Determinación de aspectos carentes en un Proceso Unificado de Recuperación de Información digital.

Ana Haydée Di Iorio[1], Rita Evelina Sansevero[2],

Martín Castellote[3], Ariel Podestá[4], Fernando Greco[5],

Bruno Constanzo[6], Julián Waimann[7]

[1] Ingeniera en Informática, Instructor Informático en el Ministerio Público de la Provincia de Buenos Aires, Docente e Investigadora en Universidad FASTA, diana@ufasta.edu.ar

[2] Ingeniera en Informática. Docente e Investigador de la Facultad de Ingeniería de la Universidad FASTA. Programador/Analista de la Municipalidad del Partido de General Pueyrredon, resansevero@gmail.com

[3] Ingeniera en Informática. Docente e Investigador de la Facultad de Ingeniería de la Universidad FASTA. Bioinformático en INTA EEA Balcarce. Desarrollador web. castellotemartin@yahoo.com.ar

[4] Ingeniero Informático. Docente e Investigador de la Facultad de Ingeniería de la Universidad FASTA. Programador/Analista de la Municipalidad del Partido de General Pueyrredon, arielpodesta@gmail.com

[5] Ingeniero en Informática, Instructor Informático en el Ministerio Público de la Provincia de Buenos Aires, Docente e Investigador en Universidad FASTA, fmartingreco@gmail.com

[6] Técnico en Informática. Auxilliar de Investigación Alumno, Facultad de Ingeniería de la Universidad FASTA. Bru.constanzo@gmail.com

7 Técnico en Informática. Auxilliar de Investigación Alumno, Facultad de Ingeniería de la Universidad FASTA. Desarrollador .Net Common Sense julianw@ufasta.edu.ar

Resúmen: Con el advenimiento de la Sociedad de la Información, las nuevas tecnologías irrumpen en practicamente todos los aspectos de nuestra cotidianeidad. Es así que, la recuperación de la información digitalizada pasa a ser considerado un aspecto crítico, tanto en ámbitos privados como públicos. Vinculado este tema a la resolución de conflictos judiciales, una de las mayores problemáticas es la falta de un proceso unificado que guíe a los expertos forenses en esta tarea tan compleja, ya sea por la variedad de plataformas tecnológicas que pueden encontrarse, como por la diversidad de formas que puede tomar una evidencia digital. Se presenta en éste trabajo una propuesta de "Proceso Unificado de Recuperación de la Información", resultado de un Proyecto de Investigación de la Facultad de Ingeniería de la Universidad FASTA – Mar del Plata – Argentina, que pueda llenar este vacío de la ciencia forense informática, y colabore además en la determinación de los puntos carentes de herramientas y técnicas aplicables.

Palabras Claves: Informática Forense – Evidencia Digital - Peritaje Informático – Recuperación de la Información.

1. Introducción

En los últimos años, las sociedades en todo el mundo han experimentado un cambio progresivo en el que cada vez más se depende de sistemas informáticos para manipular información. Este cambio se debe, entre otras cosas, al aumento en la cantidad de información con la que se cuenta, a la transferencia de procesos tradicionales a sistemas informáticos, servicios y productos que buscan, simplificar tareas de la vida moderna.

Con esta migración hacia sistemas informáticos para administrar gran parte de los aspectos de nuestra vida diaria, se produce también una situación de dependencia que implica comenzar a considerar las distintas formas en que podemos recuperar nuestra información en caso de eliminarse o perderse por algún motivo.

En este contexto de complejidad considerable y velozmente cambiante, existen dificultades a sortear en la tarea de recuperación de información almacenada digitalmente, entre las que podemos mencionar: diferentes plataformas de base, diversidad de métodos de almacenamiento, dificultad para la localización de la información, tecnologías que naturalmente eliminan evidencias, mecanismos internos de protección de la información, falta de herramientas específicas, uso de criptografía, aplicaciones que cubren solo una parte del proceso de recuperación de la información y de efectividad desconocida, legislaciones no uniformes en diferentes países, falta de guías oficiales nacionales para realizar tareas de recuperación y de mecanismos de validación de lo actuado.

La tarea de recuperación de la información tiene un aspecto forense cuando se la utiliza en procesos judiciales para obtener o corroborar evidencia obtenida. En este caso, además de recuperar la información deseada, se busca que el procedimiento realizado sea reproducible y tenga la capacidad de ser auditado. Si bien existen guías o recomendaciones sobre cómo realizar procesos de recuperación de información en ámbitos forenses, no hay ningún proceso formal establecido, evaluado y oficializado por nuestro país

Desde el año 2001 diferentes autores y organizaciones han estado trabajando en guías de buenas prácticas en informática forense. En la investigación que se presenta se analizó y tomó como referencia los siguientes trabajos:

