UNIVERSITAT POLITÈCNICA DE CATALUNYA

PROJECTE FINAL DE CARRERA

Caso práctico de informe experto de análisis forense digital

Autora:
Helena TARIBÓ GÓMEZ

Tutor:
Abraham PASAMAR

DEPARTAMENT DE TELEMÀTICA

7 de julio de 2016

”*The Truth Is Out There*”

The X-Files

Agradecimientos

Me gustaría agradecer al Josep, Abraham y Oriol la oportunidad de realizar este proyecto y de introducirme en el sector.

Un especial agradecimiento a Juanvi por todo lo que me ha enseñado, por su paciencia y por los buenos consejos.

Gracias a Javi por haber sido tan buen compañero durante todos estos meses de prácticas, por lo bien que lo he pasado mientras nos formábamos.

Gracias a todo el equipo de INCIDE por lo que he aprendido con ellos y por el buen ambiente que hay en la oficina.

Gracias a mis amigos por hacer que el estudio sea siempre más ameno.

Gracias a mi familia y a mi pareja por todo el apoyo que siempre me han dado y por la confianza que han depositado en mí.

Contents

1 Introducción

La proliferación en el uso de los dispositivos electrónicos como ordenadores y teléfonos móviles nos permiten estar conectados en todo momento y con todo el mundo. Las actividades cotidianas implican, cada vez más, el uso de Internet: realizar la compra, comunicarnos con los amigos, realizar consultas médicas, etc. Si bien facilita nuestras vidas también es una herramienta que debe usarse con responsabilidad y precaución. En Internet nos podemos encontrar con estafas de todo tipo, muchas de ellas basadas en el principio de impersonar a terceros de confianza, como la famosa estafa de los emails que simulaban ser de Correos. Los métodos cambian pero los delitos siguen siendo parecidos. No sólo con las estafas, antes, cuando un empleado decidía robar información confidencial de la empresa realizaba fotocopias del material deseado y se lo llevaba; ahora, se puede enviar los documentos por correo electrónico o copiarlos a un dispositivo USB. Es por este motivo que cada vez se realizan más juicios en los que intervienen elementos tecnológicos de última tecnología y, para ello, se debe conocer el funcionamiento de todos los elementos que implican estos dispositivos.

El objetivo de este proyecto es dar a conocer el proceso realizado durante una investigación digital. Es importante conocer todo el proceso y llevarlo a cabo correctamente para que el informe resultante sea válido, ya sea a nivel judicial o interno para una empresa. El proyecto cubre la investigación digital desde el primer momento, es decir, cuando se avisa de una posible incidencia, hasta el final, cuando se presenta el informe de resultados al cliente. Antes de llegar al punto de redacción de este proyecto, se ha estudiado el proceso a seguir en una investigación, no solo a nivel teórico y académico sino a nivel práctico, es decir, se ha realizado una investigación real en la empresa INCIDE (INCIDE Digital Data, S.L.). [9]

Estas prácticas se han divido en dos etapas: una primera etapa de formación en la UPC para aprender conceptos básicos; y una segunda etapa de formación en una

empresa para seguir aprendiendo mientras se realizan investigaciones reales.

1.1 Formación en la UPC

La experiencia empieza con dos meses de formación en la UPC bajo la tutela de Juan Vera y Josep Pegueroles, del Information Security Group [3] del departamento de telemática de la Universitat Politècnica de Catalunya. Durante esos días trabajamos conceptos básicos que nos ayudaron a familiarizarnos con una investigación digital. Empezando por la instalación de sistemas operativos Windows y Linux repasamos y aprendimos a configurar un equipo con distintos sistemas operativos. No sólo para ser capaces de realizar una instalación por nosotros mismos, sino con la finalidad de conocer las distintas opciones que existen y en que lugar se almacenan los ficheros de registro para poder consultarlos cuando se analiza un ordenador durante una investigación. Además de la configuración de un equipo, utilizamos herramientas para la creación y recuperación de particiones de disco. Aprendimos a realizar copias binarias y a optimizar la velocidad de la copia en función de la configuración del equipo y también distintos comandos para calcular un *hash* y a recuperar datos eliminados. También nos familiarizamos con el entorno Linux, empezando con comandos Bash y avanzando a *scripts* tanto en Bash como en Python porque la automatización de procesos es una pieza fundamental de las investigaciones forenses. Finalmente, reproducimos casos que habían sido realizados por el Information Security Group con anterioridad a modo de práctica.

1.2 Formación en INCIDE

Finalizada la formación en UPC, empezaron las prácticas en INCIDE. Dejar el ámbito académico para pasar al ámbito profesional supone enfrentarnos a casos reales que no han sido diseñados expresamente para los alumnos sino que son situaciones en las que no se puede mirar la guía de soluciones para llegar al final del caso y en los que no existe una única solución. INCIDE es una empresa especializada en el tratamiento e investigación de la información digital. Está encabezada por Abraham

Pasamar (CEO/CTO), que cuenta con más de 12 años de experiencia en el campo de la seguridad de la información. La diversidad de casos realizados en INCIDE permite tocar varios aspectos del mundo de la seguridad digital, como servicios forenses, monitorización, *e-crime*, etc. El caso presentado en este proyecto es el primer caso real que investigué durante las prácticas realizadas en INCIDE.

1.3 Estructura del trabajo

Con el objetivo de entender y explicar ordenadamente el proceso de una investigación digital, en el capítulo *Fases de un análisis forense* (apartado 2, pág. 5), se hará un breve repaso de las fases llevadas a cabo en una investigación. A continuación, en *Descripción del caso práctico a analizar* (apartado 3, pág. 15), se hará una introducción al caso práctico realizado. Para poder realizar una buena investigación, en *Funcionamiento de un disco duro* (apartado 4, pág. 19) y *Funcionamiento del correo electrónico* (apartado 5, pág. 37) se desarrollará la teoría necesaria que se aplica en el desarrollo del caso. Seguidamente, en *Herramientas usadas para el análisis* (apartado 6, pág. 49), se expondrán las herramientas usadas para realizar este análisis, porque son esenciales para optimizar el tiempo requerido y cumplir los objetivos en los plazos previstos. Para finalizar, en *Proceso de investigación* (apartado 7, pág. 57), se aplican todos los conocimientos adquiridos a un caso real encargado a INCIDE. Se incluyen las notas de la investigación y el informe de resultados. Se trata de un caso de fuga de información en una empresa por parte de cierto trabajador.

1.4 Objetivos

Los objetivos de este proyecto son los de:

- Entender y dar a conocer el procedimiento que debe seguir un analista o perito digital para llevar a cabo investigaciones digitales.
- Realizar una investigación de un caso real en una empresa (INCIDE). Los objetivos del caso presentado son los siguientes:

- Extraer todos los fichero que, en su denominación, contengan el nombre PALABRA_A o PALABRA_B.
- Extraer todos los correos electrónicos enviados o recibidos entre DIRECCIÓN_1 y DIRECCIÓN_2.
- Verificar que los correos electrónicos extraídos son íntegros.
- Extraer todos los ficheros eliminados entre FECHA_INICIAL y FECHA_FINAL.

2 Fases de un análisis forense

En este capítulo se hará una explicación breve de las distintas fases por las que pasa una investigación forense digital, desde el momento en el que el investigador es requerido hasta el momento en el que este entrega el informe de resultados. No hay un modelo estándar que dicte todos los pasos a seguir sino que a lo largo de más de 20 años se han ido proponiendo diversos modelos. Todos ellos remarcan la importancia de preservar las pruebas y mantener la cadena de custodia. Los casos con los que se puede enfrentar un investigador son de tipología muy distinta y no todos las etapas se pueden aplicar en cada investigación. Así pues, los pasos descritos a continuación son unas directrices y no una normativa. En concreto, se ha elegido el modelo propuesto por Michael Donovan Köhn en 2012 [5] como referencia porque incluye la fase de preparación que, aunque no implica directamente al investigador, se ha considerado relevante que las personas estén mínimamente preparadas por si se produce algún percance. Según el tipo de incidente, la preparación podría ser crucial para una investigación, tanto para la empresa como para el investigador.

Figure 2.1: Fases por las que pasa un análisis forense digital.

2.1 Definiciones

Antes de explicar las fases, es conveniente conocer algunos conceptos básicos del análisis digital forense como son *digital forensics*, *digital evidence* y cadena de custodia.

Digital Forensics

El análisis forense digital o *digital forensics*, es una ciencia que utiliza métodos científicos probados y basados en un fundamento legal sólido para preservar, recoger, validar, identificar, analizar, interpretar y presentar la información que proviene de dispositivos digitales. El objetivo de un análisis forense es el de aportar información válida en un proceso legal para validar o refutar hipótesis ante los tribunales de justicia.

En 2012 se publicó la ISO/IEC 27037:2012 — *Information technology — Security techniques — Guidelines for identification, collection, acquisition, and preservation of digital evidence* [2] en el que se establecen unas directrices básicas para el tratamiento de potenciales pruebas digitales. En Europa estos principios se agruparon

en las denominadas *buenas prácticas*, *good practice* en inglés, en el que el Consejo de Europa establece los principios básicos para el manejo de evidencias digitales. Los principios básicos recogidos en el documento son la integridad, auditabilidad, soporte de expertos, formación adecuada y legalidad.

- Integridad: ninguna acción puede cambiar los medios de prueba.
- Auditabilidad: todo el proceso debe estar documentado y a disposición de terceras partes que puedan repetirlo.
- Expertos: presencia de especialistas en las intervenciones.
- Formación adecuada: los especialistas en recogida de evidencias deben tener una formación adecuada.
- Legalidad: todo el proceso debe ajustarse a la legislación.

Los principios recogidos por las buenas prácticas deben aplicarse en todas las investigaciones, adaptándose a las circunstancias particulares de cada caso. En caso de que no existan procedimientos establecidos para una tecnología concreta a analizar, el perito debe definir y anotar el procedimiento, justificando cada paso y el objetivo al que pretende llegar al llevar a cabo dicha práctica, de forma que se cualquier otro investigador sea capaz de repetir el mismo procedimientos y obtener los mismos resultados.

Digital Evidence

Según el Departamento de Justicia de los Estados Unidos de America [7], *digital evidence* son datos e información de valor para una investigación guardada, recibida o enviada mediante un dispositivo electrónico. Esta prueba es adquirida cuando los datos o el dispositivo electrónico se incauta y asegurada para ser examinado.

Cadena de custodia

La cadena de custodia es el procedimiento que controla los indicios relacionados con un delito, desde su localización, descubrimiento o aportación hasta que ha sido

analizado y valorado y la autoridad competente ordene su conclusión. El objetivo de la cadena de custodia es el de no alterar o sustituir las potenciales pruebas de un delito. Una prueba que ha sido hallada sin haber sido establecida la cadena de custodia del medio que la albergaba no tiene validez en un tribunal pues no se puede asegurar que estos datos no han sido modificados. De la misma forma que en una escena de un asesinato se debe tener especial cuidado de no tocar nada hasta que se hayan extraído las huellas dactilares, en el campo del análisis forense digital no se puede acceder a los datos hasta que el dispositivo que alberga los datos no ha sido asegurado. La cadena de custodia debe establecerse, siempre que sea posible, en presencia de un fedatario público y éste debe quedarse una copia en custodia para garantizar el derecho a la defensa de la otra parte.

Es fundamental que los analistas presenten los medios de prueba cumpliendo estas características:

- Admisibles: deben cumplir con todas las garantías legales.
- Auténticas: vinculadas directamente con los datos disponibles en una investigación
- Completas: deben proporcionar información suficiente y objetiva.
- Confiables: se deben mantener la integridad de los datos en los que se apoyan las conclusiones.
- Verosímiles: presentadas de forma clara y razonada.
- Proporcionales: adecuado a los fines perseguidos.

2.2 Documentación

La documentación es una fase o procedimiento que debe seguirse durante toda la investigación. Consiste en anotar todas las acciones realizadas y también los resultados obtenidos. Llevar un registro detallado es útil para la redacción del informe de resultados y para que otro investigador pueda seguir nuestro trabajo. También sirve para demostrar que se han seguido los principios de integridad y auditibilidad

recogidos en las buenas prácticas forenses, es decir, no alterar el material original bajo investigación y documentar las acciones realizadas para que terceras personas puedan llegar a las mismos resultados. Asimismo, todas las notas registradas son de utilidad para el propio investigador, para ser consciente de las acciones realizadas antes de, por ejemplo, ir a juicio a ratificar el informe pericial. Cuanto más detalladas sean las notas, más fácil será la redacción del informe.

2.3 Preparación

La fase de preparación, aunque podría no considerarse parte de un análisis forense digital, es importante porque es el primer paso a seguir en el momento observar una incidencia. El objetivo de la fase de preparación es el de tener una serie de normas o políticas de empresa que se deberían seguir en caso de detectar un posible incidente. Estas normas sirven para maximizar la conservación de pruebas a la vez que se minimiza el coste de la investigación para la empresa afectada. Por ejemplo, en caso de recibir un posible ataque a un servidor, el departamento de seguridad y el de producción podrían tener ideas distintas sobre cómo reaccionar. El departamento de seguridad desearía saber qué y cómo ha sucedido para corregir las vulnerabilidades; mientras que lo importante para el departamento de producción sería volver a tener el servidor en funcionamiento lo antes posible. Con unas buenas instrucciones se debería impedir que se perdiesen pruebas potenciales debidas a la mala actuación por parte de los empleados, como podría ser apagar todos los servidores de golpe.

2.4 Incidencia

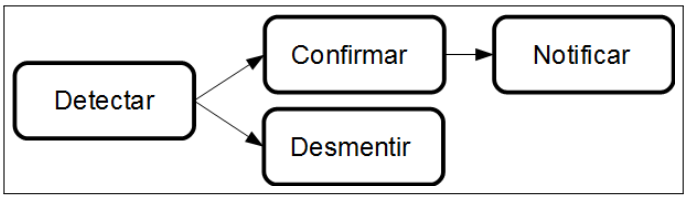

Figure 2.2: Fases de incidencia.

El primer paso de esta fase es detectar la incidencia, ya sea de manera automática o manual. Una incidencia es una acción llevada a cabo con el objetivo de compro-

meter la confidencialidad, disponibilidad e integridad de la información. Una vez detectada, ésta necesita ser evaluada para ser confirmada o desmentida. En caso de confirmar la incidencia, se debe notificar a los investigadores y a las autoridades pertinentes para empezar la siguiente fase, la de respuesta a la incidencia. En caso de desmentirla se trata de un falso positivo. Siguiendo con el ejemplo anterior, sobre un intento de acceso no autorizado al servidor de una empresa, dicho acceso generaría una notificación al encargado de sistemas. Esta persona es la que debería comprobar si se trata de un acceso controlado, como podría ser un acceso por parte de un empleado que se conecta en remoto o un empleado que ha introducido de forma incorrecta el usuario o contraseña; o, por contra, si se trata realmente de un intento de acceso malintencionado, ya sea dirigido específicamente a dicho servidor o por parte de escaneos masivos en busca de vulnerabilidades.

2.5 Respuesta a una incidencia

La fase de respuesta empieza con la llegada de los investigadores, que tienen que ser capaces de analizar la situación y definir una estrategia a seguir. Deben establecer y mantener la cadena de custodia, preservar las pruebas potenciales y llevar un registro de todos las acciones que realicen. Es importante aclarar las cuestiones que puedan ser relevantes para la investigación antes de llevar a cabo cualquier acción. Cuestiones como con qué frecuencia se realizan copias de seguridad, qué usuarios tienen acceso a un determinado dispositivo o si el correo electrónico se almacena en local o en remoto pueden ser vitales para una investigación. Se hará una clonación de los dispositivos *in situ* o en presencia de agentes judiciales según el tipo de investigación. Esta copia tiene que ir acompañada de su correspondiente *hash*, para asegurar la cadena de custodia de todos los datos. Un *hash* de un fichero es un cálculo que da como resultado una combinación de números y letras. Tiene la peculiaridad de que cualquier cambio en la información, por pequeño que sea, altera totalmente el resultado del *hash*.

2.6 Análisis forense digital

Esta es la fase en la que el investigador tiene que analizar todos los dispositivos que forman parte de las denominadas fuentes de información para dar respuesta a los objetivos de la investigación. Los objetivos suelen responder a qué ha ocurrido y quién es el responsable o autor de los hechos ocurridos y que han llevado a la investigación. Se puede dividir en distintas etapas, que se explican a continuación.

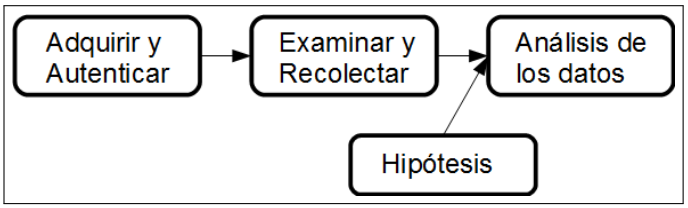

Figure 2.3: Fases de análisis forense digital.

2.6.1 Adquirir y autenticar

El primer paso consiste en tomar posesión de los dispositivos a analizar. Siguiendo las buenas prácticas forenses, lo primero que se debe hacer es realizar una copia bit-a-bit del contenido de los dispositivos a analizar, ya sea un ordenador, un teléfono móvil, un buzón de correo o el contenido de una página web. Al realizar la copia nos aseguramos de disponer de todo el contenido y dejamos el dispositivo original intacto. El dispositivo original (o una segunda copia) debe quedarse bajo custodia para garantizar el derecho a la defensa por cualquier otra parte que quiera ejercer este derecho. A cada copia se le calcula el correspondiente *hash*, que es una operación matemática que da como resultado una secuencia de números y letras. Si se realiza una modificación del contenido del dispositivo, por pequeña que sea, el valor del *hash* variará sustancialmente. El *hash* calculado también quedará guardado bajo custodia. Generalmente se usa el cálculo del *hash MD5* o *SHA256*.

2.6.2 Examinar y recolectar

Es en este momento en el que se empiezan a tratar todos los datos reunidos. En esta etapa el investigador debe tener claro cuál es su objetivo para, a partir de la copia

realizada, sacar todos los datos que puedan ser relevantes para la investigación. Esto puede incluir la recuperación masiva de ficheros, la extracción de lineas temporales del sistema de ficheros, la extracción de las cadenas del disco duro, el *parseo* del contenido del dispositivo o la extracción de todos los correos de un buzón de correo electrónico. Aunque las circunstancias de cada caso son distintas, estás acciones suelen estar automatizadas para facilitar el trabajo del investigador y evitar errores humanos.

2.6.3 Análisis de los datos

Una vez se han extraído los datos del dispositivo llega el momento de analizarlos. Es recomendable separar los datos que nos pueden interesar de los que no son importantes para la investigación porque, en general, se trabaja con grandes cantidades de datos y nos interesa ahorrar tiempo y recursos. Una vez identificados los datos relevantes (por contener potenciales pruebas), estos deben ser catalogados y organizados para realizar un análisis detallado de todos ellos. El investigador no debe olvidarse de llevar un registro de acciones y resultados. Además, si se dispone de un registro de incidentes previos, se pueden comparar los indicios obtenidos con investigaciones anteriores como método de soporte a la investigación. Durante el proceso de análisis, el investigador debe formular una hipótesis y tratar de probarla o de refutarla con evidencias o por falta de ellas. Si las pruebas desmienten la hipótesis inicial, debe formularse otra hipótesis hasta que los datos obtenidos la respalden o hasta que no haya indicios que sugieran lo contrario.

Una vez más, en función del tipo de investigación se tendrá más o menos información para analizar. Por ejemplo, si el objetivo del análisis es encontrar un tráfico sospechoso detectado en cierto ordenador, se analizarán los registros o *logs* del sistema en busca de rastros que indiquen una conexión externa o indicios que apunten hacia la anomalía y, por otra parte, también se pueden intentar reproducir las mismas condiciones que llevaron a ese tráfico con el fin de detectar su procedencia. Otro ejemplo puede ser una investigación enfocada a buscar determinados correos electrónicos y probar su autenticidad. En este caso el investigador se centrará en analizar los gestores de correo electrónico del dispositivo y el contenido residual de

las consultas a Internet con el fin de encontrar dichos correos electrónicos.

