A Survey on Digital Forensics Trends

Mohsen Damshenas, Ali Dehghantanha, Ramlan Mahmoud

Faculty of Computer Science and Information Technology, University Putra Malaysia

Corresponding Author

moansen@zhoeinix.com, {alid, ramlan}@upm.edu.my

1. ABSTRACT

Digital forensic has evolved from addressing minor computer crimes to investigating complex international cases with massive effect on the world. This paper studies the evolution of the digital forensic; its origins, its current position and its future directions. This paper sets the scene with exploring past literature on digital forensic approaches followed by the assessment and analysis of current state of art in both industrial and academic digital forensics research. The obtained results are compared and analyzed to provide a comprehensive view of the current digital forensics landscape. Furthermore, this paper highlights critical digital forensic issues that are being overlooked and not being addressed as deserved. The paper finally concludes with offering future research directions in this area.

Keywords

Digital Forensic, Mobile Device Forensic, Forensic Framework, Network Forensic, String Analysis

2. INTRODUCTION

The term computer crime, first used in 1976 in a book by Donn Parker titled “crime in computer” [1],[2] stepped into the legal system by Florida Computer Crimes Act 1978 dealing with unauthorized deletion or modification of data in a computer system [3]. However, the first actual computer analysis and response team was established by FBI in 1984 to conduct advanced digital forensic investigation of the crime scenes [4].

One of the first complicated digital forensic investigation cases was performed in 1986, pursuing a hacker named Markus Hess [5]. Hess had gained unauthorized access to Lawrence Berkeley National Laboratory (LBL) and was detected and investigated by Mr. Clifford Stoll. At the time of the incident, there was not any standard digital forensic investigation framework in place so Clifford had to do the investigation on his own. As Clifford’s objective was discovering the identity of the hacker, he did not change anything in the system and only collect the possible traces. By tracking the hacker for months using so called alarms which send notification when the attacker was active, he finally managed to discover the identity and location of the attacker by cooperation with the FBI and the Telco Company. Since the case was involved different military, academic and individual bodies in U.S and Germany the jurisdiction of the case became a big issue [6].

The day by day improvement of digital devices makes digital crime way more complicated than it was back in 1986. Nowadays crimes are happening over cloud which mandates cross national forensic investigation. It is therefore an essential demand for the security experts to realize their strengths in investigation of complex digital crimes via studying the history and current trends in the field. Specialists need to understand that digital forensics is not about looking at the past because of having an attack history; neither looking at the present in fear of being attacked; nor about looking at the future with uncertainty about what might befall us but about to being ready all the times for the moving target.

This survey offers a critical review and investigation over:

- Origins of digital forensics
- its evolvement to its current the current position
- And its future trends and directions

Vividly, using a map without knowing the current position is difficult; realizing the future can only be possible if the current and the past are clear. Hence, section 2 of this paper rectifies history of events and researches in the field in...
the period of 2002 to 2007 while section 3 concentrates only on recent research works in the period of 2008 and 2013 and offers categorization of major fields and a statistical comparison between industrial and academic in the field. Finally, we look into overlooked issues and offer several future research directions in the field.

3. DIGITAL FORENSICS: THEN
The formal beginning of academic community research in the area of digital forensic investigation was in 2002 with an article called “Network Forensics Analysis” authored by Corey et al. [7] who studied Network Forensic Analysis Tool (NFAT) and highlighted its benefits in regard to traffic capture, traffic analysis and security issues.

In 2004, Stevens [8] illuminated the issue of comparing and correlating time stamps between different time sources and proposed a clock model to address these timing issues by simulating the behavior of each independent time stamp (A.K.A independent clock). This model can be used to remove the expected clock errors from time stamps and develop a more accurate timeline analysis of the events. Forte [9] studied tools and techniques which were widely used for digital forensic investigation namely “GLIMPSE” and “GREP” and explained their syntax and applications. GREP as one of the most popular text search tool was covered comprehensively in this article. Carrier and Grand [10] studied the essential requirements for volatile memory acquisition and proposed a hardware based method to acquire memory data with the least possible changes. This method employed a hardware expansion card (PCI slot card) to create a forensic image of volatile memory with the push of a button. However, this technique required pre-installation of the card on the victim machine. Mocas [11] identified basic properties and abstractions of digital forensic investigation process and proposed a comprehensive framework for unifying all characteristics of digital forensic. Vaughan [12] presented a methodology for evaluating the evidential value of an Xbox game console system. Moreover, he proposed a methodology for evidence extraction and examination from a suspect Xbox system.

Nikkel [13] outlined the forensic investigation analysis of IP networks and domain names. This article defined point of concerns of a presence which automatically collects evidences related to the Internet presences, time-stamped these evidences, store the evidences in a neat manner, generate the integrity hash checksum of the evidence and finally produced an official report of the discovered information. Buchholz and Spafford [14] studied the effects of metadata on digital forensics to find out which information can be useful in a computer forensic investigation. Moreover, they demonstrated potentials of metadata in computer forensic investigation and analyzed issues in obtaining and storing these data.

In 2005, Francia and Clinton [15] outlined the required resources and procedures to establish a forensic lab and presented a cost efficient Computer Security and Forensic Analysis (CSFA) experimental lab design and implementation. Nikkel [16] discussed forensic investigation challenges in file and contents recovery of magnetic tape data acquisition and analysis and proposed a methodology for determination of tapes content. Jansen and Ayers [17] provided an overview of the available forensic investigation tools for Personal Digital Assistants (PDA) devices and PDAs software and hardware. They utilized some basic scenarios to evaluate those tools in a simulated environment using simulated evidentiary situations and offered a snapshot of how these tools work under provided situations. Bedford [18] explained the forensic investigation challenges of Host Protected Area (HPA) in IDE drives. The HPA is a part of a hard drive that is hidden from the operating system and the user which is often used to hide sensitive data and as such is a good source of evidence.

