Estudio de Funcionalidades en Herramientas de Informática Forense de Propósito General desde la Perspectiva del Analista

Enrique A. Miranda[1,2], Ana Garis[2] y Daniel Riesco[2]

[1]*Poder Judicial de San Luis - Departamento de Investigaciones de Delitos Complejos*

[2]*Universidad Nacional de San Luis - FCFMyN - Departamento de Informática*
{*eamiranda, agaris, driesco*}*@unsl.edu.ar*

Resumen

Los dispositivos tecnológicos de uso cotidiano (como smartphones, notebooks, CPU, etc.) se han convertido en una pieza crucial dentro de los procesos actuales de investigación, cualquiera sea el contexto y/o el delito, debido a que son potenciales contenedores de evidencia digital. Esto ha generado que el campo de la Informática Forense (IF) surja como una parte central dentro de las investigaciones modernas. Sin embargo, además de colocar a esta disciplina como uno de los ejes centrales dentro de cualquier organismo encargado de hacer cumplir la ley, también ha generado una sobrecarga de trabajo producto de la complejidad que presentan los dispositivos tecnológicos actuales así como también el volumen de datos que estos son capaces de almacenar. En este contexto se han propuesto diferentes aristas que intentan atenuar el problema principal, una de las más destacadas es la implementación de herramientas para asistir al analista forense en las tareas correspondientes. En este trabajo se presenta una descripción general de las funcionalidades que ofrecen las herramientas más frecuentemente utilizadas dentro de los laboratorios de IF, específicamente aquellas que reducen la sobrecarga de procesamiento y análisis de evidencia digital que se registra en dichas instituciones. Por otra parte, se brinda una clasificación elemental de la mismas en "primarias" y "avanzadas" de acuerdo al tipo, y en cierta forma, la complejidad y efectividad de cada funcionalidad descripta.

1. Introducción

Con el avance tecnológico actual estamos viviendo verdaderamente en una era digital. Sin duda alguna este aspecto facilita en gran medida los quehaceres de nuestra vida diaria, al punto tal que no somos verdaderamente conscientes de la cantidad de tareas que realizamos utilizando diversos dispositivos y recursos digitales. Este amplio uso de la tecnología informática en casi todos los ámbitos, sumado al avance revolucionario de las telecomunicaciones, han favorecido la generación de un "entorno virtual", denominado *ciberespacio* en el que se llevan a cabo un sinnúmero de actividades y en donde, de alguna manera, se gestiona un gran volumen de información relacionado a las mismas [37, 40].

Sin embargo, este entorno también es utilizado para perpetrar distintos tipos de delitos, ya sean aquellos que antes se llevaban a cabo en otros entornos (como por ejemplo las estafas) como así también nuevos delitos perpetrados exclusivamente en el ciberespacio [15]. Teniendo esto en cuenta, las investigaciones y los procesos judiciales han tenido que adaptarse a este nuevo paradigma. Uno de los aspectos más destacados de esta adaptación está relacionada con el resguardo y obtención de la información que es posible relevar a partir del análisis de distintos medios utilizados para interactuar con dicho entorno. En pocas palabras, se hace referencia a la evidencia digital, la cual se torna cada vez más esencial y se encamina a convertirse en un medio de prueba fundamental dentro del contexto de cualquier investigación, posiblemente desplazando con el tiempo en gran medida a la evidencia física [42, 8]. En este contexto se torna trascendental una disciplina denominada Informática Forense (IF), la cual posibilita la detección, resguardo y recuperación de la información digital que sirve de evidencia a la hora de reconstruir un hecho o sucesión de ellos [18, 4, 42].

Tomando como base lo mencionado en párrafos precedentes, está claro que la era digital trajo aparejadas muchas oportunidades pero también ciertos desafíos para la IF y en consecuencia para las agencias/laboratorios que asisten a las fuerzas de la ley (o como se las referencia en inglés *law en-*

forcement agencies) mediante el uso de métodos, técnicas y herramientas relativas a dicha disciplina.

En este sentido, el autor Simson Garfinkel [18] vaticinaba en el año 2010 que la "era dorada" de la IF finalizaba de alguna manera y que se avecinaba una especie de crisis que pondría en estrés este tipo de instituciones. Básicamente el autor hace referencia a los distintos retos que se han ido sumando con el paso del tiempo y que, entre otras cosas, se encuentran vigentes en la actualidad para cualquier laboratorio de IF; dentro de los más relevantes es posible destacar la proliferación de medios de comunicación, la cantidad de dispositivos tecnológicos de uso cotidiano con capacidad de almacenamiento y transmisión de datos, la encriptación de información almacenada, el almacenamiento en la nube, etc. [17, 40, 18]

A la par de los desafíos mencionados, también se dio un fuerte crecimiento en la oferta de herramientas que asisten a las actividades más demandadas en la disciplina [48, 1]. Muchas de estas herramientas intentan atacar los desafíos planteados anteriormente y contrarrestar las dificultades que se presentan al intentar realizar pericias informáticas en esas condiciones. Dentro de la gran variedad de soluciones disponibles, existen herramientas que ofrecen funcionalidades para asistir al analista en diversas tareas dentro del contexto de IF. Como ejemplos se pueden mencionar herramientas que facilitan la extracción y análisis de información en memoria RAM, análisis del registro de los distintos sistemas operativos, asistencia en el proceso de *triage*[1], adquisición forense remota (*remote forensics*), acceso y análisis de información en la nube (*cloud forensics*) etc. [17, 37, 13, 44]. Sin embargo, las herramientas más utilizadas dentro de cualquier laboratorio de IF son aquellas que posibilitan el análisis de los dispositivos más frecuentemente encontrados en las investigaciones: CPU, notebooks, teléfono celulares, dispositivos de almacenamiento masivo (pendrive, microSD, disco rígido, etc.), entre otros. Este tipo de herramientas, que se pueden describir como "herramientas de IF de propósito general", son utilizadas la mayor parte del tiempo por el analista cuando el objeto de investigación involucra dispositivos tecnológicos [4, 38, 48].

Si bien la oferta de este tipo de herramientas es cada vez más numerosa y la misma cubre una amplia gama de tareas y especialidades, desde la comunidad se refleja en distintos ámbitos las grandes dificultades en las que se traducen estos nuevos escenarios; entre los más resonantes se pueden mencionar la capacidad de recuperación de información sobre el volumen de dispositivos a peritar, inconvenientes respecto a la capacidad de almacenamiento de la información procesada, tiempo de procesamiento y análisis para concretar los informes, recuperación e interpretación de información encriptada, entre otros [1, 17]. Dentro de este conjunto de retos y dificultades que se plantean en el área, se puede señalar como uno de los más recurrentes la demora asociada a la cantidad de dispositivos a peritar y consecuentemente, el volumen de información que debe ser analizada.

Con el fin de contribuir a la disciplina y dar tratamiento a las problemáticas antes mencionadas, este trabajo tiene como principales objetivos i) llevar a cabo un estudio de las funcionalidades principales presentadas por las herramientas más utilizadas en los laboratorios de IF y ii) realizar una clasificación de dichas funcionalidades en primarias y avanzadas para resaltar aquellas que proporcionan mejores capacidades para reducir el tiempo de análisis por parte del examinador forense.

