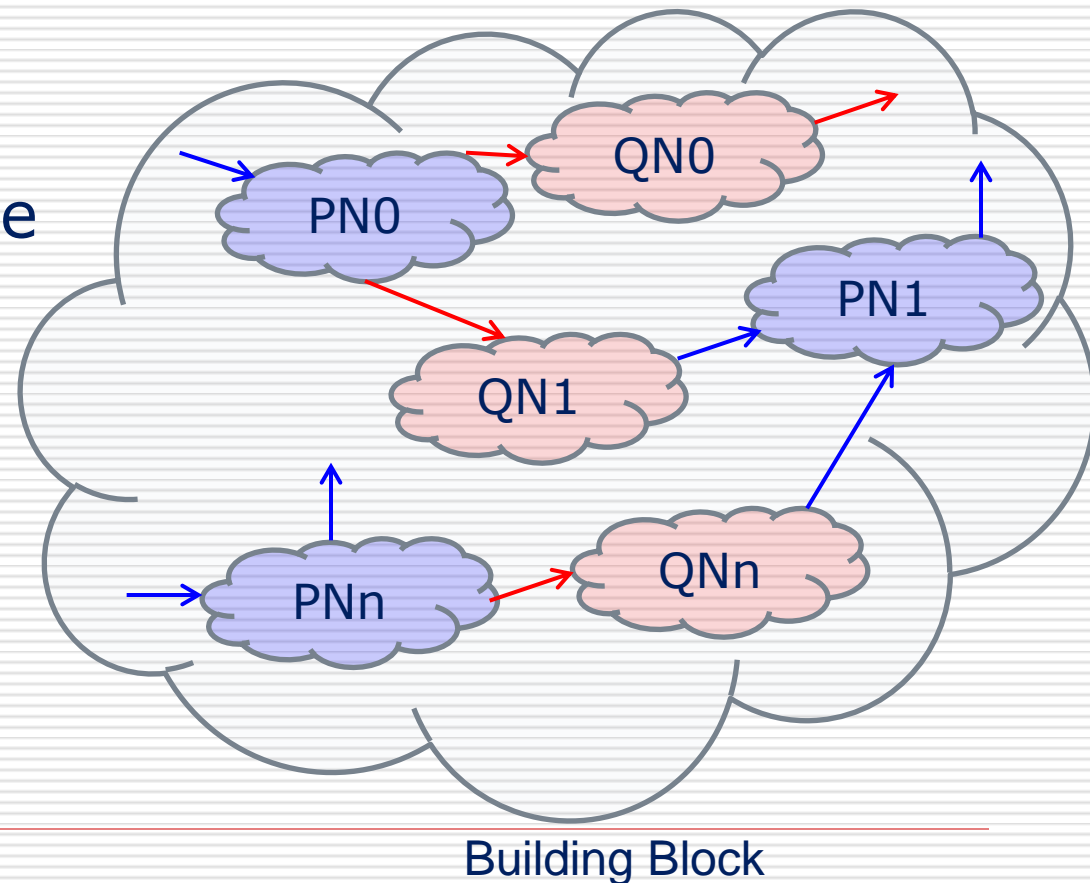


Building Block

Hybrid Petri Queuing Networks Simulator

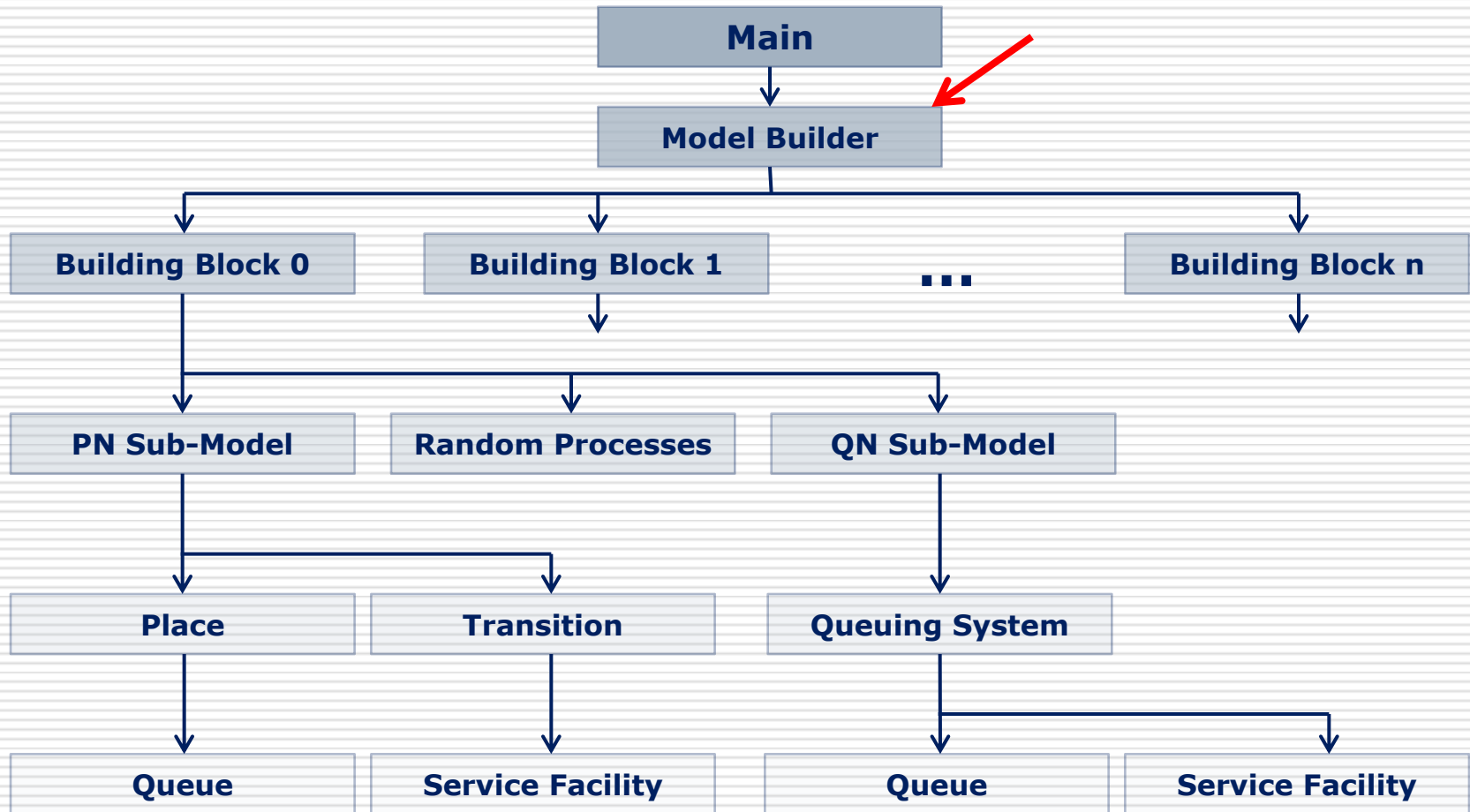
Building Block (BB)

- ✓ Combined PN and QN sub-models
- ✓ Sub-models may inter-communicate
- ✓ Represents intra-communications
- ✓ Cross connections
 - ✓ $T \rightarrow Q$ or SF
 - ✓ $SF \rightarrow P$



Hybrid Petri Queuing Networks Simulator

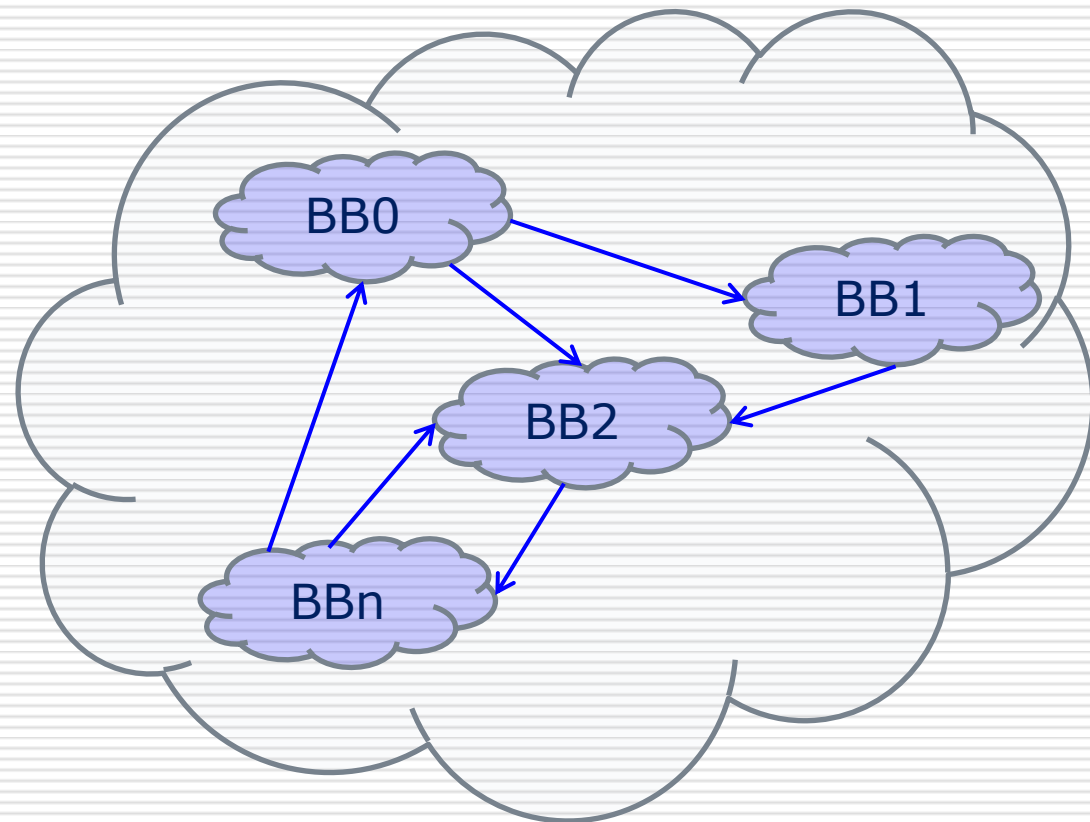
Class Hierarchy



Hybrid Petri Queuing Networks Simulator

Model Builder ^(MB)

- ✓ Inter BB-to-BB communications

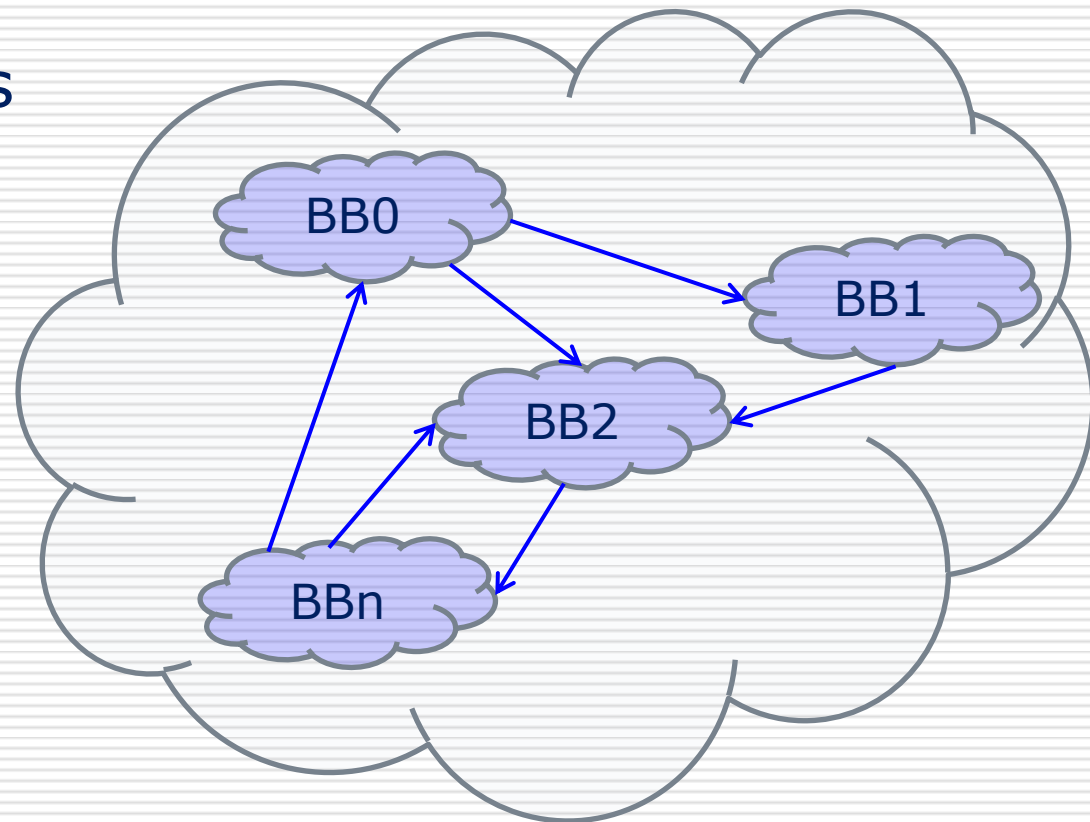


Model Builder

Hybrid Petri Queuing Networks Simulator

Model Builder ^(MB)

- ✓ Inter BB-to-BB communications
- ✓ Homogeneous BBs



Model Builder

Hybrid Petri Queuing Networks Simulator

Conclusion

- ✓ Researchers would be better off being equipped with a flexible, powerful and reliable framework rather than a rigid, finalized application solution
-

Hybrid Petri Queuing Networks Simulator

Future Developments

- ✓ Model Builder with Homo/Heterogeneous BBs
-

Hybrid Petri Queuing Networks Simulator

Future Developments

- ✓ Model Builder with Homo/Heterogeneous BBs
 - ✓ Extension of PDFs for Arrival/Service Times
-

Hybrid Petri Queuing Networks Simulator

Future Developments

- ✓ Model Builder with Homo/Heterogeneous BBs
 - ✓ Extension of PDFs for Arrival/Service Times
 - ✓ Extension of QDs/SSDs, if required
-

Hybrid Petri Queuing Networks Simulator

Future Developments

- ✓ Model Builder with Homo/Heterogeneous BBs
 - ✓ Extension of PDFs for Arrival/Service Times
 - ✓ Extension of QDs/SSDs, if required
 - ✓ Design and implementation of a GUI
-

Hybrid Petri Queuing Networks Simulator

Future Developments

- ✓ Model Builder with Homo/Heterogeneous BBs
 - ✓ Extension of PDFs for Arrival/Service Times
 - ✓ Extension of QDs/SSDs, if required
 - ✓ Design and implementation of a GUI
 - ✓ Code efficiency improvement
-

Analytical Modelling of Coordinated Multi-Point (CoMP) based handover for Next Generation Wireless Networks (NGWN)

R. R. Ahmed

School of Engineering and Informatics, University of Bradford,
Bradford, BD71DP, UK

Email: rrahmed@bradford.ac.uk

Abstract. A Coordinated Multi-Point (CoMP) based handover in Next Generation Wireless Networks (NGWN) influences the increase of the user throughput of the cell edge and minimisation of the inter-cell interference in order to increase the overall system throughput. Different deployment scenarios of CoMP in homogeneous and heterogeneous networks are considered to reduce the cell edge interference and increase the overall system capacity to perform seamless handover in Long Term Evolution - Advanced (LTE-A) 4G wireless networks. Furthermore, a CoMP based handover analytical model for LTE-A is proposed in order to optimize the system throughput and minimise the packet loss and system delay. In addition, the model may also be used to overcome the unnecessary handover and improve the energy consumption.

Keywords: Next Generation Wireless Networks, Coordinated Multi-Point, Long Term Evaluation-Advanced, CoMP based handover analytical model

Analytical Modelling of Coordinated Multi-Point (CoMP) based handover for Next Generation Wireless Networks (NGWN)

Presented by: R R Ahmed

**School of Engineering and Informatics,
University of Bradford**

Outline

- Introduction
- Aim/objectives
- Motivation
- Cell edge interference
- Enhanced inter-cell interference coordination
- Coordinated Multi-Point
- Coordinated methods
- CoMP element
- CoMP Handover (idea)
- Future Work
- Conclusion
- References

Introduction

- ITU defined standard for 4G called IMT-A
- Core milestones of IMT-A are:
 - Reduced cell edge interference
 - Reasonable power consumption for the mobile terminal
 - High Data rate (1Gb/s: UL/DL)
- 3GPP and WiMAX are participants
- 3GPP achieved maximum target of 4G
 - Long Term Evaluation- Advanced (4G)

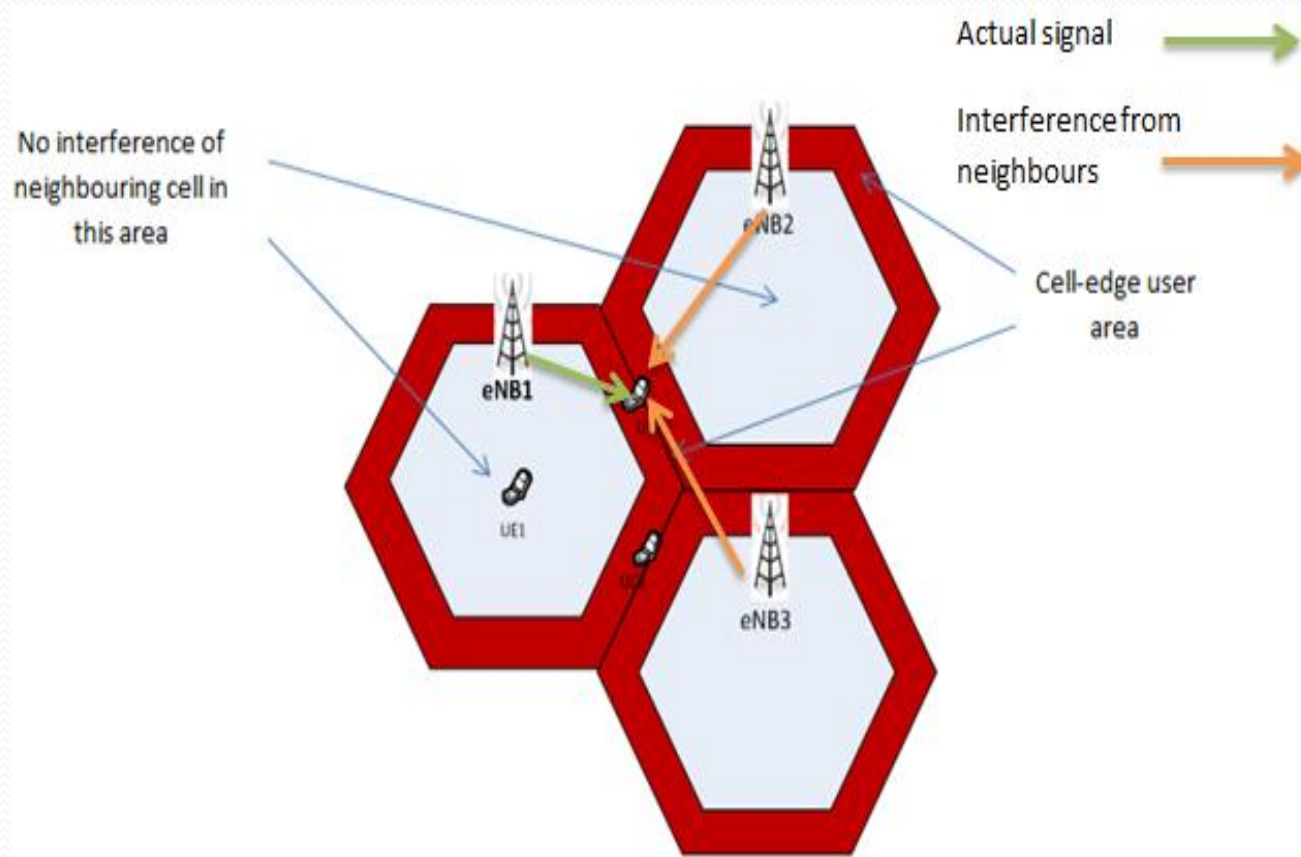
Aims/Objectives

- Research and investigate the effectiveness of coordinated multipoint technique for next generation wireless network.
- Research and evaluate existing handover techniques used in 4G networks.
- To design analytical model of CoMP for next generation wireless networks to
 - Optimize the system throughput
 - Minimise the packet loss and system delay
 - Reduce unnecessary handovers .

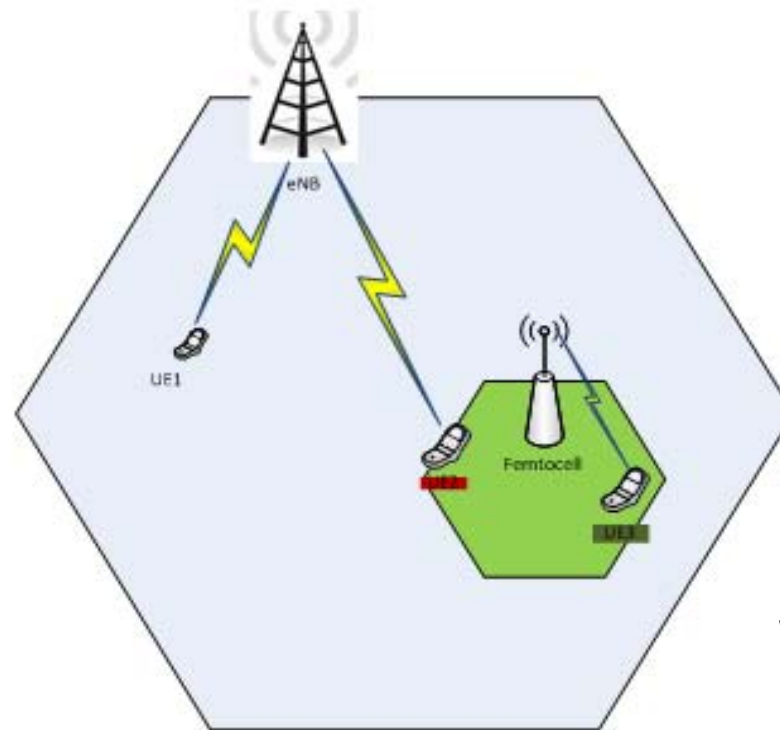
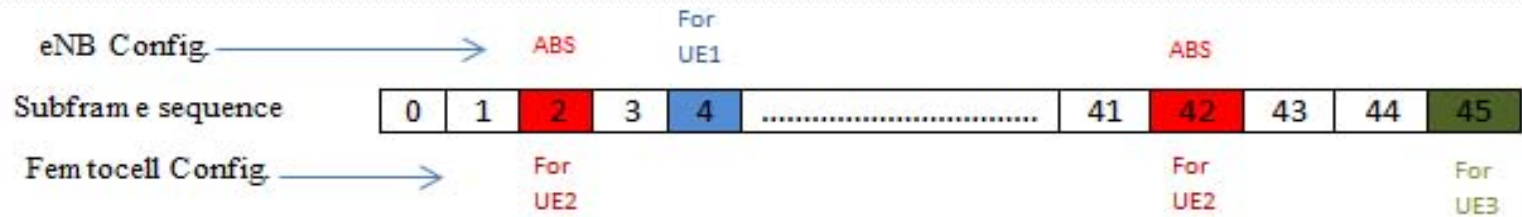
Motivation

- ▶ Cell edge interference
- ▶ Reasons
 - Signal strength is weak because of the UE distance from the base station (eNB)
 - UE is closer to neighbouring eNBs that's why Interference levels is higher.
- ▶ CoMP
 - UE at the edge of a cell is able to be served by two or more eNBs to improve signals reception / transmission and increase throughput particularly under cell edge conditions
 - Communication is made with an exchange of control among several coordinated entities
 - Scalable Radio frequency (OFDMA)

Cell Edge Interference



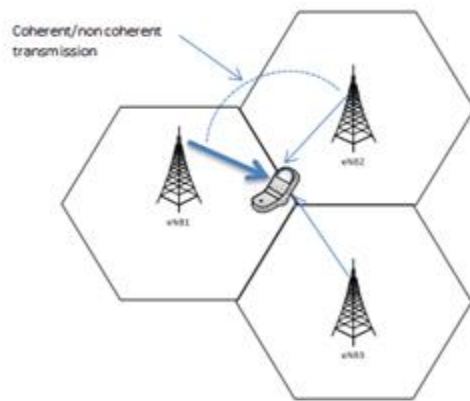
Enhanced inter cell interference coordination



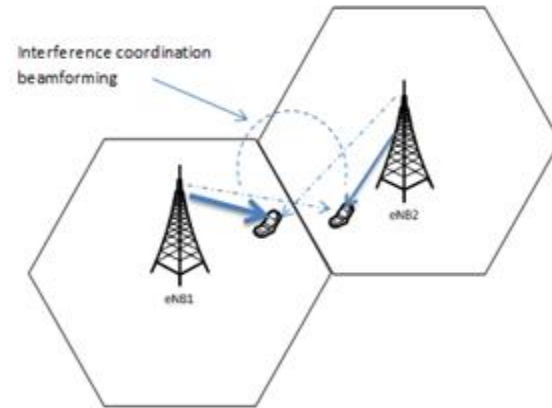
Coordinated Multipoint

- Consists on different transmission points
- Perform coordinated methods such that
 - Joint transmission
 - Joint processing
 - Coordinated scheduling

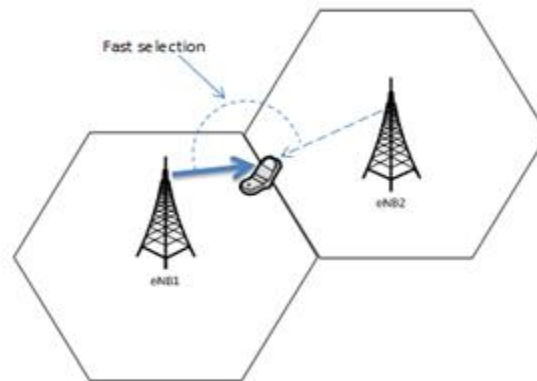
Coordinated methods



(a)



(b)



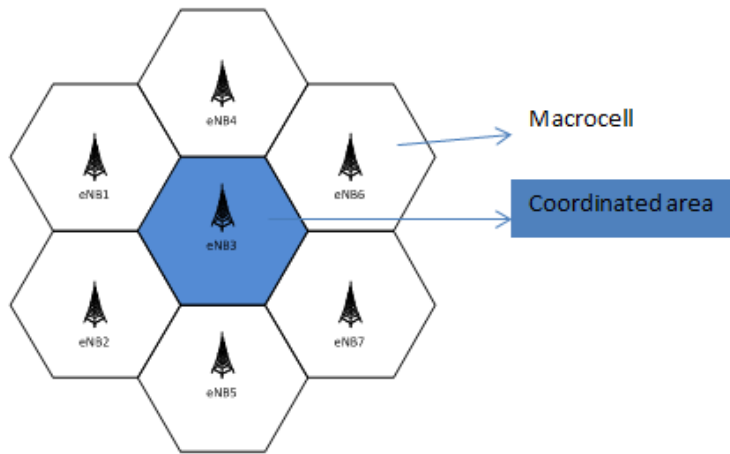
(c)

WP06-9

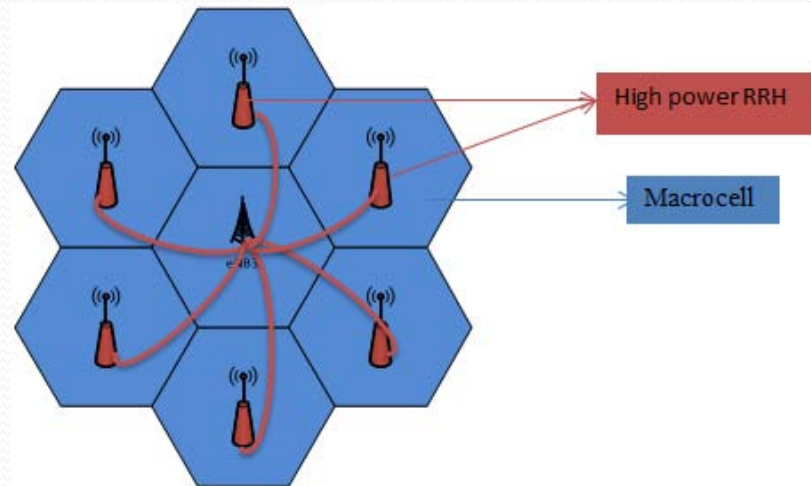
(a) Joint transmission (b) CS/CB (c) dynamic cell selection

CoMP Deployments Scenarios

- Different deployments scenarios:
 - homogeneous macro network with intra-site CoMP
 - homogeneous macro network with inter-site CoMP



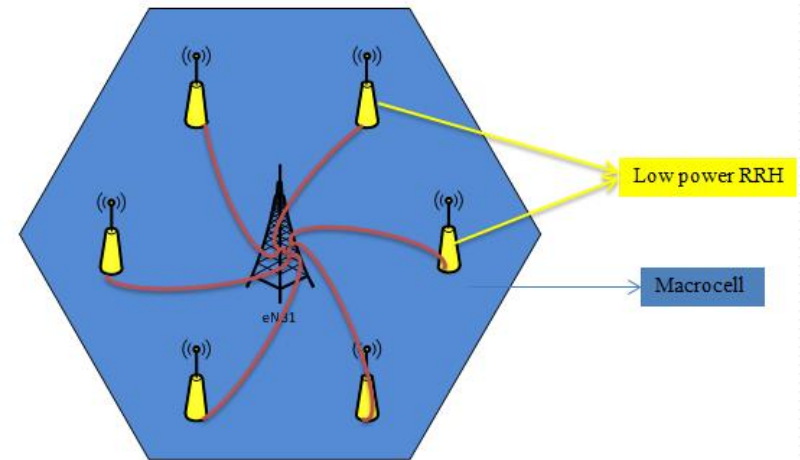
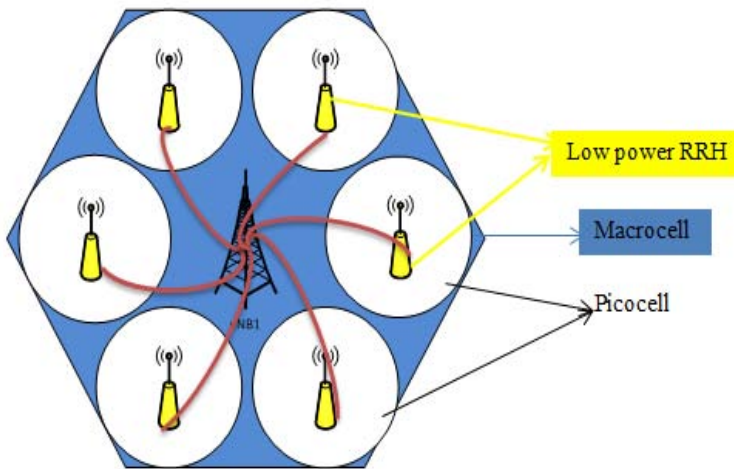
Homogeneous macro network with intra-site CoMP



Homogeneous macro network with inter-site CoMP

CoMP Deployments Scenarios (cont..)

- heterogeneous network with low power remote heads (picocells) within the macrocell coverage area
- heterogeneous with lower power remote radio heads (RRH) within the macrocell coverage area



Heterogeneous network with low power remote heads (picocells)

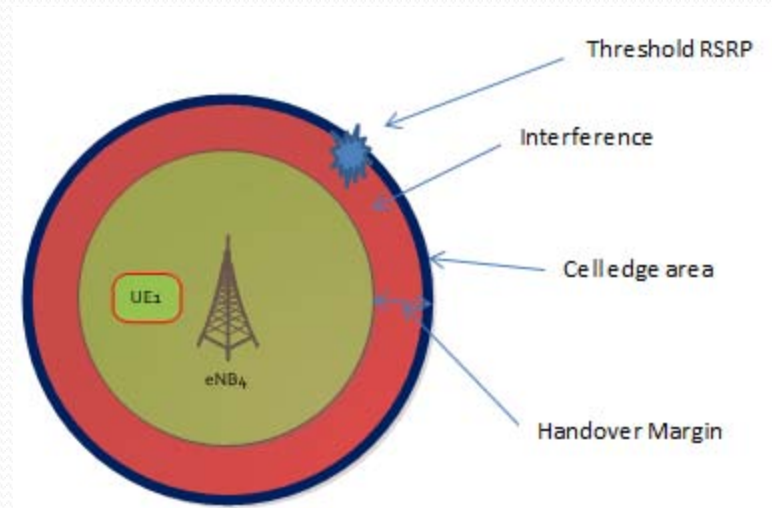
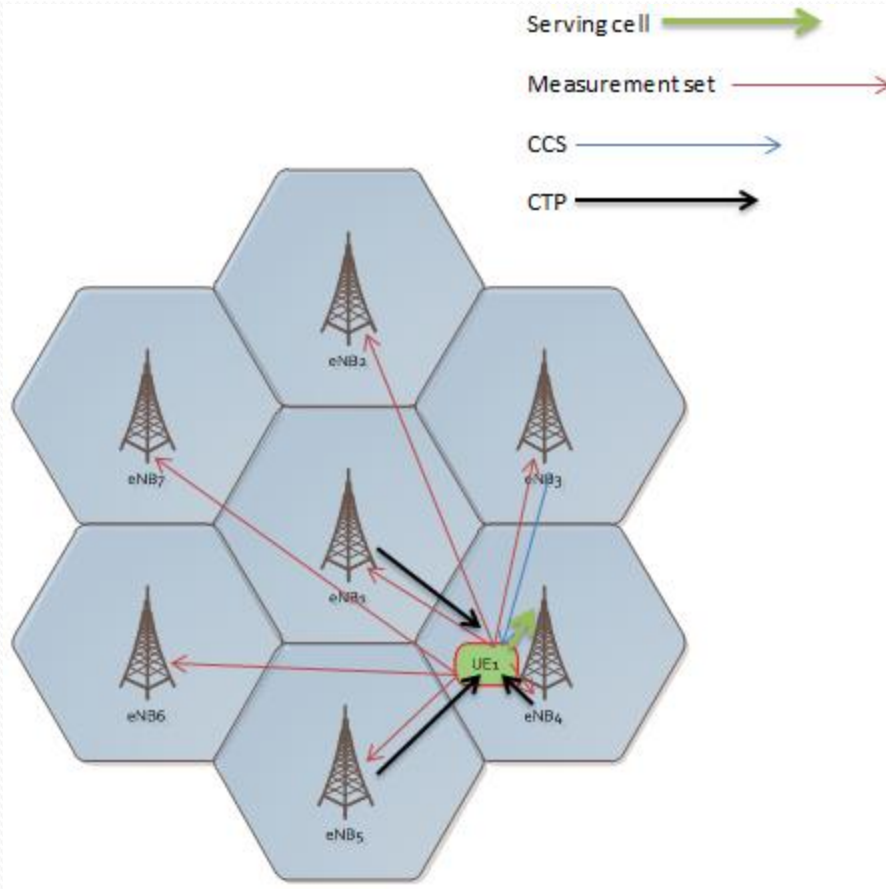
Heterogeneous with lower power remote radio heads (RRH)

CoMP elements

- Serving cell
 - Cell responsible of PDCCH(physical downlink control channel) to UE
 - UE has one serving cell
- Measurement set
 - UE perform periodical channel state information(CSI)
- CoMP transmission points (CTP)
 - Cell directly sends PDSCH to UE
 - CTP is selected from CCS

- CCS (CoMP cooperating set)
 - Directly or indirectly sends PDSCH to UE.
 - CCS is selected from the measurement cell set.

CoMP Handover (Design)



Future Work

- At first stage develop M/M/1 model with:
 - Arrival rate of ' λ '
 - Service time ' μ '
- Model has two mode
 - CoMP mode (Received Signal Strength is low)
 - None CoMP mode (Received Signal Strength is high)
- Calculate utilization of resource blocks
- If Utilization of RB higher
 - No guarantee of Successful handover
- Calculate the Bit error rate
 - For calculating the throughput of the system
 - Cell edge throughput
- Absolute blank Frame (ABF) to calculate the energy efficiency
- At advance level the model will extend to GE/GE/1 model

Conclusion

- CoMP is constructive solution of many issues
- Reduced energy consumption at eNB level.
- Increased the cell edge user throughput
- CoMP is reducing the no of unnecessary handover
- Reduce extra burden on the network

References

- L. Juho, K. Younsun, L. Hyojin, L. Boon, D. Mazzaresse, L. Jianghua, X Weimin, Z. Yongxing, "Coordinated multipoint transmission and reception in LTE-advanced systems" *Communications Magazine, IEEE* , Volume:50 , Issue: 11, pp. 44-50, Nov 2012
- Enhanced ICIC for LTE-A HetNet [Online] Available at: http://www.zte.com.cn/cn/events/lteworld2012/solutions/201205/t20120516_364207.html [Retrieved on 10-06-2013]
- I. Ahmad, and D. Habibi, "Resource Management in 4G Wireless Communications at Vehicular Speeds: A Game Theory Solution" *IEEE Vehicular Technology Conference (VTC Fall)*, pp.1-5, September 2012
- 3GPP TR 36.814 V9.0.0, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects" (Release 9), 2010-03



THANK YOU

Lightweight Clustering of Cell IDs into Meaningful Neighbourhoods

Marios Fanourakis, Katarzyna Wac

University of Geneva, Center for Informatics, Geneva, Switzerland

Abstract. Location-based services are becoming a norm in today's mobile phones, leveraging sensor data such as GPS, inertial sensors, and magnetic field sensors, or utilize up-to-date databases of RF beacons' locations such as WiFi access point (AP) and cellular towers (cell ID). With the exception of cell ID-based localization, where the cellular antenna is enabled for the core functionality of the phone, the other methods require the use of extra resources, increasing the processing and power consumption of the mobile phone. This paper presents a lightweight method to form semantically meaningful clusters of cell IDs from cell ID sequences, via analysis of 15 weeks of data collected from a real user.

1 Introduction

Modern mobile devices are equipped with a number of sensors that have been adapted to provide location information. GPS sensors were first included in mobile devices for the purpose of providing a position to a navigation system. As the importance of location awareness rose, so did the effectiveness of the methods and algorithms to provide the necessary location information. Location awareness has become an essential feature of mobile devices, also used for advertising or social networking, among other services. Many of these services do not always require fine location readings such as GPS, and can utilize a number of other sensors such as WiFi AP and cell ID, that are more energy efficient yet provide all the necessary information. An example is the 'Google Now' service (Android OS), that is able to recognize semantically meaningful locations and track the patterns of a user between those locations to provide traffic and weather information proactively. Google leverages their own databases of cell ID locations and WiFi AP locations to infer the position of a given user when GPS is not available. In our research we focus on the data collected from the user himself and analyze his location context in terms of neighborhoods that can be extracted in a lightweight way, just by analyzing the cell IDs he connects to.

1.1 Cell ID Sequence Mining

We can look at cell IDs as both nodes in a graph and states in a state sequence. A traditional connection between two states is formed when the probability of transition from one state to the other is nonzero. However, for the cell ID sequences there can be two major types of transitions occurring: transitions due to

user mobility from one place to another and due to *cell ID oscillation*. The first one occurs when a user moves out of the range of one cell ID, and into another. Cell oscillation occurs even though a user may be staying at a particular location; the connection to a cell ID cycles through a set of cell IDs that are within range, depending on the number of other users in range, network traffic, etc. A common approach is to use user's GPS sensor readings or databases of GPS locations of the cell IDs to simplify the clustering approach [6] [10] [12] [14]. Shad [13] used labeled data to overcome this issue by creating cell ID neighborhoods where all the cell IDs in each neighborhood have similar labels. Meneses[11] used a fingerprinting technique which required the offline step of staying at particular locations for a certain amount of time to retrieve statistics about the cell IDs in range and their RSSIs. Similarly, CellSense[4] leveraging the fingerprinting executed by the user, creates models for several spots, and cluster these into significant locations using certain distance measures. There are very few works that use exclusively the cell ID data without any enhancements. Among them, Laasonen[8] [9] used a graph-theoretic approach to analyze the graph connection characteristics between cell IDs and infer neighborhoods.

In our approach, we also only leverage the cell ID sequences as 'naturally' collected from the user and will assume that a person does not, in general, move from one semantically significant location to another and then back within a certain (short) time T , thereby, differentiating between cell oscillations and actual movement from one place to another.

