ISSN 2238-9121

CONSIDERACIONES PROCEDIMENTALES PARA LA RECOLECCIÓN Y ADQUISICIÓN DE ARREGLOS DE DISCOS

PROCEDIMENTAL CONSIDERATIONS FOR THE COLLECTION AND ACQUISITION OF DISK ARRAYS

Ana Haydée Di Iorio [1]
Bruno Constanzo [2]
Juan Ignacio Iturriaga [3]

RESUMEN

La proliferación de entornos distribuidos en la informática ha generado un cambio de paradigma que influye prácticamente en todas las actividades asociadas, incluso la informática forense.
El Grupo de Investigación de Sistemas Operativos e Informática Forense, de la Facultad de Ingeniería de la Universidad FASTA, ha desarrollado un Proceso Unificado de Recuperación de Información, PURI, que sirve de guía, tanto a informáticos forenses como operadores judiciales, en los pasos a seguir para recuperar la información almacenada digitalmente en un equipo de computación.
Se presenta en éste trabajo una propuesta de los pasos a seguir por los informáticos forenses en el caso particular de encontrarse con un arreglo de discos RAID, del cual se debe realizar la recolección o adquisición.

Palavras-chave: informática forense; PURI; RAID; recuperación de información

ABSTRACT

The proliferation of distributed environments in information technology has generated a paradigm shift that affects practically in every associated activities, even digital forensics. The Operating Systems and Digital Forensics Research Group of Universidad FASTA has developed a Unified Process for Information Recovery (PURI) which serves as guide, both for digital forensics experts and judicial employees, in the steps to follow to recover the information that is digitally stored in a computer.
This work proposes the steps that a digital forensics expert must follow in the special case of finding a RAID array, which must be collected or acquired.

Key-words: digital forensics; PURI; RAID; information recovery

INTRODUCCIÓN

[1] Ingeniera en Informática, Docente e Investigadora en Universidad FASTA, diana@ufasta.edu.ar
[2] Ingeniero en Informática, Investigador en Universidad FASTA, bconstanzo@ufasta.edu.ar
[3] Ingeniero en Informática, Docente e Investigador en Universidad FASTA, Juan@ufasta.edu.ar

ISSN 2238-9121

Los protocolos y guías de recomendaciones procedimentales para peritos informáticos fueron, y siguen siendo, una necesidad tanto para las autoridades, que necesitan una herramienta para evaluar la calidad de las pericias, como para los mismos peritos que necesitan justificar su trabajo. Para responder a esa necesidad el Grupo de Investigación en Sistemas Operativos e Informática Forense de la Universidad FASTA desde hace años trabaja en un Proceso Unificado de Recuperación de Información (PURI)[1], el cual es evaluado por peritos, investigadores y profesionales afines a la informática forense, que brinda bases generales para la actuación y puede ser expandido y mejorado.

Durante la investigación fueron detectados aspectos de la informática forense con carencias de herramientas y técnicas, que se convino denominar “nichos carentes”. Desde la Universidad FASTA se fomenta que los alumnos de proyecto final de Ingeniería en Informática realicen trabajos relacionados con éstos aspectos de la informática forense[2], lo que permite mejorar PURI, brindar respuestas a cuestiones propias de la temática y formar profesionales altamente capacitados.

Además, desde el año 2014, el Grupo de Investigación original participa de el Laboratorio de Investigación y Desarrollo de Tecnología en Informática Forense (InFoLab), un ente donde participan la Universidad FASTA, la Municipalidad de General Pueyrredón y el Ministerio Público Fiscal de la Provincia de Buenos Aires. Esta colaboración permite que PURI sea puesto a prueba y evaluado en forma más intensiva, fomentando los ajustes y modificaciones necesarios para llevarlo y aplicarlo en entornos reales. Este proyecto del InFoLab lleva el nombre de Protocolo de Actuación en Informática Forense PURI/PAIF.