El trabajo se organiza de la siguiente manera. En la próxima sección se detalla sobre una de las problemáticas más relevantes a resolver dentro de la IF: el volumen de información contenido en los dispositivos que son peritados, lo cual impacta en los tiempos de procesamiento y en el análisis de los casos investigados. En la Sección 4 se describen algunas de las funcionalidades más destacadas que presentan las herramientas de IF de propósito general, proporcionando una clasificación primaria que tiene en cuenta el tipo, la complejidad y efectividad de cada una. En la sección 5 se discute sobre las temáticas mencionadas previamente, remarcando los aspectos más significativos. Por último, en la Sección 6 se exponen las conclusiones y algunas líneas de trabajo futuro.

2. El Volumen: un Gran Desafío en Informática Forense

Con el aumento de dispositivos digitales involucrados en diferentes ilícitos, la demanda para los laboratorios de IF ha sido cada vez más elevada [16, 3]. Los organismos encargados de hacer cumplir la ley de todo el mundo han experimentado dificultades para abordar esta demanda. Si bien hay un conocimiento de la situación por parte de dichos organismos, el problema no sólo persiste sino que cada vez es más complejo. [37]

Muchos de los trabajos en IF de hace algunos años ya identificaban como uno de los principales desafíos de la disciplina al tamaño de los medios de almacenamiento de los dispositivos que se peritaban en ese momento [8, 3].

Como un claro ejemplo de esto, Garfinkel [18] en el año 2010 vaticinaba un problema que sería uno de los más complejos de resolver incluso en la actualidad: "el tamaño cada vez mayor de los dispositivos de almacenamiento implica que con frecuencia no hay tiempo suficiente para crear una imagen forense de un dispositivo o incluso para procesar todos los datos una vez que se obtienen".

Si bien los motivos de este abrupto crecimiento no son objeto de investigación en este trabajo, es posible mencio-

[1] *Triage* es una técnica que en términos generales permite la inspección preliminar de fuentes potenciales de evidencia digital, con el fin de identificar rápidamente aquellos elementos más factibles que podrían contener evidencia relevante para el objeto de investigación [8].

nar algunos factores que explican en gran medida el crecimiento en la capacidad de almacenamiento de los dispositivos, como por ejemplo la facilidad en términos económicos de acceder a dispositivos con gran capacidad de almacenamiento [39], el uso masivo de smartphones y *wereables* [14], incluso también factores más indirectos como por ejemplo las resoluciones cada vez más grandes en las cámaras de los smartphones [46]. Analizando cualquiera de estos factores, es claro que hay una marcada tendencia a que el panorama sea cada vez más complejo para el proceso de análisis forense.

Aunque este problema puede ser asociado a diversos aspectos, uno de los más determinantes es que la capacidad para recuperar, procesar, analizar y almacenar datos no aumenta al mismo ritmo que la capacidad de almacenamiento de los dispositivos modernos [37]. Con el paso del tiempo esto ha ido abriendo una brecha entre los grandes volúmenes de datos que se deben analizar y los medios y herramientas de IF, las cuales en su gran mayoría, proponen como principal recurso el poder de procesamiento computacional. Si bien este aspecto es esencial en cualquier herramienta informática, está claro que para poder contrarrestar el problema referido anteriormente es necesario el estudio minucioso de todo el proceso que conlleva cada tipo de análisis forense, poniendo especial énfasis en las herramientas de IF que asisten al experto en estas tareas.

Las herramientas actuales que se utilizan en una investigación digital como por ejemplo Encase, Magnet Axiom, Cellebrite UFED, FTK, Sleuthkit Autopsy, XRY, ProDiscover, entre otras, brindan a los analistas diversas funcionalidades tales como explorar sistemas de archivos, realizar búsquedas de palabras clave y emplear una variedad de otras técnicas de análisis. Pero estas herramientas, las cuales son referidas por algunos autores como "herramientas de IF de propósito general", están en constante lucha por mantenerse al día con las cargas de trabajo de las demandas que exige el análisis forense moderno [3, 40].

En la próxima sección se mencionan las herramientas que han sido consideradas para análisis en este trabajo y la metodología utilizada para relevar las distintas funcionalidades que las mismas proporcionan.

3. Herramientas Consideradas y Metodología

Desde los inicios de la IF se han propuesto numerosas herramientas para asistir en las diferentes tareas que el especialista debe llevar a cabo, algunas se han mantenido vigentes y otras por diversos motivos se han discontinuado [48, 1]. Dentro del conjunto que se encuentra vigente, es posible analizar herramientas con diferentes alcances y objetivos, donde muchas se enmarcan dentro de ciertas áreas específicas en IF. Como ejemplo puntual de este tipo de herramientas se puede mencionar a Volatility[2] o Network Miner[3], las cuales facilitan ciertas tareas puntuales, la primera para inspección de información de memoria RAM y la segunda para forensia de redes. El tipo de herramientas seleccionadas para ser analizadas en este trabajo son las consideradas como "de propósito general"', las cuales podrían clasificarse como aquellas que asisten al usuario en las tareas más significativas y recurrentes dentro del proceso de análisis forense. Debido a la extensión del trabajo, este no pretende ser un listado exhaustivo sino una descripción de aquellas funcionalidades más esenciales que facilitan la labor diaria y principalmente hacer hincapié en las funciones que apuntan a reducir el tiempo de cada investigación y por consiguiente, la sobrecarga en los procesos dentro del laboratorio de IF.

Para el desarrollo de este trabajo se han analizado los siguientes productos de software: Magnet Axiom Process y Examine, FTK, Encase, XRY, Oxygen Forensic Detective, ProDiscover, Cellebrite Physical Analyzer, Cellebrite Pathfinder, Sleuthkit Autopsy y Mobile Edit. Si bien no estarían dentro de la misma clasificación, también se destacan algunas funcionalidades de Cellebrite UFED Cloud, Magnet Axiom Automate y XEC Director. Las versiones consideradas de todos los productos son aquellas disponibles hasta marzo del 2021.

El estudio de las funcionalidades descriptas en este trabajo se realizó mediante distintas modalidades. Cierto conjunto de herramientas de IF se vienen relevando dentro del contexto de la actividad pericial. Para algunos de los productos que requieren licencia, se solicitaron los demos correspondientes y también mediante el desarrollo y estudio de casos de prueba sobre los mismos. En determinados casos, para el análisis de las funcionalidades se inspeccionó la documentación, instructivos y capacitaciones disponibles en la web sobre dichas herramientas.

En la próxima sección se analizarán algunas de las principales funcionalidades que las mismas prestan para facilitar las tareas al analista forense, poniendo especial énfasis en aquellas que atenúan de alguna manera la demora en el laboratorio de IF.