- ACPO (Association of Chief Police Officers) – England, Wales and North Ireland. Good Practice Guide for Computer-Based Electronic Evidence. Oficial release version. [23]
- A guide to basic computer forensics – TechNet Magazine [24]
- NIJ (National Institute of Justice) Report – United States of America, Department of Justice. Forensic Examination of Digital Evidence: A guide for Law Enforcement. [15]
- Law Enforcement Investigations - Active Army, Army National Guard, and US Army Reserve. FM 3-19.13. Chapter 11: Computer Crimes [25]
- Metodologías, Estrategias y Herramientas de la Informática Forense aplicables para la dirección nacional de comunicación y criminalistica de la policia Nacional de Ecuador. [26]
- RFC 3227: Guía Para Recolectar y Archivar Evidencia" (Guidelines for Evidence Collection and Archiving) [27]
- Guía de la IOCE (International Organization on Computer Evidence) "Guía para las mejores practicas en el examen forense de tecnología digital" [28]
- Guía de Mejores prácticas de la ISFS (Information Security and Forensic Society (Sociedad de Seguridad Informatica y Forense) Hong Kong . [29]

– Guía Para El Manejo De Evidencia En IT - Estándares de Australia. APEC Telecommunications and Information Working Group.[30]

Al analizar estas guías de buenas prácticas se detectó que si bien estas guías constituyen un excelente aporte procedimental, muchas abarcan solo una parte del proceso [15] [23], otras son muy generales [24] [28] y otras focalizan en los temas delictivos [26] [27] [28]. Por otro lado desde el punto de vista práctico no abordan las técnicas existentes para realizar ciertas tareas, las herramientas disponibles en el mercado, así como tampoco las diferentes alternativas de acuerdo a la plataforma de software del equipo a periciar.

Considerando esta situación, se comenzó en la Universidad FASTA el Proyecto de Investigación "PURI - Proceso Unificado de Recuperación de la Información", con el fin de proveer un proceso que pueda ser probado, evaluado, utilizado y mejorado por expertos en informática forense.
La presentación de este proceso es para que sea discutido, evaluado, ampliado, mejorado, y que sirva, en definitiva, a la tarea cotidiana del informático forense y del operador de justicia.

2. El Proyecto PURI

El proyecto PURI nace de la conjunción de dos cátedras de la facultad de Ingeniería de la Universidad FASTA, la cátedra de Sistemas Operativos y la de Informática y Derecho.
El objetivo principal del proyecto es crear un proceso unificado de recuperación de información a partir de la estructuración y organización de las fases, tareas, técnicas y herramientas que lo componen, respetando las buenas prácticas propuestas por los organismos internacionales, de manera que se convierta en un elemento de guía y consulta para los profesionales forenses, tanto en el ámbito comercial como en el judicial.
A lo largo del desarrollo del proyecto se detectó que ciertas tareas definidas en el proceso PURI carecían de técnicas y herramientas que pudieran efectuarlas. Es así que se seleccionaron dos casos y se trabajó en el desarrollo de prototipos de las herramientas ad hoc.
Se presentan en este trabajo la totalidad de las áreas de vacancia de técnicas y herramientas con el fin de darlas a conocer a la comunidad científica, y que de esta manera, puedan aportar a la solución de las tareas referidas.
En el detalle del proceso PURI se sugieren también ciertas herramientas existentes en el mercado, que se seleccionaron teniendo en cuenta las siguientes variables: las pruebas de efectividad técnica realizadas por el equipo de investigación, el tipo de licencia del software (es decir, si se trata de software libre o propietario, si es Open Source o no), la compañía de respaldo, y el grado de madurez de la técnica.

3. El Proceso Unificado de Recuperación de la Información

El proceso PURI se define entonces como una secuencia de fases compuestas por etapas que involucran tareas a llevar a cabo para recuperar información almacenada digitalmente, aplicando técnicas implementadas en herramientas concretas que permiten ejecutar dichas tareas. A continuación se presenta un resumen del proceso propuesto. [13]

3.1 Fase de Adquisición

Esta fase comprende toda actividad vinculada con la generación de una réplica exacta de todo el contenido digital alojado en el dispositivo original.
La fase de adquisición comprende etapas que de acuerdo al entorno en el que se deba llevar a cabo la recuperación de la información, aplicará o no involucrarlas en el proceso. Es así que se procedió a dividir la adquisición de dispositivos móviles de otros dispositivos por sus características altamente diferenciadoras a todo nivel, tanto físico (hardware), cómo lógico (software).

3.2 Fase de Preparación

Esta fase involucra todos los procedimientos necesarios para generar el entorno de pruebas preciso para llevar a cabo en primer lugar la inspección, y eventualmente la recuperación de la información.
Como primera etapa, la fase de preparación contempla la restauración de la imagen. Esto significa que si la misma se encontrara dividida, encriptada o comprimida deberá realizarse el proceso contrario, a fin de lograr el original.
A continuación se deberá validar que la restauración ha sido exitosa mediante un algoritmo de hash, como se mencionó previamente.
Si la imagen que se obtuvo es de un sistema de archivos de un determinado sistema operativo, entonces será útil generar una máquina virtual que tome dicha imagen como su disco principal. Al hacerlo se debería realizar una copia a fin de no alterar la imagen original.
Finalmente esta etapa contempla la identificación de tipos de sistemas de archivos y sistemas operativos contenidos en los medios de almacenamiento originales.