2.7 Presentación

Finalmente, se tiene que presentar por escrito un documento de resultados. Este documento debe incluir los procedimientos seguidos desde el establecimiento de la cadena de custodia hasta las conclusiones a las que se ha llegado y un análisis que interprete los resultados obtenidos. Tanto si forma parte de un proceso judicial como si se trata de una investigación interna de una empresa, el informe debe contener explicaciones dirigidas a personas sin conocimientos técnicos, como un juez o el directivo de una empresa. Es decir, un buen informe no puede incluir solamente un serie de *logs*, sino que estos deben presentarse acompañados de una interpretación para que tenga sentido para personas sin altos conocimientos técnicos. Sin embargo, tampoco es correcto un informe que carezca del contenido que respalda las conclusiones a las que ha llegado el investigador. Finalmente, si el informe es pericial, el perito tendrá que ir a ratificar el informe ante el juez.

Cada caso es distinto al anterior; por esta razón, las fases que se acaban de explicar en *Fases de un análisis forense* (apartado 2, pág. 5) pueden no ser aplicables en todas las investigaciones pero sí que son una guía que se debe tener en mente. A modo de resumen se destaca el aseguramiento de las fuentes de información y el mantenimiento de la cadena de custodia así como la importancia de escribir unas detalladas notas durante todo el caso. La investigación debe llevarse a cabo siguiendo las buenas prácticas forenses y el informe debe incluir un análisis de los resultados y explicaciones sin un elevado contenido técnico.

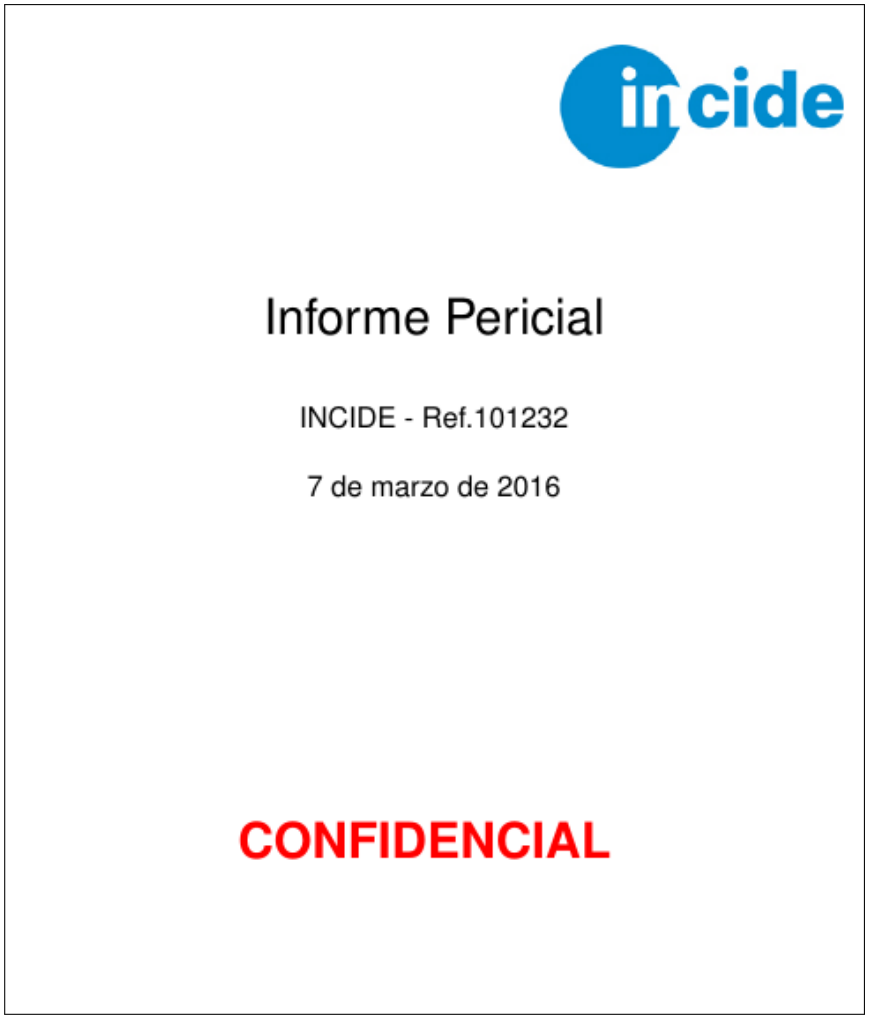

Figure 2.4: Detalle de la portada de un informe pericial realizado por INCIDE.

3 Descripción del caso práctico a analizar

A continuación se explicarán las etapas que, de forma general, se aplican en INCIDE en la gestión de cualquier caso de prueba electrónica. Estas etapas son el resultado de años de experiencia, que han permitido aplicar una metodología que se ajusta a las buenas prácticas forenses descritas en el apartado *Fases de un análisis forense* (apartado 2, pág. 5) y que resultan en la emisión de un dictamen con los resultados de toda la investigación. Antes de entrar en detalle, se introducirá el caso realizado en INCIDE así como los conceptos que han permitido una buena resolución del mismo.

La primera toma de contacto con un caso siempre es a través de la persona que contrata a INCIDE, ya sea personalmente o a través de un abogado. A partir de uno o varios encuentros con el cliente, se realiza un estudio preliminar para comprender sus necesidades, definir los objetivos de la investigación e identificar los elementos relevantes para la investigación, como por ejemplo un ordenador, un teléfono móvil o la información publicada en cierta página web. A continuación, se suele concertar cita con el notario para asegurar todos los elementos bajo análisis y, así, establecer la cadena de custodia. El notario custodiará una de las copias realizadas para garantizar el derecho a la defensa por cualquier parte involucrada en el proceso judicial. En el caso de los contenidos de una página web, se suelen realizar capturas de pantalla del contenido bajo investigación.

En el caso que se explica en este proyecto, se analizaron cinco ordenadores. Para anonimizar el caso no se utilizaran nombres propios, fechas concretas ni lugares. De esta forma, se protege la intimidad de todas las personas y empresas que formaron parte de la investigación y, además, se cumple con el contrato de privacidad firmado con INCIDE antes de empezar a realizar las prácticas en esta empresa.

En este caso, el Cliente contactó a través de su abogado y ambos estuvieron presentes en la reunión realizada con INCIDE para realizar el estudio preliminar. El Cliente explicó a INCIDE que había sido denunciado por su Empresa por irregularidades detectadas en las cuentas de la empresa. En concreto, por alteración de los precios en las facturas con un proveedor con el que él se encargaba de gestionar los pedidos. La Empresa aportó cinco ordenadores que habían sido asignados al Cliente durante su periodo laboral en la Empresa y con los que él realizaba su labor en la empresa. Toda esta información ya había sido recogida en las diligencias previas número XXX del juzgado de instrucción número 1 de XXX, proceso que acababa de ser abierto y que estaba en manos del juez de instrucción. El cliente solicitó los servicios de INCIDE para que realizase un análisis cuyos objetivos finales eran los mismos que habían sido autorizados para llevar a cabo por la Policía con el fin de que su abogado tuviera el máximo de información posible para encarar la defensa del Cliente. La adquisición fue llevada a cabo en una notaría de Barcelona, realizando dos copias de cada disco, una para el análisis de la Policía y la otra copia para el análisis de INCIDE, que actuaba como mandatario verbal del Cliente. Los discos duros originales quedaron depositados en sede notarial.

Los objetivos que dictaminó el juez de instrucción para el análisis de los discos son los siguientes:

- Extraer todos los fichero que, en su denominación, contengan el nombre PALABRA_A o PALABRA_B.
- Extraer todos los correos electrónicos enviados o recibidos entre DIRECCIÓN_1 y DIRECCIÓN_2.
- Verificar que los correos electrónicos extraídos son íntegros.
- Extraer todos los ficheros eliminados entre FECHA_INICIAL y FECHA_FINAL.

Para alcanzar los objetivos de este caso, en el siguiente capítulo (*Funcionamiento de un disco duro* (apartado 4, pág. 19)) se explicará el funcionamiento de un disco duro, haciendo especial hincapié en el sistema de ficheros NTFS utilizado por Windows. En el capítulo *Funcionamiento del correo electrónico* (apartado 5, pág. 37) se

explicará el funcionamiento de un correo electrónico, que es esencial para la resolución del caso. En el capítulo *Herramientas usadas para el análisis* (apartado 6, pág. 49) se repasan las herramientas usadas en el caso.

4 Funcionamiento de un disco duro

Esta sección explica el funcionamiento de los discos duros, cómo son y cómo se almacena y recupera la información. Se centra en el sistema de ficheros NTFS, que es el que utilizan los sistemas operativos Windows.

4.1 Geometría de un disco

Un disco duro está formado por varias capas circulares o discos uno encima del otro y que giran a la vez. Cada disco tiene dos caras que están recubiertas de un medio magnético y, para cada cara, hay un cabezal magnético que es el encargado de leer los bits. Un disco se puede dividir en pistas, cilindros y sectores. Las caras se dividen en pistas concéntricas, el conjunto de pistas que ocupan la misma posición en todas las caras se denomina cilindro y, un sector, que tiene un tamaño de 512 bytes, es la unidad de división de una pista. En la siguiente imagen se puede ver un dibujo de la estructura de un disco duro.

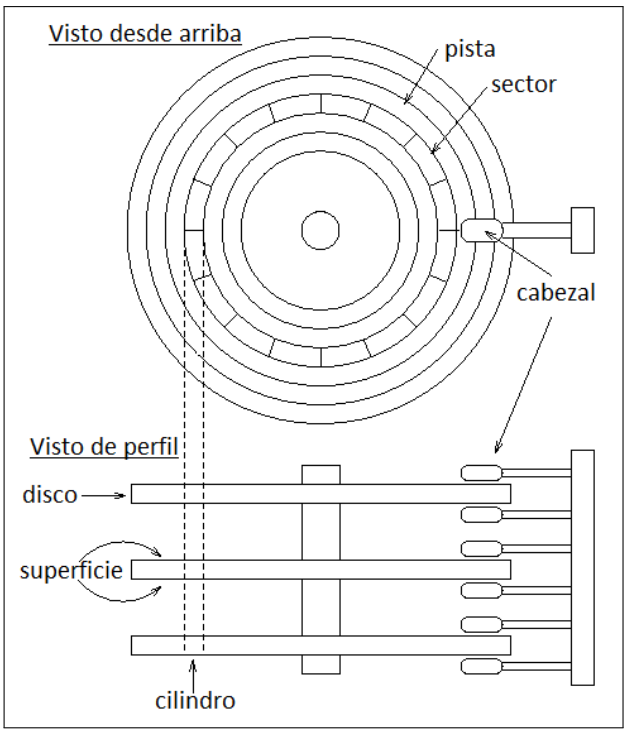

Figure 4.1: Esquema de un disco duro.

Los discos duros se conectan al ordenador mediante el controlador ATA (*Advanced Technology Attachment*) o las evoluciones del controlador, como *Parallel ATA* y *Serial ATA*. SATA es una interfaz que conecta el adaptador de bus del *host* con los dispositivos de almacenamiento como los discos duros. También existen discos duros que se conectan mediante USB o *Serial Attached SCSI* (SAS). Los discos más comunes son los de 3.5 pulgadas para ordenadores de sobremesa y 2.5 pulgadas para ordenadores portátiles.

Las ventajas principales que ofrece SATA respecto a sus predecesores son la del aumento de velocidad de transmisión (mayor ratio de señal) y mayor eficiencia (cola input output opcional), reducción del coste y del tamaño del cable que pasó de utilizar 40-80 conductores a utilizar sólo 7.

En un disco duro convencional los bits de información se escriben uno a continuación de otro, de forma secuencial, llenando *clusters* de datos. En cambio, en los dispositivos de almacenamiento de datos flash, como los USB o los discos SSD, los datos se guardan en bloques no consecutivos porque las memorias flash tienen un número limitado de ciclos de borrado/reescritura y de esta forma se reparte el uso de la memoria. El tipo de dispositivo debe tenerse en cuenta cuando se recuperan datos,

porque en caso de un disco convencional se pueden seguir los bloques e ir recuperando datos pero en un dispositivo SSD los ficheros pueden quedar fragmentados en bloques muy dispares.

A continuación se hará un resumen de los tipos de partición más usados.

4.2 Particiones DOS

Microsoft denomina *Master Boot Record* (MBR) a los discos que utilizan las particiones DOS. También existen los discos *GUID Partition Table* (GPT) que se utilizan con *Extensible Firmaware Interface* (EFI). A partir de Windows 2000, Microsoft también distingue entre discos básicos, que son discos con MBR o GPT en los que las particiones del disco son independientes y *stand-alone*; y discos dinámicos, que son aquellos discos de tipo MBR o GPT en los que las distintas particiones se pueden entrelazar para formar particiones de mayor tamaño. La información expuesta a continuación se refiere a discos MBR básicos, aunque por brevedad se utilizará el termino partición DOS sin hacer más distinciones. Las particiones de tipos DOS se utilizan en distintos sistemas operativos, como por ejemplo Microsoft Windows y Linux. Las particiones DOS contienen una MBR en el primer sector, que es de 512 bytes y puede tener hasta 4 particiones. Cada partición es una entrada en la tabla, y estas, a su vez, contienen campos con información sobre dónde empieza y acaba cada partición así como el tipo de partición y un *flag* que indica si la partición es *bootable* o no.

La limitación de las 4 particiones se puede solucionar mediante las particiones extendidas. Es decir, la tabla de particiones almacenada en una MBR puede llegar hasta 4, para crear una partición extendida se crearán hasta 3 particiones primarias no extendidas y una partición primaria extendida.

4.3 Particiones Apple

Apple no tiene un número limitado de particiones y las *data structures* se almacenan en sectores consecutivos del disco. Las particiones Apple se describen en el mapa de particiones situado al inicio del disco. El *firmware* que contiene el código procesa la

estructura; por ese motivo, el mapa de particiones no lleva código de arranque como las tablas de particiones DOS. En cada entrada del mapa de particiones se describe el sector inicial de la partición, el tamaño, el tipo y el nombre del volumen. Apple crea particiones para guardar los *drivers* del *hardware*; por esta razón, el disco principal de un sistema Apple está formado por diversas particiones con *drivers*. Un fichero de imagen de disco Apple es muy similar a un fichero de tipo *zip* en Windows o *tar* en Unix.

Para poder identificar las particiones, se tiene que leer la estructura de datos del segundo sector. La herramienta `mmls` del programa The Sleuth Kit nos mostrará las particiones, sin embargo, la herramienta `fdisk` de Linux no mostrará los contenidos de la partición del mapa. El programa The Sleuth Kit se explicará en el apartado *Herramientas usadas para el análisis* (apartado 6, pág. 49)

4.4 NTFS

A continuación, se hará una explicación del sistema de ficheros NTFS. Es importante conocerlo porque es el utilizado por Windows, el sistema operativo más extendido.

NTFS, New Technologies File System, es el sistema de ficheros diseñado y utilizado por Windows desde el año 1993. Fue concebido para ser escalable y para utilizarse en sistemas con discos duros de mayor capacidad, 400 MB en aquella época. La escabilidad viene dada por la estructura de datos interna que se adapta mientras que el envoltorio general se mantiene constante. Con envoltorio general nos referimos, por ejemplo, a que cada byte de datos del sistema se asigna a un fichero. Este concepto es muy importante porque lo diferencia de otros sistemas de ficheros. En NTFS todos los bytes de datos forman parte de ficheros, incluyendo los datos administrativos del file system y que pueden estar asignados (*allocated*) en cualquier sitio del disco. Los únicos datos que tienen un emplazamiento predeterminado son el sector de arranque (*boot sector*) y el código de arranque, que se encuentran en los primeros sectores del volumen. Por este motivo la tabla maestra de ficheros (MFT) es tan importante para el sistema de ficheros NTFS.

4.4.1 MFT

La *master file table* contiene información sobre todos los ficheros y directorios del sistema. Cada fichero y directorio tiene, por lo menos, una entrada en la tabla, incluyendo la propia MFT.

Las entradas de la MFT ocupan 1 KB, los primeros 42 bytes están destinados a 12 campos con información como la signatura, en la que hay la palabra 'FILE' escrita en ASCII en una entrada estándar; un campo que indica el número de secuencia, otro que indica el tamaño *allocated* de la entrada MFT, etc. Los bytes restantes se usan para almacenar atributos, que son pequeñas estructuras de datos usadas para un fin concreto, como el nombre del fichero, el contenido del fichero, etc. y que pueden ser residentes (su valor se encuentra en la MFT) o no residentes (en la MFT se indica en qué lugar de la zona de datos se encuentra su valor). Las primeras entradas de la tabla están reservadas para registros que contienen información general del sistema de ficheros que no suele estar asociada a un fichero de usuario específico y se denominan *file system metadata files*. Las entradas reservadas que no se usan están marcadas como *allocated* pero sólo contienen información genérica. Todos los ficheros de metadatos del sistema de ficheros se encuentran en el directorio raíz, aunque no suelen estar visible para los usuarios. Los nombres de los ficheros de metadatos del sistema de fichero empiezan con el símbolo $ y mayúscula como se podrá ver en la siguiente tabla. Una característica a destacar de estos ficheros es que, al igual que el resto de ficheros del file system, tienen marcas de tiempo, que son de utilidad para el investigador para saber la fecha en la que fueron creados, por lo tanto indican la fecha en la que el sistema de ficheros fue creado.

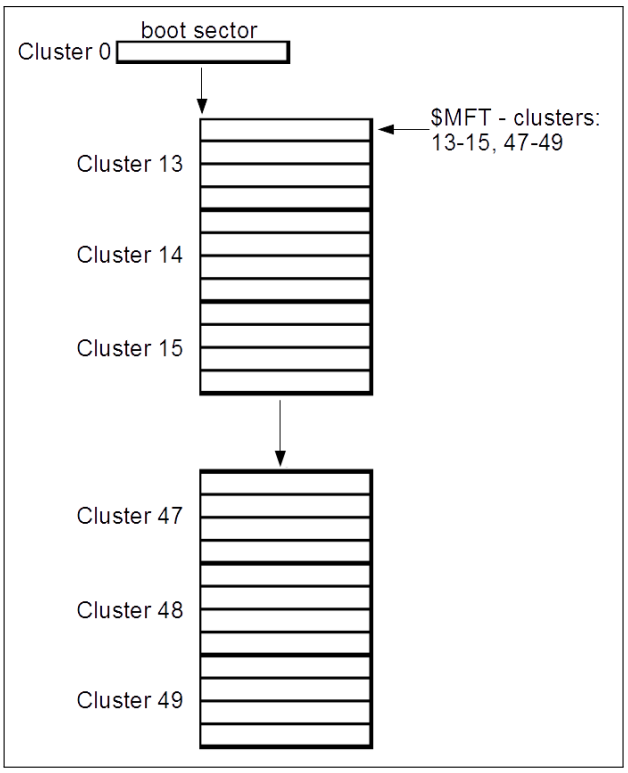

Figure 4.2: Master file table

Cada entrada se direcciona de forma secuencial usando 48 bits. La dirección máxima de la MFT va creciendo a medida que se añaden entradas. Las entradas de la MFT disponen de un número de secuencia de 16 bits que se incrementa cuando la entrada pasa a estar allocated/reallocated. La entrada de la MFT y el número de secuencia se combinan para crear una dirección de referencia de fichero de 64 bits. El sistema de ficheros NTFS utiliza este direccionamiento para referirse a las entradas de la MTF porque así le es más fácil determinar si se encuentra en un estado de corrupción. Esto puede ser útil para recuperar ficheros borrados: si tenemos datos unallocated con un número de referencia podemos determinar si la entrada de la MFT ha sido reasignada desde que se usó para los datos que tenemos.

Ficheros de metadatos del sistema de ficheros

Como se acaba de comentar, existen los *file system metadata files* en los que se almacenan los datos administrativos del sistema de ficheros. A continuación se muestra una tabla que resume los *file system metadata files*. Para ser más exactos, estas son

las entradas halladas en la MFT de un Windows 7 aunque se encuentran en todos los sistemas de ficheros NTFS.

