Emerging trends and research in digital forensics

Anamika Joshi, D.S. Bhilare
School of computer science Devi Ahilya University Indore, India

Abstract

Digital Forensics is the field of forensics science that deals with digital crimes and crimes involving computers. Digital forensics is used not only to investigate computerized crimes, such as network intrusion, fabrication and unauthorized access of data and illegitimate material distribution through digital services, but also to investigate crime where evidence is stored in any digital format on any digital device. Advancements in information technology such as cloud computing, social networking, personal devices such as smart phones; BlackBerrys, web 2.0 etc. is on one side represent an opportunity and benefits to the organisation and on other side have posed new challenges for those policing cybercrimes. This paper discusses the research category in digital forensics and identifies key research issues of each of the category.

KEYWORDS: Digital forensics, digital evidence, network intrusion, information security.

1. Introduction

Cyber-crimes or digital crimes have increased in frequencies, and their degrees of sophistication have also advanced. It poses major implications for national and economic security [1, 5]. Many industries and institutions, public-and private-sector organizations are at significant risk. This statement has been proved by the number of complaints received and processed by the Federal Bureau of Investigation (FBI) in collaboration with Internet Crime Complain Center (IC3) [2]. In 2012, the total complaints received are 289,874, in which the complaints reporting loss are 114,908. The overall statics are illustrated in the following table.

<table>
<thead>
<tr>
<th>Overall Statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total complaints received</td>
<td>289,874</td>
</tr>
<tr>
<td>Complaints reporting loss</td>
<td>114,908</td>
</tr>
<tr>
<td>Total Loss</td>
<td>$ 525,441,110</td>
</tr>
<tr>
<td>Median dollar loss for those reporting a loss</td>
<td>$ 600</td>
</tr>
<tr>
<td>Average dollar loss overall</td>
<td>$ 1,813</td>
</tr>
<tr>
<td>Average dollar loss for those reporting a loss</td>
<td>$ 4,573</td>
</tr>
</tbody>
</table>

From the cyber-crime or complaints reported, it indicates that the number of crimes involving computers and internet have grown over the last decades and it needs some mechanism that can assist law enforcement to determine the Who, What, Where, When, and How for crimes. As a result, digital forensics (DF) has evolved to assure
proper and sufficient presentation of digital crime evidentiary data into court and the role of digital forensic becomes important to get digital evidence.

Digital investigations are distinguished from other types of investigations in two very important ways. First, they may be remote crimes. That means that the attack was initiated at some indeterminate distance from the target. The attacker may have used any of a number of techniques to obfuscate his or her true location or/and can use anti-forensics methods that prevent forensic tools, investigations, and investigators from achieving their goals [3]. The crime scene, literally, could extend around the world from an organization. The second distinguishing factor is the amount of data available to analyse. In a serious digital incident there can be terabytes of data that may (or may not) contain even bytes of evidences [4].

The current trends and new technologies such as cloud computing, social media, personal devices such as smart phones; web 2.0 etc. represent an opportunity for IT to deliver significant benefits to the organisation. However, new technology also means new enhanced and diversified risk. The digital forensic also facing new challenges because of it and to meet this challenges there is a need to build/modify digital forensics tools and techniques and identify the new research and development requirements.

This paper is organized as follows. Section 2 presents the related work in DF community. Section 3 briefly categorized the research issues in digital forensic according to recent review and concluded in section 4.

2. Related work

Unlike many research areas, digital forensics is a largely practitioner-driven field. As a result, the majority of the tools and practices have been developed in response to a specific incident or class of incidents, rather than as the result of a research plan.

There have been several efforts to define and find research areas in Digital forensics. On August 2001 the first Digital Forensic Research Workshop [6] was held with the objectives to begin forming communities of interested individuals and to start a meaningful dialog for defining the field and identifying the difficult, high-priority challenges that lie ahead. The major goal was to establish a research community that would apply the scientific method in finding focused near-term solutions driven by practitioner requirements and addressing longer term needs, considering but not constrained by current paradigms. The research problems identified are Framework for Digital Forensic Science, the Trustworthiness of Digital Evidence, Detection and Recovery of Hidden Data, and Digital Forensic Science in Networked Environments (Network Forensics).This was first and important attempt which gives proper direction to the Digital Forensics.

