Restauración de datos en dispositivos de estado sólido dañados por aplastamiento y caida, mediante técnicas File Carving

Restoring data in solid state devices damaged by crushing and falling, using File Carving technique

Geovanni Ninahualpa[1,4], Carolina Pérez[1], Sang Guun Yoo[1], Teresa Guarda[1,2,3], Javier Díaz[4], Dario Piccirilli[4]
[1]Departamento de Ciencias de la Computación, Universidad de las Fuerzas Armadas ESPE, Sangolquí, Ecuador
[2] Universidad Estatal Peninsula de Santa Elena, La Libertad, Ecuador
[3]Algoritmi Centre, Minho University, Guimarães, Portugal
[4] Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina
gninahualpa@espe.edu.ec, ccperez1@espe.edu.ec, yysang@espe.edu.ec, tguarda@gmail.com, jdiaz@unlp.edu.ar, dariopiccirilli@gmail.com

Resumen — **El presente documento propone analizar el aporte de las técnicas File Carving en la recuperación de datos en dispositivos de estado sólido (Solid State Drive – SSD) que han sido afectados por escenarios de aplastamiento y caída con diferentes intensidades de fuerza. Para el efecto se han utilizado las técnicas Semantic Carving y Fragment Recovery Carving. Obteniendo resultados categorizados en las métricas: cantidad de archivos recuperados, cantidad de archivos válidos recuperados, cantidad de archivos parcialmente recuperados, cantidad de falsos positivos. Finalmente se comparan los resultados obtenidos aplicando el modelo matemático de regresión lineal.**

Palabras Clave – Dispositivos de estado solido, semantic carving, fragement recovery carving, métricas, regresión lineal.

Abstract — **This document proposes to analyze the contribution of File Carving techniques in the recovery of data in solid state devices (Solid State Drive - SSD) that have been affected by crush and fall scenarios with different strengths. For this purpose, Semantic Carving and Fragment Recovery Carving techniques have been used. Obtaining results categorized in the metrics: number of files recovered, number of valid files recovered, number of partially recovered files, number of false positives. Finally, the results obtained are compared by applying the mathematical model of linear regression.**

Keywords - Solid state devices, semantic carving, fragment recovery carving, metrics, linear regression.

I. INTRODUCCIÓN

La gran importancia que ha alcanzado la información digital en la sociedad en función de su computabilidad [18], virtualidad y capacidad sumado a la masificación de los dispositivos de almacenamiento de estado sólido SSD (Solid State Drive), que han generalización la información digital en la sociedad. En tal sentido, se encuentra presente en entornos; educativos, empresariales, productivos, entre otros.

Actualmente los SSD tienen gran acogida por sus características de velocidad en la transferencia de información, mejor resistencia a factores ambientales, debido a que no cuenta con elementos mecánicos [15], sin embargo todo dispositivo es vulnerable a sufrir algún tipo de daño y cuya información podría verse afectada o perderse de forma definitiva.

En la conferencia [13], se presenta un alto porcentaje de daños en dispositivos de almacenamiento por factores físicos como caídas y golpes, siendo este un valor significativo para poner énfasis en la recuperación de datos y técnicas que permiten recuperar información de forma más efectiva.

II. LOS SSD Y LA RECUPERACIÓN DE DATOS

La recuperación de datos es el proceso que trata de minimizar el impacto de los factores causantes de pérdida de la información según [13], en donde se evidencia que los golpes y caídas representan el mayor porcentaje entre las causas de pérdida.

En la actualidad el creciente uso de los SSD como dispositivos de almacenamiento de información debido a la velocidad de operación, han hecho que sean vulnerables a daños físicos y lógicos que podrían provocar pérdida en la información contenida en ellos [13].

A. Dispositivos de Estado Sólido

Según IBM (2017) [12], los SSD son dispositivos de almacenamiento que utilizan memoria de estado sólido no volátil (memoria flash), al no tener partes electromecánicas reducen considerablemente el tiempo de acceso a los datos almacenados, en tal sentido la ventaja más latente es que las operaciones de lectura se pueden realizar más rápido que las operaciones de escritura.