Entrada	Nombre	Descripción
0	$MFT	La entrada de la propia MFT.
1	$MFTMirr	Copia de seguridad de las primeras entradas de la MFT y que se encuentra en medio del sistema de ficheros.
2	$LogFile	Contiene el *journal* que registra las transacciones de metadatos.
3	$Volume	Contiene información del volumen, como la versión.
4	$AttrDef	Contiene información de los atributos, como nombre y tamaño.
5	.	Contiene el directorio raíz del sistema de ficheros.
6	$Bitmap	Contiene el estado de asignación (allocated/unallocated) de todos los *clusters* del sistema.

7	$Boot	Contiene el sector y código de boot del sistema de ficheros.
8	$BadClus	Contiene los clusters con sectores dañados.
9	$Secure	Contiene información sobre la seguridad y control de acceso de los ficheros (Windows 2000 y Windows XP).
10	$Upcase	Contiene los caracteres Unicode en mayúscula.
11	$Extend	Un directorio que contiene ficheros para extensiones opcionales.

- $MFT: es la entrada de la tabla. En ella se indica la disposición en el disco de la tabla entera. En el sector de arranque se indica dónde se encuentra el inicio de la tabla. El atributo $DATA de la entrada $MFT contiene los *clusters* utilizados por la MFT. Como cualquier otra entrada también dispone del atributo $BITMAP, que como se verá más adelante en esta sección, controla el estado de asignación de las entradas de la MFT. El fichero $MFT se crea ocupando el

mínimo espacio posible y va creciendo a medida que se van creando ficheros.

- $MFTMirr: esta entrada tiene un atributo $DATA no residente que contiene una copia de seguridad de, por lo menos, las cuatro primeras entradas de la MFT en medio del file system. En caso de existir algún problema para determinar el layout de la MFT, se tendría que buscar el sector ubicado en la mitad del file system para leer los datos de esta entrada.

- $LogFile: contiene información utilizada por el sistema de ficheros para una rápida recuperación del sistema. El tamaño del *log* depende del tamaño del volumen y puede llegar a los 4 MB. Con el comando `chkdsk` se puede modificar el tamaño del *log*.

- $Volume: contiene la etiqueta e información de la versión del volumen. Tiene dos atributos que sólo usa este fichero: $VOLUME_NAME, que tiene el nombre del volumen en Unicode; y $VOLUME_INFORMATION, que contiene la versión del NTFS y *dirty status*.

- $AttrDef: el atributo $DATA de este fichero define los nombres y tipos de identificadores para cada tipo de atributo. Este fichero permite a cada file system tener atributos únicos para sus ficheros y también permite redefinir identificadores estándar de los atributos.

- Directorio raíz (.): es el directorio raíz del sistema de ficheros.

- $Bitmap: este fichero contiene información sobre el estado de asignación de cada uno de los clusters del sistema. En el atributo $DATA hay un bit reservado a cada cluster del file system de forma ordenada, lo que significa que el primer bit (bit 0) es para el cluster 0, el bit 1 es para el cluster 1, etc. Si el bit está a 1 significa que el cluster está allocated.

- $Boot: esta entrada contiene el sector de boot del sistema. El atributo $DATA se encuentra siempre en el primer sector del file system porque se necesita para arrancar el sistema. Es el único fichero de metadatos del file system que tiene una ubicación estática. La signatura del sector de arranque es la misma que para los sistemas de ficheros FAT, 0xAA55. El sector de boot aporta información

básica del tamaño de los clusters, el número de sectores en el file system y el cluster en el que empieza la MFT. En el atributo $DATA también se guarda el *boot code*, imprescindible para que el file system sea bootable y ubica los ficheros necesarios para cargar el sistema operativo. En algún sector del volumen (último sector o mitad del volumen en función de la versión de Windows) se encuentra una copia de seguridad del sector de arranque.

- $BadClus: NTFS lleva un registro de los clusters dañados del sistema. Para ello, cuando un cluster se reporta como dañado, se añade al atributo $DATA llamado $Bad (sparse file).

- $Secure: este fichero utiliza el parámetro *Security ID* del atributo $STANDARD_INFORMATION de cada fichero como índice para encontrar el *descriptor* adecuado. Este *Security ID* son diferentes de los *Windows Security Identifiers* (SID) asignados a los usuarios en Windows.

- $Upcase: para convertir un carácter Unicode en minúscula al mismo carácter Unicode en mayúscula.

- $Extend: se utiliza para extensiones opcionales como *quotas*, *reparse point data* y identificadores de objeto.

Para consultar información de estos ficheros se puede usar la herramienta $istat de The Sleuth Kit junto con el número de entrada, como por ejemplo si queremos que nos muestre información sobre el fichero de metadatos $MTF ejecutaremos `istat -f ntfs nom_img.dd 0`

Atributos

Un atributo es una estructura de datos que guarda un tipo de datos específico. Hay varios tipos de atributo y cada uno tiene una estructura interna distinta. Los atributos se componen de dos partes: la cabecera y el contenido. La cabecera es igual para todos los atributos pero el contenido es distinto para cada atributo, por lo que el tamaño también varía para cada atributo. Como hemos dicho anteriormente, las entradas de

la MFT destinan la mayor parte de sus bytes a almacenar los atributos del fichero/directorio que representan. En caso de que el contenido del atributo ocupe más de unos 700 bytes, la cabecera del atributo indica el lugar en el que se almacena en contenido de dicho atributo. La cabecera identifica el tipo de atributo, su nombre y su tamaño. También tiene *flags* que identifican si el valor está comprimido o cifrado. Una entrada de la MFT puede tener múltiples atributos del mismo tipo y para diferenciarlos se usa un identificador único para cada atributo. El contenido de un atributo, como se acaba de comentar, puede tener cualquier tamaño y formato, por lo que puede haber atributos que ocupen hasta gigabytes y que, por lo tanto, no se pueden alojar en la entrada de la MFT. Para solucionar este problema, el sistema de ficheros NFTS tiene dos sitios en los que se puede alojar el contenido de un atributo: en la misma entrada de la MFT (resident) o en un cluster externo en el file system (non-resident). En la cabecera del atributo se indica si es residente o no. En caso de no serlo, en la cabecera se indica el cluster en el que se encuentra el contenido junto con el número de clusters que ocupa.

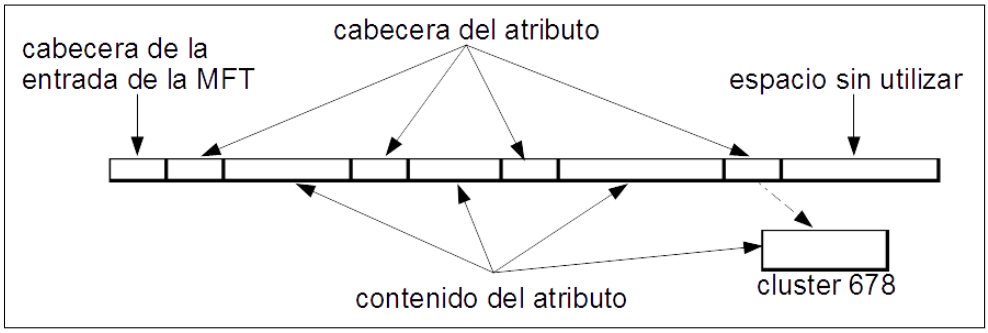

Figure 4.3: Entrada de la MFT

Los atributos se distinguen gracias a un número que identifica el tipo de atributo, además, se les da un nombre que se escribe en mayúsculas y empieza por $. Con la única excepción de las entradas non-base, todas las entradas allocated de la MFT tienen los atributos $FILE_NAME y $STANDARD_INFORMATION. El atributo $FILE_NAME contiene información del nombre del fichero, tamaño y tiempos. El atributo $STANDARD_INFORMATION, que existe en todos los ficheros y directorios, contiene información sobre tiempo de creación, acceso, etc, sobre permisos y de seguridad. Ambos atributos son residentes (se encuentran en la entrada de la MFT). Todos los ficheros tienen el atributo $DATA, en el que hay el contenido del

fichero. Si el fichero ocupa más de unos 700 bytes, se aloja fuera de la entrada (non-resident). Si el fichero tiene más de un atributo de tipo $DATA, se denomina alternate data stream (ADS) a estos atributos adicionales que, además, tienen que tener un nombre de atributo (que no es lo mismo que el nombre del tipo de atributo). Cada directorio tiene un atributo de tipo $INDEX_ROOT que contiene información sobre los subdirectorios y ficheros que contiene. Si se trata de un directorio grande, los atributos $INDEX_ALLOCATION y $BITMAP también se usan para almacenar información. Los directorio también pueden tener el atributo $DATA, en el que se guarda el contenido que una aplicación o usuario desee. Los atributos $INDEX_ROOT y $INDEX_ALLOCATION suelen tener el nombre $I30. Hay ficheros que tienen más atributos de los que caben en una sola entrada de la MFT, por lo que ocupan más entradas. En la entrada 'base', se encuentra el atributo $ATTRIBUTE_LIST, en el que se indica todos los atributos del ficheros y la entrada en la que se encuentran, sólo la entrada base tiene los atributos $FILE_NAME y $STANDARD_INFORMATION. Hay *flags* que indican si un atributo está comprimido o cifrado. Sólo se comprime o cifra el atributo no residente $DATA. La cabecera no se cifra.

NTFS usa índices (*index data structure*), que son colecciones de atributos que se guardan en un orden determinado. Los directorios, por ejemplo, usan índices para ordenar los atributos $FILE_NAME.

Los índices ordenan los atributos en lo que se denominan *B-tree*. Un árbol es un grupo de estructuras de datos (nodos) que se enlazan entre sí de forma que hay un nodo raíz que se ramifica creando padres e hijos. Los árboles son útiles porque permiten ordenar y encontrar los datos de forma fácil. El inconveniente está en la forma de crear el árbol. Como tiene que estar siempre ordenado, añadir o borrar un solo fichero puede cambiar completamente el árbol para poder cumplir con las condiciones de número de nodos permitidos por rama, cosa que puede dar problemas en una investigación porque con el movimiento podemos pensar que un fichero está eliminado pero simplemente ha cambiado de nodo para balancear. Los árboles tienen entradas de índice. Estos índices se almacenan en la MFT en los atributos $INDEX_ROOT o $INDEX_ALLOCATION en función del número de nodos. Se utiliza el atributo $BITMAP para encontrar un entrada de índice disponible para un nuevo nodo. Si no encuentra ningún indice disponible, se añade espacio. A cada índice se le

da un nombre, y a los atributos $INDEX_ROOT, $INDEX_ALLOCATION y $BITMAP se les asigna ese mismo nombre en sus cabeceras. En la entrada de un índice también existe un *flag* que indica si tienen nodos hijos.

Al formatear un equipo, la MFT nueva tendrá muy pocas entradas por lo que las entradas pertenecientes a la MFT de antes del formateo aún tendrían que existir en espacio unallocated. Se pueden buscar estas entradas para tratar de recuperar los atributos originales y, así, recuperar el contenido anterior al formateo. La herramienta `blkls` de The Sleuth Kit (*Herramientas usadas para el análisis* (apartado 6, pág. 49) nos muestra información sobre las unidades de datos del sistema de ficheros. Por defecto extrae las unidades de datos unallocated. A partir de esta información se puede usar la herramienta `sigfind` para buscar '4649c45' que significa 'FILE' en hexadecimal y nos servirá para encontrar las entradas de la MFT que, como hemos visto antes, empiezan por la signatura. El resultado de este proceso nos dará los sectores en los que se encuentran signaturas (FILE) y con la herramienta `dd` examinamos el contenido de los sectores.

A continuación se enumeran algunos de los tipos de atributos que pueden contener las entradas de la MFT.

Identificador	**Nombre**	**Descripción**
16	$STANDARD_INFORMATION	Información general, como por ejemplo fechas de modificación y acceso, propietario y flags.
32	$ATTRIBUTE_LIST	Indica dónde se pueden encontrar los atributos.
48	$FILE_NAME	Nombre del fichero (en Unicode) y fechas de modificación y acceso.

64	$OBJECT_ID	Identificador único de 16 bytes para el fichero/directorio.
80	$SECURITY_DESCRIPTOR	Control de acceso y propiedades de seguridad de un fichero.
96	$VOLUME_NAME	Nombre del volumen.
112	$VOLUME_INFORMATION	Versión del *file system* y *flags*.
128	$DATA	Contenido del fichero.
144	$SINDEX_ROOT	Nodo raíz para un árbol índice.
160	$INDEX_ALLOCATION	Utilizado para la implemetación de directorios y otros índices.
176	$BITMAP	Un bitmap para el fichero $MFT y para los índices.

- $STANDARD_INFORMATION: Este atributo existe para todos los ficheros y directorios del sistema de ficheros y contiene los metadatos más importantes. En este atributo se guardan las marcas de tiempo, información de propietario y de seguridad. No hay información indispensable para guardar los ficheros, pero muchas características del nivel de aplicación de Microsoft dependen de este atributo. Cuando el usuario selecciona las propiedades del fichero en Windows, los tiempos que se visualizan son los de creación, modificación y acceso. El tiempo de modificación de la MFT guardado en este atributo no se muestra en las propiedades. Además, a partir de Windows Vista[12], la fecha de último acceso no se actualiza al acceder a un fichero para ahorrar recursos al sistema. En este atributo existe un *flag* que indica si el fichero es sólo de lectura, si el fichero está comprimido o si está cifrado. También hay un identificador de seguridad que usa de índice el fichero $Secure y se usa para determinar qué normas de control se aplican a este fichero. Si se está usando el journal, hay el update sequence number (USN) para el último registro creado en este fichero. En concreto, las fechas que se almacenan en este atributo son las siguientes:
  - mtime: *modification time*, se guarda la fecha y hora en la que el fichero fue modificado por última vez.
  - atime: *access time*, se guarda la fecha y hora en la que el fichero fue accedido por última vez. A partir de Windows vista, esta fecha, por defecto, no se actualiza por temas de rendimiento del equipo.
  - ctime: *creation time*, este campo, aunque se llame de creación, registra la fecha y hora de última modificación en la entrada de la MFT correspondiente al fichero o directorio.
  - btime: *birth time*, se guarda la fecha y hora en la que el fichero fue *creado* en el sistema de ficheros, es decir, el momento en el que el fichero apareció en el volumen. Si el fichero fue creado en otro volumen (otro ordenador por ejemplo), la fecha de nacimiento (*btime*) será posterior a la fecha de modificación (*mtime*).

- $ATTRIBUTE_LIST: este atributo se utiliza cuando un fichero o directorio necesita más de una entrada en la MFT para guardar todos sus atributos. Aunque los atributos pueden ser no residentes, las cabeceras de todos los atributos se almacenan en la propia entrada de la MFT y este atributo contiene una lista de todos los atributos del fichero (menos él mismo). Cada entrada de la lista contiene el tipo de atributo y la dirección de la entrada en la que se encuentra.

- $FILE_NAME: en este atributo se encuentra el nombre del archivo y la dirección del directorio padre. Windows no suele actualizar la información temporal de este atributo, por lo que las fechas reflejadas en el atributo $STANDARD_INFORMATION y las fechas de este atributo pueden no coincidir. En general actualiza las fechas de este atributo sólo en caso de crear, renombrar o mover el fichero. En este atributo también se almacena información de los mismos flags que se almacenan en $STANDARD_INFORMATION con información sobre si la entrada es un directorio, read only, etc,

- $OBJECT_ID: este atributo contiene un identificador único para los ficheros. Este ID es utilizado por el *Distributed Link Tracking Service* y se utiliza, por ejemplo, para los accesos directos a ficheros. No todos los ficheros disponen de este atributo.

- $SECURITY_DESCRIPTOR: este atributo almacena la información de seguridad de un fichero. Sin embargo, en nuevas versiones de NTFS, Microsoft ha cambiado lo localización de esta información en un fichero denominado $SECURE. Una de las ventajas que conlleva el fichero $SECURE es la agrupación de permisos, varios ficheros con el mismo nivel de seguridad no necesitan almacenar esta información en ficheros individuales para cada uno.

- $VOLUME_NAME: este atributo sólo existe en el fichero $Volume. Como su nombre indica, almacena el nombre del volumen.

- $VOLUME_INFORMATION: este atributo también es utilizado sólo por el fichero $Volume. Contiene información de la versión del sistema de ficheros y un campo para *volume flags*.

- $DATA: como su nombre indica, en este atributo se almacenan los datos y no tiene un tamaño preestablecido. Puede haber más de un atributo $DATA en una misma entrada de la MFT. Por ejemplo, se puede añadir información de resumen a un fichero haciendo click con el botón derecho. Esta acción crea un segundo atributo $DATA a la entrada. Los directorios también pueden tener este atributo. Estos atributos adicionales, o alternate data streams ADS, no se muestran cuando se lista el contenido de un directorio, por lo que pueden servir para esconder información.

- $SINDEX_ROOT: todos los directorios tienen índices con información sobre los ficheros y subdirectorios que alojan. Si el directorio aloja pocos ficheros, la información se almacena en este atributo. Cuando el directorio alberga más ficheros, se almacenan en el atributo $INDEX_ALLOCATION. Estos índices forman un *B-tree*.

- $INDEX_ALLOCATION: como se acaba de mencionar, almacena información del árbol que empieza en el atributo $INDEX_ROOT.

- $BITMAP: este atributo también forma parte de la estructura de indexación. Mantiene un registro de qué partes del índice que estan allocated y cuáles están libres para ser reutilizadas. También para $MFT.

En el Instituto SANS (SysAdmin Audit, Networking and Security Institute) han elaborado un resumen de lo que significa la actualización de las fechas en los atributos $STANDARD_INFORMATION y $FILE_NAME.[4]

Figure 4.4: Cambio en los metadatos de un fichero. Fuente: SANS.

5 Funcionamiento del correo electrónico

Un correo electrónico es un servicio que permite a los usuarios enviar y recibir mensajes mediante sistemas de comunicación electrónica. Un correo electrónico es susceptible de ser alterado o manipulado con facilidad mediante programas de edición de imagen como Photoshop o con un editor de textos como el Bloc de notas de Windows sin que su contenido parezca alterado. Por esta razón, la presentación de un correo electrónico debe hacerse en formato digital y con determinadas garantías como la custodia del medio que los alberga, ya sea el buzón de correo electrónico en el que se encuentra o el correo web (*webmail*) en el que está contenido el mensaje. El estudio de la *'adveración de correos'* consiste en un análisis de las fuentes de información disponibles en cada caso (cabeceras de correo, propiedades del buzón de correo, registros del servidor, etc.) para determinar si los correos conservan su integridad o si, por el contrario, se observan incoherencias en los datos existentes en dichas fuentes de información que puedan indicar que esos correos electrónicos han podido sufrir alguna manipulación.

Los buzones de correo de los distintos gestores de correo que existen en el mercado (Microsoft Outlook, Lotus Notes, Thunderbird, Evolution, etc.) contienen una serie de metainformación introducida por el gestor de correo que permite su correcto funcionamiento. Esta información es incorporada de forma transparente por el programa gestor de correo y no puede ser manipulada directamente por el usuario, por lo que el análisis de sus valores permite deducir, a partir de su estudio detallado, las acciones realizadas por cada mensaje, incluyendo la presencia o ausencia de posibles modificaciones a las que el correo haya podido ser sometido tras ser enviado/recibido.

En el caso de que el servicio de correo electrónico sea proporcionado directamente

mediante el explorador web (Firefox, Chrome, etc), el gestor de correo no es un programa independiente usado para tal fin sino que se utiliza un correo web o *webmail*, que es un cliente de correo electrónico que provee una interfaz web por la que se accede al correo electrónico, como por ejemplo Gmail, el cliente de correo de Google.

En el caso de disponer de un gestor de correo, la adveración del correo se realizará a partir del correo almacenado en el gestor. En caso de no disponer de uno, la información que debe ser asegurada de un correo electrónico enviado o recibido a través de webmail, consiste en el contenido completo del correo bajo investigación, es decir, el cuerpo del correo en el que está el contenido del mismo, los ficheros adjuntos y la cabecera completa.

La cabecera de un correo electrónico está formada por una serie de datos identificativos del mensaje, tanto del del cuerpo del correo como de los ficheros adjuntos. En realidad, un correo electrónico es el conjunto de las cabeceras, cuerpo y datos adjuntos, pero debido a que no aportan demasiada información al usuario y que, dada su extensión y complejidad técnica, suponen más bien una molestia para el mismo, la mayoría de los programas y sistemas de correo electrónico no muestran los datos de la cabecera por defecto. Sin embargo, permiten diferentes opciones para poder visualizar la cabecera completa. Existen dos tipos de cabecera en un correo electrónico: la cabecera simple y la cabecera técnica.