In June 2008 a group of digital forensics researchers, educators and practitioners met at CISSE 2008 with the goal of collecting ideas for research categories, topics and problems in digital forensics. The result of this was an article by Nance, Hay and Bishop that attempted to define a Digital Forensics Research Agenda [7]. The authors identified six categories for digital forensics research: Evidence Modelling, Network Forensics, Data Volume, Live Acquisition, Media Types, and Control Systems.
Digital forensics has evolved significantly over period of time and was no longer a niche discipline. It is now considered a mainstream knowledge by professionals and industry alike, thus establishing the commercial and theoretical as well as legal point of view that the digital footprints that remain after interactions with computers and networks are significant and probative.

Digital forensics was once a niche science that was leveraged primarily in support of criminal investigations, and digital forensic services were utilized only during the late stages of investigations after much of the digital evidence was already spoiled. Now digital forensic services are sought right at the beginning of all types of investigations. Even popular crime shows and novels regularly incorporate digital evidence in their story lines, which give DF a popular cult like standing in general public’s mind.

Beebe, in 2009[8] admitted as an area of future growth and improvement that digital forensics largely lacks standardization and processes, and what little widespread knowledge that we have is “heavily biased towards Windows platform, and to a lesser extent, standard Linux distributions.”

Beebe elaborated the unaddressed issues of digital forensics, highlighting that the problem of scalability, lack of intelligent analytics beyond full-text search that could link up multi part scenarios, non-standard computing devices especially small and hand held devices, ease-of-use, and a laundry list of unmet technical challenges are the work that needs to be addressed in future with new technologies and research.

In 2010 Garfinkle argued in his paper entitled “Digital forensics research: The next 10 years” [4] that we have been in a “Golden Age of Digital Forensics,” and that the Golden Age is quickly coming to an end. Increasingly organizations encounter data that cannot be analysed with today’s tools because of format incompatibilities, encryption, or simply a lack of training. He believes that Digital Forensics is facing a crisis as the result of advances and fundamental changes in the computer industry like the growing size of storage devices, the increasing prevalence of embedded flash storage, the proliferation of operating systems and file formats, pervasive encryption, use of the cloud for remote processing and storage and digital forensic analysis shifted from single device to multiple devices. At the end this article summarizes current forensic research directions and argues that to move forward the community needs to adopt standardized, modular approaches for data representation and forensic processing.

The development of digital forensics tools is strongly driven by practitioners who can readily adapt cutting-edge research, while a variety of systematic barriers challenge these practitioners as per Walls [9] who develops digital forensic tools for use by law enforcement organizations in United States. In 2012 Garfinkle shares his 14 years’ experience in developing DF tools in a paper entitled “Lessons learned writing digital forensics tools and managing a 30TB digital evidence corpus”[10]. In this he explains that writing digital forensics tools is difficult than other kinds of software because of the diversity of data types that needs to be processed, the need for high performance, the skill set of most users, and the requirement that the software run without crashing.
3. Research Category in Digital Forensics

Digital forensics is a largely practitioner-oriented field; because of less standardization, temporal diversity, never-ending upgrade cycle and high degree of uncertainty it is difficult to predict how the field of digital forensics will evolve. Based on various papers and Digital Forensics Research Workshop technical reports that we have reviewed, it seems that research in digital forensics can be categorized into six categories as shown in Fig1.

3.1 Framework:

Generally accepted digital forensics process framework is actively needed by digital forensics researchers, practitioners, and customers. A framework will provide a common starting point from which researchers and practitioners can identifies the research, new theories and development requirements [6].

![Figure 1: The six categories of Digital Forensics](image)

In many digital crimes, the procedures used before 2001 to accomplish digital investigation were neither consistent nor standardized. They were written with a focus on the details of the technology and without consideration for a generalized process [11]. In 2001, The Digital Forensics Research Working Group [6] defined a generic investigation process that can be applied to the majority of investigations involving digital systems and networks. This framework was an important foundation for future work.

A number of models and methodologies have been developed in the digital forensics field. The common elements shared by these models are identification, collection, preservation, analysis, and presentation. Some of them added addition detailed steps into the element to make the step more clear and precise.

Most of the models [6,11,12,22,13,14,15,17,19,20,24, and 28] are reactive in nature as the investigation is performed after the incident is reported. Only some of these models have explicitly mentioned the proactive process which gather and preserve digital evidence before and at the time of an incident occurs [26 and 27], some of the models such as [16, 18, 21, 23, and 25] also implicitly includes proactive process.
However, the nature of incidents and attacks has changed. As future work many papers [3,4,29,30] suggested that in order to address anti-forensics, to determine the root-cause of an incident and to successfully prosecute the perpetrator, the investigator needs relevant, admissible live digital evidences (for example volatile evidence, swap files, network processes etc.) which is possible by proactive forensics investigation.