Una cuestión investigada por el Grupo de Investigación son los Sistemas Distribuidos y las diferencias que presentan con respecto a los sistemas “clásicos” sobre los que se trabajó en un primer momento. En un proyecto de actualización de PURI, se estudiaron las cuestiones relacionadas con sistemas distribuidos y servidores de median y gran tamaño, sus particularidades y cómo incorporarlas al Modelo PURI.

El uso de arreglos RAID, en sistemas de mediano a gran tamaño, presenta varios desafíos a los informáticos forenses. El problema más obvio que surge es no contar con la capacidad de almacenamiento como para realizar una adquisición en vivo de un sistema, que con los avances en tecnologías de almacenamiento se ha convertido en una cuestión trivial. Otra cuestión, no tan obvia pero mucho más dañina, es la de la forma en que se realizan la recolección y adquisición de un arreglo RAID. Cuando no se siguen

procedimientos correctos, el resultado podría ser una pila de discos (o imagenes de discos) de valor nulo para la investigación ya que no se puede acceder a los contenidos del disco virtual en forma coherente.

En este trabajo se expone el modelo PURI, la cuestión de la recuperación de información y las características propias de la tecnología RAID. Luego se evalúa cómo éstas características afectan la tarea pericial. Finalmente se propone un procedimiento general, que para incorporar a PURI en el caso específico de llevar a cabo la correcta recolección y adquisición de arreglos de discos RAID.

1 MARCO TEÓRICO

1.1 Modelo PURI

El objeto de estudio de la informática forense es la evidencia digital, información o datos que pueden ser utilizados como evidencia, físicamente almacenados, campos magnéticos, pulsos electrónicos, u otro medio tecnológico, que pueden ser recolectados y analizados con herramientas y técnicas especiales, y la ciencia de extraer información significativa de la evidencia digital se lo denomina análisis forense.

El Proceso Unificado de Recuperación de Información (PURI) define un modelo[1] a partir de la estructuración y organización de distintas técnicas y herramientas, respetando las buenas prácticas propuestas por los organizamos internacionales, de modo que se convierta en una guía y consulta para profesionales tanto para la obtención evidencia digital como el análisis forense.

El Modelo PURI está formado una secuencia de etapas compuestas por fases que involucran tareas para garantizar que el proceso sea auditable, repetible, reproducible y justificable.

ISSN 2238-9121

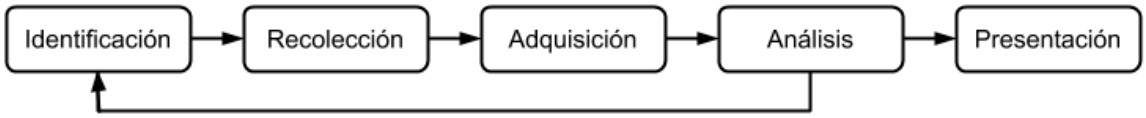

Ilustración 1 – Fases del modelo PURI.

Tal como muestra la ilustración 1, las fases del modelo son:

- **Identificación:** consiste en reconocer y documentar las fuentes de potencial evidencia digital. Se considera la volatilidad de los datos y se prioriza los objetos de mayor interés teniendo en cuenta un principio de suficiencia de la evidencia, pero considerando que muchos objetos informáticos pueden ocultar información, o su función no ser evidente.
- **Recolección:** consiste en reunir los objetos posibles de contener evidencia digital, usualmente tomando un dispositivo físico de una escena.
- **Adquisición:** toda actividad vinculada con la generación de una réplica exacta del contenido digital de interés alojado en el dispositivo original, es decir la obtención de una imagen forense.
- **Preparación:** involucra todos los procedimientos necesarios para generar un entorno de pruebas preciso para llevar a cabo la inspección y el análisis.
- **Análisis:** comprende el trabajo de analizar el contenido adquirido en busca de vestigios de lo que se quiere hallar. El objetivo final de esta fase, en el caso de un proceso judicial o pre-judicial, es encontrar la evidencia digital relacionada con el hecho ocurrido.
- **Presentación:** Consisten en el armado del informe y la presentación de la información a presentar.