- La cabecera simple está formada por los campos que se muestran al abrir el correo e incluso en la previsualización del mismo, que se encuentra en la lista de correos de la bandeja de entrada, salida, etc. Estos campos suelen mostrarse en el idioma en el que está configurado el gestor o cliente de correo.
  - De:
  - Para:
  - Asunto:
  - Fecha:
- La cabecera técnica es la que aporta más información al investigador por la cantidad de campos que contiene y le permite establecer una trazabilidad del correos en Internet. Por ejemplo, permite determinar por qué sitios pasó el mensaje antes de ser recibido y la hora exacta en la que éste fue recibido. Esta

cabecera no se muestra por defecto en los gestores de correo ni desde un cliente web pero se pueden ver a través de simples opciones como ”Muestra el original” disponible para cada correo en Gmail. A diferencia de los campos de la cabecera simple, estos siempre se muestran en inglés, que es el idioma de estandarización y todos acaban con dos puntos (:). Algunos de los campos de la cabecera técnica son los siguientes:

- Delivered-To:
- Received:
- Return-Path:
- In-Reply-To:
- Message-ID:

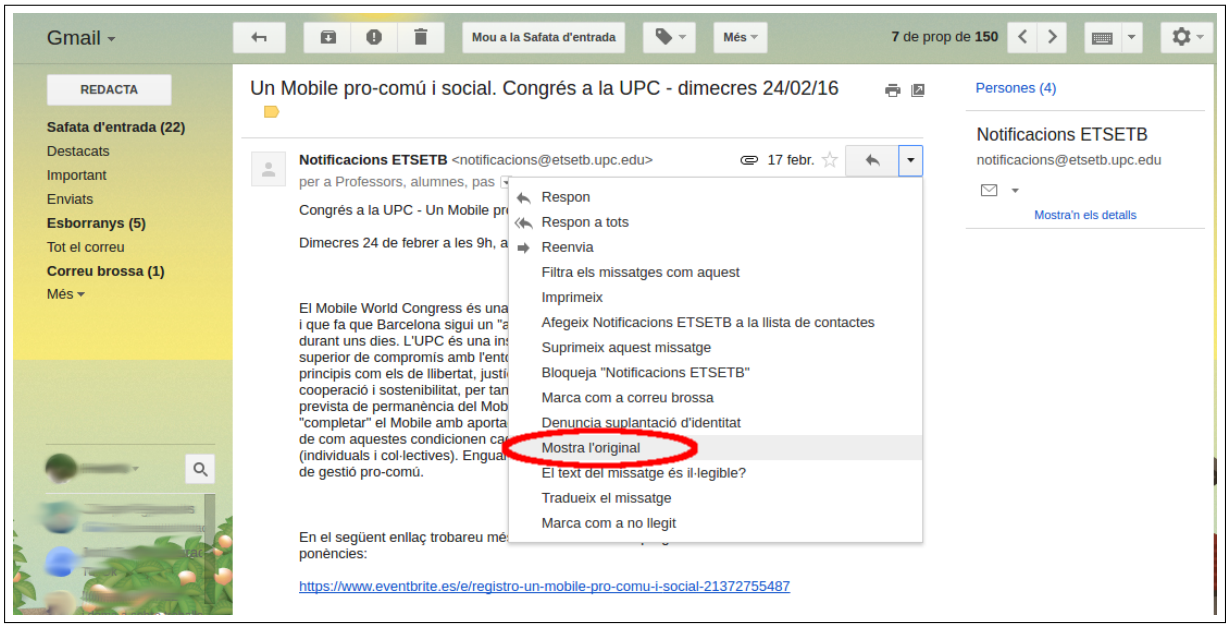

Figure 5.1: Opción para poder visualizar el correo electrónico en su totalidad.

Este aseguramiento debe realizarse en el formato original, es decir, en formato digital puesto que se trata de un correo electrónico, almacenándolo en un dispositivo adecuado tal como una memoria USB u otro dispositivo de almacenamiento de datos externo. De forma complementaria, un juez puede enviar un requerimiento judicial al proveedor de servicio de Internet (ISP por sus siglas en inglés) para obtener los

registros de actividad asociados a esta cuenta de correo. El análisis de estos *logs* permitirá estudiar la coherencia de los datos aportados respecto a aspectos tan relevantes como fecha y hora de acceso a la cuenta de correo o envío y recepción de mensajes, entre otros.

Para poder analizar una cabecera se debe conocer el funcionamiento de un correo electrónico, desde que un usuario lo envía hasta que llega a su destinatario. Los correos electrónico se envían vía SMTP (Simple Mail Transfer Protocol), que en castellano se puede traducir por protocolo simple de transferencia de correo. A continuación se resumen los pasos que sigue un correo electrónico desde su origen hasta su destino:

1. El programa de correo del emisor entrega el mensaje al servidor de correo que tenga establecido.
2. El servidor de correo del usuario entrega el mensaje al servidor de correo de destino.
3. El servidor de correo de destino almacena el correo en el buzón del usuario.
4. El destinatario recibe el mensaje de su servidor de correo.

La siguiente imagen ejemplifica el proceso que se acaba de describir.

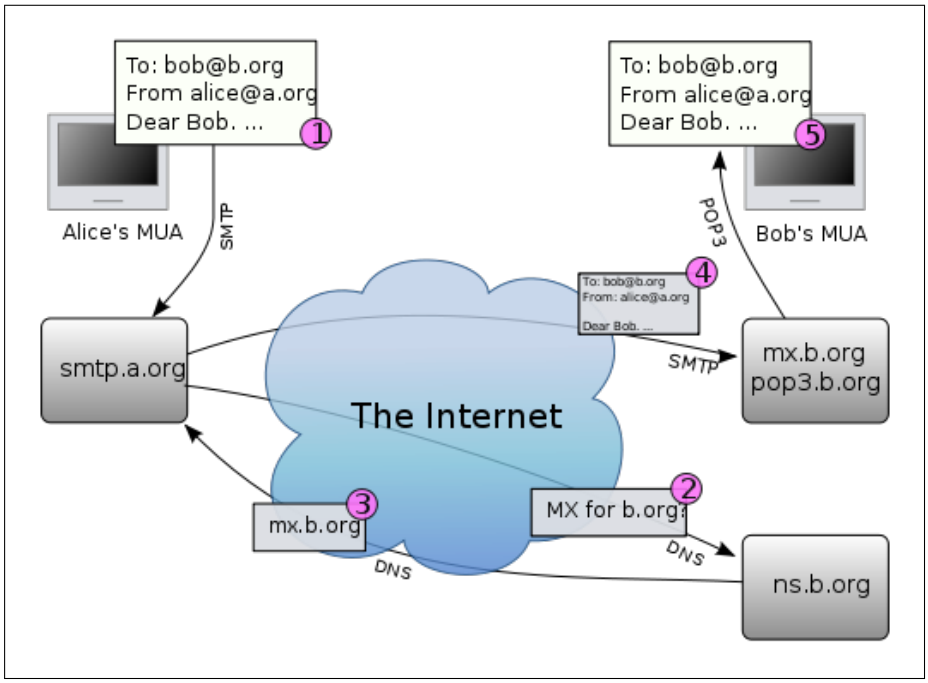

Figure 5.2: Transmisión de un correo electrónico a través de diversos servidores de correo en Internet.

En cada paso del proceso de envío del correo, se generan ciertos campos que se van añadiendo a la cabecera técnica final que recibe el destinatario del mensaje. A continuación se explica de forma breve el proceso en el que se generan los campos de la cabecera.

- Cuando el emisor escribe un correo electrónico tiene que rellenar, de forma obligatoria, el campo del receptor con la dirección de correo a quien vaya dirigido el mensaje. De forma opcional el emisor rellena el asunto, el cuerpo del correo y añade ficheros adjuntos, aunque sería posible enviar un correo completamente vacío, sólo con el destinatario.

- El cliente de correo que esté utilizando el emisor, antepone una cabecera a dicho mensaje. A continuación, cuando el emisor decide enviar el correo, el cliente de correo lo envía al servidor de correo que tenga asociado, como por ejemplo a de su proveedor de servicios de Internet.

- El servidor de correo añade, al principio, una cabecera *Received* y lo pasa al siguiente servidor de correo que corresponda.
  - Este proceso puede repetirse varias veces porque el correo puede pasar por varios servidores de correo intermedios de Internet antes de llegar a su destino
  - Normalmente, un servidor de correo intermedio sólo añade cabeceras Received, pero ocasionalmente también puede insertar otras como las que introducen los antivirus, antispam, etc.
  - Nótese que cada servidor inserta su propia cabecera al principio de las cabeceras, y por tanto antes que las demás.
  - En dicha cabecera se incluye la hora local del servidor que recibe el correo, siendo posible determinar el tiempo que tarda un mensaje en llegar a su destino.

- Cuando llega al último servidor, éste, además del campo Received puede escribir otras cabeceras. Finalmente deposita el mensaje en el buzón del usuario.

A continuación se puede ver un ejemplo, extraído de Wikipedia [13], de un correo enviado vía SMTP. En este ejemplo el correo electrónico se envía desde la dirección *bob@example.org* a dos destinatarios, a *alice@example.com* y a *theboss@example.com*. En el ejemplo se identifica el usuario con una C de cliente y el servidor, con una S. La transmisión empieza con el servidor indicando el nombre del dominio en el que se encuentra (smtp.example.com). El usuario empieza la conversación con un ”HELO”. Finalmente, la conversación acaba con un ”Bye” por parte del servidor.

```
S: 220 smtp.example.com ESMTP Postfix
C: HELO relay.example.org
S: 250 Hello relay.example.org, I am glad to meet you
C: MAIL FROM:<bob@example.org>
S: 250 Ok
C: RCPT TO:<alice@example.com>
S: 250 Ok
C: RCPT TO:<theboss@example.com>
S: 250 Ok
C: DATA
S: 354 End data with <CR><LF>.<CR><LF>
C: From: "Bob Example" <bob@example.org>
C: To: "Alice Example" <alice@example.com>
C: Cc: theboss@example.com
C: Date: Tue, 15 January 2008 16:02:43 -0500
C: Subject: Test message
C:
C: Hello Alice.
C: This is a test message with 4 header fields and 5 lines
   in the message body.
C: Your friend,
C: Bob
C: .
S: 250 Ok: queued as 12345

C: QUIT
S: 221 Bye
```

Los gestores de correo pueden añadir campos adicionales a la cabecera técnica. Este es el caso del gestor de correo Microsoft Outlook, que añade información muy relevante para el investigador sobre el envío de cada correo, en concreto, uno de los campos que añade es el denominado `x-originating-ip:` en el que se indica la dirección IP desde la que se envió el correo electrónico. Disponer de la dirección IP de

origen no sólo es útil para localizar el lugar desde el que se envió el correo electrónico sino que además, aporta información sobre el proveedor de Internet utilizado. En caso de considerarse necesario, se puede solicitar, mediante requerimiento judicial al proveedor de servicio de Internet al que pertenece la dirección IP, los datos personales e información de la utilización del servicio del usuario asociado a esta dirección IP.

A continuación se realizará el análisis de un correo electrónico a modo de ejemplo. La cabecera de este correo ha sido modificada para anonimizar los correos reales analizados, por lo que podría haber alguna pequeña incoherencia debida a la modificación realizada para este proyecto. En esta investigación no se detectó ninguna modificación de los correos electrónicos extraídos que habían sido enviados o recibidos entre DIRECCIÓN_1 y DIRECCIÓN_2.

En este ejemplo podemos ver un correo que fue enviado desde el gestor de correo Microsoft Outlook. Se han substituido valores de algunos campos por tres puntos (...) porque estos valores no aportan información relevante para la demostración y así no se revelan datos confidenciales y se facilita la lectura.

```
Delivered-To: direccion1@empresa1.com
Received: by 10.194.33.39 with SMTP id o7csp492541wji;
        Tue, 15 Nov 2011 07:19:07 -0700 (PDT)
X-Received: by 10.202.242.137 with SMTP id q131mr17697382oih.137.1464704347461;
        Tue, 15 Nov 2011 07:19:07 -0700 (PDT)
Return-Path: <direccion2@empresa2.com>
Received: from EUR01-HE1-obe.outbound.protection.outlook.com
          (mail-he1eur01on0065.outbound.protection.outlook.com. [104.47.0.65])
        by mx.google.com with ESMTPS id z194si24044937oia.58.2011.11.15.07.19.06
        for <direccion1@empresa1.com>
        Tue, 15 Nov 2011 07:19:07 -0700 (PDT)
Received-SPF: pass (google.com: domain of direccion2@empresa2.com designates
              104.47.0.65 as permitted sender) client-ip=104.47.0.65;
Authentication-Results: ...
DKIM-Signature: ...
Received: from AM2PR07MB0865.eurprd07.prod.outlook.com (10.161.71.151) by
 AM2PR07MB0867.eurprd07.prod.outlook.com (10.161.71.153) with Microsoft SMTP
 Server (TLS) id 15.1.501.7; Tue, 15 Nov 2011 14:19:04 +0000
 from AM2PR07MB0865.eurprd07.prod.outlook.com ([10.161.71.151]) by
 AM2PR07MB0865.eurprd07.prod.outlook.com ([10.161.71.151]) with mapi id
 15.01.0506.013; Tue, 15 Nov 2011 14:19:04 +0000
```

```
From: "direccion2@empresa2.com" <direccion2@empresa2.com>
To: "direccion1@empresa1.com" <direccion1@empresa1.com>
Subject: Material
Thread-Topic: Material
Thread-Index: AdG7RyTeAc80vwCdQg+45mjHWVKTUw==
Date: Tue, 15 Nov 2011 14:19:04 +0000
Message-ID: <AM2PR07MB0865B8A0143AED17493EEF30865.euasdsd07.prod.outlook.com>
Accept-Language: es-ES, en-US
Content-Language: es-ES
X-MS-Has-Attach: yes
X-MS-TNEF-Correlator:
authentication-results: ...
x-originating-ip: [212.145.159.148]
x-ms-office365-filtering-correlation-id: 8bf04447-0847-45ac-915f-08d3895e852d
x-microsoft-exchange-diagnostics:...
x-microsoft-antispam: UriScan:;BCL:0;PCL:0;RULEID:;SRVR:AM2PR07MB0867;
x-microsoft-antispam-prvs:...
x-exchange-antispam-report-test:...
x-exchange-antispam-report-cfa-test:...
x-forefront-prvs:...
x-forefront-antispam-report: ...
```

```
spamdiagnosticoutput: ...
spamdiagnosticmetadata: NSPM
Content-Type: multipart/related;
boundary="_004_AM2PR07MB0865B8A0143AED17493EEF3CF0460AM2PR07MB0865eurp_";
type="multipart/alternative"
MIME-Version: 1.0
X-OriginatorOrg: empresa2.com
X-MS-Exchange-CrossTenant-originalarrivaltime: 15 Nov 2011 14:19:04.2565 (UTC)
X-MS-Exchange-CrossTenant-fromentityheader: Hosted
X-MS-Exchange-CrossTenant-id: 1c6d7ece-5a37-467d-ae24-57ffa0901acd
X-MS-Exchange-Transport-CrossTenantHeadersStamped: AM2PR07MB0867

--_004_AM2PR07MB0865B8A0143AED17493EEF3CF0460AM2PR07MB0865eurp_
Content-Type: multipart/alternative;
boundary="_000_AM2PR07MB0865B8A0143AED17493EEF3CF0460AM2PR07MB0865eurp_"
--_000_AM2PR07MB0865B8A0143AED17493EEF3CF0460AM2PR07MB0865eurp_
Content-Type: text/plain; charset="iso-8859-1"
Content-Transfer-Encoding: quoted-printable

Hola,
"SIGUE EL CUERPO DEL CORREO"
```

Para analizar la cabecera técnica de un correo electrónico nos centramos en dos elementos: los valores de los distintos campos *Received* y las fechas y horas que van apareciendo a lo largo de la cabecera del correo electrónico. Los campos *Received* aportan información para trazar el recorrido del correo electrónico desde que fue enviado hasta que fue recibido. Cada vez que el correo pasa por un servidor, añade un campo *Received* a la cabecera. La lectura de una cabecera se realiza de abajo hacia arriba, siendo el campo *Received* de más abajo el que ofrece información del emisor y, el campo *Received* de más arriba, el que ofrece información sobre el buzón de entrega del receptor. En este caso concreto, el correo ha sido enviado desde el gestor de correo Microsoft Outlook, hecho que se observa por los datos que se pueden leer en la cabecera sobre Microsoft. Que el correo haya sido enviado desde Microsoft Outlook es una buena noticia para el investigador porque este gestor de correo añade el campo *x-originating-ip* en la cabecera del correo. El valor de este campo es el dato más relevante para la trazabilidad del origen de un correo electrónico puesto que se trata de la dirección IP de origen del correo. Por motivos de privacidad del usuario, la mayoría de gestores y clientes de correo electrónico no añaden este campo de información a la cabecera, sin embargo, Microsoft Outlook sí que lo hace.

Los campos con información temporal del correo son la principal fuente de información para determinar la veracidad del correo. En este caso observamos que la diferencia de tiempo entre el primer y el último *Received* es de 3 segundos,siendo 14:19:04 +0000 la hora que se muestra en el primer (emisor) *Received* y 07:19:07 -0700 (PDT) la hora que se muestra en el último (recepción) *Received*, que equivale a 14:19:07 +0000. Se observa una diferencia de dos décimas de segundo entre los campos *X-MS-Exchange-CrossTenant-originalarrivaltime* y el primer *Received*, es una diferencia temporal muy pequeña que puede ser atribuida a tiempo de carga del correo electrónico. Además, se encuentra en procesos del emisor, por lo que no se considera una alteración que deba ser analizada en más profundidad.

6 Herramientas usadas para el análisis

Aunque es cierto que cada caso es distinto y se necesitan personas que analicen e interpreten los datos hallados durante una investigación, es conveniente tener procesos automatizados. No sólo porqué de esta forma el investigador ahorra tiempo, sino porque así se evitan errores humanos. En este capítulo se hará un resumen de las herramientas usadas para el caso que nos ocupa. Esto incluye herramientas de *software libre* y los realizados para este caso.

6.1 Adquisición y aseguramiento

Tal y como se ha comentado en el apartado*Fases de un análisis forense* (apartado 2, pág. 5), lo primero que se debe realizar antes de empezar con el proceso de investigación es una copia binaria o forense de los dispositivos y que, además, se deben presentar acompañados del *hash* correspondiente para que cualquier otra persona pueda contrastar los resultados obtenidos.

- Copia: las herramientas más comunes son `dd` y `dcfldd`. La diferencia entre ambas es que la segunda realiza el cálculo del *hash* a la vez que realiza la copia.

- *Hash*: El *hash* de una información (un texto, un fichero, etc.) es un cálculo matemático cuyo resultado es una combinación de números y letras. Cualquier cambio en la información, por pequeño que sea, altera totalmente su *hash*. Además, es prácticamente imposible encontrar otra información que tenga el mismo *hash* que la original. Por esta razón el *hash* es un instrumento muy

valioso en el aseguramiento de prueba electrónica. Existen distintos tipos de *hash*, siendo los más conocidos SHA-1 y MD5. Si durante la adquisición de datos se deja constancia del *hash* de la información, más adelante cualquiera que tenga acceso a la información puede volver a calcularle el *hash*, y si éste no ha variado se podrá asegurar que la información de que se dispone es exactamente la misma que se obtuvo durante la adquisición de datos.

```
sudo dd if=/dev/sdg conv=sync,noerror,notrunc 2>/media/caso/img.dd.log |
tee nom_img.dd | md5sum > /media/caso/img.dd.md5"
```

- Clonadora: Se trata de un equipo que realiza la copia y el cálculo del *hash*. Se coloca el disco o discos vacíos destinatarios de la copia, el disco original se conecta a la clonadora y, gracias a la pantalla se selecciona la opción de realizar la copia y calcular el *hash*, que se almacena en el disco receptor en un fichero separado. En la pantalla de la clonadora se indica el tiempo que queda para terminar y, al finalizar, muestra el resultado del *hash* por pantalla. Se realizan fotografías del proceso para añadirlas al informe porque dan soporte a la cadena de custodia.