De acuerdo con Toshiba (2018) [22], los dispositivos SSD están basados en unidades basadas en memorias flash que no disponen de partes móviles, es decir, no poseen piezas que se

están moviendo físicamente como un disco que gira junto a un cabezal que busca sectores, permitiendo que la nueva tecnología sea de menor tamaño físico y presente una serie de otras ventajas que la colocan por sobre el disco duro tradicional.

Para SNIA (2018), las características de los dispositivos SSD son[20]:

1) Interfaz del Conductor – Driver Interface

Medio (controlador software) por el cual el dispositivo es reconocido por el sistema operativo host.

2) Bus Host – Host Bus.

El Bus Host (componente periférico expreso interconectado), es la interfaz física implementada para dispositivos SSS.

3) Interfaces y Conectores del Dispositivo.

Son los estándares de interfaz de transmisión.

4) Interfaz NAND Flash.

Diseño de la configuración de la memoria flash.

5) Factores de Forma.

Definen los formatos y tamaños disponibles, categorizándolos en unidades de estado sólido (SSD), tarjetas de estado sólido (SSC) y módulos de estado sólido (SSM).

6) Seguridad de los Datos.

Capacidad de encriptar los datos que se almacenan en el dispositivo.

7) Testeo SSS.

Métodos de prueba estandarizados para comparar las características de los dispositivos de almacenamiento de estado sólido.

8) Memoria no volátil (NVM).

Operando desde la RAM y utilizando Flash para la copia de seguridad, proporcionando almacenamiento no volátil de alta velocidad que se conecta directamente al bus de memoria del sistema.

B. Técnicas de Recuperación de Datos – File Carving

Según Anandabrata, P., & Nasir, M. (2009), es necesario acudir a las técnicas forenses de recuperación de datos, cuando las técnicas convencionales de recuperación de datos no brindan la solución esperada [6].

File Carving es una técnica forense que recupera archivos basados simplemente en la estructura y el contenido del mismo, sin tomar en cuenta a ningún metadato del sistema de archivos. También se usa con mayor frecuencia para recuperar archivos del espacio no asignado en una unidad [6].

Para Constanzo, B., & Waimann, J. (2012) y Anandabrata, P., & Nasir, M. (2009)se propone la siguiente clasificación de las técnicas File Carvin g [8] [6]:

1) Basadas en las Características del Archivo

- Técnicas Header-Footer File: Técnicas basadas en dos identificadores “Encabezado” y “Pie/Fin”, pues si se conoce el inicio y el final, los bloques útiles entre estos identificadores son considerados como el archivo a restaurar [10].
- Técnicas File Structure and Block Content: Técnicas basadas en la estructura del archivo a restaurar y los bloques útiles contenidos en él [5].
- File Structure Based Carving: Técnica que analiza las estructuras internas del tipo de archivo, además de las que se encuentran en la [8].
- Semantic Carving: Identifica el idioma utilizado en un bloque, y lo relaciona con bloques en el mismo idioma de forma coherente para reconstruir un texto [8].

2) Basadas en la Framentación del Archivo

- Técnicas Fragmentation Issue: Restauran el archivo que se encuentra fragmentado, sin contar con: identificadores o cadenas de secuencia para reconocer al fragmento [10].
- Técnicas Predictive File: Crean nueva metadata que referencia a los bloques de almacenamiento válidos, utilizando un sistema de archivos (Filesystem) virtual como forma de predecir y acceder a los mismos [5].
- Técnicas Fragment Recovery Carving: Describe el método carving reensambla dos o más fragmentos para formar el archivo original [19].
- Statistical Carving: Busca similitudes para reconstruir los archivos [8].