3.3 Fase de Análisis

Esta fase comprende el fuerte del trabajo en donde se analiza el contenido adquirido en busca de vestigios de lo que se quiere hallar. El objetivo final de la fase de análisis en el caso de un proceso judicial o pre-judicial es encontrar la denominada Evidencia Digital, es decir, aquello que relaciona el hecho ocurrido con el "imputado" y la "victima". Entonces, se piensa en la evidencia digital como en un tipo de evidencia física que está construida de campos magnéticos y pulsos electrónicos que pueden ser recolectados y analizados con herramientas y técnicas especiales.
La fase de análisis comprende las siguientes etapas:

a) Extracción lógica

b) Extracción física

c) Análisis de relaciones

La extracción lógica representa la recuperación de información eliminada a partir del sistema de archivos. Por esa razón se denomina "lógica", ya que no se accede en forma directa a los bloques, sino a través del Sistema de Archivos, y del Sistema Operativo como intermediario.
La extracción física, en cambio, va directo al dispositivo, eludiendo el Sistema Operativo. La mayoría de los sistemas operativos no eliminan la información en el momento en el que un Usuario solicita el borrado de un archivo determinado, sino que, de alguna manera, dejan registrado que el espacio que ocupaba dicho archivo ahora se encuentra disponible. De esta manera, por ejemplo, si fuese posible hallar tal espacio entonces sería posible reconstruir la información original.

La etapa de análisis de relaciones trata justamente de identificar relaciones entre conjuntos de archivos, con el fin de obtener una conclusión. Esto involucra puntualmente la Identificación de relaciones entre conjunto de archivos vinculados a una actividad en particular (ej: archivos relacionados a la navegación por internet) y la verificación de aplicaciones instaladas, entre otros.

Un punto interesante es comparar el proceso PURI con las guías de procedimiento mencionadas, sin embargo, este objetivo excede la finalidad de este paper.

4. Aspectos carentes detectados en el proceso

En el proceso detallado en la sección anterior se detectaron áreas carentes de técnicas y/o herramientas que se detallarán a continuación. Luego, en las secciones 5 y 6 se tratarán específicamente los aspectos carentes que están siendo abordados por el equipo de investigación a fin de proponer un prototipo de herramienta que venga a llenar este vacío.

4.1 Carencias en Fase de Adquisición

4.1.1 Adquisición de memoria volátil en dispositivos móviles:

Al momento de investigar las herramientas disponibles en el mercado durante el año 2011 no se contaba con ningún programa que pudiera realizar la adquisición de la memoria volátil (RAM) de un dispositivo móvil. A comienzos del año 2012, Sylve, Case, Marziale y Richard presentaron DMD (Droid Memory Dumper), una herramienta para realizar la adquisición de la memoria de dispositivos con el sistema operativo Android. Aún hacen falta, sin embargo, herramientas para poder realizar la adquisición en dispositivos con iOS, Windows Phone, Symbian y Blackberry OS, mas allá de que Android es, al día de hoy, el sistema operativo móvil con mayor market share [1][2].

4.2 Carencias en Fase de Preparación

4.2.1 Identificar y quebrar medios de encriptación de información – Bitlocker

BitLocker es una funcionalidad de Windows Vista, 7 y 2008 Server que permite el encriptado de volúmenes completos. Su finalidad es proveer un potente método de protección de datos sin obstaculizar el acceso a los usuarios admitidos.

El principio de BitLocker es realizar las verificaciones de seguridad antes de integrar la unidad de almacenamiento al sistema. Esto también incluye el mismo disco principal en el que se encuentra instalado el sistema operativo. De esta manera, si el disco es extraído y conectado en otro equipos se interrumpiría el inicio del sistema, al no poder validarse las credenciales.

Entre las modalidades que BitLocker ofrece al usuario se encuentran “Modo de autenticación de usuario” y “Modo en dispositivos USB” los cuales pueden o no hacer uso de un TPM (Trusted Platform Module) para robustecer la seguridad de estos modos.

- Modo de autenticación de usuario: Este modo requiere que el usuario proporcione alguna contraseña de autenticación durante el arranque del sistema. Si la validación es exitosa entonces el TPM (Trusted Platform Module, dispositivo que almacena las claves del algoritmo de encriptado) libera la clave de desencriptado del volumen. Aquí se presenta la oportunidad de simular tal validación introduciendo bootkits.

- Modo en dispositivos USB: El usuario debe insertar un dispositivo USB que contenga una clave de inicio en el equipo para poder arrancar el sistema operativo protegido. Esta modalidad requiere que la BIOS de la máquina protegida acepte el arranque desde dispositivos USB. Del mismo modo, existe la posibilidad de utilizar un bootkit que emule tal validación sin llevarla a cabo realmente.