4. Funcionalidades en Herramientas de Informática Forense

4.1. Funcionalidades Primarias

Desde la óptica de la IF, es posible hablar de ciertas características que toda herramienta de propósito general debería poseer. Actualmente, la mayoría de estas funcionalidades están incluidas en las herramientas forenses de propósito

[2] https://www.volatilityfoundation.org/

[3] https://www.netresec.com/?page=networkminer

general, como por ejemplo UFED Physical Analyzer, Encase, Autopsy, Magnet Axiom, ProDiscover, FTK, XRY, Mobile Edit, etc. [3]. A continuación se mencionan las características que, de acuerdo a las problemáticas planteadas anteriormente, consideramos deben estar presentes.

4.1.1. Búsquedas por palabras claves

Cualquier herramienta forense que realice análisis de información digital define dentro del flujo de procesamiento una etapa donde indexa cada uno de los artefactos relevados[4]. Uno de los objetivos de esto es facilitar distintos tipos de búsquedas a través de la interfaz con el usuario . Teniendo en cuenta esta funcionalidad, es posible identificar una amplia gama de capacidades de búsqueda y visualización que mejoran la función esencial que sería buscar una cadena de caracteres sobre toda la información analizada. Como ejemplo de estas capacidades se pueden mencionar, compatibilidad con expresiones regulares, sensibilidad a mayúsculas, soporte de distintas codificaciones y caracteres especiales, búsquedas concurrentes, búsqueda automática por listado de palabras claves, subcadenas, resaltado de palabras claves, soporte de enmascaramiento de campos sensibles, búsqueda por hash, uso de operadores booleanos y de string, entre otras [22].

4.1.2. Clasificación y filtrado de artefactos

Actualmente una de las funcionalidades esenciales que toda herramienta de IF de propósito general debe poseer es la identificación, clasificación y visualización de la información relevada. Durante el procesamiento de la información que se encuentra almacenada en un dispositivo se van identificando archivos, cadenas de caracteres o configuraciones, que deben ser indexados y clasificados para ser visualizados y así facilitar la consulta y filtrado por parte del analista [4, 38]. A modo de ejemplo, durante el procesamiento de una imagen forense de un disco rígido, se pueden encontrar una gama diversa de archivos con información del sistema operativo, multimedia, múltiples registros con datos de sesiones, etc. Cada uno de estos, son denominados artefactos, y deben ser clasificados por la herramienta y puestos a disposición del analista forense [23]. De esta manera, en general las herramientas ofrecen una vista de archivos del dispositivo peritado, pero como opción primordial brindan una vista de artefactos, donde los mismos se encuentran categorizados y agrupados por tipos. Por ejemplo, en herramientas de análisis forense de móviles, en un primer nivel es posible encontrar categorías como “Chats”, “Multimedia”, “Configuraciones”, “Aplicaciones instaladas”, etc. De esta manera, quien consulta la información debe identificar cuales serán las categorías de interés y buscar dentro de las mismas los artefactos encontrados por la herramienta. En este sentido, también es primordial que la herramienta ofrezca opciones de filtrado, búsqueda y ordenamiento sobre los artefactos recuperados, de acuerdo a las características de cada tipo [31, 20]. Como por ejemplo, a la hora de filtrar, ordenar y/o buscar imágenes, es importante que la herramienta proporcione opciones de filtrado de acuerdo a los metadatos asociadas a las mismas.

En la Figura 1 se puede observar una captura con parte de la GUI de la herramienta Axiom Examine de Magnet. Específicamente el panel de clasificación de artefactos recuperados a partir del medio analizado.

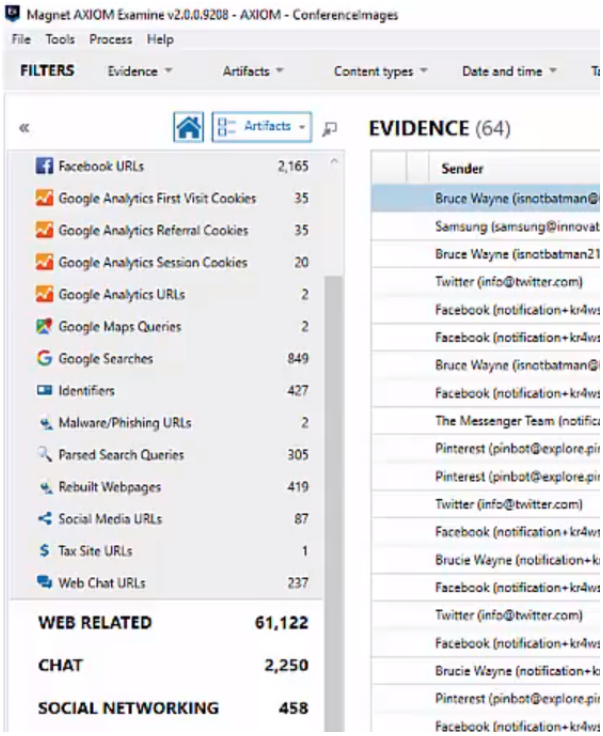

Figura 1: Vista parcial con clasificación de artefactos de Magnet Axiom Examine.

4.1.3. Data carving

Data carving es un término general ampliamente utilizado en el contexto de IF para hacer referencia a la recuperación de información a partir del espacio no asignado de la memoria, según las características específicas del formato de archivo presentes en los mismos [2, 38]. Este espacio de memoria no asignado aún puede contener información relevante para una determinada investigación producto de una eliminación intencional o eliminaciones automáticas

[4]En este contexto, la palabra *artefacto* hace referencia a cualquier elemento que puede ser recuperado de un dispositivo bajo análisis, como por ejemplo imágenes, conversaciones, entradas en un registro de un sistema operativo, links web, un sector de la memoria no asiganda con información, etc.

de archivos temporales. Desafortunadamente, estos datos no siempre son de fácil acceso. Actualmente, existen un gran número de algoritmos de *data carving* y la mayoría de las herramientas de IF de propósito general poseen esta funcionalidad [18]. Claramente, hay herramientas más robustas que otras con respecto a esta característica, sin embargo, las más utilizadas poseen funcionalidades, cuanto menos capacidades básicas, relacionadas con el análisis del espacio de almacenamiento no asignado [38].

4.1.4. Línea de tiempo

En el contexto de IF análisis de línea de tiempo juega un rol fundamental, ya que aprovecha la propiedad temporal única de los datos para ordenar una investigación y reconstruir eventos pasados [26]. Un evento es una acción única que ocurre en un momento dado y durante un tiempo determinado. Un evento puede ser la redacción de un documento, la visita a un sitio web o una conversación de chat con alguien. La línea de tiempo permite a los investigadores tener una visión global del caso y saber, por ejemplo, qué dispositivos se utilizaron, qué sistemas se estaban ejecutando o qué archivos se han modificado en un momento determinado. Cuando se mejora con múltiples fuentes, el análisis de línea de tiempo puede ayudar a descartar hipótesis específicas, identificar pruebas que necesitan un procesamiento adicional e incluso relevar con certeza evidencia potencialmente crítica [10]. Muchas herramientas de IF actuales poseen distintas funcionalidades para analizar líneas de tiempo, como por ejemplo capacidades de filtrado y ordenamiento con criterios temporales, vistas con grafos temporales, entre otros. En la Figura 2 se puede observar un ejemplo de visualización de línea de tiempo en la herramienta UFED Physical Analyzer.