1.2 Related Work

RF Beacon to Location. PlaceLab[10][14] is a commonly used database that maps RF beacons to a geographical location. CellSense[4][5] is a fingerprint based technique to map cell IDs into locations in two phases: an offline phase where data is collected at certain positions such as cell IDs in range and their signal strength (RSSI), and the localization phase where a reading of neighboring cell IDs and their RSSI is mapped to an estimated position. Similar fingerprinting techniques are not uncommon.

Significant Places and Daily Life Patterns. Very common application of mobile positioning is in the field of discovery of meaningful places and analyzing mobility patterns of users. There are many works in those fields since the advent of location based services. Some methods can be seen in, e.g., [1][3][6][7][16][17], leveraging many sensors in the phone or data from external databases.

2 Cell ID Clustering into Neighbourhoods: Dataset and Method

2.1 mQoL Living Lab Dataset

Much of the research performed in this field uses the Reality Mining[2] dataset; a relatively old dataset with a sampling rate of $0.2\text{samples/minutes}$ (5 minutes

between samples). In our research we leverage our own dataset, the University of Geneva mQoL Living Lab[15] dataset involving 38 users for the last 1.5 years, and having an average sampling rate of $1\text{samples}/\text{minute}$ (depending on the connectivity of the user or his mobility). Only a randomly selected subset was used in this research - approximately 15 weeks of data for one random mQoL user (Feb-Apr 2013). The data set had over 700 different cell IDs, with a very little data between the hours of midnight and 7am (user switched OFF his phone). Furthermore, a moving window of 3 weeks of data was used to analyze the cell ID sequences. Since the data is unlabeled, the evaluation method of our method was mainly by visual inspection and comparison, pending the implementation of an automated analytic method.

2.2 Cell ID Clustering from Detected Cell Oscillation Events

Algorithm 1 is used to detect a cell ID oscillation and update a *connection matrix*, which describes the vertices of a graph where connections form only between a given cell ID (denoted a *base cell ID*) and the cell IDs contained in its oscillation event. The algorithm increments the count of connections in a connection matrix whenever there is a round trip of a base cell ID within a time period T . For example, let a sequence be $ABCDDDB$. If the time between the first B to the second B is less than T , then we say that B is the base cell ID in an oscillation event and the connection counts between $B-C$ and $B-D$ would be incremented. The matrix is forced to be symmetrical by also incrementing the count of $C-B$ and $D-B$, i.e., the connection is directionless.

Algorithm 1: Cell ID oscillation event detection

Data: *subsequence*: $current - T$ up to $current$ measurement, connection matrix: *conMat*.

Initialize: $i \leftarrow 1$

```

while  $i < \text{length}(\text{subsequence})$  do
     $base \leftarrow i$ 
     $j \leftarrow i + 1$ 
    while no match found and  $j \leq \text{length}(\text{subsequence})$  do
        if  $\text{subsequence}(base) = \text{subsequence}(j)$  then
            match found
             $end \leftarrow j$ 
             $j \leftarrow j + 1$ 
        if match found and  $end - base > 1$  then
             $k \leftarrow base + 1$ 
            while  $k < end$  do
                 $conMat(base, k) \leftarrow conMat(base, k) + 1$ 
                 $conMat(k, base) \leftarrow conMat(k, base) + 1$ 
                 $k \leftarrow k + 1$ 
             $i \leftarrow i + 1$ 

```

Determining cell ID Neighborhoods Based on the Connection Matrix.

From the connection matrix we can form cell ID neighborhoods by looking at the *cliques* formed in the graph that this connection matrix represents. In figure 1 we see an example of two cliques in a connection matrix.

1	1	0	0
1	1	1	1
0	1	1	1
0	1	1	1

Fig. 1: Example of Cliques

Merging of Cell ID Neighborhoods. Since a cell ID can belong to multiple neighborhoods that can be similar, the neighborhoods do not completely solve the cell ID oscillation issue. For that reason we merge neighborhoods involved in what now is a *neighborhood oscillation effect*; detected in the same way as the cell ID oscillation, for a different T parameter. The main difference is that instead of updating a connection matrix, we merge the neighborhoods involved in the detected neighborhood oscillation event and relabel the cell ID sequence. As a further step, we apply this algorithm recursively in the regions of the cell ID sequence where the neighborhoods are redefined due to the merge. We define those regions to be $\pm T$ from each affected cell ID neighborhood label.

2.3 Assigning Neighborhoods to the Cell IDs in the Sequence

Since a cell ID can belong to multiple neighborhoods, assigning a neighborhood to a cell ID in the sequence is not trivial. First we find all the neighborhoods it belongs in and compare each of them with a list of recently seen cell IDs (the number of cell IDs is the same as the size of the neighborhood). The largest neighborhood which includes the most of the recently seen cells is then assigned to that cell ID.

3 Results and Discussion

For the following results the T for cell and neighborhood oscillation were set to 3.01 and 2.01 minutes respectively. The 3 week window of the data shown in the figures had 188 different cell IDs, which reduced to 80 neighbourhoods after the clustering.

Figure 2 presents a selected 6 days of user data in the sequence: raw cell ID sequence (left) and their clustering into neighbourhoods (right). As we can see, the neighbourhoods present a much clearer view on the user mobility, and, based on these results in context of time, one can reason about the meaningful neighbourhoods for this user like house, office, other.

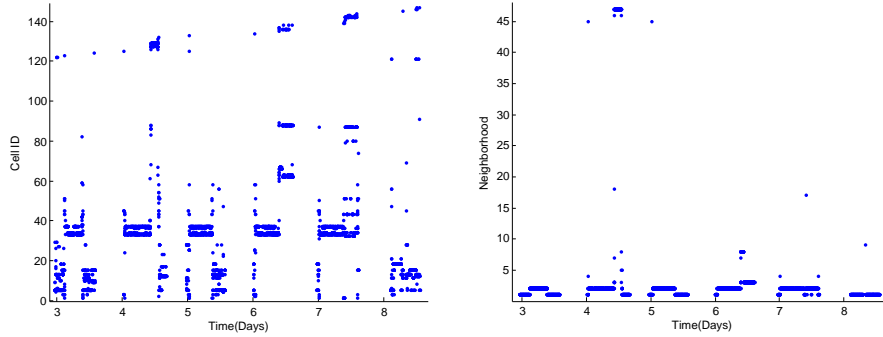


Fig. 2: Cell ID sequence (left: raw cell IDs, right: neighbourhoods)

Figure 3 presents the average user staying times in a cell ID (left) and corresponding neighbourhoods (right) for 3 weeks of data. The neighbourhood-based view may enable a more accurate reasoning upon which neighborhoods are more significant to the user based on the uninterrupted time he spends there, thus enabling potentially to leverage location-aware features and services. The longest time he spends continuously in a particular cell is less than 2 hours; very low for the most significant place. However, the longest time he spends continuously in a particular neighbourhood is a little over 8 hours, which corresponds nicely with work. Second longest is approximately 6 hours which corresponds to the time he spends at home after work before turning off the phone for the night.

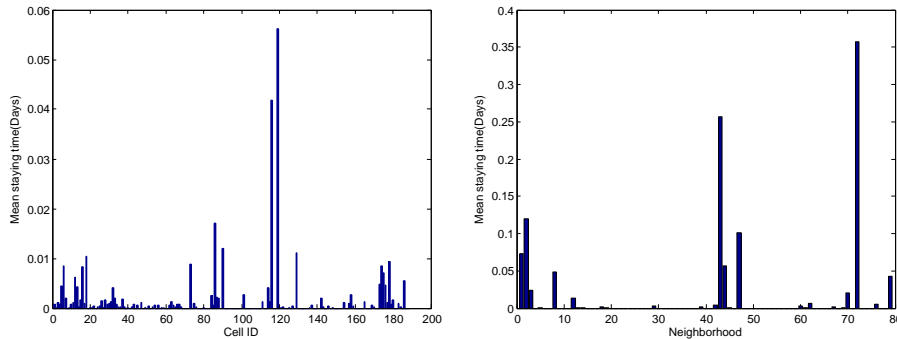


Fig. 3: Average staying times (left: raw cell IDs, right: neighbourhoods)

Our research results are promising as a step towards a lightweight method, running on a mobile phone itself and enabling the accurate clustering of cell IDs into meaningful for the user neighbourhoods. Yet, we are aware of potential limitations of our method, developed for a single user, and evaluated for a limited dataset, without labels for a ground truth about the neighbourhoods the user visited.

4 Conclusive Remarks and Future Work

We have illustrated a lightweight method for the determination of neighbourhoods from raw cell ID data collected on a phone. The method was tested on

the 15 weeks of data collected from a real user in his natural daily environments. In the future we'd like to develop a robust method for evaluating this algorithm using labeled data or through other statistical analysis, fine tune the parameters to form optimal neighborhoods, and verify this algorithm against all the other users in the mQoL Living Lab dataset. Furthermore, once the algorithm becomes accurate enough, we plan to deploy it directly on the mQoL Living Lab participants' phones towards providing them location-based services.

Acknowledgments This work was partially supported by EU AAL projects WayFiS, MyGuardian, and CaMeLi.

References

1. Bhattacharya, S.: Place Identification: A Comparative Study. Ph.D. thesis, University of Helsinki, Finland (2009)
2. Eagle, N., Pentland, A.: Reality mining: sensing complex social systems. *Personal and ubiquitous computing* 10(4), 255–268 (2006)
3. Farrahi, K., Gatica-Perez, D.: Learning and predicting multimodal daily life patterns from cell phones. In: *ACM MMI*. pp. 277–280 (2009)
4. Ibrahim, M., Youssef, M.: Cellsense: a probabilistic rssi-based gsm positioning system. In: *IEEE GLOBECOM*. pp. 1–5 (2010)
5. Ibrahim, M., Youssef, M.: Cellsense: An accurate energy-efficient gsm positioning system. *IEEE TVT* 61(1), 286–296 (2012)
6. Kim, D.H., Hightower, J., Govindan, R., Estrin, D.: Discovering semantically meaningful places from pervasive rf-beacons. In: *UbiComp*. pp. 21–30. ACM (2009)
7. Kim, D.H., Kim, Y., Estrin, D., Srivastava, M.B.: Sensloc: sensing everyday places and paths using less energy. In: *ACM CENSS*. pp. 43–56 (2010)
8. Laasonen, K., Raento, M., Toivonen, H.: Adaptive on-device location recognition. In: *Pervasive Computing*, pp. 287–304. Springer (2004)
9. Laasonen, K., et al.: Mining cell transition data. Ph.D. thesis, University of Helsinki, Finland (2009)
10. LaMarca, A., Chawathe, Y., Consolvo, S., et al.: Place lab: Device positioning using radio beacons in the wild. In: *Pervasive Computing*, pp. 116–133. Springer (2005)
11. Meneses, F., Moreira, A.: Using gsm cellid positioning for place discovering. In: *IEEE Pervasive Health Conference*. pp. 1–8 (2006)
12. Nicholson, A.J., Noble, B.D.: Breadcrumbs: forecasting mobile connectivity. In: *ACM MobiCom*. pp. 46–57 (2008)
13. Shad, S.A., Chen, E., Bao, T.: Cell oscillation resolution in mobility profile building. *International Journal of Computer Science* 9(3), 205–213 (2012)
14. Sohn, T., Griswold, W.G., Scott, J., et al.: Place laban open architecture for location-based computing. In: *ESEC/FSE* (2005)
15. Wac, K.: mQoL living lab: Experience of running longitudinal studies for mobile internet measurements. In: *IMC* (2013)
16. Zhou, C., Frankowski, D., Ludford, P., Shekhar, S., Terveen, L.: Discovering personally meaningful places: An interactive clustering approach. *ACM TOIS* 25(3), 12 (2007)
17. Zhou, C., Ludford, P., Frankowski, D., Terveen, L.: An experiment in discovering personally meaningful places from location data. In: *ACM CHI*. pp. 2029–2032 (2005)



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs FDEA
Federal Office for Professional Education and Technology OPET



UNIVERSITÉ
DE GENÈVE

CENTRE UNIVERSITAIRE
D'INFORMATIQUE



Lightweight Clustering of Cell IDs into Meaningful Neighbourhoods

Marios Fanourakis
University of Geneva, Switzerland
Marios.Fanourakis@unige.ch

Motivation

▶ Location Aware Services

- ▶ Proactive retrieval of information
- ▶ Proactive display of information
- ▶ Advertising
- ▶ Tracking user habits to improve a service
- ▶ E.g. Google Now

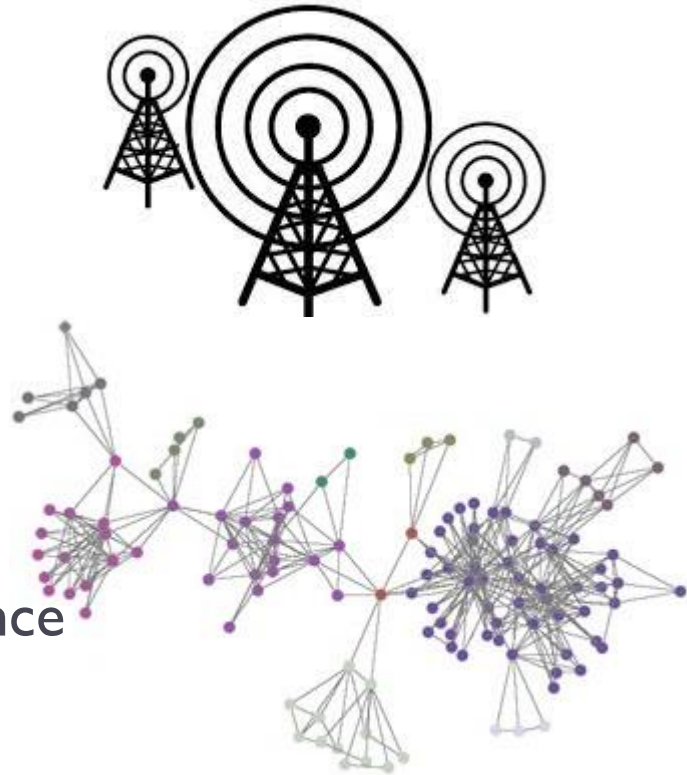
▶ Independent Device

- ▶ Can work without data connection
- ▶ Can collect data
- ▶ Analyze this data
- ▶ No need for databases
- ▶ Works out of the box
- ▶ Energy efficient



Related Work

- ▶ **PlaceLab [10][14]**
 - ▶ Database of APs and their GPS coordinates
 - ▶ Widely used
- ▶ **Fingerprinting (CellSense [4][5])**
 - ▶ Maps cell IDs to location
 - ▶ Uses a fingerprinting technique
 - ▶ Requires prior data collection
- ▶ **Cell ID Mining (Laasonen [8][9])**
 - ▶ Uses Reality Mining database
 - ▶ Only uses cell IDs
 - ▶ Creates graphs from the cell ID sequence
 - ▶ Analyzes the graphs to get clusters



Related Work cont.

- ▶ **Implementations**
 - ▶ Meaningful place discovery
 - ▶ Mobility pattern analysis
 - ▶ Location based services
 - ▶ And more...

See: [1][3][6][7][16][17]

Our Work

Related Works

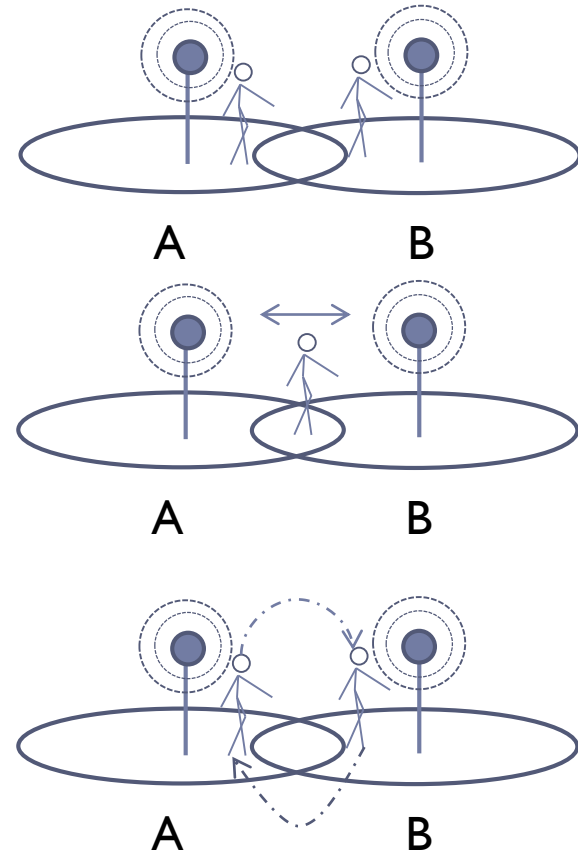
- ▶ PlaceLab, Reality Mining [2]
- ▶ Require either updated databases or some setup
- ▶ Use: GPS, Location

Our Work

- ▶ mQoL Database [15]
- ▶ Works without any prior setup
- ▶ Only uses cell IDs
- ▶ Online
- ▶ Simple
- ▶ Energy efficient

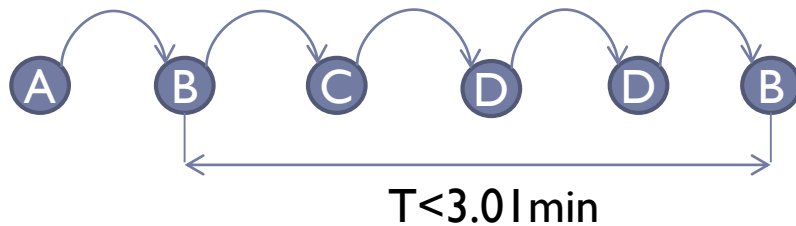
Cell ID Transitions

- ▶ **Transitions due to mobility**
 - ▶ User moves out of range of one and into the range of another
- ▶ **Transitions due to cell ID oscillation**
 - ▶ Signal strength changes
 - ▶ Network traffic
- ▶ **Our assumption**
 - ▶ User does not move from one significant location to another and then back within a certain (small) time T



Cell Oscillation

▶ Detection



- ▶ If the same cell ID is measured within a time T , claim that there has been an oscillation event
- ▶ Update the adjacency matrix at the vertices (B,C), (C,B), (B,D),(D,B)

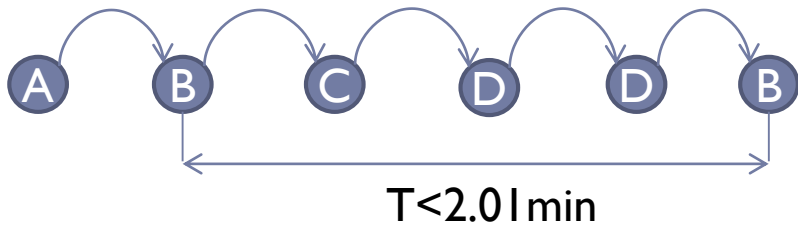
▶ Adjacency Matrix

1	1	0	0
1	1	1	1
0	1	1	1
0	1	1	1

- ▶ Updated after each oscillation event
- ▶ Determine cliques formed

Neighborhood Oscillation

▶ Detection



- ▶ If the same neighbourhood is measured within a time T , claim that there has been an oscillation event

▶ Merging

- ▶ Merge neighbourhoods

1	1	0	0
1	1	1	1
0	1	1	1
0	1	1	1

1	1	0	0
1	1	1	1
0	1	1	1
0	1	1	1

Labeling

- ▶ A cell ID can belong in two (or more) neighborhoods
 - ▶ Which of them to choose?

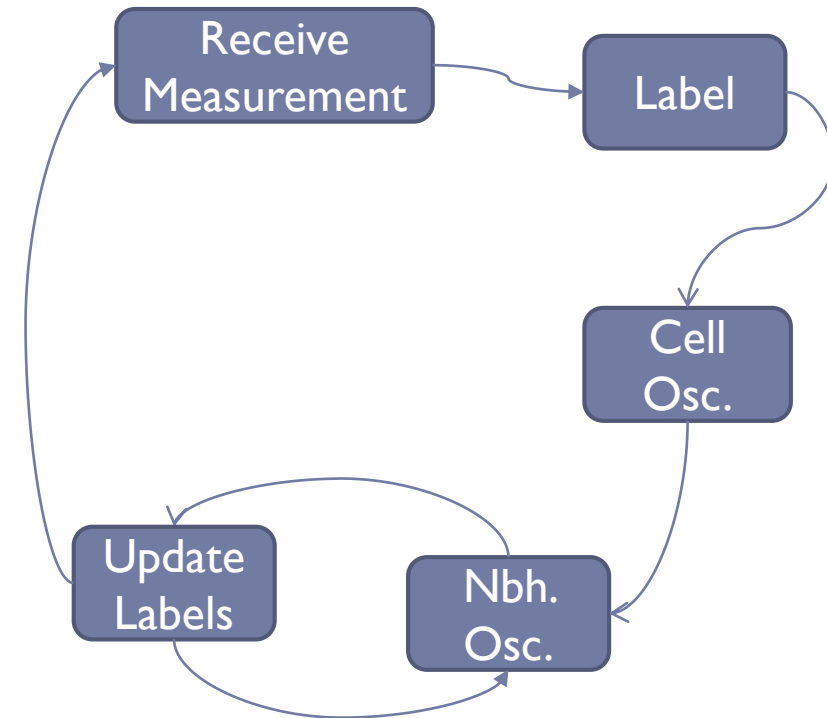
1	1	0	0
1	1	1	1
0	1	1	1
0	1	1	1

The diagram shows a 4x4 grid of cells. The top-left 2x2 area is enclosed in a thick black border, with a question mark in the top-left cell. The bottom-right 2x2 area is also enclosed in a thick black border, with a question mark in the top-right cell. The central cell (row 2, column 2) is the intersection of these two neighborhoods.

- ▶ The one that is most similar to the recent measurements
 - ▶ Compare the cell IDs of each neighborhood to an appropriate set of recent measurements
 - ▶ Choose the largest and most similar neighborhood

Main Steps

- ▶ **Cell oscillation**
 - ▶ Detect oscillation
 - ▶ Update special adjacency matrix
 - ▶ Find Cliques
- ▶ **Neighborhood oscillation**
 - ▶ Detect oscillation
 - ▶ Merge neighborhoods
- ▶ **Labeling**



Results – Dataset: mQoL Living Lab

- ▶ **TCP-RTT value every minute**
 - ▶ Cell ID, RSSI, Cell neighbors (or if WiFi), access network tech.
 - ▶ Apps running (foreground, background)
 - ▶ Screen ON/OFF, brightness, orientation and screen touches, light
 - ▶ Bytes send/received
 - ▶ Calls, SMSes
- ▶ **1.5 years, 30-40 participants (now 34)**

Results – Data Used

▶ Data:

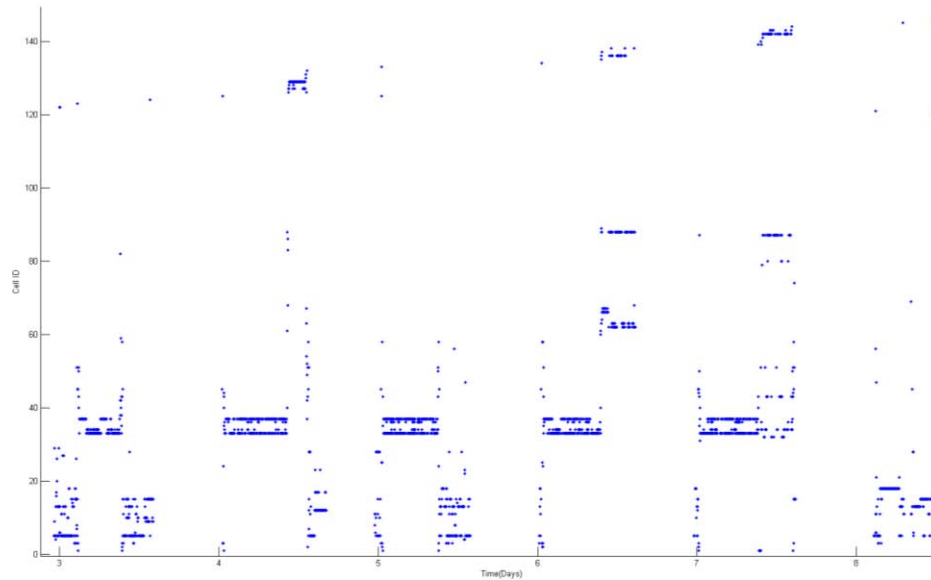
- ▶ From a single user
- ▶ From February 2013 to April 2013
- ▶ >700 different cell IDs
- ▶ Little data between 12am and 7am (phone switched off)
- ▶ Unlabeled w.r.t. neighbourhoods
- ▶ Used only cell ID measurement and its timestamp

▶ Analysis Window:

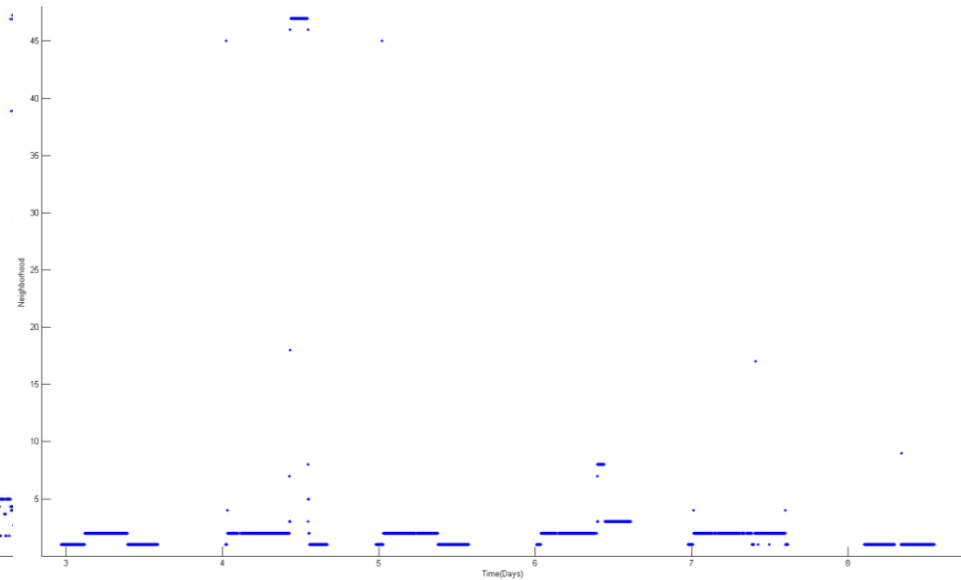
- ▶ 3 week sliding window

Results – Cell ID Sequence

Original

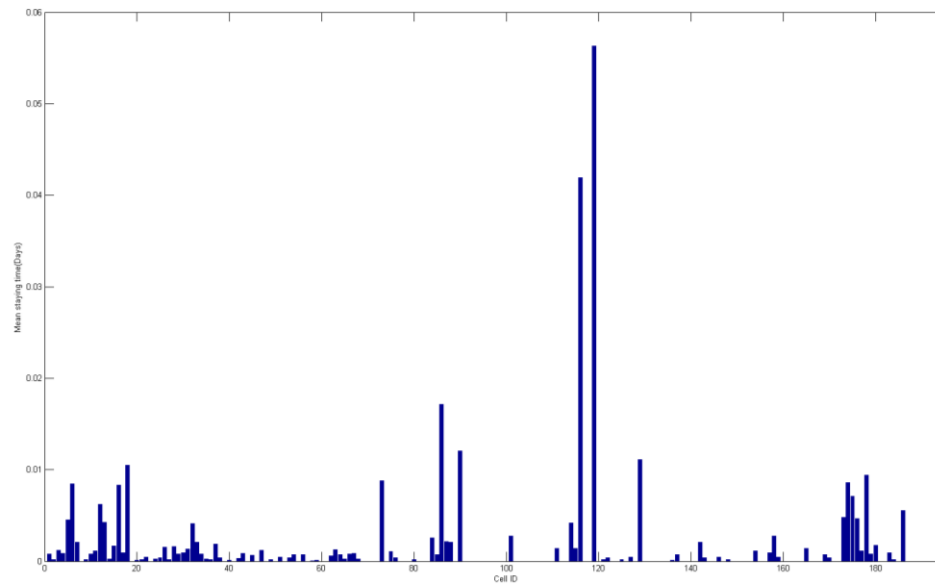


Clustered

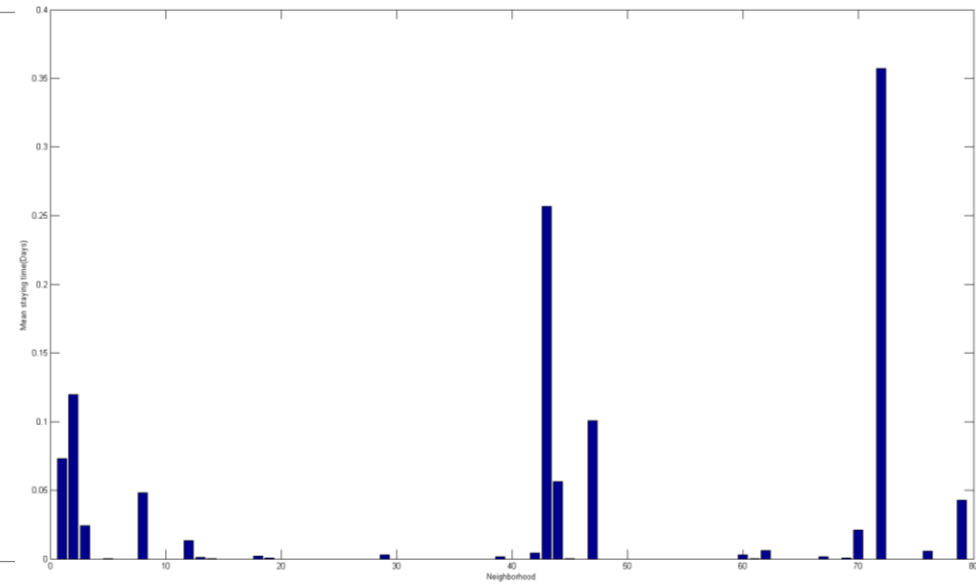


Results – Average Staying Times

Original



Clustered



Conclusion and Future Work

▶ Conclusion

- ▶ Semantically meaningful clusters of cell IDs can be extracted from a sequence
- ▶ No need for databases
- ▶ No prior setup
- ▶ Very simple

▶ Future Work

- ▶ Fine-tune the parameters (adaptive parameters?)
- ▶ Verify with more of our data
- ▶ Evaluate using labeled data or other analytics
- ▶ Build mobility models
- ▶ Ultimate goal: Predict personal QoS expectations



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs FDEA
Federal Office for Professional Education and Technology OPET



**UNIVERSITÉ
DE GENÈVE**

CENTRE UNIVERSITAIRE
D'INFORMATIQUE



Thank you!

Marios.Fanourakis@unige.ch

www.qol.unige.ch

Questions?

References

- ▶ 1. Bhattacharya, S.: Place Identification: A Comparative Study. Ph.D. thesis, University of Helsinki, Finland (2009)
- ▶ 2. Eagle, N., Pentland, A.: Reality mining: sensing complex social systems. *Personal and ubiquitous computing* 10(4), 255{268 (2006)
- ▶ 3. Farrahi, K., Gatica-Perez, D.: Learning and predicting multimodal daily life patterns from cell phones. In: *ACM MMI*. pp. 277{280 (2009)
- ▶ 4. Ibrahim, M., Youssef, M.: Cellsense: a probabilistic rssi-based gsm positioning system. In: *IEEE GLOBECOM*. pp. 1{5 (2010)
- ▶ 5. Ibrahim, M., Youssef, M.: Cellsense: An accurate energy-efficient gsm positioning system. *IEEE TVT* 61(1), 286{296 (2012)
- ▶ 6. Kim, D.H., Hightower, J., Govindan, R., Estrin, D.: Discovering semantically meaningful places from pervasive rf-beacons. In: *UbiComp*. pp. 21{30. *ACM* (2009)
- ▶ 7. Kim, D.H., Kim, Y., Estrin, D., Srivastava, M.B.: Sensloc: sensing everyday places and paths using less energy. In: *ACM CENSS*. pp. 43{56 (2010)
- ▶ 8. Laasonen, K., Raento, M., Toivonen, H.: Adaptive on-device location recognition. In: *Pervasive Computing*, pp. 287{304. *Springer* (2004)
- ▶ 9. Laasonen, K., et al.: Mining cell transition data. Ph.D. thesis, University of Helsinki, Finland (2009)
- ▶ 10. LaMarca, A., Chawathe, Y., Consolvo, S., et al.: Place lab: Device positioning using radio beacons in the wild. In: *Pervasive Computing*, pp. 116{133. *Springer* (2005)
- ▶ 11. Meneses, F., Moreira, A.: Using gsm cellid positioning for place discovering. In: *IEEE Pervasive Health Conference*. pp. 1{8 (2006)
- ▶ 12. Nicholson, A.J., Noble, B.D.: Breadcrumbs: forecasting mobile connectivity. In: *ACM MobiCom*. pp. 46{57 (2008)
- ▶ 13. Shad, S.A., Chen, E., Bao, T.: Cell oscillation resolution in mobility prole building. *International Journal of Computer Science* 9(3), 205{213 (2012)
- ▶ 14. Sohn, T., Griswold, W.G., Scott, J., et al.: Place laban open architecture for location-based computing. In: *ESEC/FSE* (2005)
- ▶ 15. Wac, K.: mqol living lab: Experience of running longitudinal studies for mobile internet measurements. In: *IMC* (2013)
- ▶ 16. Zhou, C., Frankowski, D., Ludford, P., Shekhar, S., Terveen, L.: Discovering personally meaningful places: An interactive clustering approach. *ACM TOIS* 25(3), 12 (2007)
- ▶ 17. Zhou, C., Ludford, P., Frankowski, D., Terveen, L.: An experiment in discovering personally meaningful places from location data. In: *ACM CHI*. pp. 2029{2032 (2005)

PART EIGHT Traffic Modelling and Engineering

Heterogeneous Traffic Scheduling for Multi-Service Communication Networks

Geyong Min

Department of Mathematics and Computer Science
College of Engineering, Mathematics and Physical Sciences
University of Exeter, U.K.

Email: g.min@exeter.ac.uk

Abstract. The traffic in modern multi-service communication networks often exhibits the heterogeneous Long-Range Dependent (LRD) and Short-Range Dependent (SRD) properties. The hybrid scheduling mechanism is a promising scheme for traffic management, resource allocation, and Quality-of-Service (QoS) differentiation in multi-service networks. This talk will first report the method for modelling the heterogeneous network traffic characteristics and then present the analytical performance models of the priority scheduling and hybrid scheduling schemes in the presence of multi-class LRD and SRD traffic. The accuracy of the analytical models is validated through comparison between the analytical results and those obtained from simulation experiments of the actual system. The analytical models can be used to investigate the issues of resource allocation in multi-service networks subject to specific QoS constraints. Finally, this talk will present the emerging issues and future directions of analytical modelling of multi-service communication networks.

Heterogeneous Traffic Scheduling for Multi-Service Communication Networks

Professor Geyong Min

Department of Mathematics and Computer Science
 College of Engineering, Mathematics and Physical Sciences
 University of Exeter, U.K.

Email: g.min@exeter.ac.uk

Outline

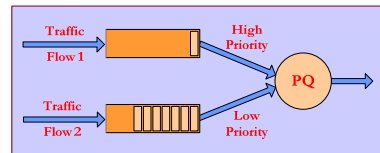
- Introduction
- Heterogeneous Network Traffic Characteristics
- Analytical Model for Priority Queueing Systems
 - Queue length distribution
 - Loss probability
- Analytical Model for Hybrid Scheduling
 - Queue Decomposition
- Resource Management and Optimization
 - Buffer Allocation
 - Minimum Bandwidth Requirement
- Conclusions

Introduction

- Advanced **multimedia applications** with diverse QoS requirements in modern communication networks
 - IP telephony, IPTV, video conferencing, interactive game, and distant learning
- **Traffic scheduling** plays a key role in the user-perceived QoS
 - Priority Queueing (PQ)
 - Generalized Processor Sharing (GPS)
 - Weighted Fair Queueing (WFQ) - Packet-based GPS (PGPS)

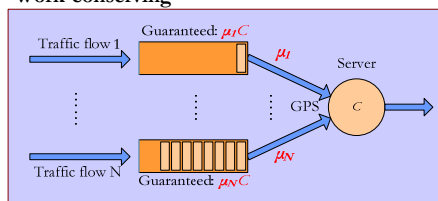
Introduction - Scheduling

- Priority Queueing (PQ)
 - PQ is an important and popular mechanism for provisioning of differentiated QoS.
 - Many research efforts have been made on its performance analysis, development, and applications.



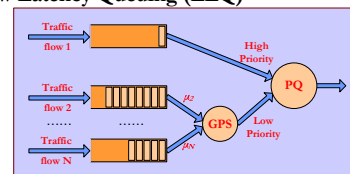
Introduction - Scheduling

- Generalized Processor Sharing (GPS)
 - Fairness
 - Traffic isolation
 - Work-conserving



Introduction - Scheduling

- Hybrid PQ-GPS Scheduling
 - Priority Queueing (PQ) + Generalized Processor Sharing (GPS)/its variants
 - Cisco: IP Real-time Transport Protocol (RTP) Priority, Low Latency Queueing (LLQ)



Introduction – Network Traffic

- **Heterogeneous traffic** in multi-service networks
 - **Long-Range Dependent (LRD) self-similar traffic**
 - Generated by multimedia applications; Strict QoS requirements
 - Scale-invariant burstiness
 - Large-lag correlation
 - **Short-Range Dependent (SRD) Poisson traffic**
 - Generated by the text communications

7

Introduction - Methodologies

- Performance evaluation studies can be achieved through either **simulation** or **analytical modelling**.
- The **convergence** of simulation to a steady state in the presence of self-similar traffic is often very **slow**.
 - **Analytical modelling** is a cost-effective method.

8

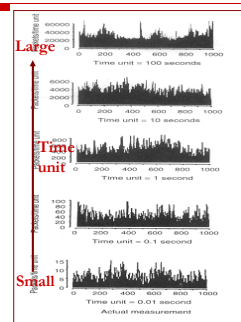
Introduction - Objectives

- Existing research problem
 - **No any analytical model** has been developed for the hybrid scheduling scheme due to the **interdependent relationship** between traffic flows and the **high complexity** of modelling **heterogeneous multimedia traffic**.
- Objectives
 - To develop **the first comprehensive analytical model** that can derive the **queue length distribution** and **loss probability** in PQ-GPS under **heterogeneous traffic**;
 - To use the model as a cost-efficient tool for **performance analysis** and **network resource management**.

9

Modelling LRD Self-Similar Traffic

- **Scale-invariant burstiness**
- **Being ubiquitous**
 - LAN, WAN, WWW
 - TCP, FTP, and TELNET
 - VBR video and multimedia systems
 - Wireless LAN, Ad-Hoc network traffic
- **Degrade network performance and QoS.**



10

Modelling LRD Self-Similar Traffic

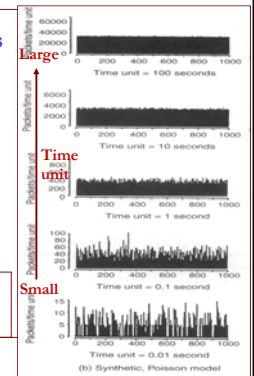
- **fractional Brownian motion (fBm)**
 - An efficient way for modelling and generating self-similar traffic
 - fBm is Gaussian in essence.
 - Formulation: $A_f(t) = mt + \sqrt{am}\bar{Z}_f(t)$
 - $A_f(t)$: The cumulative number of packets arriving up to time t
 - m : Mean arrival rate
 - a : Variance coefficient
 - Variance: $v_f(t) = amt^{2H}$
 - H ($0.5 \leq H \leq 1$): Hurst parameter

11

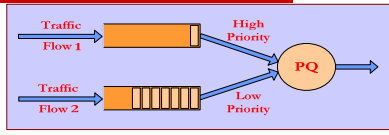
Modelling SRD Poisson Traffic

- **Burstiness over small time scales**
- **Poisson traffic**
 - Formulation: $A_p(t) = \lambda t + \bar{Z}_p(t)$
 - Variance: $v_p(t) = \lambda t$
 - Gaussian approximation: accurate when the arrival rate is large and the processing time tends to infinity

J. D. Cohn, Power spectrum and correlation function errors: Poisson vs. Gaussian shot noise, New Astronomy, 11:4(226–239), 2006.