In 2006, Harrison [19] described a project for explaining real-life issues in digital forensic investigation and utilized a group-based project for practicing computer investigation in academic environments. Laurie [20] studied well-known Bluetooth security flaws and available techniques and tools to exploit those flaws; and then discussed the effects of these attacks on common digital forensic investigation practices. Nikkel [21] described concepts of distributed network-based evidences, forensic
techniques of evidence acquisition over the network and the weaknesses of these techniques. In addition, he suggested some improvements to conquer challenges in network data collection. Ieong [22] highlighted major principles of digital forensic investigation and outlined eight roles with their responsibilities in common digital investigation frameworks. The paper proposed a framework named FORZA which is based on Zachman’s framework [23]. Harris [24] studied and analyzed several anti-forensic techniques and classified them to achieve a standard method for addressing anti-forensics issues and then outlined general strategies to preserve forensic integrity during investigation. Garfinkel [25] proposed Forensic Feature Extraction (FFE) and Cross Drive Analysis (CDA) techniques to analyze large data sets of disk images and other types of forensic data. These two methods were used for prioritizing and systematically identifying social networks usage in the suspect’s system. Schuster [26] explored the structure of memory containing processes’ data and then proposed a search pattern to scan the whole memory dump for traces of physical memory objects independent of kernel’s list of memory objects. As a proof of concept, an implementation of the technique successfully revealed some hidden and terminated processes and threads even after rebooting the system. Jeyaraman and Atallah [27] proposed an approach to examine the accuracy and effectiveness of automated event reconstruction techniques based on their capabilities to identify relations between case entities. They then quantified the rate of false positives and false negatives and scalability in terms of both computational burden and memory-usage. Alink et. al., [28] introduced a novel XML based approach for storing and retrieving forensic traces derived from digital evidences which was implemented in a prototype system called XIRAF. This system runs the available forensic analysis tools against the evidence file and then exports the result in an XML database. Johnston and Reust [29] outlined procedures for investigation of an attack which includes compromising of several systems with sensitive data and discussed challenges involved in this process. Mead [30] studied the National Software Reference Library (NSRL) repository of known software, file profiles, and file signatures by examining the uniqueness of the signatures produced in NSLR repository. This repository is essentially vital for digital forensic investigators to improve the speed of data analysis by omitting the file analysis of trusted files according to the retrieved hash checksum from NSRL. Nikkel [31] presented a forensic evidence collector tool using promiscuous packet capture on an embedded hardware using open source software. This tool operates as a standalone tool in different modes and can be pre-configured by the investigator. In 2007, Wang et al. [32] analyzed methods and applications of cryptography in digital forensic investigation, and highlighted differences between these methods. Afterwards, the authors discussed the weaknesses of SHA-1 and approaches to crack SHA-1 in order to highlight the issue of potential clashes in checksum verification and possible effects on related applications. Peisert et al. [33] presented the importance of examining the order of function calls for forensic analysis and showed its usefulness in isolating the causes and effects of the attack through intrusion detection systems. This analysis, not only detects unexpected events in the order of function calls, but also detects absence of expected events. Castiglione et al. [34] investigated the issues of hidden metadata in compound documents that use opaque format and could be exploited by any third party. The authors proposed a steganography system for Microsoft Office documents and introduced FTA and StegOE as tools to improve the forensic analysis of Microsoft Office documents. Murphey [35] proposed a method to automatically recover, repair and analyze Windows NT5 (XP and 2003) events logs. Authors implemented a proof of concept code to repair common corruptions of multiple event logs in one simple step without any manual user intervention. Spruill and Pavan [36] studied the U3 technology for portable applications and illustrated different artifacts left behind from a committed crime through a portable application. This research also investigated some of the common applications used on U3 drives. Turner [37] questioned usefulness of the common digital forensic investigation approaches in live incident
investigation and demonstrated the applications of Digital Evidence Bag (DEB) storage format in dynamic environments. Richard et al. [38] explained Data Evidence Container (DEC) format to bundle arbitrary metadata associated with evidence along with the actual evidences. Authors then explored the challenges of utilizing this container and proposed Forensic Discovery Auditing Module (FDAM) as a complementary mechanism. Nikkel [39] demonstrated digital forensic investigation principles for IPv6 networks and discussed about IPv6 addressing, packet structure, supported protocols and information collection from public registrars such as WHOIS. They finally presented some IPv6 tools for collection and analysis of network traffic and investigation of remote IPv6 nodes. Masters and Turner [40] demonstrated techniques of magnetic swipe card data manipulation in different types of devices and proposed the application of Digital Evidence Bag (DEB) as a suitable format for bulking the evidences obtained from a magnetic swipe card. Lyle and Wozar [41] studied challenges in making a forensic image of a hard disk and addressed practical problems like resolving some faulty sectors causing difficulties in imaging process. Arasteh et al. [42] proposed a formal model checking technique for forensic analysis of system logs and provided a proof of concept system using tableau-based proof checking algorithm. Arasteh and Debbabi [43] presented a forensic analysis method to extract threads history by using threads stack. The technique used a process model to retrieve data from a thread stack and then compared against an assembly model to verify extracted properties. Schatz [44] proposed a technique for obtaining volatile memory from arbitrary operating systems and provide a “point in time” snapshot of the volatile memory. BODY Snatcher, an implementation of this method was used for presenting the proof of concept of the proposed technique. Barik et al. [45] studied the issues of Ex2 file system in terms of authenticity of the OS created timestamps and proposed a solution to preserve authentic date and time stamp for the studied issues using Loadable Kernel Modules (LKM).


Figure 1: number of articles in digital forensics in the period of 2002 – 2008
(No articles matched our search criteria in 2003)
All the studied articles are extracted by searching the term “forensic” in the abstract of 4 main computer science indexing services known as Science Direct, Springer, IEEE and ACM; though the search was limited only to articles published in computer science during the period of 2002 – 2008.

