Análisis de metodologías de recolección de datos digitales en servidores web

Mónica D. Tugnarelli [(1)], Mauro F. Fornaroli [(1)] , Sonia R. Santana [(1)], Eduardo Jacobo [(1)], Javier Díaz [(2)]

[(1)] Facultad de Ciencias de la Administración – Universidad Nacional de Entre Ríos
[(2)] Facultad de Informática – Universidad Nacional de La Plata
e-mail: montug, maufor,ssantana,ejacobo [@fcad.uner.edu.ar]
jdiaz[@unlp.edu.ar]

Abstract. Cuando se produce un incidente o amenaza de seguridad, en el cual un recurso del sistema queda comprometido o potencialmente expuesto a accesos no autorizados, las técnicas y metodologías de forensia informática deben asegurar que se pueda determinar adecuadamente el qué, quién, cuándo y cómo sucedió el incidente, así como también ocuparse del aseguramiento y preservación de la evidencia recolectada. Este trabajo aborda dos metodologías de recolección de datos digitales, la primera llamada *Enfoque preventivo-Recolección de datos a priori* o *Forensic Readiness* y la segunda *Enfoque reactivo - Recolección de datos a posteriori de un evento de seguridad,* para analizar comparativamente sus prestaciones en base a determinados criterios y puntos de control establecidos sobre servidores web HTTP.

Keywords: seguridad, incidente, forensia, metodologías, HTTP.

1 Introducción

Si una arquitectura de seguridad informática está correctamente definida debe ofrecer un plan y un conjunto de políticas que describan tanto los servicios de seguridad ofrecidos a los usuarios como los componentes del sistema requeridos para implementar dichos servicios. Estas políticas de seguridad se aplican a los activos de información identificados por su relevancia con los objetivos de la organización, conociendo como se gestionan y cuáles son sus riesgos para implementar estrategias y mecanismos que aseguren la confidencialidad, la integridad y la disponibilidad de los mismos [1].
Cuando se produce un incidente o amenaza de seguridad, en el cual un recurso queda comprometido o potencialmente expuesto a accesos no autorizados, esta

arquitectura de seguridad se ve vulnerada. A modo general, como amenazas del entorno, pueden considerarse aspectos que incluyan desde la seguridad administrativa, la seguridad de las comunicaciones, la seguridad informática, la seguridad ambiental hasta la seguridad física. Entonces, la arquitectura de seguridad debe poder afrontar tanto amenazas intencionales como accidentales. Implementar un programa sistemático de monitoreo y gestión de incidentes, basado en el empleo de metodologías, puede proporcionar un enfoque estructurado y organizado para minimizar el impacto del incidente de seguridad y ayudar a proporcionar una respuesta rápida y adecuada.

Diariamente cientos de equipos se encuentran expuestos a potenciales incidentes, consideremos como ejemplo el avance de Internet de las Cosas (IoT) y sus características de trabajo para llegar a dimensionar el grado de posibilidad y el riesgo de ocurrencia de un incidente y su consecuente impacto [2,3]

En este entorno tecnológico las metodologías de forensia informática deben asegurar que se pueda determinar adecuadamente el qué, quién, cuándo y cómo sucedió en relación a ese incidente de seguridad, así como también ocuparse de la correcta preservación y trazabilidad de los datos recolectados.

La definición brindada por la primera Digital Forensics Research Workshop (DFRWS), acuerda que el análisis forense digital o forensia informática es "*El uso de métodos científicamente probados y derivados hacia la preservación, recolección, validación, identificación, análisis, interpretación, documentación y presentación de evidencia digital derivada de fuentes digitales con el fin de facilitar o promover la reconstrucción de los hechos, que pueden constituirse en evidencia legal, o ayudando a anticipar acciones no autorizadas que han demostrado ser perjudiciales para operaciones planeadas*." [4].

Las fuentes proveedoras de datos son numerosas, abarcan desde computadoras, teléfonos celulares, tarjetas de cámaras digitales, chips embebidos, drones, *snapshots* de memoria hasta consolas de videojuegos, es decir cualquier tipo de dispositivo que produzca datos digitales.