Analytical Model for PQ:



- **Traffic Flow 1: High Priority, Multimedia, fBM**
- **Traffic Flow 2: Low Priority, Text, Poisson**
 - Total Queue Length Distribution
 - Individual Queue Length Distribution
 - Loss Probability

13

Analytical Model for PQ: Total Queue Length Distribution

- **Lower and upper bounds of the total queue length distribution of PQ systems**

■ Total queue length:

$$Q(t) = Q_f(t) + Q_p(t) = \sup_{s \leq t} \{A_f(s, t) + A_p(s, t) - C(t - s)\}$$

- Large Deviation Principle

■ Bounds:
$$\frac{\exp(-\frac{1}{2}Y(t_x))}{\sqrt{2\pi(1 + \sqrt{Y(t_x)})^2}} \leq P(Q > x) \leq \exp(-\frac{1}{2}Y(t_x))$$

$$t_x = \arg \min_t Y(t) \text{ and } Y(t) = \frac{(-x + (C - \sum_{i=1}^N \lambda_i)t)^2}{\sum_{i=1}^N v_i(t)}$$

- Queue length distribution: (Geometric mean)

$$P(Q > x) \approx \frac{\exp(-\frac{1}{2}Y(t_x))}{\sqrt{2\pi(1 + Y(t_x))^2}}$$

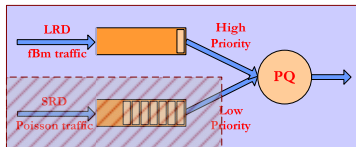
14

Analytical Model for PQ: Individual Queue Length Distributions

- **High priority fBm traffic**

- Fact:

- The high priority traffic in a PQ system is served in a manner as if the low priority traffic does not exist.



- Queue length distribution:

$$\frac{\exp(-\frac{1}{2}Y_f(t_x))}{\sqrt{2\pi(1 + \sqrt{Y_f(t_x)})^2}} \leq P(Q_f > x) \leq \exp(-\frac{1}{2}Y_f(t_x))$$

$$t_x = \arg \min_t Y_f(t) \text{ and } Y_f(t) = \frac{(-x + (C - m)t)^2}{v_f(t)}$$

15

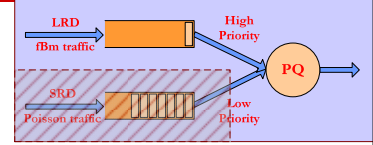
Analytical Model for PQ: Individual Queue Length Distributions

- **Low priority Poisson traffic**

- The total queue in a PQ system is almost exclusively composed of the low priority traffic.

- Empty Buffer

- Approximation: The queue length distribution of the low priority traffic can be approximated by the total queue length distribution.



- Queue length distribution:

$$\frac{\exp(-\frac{1}{2}Y_f(t_x))}{\sqrt{2\pi(1 + \sqrt{Y_f(t_x)})^2}} \leq P(Q_f > x) \leq \exp(-\frac{1}{2}Y_f(t_x))$$

$$t_x = \arg \min_t Y_f(t) \text{ and } Y_f(t) = \frac{(-x + (C - m)t)^2}{v_f(t)}$$

16

Analytical Model for PQ: Model Validation

- **Methodology:** Comparison between analytical and simulation results

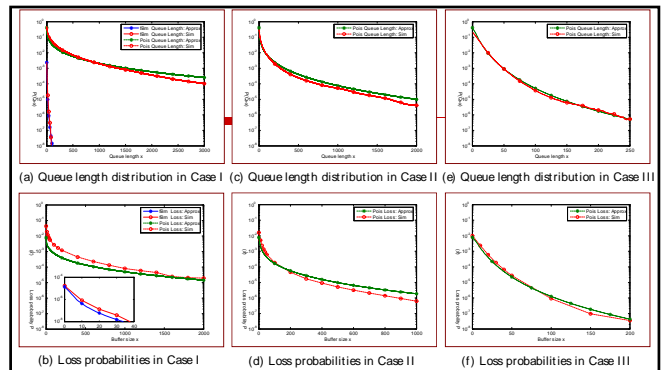
- A simulator of the PQ system: C++ language

- Two typical scenarios

- Hurst parameter $H_f \in \{0.85, 0.75\}$
- Service capacity $C = 120$
- Mean arrival rates
 - Case I: $\lambda_f = 90$ and $\lambda_p = 20$
 - Case II: $\lambda_f = 55$ and $\lambda_p = 55$
 - Case III: $\lambda_f = 20$ and $\lambda_p = 90$

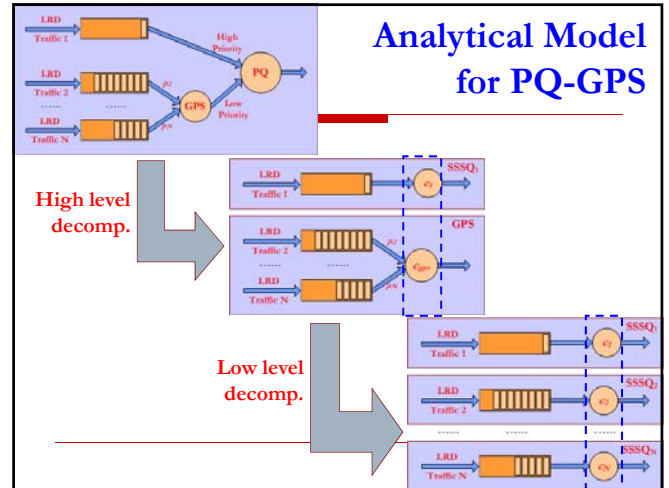
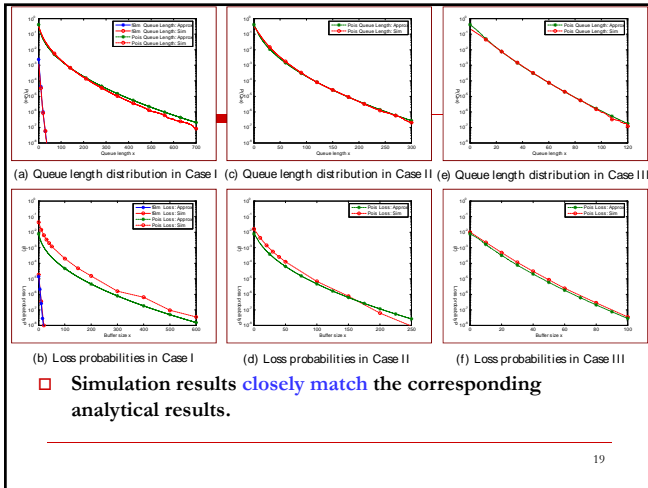
- Dominative traffic flow: fBm in Case I; Poisson in Case III

17



- Simulation results closely match the corresponding analytical results.

18



Analytical Model of PQ: Application

- Buffer allocation
 - Total buffer fixed: 400
 - QoS requirements:
 - $P_{L,f}(x_f) < 10^{-8}$
 - $P_{L,p}(x_p) < 10^{-2}$
 - How to allocate the buffer to fBm and Poisson traffic flows such that their QoS requirements are satisfied?
 - Admissible region:
 - $73 < x_f < 367$ and $33 < x_p < 327$

Analytical Model of PQ-GPS: Minimum Bandwidth Requirement

- Problem description
 - Given: buffer sizes χ_i ($1 \leq i \leq 4$) for different traffic flows
 - Objective: To determine traffic weights μ_i ($2 \leq i \leq 4$) such that the required service capacity C of the PQ-GPS system is minimized
 - Mathematically,

$$\text{Minimize } C(\mu_2, \mu_3, \mu_4)$$

subject to

$$\sum_{i=2}^4 \mu_i = 1.0;$$

$$P_{L,i}(\mu_2, \mu_3, \mu_4) \leq \ell_i$$

Where ℓ_i ($1 \leq i \leq 4$) is the QoS requirement of traffic flow i in terms of loss probability

Analytical Model for PQ-GPS: Minimum Bandwidth Requirement

- Parameter Settings:
 - fBm 1: $m_1=1200, H_1=0.8$; fBm 2: $m_2=300, H_2=0.7$;
 - fBm 3: $m_3=500, H_3=0.8$; fBm 4: $m_4=700, H_4=0.7$;
 - Buffer sizes: $\chi_1=1, \chi_2=20, \chi_3=20$, and $\chi_4=50$
 - Loss requirements: $\ell_1 < 10^{-8}, \ell_2 < 10^{-6}, \ell_3 < 10^{-5}, \ell_4 < 10^{-4}$
- Method:
 - To adopt the model to explore the possible solution space of C and examine the possible combinations of μ_2, μ_3 , and μ_4
- Results:
 - Minimum service capacity: $C=2920$
 - Traffic weights: $\mu_2=0.25, \mu_3=0.34$, and $\mu_4=0.41$

Conclusions

- Developed and validated the comprehensive analytical model for PQ and the hybrid PQ-GPS systems subject to heterogeneous traffic
- Demonstrated the application of the model to performance analysis and resource management

Bi-Greedy Cognitive Radio MAC Protocol for Supporting TCP/UDP Traffic

Humaira Afzal^{*a}, Muhammad Rafiq Mufti^a and Irfan Awan^a

^aSchool of Computing, Informatics and Media
University of Bradford, UK

Abstract — The Cognitive Radio technology is an auspicious area that has been developed as a new innovation to overcome future wireless demands. In Cognitive Radio Network, the vacant opportunities of licensed spectrum are dynamically accessed by unauthorized users to increase the bandwidth utilization of licensed spectrum. Cognitive Radio is a way to improve the efficiency and make best use of radio spectrum. Several MAC protocols exist in the literature to share the radio spectrum among multiple cognitive radio users. In this paper, a Bi-Greedy MAC protocol has been proposed to evaluate the performance of cognitive radio network in case of “Spectrum Handoff” and also eliminates the requirement of common control channel for the entire network. Since TCP and UDP are the basic transport protocol for internet applications today, therefore, the performance of the proposed protocol has been evaluated for TCP and UDP traffic. The significance of the proposed Bi-Greedy MAC has been investigated with the Simple Sequential MAC (SS-MAC) in terms of throughput as well as channel utilization.

The simulation results demonstrate the effectiveness of the Bi-Greedy MAC on TCP and UDP performance of cognitive radio users.

Index Terms — Cognitive radio, Cognitive radio network, Common control channel, Media access control.

I. INTRODUCTION

The spectrum bandwidth used by various radio technologies is fixed and is regulated by different authorities in different areas such as Federal Communication Commission (FCC) in United States and National Communication Commission in Taiwan. Under fixed spectrum allocation policies, most of the spectrum remains underutilized because of not allowing any unauthorized user to use, even if it remains vacant for most of the time ([1], [2]). On the other hand, some portion of unlicensed spectrum i.e. the ISM band, is being exhausted by evolving wireless services and applications, leading to so-called, the spectrum scarcity problem. The solution to this problem has been emerged in the form of Cognitive Radio (CR) technology where the unauthorized users are allowed to

use the vacant spaces also called white spaces/spectrum holes in the licensed spectrum in an opportunistic and non-interfering basis.

The Cognitive Radio Network (CRN) has two types of users; the Primary users (PUs) having authority to use the spectrum and the Secondary users (SUs) that have unlicensed access. The network formed by PUs is termed as Primary network and the network of SUs is called Cognitive Radio network (CRN) or Secondary network. The main functions of CRN are spectrum sensing, spectrum sharing, spectrum mobility, and spectrum decision. For an extended detail in CRN, readers are referred to [3] and [4]. The basic idea behind the CRN is to access the spectrum holes dynamically while preventing interference with PUs. The idea stirs the researcher to design protocols for successful operation of any CR network. A number of MAC protocols have been developed so far and through wide simulations, many MAC protocols have shown improvements in throughput by considering various aspects but a little attention has been paid by the researchers to assess the performance of TCP and UDP over their CR MAC protocols. We have focused on throughput improvement by taking TCP and UDP protocol performance into account in an ad-hoc CRN. The contribution of this paper is two-folds: first is the design of Bi-Greedy MAC protocol using TCP and UDP traffic for the system without common control channel requirement. Secondly, the proposed protocol provides the solution for co-existence/self co-existence problems avoidance.

The rest of the paper is organized as follows. The important related work from the literature is briefly discussed in Section-II, proposed Bi-Greedy MAC is explained in Section-III, simulation results and performance evaluations are given in Section- IV. Finally the Section-V concludes the paper and proposes some future work.

II. RELATED WORK

Different areas of CRNs have been explored and various research problems are being addressed in the literature. One of the research problems is dynamic spectrum sharing whose solution come up in the form of MAC protocols. In CRNs, since the channel availability may change quickly so

synchronizing the entire network and acquiring up-to-date environmental information are the key objectives that must be taken into account while designing the MAC protocols. The characteristic aspects, advantages and limiting factors of MAC protocols are extensively discussed in [5-6] for CRN. The protocols are generally classified into three categories: random access, time slotted, and hybrid. The further classification is primarily based on number of radio transceivers, time synchronization, infrastructure assistance, and complexity.

In [7], a cross layer based opportunistic MAC protocol is proposed for cognitive radio ad hoc network (CRAHN) using Markov chain model. The scheme combines the spectrum sensing functionality of the physical layer with the packet scheduling functionality of the MAC layer. Each SU is facilitated with two transceivers; one is reserved for control channel and other is used as CR, which periodically sensed the spectrum for unused channels. The spectrum sensing capability is performed through random sensing and negotiation based sensing policies. The protocol is claimed to exhibit a lesser amount of control overhead but does not provide solution of multichannel hidden terminal problems. The same strategy is also exploited in energy efficient multichannel protocol called ECR-MAC [8]. The same network architecture is utilized where each node is attached with a single half duplex radio transceiver. Power can be saved if the nodes not involving in the communication go into the dose mode. The protocol requires global time synchronization in order to avoid hidden terminal problem. With the help of periodic beaconing, each node is aware of available channels within two hop neighbours. In [9], a distributed MAC protocol (DOSS-MAC) provides a solution to address hidden and exposed terminal node problems. Three different transceivers are required to manage control, data and busy-tone channel. Whenever a node wants to transmit or receive data on a given data channel, both sender and receiver send a busy-tone signal in the corresponding busy-tone channel. Using the above mentioned scheme, the spectrum is not effectively utilized as part of the spectrum is wasted in control and busy-tone information propagation. The requirement of multi transceivers is not reasonable as two of them do not contribute any participation in data communication. Moreover, DOSS-MAC protocol is not feasible for the nodes having only half duplex radio. In 'SCA-MAC' (Statistical Channel Allocation) spectrum access scheme, proposed in [10], the SU selects the channel that has the highest successful transmission probability. However, the computational cost for finding the successful transmission probabilities, increases exponentially with increasing the licensed channels.

In [11], a multichannel MAC protocol (CREAM-MAC) is developed in which each node is equipped with one CR enabled transceiver containing N sensors to sense N channels. Six-way handshake scheme (RTS/CTS/CST/CSR/DATA/ACK) is employed to ensure successful transmission. The co-existence and self co-existence issues are controlled by RTS/CTS and CST/CSR respectively but the effectiveness of the scheme is dependent upon the licensed channels being selected cooperatively. Furthermore, the assumption of prior availability and

reliability of CCC, makes CREAM-MAC unjustified. A cooperative relay and various related issues are discussed in [12] for CRN. A distributed cooperative MAC is also designed and implemented on a test bench. Cooperative transmission is basically comprised of three components: source, destination and relay. The relay node receives the transmission from source node and forwards it to intended destination node. In the proposed scenario, the spectrum rich node is selected as a relay node and it is proved that the cooperative relay node helps the SUs transmission and improves the spectrum efficiency. However, if a common channel is available between source and destination then there will be no need of relay node. For such a system, the transmission requirement of each CR user has to be considered [13].

Opportunistic spectrum access facilitates PUs with sufficient protection and SUs should not create any interference to the PUs [14]. Spectrum handoff is a good tool that provides not only interference avoidance for the PUs but also offers extra transmission opportunities for the SUs. However, it exerts inherent transmission delay that may affect the throughput of SUs. In addition of PUs protection, the basic purpose is to optimize the throughput of CR network dependent on spectrum sensing results. A new research work is reported in [15], where OSA dynamics is studied to optimize the MAC frame structure for OSA which has more concerned with physical layer tradeoff and MAC layer tradeoff. Both tradeoffs need to be balanced for optimized MAC frame structure. At the physical layer, tradeoff is between false alarm and misdetection whereas, at the MAC layer, tradeoff is between optimizing throughput of SUs and decreasing collisions with PUs. Therefore, based on these tradeoffs, a throughput model for CR network has been developed. The simulation results validate the effectiveness of the proposed design.

Most of the protocols discussed in this section exploit either TCP traffic or UDP traffic but our Bi-Greedy protocol considers both types of traffic and eliminates the need of CCC for the network.

III. PROPOSED BI-GREEDY MAC PROTOCOL

The proposed network is made up of PUs, SUs and multiple non-overlapping channels with a single interface- multiple channels environment. The network is assumed without common control channel (CCC), and each SU is equipped with a single radio. Before actual data transmission, each SU is made synchronized. Channels are fixed for PUs but SUs are assigned channels during run time after performing extensive channel search which is basically greedy and is based upon the proposed algorithm called Bi-Greedy algorithm. The algorithm performs searching for the channels greedily using TCP/UDP traffic. The SUs search the channel based on long term observations and channel usage statistics. The PUs traffic is expressed by ON/OFF pattern where ON period denotes the channel being used by the PU dynamically and OFF pattern represents no PU activity and therefore, the channel is never considered as spectrum hole, and will be available for the SUs. For clarity, the key features of the proposed network are listed below:

- Each channel can accept two types of users; PUs and SUs. PUs have the higher priority than SUs.
- The PUs have pre-emptive priority to use the channel and can interrupt the SU's transmission. The interrupted SU can resume its remaining work instead of retransmitting the whole work [16]. The target channel for the interrupted SU can be different from the current operating channel.
- During the transmission, the SU may face multiple interruptions from the PUs.
- Only one user can use the channel at a time.
- If PU receives a packet from SU on its native channel, then this packet would be dropped by the PU.

The algorithm for channel selection spectrum hand off is as follows:

Bi-Greedy Algorithm for Channel Selection	
1	Listen to the channel
2	If a channel is found to be idle then
3	Transmit data packets
4	else
5	goto step 1
6	If PU is detected on the channel then
7	If simulation time > t then
8	Find the channel with threshold time maximum received packets
9	Jump to the channel selected in the above step
10	else
11	Calculate the number of packets received on the channel
12	Increment in the channel number
13	goto step 1
14	else
15	Send data packets
16	goto step 1
17	end if
18	end if
19	end if

On boot up, each SU is tuned to listen to the channels. The channel is selected randomly if it is found idle and packets are then transmitted on the selected channel. In order to get the channels statistics, Simple Sequential MAC [16] is executed for t seconds. Channel statistics refer to the most available channel where SU has spent maximum time duration. The channel list is then sorted out in such a way that the last element of the list represents the channel lying free for most of the time. The sender is synchronized with receiver after the threshold time t is over. During data transmission, if SU receives a packet from PU then spectrum handoff is performed on the channel having maximum channel usage statistics in the first t seconds.

IV. SIMULATION RESULTS AND PERFORMANCE EVALUATION

The effectiveness of the proposed Bi-Greedy MAC is evaluated by simulations using NS2.31 with CRCN patch. We

compare its performance with Simple Sequential MAC (SS-MAC) under TCP and UDP traffic. The main parameters used in simulations are given in Table-1. The Bandwidth of the network is divided into 12 different data channels and traffic load generated by each pair of PUs is taken as variable. We consider five different scenarios where low data rate and high data rate PUs are taken into account. The low data rate is assumed less than 200 kbps and high data rate is supposed greater than 800 kbps. These scenarios are carried out using transport protocols by keeping the SU data rate constant throughout the simulation period. The simulation is first run by taking current time less than 40 seconds and then taking current time less than 60 seconds. The number of channels selected along with data rates are listed in Table 2.

Table1: Simulations Parameters.

Parameter Name	Value
Number of Data Channel	12
Number of PUs	24
Number of SUs	2
Number of radios per user	1
Simulation time	100 sec
Transport Protocols	TCP, UDP
Traffic	CBR, FTP
Rate of secondary user pair	11 Mbps
Threshold time	$\leq 40, \leq 60$

Table 2: Scenarios description.

	Channels	Channel data rate (Kbps)
Scenario-1	1	> 800
	11	< 200
Scenario-2	3	> 800
	9	< 200
Scenario-3	6	> 800
	6	< 200
Scenario-4	9	> 800
	3	< 200
Scenario-5	11	> 800
	1	< 200

The Figures 1-2 illustrate the throughput when time is kept less than 40 seconds. The average throughput of SS-MAC is 45.51 kbps whereas the average throughput of Bi-Greedy MAC is 65.87 kbps which is 30.91 % more than SS-MAC under TCP traffic. In case of UDP flow, the average throughput of SS-MAC is 386.08 kbps and average throughput of Bi-Greedy MAC is 417.80 kbps which is 7.59 % higher than SS-MAC. Therefore Bi-Greedy MAC exhibits significant improvement over SS-MAC.

The similar performance of Bi-Greedy MAC is observed when threshold time is increased to 60 seconds in Figures 3-4. The average throughput of SS-MAC is 47.17 kbps and the average

throughput of Bi-Greedy MAC is 59.65 kbps which goes up to 20.92 % of SS-MAC under TCP traffic. Using UDP traffic, the average throughput of SS-MAC is 359.30 kbps and average throughput of Bi-Greedy MAC is 397.52kbps, which is 9.61 % higher than SS-MAC. Therefore, either using TCP or UDP traffic, the performance of Bi-Greedy MAC is far better than SS-MAC.

SS-MAC	45.51	386.08	47.17	359.30
--------	-------	--------	-------	--------

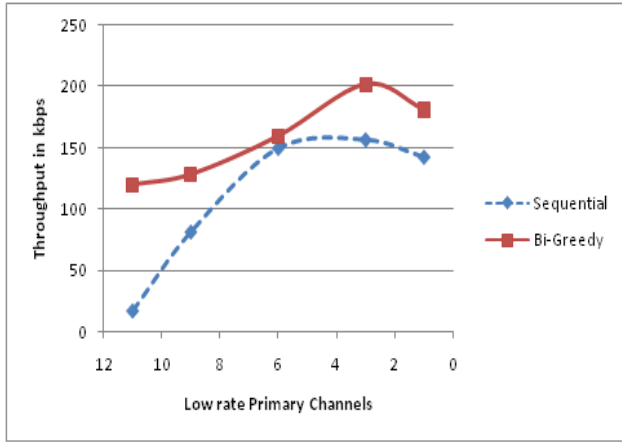


Figure 1: Throughput with Current Time less than 40 using TCP traffic.

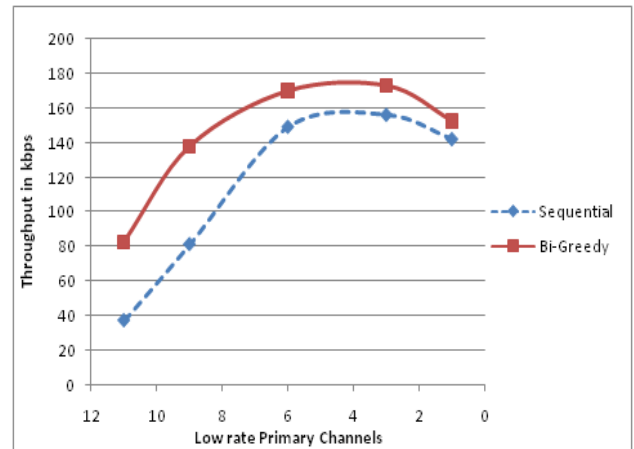


Figure 3: Throughput with Current Time less than 60 using TCP traffic.

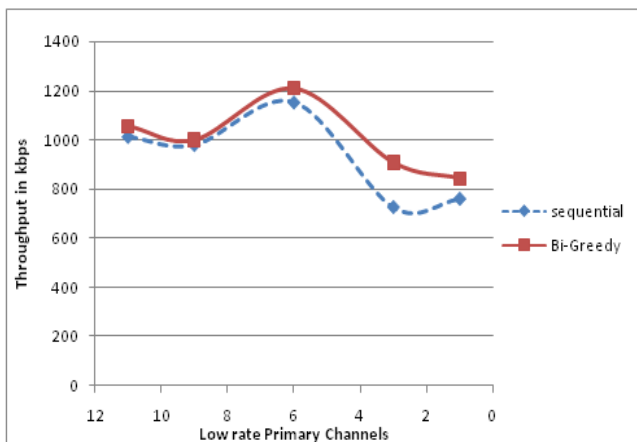


Figure 2: Throughput with Current Time less than 40 using UDP traffic.

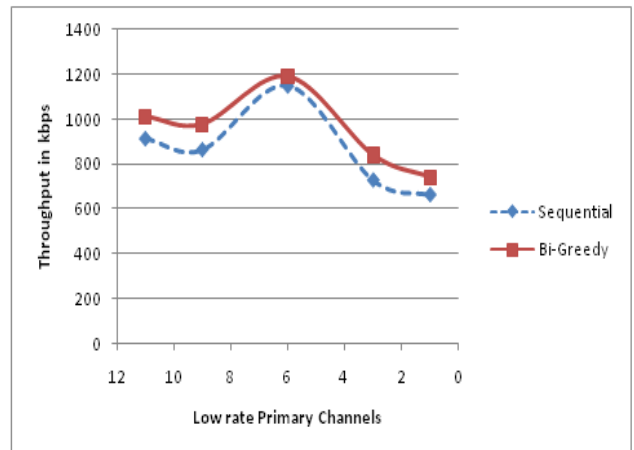


Figure 4: Throughput with Current Time less than 60 using UDP traffic.

Table 3: Performance comparison of proposed Bi-Greedy MAC with SS-MAC.

Protocols	less than 40 seconds		less than 60 seconds	
	Average Throughput (kbps)		Average Throughput (kbps)	
	TCP	UDP	TCP	UDP
Bi-Greedy MAC	65.87	417.80	59.65	397.52

It is noteworthy that the throughput range of TCP is lower than UDP in all the above cases. This is due to the congestion control mechanism and bi-directional (data/ack cycle) behavior of TCP. Therefore, TCP is sensitive to packet losses. Moreover, due to slow nature, TCP cannot perform well for bandwidth utilization in collision prone environment. Although, TCP has some shortcomings but we cannot ignore its reliable nature. The reliability, in time delivery and non-duplication characteristics overcomes the shortcomings of having low values of throughput when compared to UDP.

V. CONCLUSION AND FUTURE WORK

We proposed Bi-Greedy MAC for CRN and compared its performance with SS-MAC under TCP and UDP traffic. Simulation results have shown that Bi-Greedy MAC could

increase the average throughput of CRN by performing spectrum handoff proactively on the channels having better packet reception rate. The suggested scheme can perform spectrum handoff in an efficient way improving the spectrum utilization and also avoids the need for CCC for the entire network. The concept presented can be extended by incorporating the issue of spectrum handoff delay. Furthermore, the reactive approach to perform spectrum handoff can also be embedded in the proposed scenario with multi-channel multi-interface environment.

ACKNOWLEDGEMENT

This research is financially supported by Bahauddin Zakariya University Multan, Pakistan. The authors are grateful to the research group at University of Bradford.

REFERENCES

- [1] Report of the Spectrum Efficiency Working Group, Federal Communications Commission, Spectrum Policy Task Force, Nov 15, 2002.
- [2] V. Valenta, Z. Fedra, R. Marsalek, G. Baudoin, and M. Villegas, "Towards cognitive radio networks: Spectrum utilization measurements in suburb environment", in Proc. IEEE, Radio and Wireless Symposium, San Diego, CA, USA, 18-22 Jan. 2009, pp.352-355.
- [3] I. F. Akyildiz, W. Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation / dynamic spectrum access / cognitive radio wireless networks: A survey," Computer Networks Journal (Elsevier), vol. 50, pp. 2127-2159. 2006.
- [4] D. Niyato, E. Hossain, "Cognitive radio for next-generation wireless networks: an approach to opportunistic channel selection in IEEE 802.11-based wireless mesh", IEEE Trans. On Wireless Communications, vol. 16, no.1, pp.46-54, February 2009.
- [5] C. Cormio and K. R. Chowdhury, "A survey on MAC protocols for cognitive radio networks", Ad Hoc Networks, vol. 7, no. 7, pp. 1315-1329, September 2009.
- [6] A. D. Domenico, E. C. Strinati, and M. D. Benedetto, "A Survey on MAC Strategies for Cognitive Radio Networks", IEEE Communications Surveys & Tutorials, vol. 14, no. 1, First Quarter 2012.
- [7] H. Su, Xi Zhang, "Cross-Layer Based Opportunistic MAC Protocol for QoS Provisioning Over Cognitive Radio wireless Networks", IEEE Journal on Selected Areas in Communications, Vol. 26, No. 1, pp. 118-129, 2008.
- [8] S. M. Kamruzzaman, "Energy efficient Multichannel MAC protocol for Cognitive Radio ad-hoc Network", International Journal of Communication Networks and Information Security, Vol. 2, No. 2, 2010.
- [9] L. Ma, X. Han and C. Shen, "Dynamic open spectrum sharing for wireless ad hoc networks", in Proc. IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, November 2005, pp. 203-213
- [10] A. Hsu, D. Wei, and C. Kuo, "A cognitive MAC protocol using statistical channel allocation for wireless ad-hoc networks", in Proc. IEEE, Wireless Communications and Networking, March 11-15, 2007, pp. 105-110.
- [11] Zhang, Xi, and Hang Su. "CREAM-MAC: Cognitive radio-enabled multi-channel MAC protocol over dynamic spectrum access networks", IEEE Trans. on Selected Topics in Signal Processing, vol. 5, no.1, pp.110-123, 2011.
- [12] T. Lou, H. Chao and T. Tsai, "On synchronized channel sensing and accessing for cognitive radio users in IEEE 802.11 wireless networks", in Proc. IEEE PIMRC 2009, Tokyo, Japan, 13-16 Sept. 2009, pp.757-761.
- [13] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green Cellular Networks: A Survey, Some Research Issues and Challenges", IEEE Communications Surveys & Tutorials, vol. 13, no. 4, fourth quarter 2011.
- [14] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access: signal processing, networking, and regulatory policy", IEEE Signal Process. Mag., vol. 24, no. 3, pp. 79-89, May 2007.
- [15] J. Zhang, L. Qi, and H. Zhu, "Optimization of MAC Frame Structure for Opportunistic Spectrum Access", IEEE Transactions on Wireless Communications, vol. 11, no. 6, June 2012.
- [16] S. Srinivasa and S. A. Jafar, "The Throughput Potential of Cognitive Radio: A Theoretical Perspective," IEEE Communications Magazine, pp. 73-79, May 2007.
- [17] U.S. Khan and T. Maqsood, "CRN survey and a simple sequential MAC protocol for CRN learning", in Proc. IEEE, The Second International Conference on Advances in Cognitive Radio, COCORA 2012, pp. 22-27.

Policy Based RAT Selection incorporating IEEE 802.21 Framework for Analysis of Cellular Traffic Offloading through Wi-Fi

A.Ahmed, M.Ali, Y.F.Hu, P.Pillai

School of Engineering, Design & Technology, University of Bradford,
Bradford, West Yorkshire BD7 1DP, UK.

e-mail: {a.ahmed84, m.ali70, y.f.hu, p.pillai}@bradford.ac.uk

Abstract. The rapid growth in the mobile communication data has posed a challenge to the cellular networks to meet the high rising data traffic demands from the end users. The data traffic from the mobile devices is increasing in double folds every year and this rising tendency is anticipated to grow further with explosive growth and development of mobile applications. It is crucial for the mobile network operators (MNOs) to handle this capacious data traffic by effectively addressing the bottlenecks imposed by such data traffic. In this paper, we evaluate a heterogeneous network environment, of Universal Mobile Telecommunication System (UMTS) and Wi-Fi access points (APs), to offload the UMTS network through application based traffic migration to Wi-Fi APs. This traffic migration is carried out by utilizing the IEEE 802.21 Media Independent handover (MIH) framework. The benefits of such deployment are quantified by analysing the scenario with 200 mobile users with the active connections of mobile data traffic comprising of voice, video streaming and file transfer applications. A new Policy Based RAT selection strategy has been proposed and comparison between the proposed strategy and the base line strategy has also been analysed through simulation results. Results have revealed the optimised performance of proposed policy based RAT selection using IEEE 802.21 to offload cellular networks through Wi-Fi.

Keywords; Policy Based RAT Selection, IEEE 802.21, MIH, Cellular Network Offloading

1 Introduction

The access of the mobile users to web contents has been extensively enhanced by the provision of ubiquitous internet access in smart phones and this has greatly served in the popularity and commercialization of 3G networks. The extensive growth in the number of 3G users and their demand for high bandwidth multimedia communication has posed the cellular network operators with the challenge of maintaining user satisfaction and their service demands. The recent studies have revealed that 3G traffic capacity has been exceeded by the tremendous data traffic generated by mobile users, hence degrading the quality of service (QoS) offered to mobile users [1]. Cisco forecast [2], published under Visual Network Index (VNI), has revealed that the traffic generated by the mobile users is growing rapidly with percentage compound annual

growth rate (CAGR) of 131 and the monthly mobile data traffic in 2013 is expected to rise above 2 Exabyte. This exponential growth of mobile data traffic urges the MNOs to address the capacity bottlenecks in the cellular networks.

The most straight forward solution to this challenge is to expand the capacity of current 3G networks but this solution results in higher capital expenses (CAPEX) and lead to higher operating expenses (OPEX) as well. The MNOs have upgraded their networks with High Speed Packet Access (HSPA) and Long Term Evolution (LTE), however the bandwidth might not be enough even with 4G [3]. The other desirable solution is to effectively offload the cellular networks. As most of the mobile network operators support flat rate charging regime for unlimited internet access, it would be beneficial for the MNOs to offload the mobile data traffic to other low cost radio access technologies. This offloading solution becomes more feasible in the current radio environment which is a heterogeneous network environment with multiple radio access technologies present in same environment. This offloading solution is further complimented by the availability of multi radio interfaced portable devices that have also served in the increased popularity of mobile data communication. Previously the networks have been replaced by the other networks to facilitate the user demands but the present trend in rapidly growing mobile data communications demands interoperability between the different networks. Since each network has its own merits and limitations, the coexistence of these networks in an all IP based heterogeneous network will require each network to compliment the other networks to ensure seamless connectivity anytime, anywhere while utilizing the best available network. This seamless connectivity is provided by the switching of multi interfaced devices from one access network to other access network.

In this paper we propose and evaluate the offloading of data traffic through seamless transfer of mobile data traffic from 3G network to Wi-Fi network. The seamless data transfer is facilitated with the utilization of MIH framework. We focus our analysis on an urban environment with high data traffic demands. In our scenario we have considered the abundance of 3G network coverage with presence of Wi-Fi APs along the route of mobile users. The end user terminals are assumed to support multi radio technologies, which traverse through the multi Radio Access Technology (RAT) environment with active data connections. This is assumed to be a common scenario in the typical urban 3G coverage.

2 Cellular Traffic Offloading

In this section we present an overview of solutions to the offloading of cellular networks that includes the Delay Tolerant Networking (DTN), peer-to-peer opportunistic offloading, Femtocells and Wi-Fi hot-spots. Our analysis is based on the integrated concept of Wi-Fi for data offloading along with the session continuity during the whole offloading cycle.

2.1 Delay Tolerant Networking (DTN)

The traffic in a Delay Tolerant Networking (DTN) is categorized into three classes based on their delay tolerance. The three classes of categorization, in the ascending order of prioritization, are bulk, normal and expedited. Packets are processed based on the priority level that means the packets belonging to the bulk class of traffic, e.g., FTP traffic is dealt once there is no packet left from other classes. As majority of data from mobile traffic is from the bulk class, some MNOs have started using DTN to offload the bulk data across the internet [4].

2.2 Peer-To-Peer Opportunistic Offloading

This opportunistic offloading technique is proposed by Han et al. [5] by choosing k mobile users for pushing the contents in a peer-to-peer mobile network and these k users form the initial set for content forwarding. The users present in the initial set, facilitate the transfer of contents to other users through utilization of low range attachments (e.g., ad hoc WLAN or Bluetooth). The service delivery can be improved by the identification of social groups [6] in the network and transferring group specific contents to individual groups. It has been shown by Han et al. that a large percentage of cellular traffic can be offloaded by choosing initial set from straightforward heuristic based on precedent information.

2.3 Femtocells

The Femtocells were introduced in the cellular networks to optimize the voice and data services in an indoor environment [7] and introduction of Femtocells does not require any hardware upgrade on the end user terminal as they operate in the same frequency spectrum of the MNO. Although this offloads the outdoor cellular network but it cost in terms of CAPEX and OPEX. Large deployment of Femtocells also poses the challenge of high interference amongst the Femtocells, hence degrading the QoS offered by the MNO to end mobile user. Femtocells only address the indoor environment and do not handle mobility in the macroscopic environment [3].

2.4 Wi-Fi Hotspots

The presence of Wi-Fi in an urban and sub-urban environment is ubiquitous and Wi-Fi operational frequency belongs to unlicensed frequency band. Wi-Fi do not pose any interference issues to the cellular network and several interworking solution for 3G offloading have been presented from the standardization bodies and also from the industry. Certain applications have been developed to initiate services on Wi-Fi network rather than 3G networks for example an iPhone application called Line2 commences voice calls through the Wi-Fi network. The IEEE 802.21 (MIH) standard provides the seamless mobility framework across 3G and Wi-Fi networks. To supplement the 3G network through Wi-Fi for the applications that are delay tolerant, Wiffler scheme is proposed by Balasubramanian et al. [8].

Non-seamless offload is still possible between the 3G and Wi-Fi network but the connection is dropped from the first access network and then re-established onto the other network. In this paper the target analysis is based on the augmentation of 3G network

by seamlessly switching the traffic from the 3G network to Wi-Fi network through MIH.

3 MIH Frame Work

The IEEE 802.21 Working Group put an effort to ratify the Media Independent Handover (MIH) standard, to enhance the user centric mobility handovers and to enable handovers across heterogeneous networks. IEEE 802.21 standard [9] defines the media independent handover function (MIHF) to facilitate the seamless roaming of the mobile devices across heterogeneous access networks [10]. MIHF is logically designed as a shim layer between the link layer (L2) and upper layers in the protocol stack of both the Mobile Node (MN) and network element [11]. IEEE802.21 provides network information to higher layers by enabling link layer intelligence to optimize the inter radio access technology handover. Figure 1 below illustrates the position of MIHF and also represents the elements of the IEEE 802.21 MIH standard.

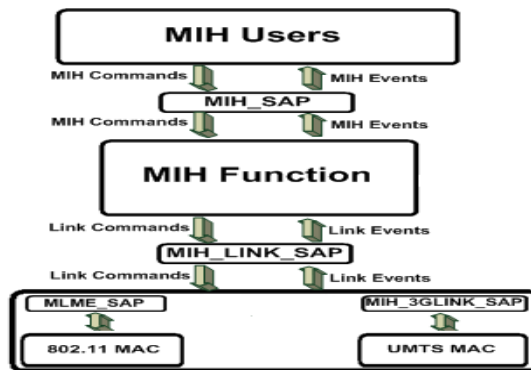


Fig. 1. MIH Elements

4 Proposed RAT Selection Policy

The purposed RAT selection approach follows the policy where all the user with active voice connections only are bounded to stay on the UMTS network wherever UMTS is available and users with other active connections such as video and file transfer should handover to the Wi-Fi networks on availability.

Handover of only selected users from UMTS to Wi-Fi, who are consuming large bandwidth, slightly deviates away from the Always Best Connectivity (ABC) concept [12]. But this approach gives better results as the UMTS network gets offloaded to a considerable amount. This strategy can offload the UMTS network to a large extent in the areas where Wi-Fi is available such as Wi-Fi hot-spot like train stations, airports and shopping centres etc.

4.1 Handover Procedures

This section explains the steps involved in the message sequence charts for the handover between UMTS and the Wi-Fi access technologies. The seamless handover is achieved when moving the mobile users between UMTS and Wi-Fi, as the IEEE 802.21 MIH is utilized for the vertical handovers between considered access technologies.

Handover from UMTS to Wi-Fi.