C. Herramientas de Recuperación de Datos – File Carvers

Las técnicas descritas guían el desarrollo de aplicativos que implementen las mismas. En tal sentido se proponen:

1) Herramientas Linux

- TestDisk [7]. Recupera particiones perdidas.
- Scalpel [1]. Implementa algoritmo Semantic Carving.
- Foremost [2]. Implementa algoritmo Fragment recovery Carving.
- GNU Ddrescue [3] [2], Copia los datos de un archivo o dispositivo de bloques a otro.
- SafeCopy [21] [2], Recupera de modo genérico, datos de cualquier tipo de fuente.

2) Herramientas Windows

- PhotoRec [21] [2]. Recuperar archivos de discos duros y tarjetas de memorias de las cámaras digitales.
- Forensic Toolkit (FTK) [9]. Software forense de recuperación de datos.
- EVTXtract [23]. Comandos de Python que intenta recuperar y reconstruir fragmentos de archivos de registro de eventos de Windows a partir de datos binarios en bruto.
- WinHex [4]. Editor hexadecimal que permite ver datos, inclusive los ocultos.

III. RECUPERACIÓN DE DATOS EN SSD

En la recuperación de datos en dispositivos de almacenamiento SSD, se propone nueva metodología de recuperación de información con base en la presentada por [17] para dispositivos electromecánicos y actualizada mediante experimentación para operar en dispositivos SSD.

A. *Propuesta Metodológica de Recuperación de Datos en SSD*

A la propuesta se la ha dividido en fases y a cada una en procesos.

1) Fases de Análisis Físico

En esta fase se realizarán los procesos de: Inspección visual del SSD y la verificación de la alimentación de poder.

2) Fase de Validación de Daños

La fase de validación de daños iniciará con el proceso de verificación del daño físico a continuación y dependiendo del proceso previo se realiza el reconocimiento del SSD en el BIOS.

3) Fase de Recuperación Física Temporal

En esta fase se realizan procesos de: identificación del tipo de daño, propuesta y ejecución de la solución y verificación de resultados.

4) Fase de Obtención de Imagen

Los procesos de obtención de imagen son: Por Hardware o Software, en tal sentido en cada uno de estos se realizarán las tareas de: preparación del dispositivo de almacenamiento de destino, creación y verificación de la imagen.

5) Fase de Análisis Lógico y Recuperación de Datos

Para proceder con la recuperación de datos se realizan: proceso de verificación del tiempo de uso, verificación de procesos de limpieza del SSD, para luego realizar la ejecución de la recuperación por medio de software especializado y finalmente revisión de la integridad de los archivos.

B. *Selección y Preparación de Escenarios*

En la preparación de los escenarios se realizaron encuestas a una muestra de profesionales y estudiantes de un universo de 20200 participantes, sobre los principales eventos que pueden generar pérdida de datos en los dispositivos de almacenamiento. Así 53% de los encuestados señalaron haber tenido daños físicos y el 47 % por daños lógicos, en tal sentido llama mucho la atención que el 50% sean por efectos de caídas, impactos o aplastamientos.

- Escenario Daño por Caídas: Este tipo de escenarios responden a la información recabada en las encuestas así el 69% de los encestados indicaron que el dispositivo se cayó a menos de un metro, 28% entre uno y dos metros y 3 % con caídas superiores a 2 metros.
- Escenario Daño Aplastamiento por Impacto: El aplastamiento por impacto es el escenario con que continua con el 81% de incidencias en las causas que provocan la pérdida de información en los dispositivos de almacenamiento SSD.

Por consiguiente los escenarios descritos serán los que permitirán la realización del presente trabajo.

C. *Desarrollo de la Investigación por Escenarios*

El presente trabajo se realizó sobre dispositivos WD GREEN 120GB sata SSD, en los siguientes escenarios:

1) Escenario Daño por Caídas

Para la experimentación se realizaron las tareas: pesaje de los dispositivos a emplear, determinación de las alturas, cálculo de las energías cinética y potencial del cuerpo, cálculo de la fuerza ejercida.

Se realizan las experimentaciones lanzando los dispositivos SSD de las diferentes alturas y llenando la bitácora de los datos obtenidos, tanto de energías y fuerza.