4. NOW
From the very first days of digital forensic investigation’s life up to now, there were vast improvements in digital forensics techniques ranging from recovering deleted evidences and searching the megabytes storage devices to deal with petabytes storage devices [46], cloud based investigations [47], mobile device examinations [48], wireless network investigations [49], and database forensics [50]. Generally, the current perspective of digital forensic investigation can be categorized into four main types namely Computer Forensic, Smart Device Forensic, Network Forensic and Database Forensic. Among mentioned categories, computer forensics has attracted the most attention of academicians and professionals. On the other hand, digital criminals and intruders are trying to minimize footprints of their actions utilizing anti-forensic techniques. Some of the common approaches of anti-forensics are using cryptography [51], steganography [52], meta-

Table 1 – Final result of participated Journals with number of related papers in each

<table>
<thead>
<tr>
<th>Journal Name</th>
<th>Number of Related Papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital Investigation</td>
<td>117</td>
</tr>
<tr>
<td>IEEE Transactions on Information Forensics and Security</td>
<td>20</td>
</tr>
<tr>
<td>Computers &amp; Security</td>
<td>9</td>
</tr>
<tr>
<td>Computer Law &amp; Security Review</td>
<td>4</td>
</tr>
<tr>
<td>Multimedia Tools and Applications</td>
<td>3</td>
</tr>
<tr>
<td>ACM Computing Surveys</td>
<td>2</td>
</tr>
<tr>
<td>Computer Standards &amp; Interfaces</td>
<td>2</td>
</tr>
<tr>
<td>Computers &amp; Mathematics with Applications</td>
<td>2</td>
</tr>
<tr>
<td>Information Sciences</td>
<td>2</td>
</tr>
<tr>
<td>Personal and Ubiquitous Computing</td>
<td>2</td>
</tr>
<tr>
<td>IEEE Journal on Selected Areas in Communications</td>
<td>2</td>
</tr>
</tbody>
</table>
Figure 2 – Number of forensics relevant papers from Jan 2008 to Mar 2013

Figure 2 offers a ratio of the position; “Digital Investigation” journal (indexed in “Science Direct”) with 117 published articles (66% of total articles in the studied period) has the highest number of published articles followed by “IEEE Transactions on Information Forensics and Security” journal (indexed in “IEEE”) with 20 published articles (12% of total articles in the studied period) and “Computers & Security” journal (indexed in “Science Direct”) with 9 articles (5% of total articles in the studied period).

Not being an exception, this research faced some limitations. The main limitation of this study was to include all perspectives of digital forensic using trust worthy sources. The most important selection factor for us was creditability of our sources so we can provide a trustable review of current forensics landscape. As such, we just selected papers from reputable sources which indirectly implicates that we ignored possibly relevant papers which were not published in reputable sources. Selecting journals was not only based on the mere number of articles they published (2 or more) in a specific period (2008-2013), but to make sure only papers of very relevant journals which are reflecting realistic view of the field are selected. Moreover, to improve the reliability of the given perspective in addition to reputable academic sources some articles from the SANS reading room were included as the representative of the industrial research.

4.1. Results Obtained
The data collection process began with brainstorming among authors to identify major types of digital forensic investigation to classify and rank research papers appropriately. For example, all topics related to smartphone investigation, iPad data acquisition, GPS data analysis and similar subjects were categorized under mobile forensics. ISO9660 analysis, meta-data analysis, cluster and all investigations relevant to the science of file systems grouped as file system forensics. The same approach was taken for all other categories as well as shown in Figure 3. Finally, papers in topics like digital audio forensics, malware forensics [57], database forensic and cloud forensic which were not fall in previously mentioned categories and there were 5 or lesser papers in those topics were grouped as “Others” as shown in Figure 4.

Identification of general categories was more difficult than it was expected as there were difficulties in categorizing inter-disciplinary papers or papers with wide range of focus. For example, papers under collecting volatile
memory of mobile phones could be categorized in both volatile memory acquisition and in mobile phones investigation category. Obviously, we could not count a paper multiple times and in such cases, the team carefully read the paper thoroughly to find out the major focus of the paper with specific attention to the paper abstract, keywords and conclusion. Afterwards, we categorized the paper in the most relevant category. It is notable that, the classification of this study is exclusively the opinion of the authors and although all relevant scientific and statistical techniques were employed to ensure correctness and comprehensiveness of the study but all conclusions are subjective to author’s view-points only. Moreover, there were several publications contain news section or discussion that cannot be qualified to be called full article. Each of these papers were analyzed carefully and included in this survey if they offer valuable knowledge or results about the current state of the art in digital forensic.

When investigating each of the journals separately, it is interesting to note different topics emphasized in each journal. Table 2 lists the number of publications in each journal in each topic. Outstanding result indicates that the “Digital Image Forensic” topic took the lead with 20 articles, followed by “Mobile Device Forensic” with 14 articles, “forensic framework” at 12, “forensic tools” and “file system forensic” both at 9 that constitute the top six topics. In “IEEE Transactions on Information Forensics and Security”, articles on anti-forensics were more than all other categories (3 articles), followed by “digital image forensic”, “forensic tools” and “string analysis” with only 1.

![Figure 3 – Digital forensic investigation categories in period of 2008 - 2013](image-url)
Figure 4 – Digital forensic investigation topics with less than 5 papers (“Other” category)

Table 2 – List of journals and number of published articles grouped by the category

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti Forensic</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application Forensic</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital Image Forensic</td>
<td>20</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>File System Forensic</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forensic Framework</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forensic Tools</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile Device Forensic</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network Forensic</td>
<td>5</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>String Analysis</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volatile Memory Forensic</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.1.1. **Topics covered by SANS**
This subsection of the study analyzes articles published in the SANS reading room as a well-established organization known as a pioneer in non-academic digital forensic investigation research. SANS papers are considered as a reliable source for studying industrial digital forensic research trends. However, the quantity of SANS forensics articles is just a few in compare to the academic journals as publication is not included in the main duties of the forensics practitioners.