Figure 2 shows the SAP primitives exchanged between different entities in the handover procedure from UMTS to Wi-Fi. As represented by Figure 2 the 802.11 MAC layer in the mobile node, after detecting and registering with Wi-Fi network, it sends MLME-LinkUp.indication message to the MIHF. In step 3, MIHF sends MIH_Link_UP.indication to MIH User. A set of messages from step 4 to 7 acquire the neighbouring networks information. Step 8 checks for the required resources in Wi-Fi for handover. The MIH User decides whether to perform handover or not in step 9. Steps 10 to 12 show the handing over of the connections to the Wi-Fi network. Finally, step 13 and step 14 make sure that traffic flow has been re-establish between mobile node and source and then release bindings with UMTS network.

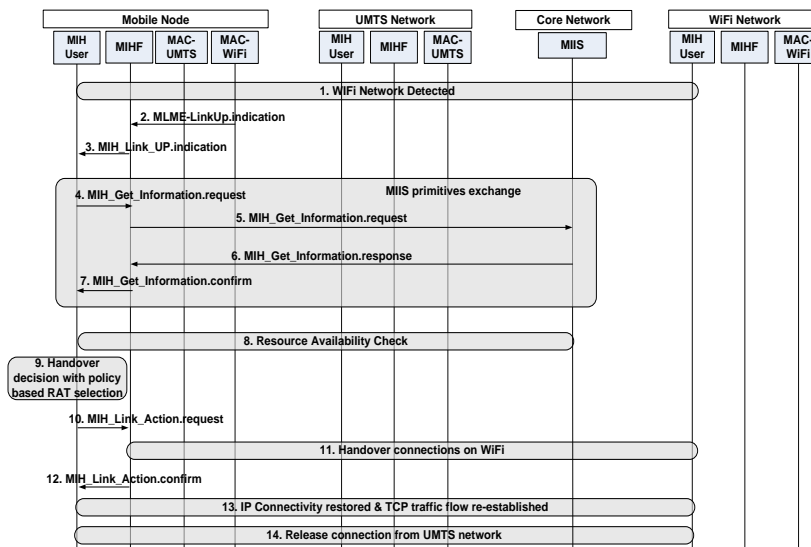


Fig. 2. Handover from UMTS to Wi-Fi

Handover from Wi-Fi to UMTS.

Figure 3 shows the handover procedure when mobile user moves away from the Wi-Fi coverage area and enters a UMTS coverage area. The first step in Figure 3 is the message MLME_MREPORT.indication from MAC Wi-Fi to MIHF. This is the periodic message which carries parameters of Wi-Fi link.

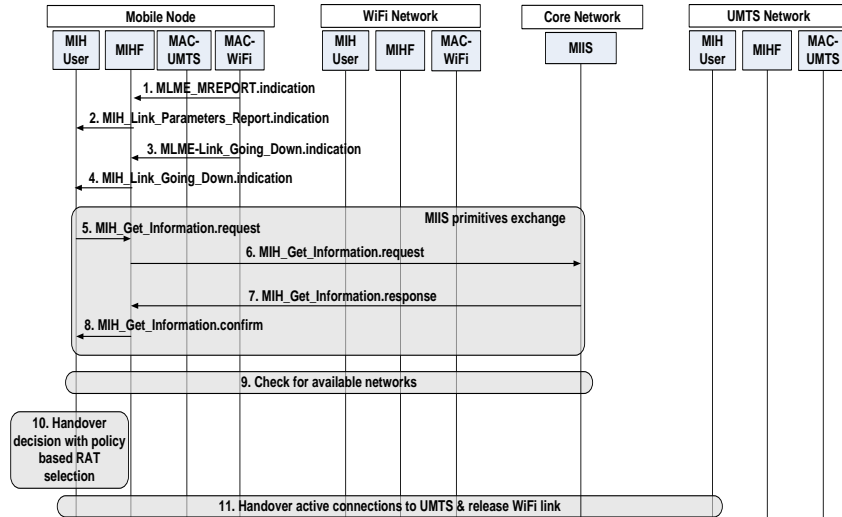


Fig. 3. Handover from Wi-Fi to UMTS

In step 2 the MIH User is being updated with link parameters report. Step 3 shows the message link-Going_down from Wi-Fi MAC to MIHF, which represents that mobile node, is gradually losing the connectivity with Wi-Fi. Step 4 informs the MIH User about link going down event. From step 5 to step 8 the messages are used to acquire neighbouring networks information from MIIS. Step 9 shown as bubble represents the process of scanning on all interfaces supported by mobile node. Step 10 and step 11 are for selecting the UMTS network and handover all active connections to UMTS.

5 Simulation Scenario

The simulation scenario is designed in such a way that multiple Wi-Fi networks overlap the coverage area of the UMTS coverage area so that the effects of purposed RAT selection strategy can be observed in ideal context. The Figure 4 shown blow represents the simulation topology. In this scenario a large group mobile users start moving from the UMTS coverage area and passes through the common coverage areas of UMTS and other Wi-Fi networks.

Figure 4 shows the simulation scenario where UMTS has the global coverage in the simulation grid whereas each Wi-Fi has the 25meters radius or a range of 50 meters of coverage. All the wired links (shown in dotted line) are 1Gbps. Wi-Fi provides the data rate of 11Mbps whereas the UMTS provides a data rate of 384Kbps for Up-link and Downlink. In simulation the group of mobile users are set to move from UMTS coverage area and pass through the common coverage areas of UMTS and different Wi-Fi networks. There are 200 mobile nodes (MNs) used in the simulation and there are three different types of connections namely: Video, Voice and File transfer. Video and Voice connections are setup as a UDP connection whereas file

transfer is set up using the TCP connection. Each mobile node has one of the three connections.

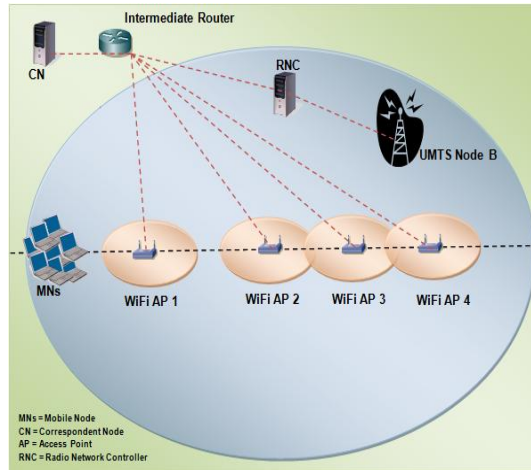


Fig. 4. Simulation Scenario

6 Results

The simulation results have been analysed to observe the efficiency of the proposed RAT selection strategy as compared to the baseline (ABC) RAT selection strategy. Figure 5 presents the packet rate of UMTS and Wi-Fi networks for baseline RAT selection strategy. The plot area in Figure 5 is categorised by region 1 ($X1 \times Y1$) and region 2 ($X2 \times Y2$).

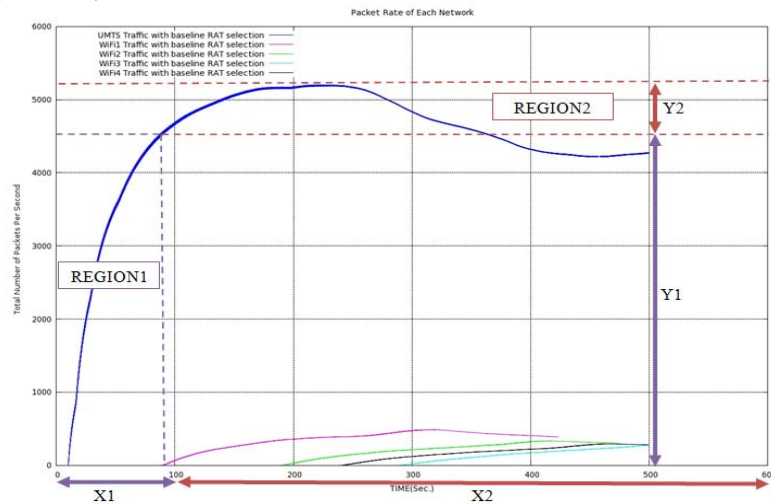


Fig. 5. Data Rate of Each Network Using Baseline RAT Selection

In region1 where it is only the UMTS network that is serving the available users, it is clear that the rate $\Delta Y1/\Delta X1$ is very high for time and the UMTS network is highly loaded. But as soon as the users entered a common coverage area of UMTS and Wi-Fi, the rate $\Delta Y2/\Delta X2$ for UMTS is greatly reduced; hence Wi-Fi is effectively off-loading the UMTS network through seamless handover of traffic from UMTS to Wi-Fi networks using the IEEE 802.21 MIH framework.

The baseline RAT selection procedure for offloading the UMTS network has been optimised by introducing the policy based IEEE 802.21 framework for RAT selection. Figure 6 shows the improved packet rate for the scenario where policy based RAT selection is applied as the policy based RAT selection does not moves users with active voice connection from UMTS to Wi-Fi. On other hand where baseline RAT selection is applied, all the users go for the best available network and handover to the Wi-Fi network when they enter Wi-Fi coverage area.

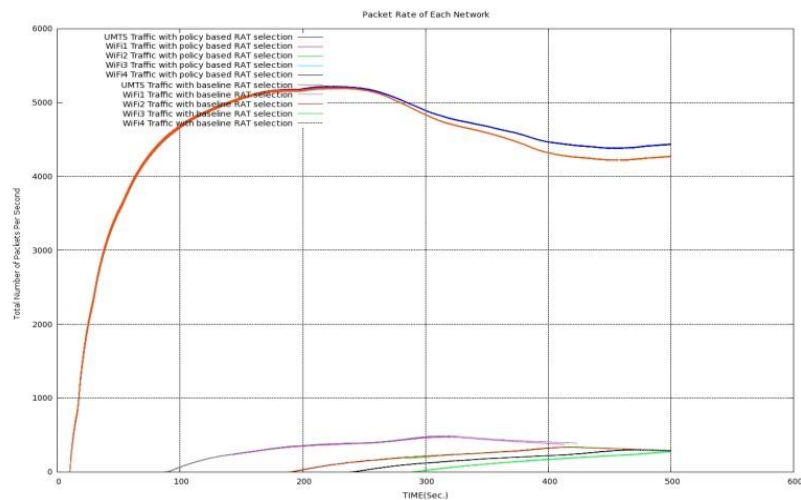


Fig. 6. Baseline Vs Policy Based RAT selection Data Rate

The performance efficiency of policy based RAT selection is clearly reflected through the comparison of handover latencies for policy base RAT selection and baseline RAT selection strategies. Figure 7 shows that the handover latencies observed by average MN has been greatly reduced by using the policy based RAT selection as compared to the baseline RAT selection strategy. The improved performance in handover latencies has been observed due to the fact that policy based RAT selection do not shift all the traffic to Wi-Fi, in common coverage area, whereas baseline RAT selection shifts all the traffic to Wi-Fi network.

In UMTS, each user is provided with a separate dedicated channel for communication as long as the UMTS network has the capacity to do so whereas in Wi-Fi all the users share the network resources. Baseline RAT selection shifts all the traffic to Wi-

Fi which leads to congestion in the Wi-Fi network and increased handover latencies. Policy based RAT selection, shifts the traffic based on the data rate of the application and keeps the voice connection always in the UMTS network. This controlled hand-over leads to lesser congestion in the Wi-Fi network and hence an improvement in handover latency is observed.

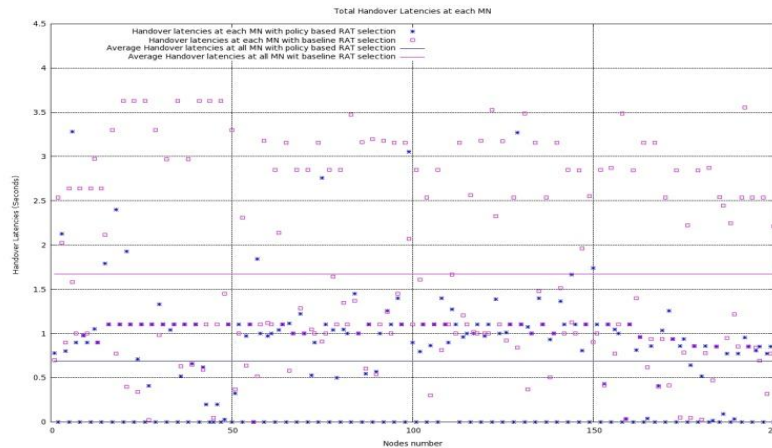


Fig. 7. Handover Latencies for PolicyBased and BaseLine RAT Selection

The performance efficiency of policy based RAT selection is also reflected through the comparison of drop rate of both strategies. Figure 8 shows the reduction in drop rate by 0.15%, hence a higher performance in successful reception of data packets. The improved performance in drop rate is also due to the controlled handover strategy used in the policy based RAT selection. As policy based RAT selection keeps the voice connection in UMTS network which means that less number of users shared the Wi-Fi network resources. Hence the Wi-Fi network observes fewer collisions which lead to higher number of successfully received packets and reduced number of dropped packets. This results in an improved performance of policy based RAT selection as shown in figure 8.

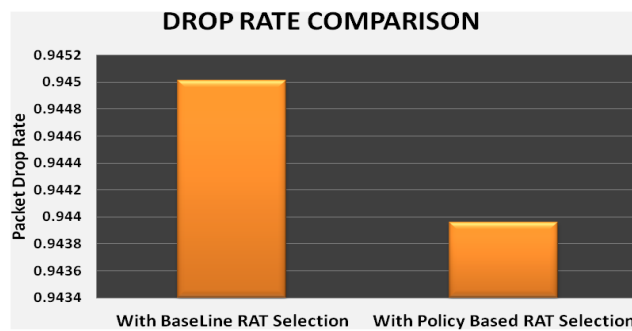


Fig. 8. Drop Rate of Base Line and Policy Based RAT Selection

7 Conclusion

In this paper, a policy based RAT selection by incorporating the IEEE 802.21 MIH framework has been presented to offload the cellular network traffic to Wi-Fi. The simulation has proved that the Wi-Fi networks can facilitate the cellular network traffic offloading and IEEE MIH framework provides the seamless offloading of UMTS network through Wi-Fi. The IEEE 802.21 framework has been optimised through the policy based RAT selection strategy. The proposed policy based RAT selection scheme has been compared with baseline i.e. ABC strategy and the new proposed, policy based RAT selection strategy proved to be better than the baseline strategy in terms of reduced handover latencies and reduced packet drop rate. Proposed RAT selection strategy reduced the number of handovers to large extent and the use of IEEE 802.21 MIH framework further improved the handover latencies observed by each MN.

References

1. Xuejun Zhuo; Wei Gao; Guohong Cao; Yiqi Dai, "Win-Coupon: An incentive framework for 3G traffic offloading," *Network Protocols (ICNP)*, 2011 19th IEEE International Conference, vol., no., pp.206-215,17-20.Oct.2011 doi: 10.1109/ICNP.2011.6089054.
2. (2009, January) Cisco visual networking index: Global mobile data traffic forecast update.
3. Dimatteo, S.; Pan Hui; Bo Han; Li, V.O.K.; , "Cellular Traffic Offloading through Wi-Fi Networks,"*Mobile Adhoc and Sensor Systems (MASS)*, 2011 IEEE 8th International Conference on , vol., no., pp.192 201,17-22Oct.2011doi:10.1109/MASS.2011.26
4. N. Laoutaris, G. Smaragdakis, R. Sundaram, and P. Rodriguez, "Delay-Tolerant Bulk Data Transfer on the Internet," in *ACM SIGMETRICS 2009*, June 2009.
5. B. Han, P. Hui, V. Kumar, M. V. Marathe, G. Pei, and A. Srinivasan, "Cellular Traffic Offloading through Opportunistic Communications: A Case Study," in *ACM CHANTS 2010*, Sep. 2010.
6. P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: socialbased forwarding in delay tolerant networks," in *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*, ser. *MobiHoc '08*. New York, NY, USA: ACM, 2008, pp. 241–250.
7. V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: A survey," *IEEE Communications Magazine*, vol. 46, no. 9, pp. 59–67, September 2008.
8. A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting Mobile 3G Using Wi-Fi," in *ACM MobiSys 2010*, Jun. 2010, pp. 209–222.
9. IEEE 802.21 Working Group, "IEEE Standard for Local and Metropolitan Area Networks , Part 21: Media Independent Handover Services", IEEE Std. 802.21-2008, Jan. 2009.
10. A. Dutta, S. Das, D. Famolari, Y. Ohba, K. Taniuchi, T. Kodama and H. Schulzrinne, "Seamless Handover across Heterogeneous Networks -An IEEE 802.21 Centric Approach". In *Proc. WPMC*, Aalborg, Denmark, September 2005.
11. Moon Kim, Tae-wook, Sung-joon Cho , "A study on IEEE 802.21 MIH frameworks in heterogeneous wireless networks," *11th International Conference on Advanced Communication Technology (ICACT)*, vol.01, pp.242-246, 15-18 Feb. 2009.
12. E. Gustafsson and A. Jonsson. Always best connected. *IEEE Wireless Commun. Mag.*, 10(1):49–55, February 2003.

PART NINE Trust-based Modelling of MANETS

A Recommendation-Based Trust Model for MANETs to Enhance Dynamic Recommender Selection Using Multiple Rules

Antesar M. Shabut^{1,2}, KeshavDahal³ and Irfan Awan²

1 University of Benghazi-Libya

2 University of Bradford

3 University of Glasgow

Abstract. Trust and reputation models are utilised by several researchers as one vital factor in the security mechanisms in MANETs to deal with selfish and misbehaving nodes and ensure packet delivery from source to destination. However, in the presence of new attacks, it is important to build a trust model to resist countermeasures related to propagation of dishonest recommendations, and aggregation which may easily degrade the effectiveness of using trust models in a hostile environment such as MANETs. However, dealing with dishonest recommendation attacks in MANETs remains an open and challenging area of research. In this work, we propose a dynamic selection algorithm to filter out recommendations in order to achieve resistance against certain existing attacks such as bad-mouthing and ballot-stuffing. The selection algorithm is based on three different rules: (i) majority rule based; (ii) personal experience based; and (iii) service reputation based. Recommendations are clustered, filtered, and selected based on these three rules in order to give the trust and reputation model greater robustness and accuracy over the dynamic and changeable MANET environment.

Keywords: trust metric, trust rating, trust management, mobile ad hoc networks



A Recommendation-Based Trust Model for MANETs to Enhance Dynamic Recommender Selection Using Multiple Rules

¹Antesar M.Shabut, ²Keshav Dahal, and ¹Irfan Awan

¹ School of Computing, Informatics and Media, University of Bradford, UK

²School of Computing, University of the West of Scotland, Glasgow, UK

HET-NETs 2013



Outlines

- INTRODUCTION
- RELATED WORK
- DISHONEST RECOMMENDATION PROBLEMS AND ATTACKS
- PROPOSED TRUST MODEL
 - TRUST COMPUTATION COMPONENT
 - RECOMMENDATION MANAGER COMPONENT
 - CLUSTER MANAGER COMPONENT
- SIMULATION AND RESULTS
- CONCLUSION & FUTURE WORK



Introduction

- Trust is the basis of all social interactions, for example buy or sell a product on ebay!
 - To cope with systems complexities and uncertainties
- Trust and reputation management has attracted lots of research interests
 - To gain trust among systems' entity
- Distributed systems like MANETs demand such these techniques
 - To reduce risk and guarantee the network activity



Continue ...

- Trust is one of the most approved mechanisms to improve security issues in MANETs [1]
 - To deal with misbehaving nodes and stimulate them to cooperate
- Trust in MANETs is the opinion held by one node (known as evaluating node) that can put on another (known as the evaluated node) using
 - Past behaviour observation and recommendations
- Dealing with dishonest recommendation is an open and challenging problem of trust models in MANETs
 - Recommender selection, propagation and aggregation of received recommendation



Related work

- Different trust and reputation schemes are proposed to enhance the security level of MANET by focusing on
 - Offering secure ad hoc routing (detecting selfish or malicious nodes), and
 - Encouraging collaboration among them
- Mostly, methods used to filter out recommendations are still unable to effectively mitigate the influence of dishonest recommendations [2, 3, 4] or no recommendation information is used [5, 6]
- MANETs with no central authority or hierarchical relationships need careful consideration of using recommendation's filter algorithms



Continue ...

- *Three different rules is used by the literature to filter out recommendation:*
- **Majority Rule [7]**
 - Used by most of the literature
 - Evaluates the honesty of the recommender by choosing the majority opinion between recommendations
 - Classify recommendations that deviate too much from the majority opinion as untrustworthy, and consequently exclude them from calculation
 - Can cause serious problems when recommenders collude with each other to accomplish a malicious goal



Continue ...

➤ **Personal Experience Based Rule [8]**

- Evaluates the honesty of the recommender by using the personal information gained by the evaluating node
- Aims to filter out any recommendation that is considered as incompatible with the opinion of the evaluating node
- Usually based on applying the deviation test to the receiving recommendations and excluding any that deviate too much from the opinion held by the evaluating nodes using a deviation threshold
- Can be unfruitful when the evaluating node has no prior experience with the evaluated node



Continue ...

➤ **Service Reputation Based Rule [9]**

- Assumes that there is a consistency between the trustworthiness of a node as a service provider and as a recommender
 - For example, the evaluating node gives more weight to recommendations received from highly reputed nodes for service providing such as packet forwarding and treats them as trustworthy recommenders
 - Can be harmful to the accuracy of the trust model because nodes can behave differently in providing different services
- *Due to the previous challenges, our model considers the combination of the three types of rules to come up with an integrated measure to address the problem of dishonest recommendations*



DISHONEST RECOMMENDATION PROBLEMS AND ATTACKS

- Recently, trust management models has been proved to be resistant against attacks related to forwarding packets such as blackhole, and wormhole attack [10]
- Attacks related to dishonest recommendation are still open and challenging problems in MANETs [11]
 - *Bad-Mouthing Attack*: In this kind of attack, the misbehaving node propagates a negative recommendation about good nodes in the network and consequently decreases their reputation ratings
 - *Ballot-Stuffing Attack*: This attack can be considered as a type of conflicting behaviour attack in which the attacker propagates positive recommendations about bad nodes in the network to obtain falsely high ratings
 - *Collusion Attack*: In this attack, more than one attacker working together to promote or degrade the trust value of a node



Proposed Trust Model

- We propose a recommendation-based trust management model to secure the routing protocol between source and destination nodes based on the trust value of each node in the path
- The model considers the problem of dishonest recommendations that might cause some of the attacks reviewed above which relate to misbehaviour issues in MANETs.
- The model combines three different rules: *majority rule*, *personal experience rule*, and *service rule* accompanied by a *clustering technique* to evaluate the recommender trustworthiness
- The model makes use of a Bayesian statistical approach and trust values are computed based on the assumption that they follow Beta Probability Distribution



Proposed Trust Model Components

➤ Trust Computation Component

- *Direct trust* T_{ij}^d : this value is obtained from two nodes that have already initiated a trust relationship

$$T_{ij}^d = \frac{\alpha_{ij}}{\alpha_{ij} + \beta_{ij}}$$

- *Indirect trust* T_{ij}^i : this value is needed when two nodes have not established a previous trust relationship through any kind of communication, or when there are few interactions between the two nodes

$$T_{ij}^i = \sum_{k=1}^n \frac{\alpha'_{kj}}{\alpha'_{kj} + \beta'_{kj}}$$

- Overall trust T_{ij} : this value is calculated by combining both direct and indirect trust values with different weights

$$T_{ij} = w_d * T_{ij}^d + w_i * T_{ij}^i$$

Continue ...

➤ Recommendation Manager Component

- Works as an intermediate component between the indirect trust computation component and the cluster manager component
- Helps in detecting and eliminating false recommendations and it has some important roles as in Algorithm 1.

Algorithm 1: *Recommendation Manager Algorithm*

1. **For** each recommendation request **Do**
 2. **Send** request to neighbours
 3. **Collect** received recommendation
 4. **Construct** $L = \{k_1, k_2, k_3, \dots, k_N\}$
 5. **Send** L to the cluster manager for processing
 6. **Receive** trustworthy cluster $C^{Trustworthy} = \{k_1^{Tr}, k_2^{Tr}, k_3^{Tr}, \dots, k_N^{Tr}\}$
 7. **Send** $C^{Trustworthy}$ to the requested node
 8. **End For**
-

Continue ...

➤ Cluster Manager Component

- Performed by the evaluating node on all the received recommendation
- The nearest neighbour clustering technique is applied on trust values provided by recommenders

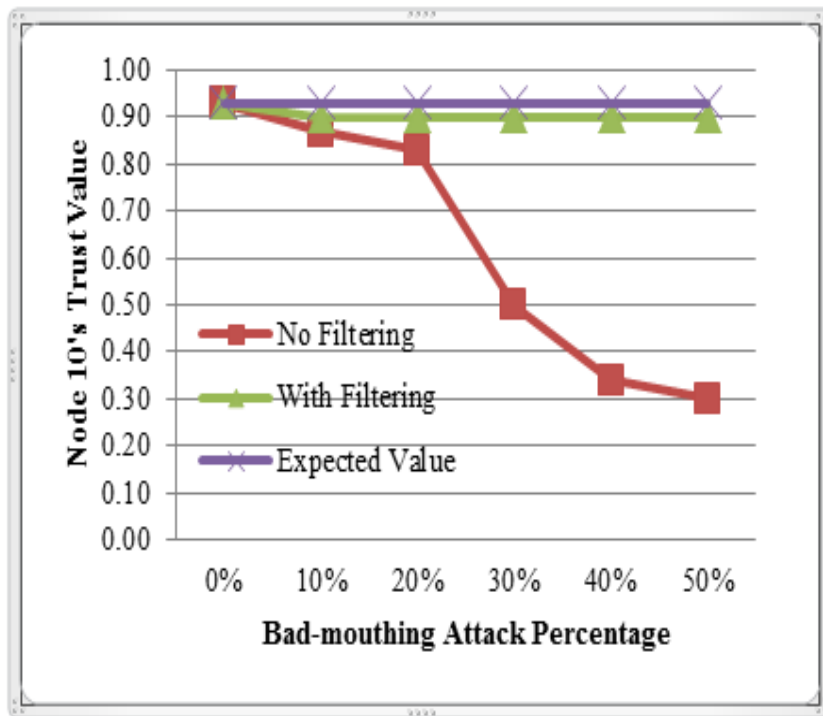
Algorithm 2: Cluster Manager Algorithm

1. **For** each recommendation list L **Do**
2. **For** each rating vector in the list (α^r, β^r) **Do**
3. **Calculate** trust value for the recommender as

$$T_{kj}^r = \frac{\alpha_{kj}^r}{\alpha_{kj}^r + \beta_{kj}^r}$$
4. **Calculate** deviation value as $d_{ik} = |T_{ij}^d - T_{kj}^r|$
5. **Calculate** the closeness value V_{ij}^{close} as Euclidean distance
6. **Weight** T_{kj}^r as $TW_{kj}^r = T_{kj}^r * w_{kj}$ based on d_{ik} value
7. **End For**
8. **Initialize** each vector as a unique cluster
9. **Repeat**
10. **For** each vector **Do**
11. **Merge** two clusters with the shortest Euclidean distance
12. **End For**
13. **Until** number of clusters = K
14. **For** each cluster C_i appeared in the previous iteration **Do**
15. **Calculate** the average of the deviation value d_{ik}^{Avg}
16. **If** $(d_{ik}^{Avg} \leq D)$ **Then**
17. **Merge** C_i and C_{i+1}
18. **End If**
19. **End For**
20. **Apply** the majority rule
21. **Select** trustworthy cluster with the highest TW_{kj}^r
22. **Return** trustworthy cluster $C^{Trustworthy}$
23. **End For**

Simulation and Results

➤ *The Effect of Bad-mouthing Attack*



- Increased percentage of attackers distorts the value of the trust value of node 10 held by other nodes in the network when there is no filtering algorithm incorporated
- The proposed filtering algorithm is able to keep the trust value close to the expected level even if half of the nodes in the network are regarded as bad-mouthing attackers

Figure 1. Good-Node: 10's Trust Value in the Presence of a Bad-mouthing Attack.

Continue ...

The Effect of Ballot-stuffing Attack

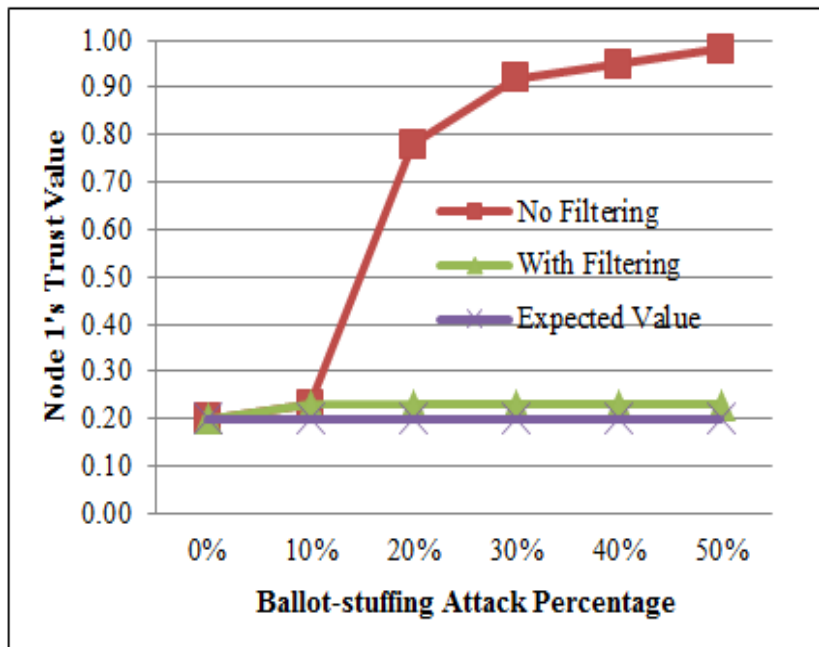


Figure 1. Bad-Node: 1's Trust Value in the Presence of Ballot-stuffing Attack

- Misleading of the decision made by other nodes to trust node 1 (*bad node in this case*) to more than 90% when the attackers increase to 50%
- The proposed model, which is equipped with a filtering algorithm can be more capable of mitigating the influence of dishonest recommenders



Conclusion

- A trust-based recommendation model for MANETs is investigated to cope with attacks related to recommendations using the dynamic recommender selection algorithm to filter out dishonest recommendations
- The model utilises the clustering technique accompanied with a deviation detection to filter out unfair recommendations exchanged by nodes in the network
- The proposed model has been tested by simulation against both bad-mouthing and ballot-stuffing attacks. The results of the simulation indicate that our model can safely incorporate honest recommendations received by recommenders and eliminate untrustworthy ones



Future Work

- Weighting recommenders based on time and location of receiving these recommendations, to mitigate the influence of location and time dependent attacks (recommending nodes differently according to time and location).
- Selecting trustworthy recommender is based not only on its trust value but also on other important factors like mobility, battery life, and closeness of recommender.
- Social network properties like friendship degree, closeness centrality and betweenness can be applied to enhance the recommender selection methods.



References

- [1] Li, W., J. Parker, et al. (2012). "Security through collaboration and trust in manets." Mobile Networks and Applications 17(3): 342-352.
- [2] Buchegger, S. and J. Y. Le Boudec (2004). "A Robust Reputation System for P2P and Mobile Ad hoc Networks P2P and Mobile Ad-hoc Networks."
- [3] Ganeriwal, S., L. K. Balzano, et al. (2008). "Reputation-based framework for high integrity sensor networks." ACM Transactions on Sensor Networks (TOSN) 4(3): 15.
- [4] Zakhary, S. R. and M. Radenkovic (2010). Reputation-based security protocol for manets in highly mobile disconnection-prone environments. Wireless On-demand Network Systems and Services (WONS), 2010 Seventh International Conference on, IEEE.
- [5] Pirzada, A. A. and C. McDonald (2004). Establishing trust in pure ad-hoc networks. Proceedings of the 27th Australasian conference on Computer science-Volume 26, Australian Computer Society, Inc.
- [6] Pissinou, N., T. Ghosh, et al. (2004). Collaborative trust-based secure routing in multihop ad hoc networks. NETWORKING 2004. Networking Technologies, Services, and Protocols;
- [7] Yu, H., S. Liu, et al. (2011). Dynamic witness selection for trustworthy distributed cooperative sensing in cognitive radio networks. Communication Technology (ICCT), 2011 IEEE 13th International Conference on, IEEE.
- [8] Buchegger, S. and J. Y. Le Boudec (2005). "Self-policing mobile ad hoc networks by reputation systems." Communications Magazine, IEEE 43(7): 101-107.
- [9] Zouridaki, C., B. L. Mark, et al. (2005). A quantitative trust establishment framework for reliable data packet delivery in MANETs. Proceedings of the 3rd ACM workshop on Security of ad hoc and sensor networks, ACM.
- [10] Cai, J., P. Yi, et al. (2010). An adaptive approach to detecting black and gray hole attacks in ad hoc network. Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on, IEEE.
- [11] Yu, Y., K. Li, et al. (2012). "Trust mechanisms in wireless sensor networks: Attack analysis and countermeasures." Journal of network and computer applications 35(3): 867-880.



Questions

Thank You

A Trust-Based Monitoring Model to Secure Routing Protocol in MANETs Using Enhanced Trust Metric

Antesar M.Shabut^{1,2}, KeshavDahal³ and Irfan Awan²

1 University of Benghazi-Libya

2 University of Bradford

3 University of Glasgow

Abstract. Mobile ad hoc networks suffer from several security deficiencies and vulnerabilities that have not yet been accurately solved, and consequently this can degrade the performance of such networks. We analyse and develop a trust-based monitoring model for mobile ad hoc networks (MANETs) using a probabilistic trust management technique based on beta distribution function. The goal is to establish a trust relationship between mobile nodes to efficiently allow nodes to acquaint with each other without previous interactions. We seek to utilise any available information in the network in constructing the trust relationships between any two nodes would like to interact. We make use of the beta distribution function to compute three types of trust values: direct, which come from the node's direct experiences, indirect, via other nodes which report about their own experiences, and opinion trust, which shows how honest node is as a recommender in the trust and reputation system.

Keywords: trust metric, trust rating, trust management, mobile ad hoc networks



A Trust-Based Monitoring Model to Secure Routing Protocol in MANETs Using Enhanced Trust Metric

¹Antesar M.Shabut, ²Keshav Dahal, and ¹Irfan Awan

¹ School of Computing, Informatics and Media, University of Bradford, UK

²School of Computing, University of the West of Scotland, Glasgow, UK

HET-NETs 2013



Outlines

- INTRODUCTION
- RELATED WORK
- COMPONENTS OF TRUST IN MANETS
- PROPOSED TRUST MODEL
- PROPOSED TRUST MODEL PROCEDURE
- SIMULATION AND RESULTS
- CONCLUSION & FUTURE WORK



Introduction

- Trust is the basis of all social interactions, for example buy or sell a product on ebay!
 - To cope with systems complexities and uncertainties.
- Trust and reputation management has attracted lots of research interests
 - To gain trust among systems' entity
- Distributed systems like MANETs demand such these techniques
 - To reduce risk and guarantee the network activity

Continue ...

- Trust is a multidimensional, complex, and context dependent concept
- Trust can be defined as *“the subjective probability by which an individual, A, expects that another individual, B, performs a given action on which its welfare depends”*

(Josang et al, 2007, “A survey of Trust and Reputation for online service provision”)

- reputation can be defined as “an expectation about an agent’s behaviour based on information about or observations of its past behaviour”

(Abdul-Rahman, 1997, “The PGP Trust Model”)



Continue ...

- Trust management system has been proved its applicability in enhancing security level in distributed systems includes
 - *e-commerce like ebay*
 - *social networks like facebook*
 - *peer-to-peer networks, and*
 - *mobile ad hoc networks*

- Trust management system can offer these systems
 - *quality,*
 - *reliability, and*
 - *trustworthy information*



Related work

- Different trust and reputation schemes are proposed to enhance the security level of MANET.
- The aim of trust and reputation management in MANETs have heavily focused on
 - Offering secure ad hoc routing (detecting selfish or malicious nodes), and
 - Encouraging collaboration among them
- MANETs with no central authority or hierarchical relationships need moving from hard evidence to soft evidence and credential
- Monitoring-based trust management models could be the proper solution



Continue ...

➤ **CONFIDANT mechanism:**

- Supports cooperation in MANETs by detecting and isolating malicious nodes
- Monitoring component runs by each node to monitor its neighbors
- Uses observation or reports about many types of attacks.
- Cannot prevent the dissemination of false information by malicious nodes, and
- Trustworthy nodes cannot be always truthful.
- Depends exclusively on negative ratings.

(Buchegger and Le Boudec, 2003, "A Robust Reputation System for Mobile Ad-hoc Networks")



Continue ...

➤ **Watchdog and path rater protocol**

- ***A watchdog***
 - complements the DSR protocol
 - runs on every node in the network to detect misbehaving nodes
- ***A path rater***
 - manages trust and routing policy, and
 - rates every path used in the network

- Do not exchange the reputation information with other nodes
- Cannot truthfully detect misbehaving nodes in circumstances like packet collisions and collusion of malicious nodes

(Marti et al, 2000, "Mitigating routing misbehavior in mobile ad hoc networks")

Continue ...

➤ **CORE Proposal:**

- A watchdog component complemented by a reputation system
- Distinguishes between three types of information
 - Observations,
 - Indirect reputation, and
 - Functional reputation
- Positive reputation is the only information that is exchanged by nodes.
- permitting the attacker to evade the reputation system such as spoofing attacks

(Michiardi and Molva, 2002, "CORE: a COllaborative Reputation mechanism to enforce node cooperation in mobile ad hoc networks")



Continue ...

- Trust and reputation management in MANETs face several challenges because of unique characteristics such:
 - *severe resource constraints,*
 - *the nature of the wireless medium,*
 - *mobility ,and*
 - *lack infrastructure*

- Most of the existing trust management mechanisms do not model trust in MANETs based on all these unique characteristics

- Therefore, the building of a dynamic monitoring-based trust management model for MANETs concerning their unique characteristics is demanded



Components of Trust in MANETs

- (a) Evidence manager to collect evidence about the behaviour of each node in the network,
- (b) Mathematical model to translate evidence into opinion, and then to use them to predict the behaviour of nodes in the future interactions, and
- (c) Policy manager to identify the decision rules and policies to enable nodes making decisions.



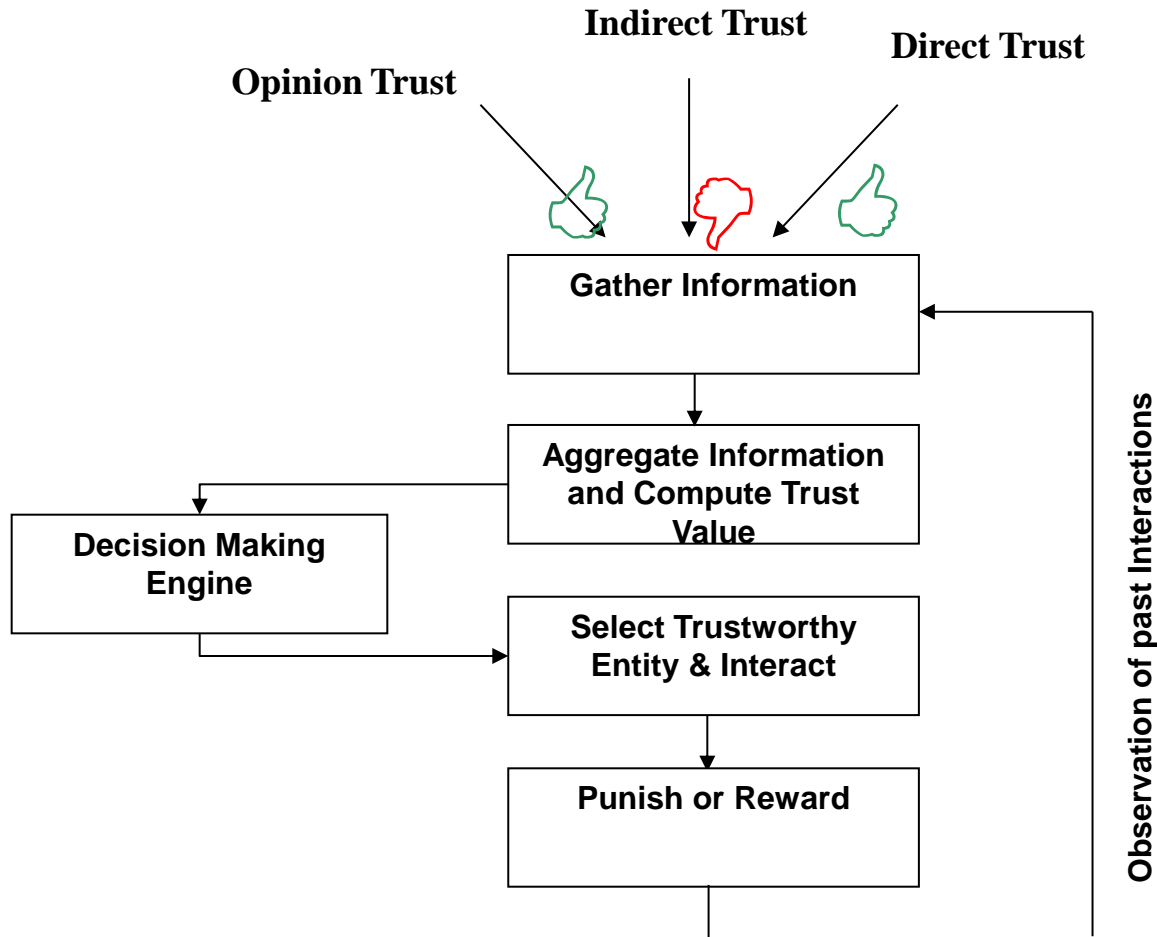
Proposed Trust Model

➤ Our main contribution

- First, making use of any available information in the network to construct the trust relationships between any two nodes would like to interact
- Second, adding another component known as opinion trust that expresses node's opinion about how honest node is as a recommender in the trust and reputation system
- This value is used in calculating the overall trust value of a particular node in addition to direct and indirect trust with various different weights.