La forensia informática requiere entonces, una correcta aplicación de métodos científicos, técnicas y herramientas para cumplimentar las etapas relacionadas con la identificación, preservación y análisis de la evidencia digital la cual, llegado el caso, puede ser considerada legalmente en un proceso judicial por lo cual además se necesita asegurar la calidad y trazabilidad de estos datos.

Planteado este panorama, el Proyecto de Investigación y Desarrollo PID 7052 realizado en la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos busca avanzar en el estudio comparativo de metodologías de recopilación de datos asociados a incidentes de seguridad y, particularmente, analizar la perfomance de estas metodologías en entornos de servidores web. Para ello se han establecido cuatro etapas: Estudio exploratorio de las metodologías de recolección de datos; Análisis del nuevo protocolo HTTP/2; Configuración del entorno de testing y determinación de puntos de control y captura; y por último la construcción de una matriz comparativa entre las metodologías y sus aspectos más relevantes en cuanto a calidad, trazabilidad, disponibilidad de datos y el tiempo de respuesta ante un incidente.

Las tres primeras etapas se han cumplimentado y en los puntos siguientes se presentan de manera resumida los resultados de las mismas. La última etapa de análisis de datos, se encuentra en ejecución y se presentan resultados parciales.

2 Estudio exploratorio de las metodologías de recolección de datos

Actualmente, y a modo general, las metodologías de recolección de datos pueden clasificarse en dos enfoques:

2.1- Enfoque preventivo-Recolección de datos a priori de un evento de seguridad. También conocido como *Forensic Readiness* [5,6,7]. Este enfoque introduce el concepto de resguardar la posible evidencia antes de que ocurra el incidente para cubrir principalmente dos objetivos: maximizar la capacidad del entorno para reunir evidencia digital confiable y minimizar el costo forense durante la respuesta a un incidente. La premisa es que esos datos puedan ser pasibles de ser utilizados no solo como insumo para el análisis de posibles incidentes de seguridad y de recuperación para la continuidad del negocio, sino también como prueba legal lo que involucra el aseguramiento de la prueba a medida que se realiza la recolección activa de los datos.

No solo es un enfoque beneficioso para organizaciones, sino que también puede aplicarse en otros ámbitos, como por ejemplo en los sistemas de voto electrónico, en los cuales no solo hay que ocuparse de la seguridad del software de votación y de los protocolos criptográficos sino que también se requiere de un entorno de confianza que tenga la capacidad de preservar la evidencia para mostrar a los votantes que el proceso funcionó como se esperaba.

Este enfoque, además, plantea que estar preparado para reunir y utilizar evidencia también puede tener beneficios como elemento disuasorio considerando los altos porcentajes de infracción de políticas internas de seguridad.

Algunas de las actividades clave en la planificación de la Forensic Readiness son:

a. Definir los escenarios o activos que pueden requerir de pruebas digitales;
b. Identificar las fuentes disponibles y los diferentes tipos de posibles pruebas;
c. Establecer una forma segura de obtención de pruebas para cumplir con el requisito de admisibilidad legal;
d. Establecer una política para el almacenamiento seguro y el manejo seguro de las evidencias;
e. Garantizar el seguimiento para detectar y prevenir incidentes mayores;

f. Capacitar al personal de modo que todos entiendan su papel en el proceso de pruebas digitales y la sensibilidad jurídica de las mismas;
g. Garantizar el control jurídico para facilitar la acción en respuesta al incidente.

En resumen, la capacidad de una organización para explotar estos datos y la anticipación de la respuesta a un incidente son el foco de la Disponibilidad Forense complementando y mejorando las actividades de la organización en cuanto a seguridad de la información y evaluación de riesgos.

2.2- Enfoque reactivo - Recolección de datos a posteriori de un evento de seguridad. En este enfoque se trata de recuperar la evidencia luego de la detección de un incidente de seguridad, con el objetivo de realizar un análisis forense para determinar lo ocurrido, Este examen forense debe preservar el escenario y garantizar la admisibilidad de las pruebas. Piccirilli [8] aporta en su tesis doctoral, una descripción de las etapas que se pueden aplicar en todos los casos periciales en los que intervengan elementos vinculados a la informática, las cuales incluyen:

- el estudio y análisis del entorno, para identificar la evidencia digital a obtener;
- el análisis de los puntos de pericia, que establecen el objetivo que debe cumplir la evidencia digital;
- la adquisición de la evidencia digital;
- el análisis de la evidencia obtenida, conforme a los lineamientos del cuestionario pericial ordenado;
- la forma de exponer la evidencia digital obtenida en la investigación realizada;
- la preservación de la evidencia digital tratada (para eventuales futuras etapas de investigación, cuya fuente sería la misma evidencia digital).