1.2 Recuperación de Información

Un proceso de recuperación de información usualmente comienza con el contacto con un dispositivo de almacenamiento sobre el cual se debe trabajar. El contacto inicial con el dispositivo de almacenamiento permite determinar el modo de realizar el proceso de recuperación. Usualmente se toma el dispositivo físico de la escena, este proceso que

ISSN 2238-9121

se denomina **recolección**. Acto seguido, se establece la cadena de custodia del mismo y luego, en un ambiente controlado, se realiza la **adquisición** del mismo. Dependiendo de las necesidades, restricciones y consideraciones particulares del caso, es posible omitir la recolección del dispositivo físico y realizar directamente una adquisición en sitio.

La adquisición del dispositivo es la acción mediante la cual se obtiene una imagen forense del dispositivo de almacenamiento. Se denomina imagen forense a una copia idéntica a nivel de bit de toda la información contenida en un medio de almacenamiento. Para una computadora, no es posible diferenciar una imagen forense del dispositivo original. En casos especiales, cuando la información que se debe recuperar es muy específica y está contenida en un conjunto de archivos, o no se cuenta con la capacidad de almacenamiento para generar una imagen, se realizan copias lógicas, es decir, copia de los archivos como se pueden acceder en la computadora.

La ventaja de trabajar con una imagen forense es que, además de obtenerse una copia de todo lo accesible, se obtiene una copia de todo el contenido remanente en el dispositivo de almacenamiento. Ésto permite, por ejemplo, extraer registros de información, archivos ocultos o eliminados o partes de los mismos que hayan quedado de actividad anterior en el dispositivo de almacenamiento.

Además, el trabajo con una imagen forense permite que se valide que es copia idéntica del dispositivo original, por medio de funciones de digesto matemático. Una vez validada la copia, se puede trabajar sobre la misma sin tener contacto con el original, lo que refuerza el principio forense de preservación de la evidencia.

Las tareas de recuperación de información luego consisten en tareas de búsqueda y análisis sobre la imagen del dispositivo adquirida. Ésta búsqueda se orienta de acuerdo a lo que interesa recuperar. Usualmente se utilizan herramientas que permiten hacer un análisis automatizado en busca de formatos de archivo, patrones, texto o puntos de interés.

1.3 Arreglos RAID

RAID (*Redundant Array of Independent Disks*) es una tecnología que permite combinar múltiples dispositivos de almacenamiento y los presenta al host como un solo volumen virtual[3][4]. RAID establece una sinergia entre los dispositivos que conforman el arreglo, brindando las siguientes ventajas:

ISSN 2238-9121

- **Rendimiento:** el funcionamiento en conjunto de los múltiples dispositivos abre la posibilidad de realizar operaciones de lectura y escritura en forma paralela, que no serían posibles si se tratara de un único dispositivo.
- **Tolerancia a fallos:** RAID, en algunos de sus modos de operación, permite tener redundancia en los datos. En estos casos la falla de un disco no compromete la información, pero el rendimiento se ve degradado hasta reemplazar el dispositivo y restaurar el arreglo.
- **Capacidad:** como consecuencia de combinar los dispositivos, se obtiene un dispositivo virtual de igual o mayor tamaño que cada uno de los dispositivos individuales.
- **Economía:** éstas características se obtienen de combinar discos reales con un costo relativamente bajo. Si se buscara un dispositivo único real con las mismas características que un arreglo de discos, en caso de existir, probablemente sería mucho más costoso.

Dependiendo la configuración de RAID que se utilice, se refuerzan en mayor o menor medida estos aspectos. Como se verá más adelante, hay configuraciones que sacrifican la capacidad de almacenamiento por redundancia, o al revés, sacrifican la redundancia por mayor capacidad de almacenamiento. También hay configuraciones que establecen un balance entre capacidad y redundancia, a un costo de rendimiento.