Proposed Model Steps



Proposed Trust Model Components

➤ Direct Trust

- Obtained by immediate observations

$$T_{ij}^d = \frac{\alpha}{\alpha + \beta}$$

➤ Indirect Trust

- Obtained by recommendations from other nodes

$$T_{ij}^i = \frac{\alpha'}{\alpha' + \beta'}$$

➤ Opinion Trust

- Show how honest node is as a recommender in the trust system

$$T_{ij}^o = \frac{\alpha''}{\alpha'' + \beta''}$$

Continue ...

➤ Deviation Test

- Ensure the unity of view with the receiving node point of view

$$|T_{ji}^d - T_{ki}^i| \leq d$$

➤ Weighting Factor

- Give different weight to the different trust components

$$T_{ji} = T_{ji}^d + T_{ki}^i + T_{ji}^o$$

➤ Forgetting Factor

- Decrease the influence of past experience

$$s = s^{\text{old}} * \mu + s^{\text{new}}, \quad f = f^{\text{old}} * \mu + f^{\text{new}}$$

Simulation and Results

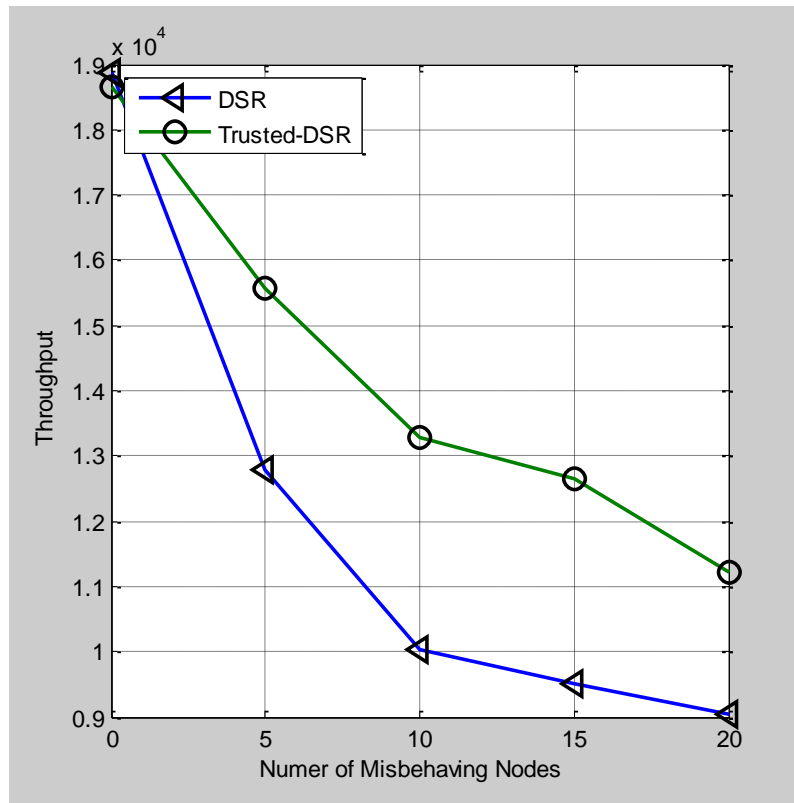


Figure 1. (a) Network throughput

- The figure1 (a) shows that the throughput for our trust mechanism gracefully drops as the number of misbehaving nodes increase, but it remains nearly 60%.
- While standard DSR falls below the 50% at the same time and same number of misbehaving nodes.
- Our trust mechanism enhanced the throughput by more than 10% than the standard DSR

Continue ...

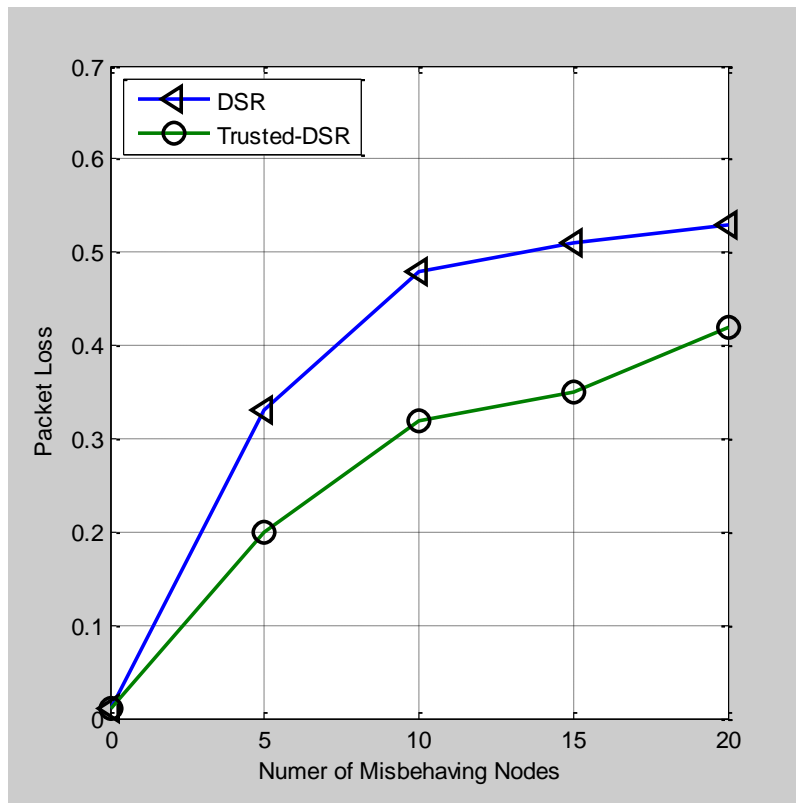


Figure 1. (b) Packet Loss

- Figure 1 (b) shows an improvement in the ratio of packet loss than the standard DSR.
- However, the results prove that our model performance exceeds the standard DSR in terms of packet delivery ratio or throughput.



Conclusion

- We analyse and develop a trust-based monitoring model for MANETs using probabilistic trust management technique based on beta distribution function to establish a trust relationship between mobile nodes to efficiently allow nodes to acquaint with each other without previous interactions.
- Our model is tested by simulation. The results prove that our model performance exceeds the standard DSR in terms of packet delivery ratio or throughput.



Future Work

- Using fuzzy logic or fuzzy inference to dynamically calculate weighting and forgetting factors instead of having constant values to adapt to network changing behaviours; and
- Applying the recent ideas about the cognitive networks to our model to benefit from the Cognition process used in such these networks to incorporate learning and reasoning;
- Constructing the trust relationship between nodes based on the concept of social network by increasing the robustness of this relationship by taking into account the common interests between nodes.



Questions

Thank You

PART TEN Performance Evaluation Algorithms for VANETs

Performance Evaluation of VANETs with Multiple Car Crashes in Different Traffic Conditions

Georgios Charalampopoulos^{1,2} and Tasos Dagiuklas¹

1. Dept. of Computer Science, Hellenic Open University, Greece,
Email: {gcharalampopoulos, dagiuklas}@eap.gr
2. Dept. of ECE, University of Patras, Greece, Email: gcharalampopoulos@eap.gr

Abstract. This paper studies the performance of VANETs with multiple car crashes under different traffic conditions. More specifically, a module has been designed and developed in a simulation environment in order to generate multiple vehicle crashes. It is shown that the VANET network statistics are affected by the node density and the number of vehicle crashes.

Keywords: VANET, Car Accident, VANET Simulation Tools

1 Introduction

There is a growing interest in deployment of warning services in Vehicular Ad-Hoc Networks (VANETs) for urban contexts allowing the communication among vehicles [1], [2]. These networks can be used for a variety of applications such as traffic management on motorways, emergency services and the prevention of collisions among vehicles. Research activities so far are mainly focused on approaches to reduce traffic congestion and provide general information services [3]. Several mobility models in VANET environment have been studied and simulated to extract useful information regarding different routing protocols, beaconing and delivery of messages etc.

The main goal of this paper is to design a module in a simulated environment allowing the creation of crashes/collisions among the vehicles. Important parameters of this car-crash module are the following: Node density, Departure Time - Arrival Time of each node.[4] The aim is to investigate the impact of the car crashes on the mobility models and the VANET network metrics. Moreover, we study the impact of the number vehicles involved in the collision on the performance of VANET. The paper structure is the following: VANET Mobility Modeling, Simulation Scenarios, Results and Conclusion.

2 VANET Mobility Modeling

For the VANET simulation scenarios, realistic mobility models in microscopic simulation (it takes into account road and node characteristics) have been used providing

road maps, the number of travelling cars and some road and car parameters such as maximum car speed, road limitations, departure and arrival times of each car [5]. Amongst others, the most common simulator environments for these mobility models are SUMO - Simulation of Urban MObility, VanetMobiSim and STRAW [6]. Without loss of generality, SUMO model has been used.

Node density is one of the factors that affect the mobility models and the vehicular networks used for the simulation. The different node densities are used to represent the different traffic conditions for peak and off peak hours. A parser has been created to calculate the road network that will be used in accordance with the street map to carry out the simulations.

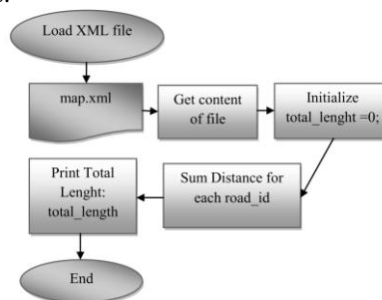


Fig. 1. Process of Road Network Calculator Parser

The figure above shows the process of calculating the road network using the XML map file that has been generated by SUMO. The road network is calculated from the sum of the distances of all the road-IDs and the distance of the road network. Then, using node density [7], and the road network, the process calculates the node density for the simulation scenarios that will be presented below.

Vehicle crashes are another important factor that has been influenced by the mobility model and the node density. When there are accidents in VANET, the vehicles around the point of interest (crash event) are mainly influenced by stopping and/or changing direction to reach the final destination using the Dynamic Re-route provided by SUMO. [6] Furthermore, the network infrastructure is influenced by the number of nodes, because emergency messages from crashed vehicles are sent after the accident in order to inform drivers and emergency response teams such as police vehicles, ambulances etc when required.

To find vehicles in a simulated environment that can crash among each other, another parser written in JAVA has been created. This parser initially selects a node randomly and then searches for other nodes that have common edges on their fixed routes with the first node and are close with it in distance. As an effect, a vehicle crash is created. The picture below shows the process of finding crashes of vehicles. More specifically, the parser uses XML route file, which is loaded for use in the simulation. Then the parser searches for the previously mentioned nodes in order to create the accident. Otherwise the first random selected node is referred as a broken-down vehicle. For

multiple car crashes, the JAVA parser will be compiled several times to extract more than one crash.

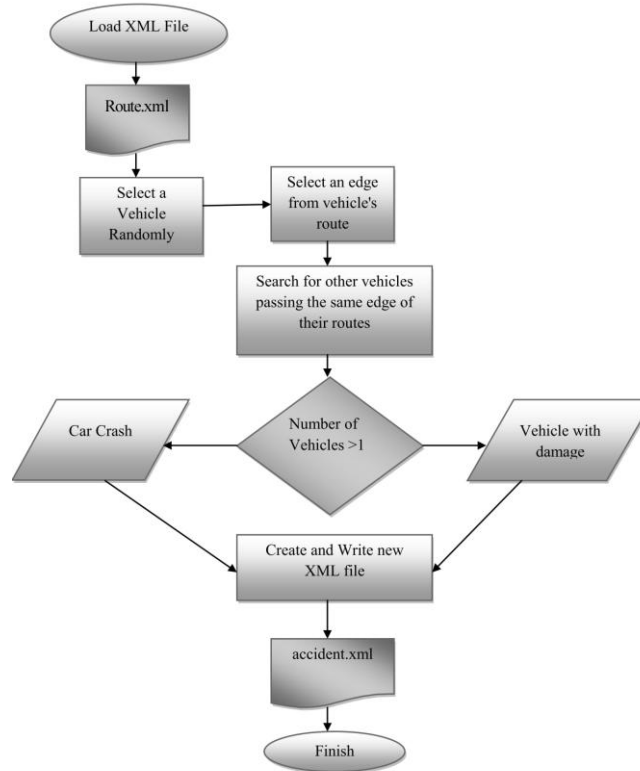


Fig. 2. Process of Car Crashes Parser

3 Simulation Scenarios

For Vehicle Crash simulation, Veins framework [8] has been selected because it integrates the OMNeT++ and SUMO [9] and can offer online re-configuration and re-routing of cars in reaction to the network simulator, IEEE 802.11p and IEEE 1609.4 DSRC/WAVE network layers, supporting realistic maps and traffics. The IEEE 1609.4 defines the multi-channel and QoS operation of radios; vehicles with a single radio will periodically switch among multiple channels. [10]

Firstly, a realistic map and the road network are imported from OpenStreetMap.com [11], the routes of the cars are generated using SUMO and then exported to the network simulator. OMNeT++ considers all the cars as nodes and simulates the scenario. If there are changes in the network, Veins can modify the scenario in SUMO and the realistic traffic of the city center is simulated as a VANET in OMNeT++. Simulation scenario is done using the configuration, dataset and parameters for 2000 seconds. About 667 cars for peak and 383 cars for off peak hours travel in the map at this time

and each scenario runs the simulation 20 times starting from zero. Three different simulation scenarios have been created with one, two and three car crashes. Each scenario has been replicated 20 times and for each scenario a different start time has been used for each node. The following table shows network parameters for our simulations.

Beaconing Rate	10 Hz	CWmin	15
Max Transmission Power	20 mW	CWmax	1023
Thermal Noise	-110dBm	CCH interval	50
Header Length	256 bit	Data Rate	18Mbps
Slot Duration	16 s	Dedicated Frequency Range	5.9 GHz

Table 1. Network Simulator Parameters

4 Results

The results from the simulations scenarios will be presented below. Especially, Received Broadcast and Total Lost packets, Busy Time and Total CO2 Emission will be discussed to check how these metrics are affected by the different number of crashes.

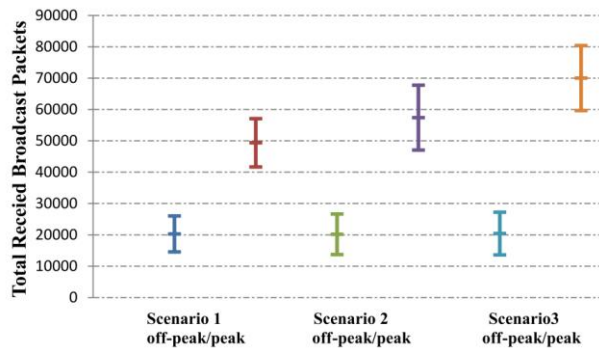


Fig. 3. Total Received Broadcast Packets

The first graph shows the Received Broadcast Packets and nodes and that the different number of vehicles influences the network infrastructure. The value of Received Broadcast Packets is higher for peak than the off peak hours. Moreover, the number of crashes increases the number of Received Broadcast packets, which is due to the fact that more and more nodes close to the area of interest must be updated after the crash event. The graph below shows that different number of vehicles impact the number of Total Lost packets because vehicles collect more information when more than one crash has been occurred. For the peak hours the number of these packets is larger than that to the off peak hours.

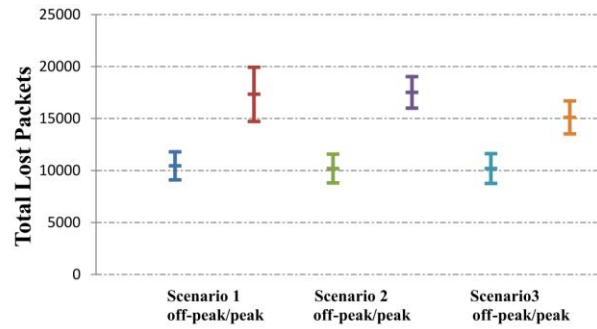


Fig. 4. Total Lost Packets

Busy Time is affected by the different number of vehicles and car crashes. When the number of car crashes increases, the Busy Time increases too.

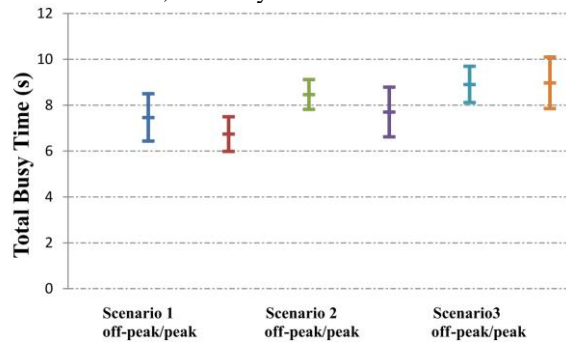


Fig. 5. Total Busy Time in seconds

The final graph shows that the Total CO2 Emission per vehicle. As expected, this parameter increases when the number of accidents and node density increase because many vehicles stop and wait during their route.

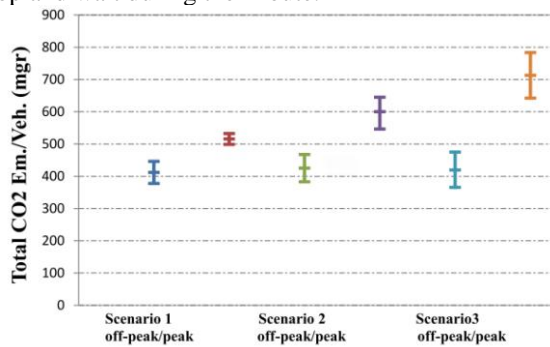


Fig. 6. Total CO2 Em. / Vehicle in mgr

The results show that the multiple car crashes and the different number of vehicles in the same area are highly affecting the mobility model and the performance of the network infrastructure.

5 Conclusion

This paper has studied how multiple car crashes can be generated in a VANET simulation environment. We have investigated the impact of node density and number of car crashes on VANET performance evaluation. Realistic scenarios in an urban area with real vehicle routes have been studied. The future work includes simulations of multiple car crashes using Vehicle-to-Infrastructure (V2I) communication and comparative study with V2V communication and implementation of other applications regarding the warning services of VANETs.

Acknowledgement. This work was supported by the *SALUS* project (grant number: 313296), funded by the EC FP7 *ICT-security* collaborative research

6 References

1. Sebastian Grafling, Petri Mahonen, and Janne Riihijarvi. Performance Evaluation of IEEE 1609 WAVE and IEEE 802.11p for Vehicular Communications. In International Conference on Ubiquitous and Future Networks, pages 344-348, 2010.
2. H. Conceicao, L. Damas, M. Ferreira, and J. Barros, "Large-scale simulation of V2V environments," in Proceedings of the 2008 ACM symposium on Applied computing. ACM, 2008, pp. 2833.
3. C. Manasseh and R. Sengupta. Middleware to enhance mobile communications for road safety and traffic mobility applications. IET Intelligent Transport Systems, 4(1):24, 2010.
4. Harri, J.; Filali, F.; Bonnet, C., "Mobility models for vehicular ad hoc networks: a survey and taxonomy," Communications Surveys & Tutorials, IEEE, vol.11, no.4, pp.19-41, Fourth Quarter 2009.
5. Karnadi, F.K.; Zhi Hai Mo; Kun-chan Lan; , "Rapid Generation of Realistic Mobility Models for VANET," Wireless Communications and Networking Conference, 2007.WCNC 2007. IEEE, vol., no., pp.2506- 2511, 11-15 March 2007.
6. SUMO, Simulation of Urban MObility <http://sumo.sourceforge.net/>
7. Clean Air Initiative, <http://cleanairinitiative.org/portal/greece>
8. Veins, The open source vehicular network simulation framework, <http://veins.car2x.org>
9. A. Kopke, M. Swigulski, K. Wessel, D. Willkomm, PT Haneveld, TEV Parker, OWVisser, HS Lichte, and S. Valentin. "Simulating wireless and mobile networks in OMNeT++ the MiXiM vision." In Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops, pages 1-8. ICST, 2008.
10. IEEE Draft Standard for Wireless Access in Vehicular Environments (WAVE) - Multi-Channel Operation," IEEE P1609.41D9, August 2010, pp. 1 -61, 2010.
11. OpenStreetMap, <http://www.openstreetmap.org>

Context-aware Rate Adaptation Algorithm for DSRC Vehicular Networks

Ayoade Ilori¹, Zuoyin Tang¹, Yue Li^{2,3}, Jianhua He^{1*}

¹School of Engineering and Science, Aston University, Birmingham, UK.

Email: ({iloriao, z.tang1, j.he7@aston.ac.uk); *Corresponding Author.

²Research Innovation, Swansea University, UK (y.le@swansea.ac.uk);

³J&Y Co Ltd, Swansea, UK (y.le@swansea.ac.uk);

Abstract—Vehicular networks are heavily subjected to rapid changes in environmental conditions that affect their performance. This, coupled with the mobility of nodes involved makes rate determination challenging for this kind of network. Rate adaptation algorithms based on static networks exist but they do not perform well in vehicular environments. However, contextual information is quite readily available and with some modifications, the performance of these algorithms could be enhanced. Considering two of such algorithms, the AARF and ONOE in a vehicular environment, modifications were made to the ONOE algorithm by utilizing the context information. Simulation results were carried out using the IEEE 802.11p standard. Results show that the performance of the modified ONOE algorithm is significantly better compared to the other algorithms under consideration. Factors that could affect the performance of any rate adaptation algorithm were also identified and proven by simulation results.

I. INTRODUCTION

The application of technology today now spans every area of human endeavors from homes to offices. Key concerns to the world at large are the safety of human lives and environment pollution especially during transportation. Intelligent transport systems (ITS) have emerged to provide unconventional services unique to the different modes of transport and also to manage traffic with the aims of guaranteeing the safety of human lives during transportation and reduce pollution.

Vehicular Ad-Hoc Networks (VANETs) is a subset of Mobile Ad-Hoc Networks (MANETs) and offers an exciting advance for future ITS applications. Dedicated Short Range Communications (DSRC) is specifically designed for automotive use and offers a set of standards and protocols for bidirectional or unidirectional short to medium range wireless communications[1]. Operating at the 5.725-5.825MHz of the Industrial Scientific and Medical (ISM) band, DSRC systems are made up of road side unit (RSU) and on board units (OBU) with transponders and transceivers [2].

The IEEE 802.11p standard defines enhancements to 802.11 networks to support Intelligent Transport System (ITS)

applications [3]. Also, vehicular networks have started finding applications in toll collections, broadcast of red light duration at traffic lights and hotspots for transferring and downloading maps, just to mention a few.

The performance of networks could be characterized by the number of bits or packets of data that could be delivered successfully within a given period. Wireless networks have the capability to transmit data at multiple data rates and it is expected that transmission would be carried out at an optimum rate depending on existing channel conditions. As 802.11 wireless networks are mainly based on collision avoidance for random channel access rather than collision detection as in Ethernet networks meaning packet loss could be due to collision or unfavorable channel conditions. Rate adaptation algorithms have long existed and aided wireless nodes in the determination of the optimum transmission rate based on existing channel conditions. Although, rate adaptation has not been specified by the 802.11 standards, it remains a crucial factor in the determination of the performance of a wireless network [4]. Channel conditions have been one of the major determinants of the overall throughput of any wireless network.

In this paper two rate adaptation algorithms have been selected for study and performance improvement of rate adaptation algorithms for DSRC vehicle networks, the Adaptive Auto Rate Fall Back (AARF) and the ONOE algorithms [5], [6]. System level simulations were carried out on wireless nodes in vehicular environments using the two representative algorithms which were hitherto developed for static networks. This provides a platform to justify which factors affect the performance of vehicular networks and also tries to improve on performance by modifying parameters in the representative algorithms. Simulations serve as a great time and cost saver, making it easy to modify variables that would have hitherto been hard or time consuming to modify in the real world scenario.

The primary contributions of this paper are:

- Development of a system level simulator for DSRC based wireless networks in vehicular environments.
- Evaluation of the performance of the two representative rate adaptation algorithms through simulation in vehicular environments.
- Improvement in the network performance by

modification of existing algorithm.

The rest of this paper is organized as follows. We establish some background to this study in section II and review existing rate adaptation algorithms in section III. Considerations made during the course of this study are discussed in section IV. Simulation results are presented in section V. The conclusion and discussion of future works is done in section VI.

II. BACKGROUND

The main function of the MAC layer is the provision of access functions such as access coordination, frame check sequence generation and addressing, for shared medium physical layers in support of the link layer control (LLC) sub layer. The carrier sense multiple access with collision avoidance (CSMA/CA) is used in 802.11 standards because collisions are undetectable in wireless LANs. In CSMA/CA, the stations listen to the medium to see if any packet is not being sent on the medium (carrier sensing)[7]. If the medium is idle, prospective sending stations wait for some predetermined period of time before proceeding to send their packets.

IEEE 802.11 standards make use of CSMA/CA with positive acknowledgement and retransmission. MAC allows for interoperability between two different physical layers on different hosts by using this CSMA/CA protocol and a random back-off timer. The MAC layer also describes the way beacon frames are sent by access points. A beacon frame enables stations monitor the presence of an access point while a management frame allows stations to actively scan for the presence of access points on any available channel. Stations choose the best suited access point based on these.

802.11 MAC also defines special functional behavior for packet fragmentation, medium reservation via request to send (RTS) / clear to send (CTS) polling interaction and point coordination (used mainly by time bound services) [8]. The MAC sub-layer, link layer control (LLC), is responsible for channel allocation procedures, frame formatting, error checking, as well as, fragmentation and re-assembly of packets.

The wireless transmission media can operate between an exclusively contention mode and it can also alternate between contention mode and contention free mode. In contention mode, all stations have equal rights to transmit and they compete for access to the channel while in contention free mode, channel access is controlled by the access point.

III. REVIEW OF EXISTING RATE ADAPTATION ALGORITHMS

A. Adaptive Auto Rate Fallback (AARF)

AARF is an improvement over the earliest rate adaptation algorithm called the Auto Rate Fallback (ARF) [5]. The AARF is based on the ARF and is only a modification of it in order to increase its efficiency in terms of changes in data rate whilst in use. AARF starts at the basic data rate (depending on the IEEE 802.11 standard being used, for example, 802.11b starts at 2Mbps) and then starts a timer. Either at the expiration of the timer or a number of successful transmission, called threshold, the data rate is increased.

The first transmission at the new data rate is crucial. If this transmission fails, the data rate is reverted to the old one and the threshold is doubled. If the first transmission at the new data rate succeeds the threshold is reset and transmission continues in both cases. With this, AARF is simple and able to stabilize data rate changes over a transmission period. This makes AARF produce fewer fluctuations in rate changes decisions when compared with ARF

B. ONOE

This was developed by the MadWifi organization for wireless adapters with Atheros chips [6]. It is the earliest implemented open source rate adaptation algorithm and it operates on a Linux driver. The Onoe algorithm is a credit based system that tries to find the best data rate with a loss ratio of not up to 50% [9]. At every one second interval, transmission statistics are collected and based on these statistics a decision is made on whether to increase the current rate or reduce it. In 802.11a and 802.11g Onoe starts transmission at 24 megabits while it starts transmission at 11 megabits for 802.11b. A flow chart describing the operations of the ONOE algorithm is shown in figure 2.

Onoe is relatively conservative and insensitive to bursty losses. Once Onoe decides that a particular data rate would not work, it does not try it again until after 10seconds, thus, making it slow in responding to fast changes in wireless conditions. Other rate adaption algorithms exists, e.g. SampleRate [10], Receiver Based Auto-Rate [4][11] and the context aware rate selection algorithm [12]. Study of the other algorithms is left for our future work.

IV. SIMULATION CONSIDERATIONS

In this section, the considerations made during the course of this work are discussed.

A. Simulation Configuration

Simulations were carried out based on assumption that the nodes were moving with a constant speed and a road side unit (access point) covering a range of distance through which the each nodes passes. There is the possibility of having more than one node within the range at any point in time and expectedly transmission is only done when these nodes are within the range of the access point. It is also worthy of note that most of our measurements are based on the 802.11a standard and this is because whilst this standard makes our simulation a little less complex, the 802.11p MAC and PHY standards are similar to the already established 802.11a standard. Also, the modulation and coding schemes as well as training sequence of both standards are the same [13]. Thus, this does not invalidate the result of our simulation in any way.

For a station to transmit, it has to follow the truncated binary exponential backoff procedure specified in the 802.11 standard. In the simulation the node(s) with the lowest value of contention window, which is generated randomly and every node in the system has the right to transmit so far it is the node with the smallest contention window size. In a case whereby more than one node with the same value of contention window, a collision occurs and the process is restarted.

During the transmission of any packet there are three possible outcomes and these are:

- A packet could be transmitted successfully, in which case an acknowledgement packet (ACK) is sent back to the sender.
- A packet may suffer from collision and in which case it would need to be retransmitted.
- A packet might also be considered unacceptable due to bit errors in which case it is discarded.

The effect of access point coverage distance is considered by varying the communication range. Based on the fact that the value of SNR changes with respect to distance and speed of vehicle, the formula in (1) has been used to calculate SNR value based on a distance power gradient of 2.

$$SNR = P_{tx}/distance^2 \quad (1)$$

Where P_{tx} is transmitted power measured in watts. The SNR of a node at any particular point in time is also dependent on its position (distance from the access point). The distance at any point in time is a function of vehicular speed and time. The amount of data that a receiver can successfully get over a link is based on the number of bits it can decode correctly. Bit error rate (BER) is a ratio of the number of bits received in error to the total number of bits transmitted over the link.

In wireless medium, bits are usually grouped into packets and this allows for sharing of the transmission media by the nodes in the wireless system. Similar to the bit error rate, the packet error rate (PER) is the ratio of incorrectly received packet by a receiver to the total number of packet that was sent by a transmitter. A packet is deemed to contain errors if at least one bit in the packet is an error. Thus, mathematically, the packet error probability for a packet containing M bits by the equation in (2)

$$P_p = 1 - (1 - P_e)^M \quad (2)$$

Where P_p is the packet error probability, P_e is the packet error rate and M the number of bits in the packet. Similarly, the throughput of a wireless system is also expressed mathematically as given in (3):

$$Throughput = (1 - BER)^M * data\ rate \quad (3)$$

Since throughput is measured in the number of successfully transmitted bits per second, BER and the data rate used for the transmission are very crucial to its determination, hence, the equation above.

Thus, for wireless simulation of which throughput is a major output to be considered, it follows that the determination of an acceptable estimate for bit errors in channel conditions and varying data rates be given. This simulation makes use of an SNR-PER plot which estimates channel errors at various data rates. The plot used takes SNR and data rate as input and gives a packet error rate as output. The plot is based on raw signal to noise and bit error rate data for 802.11a modes with channel 'a' modes.

Both rate adaptation algorithms were coded based on the algorithms described in [3] and [4] respectively, taking all factors considered in the algorithm in consideration for implementation and then measurements were made to compare the performance of both rate adaptation algorithms.

The table below summarizes some of the major parameters used in the simulation this work.

PARAMETER	VALUE
DIFS time	34 μ s
SIFS time	16 μ s
Slot time	9 μ s
Header packet Length	464 bits
ACK packet length	304 bits
Contention window size	32 bits
Average packet length	2000 bits
Node Velocity	4.5 m/s
Power distance gradient	2
Normalized Power	20mW

Table 1: Simulations Parameters

B. Modifications to Onoe

In an attempt to increase the system performance, some modifications to the original ONOE algorithm were made by exploiting the context information of the vehicle communication environment. The modification can make it suitable for use in vehicular networks with coverage distance of about 250m with a limited number of nodes. Basically, the changes made to the ONOE code are listed below:

- Reduction in rate increment/decrement threshold to 0.5 from 1 due to the mobility of the vehicles.
- Reduction in credits threshold to 3 from 10.
- Increase of start rate to 54Mbps.

The major reasons for these changes are that considering the fact that our nodes are now moving at velocities, a choice rate adaptation algorithm should be able to adapt to fast changing channel conditions and also be able to make rate changes as fast as possible so as to optimize system throughput. Transmitting at the highest rate possible gives the opportunity to send data quickly in face of rapidly changing channel conditions. Also, logically at a closer distance to the access point and with higher transmission rate, a higher throughput should be obtainable. The effect of the modification was studied whilst gradually increasing the coverage range (distance) of the access point as explained in the previous sections.

V. SIMULATION RESULTS

The performance of both rate adaptation algorithms and the modifications to ONOE, with access point (AP) coverage distance of 200m and 250m and nodes moving at a constant speed of 4.5m/s was evaluated. It should be noted here that the SNR values of transmitting nodes at any point in time during simulations changes with distance and this has already been discussed in the simulation configuration section. Nodes around the edges are certain to receive lower signal strength in comparison to those that are closer to the access point. A decrease in throughput of about 67% was observed at both distances under maximum load of eleven nodes.

Also, a reduction in the value of throughput was observed using the ONOE algorithm at distances of 200m and 250m respectively. When compared to the previous scenario, a reduction in throughput of about 48% at 250m coverage distance was observed.

In both cases, the effect of mobility on throughput has been seen and whilst ONOE still performs better than AARF, modifications were made to it with a view to increasing its performance.

A. Performance at 200m and 250m

The performance of the both AARF, ONOE and the modified algorithms are shown in the following graphs.

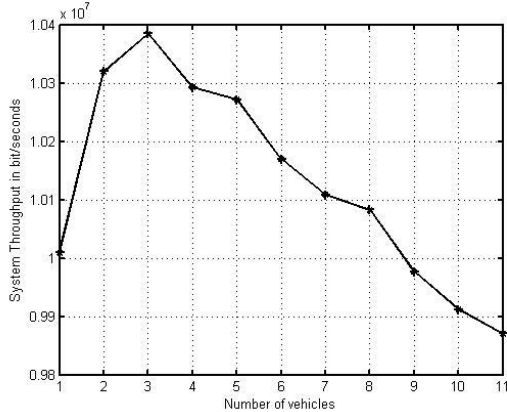


Fig. 1. Performance of ONOE with AP coverage of 200m

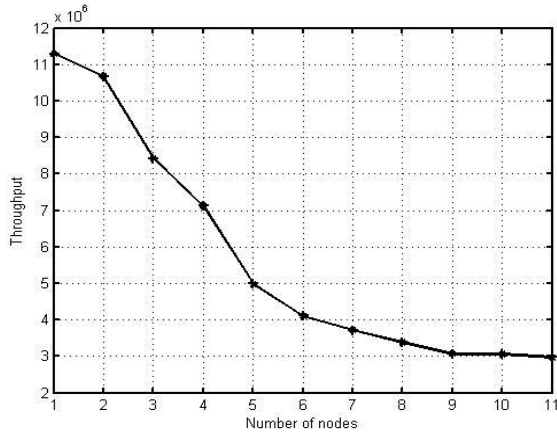


Fig. 2. Performance of AARF with AP coverage of 200m

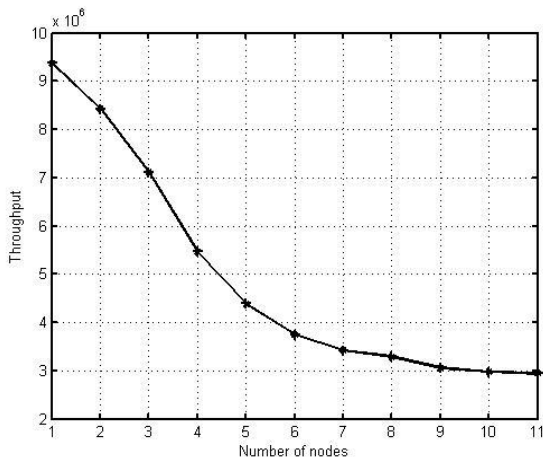


Fig. 3. Performance of AARF with AP coverage of 250m

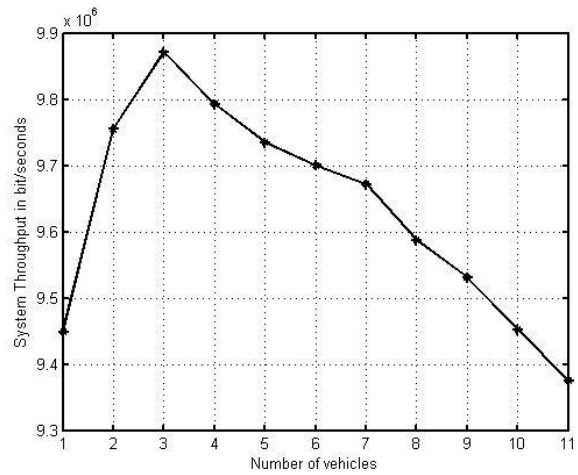


Fig. 4. Performance of ONOE with AP coverage of 250m

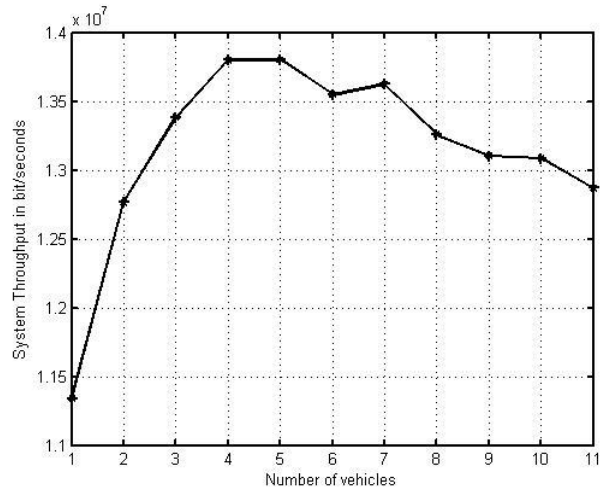


Fig. 5. Performance of modified ONOE with coverage 200m.

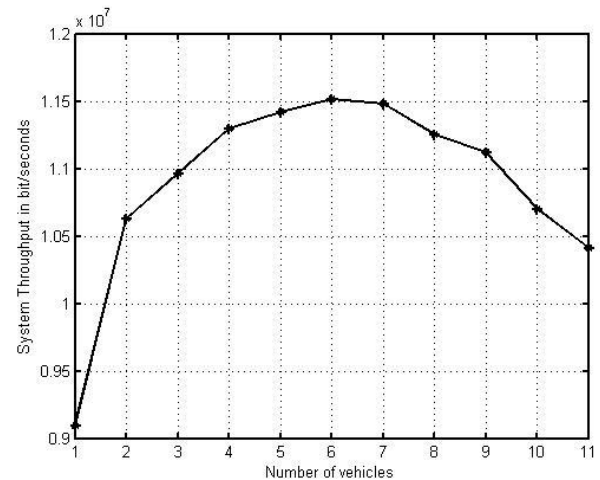


Fig. 6. Performance of modified ONOE with coverage 250m.

As can be deduced from the graphs above, with a gradual increase in antenna coverage distance, there is a reduction in overall throughput of the system and expectedly so, as the farther away a node is from the access point, the higher the probability of failure of transmission due to an extremely unstable environmental (channel) condition. Thus, the

performances of the various rate adaptation algorithms considered in this paper can be summed up in Table.3.