Figure 6.1: Clonación de un disco y cálculo del hash con la clonadora.

6.2 Montaje del disco

Para poder acceder al sistema de ficheros sin modificar ni un solo bit de información, se montan los discos sin necesidad de arrancar el sistema operativo y en formato solo lectura.

- `fdisk`, `parted` o `mmls`: son programas que sirven para visualizar y manipular las tablas de partición de un disco. Para las investigaciones se utiliza para determinar el número de particiones del disco, el file system utilizado y el offset en el que empieza el sistema de ficheros y, así, poder montarlo para acceder a la estructura de directorios y ficheros.

- `mount`: monta un sistema de ficheros al destino que se le indique. Es importante para un investigador usar la opción de sólo lectura para no comprometer los datos a analizar. De esta forma se accede al sistema de ficheros sin necesidad de arrancar el equipo y, por lo tanto, no se modifican los registros del sistema.

6.3 Tratamiento de la información

- `strings`: es una herramienta que imprime cadenas de caracteres del fichero que se le indique, ya sea un fichero o una imagen de disco. Puede imprimir caracteres en más de una codificación, como por ejemplo ASCII o Unicode.

- `grep`: es una herramienta que busca un patrón concreto dentro de un fichero y muestra por pantalla la línea del fichero en la que este patrón aparece. Entre las opciones que ofrece la herramienta existe la de ignorar mayúsculas y minúsculas, mostrar ciertas líneas antes y después de la coincidencia e invertir la coincidencia, es decir, mostrar las líneas en las que el patrón buscado no aparece.

- `find`: es una herramienta que busca ficheros en el árbol de directorios del sistema por el nombre que se le pasa, una de las opciones más interesantes es la de ignorar mayúsculas y minúsculas.

- `tree`: es una herramienta que lista los directorios y su contenido en forma de árbol, es decir, manteniendo la jerarquía de directorios y subdirectorios existente. Una de las opciones más interesantes que ofrece la herramienta es la de mostrar ficheros ocultos, que no se muestran por defecto. También se le puede indicar el nivel de 'profundidad' que se quiere mostrar, es decir, el número de subdirectorios que mostrará por pantalla esta herramienta.

- RegRipper: es una herramienta forense que sirve para extraer datos en un equipo que utilize el sistema operativo Windows. Es una herramienta muy extensa, formada por diversos paquetes que extraen información muy diversa sobre el ordenador, como por ejemplo los usuarios del equipo con la última fecha de escritura en disco, información sobre las impresoras o sobre los dispositivos externos que han sido conectados al ordenador en cuestión.

- `exiftool`: es una herramienta que se utiliza para leer los metadatos de un fichero.

6.4 Búsqueda ciega de palabras clave y análisis heurístico

El análisis heurístico persigue acceder a los contenidos de los documentos electrónicos con la seguridad de que no contienen información personal, al tiempo que reduce aún más el número de documentos a analizar manualmente, lo que repercute directamente en la simplificación de la investigación. Para esta detección de información personal suelen utilizarse diversas variantes de lo que se llama *análisis heurístico*, que tiene como base asignar, de manera ciega y automatizada, una puntuación a cada documento electrónico, basándose en ciertos criterios que pueden variar en cada caso, y descartar aquéllos sospechosos de contener información personal en base a esta puntuación. Esta técnica se utiliza de forma ruinaria en las investigaciones de informática forense, y se considera un estándar *de facto*, aunque su implementación exacta puede variar dependiendo de los criterios utilizados por cada perito para la puntuación. Aunque no se haya utilizado el análisis heurístico en esta investigación,

se ha considerado importante explicarla porque remarca la importancia de no analizar los ficheros de carácter privado de los usuarios.

6.5 The Sleuth Kit

The Sleuth Kit está formado por una colección de programas o comandos de código abierto que sirven para analizar imágenes de discos duros. En el capítulo anterior ya se han introducido algunas de estas herramientas y continuación se detallan:

Figure 6.2: The Sleuth Kit.

- `tsk_gettimes`: esta herramienta extrae los metadatos de los sistemas de ficheros de una imagen de disco. El resultado se puede usar para crear una *timeline* con la herramienta mactime.
- `mactime`: esta herramienta crea una timeline de los ficheros a partir del resultado obtenido con la herramienta tsk_gettimes.
- `fls`: esta herramienta genera una lista de todos los ficheros y directorios de file system. También puede incluir directorios o ficheros borrados. Indica el nodo al que pertenece el fichero/directorio.
- `mmls`: como se ha comentado antes, este comando lista la tabla de particiones, el sector en el que empieza cada partición, el tamaño y el tipo de partición.
- `istat`: esta herramienta muestra por pantalla información sobre el nodo seleccionado. Por ejemplo tamaño, si el fichero/directorio está borrado, los tiempos

de acceso y modificación, etc. En los sistemas de ficheros NTFS muestra los atributos de la MFT del fichero/directorio correspondiente.

- `ils`: esta herramienta lista la información de los nodos. Por defecto sólo muestra aquellos nodos correspondientes a ficheros eliminados.

- `blkls`: esta herramienta muestra los bloques de datos del file system. Por defecto muestra sólo los bloques no asignados pero también puede mostrar detalles de los que sí que están allocated.

Mediante tsk_gettimes y mactime se extraen *timelines* del dispositivo analizado, que nos puede dar bastante información sobre la actividad del equipo si se saben interpretar los datos correctamente, es decir, se conoce el significado de la mtime, atime, ctime y btime en los distintos sistemas operativos que nos podemos encontrar durante una investigación.

6.6 Recuperación de datos

A veces, en una investigación, se requiere recuperar datos eliminados ya sea como requerimiento expreso del cliente o para tener más información acerca de la actividad llevada a cabo en el dispositivo. Otras veces, nos encontramos con particiones de disco dañadas y para acceder a los datos se tiene que recuperar antes la partición del disco. En este apartado se explican dos herramientas distintas: TestDisk, PhotoRec. Ambas herramientas son *open source* y tienen la misma finalidad pero utilizan métodos distintos.

Figure 6.3: TestDisk y PhotoRec.

- TestDisk: esta herramienta fue diseñada para recuperar particiones. Actualmente, además de recuperar particiones borradas, TestDisk también puede recuperar el sector de boot de la copia de seguridad de los sistemas de ficheros

FAT y NTFS y recuperar ficheros eliminados de los sistemas FAT, exFAT, NTFS y ext2. También puede copiar ficheros de particiones FAT, exFAT, NTFS y ext2/ext3/ ext4 eliminadas. TestDisk a veces dispone de nombres de ficheros que no puede recuperar. En este caso sirve para listar ficheros eliminados.

- PhotoRec: esta herramienta fue diseñada para recuperar ficheros eliminados. Se puede utilizar en discos duros, CDs, USB, tarjetas SD y cámaras digitales. PhotoRec ignora el sistema de ficheros y se centra en los datos subyacentes, por lo que puede recuperar ficheros aunque el sistema de ficheros este dañado o haya sido formateado. Puede recuperar ficheros de, por lo menos, FAT, NTFS, exFAT, ext2/ext3/ext4 y HFS+. También recupera fotos y vídeos de varias cámaras digitales. Para recuperar ficheros, busca cabeceras de más de 480 tipos de ficheros. Si los datos no están fragmentados, puede recuperar el fichero entero. Los resultados de recuperación de ficheros que se obtienen suelen ser de muchos ficheros sin algunos de sus metadatos, como el nombre del ficheros o las marcas temporales.

6.7 Scripts

En función de la naturaleza del caso, puede ser aconsejable la creación de *scripts* específicos para automatizar ciertos procesos igual que hacen las herramientas descritas en el apartado anterior. La automatización de tareas permite:

- Minimizar errores humanos.
- Reutilizar herramientas durante el caso y en investigaciones futuras.
- Y en la misma linea que el punto anterior, optimizar el tiempo para que el investigador se pueda centrar en analizar los datos y así evitar perder tiempo en llevar a cabo tareas que no aportan valor dado que son automatizables.
- Si en el transcurso de una investigación se analiza el contenido del sistema de ficheros, no se suele arrancar el sistema operativo, sino que se *monta* el sistema de ficheros en modo lectura para poder ver los directorios y ficheros. Los discos

pueden tener varias particiones y, para montar el disco, se debe calcular el offset de cada partición. Para facilitar el trabajo, se creó una herramienta que monta las distintas particiones de un disco.

- Uno de los objetivos que perseguía el informe era el de extraer todos los correos electrónicos enviados y recibidos entre dos direcciones concretas. Por diversos factores que fueron apareciendo durante la investigación, se optó por buscar los correos electrónicos en el fichero de *strings* generado. Como se ha visto en la explicación de los discos duros, la información puede ser parcialmente eliminada, cosa que dificulta la extracción de correos eliminados. Así pues, se optó por extraer correos totales y parciales. En el anexo se puede consultar el *script* creado para esta investigación.

- Otro de los objetivos del informe era el de encontrar todos los ficheros con ciertas palabras en su nombre. Como no se revisarían uno a uno todos los ficheros de los discos, porque, recordemos, se debe mantener el derecho a la intimidad de los usuarios, se creó un simple *script* que revisaba los nombres de todos los ficheros de los discos.

7 Proceso de investigación

En este apartado se representan las fases llevadas a cabo en una investigación real realizada en INCIDE. Con la finalidad de escribir este proyecto sin desvelar información real de ninguna de las partes involucradas en el proceso judicial, se utilizan palabras genéricas como Cliente y Empresa, empezando en mayúscula para remarcar que no es una empresa ni un cliente cualquiera sino la empresa y el cliente en los que se centra el caso; y FECHA y PALABRA, escritas en mayúscula para remarcar que se trataría de fechas y palabras concretas pero que no se han escrito para hacer el informe de resultados anónimo. También se emplean 'XX' para ocultar resultados numéricos y otra información confidencial de la investigación.

7.1 Documentación

La fase de documentación empieza recopilando la información proporcionada por el Cliente sobre el caso y termina con el informe de resultados. Cuando el Cliente contactó con INCIDE, la Empresa ya había denunciado al Cliente, por este motivo, parte de la información recopilada ha sido extraída de las diligencias previas número XXX del juzgado número 1 de XXX. Esta información se resume a continuación:

- La Empresa despide a Cliente, por sospechar de irregularidades en las cuentas.
- La Empresa denuncia a Cliente por supuestas operaciones de venta de material en un periodo en el que el denunciado desempeñaba sus tareas en la Empresa.
- Concretamente por la compraventa de material a un proveedor en concreto por un precio que no se correspondía con la facturación entre ambas empresas.

El Cliente contrata a INCIDE para realizar el mismo análisis que la Policía y, de esta forma, que su abogado pueda construir una buena defensa en base a los resultados que puede encontrar la Policía. Los objetivos del análisis son los siguientes:

- Extraer todos los fichero que, en su denominación, contengan el nombre PALABRA_A o PALABRA_B.
- Extraer todos los correos electrónicos enviados o recibidos entre DIRECCIÓN_1 y DIRECCIÓN_2.
- Verificar que los correos electrónicos extraídos son íntegros.
- Extraer todos los ficheros eliminados entre FECHA_INICIAL y FECHA_FINAL.

Además de la información previa sobre el caso, todos los procedimientos que se realicen y los resultados que se extraigan forman parte de la fase de documentación. En este caso, tanto el apartado anterior *Descripción del caso práctico a analizar* (apartado 3, pág. 15), este mismo apartado, como el informe pericial que se puede consultar en *Informe pericial* (apartado 8, pág. 76), forman parte de la documentación del caso. También posibles notas intermedias que no se acaban incluyendo en el informe ya sea por tener poca relevancia para el resultado de la investigación o por ser detalles demasiado técnicos.

7.2 Preparación

La empresa no disponía de ningún protocolo para evitar que los empleados sustrajeran información confidencial. Los equipos usados por los empleados podían tener los sistemas operativos Windows o Ubuntu instalados. Los ordenadores que utilizaban Windows tenían todos el gestor de correo Microsoft Outlook instalados pero no era obligatorio usarlo para acceder al correo corporativo. Los ordenadores que utilizaban Ubuntu no disponían de un gestor de correos instalado de serie para que los empleados lo utilizaran. Tampoco había habilitado medidas para que los empleados no pudieran conectar dispositivos externos a sus equipos ni medidas de monitorización de las acciones de los empleados.

7.3 Incidencia

La incidencia por la cual se ha realizado la investigación es la alteración de los precios de compraventa entre Cliente (denunciado) y un proveedor. Fue detectada por la Empresa (denunciantes). Se desconoce cómo y cuándo lo descubrieron. La pericial ha sido encargada para poder demostrar los hechos que confirmar la incidencia.

7.4 Respuesta a la incidencia

En esta investigación no hubo respuesta a la incidencia. Para INCIDE el caso empezó cuando el Cliente ya había sido despedido y el proceso judicial contra él ya se encontraba en el juzgado de instrucción. El análisis realizado se ha efectuado sobre las copias de los equipos proporcionados por la Empresa por su relación con el Cliente.

7.5 Análisis forense digital

7.5.1 Adquirir y autenticar

Tal y como se ha comentado en la primera etapa de análisis digital (*Adquirir y autenticar* (apartado 2.6.1, pág. 11)), el primer paso consiste en tomar posesión de los dispositivos a analizar. En este caso la empresa aportó cinco equipos de sobremesa con un disco duro cada uno. Se desmontaron los equipos para extraer los discos duros y, mediante la *clonadora* se realizaron dos copias de cada disco, una para cada parte. Los discos duros originales fueron depositados ante notario junto con una memoria USB que contenía fotografías realizadas durante el proceso de clonación de los discos, incluyendo las fotografías realizadas de la pantalla de la *clonadora* al finalizar la copia, en la que se puede ver el resultado del cálculo del *hash*. Los discos duros que reciben las copias son de tamaño ligeramente superior a los originales porque la *clonadora* crea un fichero con los *hash SHA256* y *MD5* calculados que sirven para posibles comprobaciones sobre si se han alterado los datos del disco.

7.5.2 Examinar y recolectar

Una vez se tienen los dispositivos para analizar, se realiza una pequeña descripción del dispositivo para tener más datos de contexto que ayuden en la investigación. El tamaño del disco, el número de particiones, sistema operativo y sistema de ficheros así como los usuarios del equipo y fechas relevantes como la de instalación y de último uso por parte de todos los usuarios forman parte de la caracterización del equipo. Esta información se suele añadir como anexo al informe. La herramienta `mmls` de The Sleuth Kit extrae información sobre el disco: el número y tipo de particiones, dónde empiezan y cuánto ocupan. Para el equipo de Windows, la herramienta `Regripper` extrae información sobre la fecha de instalación, la versión del sistema operativo y los usuarios con las fechas de creación y de última conexión.

Como se ha visto en el apartado *Documentación* (apartado 7.1, pág. 57), se han establecido unos objetivos concretos a cumplir, por lo que antes de empezar a analizar los datos, se evalúa qué información puede ser relevante y cómo se puede extraer. En este caso, se necesitan ficheros con un nombre particular, ciertos correos electrónicos y recuperar ficheros entre dos fechas. Por este motivo, se realizan las siguientes acciones:

- Listado de todos los archivos del sistema de ficheros con la herramienta `find`.
- Listado de todos los archivos y directorios del equipo, actuales y borrados recientemente, con la herramienta `fls` de The Sleuth Kit.
- Búsqueda de gestores de correo electrónico y/o *webmail* (cliente de correo electrónico al que se accede desde el navegador de Internet).
- Extracción de las cadenas del disco con la herramienta `strings`.
- Extracción de la *timeline* o línea temporal de los equipos con las herramientas `tsk_gettimes` y `mactime` de The Sleuth Kit.
- Carving: recuperar el máximo de ficheros y directorios eliminados con las herramientas `TestDisk` y `PhotoRec`.
- Extracción de un listado de todos los ficheros contenidos en la papelera de reciclaje.

La mayoría de estas acciones están automatizadas de forma que se generan ficheros en los que se van almacenando los resultados encontrados para poder ser analizados.

7.5.3 Análisis de los datos

En esta fase se empiezan a analizar los datos extraídos en la fase anterior con el fin de poder cumplir los objetivos marcados para la realización del informe pericial. En este caso las hipótesis que se tienen que probar o desmentir son los objetivos establecidos en el proceso judicial y que se detallan en los siguientes subapartados. Todas las anotaciones realizadas forman parte de la fase de documentación y se encaminan a la redacción del informe de resultados.

Extraer todos los fichero que, en su denominación, contengan el nombre PALABRA_A o PALABRA_B

Uno de los objetivos era el de extraer todos los ficheros con ciertas palabras en el nombre. Lógicamente, no se mirarían los directorios uno a uno hasta haber repasado el nombre de todos los ficheros. No solo por la gran cantidad de tiempo que se necesitaría sino también porque vulneraría el derecho a la intimidad de los usuarios del equipo. Se utilizó la lista de ficheros *allocated* del disco (extraída en el subapartado anterior) para realizar una búsqueda ciega mediante palabras clave, siendo las palabras claves las dos indicadas por el Cliente que debían figurar en el nombre del fichero.

Para determinar si hay ficheros eliminados que se conservan en el disco y que contienen las palabras PALABRA_A o PALABRA_B en el nombre, se realizó una búsqueda en el listado de ficheros extraído con la herramienta `fls` de The Sleuth Kit. Igual que en el caso anterior, la palabras se buscan ignorando las mayúsculas y minúsculas. A continuación se muestra como, a partir de los ficheros extraídos en el apartado anterior (*Examinar y recolectar* (apartado 7.5.2, pág. 60)), se realiza una búsqueda ciega y se escriben los resultados en un nuevo ficheros.

```
echo "PALABRA_A" >> ficheros_AB.txt
grep -Ei "PALABRA_A" ficheros_disco.txt >> ficheros_AB.txt

echo "PALABRA_B" >> ficheros_AB.txt
grep -Ei "PALABRA_B" ficheros_disco.txt >> ficheros_AB.txt
```

Se comprueban y comparan los resultados de los dos procedimientos. Los ficheros no eliminados son los mismos en ambas listas. En el listado fls_AB.txt se encuentran ficheros marcados como eliminados. Se buscan estos ficheros en los resultados del procedimiento de recuperación de ficheros eliminados pero no se han encontrado todos los ficheros. Esto puede pasar porque el *inodo* en el que se almacenan los metadatos de nombre apunte a un lugar en el que ya hay otro fichero y, este, también tiene otro *inodo* con los metadatos correctos que lo apunta. Además, Photorec no permite recuperar el nombre del fichero, por lo que es posible que alguno de los ficheros recuperados sean los que se buscan pero dado que sólo se tiene el nombre no se puede asegurar y no entra en los objetivos de este caso, puesto que se solicitaba por nombres de fichero sin entrar en su contenido. Dado que se han encontrado ficheros eliminados que contenían las palabras en el nombre, se realiza una búsqueda por estas mismas palabras en todos los ficheros recuperados por si se ha recuperado algún fichero que contenga las palabras PALABRA_A y/o PALABRA_B en su nombre y que la herramienta `fls` no haya podido extraer. No se han encontrado más ficheros pero se han extraído aquellos ficheros localizados a partir de la lista de ficheros eliminados.

Extraer todos los correos electrónicos enviados o recibidos entre DIRECCIÓN_1 y DIRECCIÓN_2

El Cliente solicitó extraer todos los correos electrónicos enviados y recibidos entre dos direcciones de correo electrónico concretas. Como se ha comentado en el apartado *Descripción del caso práctico a analizar* (apartado 3, pág. 15) y en el subapartado *Documentación* (apartado 7.1, pág. 57), la Empresa quiere aportar pruebas que demuestren cómo se llevaban a cabo las irregularidades entre Cliente y el proveedor. Para ellos, el juez decretó que se podía acceder al correo laboral del Cliente (DIRECCIÓN_1) y extraer toda la correspondencia que tenga relación con la com-

praventa de material con dicho proveedor (DIRECCIÓN_2). Dada la importancia de los correos electrónicos por contener potencial información sobre el asunto que atañe al caso, los correos electrónicos extraídos tiene que ser analizados para verificarse su autenticidad tal y como se verá en el siguiente subapartado.