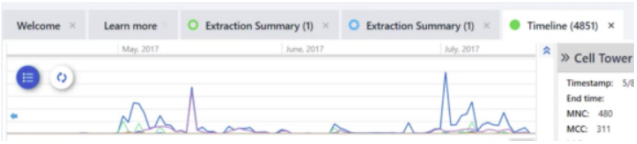

Figura 2: Ejemplo con parte del *widget* de línea de tiempo en UFED PA.

4.1.5. Múltiples vistas y visualizadores integrados

En la actualidad existen pocas herramientas de propósito general que no posean una interfaz gráfica de usuario (*GUI* por sus siglas en inglés) que haga uso de las facilidades que proporcionan las estrategias de Visualización de la Información [45, 6]. De la misma manera que en cualquier herramientas de software, la carga cognitiva del usuario se reduce siempre que se integran distintas vistas interconectadas dentro la interfaz. Las herramientas de IF propósito general actuales se caracterizan por presentar distintas vistas de la información recuperada (como por ejemplo, línea de tiempo y/o vistas en miniatura de la información multimedia) [12, 38]. Incluso, muchas de estas presentan visualizadores externos que facilitan el análisis de información. Como ejemplo de esto se puede mencionar los lectores de código hexadecimal o las herramientas de reproducción de archivos multimedia. En la Figura 3 se muestra una captura de la herramienta Autopsy y algunas de sus vistas mientras que en la Figura 4 se muestra la vista del grafo de relaciones y comunicaciones de Oxygen Forensic Detective.

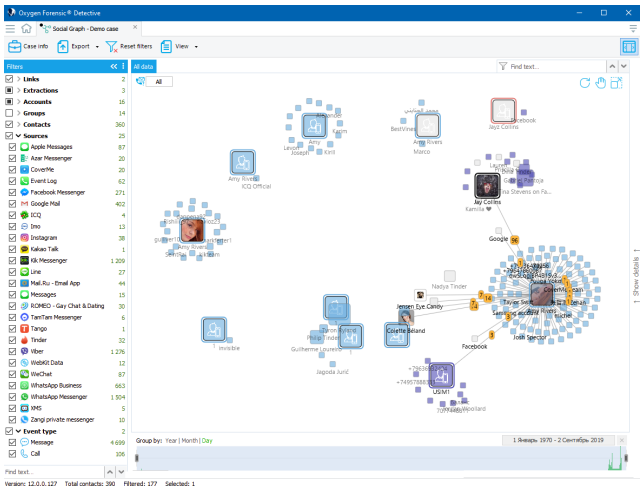

Figura 4: Vista de grafo de comunicaciones de Oxygen Forensic Detective.

4.1.6. Generación de reportes (formatos simples)

Esta funcionalidad hace referencia a la capacidad de una herramienta de IF de resumir el proceso que se llevó a cabo y la información importante del caso investigado en un documento o recurso de visualización que posibilite el análisis de la información relevante recuperada durante la investigación [3]. Para generar un reporte preciso y confiable, es crucial mantener registros de todas las etapas forenses previas. Esto depende en gran medida que exista un proceso de documentación sólido, notas, fotos y contenido producido por las herramientas correspondientes [38].

La referencia específica a “reportes simples”, es porque la gran mayoría de herramientas de IF cuenta con cierto tipos de funcionalidades de generación de reportes que podrían considerarse elementales. Como ejemplo de esto se podrían mencionar herramientas que generan reportes en formatos PDF, HTML, etc. En este sentido, se hace referencia a la

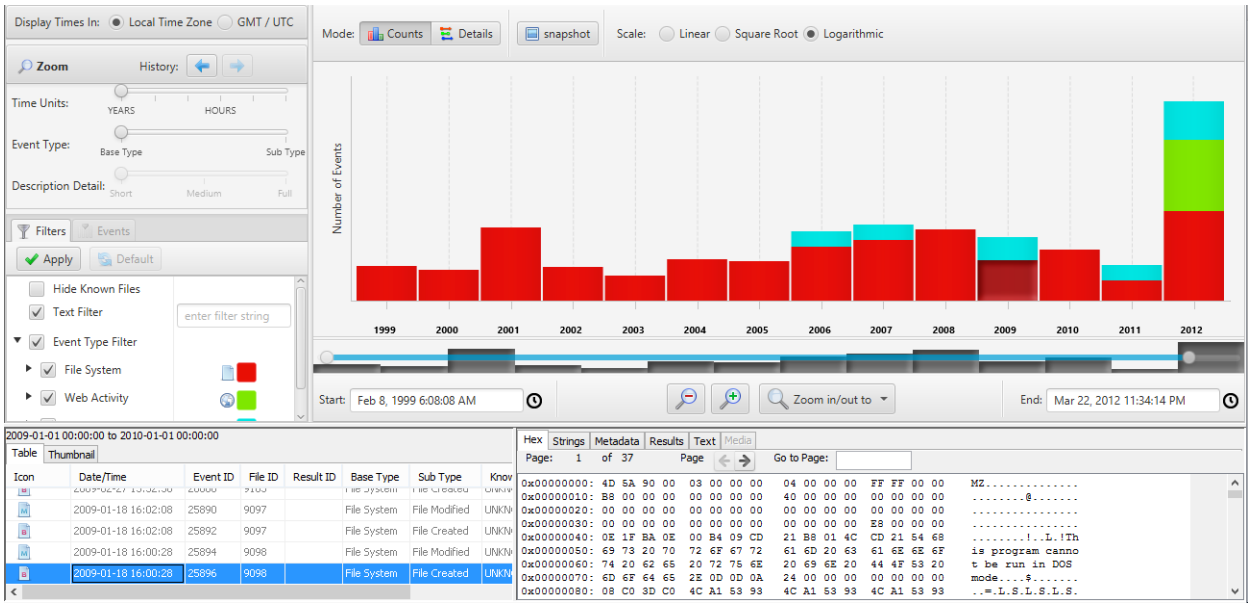

Figura 3: Algunas vistas de Autopsy.

generación de reportes con un grado bajo de interacción por parte del usuario. Esto en contraste con aquellos que se mencionan en el próximo apartado.

4.1.7. Importación de bases de datos (hash)

Debido a la gran cantidad de información que debe ser procesada por las herramientas de IF de propósito general, es muy importante contar con mecanismos de clasificación automática de artefactos *relevantes* o *irrelevantes*. En principio, esta funcionalidad es cubierta mediante la importación y análisis de bases de datos de hash de diversas fuentes [43]. De esta forma, las herramientas pueden clasificar de forma automática o semi-automática distintos artefactos que de antemano son conocidos en su entorno. Generalmente esto se realiza mediante el uso de funciones hash (MD5, SHA-1, SHA-2, etc.) [31] que posibilitan identificar unívocamente cualquier tipo de archivo. Un claro ejemplo de este tipo es la National Software Reference Library (NSRL) de NIST [35], la cual es una librería que tiene indexados hashes de artefactos ampliamente conocidos por estar asociados a sistemas operativos, programas, aplicaciones móviles, etc. Este tipo de bases de datos son también conocidos como *listas blancas* [5]. Por otra parte, también es posible importar bases de datos de hash con artefactos que representen elementos significativos bajo distintos contextos. Como ejemplo de esto se podrían mencionar las bases de datos de artefactos relacionados a identificación de material de abuso sexual infantil [28, 30] o también aquellas que permiten identificar de forma automática artefactos relacionados con malware [5]. Este tipo de bases de datos suelen ser referenciadas como *listas negras* y si bien comparten características con la identificación de artefactos con las listas blancas, poseen sus propias dificultades en el sentido en que, en general, son artefactos que cambian constantemente y por consiguiente su código hash. Más allá que esta fuera del foco de este trabajo, es importante destacar que existe una amplia bibliografía respecto al tratamiento y análisis de listas blancas/negras mediante aproximaciones con funciones de hash [29, 5, 47, 11, 20].