Algorithm	Coverage Distance (in meters)	Node Speed (in meters/second)	Throughput with 11 nodes
AARF	200	4.5	3.0 Mbps
AARF	250	4.5	3.0 Mbps
ONOE	250	4.5	9.35 Mbps
ONOE	200	4.5	9.85 Mbps
Modified ONOE	200	4.5	13 Mbps
Modified ONOE	250	4.5	10.5 Mbps

Table 3: Summary of result of simulations

Performance increases of about 30% at an access point range of 200m and 12% at an access point range of 250m have been observed when comparing the throughput of the ONOE and modified ONOE algorithms.

VI. CONCLUSION

The focus of this work has been to study the performance of existing rate adaptation algorithms on vehicular networks and propose an improvement to the existing algorithm based on the result(s) gathered. The performance of existing rate adaptation algorithms was evaluated in a vehicular environment. The more effective algorithm (in a static environment) was modified by utilizing context information with the aim of improving its throughput.

Several factors play important roles in the determination of throughput of any vehicular network and these include the speed at which the node (vehicle) is moving, the coverage distance of the antenna and channel conditions as at the time of transmission. There has been considerable improvement in throughput of vehicular networks as proven by our modified ONOE algorithm. Investigations have also been carried out to check the effect of distance on our modified algorithm and as such it was discovered that the algorithm would only work effectively with small coverage distance and this is due to the fact that at closer distances to the antenna, the probability of successful delivery of a packet is higher than when compared to that of a transmitting node that is quite far off.

Also, considering the fact that our coverage distance is quite small, it might not be too effective to transmit huge packets. Thus, it might be unsuitable for applications such as real-time video streaming. This is because since nodes are expected to be in continuous motion, the time in which each node spends in the access range would be small compared to the time it would take for a transmission of huge packet sizes and once the node is out of the range of the access point, the transmission would be interrupted.

In furtherance of this study, work is ongoing on the development of a rate adaptation algorithm that would be efficient with a vehicular environment considering the performance of recent rate adaptation algorithms, like the CARS [12].

ACKNOWLEDGEMENT

The work is partially supported by the UK Engineering and Physical Sciences Research Council (EPSRC) with grant reference number EP/1010157/1.

REFERENCES

- [1] "Dedicated Short Range Communications." [Online]. Available: <http://www.its.dot.gov/DSRC/>.
- [2] "DSRC." [Online]. Available: <http://www.etsi.org/index.php/technologies-clusters/technologies/intelligent-transport/dsrc>.
- [3] IEEE, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," in *IEEE Std 802.11-2007 (Revision of IEEE Std 802.11- ..., 2009*.
- [4] S. H. Y. Wong, H. Yang, S. Lu, and V. Bharghavan, "Robust Rate Adaptation for 802 . 11 Wireless Networks," *Mobicom*, pp. 146–157, 2006.
- [5] M. Lacage, M. H. Manshaei, T. Turletti, and S. Antipolis, "IEEE 802 . 11 Rate Adaptation: A Practical Approach," 2004.
- [6] S. Biaz and S. Wu, "Rate Adaptation Algorithms for IEEE 802 . 11 Networks : A Survey and Comparison," pp. 130–136, 2008.
- [7] A. R. Prasad and N. R. Prasad, *WLANs and IP Networking: Security, QoS, and mobility*. Artech House, London, 2005.
- [8] R. P. and L. J., *802.11 Wireless LAN Fundamentals*. Cisco Press, 2004.
- [9] MadWifi, "The MadWifi Project." [Online]. Available: <http://sourceforge.net/projects/madwifi>. [Accessed: 09-Sep-2013].
- [10] J. C. Bicket, "Bit-rate Selection in Wireless Networks," 2005.
- [11] G. Holland, N. Vaidya, and P. Bahl, "A rate-adaptive MAC protocol for multi-hop wireless networks," ... *Mob. Comput. Netw.*, no. July, 2001.
- [12] P. Shankar, T. Nadeem, J. Rosca, and L. Iftode, "CARS: Context-Aware Rate Selection for vehicular networks," *2008 IEEE Int. Conf. Netw. Protoc.*, pp. 1–12, Oct. 2008.
- [13] C. B., "Rate Control Algorithms for IEEE 802.11 wireless networks," The University of Texas, Arlington, 2007.

Vehicular Ad Hoc Networks (VANETs), Past Present and Future: A survey

Farhan Aadil, Shahzad Rizwan, Adeel Akram

farhan.aadil, Shahzad.rizwan [at]ciit-attock.edu.pk],
adeel.akram@uettaxila.edu.pk

Abstract. Vehicular Ad Hoc Networks (VANETs) are classified as a subset of Mobile Ad Hoc Networks (MANETs). VANETs have the potential of improving road safety and providing travelers comfort. This network rehearses moving cars as nodes in a network to create a mobile network. As a component and concrete application of an ITS inter vehicle communication gained attention from researchers, academics and industry leaders, especially in US, EU and Japan. Intelligent Transportation System (ITS) enables coordinated traffic management, like advanced traveler information services, Vehicle tracking and Autonomous vehicle safety audits etc. In this research paper we tried to pre-sent a thorough survey of VANETs.

Keywords: Vehicular Ad Hoc Network, Mobile Ad Hoc Network, Intelligent Transportation System

Vehicular Ad Hoc Networks (VANETs), Past Present and Future: A survey

Farhan Aadil¹, Shahzad Rizwan², Adeel Akram³

¹farhan.aadil, ²Shahzad.rizwan [@ciit-attock.edu.pk],
³adeel.akram@uettaxila.edu.pk

Abstract. Vehicular Ad Hoc Networks (VANETs) are classified as a subset of Mobile Ad Hoc Networks (MANETs). VANETs have the potential of improving road safety and providing travelers comfort. This network rehearses moving cars as nodes in a network to create a mobile network. As a component and concrete application of an ITS inter vehicle communication gained attention from researchers, academics and industry leaders, especially in US, EU and Japan [7]. Intelligent Transportation System (ITS) enables coordinated traffic management, like advanced traveler information services, Vehicle tracking and Autonomous vehicle safety audits etc. In this research paper we tried to present a thorough survey of VANETs.

Keywords: Vehicular Ad Hoc Network, Mobile Ad Hoc Network, Intelligent Transportation System

1 Introduction

In VANET vehicles are used as nodes which transmit and receive messages by using the wireless VANET based applications on the intelligent transportation system. In which vehicles are equipped with on board unit (OBU), road side unit (RSU), GPS, DGPS, application unit (AU), advanced mobile devices such as iPhone and other types of sensors. VANET turns every participating car as network router. VANET is an enhanced and dynamic form of MANET. Different type of communication in VANETs are Vehicle-to-Vehicle (V2V), Vehicle-to-Roadside (V2R) or Vehicle-to-Infrastructure (V2I). VANET provides wireless communication between moving vehicles using a dedicated short range communication (DSRC) IEEE 802.11p emended low overhead operation to 802.11p. Wireless access in vehicular environments (WAVE) referred by the IEEE to standardize the whole communication now. VANET also integrates different technologies such as WiFi IEEE 802.11p, WAVE IEEE 1609, WiMAX

adfa, p. 1, 2011.

© Springer-Verlag Berlin Heidelberg 2011

IEEE 802.16, Bluetooth, IRA, ZigBee. Different types of routing protocols are used for these communications. The purpose of VANET is to provide safety and comfort applications. Safety applications are most important for minimize the hazardous condition and road accident etc. Comfort applications include information about the weather, nearby restaurant, park and CNG station etc. Intelligent Transportation systems (ITS) can be extended for Water Transportation, Air Transportation and Rail Transportation beside Road Transportation.

2 Mobile ad hoc Networks (MANETs)

Along with their attractive properties, MANETs have other properties that make their realization a challenging task. Due to the mobile nature and the limited transmission ranges of MANETs' nodes, communication links between nodes are established and torn down in unpredictable fashion, as such MANETs are characterized by dynamic topologies. The self-control property, a powerful feature of MANETs, forces nodes to perform routing and other network administrative tasks cooperatively. The routing algorithms used in traditional wired networks generate a huge amount of overhead traffic in order to discover and maintain routes in such dynamic topologies. The limited bandwidth of the wireless channel renders all these algorithms non-feasible for MANETs routing.

3 VANETs a subtype of MANETs

Vehicular Ad-Hoc Networks (VANETs) are the special kind of MANETs that aims at providing communications among vehicles on the roads. Either VANET is a subtype or subset of MANET, but it has its own unique properties as compared with other types of MANETs, these unique characteristics of VANET include:

3.1 Predictable mobility

In VANET vehicles are constrained with layout, topology and obey road signs and traffic lights and responds other vehicles while leading predictability. The

fact that vehicles movement is predictable and constrained by the road trajectory is the feature that gives VANET predictable mobility. [6] [1] [12]

3.2 No power limitations:

As vehicles have the ability to provide continuous power to the OBU via the long life battery [6] [13]. So there is no critical challenge of power in Vehicular ad hoc networks like MANETs.

3.3 High computational ability:

Because the nodes in VANET are vehicles, they can be equipped with a sufficient number of Sensors and computational resources; such as processors, a large memory capacity, advanced antenna technology and global position system (GPS). These resources increase the computational capacity of the node, which help in increasing efficiency and accuracy.

3.4 Variable network density:

Network density depends on the traffic density because the traffic density changes in the case of traffic jam and so do network density also changes [13] [12].

3.5 Rapid changes in network topology

Network topology is directly subjective to the vehicles moving on high speeds, especially at the highways. Likewise, driver behavior is affected by the necessity to react to the data received from the network, [6] [13]. Life time of the link between vehicle depend on the radio communication range and direction of vehicle when radio range increase life time of link increase. When vehicle's movement direction is same life time of link is long, and when it is in opposite direction the life time is short [1].

3.6 Large scale network

The network scale could be large in dense urban areas such as the city centre, highways and at the entrance of the big cities [13] [12].

3.7 Providing safe driving, improving passenger comfort and enhancing traffic efficiency:

VANET provides direct communications among moving vehicles, so it provide information to drivers travelling in the same direction with warning messages about accidents. These applications also provide information about point of interest like weather and shopping malls etc [6].

4 VANET Assemblage

In intelligent transportation systems, each vehicle act as a sender, receiver, and router [4] to broadcast information to the vehicular network or transportation agency, which then uses the information to ensure safe, free-flow of traffic. The main components in this system are: Application Unit (AU), On Board Unit (OBU) and Road Side Unit (RSU). Each vehicle is equipped with an OBU and a set of sensors to collect and process the information then send it on as a message to other vehicles or RSUs through the wireless medium. RSU hosts an application that provides services, OBU uses these services. The RSU can also connect to the Internet or to another server which allows AU's from multiple vehicles to connect to the Internet.

The dispersal of RSUs and total number of RSUs are subject to the VANET protocol which is to be used. Though it is safe to assume that infrastructure exists to some extent and vehicles have access to it intermittently, it is unrealistic to require that vehicles always have wireless access to roadside units. Figures 1, 2 and 3 depict the possible communication configurations in intelligent transportation systems. These include Vehicle-to-vehicle (V2V), vehicle-to Infrastructure (V2I), and routing-based communications.



Fig. 1. Vehicle-to-Vehicle (V2V) communication

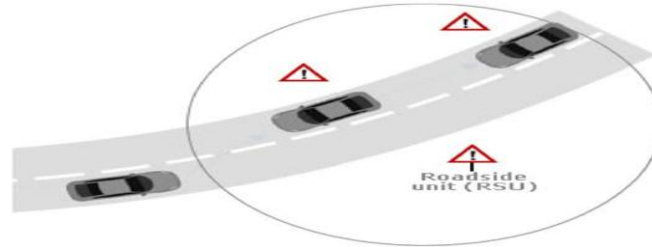


Fig. 2. Vehicle-to-Infrastructure (V2I) communication

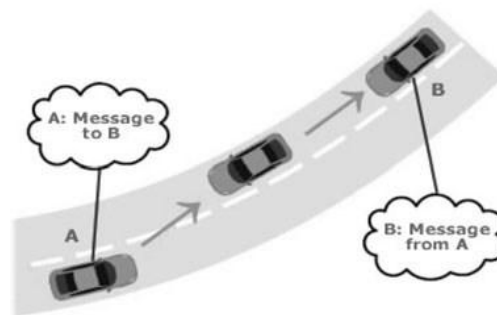


Fig. 3. Routing-based communication

4.1 On board unit (OBU):

An OBU is a wave device usually mounted on-board a vehicle used for exchanging information with RSUs or with other OBUs. It consists of a Resource Command Processor (RCP), and resources include a read/write memory used to store and retrieve information, a user interface, a specialized interface to connect to other OBUs and a network device for short range wireless communication based on IEEE 802.11p radio technology. It may additionally include another network device for non-safety applications based on other radio technologies such as IEEE 802.11a/b/g/n. These devices connected through wireless link based on the IEEE 802.11p radio frequency channel. The main functions of the OBU are wireless radio access, ad hoc and geographical routing, network congestion control, reliable message transfer, data security and IP mobility.

4.2 Application unit (AU)

The AU is the device equipped within the vehicle that uses the applications provided by the provider using the communication capabilities of the OBU. The AU can be a dedicated device for safety applications or a normal device such as a personal digital assistant (PDA) to run the Internet, the AU can be connected to the OBU through a wired or wireless connection and may reside with the OBU in a single physical unit; the distinction between the AU and the OBU is logical. The AU communicates with the network solely via the OBU which takes responsibility for all mobility and networking functions [2] [8].

4.3 Roadside Unit (RSU)

The RSU is a wave device usually fixed along the road side or in dedicated locations such as at junctions or near parking spaces dedicated for short range communication based on IEEE 802.11p radio technology, and can also be equipped with infrastructural Network. Main functions of RSU are as under:

- Extend the communication range.
- Provide internet connectivity to OBUs.
- Providing safety applications such as accident warning.

5 VANET Communication Domains

As shown in Fig. 4, the communication between vehicles and the RSU and the infrastructure form three types of domains:

- In Vehicle Domain
- Ad hoc domain (V2V & V2RSU)
- Infrastructure domain

- **In-vehicle domain:** In this domain AU and OBU are connected by WUSB (wireless universal serial bus) or UWB (Ultra-Wide Band). This domain consists of an OBU and one or multiple AUs. The connection could

be wired or wireless using WUSB or UWB; an OBU and an AU can reside in a single device. The OBU provides a communication link to the AU in order to execute one or more of a set of applications provided by the application provider using the communication capabilities of the OBU ([8] [2] Consortium; Festag et al., 2008).

- **Ad hoc domain:** The ad hoc domain on VANET is composed of vehicles equipped with OBUs and RSUs. According to [8] [2] [9], Festag et al. (2008), two types of communications are available in the ad hoc domain:
 - Vehicle to Vehicle (V2V)
 - Vehicle to Road Side Unit (V2I).

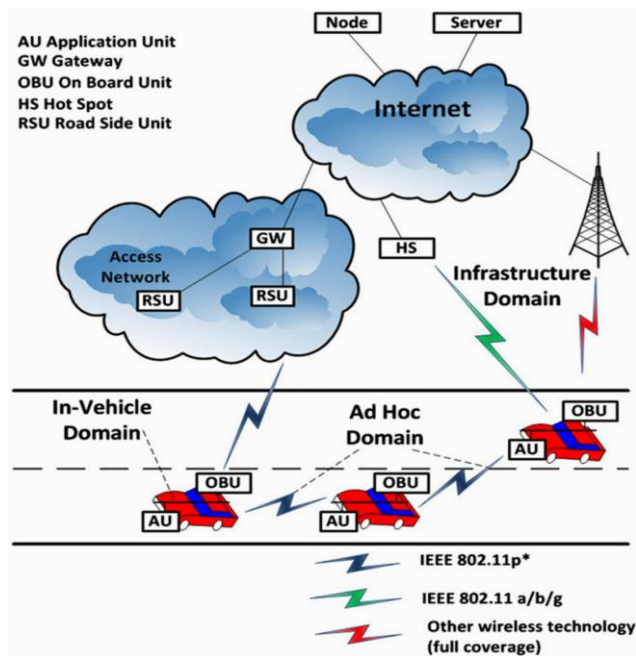


Fig. 4. Communication domains in VANET [8].

- **Infrastructural domain:** The OBU can access infrastructure network via RSU, which can connect to the infrastructural networks or to the Internet. The authors in Faezipour et al. [4] categorized the communication types in VANET into four types: intra vehicle communication which refers to the in-vehicle domain in our classification; vehicle to vehicle communication (V2V) and vehicle to roadside infrastructure communication (V2I) which

we classified them as the ad hoc domain and the last type of communications is the vehicle to broadband cloud communication, this type is similar to infrastructural domain in our classification.

Our categorization for VANET communication referred to Intra vehicle communication e.g (V2V) and (V2I) communication and Vehicle to Broadband (V2B) Include cloud communication where the vehicle communicate with a monitoring data center.

6 Wireless access technology in VANET

Wireless technology that used in VANET provide radio interface, some of these technologies rely on a centralized infrastructure to coordinate the communications between nodes. In contrast, other technologies operate in ad hoc mode (distributed coordination).

- **Cellular systems (2G/2.5G/3G):** cellular system is to reuse the limited frequency available for the service. GSM is one of the cellular systems which provide maximum data rate of 9.6 kbps .GSM uses FDMA and TDMA scheme. GPRS is general packet radio service that accommodates data transmission at high bandwidth with efficiency. High data rates are required to transmit multimedia data so UMTS is use for this purpose. Enhanced data rate evolution (EDGE) which is the advanced version of GSM is used for this purpose also provide peak data rate.
- **WLAN/Wi-Fi:** Wireless local area network (WLAN) or wireless fidelity (Wi-Fi) can provide wireless access to enable V2V communication or V2I communication. IEEE 802.11 standard can be applied to provide wireless connectivity. IEEE 802.11a works at 5 GHz and provides a data rate of 54 Mbps with a communication range of at least 38 m indoor and a 140 m range for outdoor use.
- **WiMAX:** —WiMAX or IEEE 802.16e is an amendment to the original worldwide interoperability for microwave access provides a high data rate and covers a wide transmission range with reliable communications and high quality of service (QoS), which are suitable for multimedia, video and voice over internet protocol (VoIP) applications. WiMAX achieves a

high data rate of up to 35 Mbps using multiple-input and multiple-output (MIMO), with an orthogonal frequency division multiplexing (OFDM) .

- **DSRC/WAVE:** It is licensed spectrum which is divided in to 7 channels starting from channel 172 and ending on 184. 178 is control channel used for safety applications. While the other (SHC) are both safety and non-safety communication.
- **Combined wireless access technologies:** CALM Continuous air interface for long and medium range (CALM M5) incorporating a set of wireless technologies including GSM-2G/GPRS-2.5G. Peoples using smart phones that can connect to the internet can communicate with each other via the internet forming people to people (P2P) network.

7 Challenges and requirements in VANET

- **Signal fading:** objects between two communicating devices work obstacle and preventing signal from reaching its correct destination and increasing fading.
- **Bandwidth limitations:** It is an important issue in VANET is absence of a central coordinator which controls the communication between nodes and manages bandwidth.
- **Connectivity:** owing to the high mobility and rapid changes of topology causes the NW fragmentation and lose the connectivity. Connectivity is important to increase the transmission power.
- **Security and privacy issue:** keeping privacy and security is one of the most challenge in VANET the receipt of trustworthy information from it source however this trusted information can violate the privacy needs of the sender
- **Routing protocol:** it is a critical challenge to design an efficient protocol that can deliver a packet in minimum period of time. However many researcher have concentrated on designing routing protocol suitable for dense environment that have a high density of vehicle with close distance between them .
- **Effective Protocol:** Enhance the reliability and reduced the extent of interference cause by high building taking scalability into consideration is es-

essential to avoid conflicts delivery of packet into short time specially in emergency.

8 VANET applications:

VANET applications are categorized in following table 1

Category	Sub-classification	Vehicular Application
Safety Applications	Collision avoidance	Cooperative collision warning Safe-distance notification Hazardous intersection message
	Road sign notification	Curve speed warning Cooperative violation warning
	Incident management	Emergency vehicle notification Post-crash notification
Efficiency Applications	Traffic Management	Intelligent traffic flow control Roadways planning Congested road notification
	Traffic Monitoring	Road conditions sensing Vehicles and fleet tracking
Infotainment Applications	Contextual Information	Commercial services announcement Parking assistance
	Entertainment	Media-content download Distributed games

Table 1. VANET applications

9 VANET simulation:

To test all applications in real life simulations are made to evaluate the performance of any network type, particularly wireless and ad-hoc network. These simulators emulate the network in terms of routing protocols, security constraints. Following tools are used for VANET simulations:

Network Simulator—ns-2: NS-2 is written in C++ with OTCL interpreter as a command and configuration interface. NS2 can support only the bi-directional and Omni directional antenna and needs manually programming.

NS-3: Most limitations of NS-2 are removed. NS-3 is completely written in C++. Available for different platform such as Windows, Linux, UNIX with limited coding in few hundred lines.

Global mobile information system: wireless network simulation. It was coded in Parsec. GlomoSim has the ability to run on SMP (shared-memory symmetric processor) assist to dividing NW in separate module running as separate process support millions of nodes as a single simulation.

MOVE:It is the mobility model generator based on JAVA supported by the GUI and has greater consideration on traffic levels. Mobility trace files can be generated from the Google Earth or TIGER databases. MOVE is not yet capable of simulating networks, but parses files of traces to be subsequently processed by a network simulator.

NCTUns: Built using C++ and it has high level GUI. Combine NW simulator and traffic simulator in a single module. Support ITS environment by using the SHARPE-format map file. NCTUns includes directional, bidirectional and omnidirectional antenna.

VANET MobiSim: VANET MobiSim was developed to overcome the limitations of CanuMobiSim. It supports V2V and V2I communications, which support stop signs, traffic lights and activities based macro-mobility with the support of human mobility dynamics.

TraNs: (traffic and network simulator environment) based on java and visualization tool to integrate SUMO and NS-2 but there is no feedback which fails to produce a simulation similar to real life. A separate version, called TraNs Lite, for mobility model generation only without using integrated NS-2.

References

1. Blum J, Eskandarian A, Hoffman L. Challenges of intervehicle ad hoc networks. IEEE Transactions on Intelligent Transportation Systems;5(4):347–51 (2004)
2. C.C. Communication Consortium. Car 2 car communication consortium manifesto. (<http://car-to-car.org/index.php?id=31>)

3. Chen W, Guha R, Kwon T, Lee J, Hsu Y.: A survey and challenges in routing and data dissemination in vehicular ad hoc networks. *Wireless Communications and Mobile Computing*;11(7):787–95 (2011)
4. Faezipour M, Nourani M, Saeed A, Addepalli S. Progress and challenges in intelligent vehicle area networks. *Communications of the ACM*;55(2):90–100 (2012)
5. Global Mobile Information System Simulation glomosim. (<http://pcl.cs.ucla.edu/projects/glomosim/>)
6. Jakubiak J, Koucheryavy Y.: State of the art and research challenges for vanets. In: Consumer communications and networking conference. CCNC 2008. 5th IEEE;pp 912–6 (2008)
7. Luo J, Hubaux J-P.: A survey of inter-vehicle communication. Technical Report, School of Computer and Communication Sciences. EPFL, Lausanne, Switzerland. Technical Report IC/2004/24; (2004)
8. Olariu S, Weigle MC.: Vehicular networks: from theory to practice. 1st ed. Chapman & Hall/CRC; (2009)
9. Sichitiu M, Kihl M. Inter-vehicle communication systems: a survey. *IEEE Communications Surveys & Tutorials*;10(2):88–105 (2008)
10. Transims. (<http://transims.tsasa.lanl.gov/>)
11. Wang S-Y, Lin C-C. Nctuns 5.0: a network simulator for ieee 802.11(p) and 1609 wireless vehicular network researches. In: Vehicular technology conference. VTC 2008-Fall. IEEE 68th; pp. 1–2 (2008)
12. Toor Y, Muhlethaler P, Laouiti A.: Vehicle ad hoc networks: applications and related technical issues. *IEEE Communications Surveys Tutorials*;10(3):74–88 (2008)
13. Yousefi S, Mousavi MS, Fathy M. Vehicular ad hoc networks (vanets): challenges and perspectives. In: Proceedings of the 6th international conference on ITS telecommunications; pp. 761–6 (2006)

PART ELEVEN Performance Issues in Optical Networks

Sustainable load in non-equidistant optical buffers

W. Rogiest and H. Bruneel

Department TELIN (EA07) · Ghent University (UGent) · St-Pietersnieuwstraat 41,
B-9000 Ghent, Belgium · {wouter.rogiest,hb}@telin.ugent.be

Abstract. In optical packet/burst switching, fiber delay line (FDL) optical buffers provide an effective means of contention resolution. In single-stage feedforward buffers, the delay line lengths are often chosen equal to multiples of a basic value called the granularity. Opposed to this, the more general case of non-equidistant grid structure has received little attention, despite its relevance in practical implementations.

In this contribution, we analyze the stability of an optical buffer model of infinite size with non-equidistant grid structure. For such a system, the sustainable traffic load is not one (as in a classic buffer system) but some value smaller than one, which is obtained as stability bound.

1 Introduction

Given the ever-increasing reliance on the Internet for our daily activities, the importance of a reliable Internet infrastructure can hardly be underestimated. Current backbone networks provide for reliable long-haul connections, but do so in an inefficient way, in the sense that most fiber capacity remains unused. Optical packet switching (OPS) and optical burst switching (OBS) [1] have been proposed to improve fiber utilization. In the implementation of current OPS/OBS switch prototypes, contention resolution is required whenever two or more packets head for the same outgoing channel at the same time. An effective solution is buffering; since light cannot be “frozen”, the buffer is implemented by letting packets traverse a piece of fiber, selected from a set of fiber delay lines (FDLs).

For FDL buffers, many performance models have been proposed for the case of single wavelength output buffering [2–4], with the buffer residing at the output port of a switch, which is assumed also in this contribution (see below). Here, it is further assumed that voids created on the outgoing channel are not filled by the scheduling algorithm; this corresponds to first-in-first-out (FIFO) scheduling. This is opposed to more complex algorithms, see e.g. [5], which provide improved performance, but do not preserve the original packet order. In the contributions mentioned [2–5], for the lengths of the delay lines, multiples of a basic value D are chosen. The corresponding grid is called equidistant; the delay difference between adjacent delays (neighbors in the set of delays when ordered from small to large) is constant, and equal to D . Opposed to this, as discussed in [6], in some cases, choosing non-equidistant line lengths yields superior performance. This observation motivates the interest in non-equidistant settings in [7–9] and also here. In [9], the sustainable load for this type of system was first analyzed,

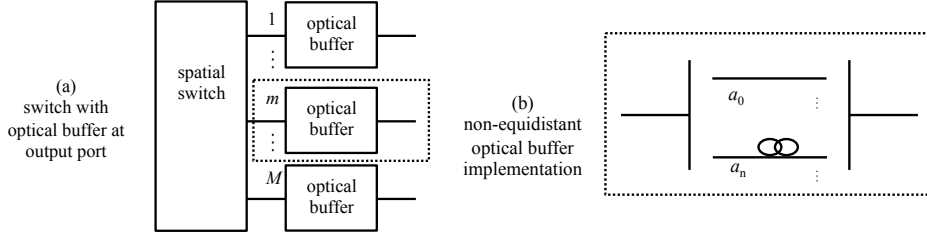


Fig. 1. The modeled output port with optical buffer (in dashed-line box) as part of a spatial optical switch (a), with (b) its implementation.

presenting sufficient (but not necessary) conditions for stability. To the best of our knowledge, the exact bound on the load (necessary and sufficient condition) has never been obtained, and is the focus of the current contribution.

The remainder of this paper is organized as follows. In the next section, the model assumptions and system equations are highlighted. Sect. 3 provides the analysis by means of probability generating functions, yielding an exact expression for the sustainable load. Sect. 4 is devoted to numerical examples of the sustainable load for three specific settings.

2 Model

Setting The setting assumed in this contribution is depicted in Fig. 1 (a). We assume a spatial optical switch with M output ports, operating in an asynchronous fashion. At each output port, an optical buffer is situated, resolving contention between packets heading for the same output port, on the same (single) wavelength. All M output ports are equivalent; in the following, we focus on the analysis of a single output port m , $m \in \{1, \dots, M\}$, chosen arbitrarily. We assume a discrete-time setting. Each time slot has a duration of a certain time unit; all variables and parameters below are expressed in multiples of this time unit which is not mentioned further below. A non-equidistant optical buffer implementation is considered, with an infinite number of delay lines. In its simplest instance, this is a single-stage buffer, with available delays pertaining to the set $\{a_n\}$, $n \in \{0, 1, \dots\}$. This is depicted in Fig. 1 (b). Further, some structure is assumed for the set of possible delays. More precisely, introducing the period $D > 0$ and a set of $P + 1$ different offsets $\{x_j\}$, $j \in \{0, 1, \dots, P\}$, we assume that

$$a_n \in \{i \cdot D - x_j\}, \quad (i, j) \in \{1, 2, \dots\} \times \{0, 1, \dots, P - 1\}, \quad (1)$$

with $n \in \{1, 2, \dots\}$, $x_0 < x_1 < \dots < x_{P-1}$, $a_0 = 0$, $x_0 = 0$ and $x_P = D$ by definition. The sequence of available delays is D -periodic, i.e., each period has length D . Each period is partitioned in P partitions with lengths $d_j = x_{j+1} - x_j$, $j \in \{0, 1, \dots, P - 1\}$, with $D = \sum_{j=0}^{P-1} d_j$.

While the analysis is provided for general P , special attention goes to the special case of two partitions ($P = 2$), for which a simple solution is obtained in Sect. 4. For the case $P = 2$, we assume as grid

$$\{0, E, D, D + E, 2 \cdot D, 2 \cdot D + E, \dots\}, \quad (2)$$

with E an additional grid parameter smaller than D , $0 < E < D$. Formally, this setting is a special case of (1), obtained for $P = 2$ (two partitions), and $x_1 = D - E$. Apart from being a simple (and thus, instructive) example of a non-equidistant buffer setting, this setting can also be used to model two-stage buffer implementations; the latter however falls beyond the scope of the current contribution. Below, we limit ourselves to the analysis of the setting given in Fig. 1 (b), with the case of general P (with grid structure (1)) and $P = 2$ (with grid structure (2)) treated separately.

Traffic assumptions and scheduling The given setting processes packets upon their arrival; packet length is modeled as an independent and identically distributed (iid) random variable (rv) B , with probability generating function (pgf) $B(z)$, defined as $B(z) = \mathbb{E}[z^B]$, $z \in \mathbb{C}$, where $\mathbb{E}[X]$ denotes the expected value of a rv X . Arrivals occur at the output port according to a Bernoulli arrival process with parameter p (the discrete-time counterpart to the Poisson process). The inter-arrival times are described by a random variable T , with expected value $\mathbb{E}[T] = 1/p$. Buffer control exercises a *horizon scheduling* algorithm to resolve contention between packets, assigning appropriate delays to packets upon their arrival. Horizon scheduling adheres to the FIFO principle (First In, First Out), always assigning the smallest delay available, the so-called waiting time, denoted W , that is larger than the time needed in order for all previously arrived packets to have left the system, the so-called scheduling horizon, denoted H . More precisely, a *delay* (or *waiting time*) $W = \min_{a \in \mathcal{A}} \{a \geq H\}$ is assigned. Here, the delay is defined as the time between arrival and start of transmission on the outgoing channel, and \mathcal{A} denotes the involved *quantization grid*, or *grid*,

$$\mathcal{A} = \{i \cdot D - x_j\} ; (i, j) \in \mathbb{Z} \times \{0, 1, \dots, P - 1\},$$

a periodic lattice with period D , which is similar but not identical to the set of delays given by (1), as realizable delays are always non-negative, while \mathcal{A} includes negative values, facilitating the analysis below.

System equations Numbering packets in the order of their arrival, the k th packet sees a scheduling horizon $(H)_k$ and receives a delay $(W)_k$, bringing an amount of work $(B)_k$. The next packet (numbered $k + 1$) arrives $(T)_k$ time slots later. Packet $k + 1$ finds horizon $(H)_{k+1}$, given by

$$(H)_{k+1} = \lceil \lceil (H)_k \rceil_{\mathcal{A}} + (B)_k - (T)_k \rceil^+, \quad (3)$$

where $[x]^+$ is a well-known notation for $\max(0, x)$ (here, with $x \in \mathbb{Z}$), and $\lceil x \rceil_{\mathcal{A}} = \min_{a \in \mathcal{A}} \{a \geq x\}$ (again with $x \in \mathbb{Z}$) denotes *the ceil of x with respect to grid \mathcal{A}* , a generalization of the classic ceil operation. The delay is only involved implicitly, but can easily be derived, by means of $(W)_k = \lceil (H)_k \rceil_{\mathcal{A}}$.

3 Analysis

On certain (well-known) conditions (irreducibility, aperiodicity), the random variables involved converge to a unique equilibrium distribution for $k \rightarrow \infty$. Correspondingly, the index k is omitted below, and all rv's are considered associated with this unique stochastic equilibrium.

Auxiliary variables and derivations For notational convenience, we introduce the rv $U = B - T$, with corresponding pgf

$$U(z) = \mathbb{E}[z^U] = \mathbb{E}[z^{B-T}] = B(z)T(1/z).$$

The random variable Y^l , $l \in \{0, 1\}$, is the ceil of the argument U with respect to the l -shifted version of grid \mathcal{A} , that is, $Y^l = \lceil U \rceil_{\mathcal{A}_l}$, where \mathcal{A}_l denotes the l -shifted version of grid \mathcal{A} , that is,

$$\mathcal{A}_l = \{i \cdot D - x_j + x_l\}, (i, j) \in \mathbb{Z} \times \{0, 1, \dots, P-1\}, \quad (4)$$

with $l \in \{0, 1, \dots, P-1\}$. Here, note that $\mathcal{A} = \mathcal{A}_0$. The pgf of Y^l , $Y^l(z)$, defined as $Y^l(z) = \mathbb{E}[z^{Y^l}]$, $l \in \{0, \dots, P-1\}$, $z \in \mathbb{C}$, can be expressed as follows (as explained below),

$$Y^l(z) = \sum_{k=0}^{D-1} \frac{1}{D} \frac{\sum_{m=0}^{P-1} \varepsilon_k^{x_m - x_l} ((z\varepsilon_k)^{d_m} - 1)}{z\varepsilon_k - 1} U(z\varepsilon_k), \quad (5)$$

Here, ε_k denotes one of the D th roots of unity, more precisely,

$$\varepsilon_k = \exp(j \frac{2\pi k}{D}),$$

with $k \in \{0, 1, \dots, D-1\}$, and j the imaginary unit (square root of -1). The obtained pgf's can each be split in P components, i.e., partial pgf's, $Y^l(z) = \sum_{m=0}^{P-1} Y_m^l(z)$. Here, each $Y_m^l(z)$, $(l, m) \in \{0, 1, \dots, P-1\} \times \{0, 1, \dots, P-1\}$, only contains corresponding powers of z (i.e., with equal l and m) present on the l -shifted grid $\mathcal{A}_l = \{i \cdot D - x_m + x_l\}$, $(i, m) \in \mathbb{Z} \times \{0, \dots, P-1\}$, with $l \in \{0, \dots, P-1\}$. Below, also other functions of z (not necessarily pgf's) are denoted accordingly:

$$W(z) = \sum_{m=0}^{P-1} W_m(z), \quad F(z) = \sum_{m=0}^{P-1} F_m(z),$$

where the components $W_m(z)$ (partial pgf's, see below) and $F_m(z)$ (not related to a pgf, see below) only contain corresponding powers of z (i.e., with equal m) present on the grid $\mathcal{A} = \mathcal{A}_0 = \{i \cdot D - x_m\}$, $(i, m) \in \mathbb{Z} \times \{0, \dots, P-1\}$.

To derive (5), we rely on a generalized interpretation of the discrete Fourier transform. More precisely, assuming $u_i = \Pr[U = i]$, $i \in \mathbb{Z}$, we define the D different partial pgf's $u_l(z)$ as follows,

$$u_l(z) = \sum_{m=-\infty}^{\infty} u_{m \cdot D - l} \cdot z^{m \cdot D - l}, \quad l \in \{0, 1, \dots, D-1\}.$$

Interestingly, the $u_l(z)$ can be expressed directly in terms of the pgf $U(z)$, through a transformation similar to a discrete Fourier transform, namely

$$u_l(z) = \sum_{k=0}^{D-1} \frac{1}{D} U(z\varepsilon_k) \varepsilon_k^l. \quad (6)$$

To now obtain (5), we derive $Y_m^l(z)$ and $Y^l(z)$ consecutively. The former is obtained as

$$Y_m^l(z) = \sum_{j=0}^{d_m-1} u_{x_m-x_l+j}(z) \cdot z^j = \sum_{k=0}^{D-1} \frac{1}{D} \frac{\varepsilon_k^{x_m-x_l} ((z\varepsilon_k)^{d_m} - 1)}{z\varepsilon_k - 1} U(z\varepsilon_k), \quad (7)$$

where we applied (6) in the second step. Summing over m , we easily obtain (5).

System equations in z -domain The next aim is to express the system equations in terms of pgf's. This is done in two steps ($H(z)$ and $W(z)$). The expression for $H(z)$ in the z -domain is

$$H(z) = W(z)B(z)T(1/z) + K \cdot C(z), \quad (8)$$

with K a real-valued constant whose value is determined by the normalization condition, $H(1) = W(1) = 1$. Further, $C(z)$ equals

$$C(z) = \frac{z-1}{z-(1-p)} = \frac{p}{(1-p)} \sum_{i=0}^{\infty} \left(\frac{z}{(1-p)} \right)^i + 1, \quad (9)$$

and comes about naturally, as part of the expression of the operator $[\cdot]^+$ in the z -domain. More precisely, as explained in e.g. [10], for $(X)_k$ a discrete rv associated with the k th arriving packet (or customer), and $(T)_k$ a rv with geometric distribution and parameter p , the expression

$$[(X)_k - (T)_k]^+,$$

has the following z -transform (for $k \rightarrow \infty$, under the assumptions mentioned at the start of this section),

$$X(z)T(1/z) + K \cdot C(z),$$

with K the real-valued constant mentioned earlier.

The relation of $W(z)$ as function of $H(z)$ is the same relation as the one between $Y^0(z)$ and $U(z)$, given by (5). By plugging (8) into the latter relation we obtain

$$W(z) = \sum_{m=0}^{P-1} W_m(z) Y^m(z) + K \cdot F(z), \quad (10)$$

with $F(z)$ a function of $C(z)$, again the same relation as the one between $Y^0(z)$ and $U(z)$, given by (5).

Solution for general P To extract an expression for $W(z)$ from (10), we consider its P components $W_m(z)$, $m \in \{0, 1, \dots, P-1\}$, by equating coefficients of similar terms. This yields following set of $P+1$ (plain algebraic) equations,

$$\begin{aligned} W_m(z) &= \sum_{l=0}^{P-1} W_l(z) Y_{m-l}^l(z) + K \cdot F_m(z), \quad m \in \{0, 1, \dots, P-1\}, \\ W(z) &= \sum_{m=0}^{P-1} W_m(z), \end{aligned} \quad (11)$$

where $Y_{m-l}^l(z)$ is shorthand for $Y_{(m-l) \bmod D}^l(z)$. For any value of P , this set of linear equations can be applied recursively to obtain an explicit expression for $W(z)$. In this expression, however, K is yet to be determined. Taking the first derivative with respect to z of (10) (where primes denote derivatives, e.g., $F'(z) = dF(z)/dz$) and evaluating in $z = 1$, one obtains following expression for K ,

$$K = - \frac{\sum_{l=0}^{P-1} \pi_l \cdot \mathbf{E}[[U]_{\mathcal{A}_l}]}{F'(1)}, \quad (12)$$

where $\mathbf{E}[[U]_{\mathcal{A}_l}] = Y^{l'}(1)$ (as a property of pgf's) and

$$\pi_l = W_l(1) = \Pr[(W + x_l) \bmod D = 0], \quad l \in \{0, 1, \dots, P-1\},$$

a set of P different steady-state probabilities, with $\sum_{l=0}^{P-1} \pi_l = 1$. As such, (12) completely and uniquely characterizes K .