The current investigation is very complex and time consuming as the single system with single small capacity disks are replaced by multiple systems with multiple large capacity disks and network storage. So in response to changing contexts there is a need a framework which can meet current challenges.

### 3.2 Network Forensics:

Network forensics is a branch of digital forensics that focuses on the monitoring and analysis of network traffic. It deals with volatile and dynamic data. It generally has two uses. The first, relating to security, involves detecting anomalous traffic and identifying intrusions. The second use, relating to law enforcement, involves capturing and analysing network traffic and can include tasks such as reassembling transferred files, searching for keywords, and parsing human communication such as emails or chat sessions and linking it to the ‘time frame’. The possible emerging network challenges and areas are [7, 48, 49, and 50]:

- Use of the “cloud” for remote processing and storage present new challenges because network data is often difficult to locate, thus acquisition might be challenging or even impossible.
- Social networking sites such as Google+, Facebook, Twitter, and YouTube have expanded rapidly in recent years, so there is a need for network forensic tools that address such an important area of usage.
- The processing of network forensic data in real time could involve many different protocols and the amount of data could potentially be very large. 10-Gbps traffic flow with a two-hour sliding window requires 10 Tbytes of storage.
- While workstation and server network forensics are somewhat well understood, it is less clear how network data from non-end-points, such as switches and routers could be collected, analysed and presented as evidence.
- The non-traditional networking devices are increasingly appearing in network. Examples of non-traditional network devices are office infrastructure (e.g. printers, copiers, scanners, fax machines), media players, game consoles, phones, and even cars and home appliances. There is a need to develop methods to determine how these devices interacted with the network during a time period of interest.

A number of researchers have worked on this area such as collect information from computer networks to support forensics investigation. But there is a need to create and improve networking tools and processes to handle above mentioned new challenges.

### 3.3 Data Volume
The digital forensics investigations were previously limited to the analysis of single systems with single small capacity disks, now increasingly investigations require analysis of multiple systems with multiple large capacity disks, network storage, and encrypted volumes. Today a 2TB hard drive can be purchased for $120 but takes more than 7 h to image; systems and individuals of interest can easily have more storage than the police crime lab responsible for performing the analysis.

The growing size of storage devices means that there is frequently insufficient time to create a forensic image of a subject device, or to process all of the data once it is found. To handle this problem many research works have been done on this direction firstly selected evidence acquisition like Kenneally and Brown [33] has suggested The Risk Sensitive Evidence Collection (RSEC) Methodology. RSEC is on-scene Pre-searching and subsequent identification of evidence/artifacts and selective artifact/evidence extraction from live or dead systems. Instead of imaging whole disk and filtered out data during the analysis of a collected image, its shift filtering from analysis to collection to reduce the overall time needed in collection and analysis, and it’s also reduce the storage requirement.

Brian Jones, Syd Pleno, and Michael Wilkinson [34] also suggest the use of random sampling in investigations and they discuss how the New South Wales Police Force, State Electronic Evidence Branch (SEEB) has implemented a “Discovery Process”. Using random sampling of files and applying statistical estimation to the results, the branch has been able to reduce backlogs from three months to 24 hours.

Secondly, applying data mining research to digital forensics [35] to gain some or all of the following benefits: (i) reduced system and human processing time associated with data analysis; (ii) improved information quality associated with data analysis; and (iii) reduced monetary costs associated with digital investigations.

And lastly Triaging [36, 37, 38, 39, and 40] has been proposed as a solution to systematically prioritize the acquisition and analysis of digital evidence.

Parallelization research could also provide benefits, the area identified in [7] are the imaging and carving process, and the development of user history timelines, including those based on multiple data sources. In addition, approaches that combine data imaging and evidence identification in parallel could also be beneficial, allowing an investigator to potentially direct the data acquisition process based on real-time results to acquire the most promising data sources during the initial phase of analysis.

### 3.4 Mobile and Embedded Devices

There have been major advances in the field as Computer Forensics has evolved into Digital Forensics. This change of name indicates the wide range of digital devices that are often part of an investigation. However, while devices such as cell phones, smartphones, including iPhone, Android and Blackberry, digital media players and game consoles may contain useful information, the law enforcement and forensics investigators have struggled to effectively manage digital evidence obtained from these devices. Some of the reasons include [51 and 52]:
• The mobile phones are the most diverse, as they tend to have no standard interface, either at the hardware or software levels, essentially making the analysis process unique to each device model.

• File systems that are contained in mobile devices operate from volatile memory or computer memory that requires power to maintain stored information.

• The short product cycles from the manufacturers to provide new mobile devices and their respective operating systems are making it difficult for law enforcement agencies to remain current with new technologies.