No se especificó cómo se accedía al correo electrónico corporativo, pero sí que se informó a este perito de que los equipos con el sistema operativo Windows tenían el gestor de correo Microsoft Outlook instalado. Por este motivo, como se ha visto en el apartado anterior, se realizó una búsqueda de gestores de correo electrónico y de uso de *webmail*. No se encontraron indicios en los ficheros temporales de Internet que apuntaran que se hubiese utilizado el correo electrónico desde el navegador de Internet. Por lo que respecta al uso de gestores de correo, se encontró Microsoft Outlook en el equipo que utilizaba el sistema operativo Windows XP y el gestor de correo Evolution en los equipos que utilizaban el sistema operativo Ubuntu.

Todos los buzones encontrados estaban asociados a una única dirección de correo electrónico, el correo corporativo del Cliente (DIRECCIÓN_1) por lo que las búsquedas realizadas son de la dirección de correo electrónico DIRECCIÓN_2 con la herramienta `grep`. A continuación se resumen los resultados:

- Disco 01: La única coincidencia de la palabra DIRECCIÓN_2 fue en la agenda de contactos, no se encontraron correos enviados o recibidos entre DIRECCIÓN_1 y DIRECCIÓN_2.
- Disco 02: La única coincidencia de la palabra DIRECCIÓN_2 fue en la agenda de contactos, no se encontraron correos enviados o recibidos entre DIRECCIÓN_1 y DIRECCIÓN_2.
- Disco 03: No se ha encontrado ninguna coincidencia de la palabra DIRECCIÓN_2.
- Disco 04: Se encontraron coincidencias en la agenda de contactos, en la bandeja de entrada y en la bandeja de correos eliminados. Para recuperar estos correos se ha usado la herramienta `undbx`, que extrae correos de las bases de datos de Outlook. Además, con la opción de recuperación no sólo extrae los correos, sino que también recupera posibles correos borrados y los presenta en carpetas separadas para distinguir su origen.

- Disco 05: Se encontraron coincidencias en la agenda de contactos y en las bandejas de entrada y de salida. Sin embargo, el comportamiento de Evolution no fue el esperado y los correos no se podían abrir y aislar de forma individual. Para poder analizar el contenido de los correos, el investigador creó una herramienta que separa los correos a partir de su cabecera.

Nótese que realizar la extracción de correos electrónicos a partir del resultado de búsquedas ciegas por palabra clave, protege los usuarios de que su correspondencia sea leída más allá de los estrictamente necesario para la investigación. Aunque se trate de una dirección de correo corporativo, sólo han sido analizados aquellos correos que cumplían las condiciones establecidas por un juez, protegiendo, así, el derecho a la intimidad de los usuarios.

El hecho de que en el disco 01 y en el disco 02 se encontrase la dirección de correo electrónico DIRECCIÓN_2 en la agenda de contactos, puede indicar que en algún momento se han intercambiado correos electrónicos con esta dirección aunque no se haya encontrado ningún correo presente en el sistema de ficheros en el momento de realizar la extracción. Por este motivo se aprovechó el *script* creado para separar los correos del disco 05 para buscar correos electrónicos eliminados en las cadenas del disco, ya que el usuario podría haber eliminado dichos correos pero podrían seguir en espacio no reasignado del disco. A continuación se resumen los resultados:

- Disco 01: Se encontraron correos enviados y/o recibidos entre DIRECCIÓN_1 y DIRECCIÓN_2.
- Disco 02: Se encontraron correos enviados y/o recibidos entre DIRECCIÓN_1 y DIRECCIÓN_2.
- Disco 03: No se ha encontrado ninguna coincidencia de la palabra DIRECCIÓN_2.
- Disco 04: Se encontraron correos enviados y/o recibidos entre DIRECCIÓN_1 y DIRECCIÓN_2.
- Disco 05: Se encontraron correos enviados y/o recibidos entre DIRECCIÓN_1 y DIRECCIÓN_2.

Estos correos electrónicos han sido extraídos de las cadenas del disco, es decir, se han extraído todos los caracteres imprimibles del disco y se han realizado las búsquedas sobre estos ficheros. Si comparamos los resultados de los dos procedimientos podemos ver que en el disco 01 y en el disco 02 no se encontraron correos en el sistema de ficheros, por lo que podemos deducir que los correos encontrados a partir de los *strings* son correos eliminados. Se han podido recuperar correos porque los datos (ficheros o directorios) que elimina un usuario no se eliminan del disco duro de forma inmediata sino que el sistema los ”marca” como eliminados. Sin embargo, los datos permanecen en el disco hasta que el espacio que ocupan se sobreescribe con datos nuevos. Dado que no todos los ficheros ocupan igual, puede que algunos correos hayan quedado parcialmente sobreescritos y que sólo se haya podido recuperar aquella parte del correo electrónico que aun no había sido sobreescrita.

Verificar que los correos electrónicos extraídos son íntegros

Se han analizado todos los correos electrónicos que han sido recuperados de forma completa y no se han detectado alteraciones tal y como se ha explicado en *Funcionamiento del correo electrónico* (apartado 5, pág. 37). Se puede afirmar que todos los correos son íntegros. Aquellos correos electrónicos que no han podido ser recuperados en su totalidad se han dividido en dos: correos con cabeceras totales y contenido parcial y correos con cabecera parcial y contenido total o parcial. Los correos de los que se dispone una cabecera técnica completa han podido ser adverados aunque no puedan ser visualizados en su totalidad. Aunque no se haya encontrado ningún correo modificado, los correos de los que no se dispone de una cabecera completa no pueden ser adverados y, por lo tanto, no se puede afirmar que no hayan sufrido ninguna modificación.

Extraer todos los ficheros eliminados entre FECHA_INICIAL y FECHA_FINAL

El Cliente solicita recuperar los ficheros eliminados entre FECHA_INICIAL y FECHA_FINAL. Para extraer los correos se han utilizado las herramientas `TestDisk` y `PhotoRec`. Sin embargo, estos resultados se tienen que filtrar para que se ajusten a las fechas indicadas. Para ello, se diferencia entre el disco 04 (Windows XP) de los discos 01,

02, 03 y 05 (Ubuntu).

Windows, como se ha comentado en *NTFS* (apartado 4.4, pág. 22), utiliza el sistema de ficheros NTFS. Este sistema de ficheros no registra datos del momento en el que se elimina un fichero. Esto ocurre porque a un usuario no le interesa la fecha de eliminación de un fichero que está borrando, ya que, precisamente, lo está eliminando y simplemente quiere que desaparezca. Por lo tanto, NTFS se *ahorra* este proceso. Sin embargo, aunque los metadatos de un fichero o directorio eliminado no incluyan fecha y hora del momento en el que este fue borrado, sí que es posible encontrar rastros indirectos de cuándo sucedió este evento. Cuando se elimina un fichero o un subdirectorio, se produce un cambio en el directorio en el que está contenido el elemento y este cambio se registra en los metadatos del directorio, en concreto, en el atributo $STANDARD_INFORMATION de la entrada de la MFT correspondiente a este directorio. El mismo campo también se actualiza si se añade un fichero en el directorio, si se renombra el directorio o si mueve el directorio, es decir, hay varias acciones que implican modificar el campo de última modificación de la MFT del atributo $STANDARD_INFORMATION. Por lo tanto, no se puede saber en qué momento fue eliminado un fichero ni un directorio o subdirectorio.

Aunque, como acabamos de ver, no hay ningún campo que registre el momento de borrado de un fichero o directorio, se puede establecer una ventana temporal en la que un directorio podría haber sido eliminado. Esto es, entre el momento de última modificación del directorio que se encuentra eliminado en el momento de realizar el análisis y el momento de última modificación de su directorio padre. Es importante entender que esta fecha sólo afecta al directorio y no a su contenido. Cada directorio dispone de un índice de su contenido, ya sean ficheros o subdirectorios. Este índice va aumentando a medida que el directorio aloja más elementos. El tamaño del índice no disminuye aunque se elimine el contenido. Por esta razón, no se puede saber el contenido que alojaba un directorio en el momento de ser eliminado. Es decir, no podemos saber la fecha en la que fue eliminado un fichero.

No obstante, en esta investigación no se necesita una fecha exacta de eliminación sino que se establece un rango de fechas. Así, todos aquellos ficheros que hayan sido accedidos, creados o modificados en una fecha posterior a FECHA_INICIAL de un directorio que se ha eliminado entre las fechas deseadas se puede afirmar que sí que

han sido eliminados en el periodo temporal entre FECHA_INICIAL y FECHA_FINAL porque se tienen constancia de que existían pasada FECHA_INICIAL. El siguiente esquema ejemplifica esta explicación.

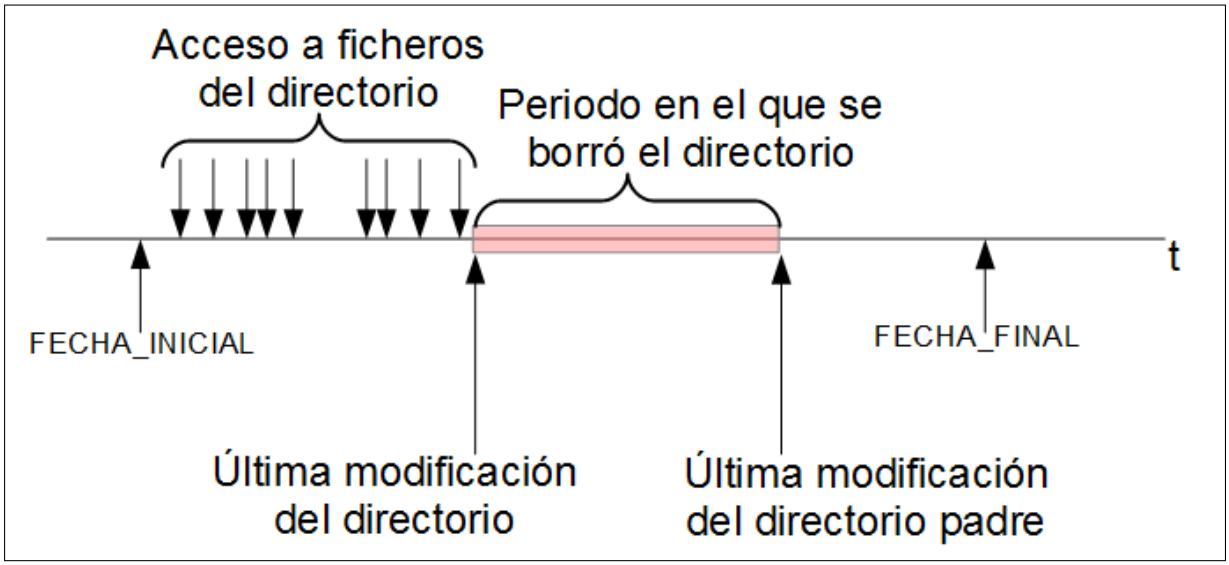

Figure 7.1: Esquema de representación del borrado de ficheros.

Se han examinado los metadatos de los directorios eliminados para encontrar todos aquellos que cumplen los requisitos indicados. Además, tal y como se acaba de explicar, se han comprobado los registros temporales de todos los ficheros y subdirectorios contenidos en estos directorios para elaborar un listado de todos los ficheros eliminados entre FECHA_INICIAL y FECHA_FINAL del disco 04.

El sistema de ficheros NTFS sí que registra cuándo un fichero ha sido enviado a la papelera de reciclaje. Se ha elaborado un listado con los ficheros enviados a la papelera entre FECHA_INICIAL y FECHA_FINAL aunque el hecho de que un fichero se encuentre en la papelera de reciclaje no puede considerarse como borrado sí que denota intención de querer eliminarlo.

Los cuatro equipos Linux venían con el sistema de ficheros EXT4. Este sistema de ficheros registra cuatro marcas temporales en sus inodos. Estas son fecha de cambio del inodo (ctime), fecha de acceso (atime), fecha de modificación (mtime) y fecha de eliminación (dtime). Gracias al registro al campo *deletion time* (dtime) se puede saber la fecha en la que un fichero fue eliminado. Se han analizado los metadatos de todos

los ficheros recuperados mediante la herramienta `TestDisk` para separar los ficheros eliminados entre FECHA_INICIAL y FECHA_FINAL.

Además, los usuarios del sistema operativo Ubuntu (y Linux en general), se caracterizan por utilizar entornos menos gráficos que en Windows, es decir, utilizan más la consola del sistema y no tanto las ventanas características de Windows. Los comandos ejecutados desde la terminal se almacenan en el fichero *bash_history* que, por defecto, viene limitado a XXX instrucciones almacenadas. Se ha analizado el historial de comandos realizados y se han encontrado varias ejecuciones del comando `rm` de remove en inglés, y que se utiliza para eliminar ficheros y directorios. Algunos de los ficheros eliminados por línea de comandos no han sido recuperados, por lo que se ha querido ubicar en el tiempo la ejecución de los comandos de borrado. Para ello, se han analizado los comandos de antes y después del borrado para establecer un marco temporal en el que se han eliminado los ficheros. Es decir, si el comando anterior y posterior al de borrado de un fichero era de creación de otros ficheros, se han buscado los nuevos ficheros y se han analizado sus metadatos para saber la fecha de creación y, así, establecer el momento en el que fue borrado el fichero. No se ha podido establecer la fecha y hora de borrado de todos los ficheros.

7.6 Presentación

Por último, todas las notas recopiladas a lo largo de las fases anteriores, se presentan en un informe de resultados. En este caso, el informe pericial del caso se puede consultar en el anexo *Informe pericial* (apartado 8, pág. 76). En el informe pericial no se incluyen los comandos realizados para obtener los resultados pero sí que se explica el procedimiento seguido para llegar a los resultados y las conclusiones que se desprenden de estos. Todo este proyecto se puede considerar que forma parte de la fase de presentación de la investigación llevada a cabo con motivo del proyecto final de carrera.

8 Conclusiones

Este trabajo ha tenido un enfoque teórico en su primera parte y práctico en la segunda.

En la explicación teórica del proyecto, hemos visto las distintas etapas por las que pasa una investigación forense digital: documentación, preparación, incidencia, respuesta a la incidencia, análisis forense digital y presentación. Aunque cada investigación tiene sus particularidades, todas se pueden adaptar a las fases explicadas: se empieza por la documentación del caso, con los antecedentes que han llevado al cliente a solicitar una investigación y que sirven para tener contexto con el que trabajar. La fase de documentación sigue viva durante todo el caso, porque el perito o analista forense anota todo el procedimiento seguido y los resultados conseguidos con el fin de presentar un informe detallado de todos los hechos. Antes de llegar a presentar nuestro informe, que se corresponde a la última fase, también se han explicado las etapas de preparación, incidencia y respuesta a la incidencia y, evidentemente, análisis forense digital en la que el investigador trabaja para confirmar la hipótesis realizada sobre los hechos que mueven el caso. Dividir la investigación en distintas fases ayuda a entender mejor el proceso en su conjunto sin perder de vista el objetivo final de la investigación. También es de utilidad a la hora de elaborar el informe de forma lógica y organizada.

A lo largo de todas las fases de la investigación, es de especial importancia y de obligado cumplimiento seguir las buenas prácticas forenses, en las que se establecen los principios básicos de:

- Integridad
- Auditabilidad
- Expertos

- Formación adecuada
- Legalidad

Con tal de garantizar estos principios, se debe establecer la cadena de custodia para asegurar que los datos no han sido modificados en ningún momento de la investigación y, así, poder presentar, de forma clara y detallada, las pruebas en un proceso legal.

En la parte práctica del proyecto, hemos visto como tanto las fases mencionadas anteriormente como los principios de buenas prácticas forenses y la cadena de custodia se pueden apreciar en la investigación llevada a cabo por INCIDE. La investigación descrita en este proyecto no ha sido una excepción. En este caso se han seguido todas las fases salvo la de respuesta a la incidencia, porque no hubo una incidencia a tratar en el momento. Se estableció la cadena de custodia por parte del perito ante notario en el momento de realizar la adquisición de datos, hecho que asegura la integridad y auditabilidad de los datos analizados, que han sido recogidos por un especialista en presencia de un fedatario público. Como se ha podido comprobar a lo largo del proyecto, todas las acciones han seguido procedimientos legales y han quedado perfectamente documentadas para quien desee consultarlas.

Los objetivos marcados para la investigación eran los siguientes:

- Extraer todos los fichero que, en su denominación, contengan el nombre PALABRA_A o PALABRA_B.
- Extraer todos los correos electrónicos enviados o recibidos entre DIRECCIÓN_1 y DIRECCIÓN_2.
- Verificar que los correos electrónicos extraídos son íntegros.
- Extraer todos los ficheros eliminados entre FECHA_INICIAL y FECHA_FINAL.

Para cumplir estos objetivos se han seguido varios procedimientos forenses adaptados a cada objetivo en particular. Como se puede ver en el anexo *Conclusiones* (apartado 3, pág. 91), se ha respondido satisfactoriamente cada uno de los objetivos planteados. Los puntos que conllevaron un mayor desempeño por parte del investigador fueron la verificación de la integridad de los correos electrónicos y la extracción de ficheros eliminados entre dos fechas concretas.

Puesto que los correos electrónicos son fácilmente manipulables, se debe tener especial cuidado en la comprobación de la concordancia y coherencia de los distintos datos que aparecen en las cabeceras de los correos. En el caso de la extracción de ficheros eliminados entre dos fechas, la tarea puede llegar a no ser concluyente debido a las características de cada sistema de ficheros. En concreto, hemos visto que NTFS, el sistema de ficheros utilizado en Windows, no registra la fecha y hora de borrado de los ficheros, por lo que sólo se puede asegurar una ventana de tiempo para la eliminación de un fichero. Todos estos factores contribuyen en gran medida al éxito de una investigación digital.

Bibliography

[1] Brian Carrier. File system forensic analysis, 2005.

[2] International Organization for Standardization. Iso/iec 27037:2012, 2012. URL https://www.iso.org/obp/ui/#iso:std:iso-iec:27037:ed-1:v1:en.

[3] Information Security Group. URL http://futur.upc.edu/ISG.

[4] SANS Institute. Windows time rules, 2012. URL https://uk.sans.org/posters/windows_artifact_analysis.pdf.

[5] Michael Donovan Köhn. Integrated digital forensic process model, 2012.

[6] The Sleuth Kit. Help documents. URL http://wiki.sleuthkit.org/index.php?title=Help_Documents.

[7] U.S. Department of Justice Office of Justice Programs. Electronic crime scene investigation: A guide for first responders, 2001. URL https://www.ncjrs.gov/pdffiles1/nij/219941.pdf.

[8] RegRipper. URL https://code.google.com/archive/p/regripper/.

[9] INCIDE Digital Data S.L. URL http://www.incide.es.

[10] Microsoft Enterprise Platforms Support: Windows Server Core Team. Ntfs file attributes, 2010. URL https://blogs.technet.microsoft.com/askcore/2010/08/25/ntfs-file-attributes/.

[11] TestDisk and PhotoRec. URL http://www.cgsecurity.org/.

[12] Forensics Wiki. Mac times. URL http://www.forensicswiki.org/wiki/MAC_times#NTFS.

[13] Wikipedia. Simple mail transfer protocol. URL https://en.wikipedia.org/wiki/Simple_Mail_Transfer_Protocol.

Informe pericial

Informe Pericial

INCIDE - Ref.101232

7 de marzo de 2016

CONFIDENCIAL

Informe pericial con relación al procedimiento de diligencias previas XXX

1 Cuestiones previas

1.1 Consideraciones previas

La siguiente información ha sido facilitada por el Cliente. Ha sido extraída de las diligencias previas XXX del juzgado número 1 de XXX:

- La Empresa despide a Cliente, por sospechar de irregularidades en las cuentas.
- La Empresa denuncia a Cliente por supuestas operaciones de venta de material en un periodo en el que el denunciado desempeñaba sus tareas en la Empresa.
- Concretamente por la venta de material a un cliente en concreto por un precio que no se correspondía con la facturación entre ambas empresas.