Solution for $P = 2$ To show that the above leads to an explicit expression for $W(z)$, we focus on the case of two partitions, $P = 2$ (of special interest also below). Indeed, departing from (11), some calculation leads to the following solution in terms of pgf's,

$$W(z) = K \cdot \frac{F_0(1 + Y_1^0 - Y_0^1) + F_1(1 + Y_1^1 - Y_0^0)}{(1 - Y_0^0)(1 - Y_0^1) - Y_1^0 Y_1^1}, \quad P = 2, \quad (13)$$

where F_m and Y_m^l , $(l, m) \in \{0, 1\} \times \{0, 1\}$, are shorthand for $F_m(z)$ and $Y_m^l(z)$. Here, the value for K (still for the case $P = 2$) can be obtained immediately from (13) through the normalization condition $W(1) = 1$ and applying de l'Hôpital's rule, finding

$$\begin{aligned} K &= \frac{K_t}{K_n}, \\ &= \frac{-[Y_1^1(1)Y_0^{0'}(1) + Y_1^0(1)Y_1^{1'}(1)]}{F'(1)[Y_1^1(1) + Y_1^0(1)] + F_0(1)[Y_0^{0'}(1) - Y_1^{1'}(1)]}, \quad P = 2, \end{aligned} \quad (14)$$

where we coined the labels K_t and K_n for the numerator and denominator of K . One can verify that (14) is indeed consistent with (12), as should, with

$$\pi_l = \frac{F'(1)Y_1^{(1-l)}(1) - F_0(1)Y^{(1-l)'}(1)}{K_n}, \quad l \in \{0, 1\}, \quad P = 2.$$

Stability condition for general P From (11) (for general P), the probability of finding an empty system upon arrival, $\Pr[W = 0]$, is obtained as

$$\Pr[W = 0] = \lim_{z \rightarrow 0} W_0(z) = \frac{K}{1 - p}. \quad (15)$$

Due to the infinite system size, the distribution of the delay does not reach steady state for $k \rightarrow \infty$ when the system is overloaded, which occurs if some critical load threshold is exceeded. With ρ denoting the traffic load, $\rho = p \cdot E[B]$, ρ_{tol} denotes the maximum tolerable load, which is smaller than one for this type of (optical) system. When this load is exceeded, the delay grows unboundedly, and $\Pr[W = 0] \rightarrow 0$. As such, given (12) and (15), ρ_{tol} is characterized by

$$\sum_{l=0}^{P-1} \pi_l \cdot E[[U]_{\mathcal{A}_l}] = 0 \quad \text{when} \quad \rho = \rho_{tol}. \quad (16)$$

Stability condition for $P = 2$ Using the above-mentioned characterization of stability, by virtue of (15), it is clear that $K_t \rightarrow 0$ as $\rho \rightarrow \rho_{tol}$. As such, ρ_{tol} is characterized by $K_t = 0$, i.e., given (14), by

$$Y_1^1(1)Y^{0'}(1) + Y_1^0(1)Y^{1'}(1) = 0 \quad \text{when} \quad \rho = \rho_{tol}, P = 2. \quad (17)$$

4 Numerical examples

In this section, we consider the stability of three system settings. For all three, we assume a system with two partitions ($P = 2$), with the set of available delays given by (2), with $0 < E < D$ as mentioned.

Across examples, the average packet length is always chosen equal to 21 time slots, $E[B] = 21$ (which, for a time slot length of 100 ns, corresponds to 2.1 μ s). This particular choice is made to obtain a ‘‘convenient’’ value for Δ , a notion introduced in [4]. There, for the case of an equidistant buffer setting, it was shown that the *largest useful delay difference* between adjacent lines (where adjacent lines are neighbors in the set \mathcal{A} ordered from small to large) equals

$$\Delta = B_{max} - T_{min}, \quad (18)$$

with B_{max} the maximum of the support of B , and T_{min} the minimum of the support of T . Here, ‘‘largest useful’’ indicates that no performance improvement or stability improvement can be expected by choosing delay differences larger than this value. As can be verified, the motivation given in [4] for (18) is valid also for the more general case of non-equidistant settings. Further, under the assumptions of this contribution, $T_{min} = 1$, and the largest useful delay difference is thus $\Delta = B_{max} - 1$, always a multiple of ten in the examples, and thus a ‘‘convenient’’ value. Below, three distributions are considered:

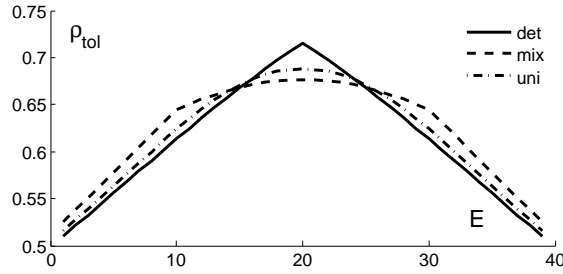


Fig. 2. Maximum tolerable load for a packet length distribution with range $Q = 10$ for *mix* and *uni*.

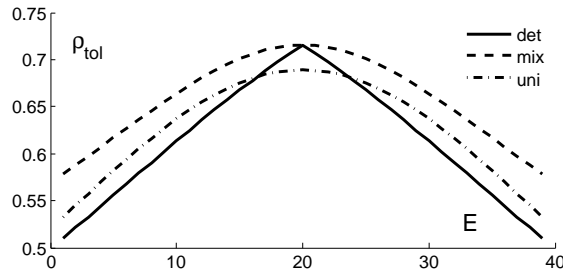


Fig. 3. Maximum tolerable load for a packet length distribution with range $Q = 20$ for *mix* and *uni*.

1. *det*. A deterministic distribution; $B = 21$,

$$B(z) = z^B,$$

and thus, $\Delta = 20$.

2. *mix*. A balanced (fifty-fifty) mixture of two packet lengths B_1 and B_2 , differentiated by a range parameter Q , with average $B = 21$, $B_1 = B - Q$, $B_2 = B + Q$, and

$$B(z) = (z^{B_1} + z^{B_2})/2.$$

Here, $\Delta = B + Q - 1$. The cases of intermediate range $Q = 10$ ($\Delta = 30$) and maximum range $Q = 20$ ($\Delta = 40$) are considered below.

3. *uni*. A uniform distribution with range parameter Q , with average $B = 21$, range $\{B - Q, \dots, B + Q\}$, and

$$B(z) = \frac{1}{2Q+1} \cdot \frac{z^{B+Q+1} - z^{B-Q}}{z-1}.$$

Just like for *mix*, $\Delta = B + Q - 1$, with the cases $Q = 10$ ($\Delta = 30$) and $Q = 20$ ($\Delta = 40$) considered below.

In Figs. 2–4, the maximum tolerable load ρ_{tol} (as expressed by (17)) is evaluated for three different settings; for each setting, the comparison between the

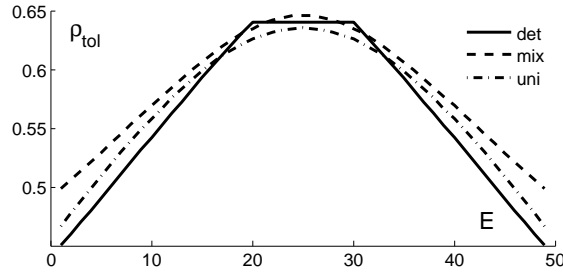


Fig. 4. Maximum tolerable load for the same setting as in Fig. 3, but with increased D .

three mentioned packet length distributions is highlighted. Each figure shows the maximum tolerable load as function of the delay parameter E , which varies between 0 and D , $0 < E < D$. In Figs. 2–3, the granularity D is set to 40, $D = 40$, and the only difference lies in the packet length range for *mix* and *uni*: an intermediate range $Q = 10$ in Fig. 2, and a maximum range $Q = 20$ in Fig. 3.

Focusing on Fig. 2, we find that, for all three distributions, the equidistant setting (with $E = D/2 = 20$) provides best stability, with highest tolerable load occurring for a deterministic distribution (*det*) in an equidistant setting, at $\rho_{tol} = 0.715$. All curves are symmetrical with respect to $E = D/2$, which comes as no surprise, given the symmetry present in the stability condition (17). Further, the choice of E appears more critical for *det* (narrow optimum) than for *mix* and *uni* (broader optimum). As can be seen on Fig. 3, a similar conclusion holds for maximum packet length range $Q = 20$, the main difference being that *det* and *mix* apparently exhibit an identical stability bound (again, $\rho_{tol} = 0.715$) for an equidistant setting ($E = D/2$).

Considering Figs. 2 and 3 together, there seems no reason to deviate from the equidistant setting ($E = D/2$) when choosing the length E . This observation is however highly related to the choice of $D = 40$, especially in the case *det*, for which the largest useful difference between adjacent delays (in the sense mentioned above) is 20, and thus, $E = 20$ yields an equidistant grid that is highly adapted to the needs of deterministic packet length traffic.

In Fig. 4, a different value $D = 50$ is considered. As can be seen, the symmetry relative to $E = D/2$ still holds, yet the stability for *det* is unaltered by the exact value of E when $E \in \{\Delta, \dots, D - \Delta\}$, with $\Delta = 20$ for *det*. This can be understood intuitively by the notion of largest useful delay difference: any delay difference exceeding Δ (while not exceeding $D - \Delta$) does not contribute to increased stability, leaving the maximum tolerable load unaltered in case of *det*. Due to this effect, the maximum tolerable load for *mix* and *uni* (for which Δ equals 40 and is thus *not* exceeded) increases along when E increases from 20 to 25. In case of *mix*, the maximum tolerable load of *mix* even surpasses that of *det* for $E = D/2$.

5 Conclusions

In this contribution, a stochastic model for a non-equidistant optical buffer was analyzed. Here, a finite sequence of partitions (with non-equidistant delay differences) is taken as basic unit and repeated, to span an infinite grid. Through the use of probability generating functions, we obtained an expression to calculate the maximum tolerable load, valid regardless of the exact number of partitions. The case of two partitions is focused on explicitly in the analysis. Numerical examples provide for additional insight in the system's functioning, including the role of the so-called largest useful delay difference.

Acknowledgment

Part of this research has been funded by the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office. The first author is Postdoctoral Fellow with the Research Foundation Flanders (FWO-Vlaanderen).

References

1. L. Xu, H. G. Perros, and G. Rouskas. Techniques for optical packet switching and optical burst switching. *IEEE Commun. Magaz.*, 39(1):136–142, 2001.
2. F. Callegati. Optical buffers for variable length packets. *IEEE Commun. Letters*, 4(9):292–294, 2000.
3. M. Yasuji. An approximation for blocking probabilities and delays of optical buffer with general packet-length distributions. *J. Lightwave Technol.*, 30(1):54–60, 2012.
4. W. Rogiest, J. Lambert, D. Fiems, B. Van Houdt, H. Bruneel, and C. Blondia. A unified model for synchronous and asynchronous FDL buffers allowing closed-form solution. *Perform. Eval.*, 66(7):343–355, 2009.
5. F. Callegati, A. Campi, and W. Cerroni. Fast and versatile scheduler design for optical packet/burst switching. *Optical Switch. and Netw.*, 8(2):93–102, 2011.
6. J. Lambert, B. Van Houdt, and C. Blondia. Single-wavelength optical buffers: non-equidistant structures and preventive drop mechanisms. *Proceedings of NAEC 2005, Italy*, pages 545–555, 2005.
7. R.C. Almeida, J.U. Pelegrini, and H. Waldman. A generic-traffic optical buffer modeling for asynchronous optical switching networks. *IEEE Commun. Letters*, 9(2):175–177, 2005.
8. W. Rogiest, D. Fiems, K. Laevens, and H. Bruneel. Tracing an optical buffer's performance: An effective approach. volume 4465 of *Proceedings of NETCOOP'07, Avignon, June 2007 – Lecture Notes in Computer Science*, pages 185–194. Springer, 2007.
9. W. Rogiest and E. Morozov et al. Stability of single-wavelength optical buffers. *European Transact. on Telecomm.*, 21(3):202–212, 2010.
10. K. Laevens, M. Moeneclaey, and H. Bruneel. Queueing analysis of a single-wavelength fiber-delay-line buffer. *Telecommun. Syst.*, 31(2-3):259–287, 2006.

A Novel OLT Based Energy Efficiency Algorithm in TDM Passive Optical Networks

O.C. Turna^{1,2}, M.A. Aydin¹, T. Atmaca²

¹Department of Computer Engineering Istanbul University, Istanbul, Turkey

²Laboratoire SAMOVAR, UMR 5157, IMT/Telecom SudParis, Evry, France

ozcantur, aydinali@istanbul.edu.tr

tulin.atmaca@telecom-sudparis.eu

Abstract. In last decade, energy efficiency is one of the most focused topics. Since, for the future of our planet, usage of the natural resources is getting greater interest. For this purpose, there are ongoing studies to develop less energy consuming products in both academic and industrial environments. Telecommunication networks are one of the major slices of the global energy consumption. Since 2009, a couple of standards for energy conservation in PONs have been developed. These standards and most of the researches are based on improvements on ONU side. In this study, a novel energy efficiency algorithm which is based on coupling two OLT to reduce energy consumption in central office is proposed. Our design employs optical switches and amplifiers to create a switch-box, which is under control of both OLT pairs. The finite state machine diagrams of control plane of this novel design, which is performed by both OLT pairs over a dedicated line, are also presented.

Keywords: PON, OLT, Energy Efficiency, Performance

1 Introduction

Passive Optical Network (PON) is the most promising solution for future access networks. Since, using optical technology in fix networks brings advantages as; reaching long-haul, less corruption of carried traffic, and far less being affected from environmental factors. Besides these advantages PON also gives the most favorable solution when we consider energy consumption, which is one of the crucial problems for future of our planet. One of the considerable energy consumer is network devices which are distributed all over the world for Internet coverage. Thus, besides other network segments, energy efficiency in PON also gets much more attention in last years. According to the study in [1], EPON systems around the world consumes 2.37TWh energy annually, which is approximately equals to 300.000 house energy necessity. In 2009, power conservation recommendations for GPON and XG-PON standardization is released to cover sort of energy conservation methods in PON systems [2,3]. In 2012, a novel approach Bit-Interleaving PON (Bi-PON) released by GreenTouch consortium [4]. GreenTouch is founded by a group of network companies and aca-

demographic institutes to accomplish the purpose that ICT energy consumption in 2010 will be decreased factor of 1000 in 2015. Bi-PON aims to reduce the power consumption on end user devices (Optical network Unit - ONU) and reveal a success that energy consumption of ONU reduced from 3.5W (XG-PON) to 0.5W (Bi-PON) in average. On the other side, central office equipment (Optical Line Terminal – OLT) consumes approximately 10W-100W according to the manufacturer [5, 6, 7, and 8]. There are a lot of ONU based studies for energy efficiency, where just a few propositions presented for energy conservation in OLT. Solutions for OLT use wavelength division in [9], grouping OLTs in [10] and summary of some sort of studies presented in [11]. When these solutions are examined, it is observed that these methods are mostly based on a major device renovation with novel technical devices or close the device within some time period of day. Our contribution provides an energy efficiency approach with a diminutive modification on central office systems and no modification reflected to the ONU part. In our approach, two OLT as a couple handle each other traffic flows under low load. Thus, one OLT can be left in deep sleep mode till a heavy traffic volume occurs on uplink or downlink. These operating modes are switched dynamically according to the traffic pattern.

2 OLT Based Energy Efficiency Algorithm

In a PON system, OLT consumes considerably high energy, which is a challenge for green networking and cause operational expenditure for service providers. In course of time, network equipment generally are not fully utilized especially while servicing home subscribers. Since, home devices can be switched off, be idle or in usage with a quite light traffic (i.e. web surfing traffic). In such cases by the use of different techniques some network equipment can be switched off or sent to sleep / deep sleep modes. In PON, there is no intermediate equipment that can be removed, and OLTs and ONUs must always work for providing service. However, owing to time division multiplexing (TDM) aspects, ONUs are actively in use less than %10 of the timeline. Through, ONUs can be put into sleep mode while the ONU is not actively in use. On the contrary, OLT must be active all the time to schedule and serve each ONU simultaneously. If we want to put an OLT into sleep mode, the subscribers must be served by another network equipment.

Optical amplifiers are used to elevate optical power to reach long-hauls with fiber lines. Also they are used to recover the decay of passive splitter on divided optical power by two or more fiber. A summary of amplifiers used in PON systems as; xDFA (...-doped fiber amplifier), Raman amplifier, and SOA (semiconductor optical amplifier) is given in [12]. According to the study a raman amplifier consumes 0.5W. Besides, by novel methods, decreasing the energy consumption of amplifiers is aimed. Optical switches are used in re-configurable optical add/drop multiplexer, optical cross-connect systems, and network switching for fault and restoration applications. In our proposition, optical switches are used for forwarding one OLT's subscribers (ONUs) to the other on power saving mode.

The proposed approach uses a switch-box that is electronically connected with the OLTs for control plane and consists of 1x2 optical switches and optical amplifier. The switch-box implementation is assumed to use far less power compared to an OLT. On standard process switch-box shuts down the amplifier and consumes energy just for the electronic control unit. In energy saving mode consumes as ~0.6W. The amplifier only in use when one OLT put into sleep mode, in another words when it is in energy saving mode to satisfy the power budget for doubled ONU count.

2.1 Developed Architecture

The developed architecture comprises two OLT (OLT-A for sleeping one, OLT-B for master one) and a switch-box consists of five 1x2 optical switches and one amplifier. OLTs and switch-box are directly connected by dedicated communication lines. System works in two modes; Normal and Energy Saving. The transition between these two modes is decided according to the traffic load over the OLTs. If one of the OLTs (A) has a weak traffic volume (%10 is taken as a reference for this study) on both uplink and downlink, it asks to OLT-B for passing to the energy saving mode. If OLT-B's traffic volume is under average load (%40 is taken as a reference), it responds with a positive response and starts processes to put OLT-A into deep sleep mode (OLT is completely shut down except wake on by signal feature). Thus OLT-B must be able to take all the responsibilities of OLT-A and gives the services without any unsatisfactory delays. If OLT-B encounter a heavy traffic volume (over %90 as a reference) in uplink or downlink, than OLT-B wakes up and prepares OLT-A to serve as in usual way.

In the implementation of the system the points below have to be considered;

- OLTs have to capable to exchange their synchronization and route tables with each other over a dedicated connection. It can be an Ethernet, a USB or over a BUS.
- A switch-box that consists of optical switches and amplifier must be implemented. Energy consumption of the switch must be at a very low level.
- OLTs can be capable of controlling the switch-box and buffering the traffic of backbone connection while the other OLT is not ready for switching process.

The general structure and switch-box design are shown in Fig.1. Each OLT connects to the switch box with a single fiber. The fiber coming from OLT is connected to a 1x2 optical switch. The fiber coming from the ONUs side also connected to another 1x2 switch at the entrance of switch-box. On normal mode, through input switches OLTs directly connected to its ONUs, on energy saving mode both ONU set directed to middle line in switch-box and it is directed to the active OLT (OLT-B).

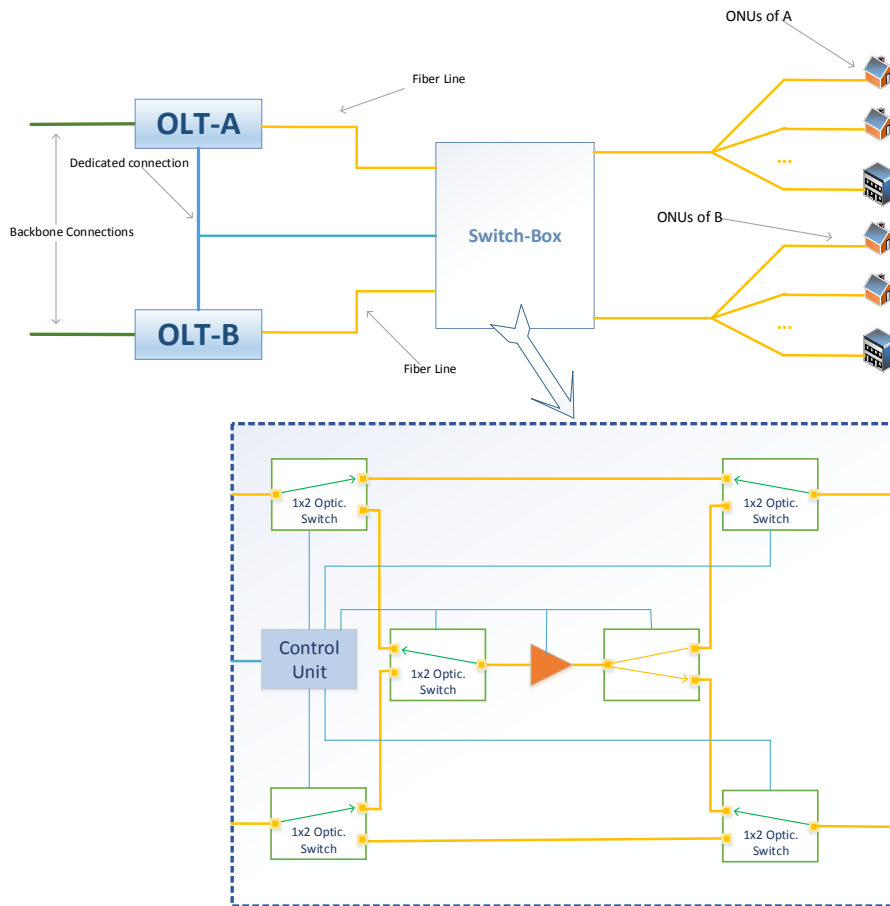


Fig. 1. General Architectural Design of OLT Couple with switch-box

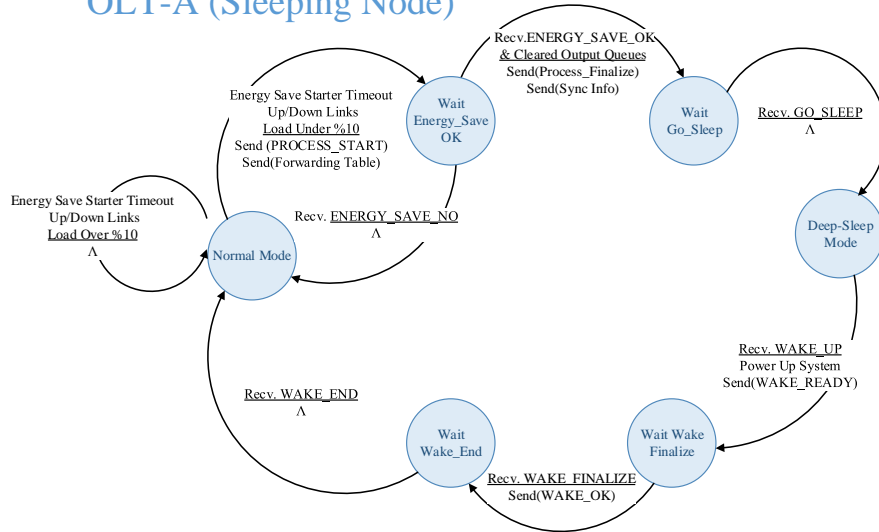
2.2 Proposed Algorithm for Control Plane

System works in two modes as Normal and Energy Saving. In normal mode, the switch-box always waits on stand-by. The amplifier is not powered and optical switches are locked in their states. The system performs as usual such as two separate OLTs work. To switch between modes, a control plane is used to schedule and share information between OLTs. The control plane of the algorithm is given as a finite state machine in Fig. 2 and summarized below. First figure shows the states of sleeping node and second on is states of active (master) pair.

When one of the OLTs (OLT-A) have less traffic than a particular lower bound in both uplink and downlink (taken as %10 for this study), it informs OLT-B with ENERGY_SAVE_REQUEST from the dedicated connection. If OLT-B's traffic vol-

ume is low enough (i.e. %40), OLT-B responses with ENERGY_SAVE_OK response otherwise ENERGY_SAVE_NO from the dedicated communication link. If OLT-B responses with affirmative result, OLT-A starts the preparations to go to deep-sleep mode. Otherwise OLT-A waits for a time period for a new try.

OLT-A (Sleeping Node)



OLT-B (Active Node)

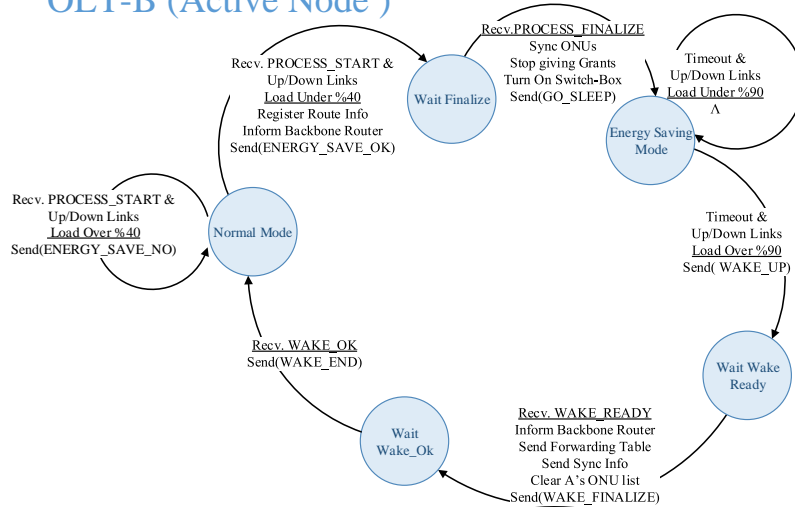


Fig. 2. Finite State Machine of Transition between two modes in Control Plane

OLT-A executes the processes below to prepare itself for deep-sleep mode;

- OLT-A must lead all connection (routing information) to transmit over itself to OLT-B. Thus, OLT-A activates OLT-B from the direct connection link by sending PROCESS_START control message and its forwarding table.
- Till the connection establishes over OLT-B (the time is about refreshing Forwarding Table of connected router) OLT-A receives incoming packets from backbone and schedule them for its ONUs. After a while, the forwarding table is refreshed (can be assumed or strictly checked from the backbone device). After sending all the queued packets, OLT-A sends PROCESS_FINALIZE message to OLT-B to inform it is ready for deep sleep.
- Just before sending PROCESS_FINALIZE message, OLT-A clears all the grant information for its ONUs and give the sync information to OLT-B. By doing so, OLT-A should be sure of its ONUs that no packets will be on fiber while switching process is on.
- After finishing all the operations OLT-A waits to get GO_SLEEP message from OLT-B. If no response returns from OLT-B, OLT-A cancel the energy saving process and wait for new commands and packets in active state.

OLT-B performs the processes below before switching to energy saving mode;

- After receiving PROCESS_START message, OLT-B waits for OLT-A's forwarding table, then update its forwarding table and inform the backbone router for new routes. Prepare itself to communicate ONUs of OLT-A, and waits for PROCESS_FINALIZE message form OLT-A.
- Before PROCESS_FINALIZE message, OLT-B gets sync information for ONU's of OLT-A, and schedules bandwidth allocation for all ONUs including A and B networks.
- With PROCESS_FINALIZE message, OLT-B arranges a time interval for switching, stops giving grants and stops downstream traffic for this interval, and turns the switch-box for forwarding packets to itself.
- While switching to energy saving mode, OLT-B stores any arrived packets that belongs to ONUs of OLT-A. After the switching procedure, these packets scheduled with appropriate grant (GATE) information.

When OLT-B reaches to a highly utilized situation (over %90 usage on downlink or uplink) then it demands OLT-A to wake up to share the ONUs again and work in normal mode by sending WAKE_UP message from dedicated link. After OLT-A receives WAKE_UP signal, it powers up and prepares its parameters (buffers, schedulers, counters etc.) and response to OLT-B with WAKE_READY. After OLT-B takes WAKE_READY message it knows that OLT-A can handle the incoming messages. OLT-B sends sync information to OLT-A and rearranges forwarding table of backbone router, itself, and OLT-A. After all arrangements, OLT-B sends WAKE_FINALIZE message to OLT-A. If preparations are finished as expected OLT-A responds with WAKE_OK and OLT-B turns the switch-box in normal work-

ing mode. After that, OLT-B triggers OLT-A to start processing as normal with WAKE_END control message.

3 Performance Evaluation

To evaluate the performance of our proposition it is compared with standard OLT implementation with OMNET++ 4.3 tool. We simulate a simple scenario where two OLT cards connected with aggregation node and switch-box that are supplied with traffic sources. Traffic sources are configured to produce different loads in successive time intervals to easily analyze the performance of the energy efficiency algorithm. The time intervals and load is exponentially random distributed. For a given time the traffic load is arbitrarily change between zero to one. Traffic generator creates different loads in exponentially distributed time intervals where 5sec selected as mean of time interval. The interval of load change can directly affect the delay time. In Figure 3 the sleep percentage (energy saving) and delay comparisons under different traffic loads are given. The simulation parameters are selected as; 1Gbps link bandwidth, 20MB output buffers for OLT and 1500Byte packets. Energy Efficiency algorithm triggered every 10ms to check the load and make decision to stay in same mode or change to other.

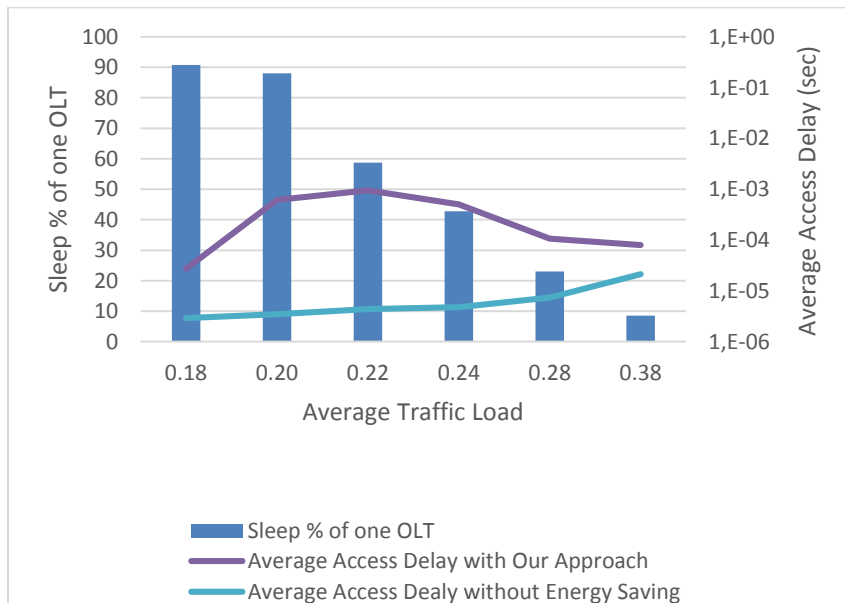


Fig. 3. Delay & Sleep Performance under different loads (With & Without Energy Saving)

The results that has traffic load over 0.38 are not shown. Since, the energy saving mechanism is unable to work under high load. Under low loads one OLT of the system can stay in sleep %90 as average. While saving energy, our proposition brings

some extra delay in overall system performance. While the delay is increased according to the increasing load values our dynamic decision mechanism can keep the system to perform without bottlenecking the bandwidth.

4 Conclusion

In this paper, a novel OLT based energy efficiency technique is presented. A control plane is designed for communications between OLT pairs and switch-box. Our strategy is simply putting one of the OLTs into deep-sleep mode as much as possible while using standard TDM messaging procedure in PON. This approach can reduce the OLTs energy consumption under low loads and differs from other OLT-based approaches. Since, it does not bring to much structural change and can be performed by a service provider without leaving PON standard implementation. Besides, our structure has no limitations for using ONU-based energy conservation approaches as XGPON, and BIPON. Besides, this approach seems to be promising for implementation expenditures compared to WDM based solutions.

As the negative aspects of our proposition, it requires a switch-box implementation and use of dedicated communication line between OLT pairs and switch-box. While the system is on the switching process the packets are delayed in OLT and ONU buffers. This waiting time is necessary because of switching time and aggregation node updates (i.e. less than 30ms).

The switching mechanism can be triggered dynamically all the time or a weighted approach can be used for day-night, weekdays-weekend pairs for a better performance and energy conservation.

As a future study, the performance of energy saving scheme will be presented with different traffic patterns. Dynamic vs. time-based schedule working modes will be compared. And also, QoS aware decision making is going to be investigated. Afterwards, we aim to figure out, 1xN OLT switch-box design and examine performance evaluation.

5 Acknowledgement

This work was partially supported by Scientific Research Projects Coordination Unit of Istanbul University (Project number 35709).

6 References

1. Nishihara, S.; Hajduczenia, M.; Mukai, H.; Elbakoury, H.; Hirth, R.; Kimura, M.; Kato, M., "Power-saving methods with guaranteed service interoperability in Ethernet passive optical networks," *Communications Magazine, IEEE*, vol.50, no.9, pp.110,117, September 2012.

2. ITU-T, "G.Sup45: GPON power conservation," 2009 [Online], Available: <http://www.itu.int/rec/T-REC-G.Sup45-200905-I/en>.
3. ITU-T, "G.987 Series: 10-Gigabit-capable passive optical network (XG-PON) systems: Definitions, abbreviations and acronyms", 2010 [Online] Available: <http://www.itu.int/rec/T-REC-G.987/en>.
4. GreenTouch Consortium, "BIPON: Bit-Interleaved Passive Optical Network Technology", 2012 [Online], Available: <http://www.greentouch.org/index.php?page=Bi-PON>
5. Klaus Grobe, "Performance, Cost, and Energy Consumption in Next-Generation WDM-based Access," BROADNETS 2010, Athens, October 2010.
6. Navid Ghazisaidi, Martin Maier, "Techno-economic analysis of EPON and WiMAX for future Fiber-Wireless (FiWi) networks", *Computer Networks* 54 (2010) 2640–2650.
7. Pulak Chowdhury, "Energy-Efficient Next-Generation Networks (E²NGN)", Thesis (Ph.D.), University of California, Davis, 2011.
8. Slavisa Aleksic, Margot Deruyck, Willem Vereecken, Wout Joseph, Mario Pickavet, Luc Martens, "Energy efficiency of femtocell deployment in combined wireless/optical access networks", *Computer Networks* 57 (2013) 1217–1233.
9. TOKUHASHI, Kazumasa and ISHII, Daisuke and OKAMOTO, Satoru and YAMANAKA, Naoaki, "Energy Saving Optical Access Network based on Hybrid Passive/Active Architecture", 電子情報通信学会技術研究報告. PN, フォトニックネットワーク, The Institute of Electronics, Information and Communication Engineers, vol.111, no:274, pp 7-12, 2011.
10. Saliou, F.; Chancelou, P.; Genay, N.; Laurent, F., "Energy efficiency scenarios for long reach PON Central Offices," *Optical Fiber Communication Conference and Exposition (OFC/NFOEC), 2011 and the National Fiber Optic Engineers Conference*, vol., no., pp.1-3, 6-10 March 2011.
11. Kani, J., "Power Saving Techniques and Mechanisms for Optical Access Networks Systems," *Journal of Lightwave Technology*, vol.31, no.4, pp.563-570, Feb.15, 2013.
12. E. Trojer, S. Dahlfort, D. Hood, and H. Mickelsson, "Current and Next-Generation PONs: A Technical Overview of Present and Future PON Technology," *Ericsson Rev.*, no. 2, 2008, pp. 64-69.

PART TWELVE Acoustic Networks

Underwater Acoustic Networks

Javier Poncela

University of Malaga, Spain

Abstract. Wireless underwater acoustic communication systems have become one of the most promising technologies for the development and deployment of future ocean observation and sensor networks. Applications range from oil prospecting and transportation to aquaculture, and include pollution control, climate recording, prediction of natural disturbances, search and survey missions, etc. The high attenuation of electromagnetic waves in underwater medium precludes them from being used as the information vehicle. Therefore, acoustic waves are the most viable alternative for the transmission. Nevertheless, the underwater acoustic channel is not free of drawbacks. On the contrary, it exhibits important obstacles that must be overcome if communication networks are to be implemented in the future. In addition, underwater systems have inherently serious problems of power supply and batteries duration. At the same time, network protocols must take into account the specific characteristics of the transmission medium, not dismissing the slow propagation speed. This talk will present an overall view of underwater acoustic networks and the impact of the acoustic channel characteristics in network protocols.

Underwater Acoustic Networks

Wireless underwater acoustic communication systems have become one of the most promising technologies for the development and deployment of future ocean observation and sensor networks. Applications range from oil prospecting and transportation to aquaculture, and include pollution control, climate recording, prediction of natural disturbances, search and survey missions, etc. The high attenuation of electromagnetic waves in underwater medium precludes them from being used as the information vehicle. Therefore, acoustic waves are the most viable alternative for the transmission. Nevertheless, the underwater acoustic channel is not free of drawbacks. On the contrary, it exhibits important obstacles that must be overcome if communication networks are to be implemented in the future. In addition, underwater systems have inherently serious problems of power supply and batteries duration. At the same time, network protocols must take into account the specific characteristics of the transmission medium, not dismissing the slow propagation speed. This talk will present an overall view of the acoustic channel characteristics and their impact on network protocols.