En función de la metodología propuesta se realizan los procesos de evaluación, análisis, rescate y validación de los datos recuperados en cada dispositivo.

2) Escenario Daño Aplastamiento por Impacto

En el desarrollo de esta experimentación se realizaron las tareas: pesaje de los dispositivos y determinación de la fuerzas a emplear.

Se realizan las experimentaciones ejerciendo las fuerzas sobre los dispositivos con la ayuda de una máquina de compresión simple y llenando la bitácora de los datos obtenidos de las fuerzas empleadas.

En función de la metodología propuesta se realizan los procesos de evaluación, análisis, rescate y validación de los datos recuperados en cada dispositivo.

La fase de validación de daños iniciará con el proceso de verificación del daño físico a continuación y dependiendo del proceso previo se realiza el reconocimiento del SSD en el BIOS.

IV. ANÁLISIS DE RESULTADOS

A continuación se detallan los resultados obtenidos, considerando las técnicas File Carving empleadas (Tab.1, Tab2), los escenarios de experimentación, condiciones de fuerza, porcentajes de recuperación de archivos y agrupación de los archivos (ofimática o multimedia).

TABLE I. RECUPERACIÓN SEMANTIC CARVING - FOREMOST

Condiciones de Recupero	**Daño por Caídas**		**Daño por Aplastamiento**	
	Ofimática	***Multimedia***	***Ofimática***	***Multimedia***
Intervalo I Fuerza (N) Recupero (%)	1 – 30 N 79.43%	1 – 30 N 79.43%	0 – 15 N 86%	0 – 15 N 88%
Intervalo II Fuerza (N) Recupero (%)	31 – 70 N 49.43%	31 – 70 N 49.43%	15.1 – 30 N 57%	15.1 – 30 N 59%
Intervalo III Fuerza (N) Recupero (%)	71 – 110 N 19.53 %	71 – 110 N 19.53 %	30.1 – 45 N 32%	30.1 – 45 N 34%
Intervalo IV Fuerza (N) Recupero (%)	111 - 136.1 0.003%	111 - 136.1 0.003%	45.1 – 65 N 0.002%	45.1 – 65 N 0.01%

TABLE II. RECUPERACIÓN FRAGMENTED - SCALPEL

Condiciones de Recupero	Daño por Caídas		Daño por Aplastamiento	
	Ofimática	*Multimedia*	*Ofimática*	*Multimedia*
Intervalo I Fuerza (N) Recupero (%)	0 – 80 N 47.43%	0 – 80 N 47.43%	0 – 4.05 N 99.3%	0 – 4.05 N 99%
Intervalo II Fuerza (N) Recupero (%)	80.1 – 174 N 0.27%	80.1 – 174 N 0.27%	4.1 – 12 N 0.64%	4.1 – 12 N 0.96%

V. CONCLUSIONES

En el desarrollo de las pruebas y procedimiento de recupero de información, se han presentado inconvenientes relacionados con replicar los factores ambientales en las condiciones más semejantes a la realidad, por lo que se ha debido repetir la experimentación para obtener datos fidedignos.

La técnica Semantic Carving representada en el carver Foremost ha demostrado ser la que mejor resultados presenta en recuperación de información, tanto en para archivos de ofimática y mejor aún para archivos multimedia, en los dos escenarios planteados.

Finalmente la técnica Fragmented Carving representada en el carver Scalpel tiene su mejor índice de recuperación si la afectación por cualquier de los escenarios propuestos es pequeña, ya que al aumentar la fuerza disminuye su fiabilidad.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] *Tecmint*. (07 de 06 de 2013). Obtenido de Tecmint : https://www.tecmint.com/install-scalpel-a-filesystem-recovery-tool-to-recover-deleted-filesfolders-in-linux/

[2] *Kali Tools.* . (18 de 02 de 2014). Obtenido de Kali Tools. : http://tools.kali.org/forensics/foremost

[3] *GNU*. (21 de 02 de 2017). Obtenido de GNU: https://www.gnu.org/software/ddrescue/

[4] *Xways*. (27 de 11 de 2017). Obtenido de Xways: https://www.x-ways.net/winhex/

[5] Alherbawi, N., Shukur, Z., & Sulaiman, R. (2016). A Survey on Data Carving in Digital Forensic. *Asian Journal of Information Technology*, 15(24), 5137-5144.