[5]Este nombre hace referencia a bases de datos (*listas*) de archivos o artefactos que son conocidos por no representar elementos significativos [29].

4.2. Funcionalidades Avanzadas

Si bien no es posible clasificar de manera unívoca qué funciones son más esenciales o simples que otras, con el paso del tiempo las herramientas de IF de propósito general han ido incorporando funcionalidades más *avanzadas* o *complejas* las cuales han facilitado muchas tareas relacionadas con los procesos llevados a cabo por el analista forense. A continuación se brinda un listado no exhaustivo de este tipo de funcionalidades.

4.2.1. Artefactos definidos por el usuario

Una de las características destacables que se observan al analizar algunas herramientas de IF de propósito general es la posibilidad no sólo de identificar y pre-clasificar artefactos conocidos de los distintos sistemas operativos, sino también que el usuario de la herramienta pudiera especificar y definir, mediante algún caso de uso, nuevos artefactos de acuerdo a sus requerimientos. De esta manera, el analista

forense puede incorporar al procesamiento ciertos aspectos que la herramienta de IF no tiene desarrollada por defecto. Como ejemplo de herramientas se pueden mencionar a Magent Axiom, la cual implementa la plataforma *Artifact Exchange* [32] que posibilita a cualquier usuario el intercambio de artefactos de desarrollo propio para la herramienta mencionada. Si bien de las herramientas analizadas, sólo Magnet Axiom presenta esta funcionalidad definida y documentada, se podrían considerar otras herramientas que posibilitan la definición y especificación de artefactos propios por medio de la implementación de plugins en la herramienta de propósito general. Una breve explicación de este tipo de funciones se desarrollará en el próximo apartado.

4.2.2. Extensibilidad con módulos adicionales (*plugins*)

Con el paso de los años, las herramientas de IF de propósito general han ido incorporando nuevas funcionalidades y capacidades de análisis para ir cubriendo la evolución de los sistemas operativos, las aplicaciones emergentes, el amplio crecimiento en los navegadores, etc. Sin embargo, dada la variedad y el tipo de sistemas, bases de datos y artefactos que en general que deben ser interpretados, es difícil que dichas herramientas puedan brindar cobertura total sobre cualquier tipo de sistema, incluso sobre los diversos flujos de análisis que un analista puede requerir. En este contexto es que algunas brindan capacidades a los usuarios para definir sus propios módulos y de esta manera ampliar la gama de funcionalidades que la herramienta ofrece, desde un punto de vista generalizado. Como ejemplo de herramientas de IF que brinden este tipo de funcionalidades se destacan Sleuthkit Autopsy [7], Cellebrite Physical Analyzer [9] y ProDiscover [19]. La primer herramienta proporciona un ecosistema de módulos adicionales y facilidades para incorporar nuevas funcionalidades aparte de las ya implementadas por defecto en la misma [41]. Uno de los aspectos que fomenta este entorno es que el proyecto Autopsy es de código abierto. Las herramientas Physical Analyzer y Prodiscover proporcionan al usuario la posibilidad de implementar nuevas funcionalidades mediante el uso de scripts en el lenguaje Python y Perl [24, 21, 19]. En contraposición con Autopsy, las herramientas Physical Analyzer y Prodiscover son del tipo software propietario.

4.2.3. Análisis integrado de múltiples recursos

Con la proliferación de dispositivos inteligentes de uso cotidiano, el analista forense se puede encontrar con una amplia gama de artefactos a peritar dentro de un mismo caso. Muchas de las herramientas de IF de propósito general posibilitan la recuperación de información desde distintos tipos de dispositivos, con diferentes sistemas operativos y proporcionan facilidades para el procesamiento de la información que los mismos contienen. En la Figura 5 se muestran las opciones de fuentes a procesar por parte de Magnet Axiom Process y UFED 4PC. Como se puede observar, las dos herramientas posibilitan el análisis de distintos tipos de recursos, tales como pendrives, tarjetas de memoria, computadoras, dispositivos móviles, etc. Particularmente, las herramientas UFED 4PC y Oxygen Forensic permiten el análisis de información contenida en drones e incluso esta última de dispositivos IoT y *wereables*. Sin embargo, dentro de estas opciones, las herramientas Magnet Axiom Process y Oxygen Forensics Detective son de las pocas que posibilita la integración de información de distintos tipos de dispositivos. A modo de ejemplo, se puede analizar en conjunto una captura de memoria RAM y la copia forense del disco de una notebook; información del dispositivo móvil en conjunto con cuentas en la nube todo correspondiente a un mismo usuario o un mismo caso. Este tipo de capacidades posibilita el análisis en conjunto de todos los elementos asociados a una investigación y permiten interrelacionar los artefactos recuperados de todas las fuentes.

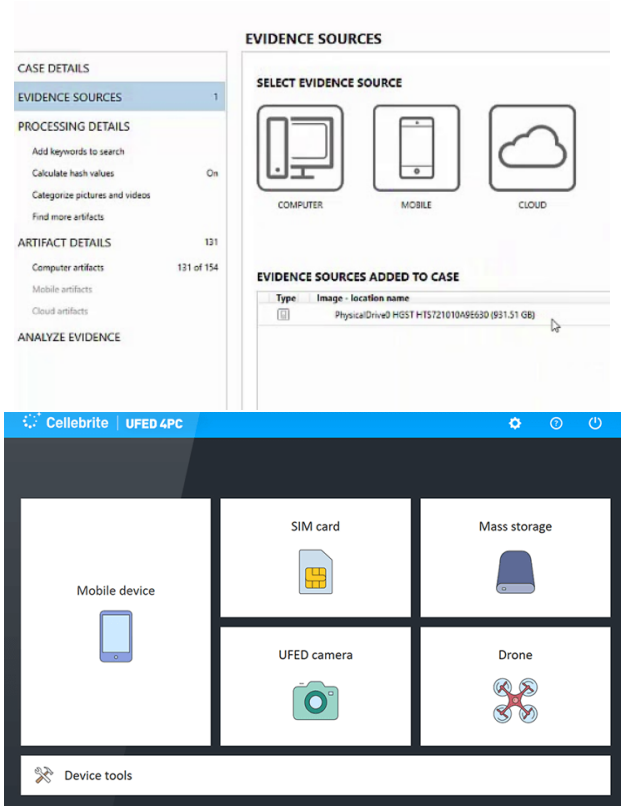

Figura 5: Selección de fuentes y dispositivos a procesar en Magnet Axiom Process (arriba) y UFED 4PC (abajo).