1.2 Solicitud de opinión

Por todo lo expuesto en el apartado anterior, Cliente ha solicitado mi opinión sobre los siguientes extremos:

- Extraer todos los fichero que, en su denominación, contengan el nombre PALABRA_A o PALABRA_B.
- Extraer todos los correos electrónicos enviados o recibidos entre DIRECCIÓN_1 y DIRECCIÓN_2.
- Verificar que los correos electrónicos extraídos son íntegros.
- Extraer todos los ficheros eliminados entre FECHA_INICIAL y FECHA_FINAL.

1.3 Fuentes de información

La información recogida en este informe deriva del análisis de los datos contenidos en los siguientes sistemas y/o soportes informáticos:

- Disco WD2500AAJS 250GB identificado como Disco 01
- Disco Seagate Barracuda 250GB identificado como Disco 02
- Disco Seagate Barracuda 250GB identificado como Disco 03
- Disco SAMSUNG HD103SJ identificado como Disco 04
- Disco Seagate Barracuda 250GB identificado como Disco 05

1.4 Adquisición de datos y cadena de custodia

Este apartado recoge las circunstancias en que se obtuvieron las diferentes fuentes de información para su análisis.

En Barcelona, el representante legal de la Empresa proporcionó cinco (5) ordenadores de sobremesa que contenían los discos mencionados en el apartado anterior (*Fuentes de información* (apartado 1.3, pág. 77)) con el fin de que el notario D. Notario, en fecha XXX, diese fe del proceso de copia realizado.

Una vez allí, se procedió a la identificación y extracción de los discos duros de los ordenadores entregados para realizar dos copias de cada disco duro, quedando los originales en depósito notarial y entregándose una de las dos copias de cada disco a INCIDE y la otra al representante legal de la Empresa. Junto con los equipos se depositó un dispositivo de memoria externa USB que contenía fotografías realizadas del proceso de identificación, extracción y clonación de los discos junto con el cálculo de la función criptográfica *hash* de cada disco.

Todo el proceso se llevó a cabo en la citada fecha en las dependencias notariales de la Notaría de D. Notario de Barcelona, quedando reflejada en el protocolo notarial número XXX. Para ver más detalles, véase el anexo de detalle de aseguramiento de pruebas *Aseguramiento de las fuentes de información* (apartado 4.1, pág. 93).

A continuación se reflejan los *hash* de los correspondientes discos duros.

```
MD5 (Disco_01.dd): c2cb76d72c5c2662d78e28c4e36ec6bb
MD5 (Disco_02.dd): a1d078a880dde94789fa4w4ba28d44c5
MD5 (Disco_03.dd): a4b69h1b11d4f5574096d569ccb20da3
MD5 (Disco_04.dd): pxj47sbf61jhd73bn6ouyasdkgh5411a
```

```
MD5 (Disco_05.dd): gsorb69sb12doo750bawn34g211dijsl
```

2 Dictamen

2.1 (a) Extraer todos los fichero que, en su denominación, contengan el nombre PALABRA_A o PALABRA_B

El Cliente requiere extraer todos los ficheros de los discos duros que, en su nombre, contengan las palabra PALABRA_A y/o PALABRA_B. Para ello, se han utilizado tres procedimientos diferentes:

- Realizar una búsqueda directamente en el sistema de ficheros actual.
- Extraer un listado de ficheros y directorios del disco actuales y borrados recientemente y, en este árbol de directorios y ficheros, realizar una búsqueda por nombre de fichero.
- Realizar una búsqueda en los directorios con la herramienta de recuperación de ficheros *TestDisk*.

Estos procedimientos se detallan a continuación.

Realizar una búsqueda directamente en el sistema de ficheros actual

El procedimiento más eficaz para encontrar ficheros que aún se encuentran en el sistema de ficheros es extraer un listado de todos los ficheros existentes y realizar una búsqueda en ese listado. Ambos procedimientos están automatizados, por lo que se evitan errores humanos. Además, dado la búsqueda se realiza por palabra clave, no se analizan todos los ficheros uno a uno y, de esta forma, se protege la privacidad del usuario. Para maximizar resultados, se ignoran las mayúsculas y minúsculas así como las tildes.

Resultados

- Disco 01: No se han encontrado ficheros.
- Disco 02: No se han encontrado ficheros.
- Disco 03: No se han encontrado ficheros.

- Disco 04: Se han encontrado tres ficheros no borrados. El contenido de estos archivos es recuperable.
- Disco 05: No se han encontrado ficheros.

Una vez confirmada la presencia de ficheros relevantes, se accede al equipo para obtener estos ficheros. Para poder ver el contenido de los discos duros sin alterar el contenido de estos, se monta el sistema de ficheros en modo sólo lectura. Nótese que al montar el sistema de ficheros no se está encendiendo el ordenador, por lo que los registros que hacen relación a encendidos/apagados y *logueos* no se modifican. El hecho de montar el disco en con la opción de sólo lectura se imposibilita la modificación de ficheros del disco.

Extraer un listado de ficheros y directorios del disco actuales y borrados recientemente y, en este árbol de directorios y ficheros, realizar una búsqueda por nombre de fichero

Se ha extraído un listado de ficheros y directorios a partir de la copia de cada disco. Al realizarse sobre todo el disco y no de la unidad montada, pueden aparecer ficheros y directorios eliminados en la lista. A partir de los listados se ha realizado una búsqueda por las palabras clave PALABRA_A y PALABRA_B. Igual que en procedimiento anterior, se ignoran mayúsculas, minúsculas y tildes para maximizar los resultados.

Resultados

- Disco 01: No se han encontrado ficheros.
- Disco 02: No se han encontrado ficheros.
- Disco 03: No se han encontrado ficheros.
- Disco 04: Se han encontrado tres ficheros no borrados. El contenido de estos ficheros es recuperable.
- Disco 05: No se han encontrado ficheros.

Los ficheros coinciden con la búsqueda realizada con el método anterior.

Realizar una búsqueda en los directorios con la herramienta de recuperación de ficheros *TestDisk*

Para recuperar ficheros borrados se ha utilizado la herramienta *TestDisk*, que permite recuperar ficheros eliminados conservando el nombre de fichero. Esta herramienta tiene ciertas limitaciones. En el caso de los equipos Linux (01, 02, 03, y 05), que utilizan el sistema de ficheros EXT4, no puede recuperar los ficheros, pero si que permite visualizar ficheros que han sido eliminados.

Resultados

- Disco 01: Se han localizado 2 ficheros pero no se han podido recuperar.
- Disco 02: No se ha detectado ningún fichero.
- Disco 03: Se han localizado 4 ficheros pero no se han podido recuperar.
- Disco 04: Se han detectado 7 ficheros eliminados y se han recuperado con *TestDisk*.
- Disco 05: Se han localizado 1 fichero pero no se ha podido recuperar.

En el proceso de recuperación de ficheros mediante la herramienta *TestDisk* se han detectado varios ficheros que, en su nombre, contenían las palabras PALABRA_A y/o PALABRA_B. Desafortunadamente, debido al tipo de sistema de ficheros que utilizan los discos Linux, no se han podido recuperar los ficheros detectados.

Se realizó una segunda búsqueda utilizando la herramienta *PhotoRec*. *PhotoRec* es una herramienta que permite recuperar ficheros eliminados y que admite ciertos filtros, es decir, permite seleccionar el tipo de fichero que se quiere recuperar (doc, png, xls jpg, ppt, etc.). El inconveniente que presenta Photorec es que los ficheros carecen de metadatos, como el nombre del fichero. Se realizaron búsquedas con las misma extensiones que los ficheros detectados con *PhotoRec*. Como resultado se han obtenido gran cantidad de ficheros sin nombre y con las extensiones deseadas. Dada la gran cantidad de resultados es imposible asociar los ficheros detectados con los resultados.

2.2 (b) Extraer todos los correos electrónicos enviados o recibidos entre DIRECCIÓN_1 y DIRECCIÓN_2

Se pide extraer todos los correos electrónicos recibidos o enviados entre las direcciones de correo electrónico DIRECCIÓN_1 y DIRECCIÓN_2. Para ello, se han seguido tres procedimientos distintos.

- Búsqueda en el sistema de ficheros.
- Búsqueda a partir de la imagen del disco.

Búsqueda en el sistema de ficheros

Se ha realizado una búsqueda de correos electrónicos en los directorios de los discos. En concreto se han buscado directorios de gestores de correo electrónico como Outlook, Thunderbird, etc. y también se han buscado en los ficheros temporales de Internet. Se ha encontrado el gestor de correo Evolution en 4 de los discos y el gestor de correo Microdoft Outlook en el disco restante. No se han encontrado indicios de correos electrónicos en el historial de navegación de Internet de ninguno de los discos.

- Disco 01: Se han encontrado dos directorios del gestor de correo Evolution, un directorio con el nombre Evolution y otro con el nombre Evolution-definitivo. Las coincidencias resultado de la búsqueda por la palabra clave DIRECCIÓN_2 fueron en el directorio Evolution-definitivo pero se trataba de la base de datos de la agenda de direcciones de correo. No se ha encontrado correspondencia entre DIRECCIÓN_1 y DIRECCIÓN_2 aunque el hecho de que se haya encontrado la DIRECCIÓN_2 en la agenda puede significar que sí se han intercambiado correos electrónicos en algún momento pero que han sido eliminados.
- Disco 02: Se han encontrado resultados de la búsqueda de DIRECCIÓN_2 en el directorio del gestor de correos Evolution. Se ha realizado una búsqueda con la palabra clave DIRECCIÓN_2 dentro de este directorio y la única coincidencia ha sido la dirección DIRECCIÓN_2 en la base de datos de la agenda de correos, igual que en el disco 01. No se ha encontrado correspondencia entre DIRECCIÓN_1 y DIRECCIÓN_2.

- Disco 03: El procedimiento seguido y los resultados obtenidos han sido los mismos que en el disco 02. No se ha encontrado ninguna coincidencia con la palabra clave DIRECCIÓN_2.

- Disco 04: Se han encontrado resultados de la búsqueda de DIRECCIÓN_2 en el directorio del gestor de correos Microsoft Outlook. En concreto, se han encontrado coincidencias en las bases de datos de la bandeja de entrada y en la bandeja de correos eliminados y también en la agenda de contactos. Para recuperar estos correos se ha usado la herramienta *undbx*, que extrae correos de las bases de datos de Outlook. Además, con la opción de recuperación no sólo extrae los correos, sino que también recupera posibles correos borrados y los presenta en carpetas separadas para distinguirlos.

- Disco 05: Se han encontrado resultados de la búsqueda de DIRECCIÓN_2 en el directorio del gestor de correos Evolution. Se ha realizado una búsqueda con la palabra clave DIRECCIÓN_2 dentro de este directorio y se han encontrado coincidencias en varios ficheros. Por alguna razón que este perito desconoce, las bases de datos no mostraban los datos de cada correo de forma individual, por lo que se han tenido que filtrar los correos a partir de la base de datos completa. Para garantizar el derecho a la intimidad de los usuarios, esta criba se ha realizado de forma automática mediante búsquedas ciegas con las palabras clave DIRECCIÓN_1 y DIRECCIÓN_2.

Resultados

- Disco 01: No se han encontrado correos entre DIRECCIÓN_1 y DIRECCIÓN_2.

- Disco 02: No se han encontrado correos entre DIRECCIÓN_1 y DIRECCIÓN_2.

- Disco 03: No se han encontrado correos entre DIRECCIÓN_1 y DIRECCIÓN_2.

- Disco 04: Se han encontrado correos entre DIRECCIÓN_1 y DIRECCIÓN_2.

- Disco 05: Se han encontrado correos entre DIRECCIÓN_1 y DIRECCIÓN_2.

Búsqueda a partir de la imagen del disco

Se han extraído las cadenas del disco, es decir, todos los caracteres imprimibles del disco. A partir del fichero resultante se han realizado dos búsquedas independientes por las palabras clave DIRECCIÓN_1 y DIRECCIÓN_2 para confirmar la presencia de estas direcciones en el disco. Una vez asegurada la presencia de las direcciones, se ha filtrado toda la información del fichero de cadenas dando como resultado los correos electrónicos enviados o recibidos entre las dos direcciones de correo. Dado que se ha realizado sobre el fichero de cadenas, hay correos que no se han podido recuperar en su totalidad.

Este procedimiento se ha usado con todos los discos y con él hemos encontrado tanto correos que se encuentran en el sistema de ficheros existentes como correos que han sido eliminados.

Resultados

- Disco 01: Se han encontrado correos entre DIRECCIÓN_1 y DIRECCIÓN_2.
- Disco 02: Se han encontrado correos entre DIRECCIÓN_1 y DIRECCIÓN_2.
- Disco 03: No se han encontrado correos entre DIRECCIÓN_1 y DIRECCIÓN_2.
- Disco 04: Se han encontrado correos entre DIRECCIÓN_1 y DIRECCIÓN_2.
- Disco 05: Se han encontrado correos entre DIRECCIÓN_1 y DIRECCIÓN_2.

2.3 (c) Verificar que los correos electrónicos extraídos son íntegros

Una vez extraídos, los correos electrónicos tienen que ser analizados para probar su integridad, es decir, para comprobar que no han sido alterados. Un correo electrónico presentado en papel es susceptible de ser manipulado con facilidad con un editor de textos como Bloc de notas o con un editor de imágenes como Photoshop antes de imprimirse y presentarse (de forma inválida según la opinión de este perito). Un correo electrónico presentado en formato digital también puede haber sido alterado

con facilidad sin que se note en el contenido. Es por este motivo que es importante la adveración de los correos electrónicos. Un correo electrónico está compuesto por la cabecera técnica, en la que hay información imprescindible para que el correo llegue a su destinatario así como de todos los saltos realizados a través de servidores entre su origen y su destino; la cabecera simple, en la que hay información sobre las direcciones de correo del emisor y del receptor, el asunto del correo y la fecha entre otros; el cuerpo del correo, en el que hay el texto enviado por el emisor; y los posibles documentos adjuntos al correo. Para más información se puede consultar el anexo *Procedimiento para análisis de correo electrónico* (apartado 4.6, pág. 97).

Los correos electrónicos extraídos han sido albergados por los gestores de correo electrónico Evolution y Microsoft Outlook. Se ha analizado la información de las cabeceras de todos los correos electrónicos extraídos en su totalidad. Todos los datos analizados son coherentes. Se puede afirmar, por tanto, que todos los correos son íntegros y no existen alteraciones. Estos correos electrónicos se pueden encontrar en el anexo *Correos electrónicos entre DIRECCIÓN_1 y DIRECCIÓN_2* (apartado 4.3, pág. 95). Los correos electrónicos parciales han sido analizados en la medida de lo posible. Se han dividido en dos casos: los correos cuyas cabeceras se han recuperado en su totalidad y que tienen un cuerpo de correo y adjuntos parciales, y los correos cuyas cabeceras no se han podido recuperar en su totalidad y que tienen un cuerpo y adjuntos totales o parciales. Los correos electrónicos de los que se dispone de una cabecera total, han sido analizados. Se puede afirmar que todos ellos son íntegros. Aquellos correos de los que no se dispone de cabeceras totales no han podido ser adverados, no se dispone de información suficiente como para afirmar que no han sido modificados, aunque tampoco se puede afirmar lo contrario.

Destacan aquellos correos que han sido enviados desde DIRECCIÓN_2 a través del gestor de correo Microsoft Outlook por incluir, en su cabecera técnica, la dirección IP de origen del correo que permite determinar el origen de la transmisión del correo electrónico.

2.4 (d) Extraer todos los ficheros eliminados entre FECHA_INICIAL y FECHA_FINAL

El Cliente requiere recuperar los ficheros eliminados entre dos fechas determinadas. Para extraer ficheros eliminados se utilizan herramientas forenses que recuperan ficheros eliminados que aun no han sido sobrescritos. Es conveniente empezar destacando que los discos 01, 02, 03 y 05 utilizan el sistema operativo Ubuntu 10.04.3, mientras que el disco 04 utiliza el sistema operativo Windows XP tal y como se detalla en el anexo *Identificación de los equipos* (apartado 4.5, pág. 95).

El sistema operativo Ubuntu, basado en Linux, tiene un campo de metadatos de los ficheros en el que se registra la fecha de borrado. Sin embargo, la recuperación de ficheros no es una tarea fácil porque los metadatos asociados a un fichero no siempre se recuperan junto a este. Se ha utilizado la herramienta TestDisk porque permite recuperar ficheros enteros, contenido y metadatos asociados. Por contra, no es capaz de recuperar tantos ficheros como otras herramientas. Dado que se han detectado ficheros eliminados que no han podido ser recuperados con TestDisk, se ha decidido utilizar también PhotoRec. Esta herramienta es capaz de recuperar el contenido de muchos ficheros pero aunque estos no llevan los metadatos asociados, por lo que se tiene el contenido de ficheros sin toda la metainformación disponible, como el nombre, propietario del fichero o referencias temporales. Con los equipos que operan con Linux, sí que se puede determinar la fecha en que un fichero fue eliminado porque existe el campo *dtime*, en el que se actualiza la fecha y hora en el momento de ser eliminado. Se ha utilizado una herramienta de *carving* para recuperar ficheros eliminados. Analizando los metadatos de los ficheros recuperados se han encontrado ficheros eliminados que cumplen el requisito de haber sido eliminados entre FECHA_INICIAL y FECHA_FINAL.

Los usuarios del sistema operativo Ubuntu y Linux en general, se caracterizan por utilizar entornos menos gráficos que en Windows, es decir, utilizan más la consola del sistema y no tanto las ventanas características de Windows. Se ha examinado el historial de comandos (acciones) realizadas desde consola y se han encontrado varias ejecuciones del comando `rm` de remove en inglés, y que se utiliza para eliminar ficheros y directorios. El historial de comandos es limitado y sólo se muestra infor-

mación temporal del último comando ejecutado. Examinando los sistemas de ficheros para buscar rastros de las ejecuciones realizadas se ha podido establecer un listado de ficheros fueron eliminados dentro del rango temporal.

Resultados

- Disco 01: XX ficheros eliminados entre FECHA_INICIAL y FECHA_FINAL.
- Disco 02: Se han identificado ficheros eliminados pero no se han podido recuperar, por lo que no se puede comprobar la fecha de eliminación de estos.
- Disco 03: XX ficheros eliminados entre FECHA_INICIAL y FECHA_FINAL.
- Disco 05: XX ficheros eliminados entre FECHA_INICIAL y FECHA_FINAL.

NTFS, el sistema de ficheros utilizado por Windows, no registra datos del momento en el que se borra un fichero dentro de los metadatos del fichero. Esto ocurre porque a un usuario no le interesa la fecha de eliminación de un fichero que está borrando, lo que él quiere es que el fichero desaparezca y, por lo tanto, NTFS se *ahorra* este proceso. Sin embargo, aunque los metadatos de un fichero o directorio eliminado no incluyan cuándo fue borrado, sí que es posible encontrar rastros indirectos de cuándo sucedió este evento. Cuando un fichero o directorio se renombra, se mueve o se elimina, el directorio en el que está contenido detecta el cambio y marca la fecha y hora en la que se produce.

Cada directorio guarda un índice de los ficheros que contiene. Y este índice va aumentando a medida que el directorio aloja más ficheros o subdirectorios. El tamaño del índice no disminuye aunque se elimine contenido del directorio. Por lo tanto, aunque se detecte un directorio eliminado con un índice de cierto tamaño, no se puede asegurar el contenido que se encontraba en el directorio en el momento de ser eliminado, sino el contenido máximo que ha llegado a tener el directorio. Lo que significa que si en el sistema de ficheros del equipo de Windows nos encontramos con un directorio que actualmente está borrado y cuya fecha de última modificación es de un día determinado, no podemos asegurar que el directorio fuese borrado aquel día en concreto, es más, hasta ese momento lo que sí que se puede asegurar es que el directorio no estaba eliminado. Podemos decir que el directorio fue eliminado entre

la fecha de su última modificación y la fecha de última modificación del directorio que lo contiene, porque al borrarse el primero (hijo), se registra una modificación en el segundo (padre). El sistema de ficheros NTFS sí que registra cuándo un fichero ha sido enviado a la papelera de reciclaje.

Para realizar el análisis de ficheros borrados entre FECHA_INICIAL y FECHA_FINAL del disco 04 se han seguido dos pasos:

1. Establecer ventanas temporales mediante la fecha de última modificación de los directorios y subdirectorios.

2. Filtrar los ficheros contenidos en los directorios por sus fechas de acceso.

De forma complementaria, se ha elaborado un listado de los ficheros enviados a la papelera de reciclaje entre FECHA_INICIAL y FECHA_FINAL. Estos procedimientos se detallan a continuación.