Aunque históricamente RAID presentaba una forma de lograr volúmenes virtuales de gran tamaño, y aún hoy lo permite, el principal atractivo se encuentra en las mejoras de performance y tolerancia a fallos que permite con respecto a un único dispositivo físico.

La especificación de RAID define siete niveles de RAID. A continuación se detallarán los cuatro niveles más utilizados:

- **RAID 0:** los discos del arreglo RAID se fraccionan en tramas (chunks) de acuerdo a bandas, también llamadas stripes, lo que permite utilizar los discos en paralelo para realizar operaciones de lectura/escritura en cada disco. Este nivel de RAID brinda excelente rendimiento, pero no cuenta con redundancia de datos.
- **RAID 1:** también llamado “arreglo espejo”, define un disco de datos y uno o más discos espejo. Los discos espejo son copias idénticas del disco de datos y pueden reemplazarlo automáticamente si ocurre una falla. Este nivel de RAID permite la

ISSN 2238-9121

lectura en paralelo de las partes espejadas, pero la ventaja principal de este modo es que realiza un "respaldo automático" del disco de datos que permite proteger la información de la rotura física de todos menos uno de los discos que componen el arreglo.

- **RAID 5:** los discos del arreglo RAID se fraccionan en bandas, lo que permite realizar lectura y escritura en paralelo en los múltiples discos. Para cada banda se calcula una función de paridad, y los resultados se distribuyen entre todos los discos para no sobrecargar uno en particular. Permite que el arreglo se recupere de la falla de un disco y reconstruirlo al reemplazarlo.
- **RAID 6:** puede considerarse como una variante de RAID 5 en donde se calculan dos funciones de paridad distintas, lo que permite recuperar el arreglo con la falla de dos discos. Nuevamente, es un balance entre capacidad, rendimiento y redundancia, más resistente a fallas que RAID 5. En particular, este nivel de RAID permite recuperarse del caso donde se produce la rotura de un segundo disco durante la reconstrucción.

RAID puede estar implementado tanto por software como por hardware. Cuando se implementa por software, es el sistema operativo el que se encarga de la distribución de la información sobre el arreglo de discos, y mantener la configuración del arreglo para que siga funcionando en forma coherente. Cuando RAID se implementa en hardware a través de una placa controladora, es ésta la encargada de la distribución de la información en los discos y la configuración del arreglo. Usualmente el hardware RAID es costoso y su uso se justifica en servidores y equipos de alto rendimiento, y las implementaciones por software cuando no se tiene acceso al hardware específico o en ambientes hogareños.

2 EL PROBLEMA

Hay dos cuestiones de suma importancia para la integridad y coherencia de un arreglo RAID, que son el orden de los discos y la configuración del arreglo. Sin esta información, cualquier recolección o adquisición realizada es prácticamente inútil, y se deberá recurrir a técnicas de reconstrucción para lograr acceder a la información del

arreglo de discos. Esto puede tomarse como un principio guía para las tareas de recolección y adquisición:

> *"Toda la información relevante que no se tome en la etapa de recolección y adquisición, tendrá que ser deducida en las etapas de preparación y análisis"*

El resultado de una mala recolección es que se contará en el laboratorio con una pila de discos, de los cuales se podrán realizar imagenes forenses por separado, pero no se podrá acceder directamente al disco virtual que conforman esos discos. El resultado de una mala adquisición es un conjunto de imagenes forenses que no conforman el disco virtual correspondiente. En ambos casos, deberá intentar reconstruirse el arreglo antes de proceder con las tareas clásicas de recuperación de información. Ésta tarea adicional de reconstrucción puede realizarse con herramientas automáticas, ser guiada en forma manual por el informático forense o resultar demasiado complicada para poder realizarse, en cuyo caso tanto la recolección como la adquisición realizadas no tendrán valor ni podrá llevarse a cabo la tarea de recuperación de la información.