UNDERWATER ACOUSTIC NETWORKS

Dr. Javier Poncela



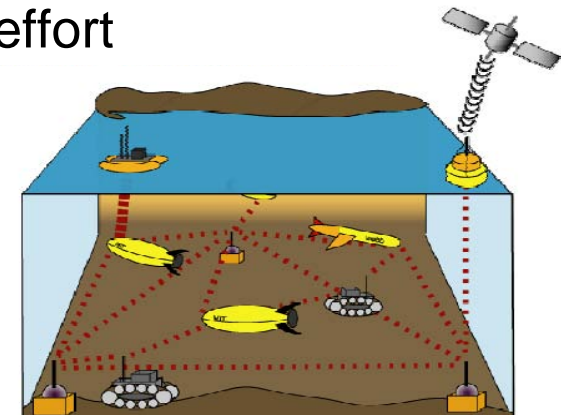
TELECOMMUNICATIONS TECHNOLOGIES
RESEARCH GROUP
UNIVERSITY OF MALAGA, SPAIN

Contents

1. Introduction
2. Underwater Acoustic Channel
3. Networks
4. Conclusions

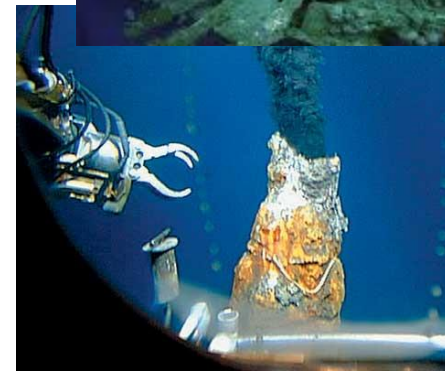
Introduction

- Oceans cover about 70 percent of the Earth's surface, and much of this vast resource remains to be explored
 - It is possible to chat from the International Space Station and make phone calls from the summit of Mount Everest, so why can't we check our email from the ocean floor?
- The volume below the sea surface has been traditionally ignored
 - It's a harsh environment that requires advanced technology
 - Resources are much easier to collect on the surface
 - Expansion has been possible without much effort
 - Even now, space resources look more tempting



UAC Applications

- Scientific
 - Submarine life monitoring
 - Natural phenomena forecasting
- Industrial
 - Aquaculture
 - Exploitation of mineral resources
- Environmental
 - Pollution control
 - Climate parameters recording
- Safety
 - Search and rescue missions,
 - Communication between divers and vehicles



Wireless Underwater Waves

- Traditionally, underwater communication is achieved via cables
 - Cables are expensive and heavy-weighted: several tens or hundreds of meters
 - Movement constraints for vehicles and divers
 - Safety issues as cables may pose dangers
- ➔ Wireless underwater communications is a must

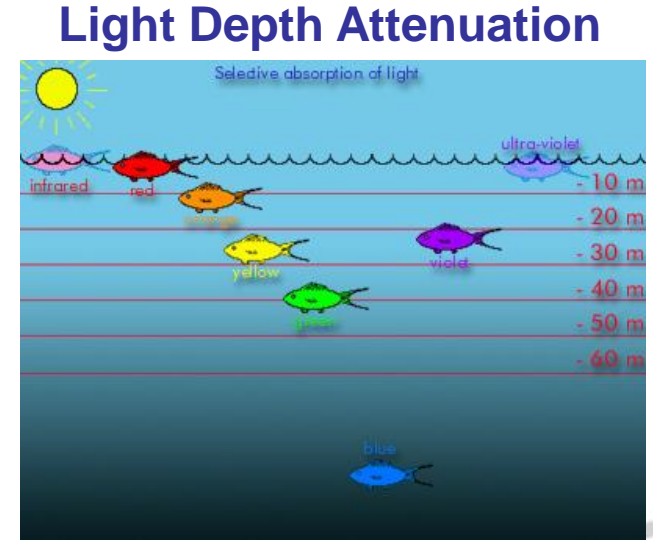


Wireless Underwater Waves

- Radio-Electromagnetic waves
 - EM waves do not travel well through thick electrical conductors like salt water
 - Strong absorption + Huge attenuation with distance

⇒ **Only for very short range communications**

- Optical communication
 - Blue-green region (450-550 nm)
 - + High bandwidth (~Mbps)
 - + Negligible delay
 - Short distance (<100 m)
 - Alignment of transmitter/receiver



Underwater Acoustics

- Used by submarine fauna
- Frequency range: 1 Hz - 500 kHz
 - A 30 kHz frequency (ultrasound) = 6 GHz in air (microwave)
(wavelength = 5 cm)
- Negative propagation characteristics
 - Limited bandwidth: 8kHz to 48-78 kHz
 - Time-varying multipath propagation:
Reflections from surface, sea floor
 - Low speed of sound underwater: ~1500 m/s



Contents

1. Introduction
2. Underwater Acoustic Channel
3. Networks
4. Conclusions

Underwater Acoustic Channel

- The underwater acoustic channel is affected by many factors
 - Salinity
 - Temperature
 - Seabed topology
 - Speed of sound
 - Surface wind-speed
 - ...
- This causes multi-paths, reverberation, Doppler, time-varying paths, ...
- The result: the communication channel has poor quality and high latency
 - Challenges are very different from terrestrial wireless

Sound Speed in Sea Water

- Propagation is realized through pressure waves
- Typical value: 1500 m/s
 - Range: 1450m/s – 1540 m/s
 - Time to travel 10 km: between 6.5s - 6.9s

- A simple model

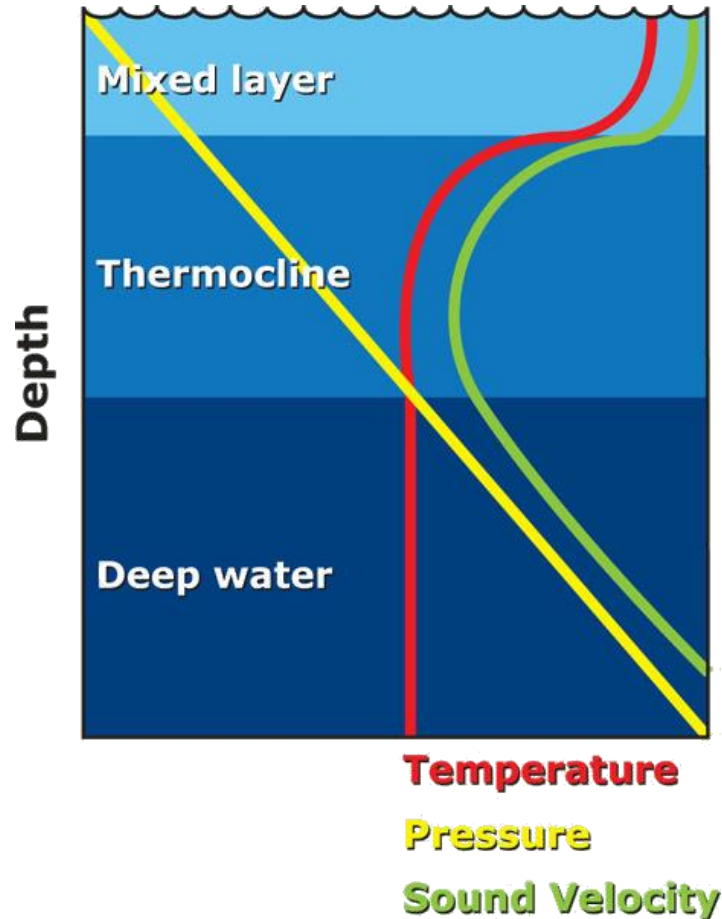
$$c = 1449.2 + 4.6T - 0.055T^2 + 0.00029T^3 + (1.34 - 0.01T)(S - 35) + 0.016z$$

T = temperature (°C), S = salinity (ppmil), z = depth (m)

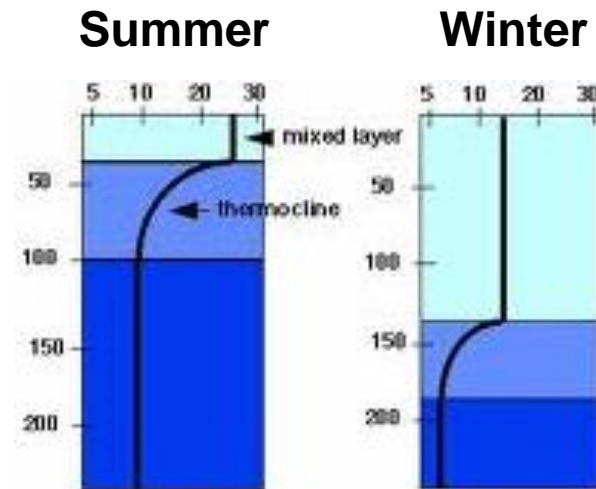
Valid for T ∈ [0°, 35°], S < 45‰, z < 1000m

➔ Speed increases with temperature, salinity and depth

Propagation Speed



- Near the surface, speed is constant
- As depth increases, speed decreases
- After 500-600 meters the increasing pressure causes an increase in speed



Attenuation

- Attenuation is strongly dependent on frequency
 - It causes a limitation in bandwidth as distance increases
 - Smaller bandwidth → Higher distances

$$A(l, f) = \left(l / l_{ref} \right)^k [a(f)]^l$$

k = 1 (cylindrical wave)
k = 2 (spherical wave)

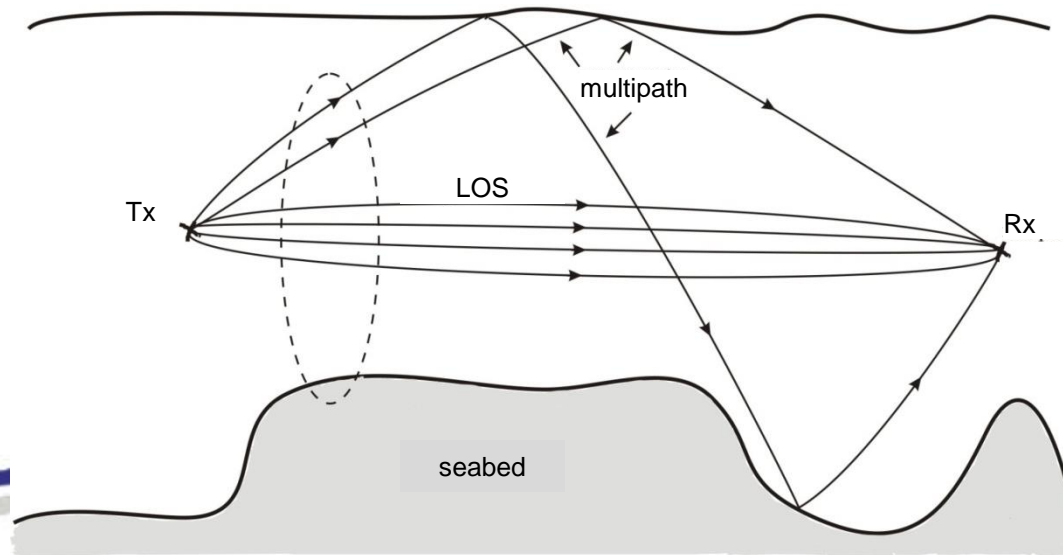
Geometrical Propagation Absorption



- Geometrical propagation Depth \ll range \implies cylindrical wave
 - The wavefront can be modeled as spherical (energy $\propto 1/R^2$) (k=2)
 - However, this spherical propagation has two limits: surface and seabed
 - ➔ Propagation is indeed cylindrical for long distances (energy $\propto 1/R$) (k=1)

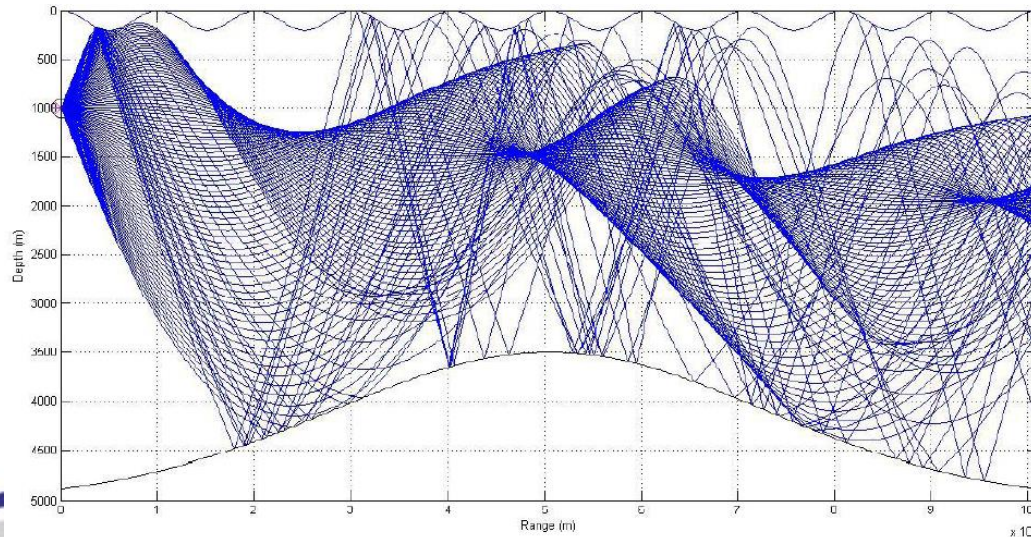
Scattering

- When the surface of the water is in movement, it causes a dispersal of the delays of the multiple reflections
- Time of coherence decreases
 - Temporal shift needed to decrease autocorrelation by 3 dB
- Experimental measurements show that scattering increases with frequency, distance and wind speed



Multipath and Fading

- When using ultrasounds, the height of the waveguide is several orders of magnitude bigger than the wavelength, and its physical modelling is quite simple using the ray theory
 - Reflections (seabed/surface) + Refractions cause multiple paths
 - Different paths cause scattering of the propagation delays



Doppler Effect

- Caused by movement of transducers and medium (turbulences, fauna, ships, ...)
- Coherence time is smaller
 - Low frequencies (≤ 5 kHz) \rightarrow seconds
 - High frequencies \rightarrow tenths of seconds (0.2s at 17 kHz)
- Causes Inter-Carrier Interference (ICI)
 - Compensation complex for Doppler dispersion higher than 3% the symbol period

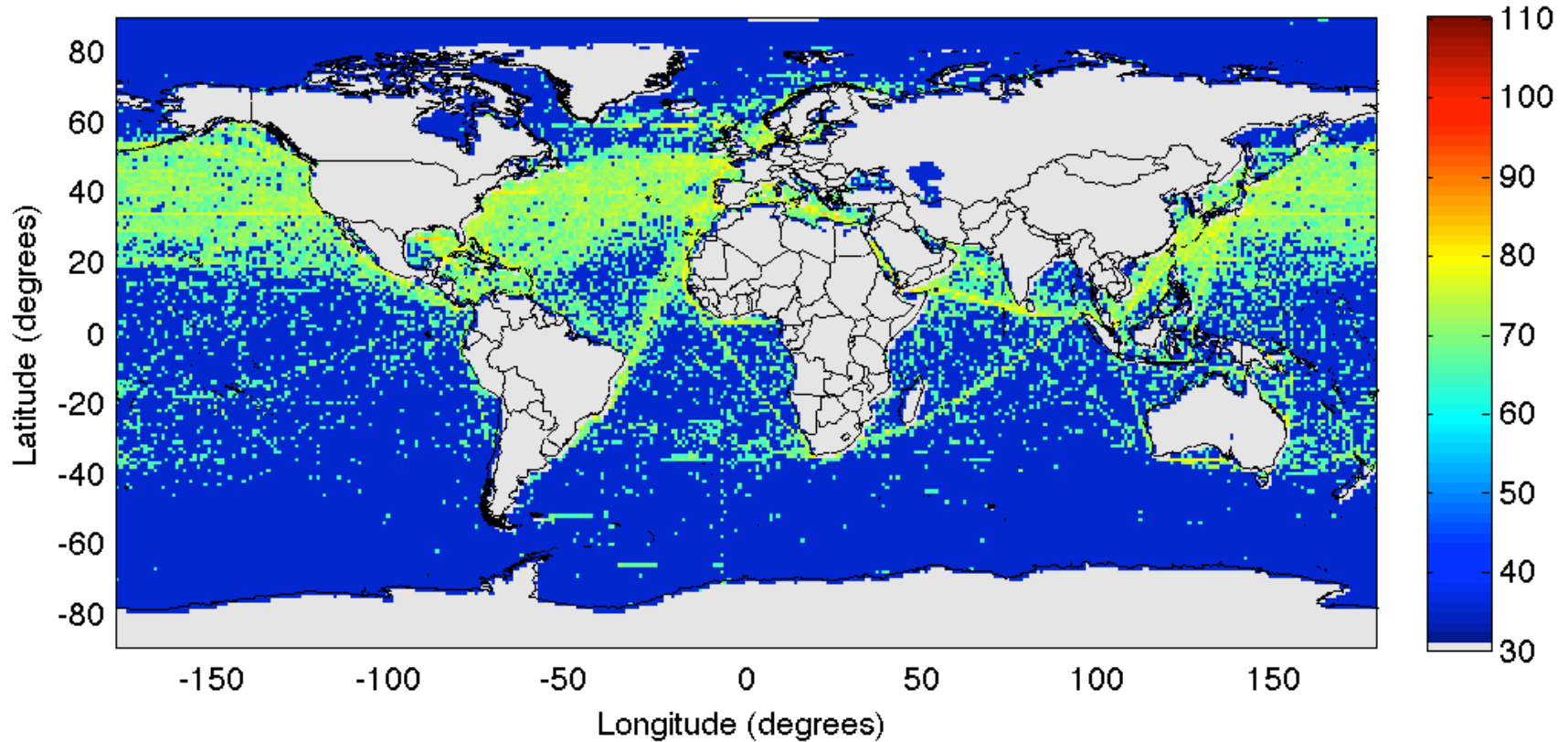


Bubbles are not So Funny

- Bubbles that appear on the surface may have a big influence on high frequency acoustic signals
- Effect: Increased attenuation of reflected signals
- Bubble density increases with wind speed
 - At 10 m/s, attenuation due to bubbles is up to 20 dB
- Bubbles underwater also create additional scattering



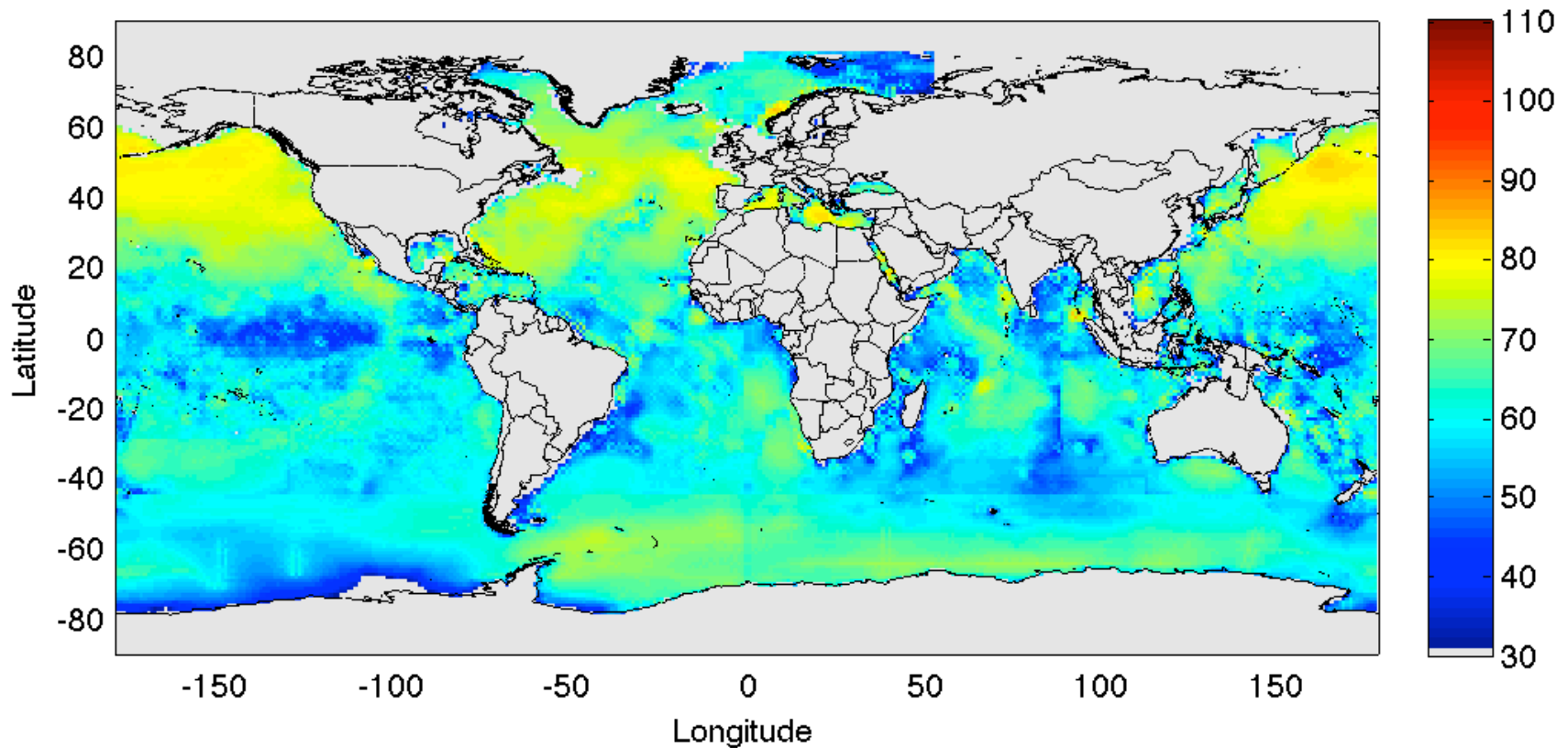
Submarine Environment



Global Shipping Noise at 200 Hz – Points of Origin

[*Ocean Acoustics Library*]

Submarine Environment



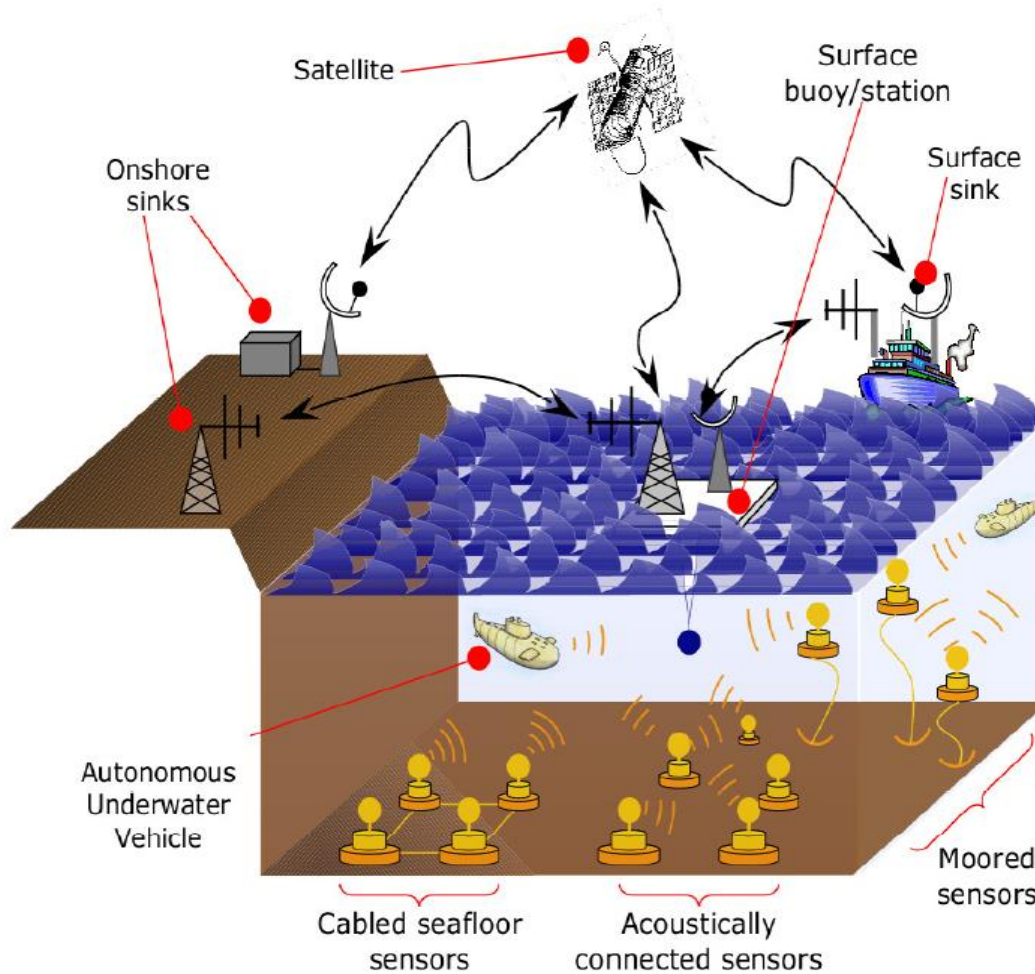
Global Shipping Noise at 200 Hz - Aggregate

[*Ocean Acoustics Library*]

Contents

1. Introduction
2. Underwater Acoustic Channel
- 3. Networks**
4. Conclusions

Network architecture: 3D

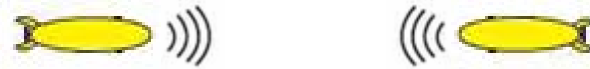


3D Networks [Stojanovic]

Networks

- Today: point-to-point acoustic links

- High delay
- Channel complexity

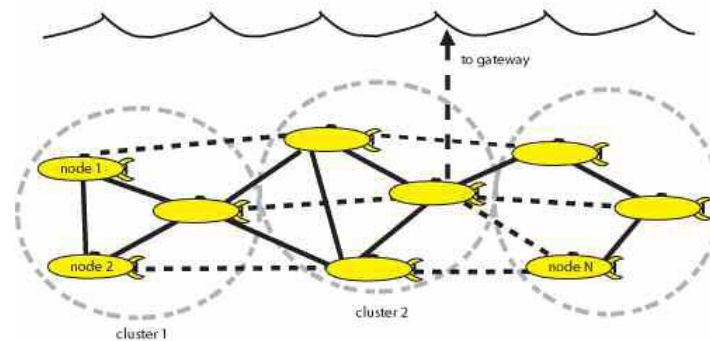


- ➔ Future: cooperative networks

- Fixed, slowly moving, mobile, sensors, relays, gateways

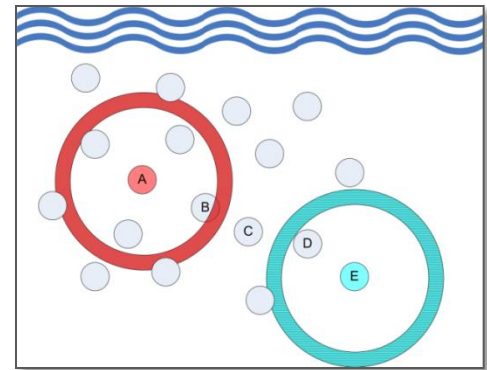
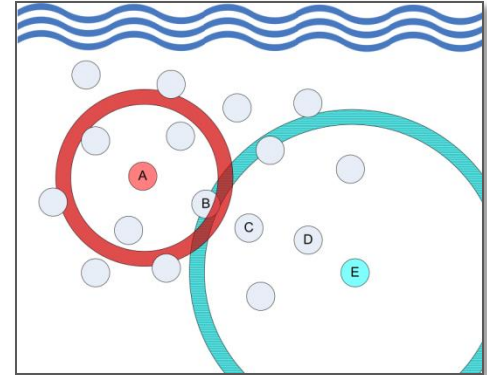
- Issues

- Shared access
- Efficient routing
- Low data rate
- Mobility



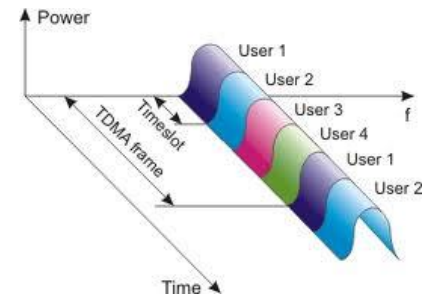
Shared Access

- Managing shared access
 - Collision detection
 - Turn sequence
 - Predefined channel allocation
- Unsynchronized protocols are simpler but explicit coordination can improve the performance
 - Requires a time reference
 - Space-time volume: overlap packets in time while they remain distinct in space



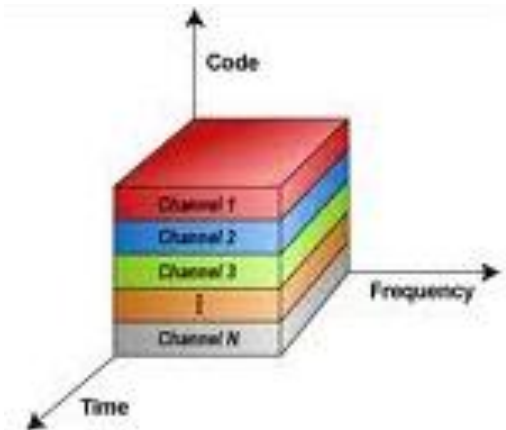
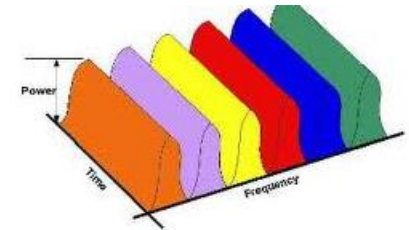
Shared Access – Static

- Static shared access
 - Nodes are allocated predetermined data channels
 - Contention-free (scheduled or deterministic protocols).
 - Problem: Inherently non-scalable.
- TDMA
 - Better performance in some aspects
 - Easy synchronization in nodes
 - More flexibility: number of allocated slots
 - Can be more flexible: number of allocated slots
 - Needs some coordination and some guard times to compensate for inconsistencies in dealing with propagation delays
 - Stable, high throughput at high loads by eliminating collision



Shared Access – Static

- FDMA
 - Inefficient for underwater applications
- CDMA
 - Causes a bandwidth expansion, especially acute in narrow channels
 - Advantages
 - No slot synchronization
 - Robustness to multipath fading
 - Power control
 - Disadvantages: complex receivers/transmitters

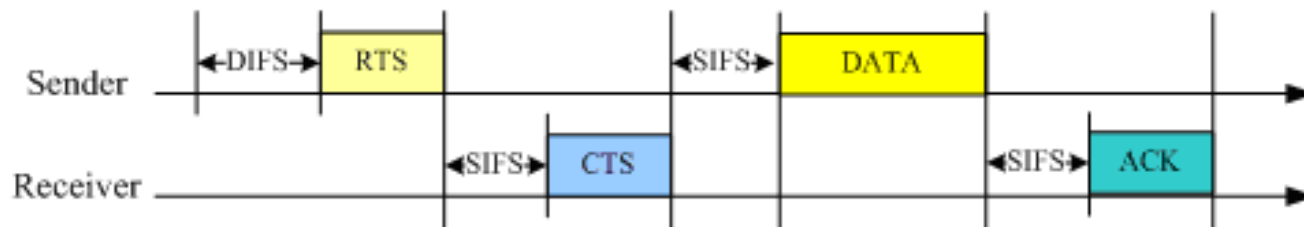


Shared Access – Dynamic

- Dynamic shared access (ad-hoc contention)
 - Nodes typically use a shared control channel over which data channels are requested
- Topologies
 - Distributed topology
 - No controlling master nodes
 - Nodes asynchronously handles data transfers
 - Dynamic MAC protocols are contention-based.
 - Centralized topology
 - A master node controls media access for nodes in neighborhood
 - May employ polling methods with no contention

Shared Access – Dynamic – Distributed Topology

- Good for scalability and no time synchronization needed
- RTS/CTS exchanges allow to measure the channel and to use optimal transmission parameters
 - Channel estimation, adaptive modulation, and power control
 - Poor performance due to latency
- Mobile AUV: DATA packet trains may pose problems
- Adaptive modulation and power control are key to maximizing channel efficiency/capacity



Shared Access – Dynamic – Distributed Topology

ALOHA

- Aloha
 - ALOHA based protocols are a good candidate for sparse low data rate networks
 - Candidate when combined with simple CSMA features
- However, Slotted Aloha degrades to pure aloha in environments with varying delay
- Proposed solution: PDT-Aloha (Propagation Delay Tolerant)
 - Add guard times to slotted ALOHA
 - Increases throughput by 17–100% compared to slotted ALOHA

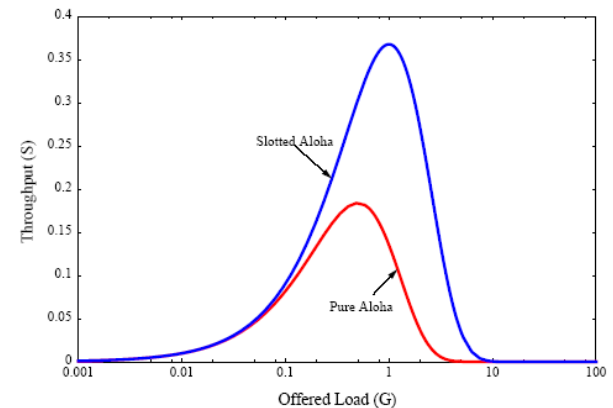


FIGURE 3.2: Throughput-Load of Pure and Slotted Aloha

Shared Access – Dynamic – Distributed Topology

CSMA

- CSMA
 - Listen-before-transmit approach
 - Requires propagation delay \ll packet duration
 - Latencies encountered make it very inefficient underwater
- Slotted MACA
 - Slot $>$ delay + RTS/CTS
- DACAP: based on MACA
 - Initial signaling exchange in order to reserve the channel
 - Adds a warning message if a RTS is overheard while waiting for a reply to its own RTS

Shared Access – Dynamic – Distributed Topology

CSMA

- T-Lohi
 - Nodes that want to transmit signal their intention by sending narrowband signals (tones)
 - The number of contenders is counted
 - Contention occurs → back-off in proportion to the contender count
 - Otherwise → Proceed with transmission
 - Achieves efficient channel utilization, stable throughput, and low energy consumption
 - Deafness and aggressive contention can cause the reservation mechanism to fail and lose packets

Shared Access – Dynamic – Centralized Topology

- Contention for requests
 - One Master allocates rights to send
 - One channel to request (shared channel), one channel to acknowledge, several data channels
 - Decrease in bandwidth
 - Delay for acknowledgment for sending because req/ack channels will be slow channels
- Contention-Free
 - Polling
 - Delay for transmission

Clustering

- Clusters
 - Cope with scalability issues
 - Good solution to enhance the network throughput performance
- Hierarchical organization using FDMA/CDMA for separating clusters and TDMA for intra-cluster
 - Spatial re-use
- Inter-cluster comms using one of the previous options

----- ALOHA - ALOHA+CDMA ×----- TDMA - CDMA, SF=1 △----- TDMA - CDMA, SF=16 ◇----- TDMA - CDMA, SF=32 -◇----- TDMA - FDMA -□----- TDMA - Opt. FDMA

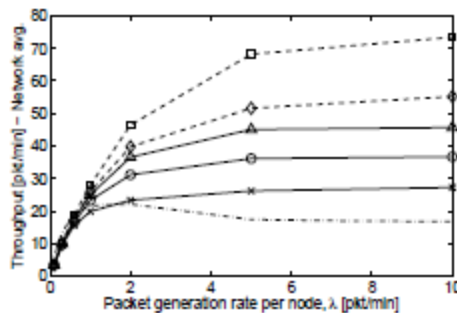


Fig. 1. Throughput (network average).

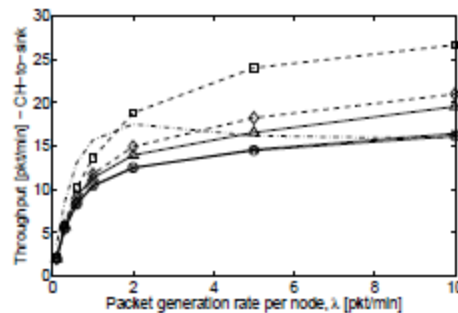


Fig. 2. Throughput (CH-to-sink).

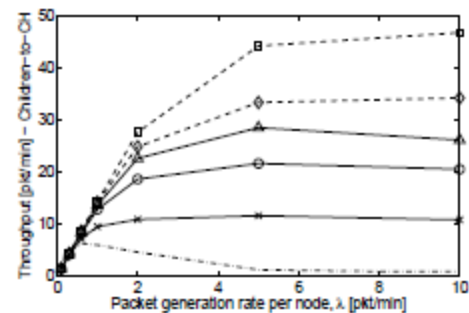


Fig. 3. Throughput (children-to-CH).

Data Link Layer Improvements

- Select the optimal packet size
 - Depends on range, rate-distance product, error probability, ...
 - Longer packets achieve better channel utilization
- Transmit packets in groups with selective acknowledgement
 - At data rates of 100 bps over 5 km links, $P_e = 10^{-3}$, packets of 256 bits, if we use groups of 16 packets, efficiency reaches 65%
- Adaptive adjustment of timeout based on measurements

Routing

- Routing overheads should be kept as minimal as possible
- In a typical clustered topology, gateways are used
 - The gateway node manages route discovery
- In distributed routing topology, all nodes perform routing
- Proposals for 3D scenarios
 - AODV-based routing
 - Location aware source routing for dynamic AUV networks
 - Energy-Efficient routing
 - Minimize the total path energy consumption by leveraging observations made on propagation characteristics of acoustic signals
 - Shortest path performs poorly ← chooses hops that are too long
 - Greedy minimum energy path also performs poorly ← ignores the advancement towards the destination

Localization

- Ranges are used to estimate local topology
- Localization has to be repeated periodically as the network topology changes due to the inherent motion of the vehicles.
- $T_{\text{error}} = 1 \mu\text{s} \rightarrow L_{\text{error}} = 15 \text{ mm}$
- ... but low number of messages is more important
- TDMA frame timings to compute ranges
 - Sufficient Distance Map Estimation (SDME) $\rightarrow \sim 1\text{m}$ at $d=139\text{m}$
 - A single moving reference beacon (high-precision clocks) \rightarrow standard deviation of 10–14m.
- Some proposals are beacon based

Efficient energy management: paramount importance

Huge propagation attenuation



High transmission power (tens of Watts)



Limited energy stored in electrical accumulators

Multihop networks



Solar energy not available



Electrical accumulators are critical system elements

Recharge or replacement of accumulators is an expensive and cumbersome process

Energy Efficiency

- Collisions should be avoided or/and detected soon
 - TDMA provides stable, high throughput
- Transmit energy costs are higher than reception ones
 - Impact on routing metrics
 - May be worth leaving nodes in an idle state and aware rather than sending them to sleep for some time
- Nodes can be waked-up by simple rx of an acoustic wave
 - Topology control: sleep but maintain network connectivity
 - However, sleep/wakeup schemes come at the cost of reduced bandwidth and additional time delays
 - long term underwater deployments
- Benefit of short-range communications

Cross-Layer

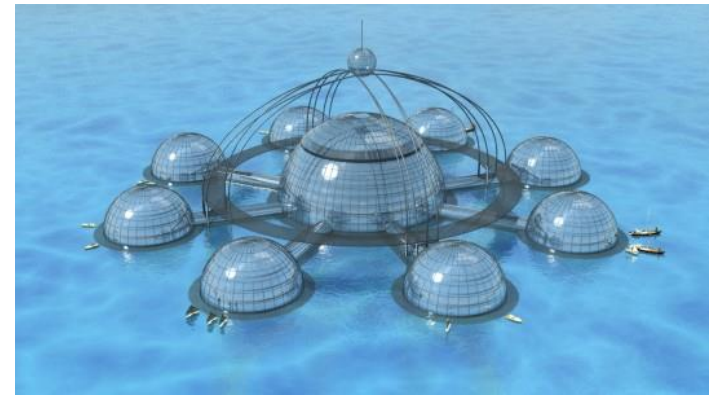
- Better data rates through cross layer optimization
 - Based on instantaneous measured link metrics
 - Based on application/service needs
- PHY + DLL + NWK
 - Adapt packet size, batch size and timers
 - Adaptive FEC code rate at the physical layer
 - Adaptive ARQ for time-varying channels
- APP + Lower Layers
 - Traffic patterns, performance requirements ↔ delay, throughput, reliability, energy efficiency

Contents

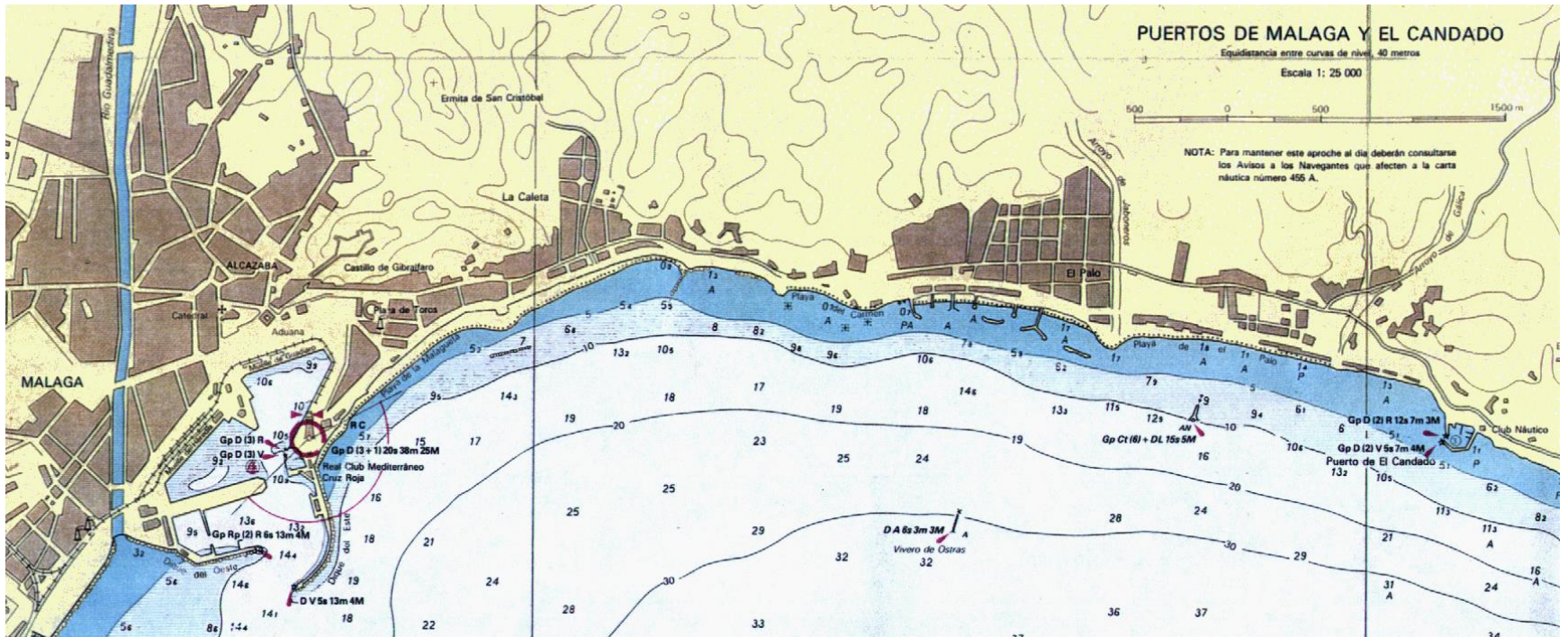
1. Introduction
2. Underwater Acoustic Channel
3. Networks
4. **Conclusions**

Conclusion

- There are no operational autonomous underwater networks, only isolated experimental demonstrations
- Underwater channel is very unfriendly
 - Long delays + Narrow bandwidth
- Terrestrial wireless techniques must be adapted
- Challenges
 - Channel modeling
 - Capacity of acoustic networks
 - Efficient and scalable channel sharing protocols



Bay of Malaga



UNDERWATER ACOUSTIC NETWORKS

Dr. Javier Poncela



TELECOMMUNICATIONS TECHNOLOGIES
RESEARCH GROUP
UNIVERSITY OF MALAGA, SPAIN