Cada modo de funcionamiento presenta algunas vulnerabilidades aprovechables que permitirían acceder al medio de almacenamiento salteando las validaciones propias de BitLocker. Una técnica para quebrar ese medio, como se mencionó anteriormente, es la inyección de "bootkits"en los que el cargador del sistema (boot loader) es reemplazado por una copia adulterada que saltea las autenticaciones requeridas.

Durante la investigación no se hallaron herramientas que ofrezcan hacer uso de tales vulnerabilidades. Por ello mismo, es que se considera esta problemática como carente de técnicas y de herramientas.

Existe un tercer modo de funcionamiento para BitLocker, que solo funciona con la presencia de una Trusted Platform Module y se denomina "Modo de funcionamiento transparente". El mismo aprovecha las capacidades del chip Trusted Platform Module el cual contiene códigos de seguridad encriptados por él mismo que no liberará hasta que ciertas verificaciones de integridad de hardware y del sistema sean realizadas en el arranque. Tales códigos son necesarios para el desencriptado del volumen.

La problemática existente con este modo es que si el TPM se encuentra integrado a la placa madre y ésta deja de funcionar, entonces no se puede acceder al TPM y se considera virtualmente imposible recuperar la información no disponiendo de un hardware que interactúe con el TPM para recuperar la clave de cifrado de la información. Dado que no se hallaron dispositivos que puedan llevar a cabo tal operación, es considerado un punto carente.

4.2.2 Identificar máquinas virtuales presentes en el equipo

Hoy en día la utilización de máquinas virtuales es una actividad que no solo está ampliamente difundida sino que ya es una herramienta en muchas plataformas empresariales. Potencialmente cada maquina virtual simula un equipo real con toda su complejidad, lo que permite que la información pueda ocultarse o perderse aún dentro de ellas. Esto implica que, cada máquina virtual hallada dentro del equipo físico a analizar, debería tratarse como un equipo por separado. Así, la problemática inicial que existe es verificar la existencia de maquinas virtuales en el equipo original.

Durante la investigación, no se halló una herramienta que permitiera efectuar tal operación.

4.3 Carencias en Fase de Extracción – Etapa Extracción Lógica

4.3.1 Extracción de archivos comprimidos

Una forma de ocultar la información es comprimiéndola. De esta manera, el archivo queda en otro formato, y se requiere un proceso de descompresión previa a su lectura o edición.

Puede suceder que el volumen de información a descomprimir sea demasiado grande y no se disponga de espacio suficiente para almacenar los datos descomprimidos, o simplemente que se quiera buscar determinado contenido u archivo tratando de evitar la descompresión. Siendo así, se requiere de una herramienta que permita trabajar con los archivos comprimidos del mismo modo que si no lo estuvieran.

Durante la investigación se halló software capaz de realizar tal operación pero solamente con determinados formatos de compresión. Se hace visible la necesidad de contar con una

herramienta que contemple la mayor cantidad posible de formatos de compresión y que permita trabajar normalmente con información comprimida.

4.3.2 Búsqueda de Información en el área de paginado

Todas las técnicas de análisis de memoria dependen de la capacidad del examinador para traducir las direcciones virtuales que utilizan los procesos y componentes del sistema operativo, a la ubicación real de los datos en la imagen de la memoria.
En general, el mecanismo de traducción de direcciones virtuales se basa en las direcciones que apuntan a los datos que había en la memoria principal, utilizadas por un programa, que no estuvieran en transición y sin modificar.
La memoria se divide en páginas o marcos. Una página que no podría ser utilizada de inmediato por un proceso porque no se encuentra cargada en memoria principal, no significa que no sea accesible para el sistema operativo y su incorporación en el análisis de la memoria crea una imagen más completa para aprovechar al máximo los recursos disponibles.
El examinador puede entonces seguir las mismas reglas que el sistema operativo para acceder a los datos en cuestión.
Una de las posibilidades es que la página se encuentra en algún archivo de paginado del sistema operativo, llamo usualmente PAGEFILE.
Cuando el sistema operativo se queda sin memoria física, almacena las páginas en el archivo de paginación y las entradas correspondientes de la tabla de páginas de dichas páginas apuntan a frames en uno de los archivos de paginación.
Para poder utilizar el archivo de paginación en relación a los procesos en ejecución, si en el momento de adquisición el equipo se encuentra encendido, es preciso que se adquiera también la memoria física. Además, se debe contar con una herramienta que permita identificar dentro del archivo los bloques correspondientes a cada página para luego mapearlos con los procesos que estaban en ejecución a los que correspondían.
Con las herramientas existentes hoy día, las búsquedas en el pagefile son por cadenas de texto (contenido) y en todo el archivo de paginado, lo que suele ser practicamente inutilizable.
Con la herramienta propuesta, que vincule el pagefile con la memoria física del equipo, no sólo se podrían recuperar de un volcado de memoria las páginas pertenecientes a un proceso que estaban en memoria principal, sino también las que se encuentran en el área de paginado, logrando así especificar más las búsquedas por cadenas en el pagefile acotándolas a las páginas de un proceso en particular.