We have identified 24 papers in SANS reading room and categorize them similar to the academic papers. The outcome includes “File Forensic”, “File System investigation”, “Forensic Framework Development”, “Forensic Tools Investigation”, “Legal Aspects of Digital Forensics”, “Mobile Device Investigation”, “Operation System Investigation”, “Database Investigation”, “Network Investigation”, “Steganography” and “String Analysis” categories as shown in Figure 5.

![Figure 5 – Categories of SANS articles](image)

The further analysis of the SANS articles indicates the differences between academic and industrial research trends. The results show that the share of forensic framework, network forensic and forensic tools from overall SANS articles is %21, %17 and %17; while they have only %8, %5 and %8 of academic journals. On the other hand, in academic environment, digital image forensic, mobile device forensic and forensic tools were the leading topics. Figure 6 is the result of comparing the common topics between SANS and academic journals while the SANS topics lacks some of the topics covered in academic environments. Figure 7 clearly indicate the results by providing the share of each topic in SANS and academic journals.
Figure 6 – Comparison of common categories in academic and SANS publications in term of number of articles

Figure 7 – SANS vs. Academic Journals (the “Other” in academic journals indicates the topics which were not covered in SANS articles)
4.2. Discussion and Analysis

In this section, we offer detailed analysis of each of 10 most categories identified in the previous sections namely digital image forensic, mobile device forensic, forensic tools, volatile memory forensics, network forensics, anti-forensic, data recovery, application forensics, file system forensics and forensic frameworks. We critically analyze trends and current state of the art in each category and rectify possible future trends.

4.2.1. Digital Image Forensics

Since the wild increase in usage of digital images, the crimes involving digital images such as image forgery are increased consequently and detection of the image tampering brought many difficulties to forensic investigators and forensic researchers. The major detectable research topics in digital image forensics are image authenticity, image correction, steganography, image processing and source detection. Image authenticity as the most popular topic in Digital Image Forensic category concerns about the reliability of the evidence. Amerini et al. [58] proposed a SIFT based algorithm to detect multiple copy-move attacks on an image. This technique extracts and matches the image features to identify similar local region on the image and then makes a hierarchical cluster of the extracted features to identify the cloned areas. In case the image is classified as unauthentic, the geometrical transformation will be identified to discover the original area and the copy-moved area. Gou et al. [59] introduced a source identification technique with the ability of recognizing various post-processing operations on scanned image. Utilizing the noise analyses through wavelet analysis, neighborhood detection and image de-noising, made this approach capable of detecting the model of scanner used to scan the image and type of the image source (scanner, digital camera or computer generated). Chen et al. [60] introduced a source digital camera identification technique with integrity verification capability based on Photo Response Non-uniformity Noise (PRNU) imaging sensor fingerprint. The PRNU is generated through the maximum likelihood principle derived from normalized model of the sensor output and then compare to a pre-generated experimental dataset. Yuan [61] introduced a novel method for detection of median filtering in digital images. The key points of this method were the capability of median filtering detection for low resolution, JPEG compressed images, and tampering detection in case a median-filtered part is inserted in a non-median-filtered image and vice versa. Mahdian and Saic [62] studied the addition of locally random noise to tampered image regions for anti-forensic purposes and introduced a segmentation technique for dividing the digital image into various partitions based on homogeneous noise levels. Their novel approach utilized tiling the high pass wavelet coefficients at the highest resolution with non-overlapping blocks, to estimate the local noise level and median based method to estimate the standard noise level of the image. Farid and Bravo [63] introduced a novel methodology for computer aided differentiation of photorealistic computer generated images and photographic image of people using images with different resolution, JPEG compression, and color mixture of the image. Kornblum [64] studied the quantization tables in JPEG compression and explained how quantization tables can be used for differentiating between images processed by software and intact images. Authors utilized other factors such as the presence or absence of EXIF data, signatures of known programs, and color signatures of real skin to increase the success rate of the detections. Mahalakshmi et al. [65] proposed an approach for detection of image manipulations through available interpolation related spectral signature method. This method could detect common forgeries like re-sampling (rotation, rescaling), contrast enhancement and histogram equalization. Moreover, a set of techniques were introduced for detection of global and local contrast enhancement and histogram equalization. Source detection explicitly tries to identify the source device with which the image is taken or edited. Tsai et al. [66], the author employed support vector machines and decision fusion to identify the camera source model of an image. The evaluation result of the proposed model indicates 91.66% accuracy for images take by 26 cameras. Choi et al. [67] proposed a new approach for color laser printer identification of
an unknown printed image, based on the noisy texture analysis and support vector machine. This method identifies the invisible noises of the schema according to wiener-filter and the 2D Discrete Wavelet Transform (DWT) filter and then the texture of the noise is analyzed via gray level co-occurrence matrix (GLCM).

**Image processing** involves the analysis of the image to find specific patterns for variety of purposes. Islam et al. [68] proposed a framework for detection of children skin in images which utilized a novel vision model based on the Markov Random Fields (MRF) prior by employing a skin model and human affine-invariant geometric descriptor to identify skin regions containing pornographic. Steel and Lu [69] proposed a system called Automated Impersonator Image Identification System (AIIIS), which tracks the used image in impersonation attack back to the original source by employing digital watermarking technique. This technique encoded access details of the image (like IP address, server and the download date-time) and download them to the image. In [70], the author utilized demosaicing artifacts to identify the digital camera model. The process includes estimating demosaicing parameters, extracting periodicity features of the image for the purpose of detecting simple forms of demosaicing and finally defining a set of image characteristics as features to be used for designing classifiers that distinguish between digital cameras.

**Image correction** as another hot topic in image forensic, mainly intends to ease the analysis of the image. Tang et al. [71] developed a Knowledge Based (KB) approach to remove JPEG blocking artifacts in skin images. The approach utilized a Markov model based algorithm and a one-pass algorithm to implement the inference; plus a block synthesis algorithm for handling the cases with no prior record in dataset. Moreover, in order to reduce the search time in the dataset, the author proposed a multi-dimension indexing algorithm. Lin et al. [72] introduced a method for detecting the image source encoder alongside the estimation of all coding parameters based on the intrinsic fingerprints of the image source encoder. The evaluation of this approach indicated the success rate of more than 90% for transform-based encoders, sub-band encoders, and DPCM encoders when PSNR is higher than 36dB.