[6] Anandabrata, P., & Nasir, M. (2009). The Evolution of File Carving. *IEEE Signal Processing Magazine*, 59-71.

[7] Cgsecurity. (18 de 04 de 2015). *CGSecurity-TestDisk*. Obtenido de CGSecurity-TestDisk: http://www.cgsecurity.org/wiki/TestDisk

[8] Constanzo, B., & Waimann, J. (2012). El estado actual de las Técnicas de File Carving y la necesidad de Nuevas Tecnologías que implementen Carving Inteligente. *Journal CADI*.

[9] Digital Inteliigence. (2017). *Forensic Software*. Obtenido de Forensic Toolkit - FTK: https://www.digitalintelligence.com/software/accessdata/forensictoolkit/

[10] Esra'a Alshammary, A. H. (2016). Reviewing and Evaluating Existing File Carving. 2016 Cybersecurity and Cyberforensics Conference (págs. 55-59). Jordan: IEEE. *IEEE Xplore*, 55-59.

[11] Guo, Y. (2010). Data recovery function testing for digital forensic tools . *SpringerLink*, 297-311.

[12] IBM. (24 de 05 de 2017). *IBM*. Recuperado el 13 de 09 de 2017, de IBM: https://www.ibm.com/support/knowledgecenter/en/POWER8/p8ebj/arebjsolidstatedrives.htm

[13] IRecovery. (25 de 09 de 2017). Curso intensivo / conviértete en data specialist en recuperación de datos y análisis forense it. Bogotá.

[14] Jedec - Global Standards for the Microelectronics Industry. (2018). *Standars & Documents*. Recuperado el 12 de Febrero de 2018, de Solid State Drives: https://www.jedec.org/standards-documents/focus/flash/solid-state-drives

[15] Mayor, J. J. (2010). 12 Discos de estado sólido: Test. (págs. 33 - 44). Computer hoy.

[16] Ninahualpa, G., Diaz, J., & Gunn, S. (2017). Data Restoration and File Carving. *IEEE Xplore*, 1-5.

[17] Pérez García, M. (2011). *Recuperación de información en discos duros eléctromecánicos a nivel físico y lógico para su análisis forense informático.* México D.F.

[18] Rosenfeld, R., & Irazábal, J. (2013). Computabilidad. En R. Rosenfeld, & J. Irazábal, *computabilidad, complejidad computacional* (págs. 5-6). La Plata: La Plata : Universidad Nacional de La Plata.

[19] Simson, G. (2007). Carving contiguous and fragmented files with fast object validation. *ScienceDirect*, S2-S12.

[20] SNIA. (2018). *The Storage Networking Industry Association (SNIA)*. Recuperado el 13 de Febrero de 2018, de Solid State Storage Standards Explained: https://www.snia.org/forums/sssi/knowledge/standards

[21] Tannhausser. (28 de 06 de 2014). *La mirada del replicante*. Obtenido de La mirada del replicante: http://lamiradadelreplicante.com/2014/06/28/los-mejores-programas-para-la-recuperacion-de-archivos-en-gnulinux/

[22] Toshiba. (2018). *Semiconductores y productos de almacenamiento*. Recuperado el 12 de Febrero de 2018, de Unidades de estado sólido (SSD): https://toshiba.semicon-storage.com/es/product/storage-products/trends-technology/ssd-0.html

[23] Windows *Forensics*. (s.f.). Obtenido de Windows Forensics: https://www.packtpub.com/mapt/book/networking_and_servers/9781784390495/7/ch07lvl1sec44/event-log-recovery-with-evtxtract