4.3.3 Extracción de archivos encriptados

En escenarios en los que los archivos a extraer se encuentran encriptados, el primer paso es la determinación del tipo de archivo que se encuentra encriptado. Esto servirá para determinar si efectivamente es procedente realizar el esfuerzo de intentar quebrar la encriptación, esto es, si existe o no la posibilidad de existencia de una evidencia en dicho archivo, en base al objetivo buscado.
Como segunda etapa se requerirá el software / algoritmo en conjunto a la clave de seguridad para acceder a los datos buscados.
Por el momento no hemos encontrado herramientas que permitan llevar a cabo estas tareas.

4.3.4 Búsqueda de Información de procesos en memoria

Durante la ejecución normal del sistema operativo, este almacena información en la memoria física que puede llegar a ser de interés como información relativa a los procesos que se estuvieron

ejecutando antes de llevar a cabo la adquisición. La técnica que nos permite concretar esta tarea consiste en analizar un volcado de memoria RAM y localizar estructuras de procesos para obtener información.
Actualmente existen en el mercado soluciones únicamente para Windows.
Sin embargo estas alternativas no permiten realizar busquedas de contenido, sino solo la visualización de las estructuras, y por otro lado, no relaciona las estructuras de procesos con las páginas del PAGEFILE.

4.4 Carencias en Fase de Extracción – Etapa Extracción Fisica

4.4.1 Semantic Carving

En la actualidad no existe una herramienta, ya sea comercial o libre, que implemente la técnica de Semantic Carving, solamente se encuentra disponible un prototipo experimental llamado S2, desarrollado por Simson Garfinkel para el DFRWS 2006.
El autor afirma en sus trabajos que el Carving Semántico representa una mejora con respecto al tradicional, dado que es un proceso de dos pasos en el cual se genera una secuencia de bloques de ensayo que luego son testeados individualmente por un validador. Aquellos bloques que pasan la validación se almacenan como un archivo completo, los demás se descartan. En el proceso se ejecutan varias pasadas.
S2 es un carver semántico escrito en C++ que utiliza una serie de estrategias para reducir el número de secuencias que se necesitan testear, entre ellas:

a) Mantiene un mapa de la imagen del disco. Los sectores de espacio asignado o que ya han sido procesadas por carving no se consideran para carving futuro.
b) Soporta carvers "conectables" de distintos tipos y validadores "conectables" (cada uno por cada tipo de archivo). El validador consiste en un programa que valida una región de bloques y marca algunos de ellos como posibles candidatos a pertenecer a un archivo.
   Cada validador presenta la siguiente información a S2:
   - Nombre
   - Nombre base y extensión para los archivos que crea el carver
   - Un código para analizar los sectores identificados como comienzo de archivo y fin de archivo.
   - Una función C++ que puede leer una cadena de bytes y devolver : C_OK (si validó), C_ERR (si no validó), or C_EOF (si comenzó a validar pero se quedó sin datos)
   - Una utilidad de linea de comandos de Unix, que puede validar si un archivo es de un tipo particular.
   - El número mínimo y máximo de sectores que puede haber en un archivo de ese tipo

Hay validadores desarrollados para la herramienta para los tipos jpeg, zip, msole, html, text, binary.
Se considera un nicho carente, dado que Semantic Carving es una técnica de carving que no tiene una aplicación que la implemente.

4.4.2 Smart Carving

El caso de Smart Carving es discutido y pueden analizarse dos opciones. La herramienta Revit07 dice aplicar “smart carving” aunque en realidad el algoritmo es un carver lineal que realiza análisis de la estructura interna de los archivos para poder recuperarlos con mayor eficacia [3].

Por otro lado, están las herramientas Adroit Photo Recovery y Adroit Photo Forensics, que realmente implementan el Smart Carving™, ya que su algoritmo basado en grafos. La primera, Adroit Photo Recovery, es un programa de uso "doméstico" orientado mayormente a la recuperación de fotos de dispositivos móviles posiblemente dañados. Si bien puede utilizarse con fines forenses, la interfaz de usuario y los casos de uso estan pensados para otro tipo de usuario. El Adroit Photo Forensics por otro lado está totalmente orientado al uso forense y además cuenta con características que le permiten clasificar y organizar los archivos recuperados en forma automática.
Orientado principalmente a la recuperación de archivos de imágenes o fotografías. Es una herramienta paga y no hay herramientas libres que implementen el mismo algoritmo.
El Smart Carving es un nicho carente desde el punto de vista que lo poco que hay, está solo orientado a recuperación de fotos.

