1. INTRODUCCIÓN

En el presente documento se expone un método que facilita la visualización de todo tipo de archivos, completos o incompletos, partiendo de la identificación del formato de los mismos.

Los delitos informáticos son una problemática creciente, y su complejidad avanza a la par de la tecnología. Para la solución de estos casos, en computación forense, los investigadores necesitan poder acceder a toda la información almacenada en una computadora sujeta de investigación pues ésta puede constituir una prueba de culpabilidad o inocencia cuando se juzga a una persona. Además de los delitos informáticos, existen otros problemas que no están ligados al ámbito jurídico pero que requieren una intervención con técnicas avanzadas que permitan recuperar archivos incompletos o no identificados.

Es importante, entonces, contar con un método flexible, que permita a los usuarios visualizar cualquier tipo de archivo y que facilite el análisis completo y exhaustivo de la evidencia digital. Este proyecto beneficia, entonces, a los investigadores en computación forense, pues provee un método de acceso a la información relevante oculta en los archivos que no pueden ser visualizados.

Inicialmente se enfrentó la problemática de desarrollar una herramienta de visualización universal de datos, a partir de un lenguaje definido en XML que permitiera crear plantillas para la estructuración y posterior visualización de los archivos. En la etapa de investigación se encontró una herramienta muy similar, aunque con la funcionalidad del uso de plantillas en fase beta. Debido a esto, el equipo de investigación decidió redefinir el alcance del proyecto, orientándolo más

a la definición de un método aplicable a problemas de visualización de archivos incompletos o de formato desconocido.

Una vez redefinido el alcance y los objetivos del proyecto, se empezó a construir el marco referencial partiendo de los conceptos básicos de computación forense y evidencia digital. Se exploró el funcionamiento de los editores hexadecimales, los cuales fueron una herramienta primordial para conocer la estructura de los archivos, y se investigó el estado del arte de la identificación y visualización de los mismos.

En el desarrollo del proyecto se identificaron tres etapas macro para el método de visualización universal de datos. Para apoyarlo, se implementaron tres programas. El primero de ellos estuvo orientado a ayudar a sugerir el formato del archivo o fragmento de archivo, a partir del cálculo de la entropía. Los dos restantes se centran en la reconstrucción del archivo, y son: un formateador BMP y un formateador WAV.

2. FORMULACIÓN DEL PROBLEMA

Cada tipo de archivo informático tiene su propio formato. Para su identificación y visualización, las herramientas de software deben conocer esa estructura interna.

En la Computación Forense[1], el investigador, al recobrar evidencia digital, puede encontrarse con fragmentos de archivos que no pueden ser visualizados por estar incompletos o porque se desconoce su tipo.

En ese caso, disponer de un método que ayude a sugerir el formato del archivo, facilitaría la recuperación o, en su defecto, el análisis del tipo de información.

Los mecanismos de identificación utilizados actualmente no son ciento por ciento confiables, pues para algunos de ellos es necesario que el archivo tenga sus encabezados completos, o porque la información necesaria para hacerlo puede ser fácilmente modificada por el usuario, como en el caso de la extensión del archivo.

Partiendo de la premisa de que tanto los avances tecnológicos como los volúmenes de información son cada vez mayores y que, por tanto, en una investigación forense se debe analizar una enorme cantidad de archivos de gran variedad[2], se hace necesario encontrar un método óptimo, adaptable a diversos escenarios, que sugiera el formato del archivo o fragmento, y en caso de estar incompleto, provea una estrategia acertada para su reconstrucción.

---

1 "Ciencia de adquirir, preservar, obtener y presentar datos que han sido procesados electrónicamente y guardados en un medio computacional".
http://www.fbi.gov/hq/lab/fsc/backissu/oct2000/computer.htm

2 VASSIL, Roussev; GOLDEN G., Richard III. "Digital Forensics Tools: The Next Generation". 2006. http://www.cs.uno.edu/~golden/Stuff/ideagroup2006.pdf

Al desarrollar un método apoyado en herramientas de software libre se facilitará la disminución en los costos de investigación relacionados con compras de licencias, y reducirá el número de horas invertidas en el proceso de identificación y recuperación de archivos, haciendo más efectiva la labor del investigador.

3. OBJETIVOS

3.1 OBJETIVO GENERAL

Proponer un método que facilite la visualización total o parcial de archivos incompletos o con formato desconocido, a partir de la definición de estructuras, estrategias, y del uso de herramientas de software libre para su análisis forense e identificación.

3.2 OBJETIVOS ESPECÍFICOS

- Definir los pasos a seguir en el análisis de archivos incompletos.
- Identificar herramientas libres que puedan apoyar el desarrollo de los pasos definidos.
- Determinar una estrategia que permita sugerir efectivamente el formato de un archivo o fragmento de archivo.
- Construir un conjunto de pruebas que permitan conocer el funcionamiento del método de visualización universal de archivos.