**Steganography** as one of anti-forensic techniques in digital images plays a significant role for identification of the evidence. Huang [73] provided a solution for detection of double JPEG compression using the “proper” randomly perturbed ratio. This approach is highly dependent on finding the correct ratio; thus a novel random perturbation strategy is utilized on the JPEG coefficients of the recompressed image. Kirchner and Bohme [74] challenged the current image tampering detection techniques by presenting types of image transformation operation that cannot be detected using the available resampling detection tools. Among these attacks, resampling with edge-modulated geometric distortion and the dual-path approach are nearly impossible to be detected.

### 4.2.2. Mobile Device Forensics

Due to the wild incline in usage of mobile devices such as smart phones, tablets and GPS devices, investigation of such devices is significantly vital; thus obtaining and analyzing the evidence from mobile devices has a great value. In continue Mobile live memory, Mobile forensic framework, Mobile forensic tools and Mobile on-scene triage topics are discussed.

**Mobile live memory** concerns about imaging and analyzing mobile devices volatile memory. In [75] the authors presented a novel approach to obtain forensic image of an Apple iPad device through iPad camera connection kit and a USB storage device. The results of the evaluation indicated that this method can obtain image up to 30 times faster than acquiring image through the usual method utilizing Wi-Fi channel. Thing et al. [76] proposed a system for real-time digital forensic analysis of mobile memory with focusing on dynamic properties of memory. The system evaluation consists of investigation of different communication variables (i.e. message length, message interval, dump interval, key-press interval) which resulted in 95.6% and 97.8% for dump intervals of 40 and 60 seconds. Sylve et al. [77] illustrated the process of live memory collection of the android devices, the new memory dump module (named DMD or LiME (Linux Memory Extractor)) and the challenges of device independent memory
acquisition; it then proposed a memory collection method which can dump memory to SD card or network storage. The proposed method operates via “rooting” the android device with methods like “Rage against the Cage” and other privilege escalation techniques. Depends on the type of device, utilizing the right mobile forensic tool always plays a noteworthy role. In [78] authors enriched their previously published methodology for smartphone evidence acquisition (Mobile Internal Acquisition Tool) in term of improvements and assessments. The MIAT can be executed from a memory card like MMC and explores recursively the file system tree and copy each entry to a backup volume. The behavior of MIAT is also evaluated by comparing to Paraben Device Seizure as a well-known tool. Vidas et al. [79] illustrated the digital forensics collection method of Android device via a modified boot loader from recovery booting. The advantage of this method is mainly because it only slightly changes the recovery partition and no user or system partition is affected during the collection stage of the investigation unlike normal approaches which “root” the android device for image acquisition.

**Mobile forensic framework** discusses different procedures and frameworks for mobile device forensics. In [80], the authors investigated the use of mobile phone flash boxes in a forensic framework and proposed a validation procedure to ensure the integrity of the acquired evidence in this new method. Unfortunately using mobile phone flash boxes do not provide forensically sound evidence; even though this proposed method increased the percentage of the reliability of the evidence, it still lacks the sufficient evidence integrity proof and demand to be lifted up by further research. Owen and Thomas [81] studied the available framework for mobile forensic investigations and highlighted the lacks and strength of each in comparison with the common digital forensic investigation practices of hard disk drives. Grispos et al. [82] compared the available mobile forensic toolkit for discovering to what extent these tools can operate and how much they are reliable. The result of this comparison indicated the limitations of using file carvers for information recovery, conflicts of diverse methods of information recovery, several differences between the documented capabilities of the CUFED and its actual performance, and the fragility of existing mobile forensic toolkits when recovering data from partially corrupted file system images.

**Mobile on-scene triage** determines the priority of evidence collection in mobile devices. In [48], the author presented a methodology for obtaining and analyzing evidences in webOS system partition; the article provides solution for obtaining different types of evidence (i.e. Call logs, contacts, calendar events and etc.). The article also provided approaches for recovering deleted files from a webOS operating system. Eijk and Roeloffs [83] discussed the challenges of obtaining evidence from TomTom GPS device and approaches to acquire volatile memory of the device via either JTAG signals or loading a small Linux distribution to the GPS memory. The article described the structure of data on GPS device and technique for analyzing the obtained data. In [84] the author studied the available framework for forensic investigation of windows mobile equipped mobile phones and showed the similarities of windows mobile investigation with normal windows investigation on PC systems. This work demonstrated that Windows mobile phone investigation obeys the common digital forensic investigation framework and the analysis of collected data is very much similar to windows operating system. Mislan et al. [85] studied the on-scene triage of mobile devices and compare its requirements with common digital forensic investigation requirements; it formalized the on-scene triage process and provides guidelines for standardization of the process. At the end, it defined the basic requirements of an automated on-scene triage.

### 4.2.3. Forensic Tools

Employing different tools in the process of digital forensic is inevitable for the investigator, yet choosing the most suitable tool can affect the result of the investigation. The identified research topics of this category would be as forensic formats, new forensic tools, and forensic tools examination.

This topic, new forensic tools, reviews novel tools discussed in forensic articles. In [86] authors presented a forensic investigation tool
named Cyber Forensic Time Lab which help the investigator to sort all evidences according to their time variables for generating Temporal analysis of the crime. Klaver [87] introduced the forensic application of available tools on Windows CE (Windows Mobile) after studying the typical hardware information and software components. The author explained the usage of current tools for forensic investigation of Windows CE mobile phones. Joyce et al. [88] discussed the Mac OS x forensic investigation approaches and introduced MEGA, a comprehensive tool for analyzing Mac OS x files from an image. MEGA offers ease of access to Spotlight metadata, content search and FileVault encrypted home directories. In [89], authors described their novel techniques to develop a memory analysis tool for variety of operating systems. Inoue et al. [90] provided a new tool for volatile memory imaging of the Mac OS x and then compared it using four metrics of completeness, correctness, speed and interference. Moreover, a visualization method called the density plot is introduced for indicating the density of repeated pages in an image.