4.5 Carencias en Fase de Extracción – Etapa Análisis de Relaciones

4.5.1 Información sobre ejecución de procesos relacionar información contenida en memoria y pagefile

Las áreas de intercambio o archivos de paginación contienen datos que son intercambiados temporalmente entre la memoria y el disco utilizándose como memoria virtual. Al igual que la memoria física, pueden contener información relacionada a los procesos ejecutados antes de la adquisición.
Al momento no hemos localizado herramientas que automaticen el análisis de volcado de archivos de paginación, de todos modos, tanto para sistemas Windows como Unix se puede utilizar el aplicativo strings para buscar información de interés[7].
Ejemplo de búsqueda de registros de mes Agosto en memoria swap de sistema operativo Linux:

```
strings < 192.168.3.10-hda9.dd | grep 'Aug ' | sort -u > logs_swap.txt
strings < 192.168.3.10-hda9.dd | grep -A3 'Aug ' > logs_swap_A3.txt
```

4.5.2 Información de impresión analizar relación entre cola de impresión, archivos de memoria

La información de memoria nos da una noción de los últimos movimientos que se llevaron a cabo en el sistema. Si obtenemos la información de los últimos procesos que estuvieron corriendo, podremos saber si alguno de ellos es el responsable de una solicitud de impresión. Combinando esto con los datos que se puedan extraer del spooler o cola de impresión sabremos de que forma se llevó a cabo la impresión y sobre que archivo.
Al momento existe solo con una aplicación para Windows, que tampoco brinda toda la información que sería de interés. El último archivo de la cola de impresión en Windows se puede visualizar recuperando los archivos con extensión SPL y SHD de la ubicación c:\windows\system32\spool\printers. El siguiente paso es analizarlos con un visor de archivos de spooling como EMF Spool Viewer[8].
Este punto se considera un nicho carente dado que sería de sumo interés recuperar información de archivos que han sido enviados a impresión.

5. Prototipo de validador de archivos recuperados

Uno de los aspectos carentes que se decide atacar, vinculado a la extracción física de la información, es la validación de los archivos devueltos por las herramientas de recuperación de archivos eliminados. Pare tal fin se propone la creación de un framework que permita realizar esta validación. La necesidad de este prototipo surge del hecho de que las herramientas de carving existentes como resultado de su operación producen una enorme cantidad de archivos que el informático forense deberá analizar uno a uno manualmente para determinar su validez y descartar los archivos recuperados erroneamente. Esto, en ocasiones, se debe a deficiencias propias de la técnica que utilice el programa para realizar la recuperación, o a que una parte de los bloques que componían originalmente al archivo fue modificada.
Esta tarea que tediosa y totamente manual, consume una gran cantidad de tiempo y se vería beneficiada de poder automatizar, a cierto nivel, su realización.
En el año 2007 Garfinkel[12] propuso una nueva técnica de *file carving* que incluía la utilización de validadores de archivos como parte integral del proceso de recuperación. Junto con la técnica de *carving*, Garfinkel propuso también un *framework* de validación de archivos, y es ésta la propuesta la que se tomó como base para la realización del prototipo de herramienta.
El diseño de Garfinkel propone realizar la validación de archivos en forma escalonada, separando la validación del archivo recuperado en cinco etapas distintas. Si un archivo no logra pasar una etapa de validación, se lo descarta como archivo inválido y no se prosigue con las siguientes etapas. Esto se realiza para maximizar la velocidad de procesamiento de los archivos, realizando primero las validaciones más básicas y rápidas de realizar. Este esquema se denomina "validadores rápidos" (*fast validation*) y sus etapas son:

1. Validación por medio de *header* y *footer*.
2. Validación de consistencia de punteros internos del archivo.
3. Validación por medio de descompresión.
4. Validación semántica.
5. Validación humana.

El aporte del prototipo está centrado en dos áreas. En primer lugar, se busca reproducir el *framework* de validación, permitiendo que se utilice como parte integral de algoritmos de carving, o que se pueda utilizar en forma independiente para validar archivos ya recuperados. En este aspecto, una herramienta prototipo recibe un listado de archivos, y los procesa por medio del framework para determinar si los archivos son válidos, emitiendo un archivo de log donde reúne los resultados.
El segundo aporte consiste en la automatización de los niveles cuatro y cinco. En el trabajo original de Garfinkel se considera que los niveles cuatro y cinco de validación representan complicaciones para la automatización. Si bien Garfinkel había trabajado en un *carver* semántico, plantea que el análisis semántico de documentos aún resultaba desafiante.
Una vez replicado el *framework* de validación, se realizará la implementación de dos validadores para documentos MS-OLE y archivos JPG, dos formatos de archivo de los cuales se cuenta con ejemplos de validadores disponibles con código libre. Además, se continuará el trabajo generando un validador semántico para realizar análisis semántico sobre documentos de texto recuperados, con el fin de evaluar su consistencia (nivel cuatro), y un validador que, a través de técnicas de visión artificial, pueda evaluar la calidad de archivos de imagen (o fotografías) recuperados (nivel cinco).
También está contemplado realizar un validador para el formato de archivo SQLite3, a modo de comprobar la utilidad del framework para soportar validadores para nuevos formatos de archivo,

y proporcionar una herramienta capaz de comprobar la integridad de éste tipo de archivos. En éste caso, el validador solamente operará en los tres primeros niveles de validación.