3 PROCEDIMIENTOS RECOMENDADOS

En cuanto a los procedimientos, deben considerarse si el equipo se encuentra prendido o apagado, y si debe hacerse recolección o adquisición. Para todos los casos se deben respetar las recomendaciones básicas a seguir durante un procedimiento, ya sea judicial, privado, un allanamiento, u otro tipo. Es decir, se debe garantizar que el estado del equipo no va a ser alterado a través de una red interna o externa, asegurar que no hay procesos ejecutándose en el equipo que puedan alterar o eliminar la información de interés. En general, se asume que la integridad y seguridad de las personas, los equipos y la información está garantizada y se puede trabajar en forma tranquila, ordenada y tomando las precauciones necesarias.

A continuación se analizará el caso de la recolección y luego la adquisición, ya que las tareas se realizan en ese orden en una experiencia real, comenzando por las situaciones más favorables hasta las de mayor complejidad junto a complicaciones que puedan surgir.

ISSN 2238-9121

3.1 Recolección

Cuando se ha identificado para la recolección un equipo que cuenta con un arreglo de discos RAID se recomienda:

1. En lo posible, recolectar el equipo completo para asegurarse que no faltará ninguno de los discos, la controladora RAID y/o el sistema operativo con la configuración adecuada para acceder al arreglo de discos.
    - Debe quedar en claro para los responsables del procedimiento que lo que se busca es garantizar que en las etapas de Preparación y Análisis se podrá acceder a la información contenida en lo recolectado. Si es necesario trasladar el equipo completo, deben buscarse los medios para realizarlo.
2. Si no es posible recolectar el equipo completo, se recomienda:
    - Tomar fotografías o hacer un esquema de cómo estan conectados los discos al hardware del equipo.
    - Si se trata de un RAID por software, revisar la configuración del equipo para detectar: el sistema operativo base, la herramienta a través de la cual se implementa RAID, las particiones de los discos que componen el arreglo y si está compuesto por archivos funcionando como dispositivos virtuales.
    - Si se trata de un RAID por hardware, revisar la configuración de la controladora y tomar nota de la configuración del arreglo. La controladora también debe recolectarse junto con los discos. Usualmente recolectar la controladora también implica recolectar el motherboard o contar un motherboard similar en el laboratorio.
    - Si el equipo se encuentra apagado, debe asumirse que todos los discos conectados al equipo participan de un arreglo RAID.
    - Deben recolectarse todos los discos que participan de alguna forma en el arreglo RAID.

3.2 Adquisición

Una vez que se tiene el acceso al dispositivo RAID, ya sea en vivo sobre el equipo o luego de realizar recolección del mismo, se debe proceder con la adquisición de la imagen, para lo que se recomienda:

ISSN 2238-9121

1. En lo posible, si el volumen RAID está montado, realizar una adquisición clásica como si se tratara de una un dispositivo único de gran tamaño.
    - Si bien RAID permite combinar varios dispositivos físicos para obtener un dispositivo lógico de mayor tamaño, en la actualidad hay soluciones de almacenamiento que brindan gran capacidad de almacenamiento y pueden almacenar el contenido de un arreglo RAID.
    - Al realizar la imagen del volumen RAID como un único dispositivo, se eliminan los problemas relacionados con la complejidad de RAID para las etapas de Preparación y Análisis.
2. Si el dispositivo no está montado sobre el equipo, debe intentarse montar el disco para accederlo como volumen virtual íntegro.
    - Para ésta tarea es importante contar con toda la información relacionada con la configuración del arreglo, es decir, el orden de los discos en el arreglo, la configuración RAID específica y, si se trataba de un RAID por hardware, la placa controladora.
    - Si se logra montar el volumen, se debe proceder con una adquisición simple del volumen virtual, como en el caso (1).
3. En caso de no poder montar el arreglo RAID como una unidad virtual, deben tomarse imagenes individuales de cada disco y proceder con un proceso de reconstrucción del arreglo RAID.
    - La reconstrucción del arreglo puede realizarse con herramientas automáticas, teniendo en cuenta las ventajas y desventajas para seleccionar la herramienta adecuada.
    - También es posible realizar la reconstrucción en forma manual, o con herramientas guiadas por el experto informático forense. Este proceso suele llevar más tiempo que una herramienta automática, pero permite resolver algunos casos que las herramientas existentes no pueden manejar.