There is always a demand for testing approaches to examine the performance and accuracy of the tools. In [91] the first generation of computer forensic tools is challenged by the authors as they have limitations and flaws in processing speed, data design, auditability, task analysis, automation and data abstraction; and then the second generation tools requirements is discussed and some available tools is suggested for handling each deficiency. Pan and Batten [92] presented a performance testing approach for digital forensic tools offering testing with good quality via a limited number of observations. The proposed method is specially designed for forensic tool testing and claimed of being the best available software among forensic tool performance testers.

Choosing the forensic format of the evidence has an important role in the forensic investigation process; in continue four novel forensic formats are proposed by different authors. In [46], the author introduced a redesign of the Advanced Forensic Format (AFF) based on the ZIP file format specification. This file format support full compatibility with previous versions of AFF while it offers the capability of storing multiple type of evidence from various devices in one archive and an improved separation between the underlying storage mechanism and forensic software. Garfinkel [93] studied the Digital Forensic XML (DFXML) language and discussed motivations, design and utilization of this language in processing forensic investigation information. Levine and Liberatore [94] introduced DEX, an XML format for recording digital evidence provenance, which enable investigators to use the raw image file and reproduced the evidence by other tools with same functionality. Conti et al. [95] studied automated tools of mapping large binary objects like physical memory, disk image and hibernation files via classifying of regions using a multi-dimensional, information-theoretic technique.

4.2.4. Volatile Memory Forensics

Extracting potential evidences from volatile memories is another challenge for forensic investigators as the basic requirements of knowing the structure of memory is not satisfied. The detectable volatile memory forensic topics are data extraction from volatile memory, volatile memory mapping and forensic imaging of volatile memory.

Data extraction from volatile memory involves utilizing different tools and techniques to analyze and obtain available evidences. In [96], authors studied the tools and techniques of extracting Windows registry data directly from physical memory dump and then presented an attack against on-disk registry analysis techniques by modifying the cached registry values in physical memory. Maartmann-Moe et al. [51] proposed a novel technique for identification of cryptographic keys stored in volatile memory and support the method by a proof of concept, Interrogate, which can identify AES, Serpent and Twofish cryptography keys. Baar et al. [97] described a novel approach for identifying and recovering files mapped in physical memory to recognize the source of the data and the usage of them via three different algorithms (carving allocated file-mapping structures, unallocated file-mapping structures and unidentified file pages). Schuster [98] studied nonpaged pool area of the physical memory and its potential in containing massive
amount of information about the cunning and even closed processes. In continue the author demonstrated the Microsoft Windows operating system pool allocation technique. **Volatile memory mapping** is an essential research area as it is the primary requirement for volatile memory analysis and data extraction. In [99], the author described an algorithm for detecting paging structure of processes in physical memory dump; this method is based on hardware layer and works on both Windows and Linux with minor tweaking. Stevens and Casey [100] introduced a methodology for recognizing and extracting user command line history in Microsoft Windows operating system after studying the structure of command line data on physical memory dumps. At the end authors provided a proof of concept tool for Microsoft Windows XP. Hejazi et al. [101] introduced a unique technique for collecting case-sensitive information from extracted memory content based on analyzing the call stack and security sensitive APIs. The method is limited to Microsoft Windows XP (SP1, SP2). In [102], authors explained the debugging structures in physical memory and Microsoft Program’s Database (PDB) to extract configurations, network activities and processes information from any Windows NT family operating system. J. Okolica and Peterson [103] demonstrated the significance of the clipboard information in digital forensic investigation and described the structure and the procedure of retrieving copy/paste information from Microsoft Windows XP, Vista, and 7. This technique can obtain information from the software which the data have copied from (i.e. Notepad or WordPad). J. Okolica and Peterson [104] proposed a novel efficient methodology for reverse engineering the Windows drivers and dynamic link libraries. In addition, the authors have revealed the network connection information structure on physical memory which eases the analysis part of the investigation.

Even though there is a lack of scientific **forensic imaging** technique for volatile memories, there is only one research found in the scope of this survey. Rabaiotti and Hargreaves [105] presented a novel method for obtaining forensic image of an embedded device by exploiting a buffer overflow vulnerability and execute customized code for creating the image of the console memory. The current work only covered Microsoft Xbox gaming console while the idea is applicable to any type of embedded device.

### 4.2.5. Network Forensics

Obtaining evidences from network traffics is a great challenge for investigators mainly due to the live characteristic of network packets. Beverly et al. [106] showed that IP packets, Ethernet frames and other network related data is available in physical memory and these data can be retrieved from hibernation files or memory images. Additionally, the authors proposed the network carving algorithm, techniques and essential tools to identify and extract the information. Thomnard and Dacier [107] presented an analysis framework for extracting known attack patterns from massive amount of honeynet data sets. This method allows the analyst to select different feature vectors and appropriate similarity metrics for creating clusters. Shebaro and Crandall [108] described the privacy related challenges of network flow recording for different purposes and then introduced a network flow recording technique by utilizing identity based cryptography and privacy preserving semantics. Zhu [109] introduced an iterative algorithm to discover network based attack patterns via network traffic logs; this method used a unique feedback mechanism to propagate the chance of being under attack or suspicious score and then pass it to the next iteration without any dependency to human supervision (i.e. defining threshold).

Tracking the source of traffic in networks is another challenge when network address translation and other types of network services interfere in the communication. In [110], the author studied the forensic investigation challenges of Network Address Translation (NAT) service enabled networks and proposed a model and algorithm for discovering of observed traffic into the source behind the NAT; the model used specific correlation of NAT gateway artifacts left overs to identify the source. Takahashi et al. [49] focused on reviewing IEEE 802.11 wireless network characteristics based on the traffic pattern of nodes; in order to recognize user fingerprints, rogue access points and media access control protocol missuses. Hsu et al.
[111] proposed a forensic approach for identifying the voice over IP (VoIP) call source via network operators (NWO) and service providers (SvP) without depending on routers logs.