6. Prototipo de "Recuperador de Información de Perfil"

Una de las problemáticas al momento de recuperar información digital, es no contar con suficientes especificaciones sobre el material a localizar, pero si en cambio, disponer del rango de tiempo en el que el usuario operó el sistema.
En el perfil del usuario del sistema se encuentra contenida mucha información útil para seguir un rastro que guíe hacia el dato objetivo.
La extracción de información de usuario requiere conocimiento de cada entorno (o sistema operativo) y es una tarea que consiste en recuperar manualmente la información buscada en distintas ubicaciones. Al momento no hemos localizado una herramienta que automatize esta extracción. Sin embargo existe información disponible acerca de la localización y comandos necesarios.
En sistemas Windows la búsqueda se realiza en los registros y sistema de archivos, particularmente en el archivo NTUSER.DAT ubicado dentro de la carpeta de usuario. Para adquirirlo, ya que no se permite mientras se encuentra el sistema corriendo, las alternativas pueden ser adquirir una imagen de disco y luego analizarlo o utilizar herramientas de parsing como Windows Shellbag Parser (sbag) [9], Erunt[10] o RegRipper[11]. También existe la posibilidad de buscar por cadenas.
Actualmente existen en el mercado varias herramientas que pueden proveer datos relacionados con la información de perfil de usuario. Sin embargo, para llevar a cabo una búsqueda a través de las acciones realizadas por un usuario durante un lapso determinado de tiempo, se deben combinar los resultados de varias aplicaciones, lidiar con los distintos formatos de salida y manualmente ordenar cronológicamente los eventos.
Para facilitar la tarea, se propone el desarrollo de un prototipo de software para recuperar información del perfil. La meta de esta aplicación es recolectar información del sistema, facilitar la visualización mostrando los eventos en una linea de tiempo y facilitar el acceso mediante búsquedas avanzadas.
El prototipo contará con módulos que permitan incorporar en un futuro distintos sistemas operativos y acceder a distintos tipos de datos.
La propuesta inicial contempla el sistema operativo Windows e incluye los siguientes módulos:

- Dispositivos conectados y drivers instalados.
- Aplicaciones instaladas y abiertas.
- Documentos abiertos, modificados y eliminados.
- Solicitudes de impresión.
- Historial de navegación web.
- Información de la máquina.
- Información de correo.
- Puntos de restauración e historial de uso de la funcionalidad.

7. Situación actual y conclusiones

El proceso propuesto presentado se validó para las plataformas de escritorio Linux Ubuntu, Microsoft Windows y Mac OSX, y en Android para dispositivos móviles.
A partir de esta validación, y de la búsqueda de técnicas propuestas por investigadores y herramientas específicas es que se detectaron áreas carentes.
Con este mapa de la realidad actual es que se seleccionaron las áreas en las que se está trabajando.
Debido a la difusión de este trabajo en el Primer CIIDDI – Congreso Iberoaméricano de Investigadores y Docentes de Derecho e Informática, surgió una vinculación con la Universidad UniAndes de Ecuador, con quienes nos encontramos trabajando en un proyecto conjunto cuyo objetivo es la adaptación y validación de PURI específicamente en Smartphones.
A raíz de esta investigación también, los alumnos Julián Waimann y Bruno Constanzo, ayudantes de investigación alumnos de este proyecto, están trabajando en el desarrollo de un Carver Inteligente [14] en el marco del Proyecto Final de su carrera de Ingeniería Informática de la Universidad FASTA, asi como el prototipo "Recuperación de Información de Perfil", que es abordado para su implementación como herramienta de software en el marco de un Proyecto Final.
Entendemos que los aspectos carentes detectados constituyen un conjunto de necesidades con potencialidad de ser resueltas desde la academia, en beneficios de la sociedad en general, y de las ciencias forenses en particular.
Las ciencias forenses deben cumplir con tres principios básicos: evitar la contaminación, actuar metódicamente y controlar la cadena de evidencia [31]. Justamente el método es el que permite garantizar el trabajo realizado, su confrontación en un juicio oral de ser necesario, y la trazabilidad.
Por las características propias de las tecnologías de la información y la comunicación, su gran dinamismo y diversidad, es necesario contar con algún proceso que sea lo suficientemente amplio, como para adaptarse a cualquier tecnología, y a su vez, tuviera guías concretas de implementación en plataformas específicas con herramientas actuales y a disposición.
Entendemos que esta primer propuesta de un proceso unificado de recuperación de la información, su difusión y uso, pueden ser el puntapié para que este proceso siga madurando y fortaleciéndose.

Bibliografía

[1] http://digitalforensicssolutions.com/papers/android-memory-analysis-DI.pdf accedido el 12 de Septiembre de 2012
[2] http://digitalforensicssolutions.com/Android_Mind_Reading.pdf accedido el 1 de Septiembre de 2012
[3] http://dfrws.org/2007/proceedings/p2-garfinkel.pdf accedido el 1 de Septiembre de 2012
[4] http://www.dfrws.org/2005/challenge/memparser.shtml accedido el 1 de Septiembre de 2012
[5] http://www.mandiant.com/resources/download/memoryze/ accedido el 1 de Septiembre de 2012
[6] http://www.mandiant.com/resources/download/audit-viewer accedido el 1 de Septiembre de 2012
[7] Román Medina y Heigl Hernández, 2003, Reto de Análisis Forense – RedIRIS accedido el 1 de Septiembre de 2012
[8] http://www.codeproject.com/Articles/10586/EMF-Printer-Spool-File-Viewer accedido el 1 de Septiembre de 2012
[9] http://www.tzworks.net/prototype_page.php?proto_id=14 accedido el 1 de Septiembre de 2012
[10] http://www.larshederer.homepage.t-online.de/erunt/ accedido el 1 de Septiembre de 2012
[11] http://regripper.wordpress.com/regripper/ accedido el 1 de Septiembre de 2012