Los estándares y recomendaciones más ampliamente considerados, tales como la RFC 3227 [9] y la ISO/IEC 27037 [10], plantean un conjunto de puntos comunes para realizar un análisis forense correcto, entre los que se incluyen: la importancia de preservar el entorno de pruebas, cómo y donde se guardan las pruebas, cómo se analizan para obtener el máximo rendimiento y la importancia de la obtención de informes claros y concisos [11,12].

Si bien muchas organizaciones son conscientes de la necesidad de contar con planes de recuperación y continuidad de negocio, la tendencia es implementar este enfoque, es decir, esperar a que ocurra un incidente, tratar de manejarlo y luego reunir las evidencias.

3 Análisis del Protocolo HTTP/2

Hypertext Transfer Protocol (HTTP) es un protocolo de nivel de aplicación con características definidas para utilizarse en sistemas de información distribuidos, colaborativos e hipermediales. Se caracteriza por ser un protocolo cliente/servidor sencillo y ampliamente aceptado, que define la estructura de los mensajes de requerimiento/respuesta así como también la forma en que se realiza el intercambio de dichos mensajes entre los clientes y los servidores web. Se han introducido

mejoras en sus diferentes versiones sobre todo tendientes a mejorar la perfomance del mismo, a reducir el consumo de recursos y la latencia y para resolver algunos inconvenientes que tiene la comunicación a través TCP [13,14].

La última versión HTTP/2 [15], presenta un protocolo binario que incorpora multiplexación y el uso obligatorio de TLS[1] conservando la misma semántica y la compatibilidad con las versiones 1.0 y 1.1 El protocolo se implementa si el cliente y el servidor tienen soporte y en el caso de que alguno de los dos no lo tengan, en la negociación de protocolo, se acuerda usar las versiones anteriores. Actualmente la mayoría de los browsers y entornos de servidor cuentan con implementaciones oficiales para la nueva versión.

En la siguiente tabla, a modo resumen, se describen las principales diferencias entre las versiones:

Versión	HTTP/1.0 (1996)	HTTP/1.1 (2000)	HTTP/2 (2015)
RFC	*RFC 1945*	*RFC 2616*	*RFC 7540*
Manejo de requerimientos	Un requerimiento entregado por vez sobre una conexión.	Mecanismo HTPP Keep Alive: varios requerimientos pueden utilizar múltiples conexiones con el servidor para reducir la latencia.	Múltiples mensajes de requerimiento/ respuesta sobre una misma conexión. Permite asignar prioridades a los requerimientos.
Header	Formato texto	Formato texto	Formato binario. - Compresión de header (Algoritmo HPACK).
Multiplexación	No permite conexiones simultáneas	No permite conexiones simultáneas	Permite múltiples solicitudes y respuestas en paralelo usando la misma conexión TCP, enviando cada requerimiento en un stream diferente.
Servidor	Descarga de recursos a solicitud del cliente (primero HTML, luego CSS, JS, imágenes, enlaces)	Descarga de recursos a solicitud del cliente (primero HTML, luego CSS, JS, imágenes, enlaces)	Tecnología server push: Permite cargar los archivos (CSS, JS, imágenes) desde el servidor al cliente sin que éste lo pida.

[1] TLS (Transport Layer Security) es un protocolo criptográfico que proporciona comunicaciones seguras en una red.

Tabla 1.Principales diferencias entre versiones HTTP

4 Configuración del entorno de testing y determinación de puntos de control y captura

Para desarrollar las actividades y pruebas se ha configurado un entorno de trabajo en una red LAN Ethernet que se conforma con un servidor donde corre un Sistema Operativo Ubuntu Server versión 15.10 y un Servidor Web Apache Server versión 2.4.12 que cuenta con una dirección IP pública. Esta red se completa con seis estaciones de trabajo con conexión a la red cableada y a red inalámbrica.
Para la ejecución de pruebas y adquisición de datos se analizaron herramientas de forensia informática de libre distribución [16,17], tales como CAINE [18], BlackArch Linux [19] y KALI Linux [20]. Se optó por Kali Linux versión 64bit 2017.1, que cuenta con más de 300 herramientas y aplicaciones relacionadas con la auditoria y la forensia.
Como guía general para las pruebas y marco de trabajo se seleccionaron los siguientes documentos:

- RFC 3227: publicado por la *Internet Engineering Task Force* (IETF) donde establece directrices para recopilar y almacenar evidencias sin ponerlas en riesgo.
- ISO/IEC 27037:2012 que proporciona directrices para el manejo adecuado de la evidencia digital gobernada por tres principios fundamentales: la relevancia, la confiabilidad y la suficiencia.
- OSSTMM (*Open Source Security Testing Methodology Manual*) [21]. Uno de los estándares profesionales más completos y utilizados para auditorías de seguridad en sistemas.

La primer actividad consistió en Identificar los puntos de control del protocolo HTTP, para lo cual se realizó la captura, análisis y resguardo de:

a. tráfico entrante y saliente de los puertos 80 y 443 TCP, con el fin de obtener patrones de tráfico.
b. estado de las conexiones establecidas a los puertos 80 y 443.
c. archivos log de Ubuntu Linux (/var/log/):
    - messages.log: registro de mensajes generales del sistema
    - auth.log: registro de autenticación
    - secure: registro de autenticación
    - utmp/wtmp: registro de login

d. httpd: archivos log de Apache: error.log y access.log. El primero proporciona información de diagnóstico y registra cualquier error que ocurre en el procesamiento de los requerimientos; mientras que el segundo almacena todos los requerimientos procesados por el servidor.
e. archivos de configuración del servidor Apache: a fin de determinar modificaciones no autorizadas en la configuración del servidor alterando su funcionamiento.

La recolección de datos contempla las particularidades de cada enfoque:

a. **Enfoque preventivo**
   - Monitoreo y recopilación de datos según los puntos establecidos y detallados en la actividad anterior.
   - Diariamente se realizan dos copias de los datos recolectados los cuales son almacenados en medios externos con protección de integridad mediante hash (MD5).

b. **Enfoque reactivo**
   - Monitoreo de datos según los puntos establecidos y detallados en la actividad anterior.
   - Se realiza un backup estándar diario de los datos

5 Análisis de datos recolectados y construcción de matriz comparativa

Si bien esta etapa esta en ejecución, un análisis inicial de los datos recolectados permite arribar a los siguientes resultados parciales:

- Identificar la fuente y tipo de evidencia requerida agiliza la recolección de pruebas y el almacenamiento de las mismas. Esto, indispensable para el enfoque preventivo, complementa el análisis de riesgo de los activos de la organización. En este caso, los puntos de control especificados sobre el servidor web concentran los recursos dedicados a la captura de datos.
- El registro centralizado es clave tanto para una detección eficiente de un incidente como para la implementación de herramientas forenses.
- El resguardo periódico de datos en un sistema externo los aísla de posibles incidentes de seguridad sobre el servidor web. El volumen de almacenamiento obviamente dependerá de los activos y puntos de control necesarios determinantes para la continuidad del negocio. En este proyecto, y por sus características, el resguardo diario cumple con estos requisitos equilibrando volumen (promedio de 2 Gb), accesibilidad y costo de almacenamiento.
- Considerando la metodología Forensic Readiness, donde lo recolectado es evidencia digital pasible de ser utilizada en instancias legales, la cadena de custodia se puede implementar con almacenamiento externo, la generación de hash por bloque de datos y la incorporación de datos administrativos del tipo fecha, hora y personal actuante.

El próximo paso de esta etapa es la simulación de un ataque activo de Denegación de Servicio (DoS) sobre el servidor de prueba, a los fines de analizar de manera completa los diversos aspectos ya mencionados relacionados con la perfomance de

las metodologías, y fundamentalmente dar respuesta sobre algunas cuestiones pendientes tales como:

- ¿Qué enfoque proporciona mejores tiempos de recuperación operativa?
- ¿Qué metodología proporciona una mejor respuesta ante un incidente de seguridad?
- ¿Qué enfoque brinda un entorno más adecuado para realizar la forensia luego de los incidentes?