Debe quedar claro que éstas recomendaciones son para facilitar el trabajo del informático forense y evitar las tareas de reconstrucción de RAID, que demandan tiempo y esfuerzo por parte del experto. El no seguir estas recomendaciones no implica una pérdida de la información, pero sí establece demoras, por el trabajo adicional que debe realizarse,

ISSN 2238-9121

y puntos en los cuales el informático forense deberá explicar y justificar su trabajo a una profundidad mayor que si hubiera realizado el trabajo en forma adecuada.

En una situación judicial es mucho más simple explicar que se realizó la adquisición de la imagen completa del disco virtual, como se podía acceder en el equipo prendido, que tener que explicar y justificar un proceso de reconstrucción de RAID, con las dificultades técnicas que implica, y explicando cómo se mantienen las garantías de no contaminación de la información. La reconstrucción es casi siempre posible, pero demanda mucho más trabajo, no sólo técnico sino también procedimental y de justificación de lo realizado.

CONCLUSIÓN

El objetivo del informático forense es, básicamente, obtener evidencia digital, y es posible que se presente el caso de tener que buscar la misma en un entorno de almacenamiento distribuido como lo es RAID. En el presente trabajo se propuso una ampliación y especialización del modelo PURI sobre las fases de recolección y adquisición en el caso específico de encontrarse con un equipo RAID.

Lo más importante a tener en cuenta, es que lo que no se tome en las fases de recolección y adquisición, deberá ser deducido en la fase de adquisición con todo lo que ello implica: mayor tiempo de análisis y menores posibilidades de éxito.

La situación más favorable, en este caso, sería poder recolectar el equipo completo, en su defecto adquirir el volumen RAID íntegro en una única imagen y, de no ser posible, adquirir imágenes individuales de los dispositivos que componen el RAID con la información adicional necesaria para la reconstrucción del volumen a partir de las mismas.

Si las imágenes de los dispositivos contienen información intacta, y no se dispone de información adicional, aun así es posible la reconstrucción por medios automáticos, ya que los mismos RAIDs guardan dicha información. Si por algún motivo esa información está dañada o no es reconocible por las herramientas automáticas se debe realizar una reconstrucción manual, la cual requiere mucho más tiempo de análisis y depende tanto del contenido y cantidad de información del volumen como en la experiencia en el tema del analista; existiendo la posibilidad que el volumen RAID no pueda ser reconstruido para extraer la evidencia digital.

ISSN 2238-9121

REFERENCIAS

[1] DI IORIO, Ana H.; SANSEVERO, Rita E.; CASTELLOTE, Martín A.; PODESTÁ, Ariel; GRECO, Fernando; CONSTANZO, Bruno; WAIMANN, Julián. **La recuperación de la información y la informática forense: Una propuesta de proceso unificado.** Congreso Argentino de Ingeniería CADI 2012.

[2] DI IORIO, Ana H.; SANSEVERO, Rita E.; CASTELLOTE, Martín A.; PODESTÁ, Ariel; GRECO, Fernando; CONSTANZO, Bruno; WAIMANN, Julián. **Determinación de aspectos carentes en un Proceso Unificado de Recuperación de Información digital.** Jornadas Argentinas de Informática Forense JAIF 2013.

[3] TANENBAUM, Andrew S. **Sistemas Operativos Modernos.** Capítulo 4. Prentice Hall Hispanoamericana, 1993.

[4] TANENBAUM, Andrew S. **Structured Computer Organization.** Pgs. 89 a 93. 5ta Edición. Pearson Prentice Hall, 2006.

[5] FAY-WOLFE, Victor. **RAID Rebuilding.** CSC-486 Network Forensics, University of Rhose Island. Enero 2008. Disponible en http://media.uri.edu/cs/csc486_wmv/RaidRebuilding_TOC.pdf.