4.2.6. Anti-Forensic

The procedure of forensic investigation can be affected by the criminals via anti-forensic action; researchers challenge the integrity of forensic process by demonstrating new types of anti-forensic method. Evidence collection and common attacks are two identified topics in this area. Distefano et al. [112] analyzed anti-forensic attacks on mobile devices and presented some fully automated examples of these attacks. The authors then examined the effectiveness and strength of the implementations, by using them against cursory examination of the device and tools for obtaining the internal memory image. Sun et al. [113] proposed two anti-forensic steganography approaches based on exploiting modification direction (EMD) technique for storing and extracting data in image. The highlight of EMD (HoEMD) and the adaptive EMD (AdEMD) are high efficient and high quality methods introduced in this work. Stamm and Liu [114] introduced anti-forensic approaches for removing forensically important indicators of compression such as image compression history from an image. This technique utilized a novel approach to identify and omit the compression fingerprints of an image’s transform coefficients. Khan et al. [115] presented a novel approach to hide data in a deniable way (to use plausible deniability) by storing sensitive data on a cluster based filesystem; to do so, a covert channel encodes the data via altering the fragmentation pattern in the cluster distribution of the secret file.

Rekhis and Boudriga [116] studied the digital forensic investigation techniques of anti-forensic attacks and then characterized secure evidence, provable and non-provable attacks. As the main contribution of this research, the authors developed an anti-forensic attack aware forensic approach using a state-based logic. Casey et al. [117] explained the challenges of full disk encryption (FDE) for digital forensic investigators and provided a guidance to take necessary items from the crime scene, to ease the access to encrypted data or live forensic acquisition of the running system. The provided measures can help obtaining evidences in an unencrypted state or finding the encryption key or passphrase.

4.2.7. Data Recovery

As one of the most essential stage of evidence identification and collection, data recovery can be affected by different variables like file system type; below some of the data recovery techniques are reviewed. In [118], the author analyzed the main properties of the Firefox SQLite database and its use in forensic investigation. Moreover, a novel algorithm is proposed to recover deleted records from unallocated space based on the fact that Firefox utilizes temporary transaction files. Jang et al. (Jang et al., 2012) described the integrity issue of forensic image corruption after acquisition and provided a novel algorithm for recovering and protecting the evidence image using recovery blocks. The recovery process is applicable when data blocks (generate by recovery blocks) are damaged. Yoo et al. [119] highlighted the challenges of multimedia file carving and NTFS compressed file carving; and proposed solutions for these challenges. The solution for multimedia file carvings is based on the characteristics of AVI, WAV and MP3 files. In [120], the author discussed the structure of Windows registry data and the behavior of Windows in deleting the registry records. In continue a technique is developed for identifying the deleted Windows registry records and recovery of deleted keys, values and other attributes of records. Garfinkel et al. [121] studied the usage of cryptographic hashes of data blocks for finding data in sectors and introduced the concept of “distinct disk sector”. Moreover it provided some approaches for better detection of JPEG, MPEG and other compressed data files. Chivers and Hargreaves [122] explored the structure of Windows search database and the challenge of recovering the deleted records after the file is deleted; it then proposed a novel record carving approach for identifying and recovering the deleted database records from database unused space or filesystem. King and Vidas [123] compared the solid state disk (SSD) with normal HDD and proved the TRIM command can affect the investigation process. The analysis from 16
different disks shows that TRIM enabled disks has almost 0% recoverable data while the SSDs with disabled TRIM has the average of %100 success rate.

4.2.8. Application Forensics

The forensic investigation of applications is quite advantageous as these applications usually store specific evidences. Identifying and collecting these evidences demands for prior research on the application behavior. In [124], the authors studied Internet Download Manager (IDM) activities recorded and effects on different files such as log files, Windows registry and history from artifacts point of view. This study demonstrated approaches and told to detect different attributes of download requests like URL, download time and login credentials. Garfinkel (Garfinkel, 2012b) shared the experience of construction a Korean Reference Data Set (KRDS) based on National Software Reference Library RDS (NSRL RDS) and developed a model for both effective importing of NSRL data sets and adding Korean specific data sets. Lallie and Briggs [125] explored three well-known peer-to-peer network clients (BitTorrent, µTorrent and Vuze) and analyzed their artifacts on Windows registry using the effects created by installation and working with these clients.

In [126], the authors outlined the significance of web browsers in forensic investigation and proposed a methodology for evidence collection and analysis from web browsers log files. Lewthwaite and Smith [127] looked into the Limewire artifacts remained in Windows registry and other log files. It also has developed a tool, AScan, to identify and recover evidences from unallocated spaces and slack spaces of hard disk drives. Fellows [128] described the significance of recovering the WinRAR temporary files and studied the behavior of WinRAR in creating these temporary files. The results of this research indicated that there is a chance to detect and recover the evidence file from deleted temporary folders while the original file is protected by cryptographic solutions.

4.2.9. File System Forensics

Conducting research on file systems structures and attributes is extremely important mainly due to the need to extract evidence from these file system which can be failed in case of unknown file systems. In [53], the authors studied the structure of file time attributes in the NTFS file system and analyzed the modifications in access, modification and creation time attributes of files/folders by the user under different operating system. Grier [129] proposed a methodology for examining filesystems and detecting emergent patterns unique to copying via stochastically modeling filesystem behavior through routine activity and emergent patterns in MAC timestamps. Beebe et al. [130] explained the functionality, architecture and disk layout of ZFS file system and then discussed the forensic investigation methods available for ZFS. This work also brought some of the forensic challenges of ZFS to light. Carrier [131] investigated the credibility of ISO9660 file system in forensic investigation and the fact that this file system can be used for data hiding. The details of data hiding process is studied and then used to create an image with hidden data for examining the available forensic toolkits.