Establecer ventanas temporales mediante la fecha de última modificación de los directorios y subdirectorios

Como se acaba de explicar, el sistema de ficheros NTFS no registra la fecha y hora de eliminación de los directorios ni de los ficheros, por lo que es imposible asegurar la fecha en la un fichero o directorio fue eliminado. Examinando los metadatos de los directorios se ha elaborado un listado de los directorio que han sido eliminados con certeza dentro del periodo establecido, entre FECHA_INICIAL y FECHA_FINAL. Por la estructura en la que se diseñó el sistema de ficheros NTFS, los ficheros y subdirectorios alojados en un directorio se van añadiendo en nodos. Los nodos permanecen aunque se elimine el fichero o subdirectorio y no se registra la fecha de eliminación del fichero o subdirectorio. Los directorios eliminados entre FECHA_INICIAL y FECHA_FINAL tienen un cierto número de nodos que indican el número máximo de ficheros y subdirectorios que han llegado a alojar en algún momento, pero no se puede asegurar o desmentir la presencia de todos estos ficheros y subdirectorios en el momento de eliminar los ficheros.

Filtrar los ficheros contenidos en los directorios por sus fechas de acceso

Para determinar qué ficheros fueron eliminados en la ventana temporal requerida se extrajo la línea temporal (o *timeline*) de la actividad de cada equipo. Como se acaba de explicar, salvo en ocasiones muy puntuales, como es el uso de la papelera de reciclaje, resulta imposible determinar con absoluta certeza la fecha exacta del borrado de un fichero. No obstante, se puede asegurar que ciertos ficheros han sido eliminados dentro de la ventana temporal que va desde FECHA_INICIAL hasta FECHA_FINAL.

Se han buscado en la *timeline* los ficheros contenidos dentro de los directorios eliminados entre FECHA_INICIAL hasta FECHA_FINAL. Se puede afirmar que todos aquellos ficheros cuya fecha de acceso sea posterior a FECHA_INICIAL han sido eliminados dentro del periodo indicado.

En concreto, se han localizado XX ficheros y XX directorios. El listado completo se puede consultar en el anexo *Ficheros eliminados entre FECHA_INICIAL y FECHA_FINAL* (apartado 4.4, pág. 95).

Análisis de los historiales de la papelera de reciclaje

Como se ha explicado anteriormente en este mismo apartado, las marcas temporales de los ficheros eliminados a través de la papelera de reciclaje sí que se actualizan, en concreto se modifican en el momento en que un fichero es enviado a la papelera, tanto en Linux como en Windows. No sólo se registra el momento en el que el fichero fue enviado a la papelera de reciclaje sino que los metadatos conservan el directorio en el que se encontraba el fichero antes de ser enviado a la papelera. No ha sido posible analizar los metadatos asociados a los ficheros de la papelera del disco 02 por la irregularidad detectada en la copia proporcionada, pero se confirma la existencia de ficheros en la papelera de reciclaje. Los ficheros que en el momento del análisis se han encontrado en la papelera se pueden ver en el anexo *Ficheros eliminados entre FECHA_INICIAL y FECHA_FINAL* (apartado 4.4, pág. 95) aunque, insistimos, estos ficheros no pueden considerarse como eliminados.

3 Conclusiones

De todo lo expuesto anteriormente se puede concluir que:

- **Se han encontrado un total de XX ficheros que, en su nombre, contienen las palabras PALABRA_A y/o PALABRA_B.**
- En concreto:
  - Se han encontrado XX ficheros con sólo la palabra PALABRA_A en su nombre.
  - Se han encontrado XX ficheros con sólo la palabra PALABRA_B en su nombre.
  - Se han encontrado XX ficheros con la palabra PALABRA_A y la palabra PALABRA_B en su nombre.
- **Se han extraído un total de XX correos electrónicos enviados o recibidos entre DIRECCIÓN_1 y DIRECCIÓN_2.**
- En concreto:
  - Se han extraído XX correos electrónicos completos. Todos ellos son veraces.
  - Se han extraído XX correos electrónicos con cabeceras completas y contenidos parciales. Todos ellos son veraces.
  - Se han extraído XX correos electrónicos con cabeceras parciales. No se puede afirmar ni desmentir su veracidad.
- **Se han recuperado XX ficheros eliminados entre FECHA_INICIAL y FECHA_FINAL.**
- De los ficheros recuperados, hay un total de XX ficheros de los que no se ha podido recuperar el nombre ni la ruta en la que se encontraban antes de ser eliminados.
- Aunque no pueden considerarse como eliminados, se han encontrado XX ficheros que fueron enviados a la papelera de reciclaje entre FECHA_INICIAL y FECHA_FINAL.

- Los resultados del disco 02 podrían ampliarse si se demuestra que la copia realizada en dependencias notariales tuvo algún error y se procede a realizar una nueva copia forense.

Y esto es todo cuanto tengo que decir a mi buen saber y entender.

4 Anexos

En los siguientes apartados se incluirían los anexos del informe en los que se muestra información más técnica de la investigación y el detalle de los resultados. Esta información se incluye en los anexos y no en el cuerpo del informe para facilitarse la lectura del mismo.

4.1 Aseguramiento de las fuentes de información

Este apartado recoge el proceso de identificación, extracción y copia de los discos. Para que este proyecto de fin de carrera no revele datos reales del caso, sólo se incluyen dos fotografías.

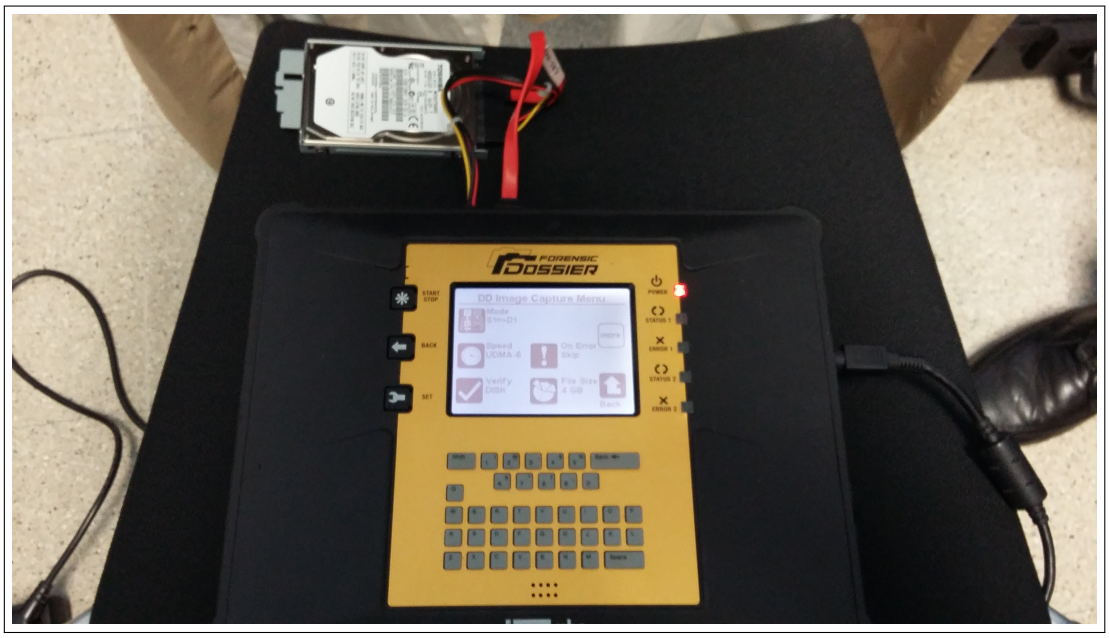

Figure 1: Clonación de un disco mediante la clonadora.

Figure 2: Sobre sellado que asegura la cadena de custodia.

4.2 Ficheros con las palabras PALABRA_A y/o PALABRA_B en el nombre

En este anexo se incluiría un listado de todos los ficheros encontrados cuyos nombre contengan la palabra A o la palabra B separados por el disco en el que se han encontrado y clasificados por formato (doc, xls, lnk, etc). De ser necesario, se entregarían los ficheros en un dispositivo de almacenamiento externo como un USB o un CD.

4.3 Correos electrónicos entre DIRECCIÓN_1 y DIRECCIÓN_2

En este anexo se reproducirían los correos electrónicos enviados y recibidos entre DIRECCIÓN_1 y DIRECCIÓN_2 separados por disco de procedencia y ordenados por fecha. En el caso de haber una gran cantidad de correos electrónicos, que es lo que pasó en esta investigación, se optó por reproducir en este anexo sólo los correos electrónicos que el Cliente consideró relevantes. El total de correos electrónicos fueron entregados en un dispositivo de almacenamiento externo USB.

4.4 Ficheros eliminados entre FECHA_INICIAL y FECHA_FINAL

En este anexo se incluiría un listado de todos los ficheros y directorios que pueden asegurarse que fueron borrados entre FECHA_INICIAL y FECHA_FINAL separados por disco de procedencia y ordenados cronológicamente. De ser necesario, se entregarían los ficheros en un dispositivo de almacenamiento externo como un USB o un CD. Como subapartado del mismo anexo, se incluiría también un listado de todos los ficheros que se encontraban en la papelera de reciclaje en el momento de realizar la investigación y que cumplen los requerimientos temporales.

4.5 Identificación de los equipos

A continuación se detallan las características principales de cada equipo:

- **Disco-01**
  - Nombre del equipo: XXX-09
  - Sistema operativo: Ubuntu 10.04.3 LTS
  - Fecha de instalación: 10 de septiembre de 2010 a las 14:05:57 (UTC +2)
  - Fecha último apagado: 14 de febrero de 2012 a las 18:38 (UTC)
  - Usuarios: XXX
  - Fecha último login del usuario: 14 de febrero de 2012 a las 18:37 (UTC)
- **Disco-02**
  - Nombre del equipo: XXX-04

  - Sistema operativo: Ubuntu 10.04.3 LTS
  - Fecha de instalación: 25 de septiembre de 2010 a las 08:34:41 (UTC +2)
  - Fecha último apagado: 14 de febrero de 2012 a las 19:41 (UTC)
  - Usuarios: XXX
  - Fecha último login del usuario: 14 de febrero de 2012 a las 18:24 (UTC)

- **Disco-03**
  - Nombre del equipo: XXX-06
  - Sistema operativo: Ubuntu 10.04.3 LTS
  - Fecha de instalación: 24 de septiembre de 2010 a las 15:36:09 (UTC +2)
  - Fecha último apagado: 14 de febrero de 2012 a las 19:22 (UTC)
  - Usuarios: XXX
  - Fecha último login del usuario: 14 de febrero de 2012 a las 18:21 (UTC)

- **Disco-04**
  - Nombre del equipo: DESKTOP
  - Sistema operativo: Windows XP SP2
  - Fecha de instalación: 3 de junio de 2011 a las 14:05:57 (UTC)
  - Fecha último apagado: 14 de febrero de 2012 a las 16:16:26 (UTC)
  - Usuarios: Administrador
  - Fecha último login del usuario: 8 de febrero de 2012 a las 07:36:43 (UTC)

- **Disco-05**
  - Nombre del equipo: XXX-05
  - Sistema operativo: Ubuntu 10.04.3 LTS
  - Fecha de instalación: 10 de diciembre de 2010 a las 08:31:22 (UTC +1)
  - Fecha último apagado: 14 de febrero de 2012 a las 16:49 (UTC)
  - Usuarios: XXX
  - Fecha último login del usuario: 14 de febrero de 2012 a las 16:47 (UTC)

4.6 Procedimiento para análisis de correo electrónico

Un correo electrónico es un servicio que permite a los usuarios enviar y recibir mensajes mediante sistemas de comunicación electrónica. Un correo electrónico es susceptible de ser alterado o manipulado. La presentación de un correo como prueba debe hacerse en formato digital y con determinadas garantías como la custodia del medio que los alberga, ya sea el buzón de correo electrónico en el que se encuentra o el correo web (*webmail*) en el que está contenido el correo electrónico. El estudio de la *adveración de correos* consiste en un análisis de las fuentes de información disponibles en cada caso (cabeceras de correo, propiedades del buzón de correo, registros del servidor o *logs*, etc.) para determinar si los correos conservan su integridad o si, por el contrario, se observan incoherencias en los datos existentes en dichas fuentes de información que puedan indicar que esos correos electrónicos han podido sufrir alguna manipulación.

En los casos en los que sea posible, es especialmente relevante disponer de buzón de correo en el que originalmente residen los correos electrónicos objeto del trabajo de adveración, ya que los buzones de correo de los distintos gestores de correo que existen en el mercado (Microsoft Outlook, Lotus Notes, Thunderbird, Evolution, etc.) contienen una serie de metainformación introducida por el gestor de correo que permite su correcto funcionamiento. Esta información es incorporada de forma transparente por el programa gestor de correo y no puede ser manipulada directamente por el usuario, por lo que el análisis de sus valores permite deducir, a partir de su estudio detallado, las acciones realizadas por cada mensaje, incluyendo la presencia o ausencia de posibles modificaciones a las que el correo haya podido ser sometido tras ser enviado/recibido.

En el caso de que el servicio de correo electrónico sea proporcionado directamente mediante el explorador de Internet (Firefox, Chrome, etc.), el gestor de correo no es un programa independiente usado para tal fin sino que se utiliza un correo web o *webmail*, que es un cliente de correo electrónico que provee una interfaz web por la que se accede al correo electrónico. Los más populares son *Gmail*, *Hotmail* o *Yahoo*. Cuando el correo electrónico que se desea analizar provenga de *webmail*, la fuente de información objeto de análisis que requiere ser asegurada es el contenido

completo de dicho correo electrónico, es decir, el cuerpo del correo en el que está el contenido del mismo, los ficheros adjuntos y la cabecera completa. La cabecera de un correo electrónico, está formada por una serie de datos identificativos del mensaje, del cuerpo y de los datos adjuntos al mensaje de correo electrónico. En realidad un correo electrónico es el conjunto de las cabeceras, cuerpo y datos adjuntos, pero debido a que no aportan demasiada información al usuario y que, dada su extensión y complejidad técnica, suponen más bien una molestia para el mismo, la mayoría de los programas y sistemas de correo electrónico no muestran, por defecto, los datos de la cabecera. Sin embargo, la gran mayoría de éstos presentan diferentes opciones para poder visualizarlas.

La cabecera simple está formada por los campos que se muestran al abrir el correo electrónico. Estos campos suelen mostrarse en el idioma en el que se ha configurado el gestor o el cliente de correo. Los campos más importantes de la cabecera simple son los siguientes:

- De:
- Para:
- Asunto:
- Fecha:

La cabecera técnica contiene información muy valiosa desde el punto de vista técnico que ayudan al investigador o perito forense a establecer una trazabilidad del correo en Internet, pudiendo determinar, en algunas de las ocasiones, desde qué dirección IP se envió el mensaje [1] (entre otras cosas). En la cabecera técnica, todos los campos están escritos en inglés, empiezan con mayúscula y acaben en dos puntos (:). Algunos de los campos que se muestran en la cabecera técnica son los siguientes:

- Delivered-To:
- Received:
- Return-Path:

[1] La mayoría de clientes de correo web, como Gmail, no proporcionan la dirección IP desde la que se ha enviado el correo para ofrecer privacidad al usuario.

- In-Reply-To:
- Message-ID:

El aseguramiento debe realizarse en el formato original, es decir, en formato digital puesto que se trata de un correo electrónico, almacenándolo en un dispositivo adecuado tal como una memoria USB (o pendrive) u otro dispositivo de almacenamiento de datos. En este caso también deben ser asegurados los registros de actividad asociados a esa cuenta de correo (los *logs*), que proporciona el servidor de correo electrónico del proveedor del servicio (Internet Service Provider o ISP), cuyo análisis permitirá estudiar la coherencia de los datos aportados respecto a aspectos tan relevantes como fecha y hora de acceso a la cuenta de correo, envío y recepción de mensajes, entre otros. Una vez aseguradas dichas fuentes, se inicia la cadena de custodia que permitirá garantizar la veracidad de las conclusiones alcanzadas tras el análisis que se pretenda realizar.

Un correo electrónico dispone de diversos campos con información, algunos de ellos están estandarizados y otros los añade el programa que gestiona el correo electrónico. A continuación se puede ver un ejemplo de cabecera de correo electrónico, para una lectura más cómoda se han eliminado y reducido campos que no tienen valor para la adveración del correo. [2]

[2]Los valores que aparecen en el siguiente ejemplo han sido inventados a modo de ejemplo.

```
Delivered-To: direccion1@empresa1.com
Received: by 10.194.33.39 with SMTP id o7cspi;
        Tue, 15 Nov 2011 07:19:07 -0700 (PDT)
X-Received: by 10.202.242.137 with SMTP id q131m;
        Tue, 15 Nov 2011 07:19:07 -0700 (PDT)
Return-Path: <direccion2@empresa2.com>
Received: from EUR01-HE1-obe.outbound.protection.outlook.com
          (mail-he1eur01on0065.outbound.protection.outlook.com. [104.47.0.65])
        by mx.google.com with ESMTPS id z194 for <direccion1@empresa1.com>
        Tue, 15 Nov 2011 07:19:07 -0700 (PDT)
Received-SPF: pass (google.com: domain of direccion2@empresa2.com designates

              104.47.0.65 as permitted sender) client-ip=104.47.0.65;
Received: from AM2PR07MB0865.eurprd07.prod.outlook.com (10.161.71.151) by
AM2PR07MB0867.eurprd07.prod.outlook.com (10.161.71.153) with Microsoft SMTP
Server (TLS) id 15.1.501.7; Tue, 15 Nov 2011 14:19:04 +0000
from AM2PR07MB0865.eurprd07.prod.outlook.com ([10.161.71.151]) by
AM2PR07MB0865.eurprd07.prod.outlook.com ([10.161.71.151]) with mapi id
15.01.0506.013; Tue, 15 Nov 2011 14:19:04 +0000
From: "direccion2@empresa2.com" <direccion2@empresa2.com>
To: "direccion1@empresa1.com" <direccion1@empresa1.com>
Subject: Material
Thread-Topic: Material

Date: Tue, 15 Nov 2011 14:19:04 +0000
Message-ID: <AM2PR07MB0865B8865.euasdsd07.prod.outlook.com>
x-originating-ip: [212.145.159.148]
Content-Type: multipart/related;
boundary="_004_AM2PR07MB0865B8A0143AE65eurp_";
type="multipart/alternative"
MIME-Version: 1.0
X-MS-Exchange-CrossTenant-originalarrivaltime: 15 Nov 2011 14:19:04.2565(UTC)
Content-Type: multipart/alternative;
Content-Type: text/plain; charset="iso-8859-1"

Este es el cuerpo del correo electrónico.
```

A continuación se explican los campos más destacados para la trazabilidad y autenticidad del correo:

- *Received:* en las cabeceras de un correo electrónico este campo suele aparecer varias veces. Cada vez que el correo electrónico pasa por un servidor se le añade un campo *Received*. Así, este campo adquiere un valor distinto para cada uno de los servidores por los que haya pasado un correo electrónico, empezando por el más cercano al origen de la comunicación y hasta el de su destino. El valor de este campo es el de la dirección IP que tengan los sucesivos servidores intervinientes. En este campo se muestra la fecha y hora en la que el correo ha pasado por cada servidor.

- *x-originating-ip:* de la cabecera añadida por el gestor de correo Microsoft Outlook. El valor de este campo es la dirección IP desde la que se ha enviado el correo electrónico.

Por tanto, es imprescindible disponer de la información de estos campos cuando se realiza cualquier tipo de consideración acerca de un correo electrónico y más aún cuando sea necesario realizar un análisis para, entre otras valoraciones, poder pronunciarse acerca de la localización desde la que el correo fue enviado o sobre la posibilidad que la integridad del correo haya podido ser comprometida. La única forma de poder disponer de esta información y realizar dichas consideraciones es preservando el correo electrónico en formato digital, lo cual se consigue mediante el proceso de aseguramiento del mismo como fuente de información que es, tal y como se ha comentado en el presente apartado.