• Pertinent data such as call histories are stored in proprietary formats in locations that will alter that data according to phone model. Even the cable used to access the mobile device’s memory will vary according to manufacturer and model.

• Different formats of information stored on mobile phones.

There are many devices that are cheaply manufactured in China and are very difficult to perform forensics by examiners. The primary reason is that inexpensive Chinese cell phones are unbranded, meaning they have no International Mobile Equipment Identity (IMEI) number and therefore, cannot be traced. Furthermore, forensics tools often cannot handle new or less commonly encountered devices, leaving an investigator to either develop custom tools, or lose the opportunity to examine the device. In addition to the number of incompatible devices of a particular type, the number of device types, especially integrated devices, is also growing rapidly.

3.5 Live acquisition

Traditional digital forensics has focused on static analysis—that is, the analysis of non-volatile data from a halted computer system. This approach maximizes result reproducibility, but it misses dynamic state information, such as processes and network connections, memory-resident malware, unlocked file system decryption keys, or data in output buffers that isn’t yet written to file. In “Live Analysis: Progress and Challenges,” Brian Hay, Matt Bishop, and Kara Nance [31] explore the challenges and opportunities of live analysis—that is, the analysis of data gathered while a system is operating. The most significant challenge here is how to gather data without introducing distortions, especially when you must rely on the running system’s integrity to execute the data collection software correctly. Even live data collection using specialized hardware comes with opportunities for introducing distortion. Virtual computing presents new challenges and opportunities. On one hand, it enables continuous recordings of a virtual machine’s complete state, without running data-gathering software inside the virtual machine itself; on the other, it doesn’t entirely eliminate the possibility of distortion or detection by an opponent.

Sasa Mrdovic, Alvin Huseinovic and Ernedin Zajko [32] propose combination of static and live analysis. Virtualization is used to bring static data to life. Volatile memory dump is used to enable offline analysis of live data. Using data from memory dump, virtual machine created from static data can be adjusted to provide better picture of the live system at the time when the dump was made. Investigator can have interactive session with virtual machine without violating evidence integrity.
The research area identified in live acquisition are RAM analysis, methods for interrupting the execution for live acquisition, and methods for performing live analysis on systems without interrupting the execution sequences.[7] Live analysis gives less assurance and is less clear in compare to static analysis, as action taken by the investigator may change the state of the target system, and the dynamic state observed may not be reproducible, repeated analysis of the same state is not possible. As a result, there is certainly integrity, trustworthiness, and legal admissibility must be considered as a part of the research effort into the live analysis.

3.6 Anti- Forensics

Anti-Forensics is a set of techniques used as countermeasures to forensic analysis. It is an attempt to negatively affect the existence, amount and/or quality of evidence from a crime scene, or make the analysis and examination of evidence difficult or impossible to conduct. In “Forensics Is So ‘Yesterday’,” Michael A. Caloyannides [41] explains that computer forensics isn’t effective against anti-forensic techniques and thus won’t be useful for catching sophisticated criminals and agents. Instead, computer forensics will catch naive crooks who don’t know how to hide their tracks and innocent people who don’t know how to protect their systems. Liu and Brown identified four primary goals for anti-forensics [42]:

- Avoiding detection that some kind of event has taken place.
- Disrupting the collection of information.
- Increasing the time that an examiner needs to spend on a case.
- Casting doubt on a forensic report or testimony.

Anti-forensics methods are often broken down into several sub-categories: data hiding (encryption, steganography), artifact wiping, trail obfuscation and attacks against the computer forensics processes and tools (evidence elimination tools).

A number of research works has been done on this area such as [41, 42, 3 and 43] examining how to define and control the anti-forensics problem and [44, 45, 46 and 47] has worked to identify and detect the availability of anti-forensics.

4. Conclusions

Recent advances in the field provide both challenges and opportunities. Advancements in information technology such as cloud computing, social media, personal devices such as smart phones; BlackBerrys, web 2.0 etc. is on one side represent an opportunity and benefits to the organisation and on other side have posed new challenges for those policing cybercrimes. The current forensics tools and processes are unable to provide reliable results in this new environment. This paper summarizes current research directions and major research areas in the digital forensics field. The major categories described in this paper are framework, network forensics, data volume, mobile and embedded devices, live acquisition and anti-forensics.

REFERENCES


[28] Siti Rahayu Selamat, Robiah Yusof, Shahrin Sahib, and Irdia Roslan, (2011).” Adapting Traceability in Digital Forensic Investigation Process”. In proceeding of
Malaysian Technical Universities International Conference on Engineering & Technology (MUICET 2011)