[12] "Carving contiguos and fragmented files with fast object validation", Garfinkel, S., DFRWS 2007, 2007 accedido el 1 de Septiembre de 2012

[13] "La recuperación de la Información y la Informática Forense – Una propuesta de proceso Unificado" Di Iorio et al. Congreso Argentino de Ingeniería, Julio 2012. www.cadi.org.ar accedido el 1 de Septiembre de 2012

[14] "El Estado Actual de las Técnicas de File Carving y la Necesidad de Nuevas Tecnologías que Implementen Carving Inteligente" Constanzo et al. Congreso Argentino de Ingeniería, Julio 2012. www.cadi.org.ar accedido el 1 de Septiembre de 2012

[15] Forensic Examination of digital Evidence: A Guide for law enforcement, NIJ Report, US Department of Justice, Office of Justice Programs, disponible en http :// www. ojp. usdoj. gov / nij

[16] Survey of Disk Image Storage Formats Version 1.0, Common Digital Evidence Storage Format Working Group, Digital Forensic Research WorkShop, 2006. disponible en www. dfrws. org / CDESF / survey - dfrws - cdesf - diskimg -01. pdf

[17] Mesfer Al-Hajri et al, "An overview of mobile embedded memory and forensics methodology" , disponible en http://ro.ecu.edu.au/ecuworks/1223/ accedido el 10 de Abril de 2012

[18] Case Andrew at al, "FACE: Automated digital evidence discovery and correlation", Science Direct, 2008, disponible en http://www.sciencedirect.com/science/article/pii/S1742287608000340, accedido el 12 de Marzo de 2012

[19] Ahmed Ibrahim , "Steganalysis in Computer Forensics", disponible en http://ro.ecu.edu.au/adf/10/, accedido el 23 de Marzo de 2012

[20] http://www.mandiant.com/resources/download/audit-viewer accedido el 10 de Junio de 2012

[21] Volatility and RegRipperUser Manual, Mark Morgan: Mark.Morgan@iarc.nv.gov , accedido el 3 de Junio de 2012

[22] http://www.codeproject.com/Articles/10586/EMF-Printer-Spool-File-Viewer accedido el 17 de Mayo de 2012

[23] ACPO (Association of Chief Police Officers) – England, Wales and North Ireland. Good Practice Guide for Computer-Based Electronic Evidence. Oficial release version. Disponible en www.acpo.police.uk , accedido el 3 de Abril de 2013

[24] A guide to basic computer forensics – TechNet Magazine, Marzo de 2008. Disponible en www.technet.microsoft.com/en-us/magazine/2007.12.forensics.aspx , accedido el 11 de Abril de 2013

[25] Law Enforcement Investigations - Active Army, Army National Guard, and US Army Reserve. FM 3-19.13. Disponible en www.armystudyguide.com, accedido el 4 de Abril de 2013

[26] María Daniela Álvarez Galarza, "METODOLOGÍAS, ESTRATEGIAS Y HERRAMIENTAS DE LA INFORMÁTICA FORENSE APLICABLES PARA LA DIRECCIÓN NACIONAL DE COMUNICACIÓN Y CRIMINALÍSTICA DE LA POLICÍA NACIONAL". Ecuador. Disponible en www.dspace.ups.edu.ec/bitstream/123456789/546/5/CAPITULO4.pdf, accedido el 4 de Julio de 2011

[27] BREZINSKI, D. y KILLALEA, T. (2002) RFC 3227: Guidelines for Evidence Collection and Archiving. Network Working Group. February. Disponible: http://www.normes-internet.com, accedido el 4 de Julio de 2011

[28] IOCE, Guidelines for the best practices in the forensic examination of digital technology, 2002. Disponible: http://www.ioce.org , accedido el 5 de Julio de 2011

[29] INFORMATION SECURITY AND FORENSICS. Computer forensics. Part2: Best Practices, 2009 Disponible: http://www.isfs.org.hk/publications/ComputerForensics/ComputerForensics_part2.pdf , accedido el 5 de Julio de 2011

[30] Guía Para El Manejo De Evidencia En IT - Estándares de Australia. APEC Telecommunications and Information Working Group, disponible en http://unpan1.un.org/intradoc/groups/public/documents/APCITY/UNPAN016411.pdf , accedido el 6 de Julio de 2011

[31]Cano,Jeimy "Computacíon Forense Descubriendo los Rastros Informáticos." Ed.Alfaomega Grupo Editor.Mexico.2009