In [132], authors presented a novel approach to identify the disk cluster size without relying on meta data of file system; instead, by detecting the difference between the entropy difference distributions of the non-cluster boundaries and cluster boundaries. Kavallaris and Katos [133] introduced a technique for identification of past pod slurping type of attacks using information stored in filesystem time stamp. This technique infers a file’s transfer rate from access time attribute and correlate it to the common rate of the suspicious USB (obtained from Windows registry) to identify the victim USB device.

4.2.10. Forensic Frameworks

Obviously, the forensic frameworks are the basis of the forensic process and developing new frameworks can guarantee the adaptation of forensic science with new technologies. In [134], authors made an extensive survey of available network forensic framework implementations and proposed a generic digital forensic investigation model with comparison to the available models. Finally, the implementation techniques of the proposed model are discussed. Beckett and Slay [135] highlighted the demand for adapting the new technologies with the science of forensic investigation so the digital forensics becomes a
true forensic science. This article examined the roots of scientific methods to establish principles of digital forensic as a science.

In [136], authors discussed the lack of standardized data sets, corpora, as a necessary requirement of forensic research; and presented a taxonomy for defining several available corpora. Guo et al. [137] studied the scientific description of computer forensic characteristics via identifying the basic functionalities of computer forensic investigation procedure; and introduced a functionality oriented validation and verification framework for digital forensic tools by using function mapping approach. Shields et al. [138] proposed PROOFS, a continuous forensic evidence collection tool, which utilizes data retrieval methods on file system forensic. PROOFS generate and save signature for files that are copied, deleted or altered over the network.

In [139], the authors analyzed the possible effects of a fictitious P2P model on committing a copyright violation crime while dealing with the illegal distribution of digital contents. Cohen et al. [140] introduced an open-source enterprise forensic investigation tool that provides remote access to memory and raw disk on multiple platforms. In continue, it described the architecture of the tool and how it performs enterprise forensic investigation regularly. Kahvedžić and Kechadi [141] presented a novel digital investigation ontology (DIALOG) for the management, reuse and analysis of digital investigation knowledge via creating a general vocabulary capable of describing the investigation at any level. Case et al. [142] explained the issue of correlation between digital investigation tools and then proposed a framework for automatic evidence identification with support of different targets. The implemented prototype presents the integrated analysis of configuration log file, memory image, disk image and network traffic captures.

In [143], the author shared the experience of forensic investigation of a commercial closed circuit television digital video recorder by using raw disk analysis of the storage disk. The result of investigation showed that with extra effort on processing the raw video data, it might be possible to extract the evidence in cases where the recording device is not functional. Kao et al. [144] proposed three analytical tools for clarifying forensic investigation issues of cybercrime using three strategies: Multi-faceted Digital Forensics Analysis (MDFA), Ideal Log and M-N model; in same order, they covered the basic elements of cyber-crime using new definitions of the forensic investigation, traceable elements of the evidence, and finally the ISP log records. Serrano et al. [145] designed a multi-agent system (MAS) debugging framework for developing a forensic analysis (in cases where MASs indicate complex tissues of connection among agents) by utilizing the pathfinder networks (PFNETs).

4.3. Critical overlooked issues

Privacy issues caused by digital forensic investigation is one of the topics which deserves more research in future as the issue rises where the investigation can threat the secrecy of unrelated data. The same challenge became even more complicated when the cloud computing and massive shared resources get involved. Neglect in conducting effective researches may lead to a direct conflict with citizens’ right of privacy can cause the digital forensic face a deadlock where the law enforcers cannot differentiate between potential evidences and other private data. Studying the available works in users right of privacy shows that the solution can be in successful identification of related evidence objects based on existing privacy policies. The current feasible solution is using formal method to tag pieces if data according to the privacy policy and only then start collecting evidences [146], [147].

Thanks to cloud computing concept, many conflicts introduced to the digital forensic investigation [148]–[150]. The jurisdiction of the data is one of the most challenging topics which seem to be overlooked. The digital forensic community should realize that this conflict will cause massive obstacle in legal aspects of the investigation. Developing a suitable cross national low is one of the solution which has been working on in the past couple of years but it demands much more affords to be feasible [151], [152].

Moreover, only a few articles discuss digital forensic investigation science and digital forensic awareness. It is utterly vital to
understand that the bests of forensic frameworks may have conflicts with the true nature of forensic science; the majority of these conflicts end up affecting the integrity of precious evidence. As an instance, imaging the physical memory in a forensic manner is one of the challenges which did not attract much of the researchers’ attention. It is for this reason that forensic science has emerged as a significant aspect of digital forensics. A well-conducted awareness campaign can help teach and make digital investigators and forensic researchers aware of these challenges. This may also help to update the investigators about the latest technologies and their new conflicts with forensic investigation disciplines on a regular bases; not only a once-off exercise.

5. CONCLUSION AND FUTURE WORKS
Digital crime is a moving target, from the era of telephone hackers up to the current state of the complex malware intrusions. With new developments and innovations, new types of crime came along. This survey result has shown that as we entered the twenty-first century, the scope of digital forensic investigation has widened and its focus is fast shifting toward mobile device and cloud based investigations. Digital forensics now requires a more coordinated and focused effort from the national and international society, governments and the private sector. It is no coincidence that the study shows a shift towards mobile device and cloud forensic while the true nature of forensic science becomes the essence of investigation frameworks. This survey results have also shown that most of today’s forensic challenges are to a greater extent in direct conflict with common digital forensic practices. All indicators points to a scientific approach in the future development of the digital forensic discipline. However, as we move forward to address the new challenges it is also critical that we continue strengthening the technologies. Finally, New research efforts is required that minimize the gap between regulatory issues and technical implementations.

6. REFERENCES


[57] M. Damshenas, A. Dehghantanha, and R. Mahmoud, “A SURVEY ON MALWARE PROPAGATION,


[75] L. Miralles and J. Moreno, “Versatile iPad forensic acquisition using the Apple Camera Connection Kit,” Computers


B. Dolan-Gavitt, “Forensic analysis of the Windows registry in memory,”