4.2.4. Análisis automático sobre archivos multimedia

Dado que frecuentemente las fuentes de evidencia contienen múltiples imágenes y videos, el análisis y la clasificación mediante la observación individual puede convertirse en una tarea altamente costosa en tiempo. Es por esto que las herramientas de IF de propósito general deben proporcionar facilidades para la pre-clasificación, filtrado y comparación de estos archivos. Anteriormente se han mencionado funcionalidades primarias y avanzadas que están directamente relacionadas y asisten en el análisis de este tipo de archivos. Como ejemplo se pueden mencionar la identificación y clasificación automática de artefactos, en este caso, clasificar las imágenes, por tipos, tamaños y fuentes. Otro ejemplo se vincula a la importación de bases de datos hash para identificar imágenes que son inherentes al sistema operativo, o también a imágenes que han sido clasificadas en bases de datos globales como material de abuso sexual infantil. Sin embargo, existen otro tipos de atributos y funcionalidades más avanzadas para asistir en el análisis de este tipo de archivos. Debido a la gran cantidad y variedad de este tipo de archivos, algunas herramientas han ido incorporando capacidades más avanzadas, como por ejemplo, estrategias para comparar archivos de una manera más "inteligente" que la comparación por valores de hash[6]. Un ejemplo de estas capacidades es la funcionalidad denominada *PhotoDNA* desarrollada por Microsoft para comparación de imágenes y videos, originalmente creada para asistir al análisis de material de abuso sexual infantil en linea [34]. Esta funcionalidad está incorporada en la herramienta Magnet Axiom Examine. Otras herramientas ofrecen funcionalidades para identificación y categorización de objetos en las imágenes y videos. Por ejemplo, las herramientas Cellebrite Pathfinder y Oxygen Forensics Detective permiten la clasificación en estos archivos de ciertos delitos o elementos que podrían ser de interés de acuerdo al caso, tales como drogas, violencia, autos, caras, pornografía, armas, etc. Otra capacidad interesante dentro de este contexto es la extracción de texto desde imágenes o videos, lo cual facilita, entre otras cosas, la búsqueda por palabras claves sobre este tipo de archivos. Algunas herramientas que de alguna manera proporcionan esta característica son Cellebrite Pathfinder, Oxygen Forensics Detective, Encase, entre otras.

4.2.5. Generación de reportes interactivos

A través del tiempo se han definido diversos modelos del proceso de análisis forense que especifican diferentes etapas [38, 1], algunos más específicos y otros más generales y abaracativos, sin embargo en todos se describe a la etapa de reporte (o comunicación de resultados) como una de las más importantes. Aún cuando las herramientas posean atributos deseables para el analista, éstas tendrán limitaciones importantes si carecen de funcionalidades para reportar los resultados obtenidos de manera efectiva. En este sentido, la mayoría de las herramientas de propósito general disponibles poseen la capacidad de generar reportes en distintos formatos como PDF, DOC, XML, HTML, etc. [25] (reportes en formatos *simples*). Sin embargo, sólo algunas poseen la capacidad de generar reportes que pueden ser visualizados en herramientas con prestaciones similares a las proporcionadas por la herramienta de IF donde se realizó el procesamiento y análisis del caso. En comparación a los reportes en formatos simples, esto ofrece distintas funcionalidades que brindan mejores capacidades de navegación de la información. A modo de ejemplo, la información que se visualiza en un tabla estática en un PDF, es la misma que sobre un tabla dinámica en una aplicación; sin embargo, las opciones de filtrado, búsqueda y visualización sobre la tabla dinámica son superiores que en las estáticas. Algunas de las herramientas que brindan esta posibilidad son Cellebrite Physical Analyzer, Magnet Axiom, XRY y Oxygen Forensic. Las primeras dos brindan la opción de crear un caso portable autocontenido que al ser ejecutado abre una aplicación con interfaz gráfica y prestaciones similares a la herramienta con la que se procesó y analizó[7]. En el caso de XRY y Oxygen Forensic brindan herramientas gratuitas separadas para visualización y analisis de reportes.

4.2.6. Análisis de información en la nube

Debido a la proliferación de plataformas de internet como servicios de almacenamiento, redes sociales, streaming, etc. la información que los usuarios almacenan en la nube hoy en día es muy valiosa en cualquier contexto forense [36]. Aunque existen fuentes en la nube que podrían ser recuperadas desde un dispositivo, es cada vez mas frecuente que las aplicaciones y plataformas con base en la nube no mantengan información en los dispositivos del usuario [13]. Es por este motivo que las herramientas de IF de propósito general han ido incorporando capacidades de procesamiento y análisis de la información de las distintas fuentes referidas. Si bien casi todas herramientas poseen características relacionadas con este tipo de fuente de información, pocas brindan módulos para extracción, procesamiento y análisis. Mientras algunas tienen dichas funcionalidades integradas dentro del producto principal (como por ejemplo Oxygen Forensic Detective y ProDiscover), otras ofrecen plugins o productos añadidos de manera externa (como por ejemplo Magnet Axiom, XRY y Cellebrite UFED Cloud).

[6] Uno de los problemas de la comparación de archivos por hash, es que cualquier ínfima alteración sobre el archivo comparado, los archivos serán distintos. Esto en el contexto del análisis de imágenes y videos es un problema recurrente

[7] Cabe aclarar que una de las desventaja de utilizar estos reportes es que los mismos requieren de la ejecución de una aplicación que corre sobre un sistema operativo en particular, mientras que un reporte en HTML o PDF puede ser visualizado en cualquier plataforma.

5. Discusión

Es posible remarcar algunos aspectos que resultan significativos a la temática desarrollada en este trabajo. En primer plano, es dable enfatizar que en períodos cortos, quizás en términos de meses, una funcionalidad descripta o clasificada como avanzada pasa a ser una capacidad primaria o elemental en toda herramienta de IF de propósito general. Esto se da debido a la alta demanda de nuevas características por parte de los usuarios en conjunto con y el ambiente competitivo que rodea al ecosistema de herramientas disponibles en la disciplina. Un ejemplo de esta transición son las capacidades de análisis de información en la nube, las cuales han sido incorporadas con el paso del tiempo por muchos productos y se espera que sean cada vez más explotadas en corto plazo, llegando en algún punto a ser un requerimiento indispensable en toda herramienta en la disciplina.