Conclusiones y trabajos futuros

En este artículo se han presentado los primeros avances del PID 7052-UNER denominado Análisis de Métodos de Recolección de Datos Digitales iniciado en abril de 2017.
En cuanto a las metodologías de recolección propuestas para el análisis, se realizó un estudio exploratorio de material y publicaciones relacionadas, logrando la identificación de las principales características de los mismos ante un incidente de seguridad.
En una segunda instancia, se analizó la nueva versión del protocolo HTTP con la finalidad de conocer inicialmente sus implicancias en la captura del tráfico de datos. Desde el punto de vista que interesa a este trabajo no se detectaron grandes cambios a nivel aplicación, sí se debería avanzar a futuro en el análisis de su relación con TCP y las restricciones de seguridad con el uso de TLS.
Para la configuración del entorno de trabajo y testing se utilizo software libre y herramientas de licenciamiento libre tales como Ubuntu, Apache y Kali Linux una distro GNU especializada en seguridad informática.
Si bien la tendencia es la respuesta reactiva, es decir, esperar a que ocurra un incidente, tratar de manejarlo y luego reunir las evidencias, los resultados iniciales del análisis de datos recolectados y tratados como evidencia digital muestran indicios de los beneficios tangibles y la importancia que puede tener para una organización adoptar el enfoque preventivo de preparación forense especialmente porque una gran parte de la evidencia requerida está disponible antes del incidente.

Referencias

[1] Incident Management and Response ISACA http://www.isaca.org/
[2] Internet Crime Complaint Center (IC3). Annual Report 2015 http://www.ic3.gov/media/annualreports.aspx
[3] Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires. CyberCrime Informe Final 2013 - Delitos Informáticos. http://delitosinformaticos.fiscalias.gob.ar/wp-content/uploads/2014/02/CyberCrime-Informe-Final-2013-flip.pdf
[4] Digital Forensic Research Workshop (DFRWS). http://www.dfrws.org/
[5] TAN, John. (2001). *Forensic Readiness*. http://isis.poly.edu/kulesh/forensics/forensic_readiness.pdf

[6] Rowlingson, Robert. *A Ten Step for Forensic Readiness*. (2004) International Journal of Digital Evidence. Volume 2, Issue 3.

[7] Pooe, A., Labuschagne, L. *A conceptual model for digital forensic readiness* (2012) http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=6320452

[8] Piccirilli, Dario. (2016). *Protocolos a aplicar en la forensia informática en el marco de las nuevas tecnologías (pericia – forensia y cibercrimen)*. Tesis de doctorado. Facultad de Informática. Universidad Nacional de La Plata. http://hdl.handle.net/10915/52212

[9] RFC 3227 Guidelines for Evidence Collection and Archiving. https://www.ietf.org/rfc/rfc3227.txt

[10] Guidelines for identification, collection, acquisition and preservation of digital evidence" ISO/IEC 27037:2012

[11]. U.S. Department of Justice. Electronic Crime Scene Investigation: A Guide for First Responders, Second Edition. https://www.ncjrs.gov/pdffiles1/nij/219941.pdf

[12] Forte, D. *Principles of digital evidence Collection* (2003) http://www.sciencedirect.com/science/article/pii/S1353485803000060

[13] RFC 1945 Hypertext Transfer Protocol - HTTP/1.0 http://tools.ietf.org/html/rfc1945

[14] RFC 2616 Hypertext Transfer Protocol - HTTP/1.1 http://tools.ietf.org/html/rfc2616

[15] Hypertext Transfer Protocol Version 2 (HTTP/2). https://tools.ietf.org/html/rfc7540

[16] Digital Forensic with Open Tools. (2011). DOI: 10.1016/B978-1-59749-586-8.00001-7. Elsevier.Inc

[17] Tugnarelli, M.; Fornaroli, M.; Pacifico, C. *Análisis de prestaciones de herramientas de software libre para la recolección a priori de evidencia digital en servidores web*. Workshop de Investigadores en Ciencias de la Computación (WICC 2015). ISBN 978-987-633-134-0

[18] Computer Aided Investigative Environment http://www.caine-live.net/

[19] BlackArch Linux. https://blackarch.org/

[20] KALI Linux. www.kali.org

[21] Open Source Security Testing Methodology Manual (OSSTMM) http://www.isecom.org/mirror/OSSTMM.3.pdf