También es significativo destacar que en la actualidad se están dedicando recursos al desarrollo de herramientas específicas que hacen a la mejora en la reducción del tiempo y esfuerzo que demanda el procesamiento y análisis de evidencia digital. Bajo este contexto, se pueden destacar herramientas de gestión de procesos forenses mediante las cuales es posible definir flujos en donde se configuran tareas correspondientes a varias de las etapas ya conocidas en la disciplina, esto es, adquisición, preservación, análisis, documentación y presentación [15, 19]. Como ejemplo de este tipo de herramientas se destacan la solución Magnet Axiom Automate, la cual brinda facilidades para automatizar partes de los distintos flujos definidos en el laboratorio forense. La herramienta posibilita el uso de nodos de manera distribuida e integra los productos de la empresa que lo desarrolla en conjunto con otras herramientas externas relacionadas, como por ejemplo Autopsy o Encase [27]. Un producto similar es XEC Director de la empresa MSAB (desarrolladora XRY) [33], la cual facilita la orquestación y monitoreo de flujos y procesos dentro del laboratorio de IF que estén relacionados con los productos de la empresa.

Por otra parte, teniendo en cuenta la clasificación propuesta, es posible afirmar que aquellas funcionalidades catalogadas como avanzadas, en general son más decisivas a la hora de reducir el tiempo de procesamiento y análisis en el laboratorio forense. Funcionalidades como el *análisis automático sobre archivos multimedia* o la *extensibilidad mediante plugins* son significativamente más efectivas que varias de las funciones primarias dado que permiten automatizar tareas que llevadas a cabo de forma manual demandan tiempo y esfuerzo por parte del analista.

Es evidente que no se han descripto todas las funcionalidades existentes en el tipo de herramientas bajo estudio, en resumen, en esta sección se brinda un panorama de aquellas funcionalidades destacadas más frecuentes en los productos que son ampliamente utilizados a nivel global por gran parte de los laboratorios de IF [19, 16]. Tomando esto como base, también es posible identificar algunas limitaciones de las herramientas bajo análisis, las cuales se presentan como demandas constantes dentro de la disciplina. Como ejemplo puntual, se puede mencionar la transcripción automática de archivos de audio recuperados, algo sumamente necesario en el análisis de información forense de dispositivos móviles. También se puede señalar, en términos más generales, el uso de técnicas de Inteligencia Artificial para reforzar las distintas funcionalidades de las herramientas. En la actualidad dichas metodologías han sido ampliamente desarrolladas y existen muchas soluciones en el campo de IT que las incorporan. Sin embargo, es una de las grandes falencias que poseen actualmente las herramientas de IF de propósito general ya que sólo unas pocas están incursionando débilmente en este tipo de capacidades. Bajo este contexto, es importante remarcar que existen varios enfoques que deben empezar a ser considerados con más determinación por las empresas o grupos que desarrollan los productos de uso cotidiano en los laboratorios de IF.

6. Conclusiones y Trabajo Futuros

Con el paso del tiempo se puede observar que la Informática Forense se encuentra en todo momento sometida a constantes desafíos que van de la mano del crecimiento abrumador de la tecnología y las comunicaciones. En este contexto, uno de los principales retos de la disciplina está relacionado con la gran sobrecarga de trabajo que poseen los laboratorios de IF producto de la cantidad de dispositivos y consecuentemente del gran volumen de información que se deben procesar y analizar a diario. Sobre este problema se viene trabajando hace tiempo y se han presentado soluciones que intentan atenuar desde distintas aristas al problema principal, como por ejemplo el uso de *triage* para realizar una selección a priori de dispositivos a peritar o también el uso de acceso remoto de dispositivos que evita la extracción masiva de información. Si bien este tipo de técnicas contribuyen a reducir un porcentaje del trabajo, en la actualidad el problema persiste ya que la cantidad de dispositivos y volumen de información que termina en el proceso de IF dentro del laboratorio es cada vez más abrumadora. Es por esto que, las herramientas utilizadas dentro de este flujo de trabajo siguen teniendo un rol preponderante para con el problema analizado.

En este trabajo se presenta una descripción general de las funcionalidades que ofrecen las herramientas que son utilizadas a diario dentro de los laboratorios de IF, específicamente aquellas que reducen la sobrecarga de procesamiento y análisis de evidencia digital que se registra a nivel global en la disciplina. Estas funcionalidades se describieron y clasificaron en dos categorías: primarias y avanzadas. Las primeras se podrían considerar como esenciales en toda herramienta de esta clase debido a que son ampliamente utilizadas en el ámbito y desde hace varios años están presentes

en la mayoría de los productos disponibles. Las segundas son aquellas que han sido incorporadas recientemente y representan desarrollos más complejos y novedosos que las anteriores. Dichas funcionalidades no son ofrecidas en todos las herramientas de IF de propósito general, sino en aquellas soluciones que se podrían considerar más avanzadas dentro del entorno. Claramente, las funcionalidades con bases en métodos inteligentes aplicadas al flujo de trabajo tienen la capacidad de mejorar aún más las calidad en las investigaciones, no solo con respecto al tiempo sino también sobre la precisión en los resultados.

Como líneas de trabajo futuro, se intentará llevar a cabo un análisis de ciertas propuestas desarrolladas en el ambiente científico-académico que describen métodos, técnicas y herramientas las cuales se presentan como funcionalidades de interés ya que pueden contribuir en gran medida a solucionar el problema de la sobrecarga de trabajo en IF. Éstas se plantean como grandes demandas dentro de ámbito del analista forense para poder automatizar determinadas tareas y reducir los tiempos de análisis de los casos. Muchas de estas funcionalidades son aproximaciones que no se encuentra en las herramientas bajo estudio en este trabajo.

El aspecto más relevante que influye directamente sobre el futuro de la Informática Forense, está fuertemente determinado por la rapidez con la que evoluciona la tecnología y las comunicaciones. La única forma de contrarrestar este ritmo vertiginoso es mediante el desarrollo de herramientas que implementen funcionalidades que estén a la vanguardia en informática.

Referencias

[1] M. Abulaish and N. A. H. Haldar. Advances in digital forensics frameworks and tools: A comparative insight and ranking. In *Cyber Warfare and Terrorism: Concepts, Methodologies, Tools, and Applications*, pages 165–191. IGI Global, 2020.

[2] N. Alherbawi, Z. Shukur, and R. Sulaiman. A survey on data carving in digital forensic. *Asian Journal of Information Technology*, 15(24):5137–5144, 2016.

[3] D. Ayers. A second generation computer forensic analysis system. *digital investigation*, 6:S34–S42, 2009.

[4] N. Beebe. Digital forensic research: The good, the bad and the unaddressed. In *IFIP International conference on digital forensics*, pages 17–36. Springer, 2009.

[5] M. Botacin, V. H. G. Moia, F. Ceschin, M. A. A. Henriques, and A. Grégio. Understanding uses and misuses of similarity hashing functions for malware detection and family clustering in actual scenarios. *Forensic Science International: Digital Investigation*, 38:301220, 2021.

[6] S. Card. *Information visualization*. CRC press, 2009.

[7] B. Carrier. The sleuthkit and autopsy. *Retrieved May*, 8:2013, 2013.

[8] E. Casey, M. Ferraro, and L. Nguyen. Investigation delayed is justice denied: proposals for expediting forensic examinations of digital evidence. *Journal of forensic sciences*, 54(6):1353–1364, 2009.

[9] L. Cellebrite. Ufed mobile forensics applications, 2015.

[10] Y. Chabot, A. Bertaux, T. Kechadi, and C. Nicolle. Event reconstruction: A state of the art. *Handbook of Research on Digital Crime, Cyberspace Security, and Information Assurance*, pages 231–245, 2015.

[11] D. Chang, M. Ghosh, S. K. Sanadhya, M. Singh, and D. R. White. Fbhash: A new similarity hashing scheme for digital forensics. *Digital Investigation*, 29:S113–S123, 2019.

[12] S. Chavhan and S. Nirkhi. Visualization techniques for digital forensics: A survey. *International Journal of Advanced Computer Research*, 2(4):74, 2012.

[13] L. Chen, N.-A. Le-Khac, S. Schlepphorst, and L. Xu. Cloud forensics: model, challenges, and approaches. *Security, Privacy, and Digital Forensics in the Cloud*, pages 201–216, 2019.

[14] T. Coughlin. Digital storage in smartphones and wearables [the art of storage]. *IEEE Consumer Electronics Magazine*, 7(2):108–120, 2018.

[15] A. H. Di Iorio, H. Curti, F. Greco, A. Podestá, M. Castellote, J. Iturriaga, S. Trigo, B. Constanzo, G. Ruiz de Angeli, and S. Lamperti. Construyendo una guía integral de informática forense. 2015.

[16] M. Dweikat, D. Eleyan, and A. Eleyan. Digital forensic tools used in analyzing cybercrime. *Journal of University of Shanghai for Science and Technology*, 23(3), 2021.

[17] V. Fernando. Cyber forensics tools: A review on mechanism and emerging challenges. In *2021 11th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, pages 1–7, 2021.

[18] S. L. Garfinkel. Digital forensics research: The next 10 years. *Digital Investigation*, 7:S64–S73, 2010. The Proceedings of the Tenth Annual DFRWS Conference.

[19] K. Ghazinour, D. M. Vakharia, K. C. Kannaji, and R. Satyakumar. A study on digital forensic tools. In *2017 IEEE international conference on power, control, signals and instrumentation engineering (ICPCSI)*, pages 3136–3142. IEEE, 2017.

[20] C. Grajeda, L. Sanchez, I. Baggili, D. Clark, and F. Breitinger. Experience constructing the artifact genome project (agp): managing the domain's knowledge one artifact at a time. *Digital Investigation*, 26:S47–S58, 2018.

[21] J. Gregorio, B. Alarcos, and A. Gardel. Forensic analysis of nucleus rtos on mtk smartwatches. *Digital Investigation*, 29:55–66, 2019.

[22] J. Hansen, K. Porter, A. Shalaginov, and K. Franke. Comparing open source search engine functionality, efficiency and effectiveness with respect to digital forensic search. *Norsk Informasjonssikkerhetskonferanse (NISK)*, 2018.

[23] V. S. Harichandran, D. Walnycky, I. Baggili, and F. Breitinger. Cufa: A more formal definition for digital forensic artifacts. *Digital Investigation*, 18:S125–S137, 2016.

[24] H. Henseler. Finding digital evidence in mobile devices. *DFRWS*, 2017.

[25] G. Horsman. The different types of reports produced in digital forensic investigations. *Science & Justice*, 2021.

[26] B. Inglot, L. Liu, and N. Antonopoulos. A framework for enhanced timeline analysis in digital forensics. In *2012 IEEE International Conference on Green Computing and Communications*, pages 253–256. IEEE, 2012.

[27] A. Jarrett and K.-K. R. Choo. The impact of automation and artificial intelligence on digital forensics. *Wiley Interdisciplinary Reviews: Forensic Science*, page e1418, 2021.

[28] J. A. Kloess, J. Woodhams, H. Whittle, T. Grant, and C. E. Hamilton-Giachritsis. The challenges of identifying and classifying child sexual abuse material. *Sexual Abuse*, 31(2):173–196, 2019.

[29] J. A. Lapso, G. L. Peterson, and J. S. Okolica. Whitelisting system state in windows forensic memory visualizations. *Digital Investigation*, 20:2–15, 2017.

[30] H.-E. Lee, T. Ermakova, V. Ververis, and B. Fabian. Detecting child sexual abuse material: A comprehensive survey. *Forensic Science International: Digital Investigation*, 34:301022, 2020.

[31] L. Liebler, P. Schmitt, H. Baier, and F. Breitinger. On efficiency of artifact lookup strategies in digital forensics. *Digital Investigation*, 28:S116–S125, 2019.

[32] MagnetForensics. Artifact exchange opens today for sharing custom artifacts, 2017. Accessed = Mayo 2021.

[33] MSAB. Xec director, 2021. Accessed = Mayo 2021.

[34] M. S. Nadeem, V. N. Franqueira, and X. Zhai. Privacy verification of photodna based on machine learning. The Institution of Engineering and Technology (IET), 2019.

[35] N. N. I. of Standards and Technology. National software reference library (nsrl), 2021. Accessed = Mayo 2021.

[36] P. Purnaye and V. Kulkarni. A comprehensive study of cloud forensics. *Archives of Computational Methods in Engineering*, pages 1–14, 2021.

[37] D. Quick and K.-K. R. Choo. Impacts of increasing volume of digital forensic data: A survey and future research challenges. *Digital Investigation*, 11(4):273–294, 2014.

[38] S. Raghavan and S. Raghavan. A study of forensic & analysis tools. In *2013 8th International Workshop on Systematic Approaches to Digital Forensics Engineering (SADFE)*, pages 1–5. IEEE, 2013.

[39] H. Riggs, S. Tufail, I. Parvez, and A. Sarwat. Survey of solid state drives, characteristics, technology, and applications. In *2020 SoutheastCon*, pages 1–6. IEEE, 2020.

[40] M. Rogers. Technology and digital forensics. *The Routledge Handbook of Technology, Crime and Justice, Routledge, Oxon*, pages 406–416, 2017.

[41] B. A. Sabernick III. Development of an autopsy forensics module for cortana artifacts analysis. *International Journal of Computer Science and Information Security*, 14(7):111, 2016.

[42] M. SALT. Nuevos desafíos de la evidencia digital. *Acceso transfronterizo y técnicas de acceso remoto a datos informáticos. Editorial: Ad-Hoc, Argentina*, 2017.

[43] M. Sjöstrand. Combatting the data volume issue in digital forensics: A structured literature review. 2020.

[44] S. Soltani and S. A. H. Seno. A survey on digital evidence collection and analysis. In *2017 7th International Conference on Computer and Knowledge Engineering (ICCKE)*, pages 247–253. IEEE, 2017.

[45] C. Ware. *Information visualization: perception for design*. Morgan Kaufmann, 2019.

[46] G. K. Webb. Price and performance trends for cellular trail cameras explained with a time trend, google keyword trends, and a use case of suburban deer management. *Issues in Information Systems*, 21(2), 2020.

[47] C. Winter, M. Steinebach, and Y. Yannikos. Fast indexing strategies for robust image hashes. *Digital Investigation*, 11:S27–S35, 2014.

[48] T. Wu, F. Breitinger, and S. O'Shaughnessy. Digital forensic tools: Recent advances and enhancing the status quo. *Forensic Science International: Digital Investigation*, 34:300999, 2020.