Chapter 2

DIGITAL FORENSIC RESEARCH: THE GOOD, THE BAD AND THE UNADDRESSED

Nicole Beebe

Abstract Digital forensics is a relatively new scientific discipline, but one that has matured greatly over the past decade. In any field of human endeavor, it is important to periodically pause and review the state of the discipline. This paper examines where the discipline of digital forensics is at this point in time and what has been accomplished in order to critically analyze what has been done well and what ought to be done better. The paper also takes stock of what is known, what is not known and what needs to be known. It is a compilation of the author's opinion and the viewpoints of twenty-one other practitioners and researchers, many of whom are leaders in the field. In synthesizing these professional opinions, several consensus views emerge that provide valuable insights into the "state of the discipline."

Keywords: Digital forensic research, evaluation, future research areas

1. Introduction

Digital forensics is defined as: "[t]he use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitation or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations" [46]. This paper presents the results of a "state of the discipline" examination of digital forensics. The focus is on digital forensic research – systematic, scientific inquiries of facts, theories and problems related to digital forensics. In particular, the paper character-

izes the current body of knowledge in digital forensics and evaluates its rigor with the goal of setting a course for future digital forensic research.

This evaluation of the current state of digital forensics attempts to convey the perspectives of researchers and practitioners. In addition to the author, the views expressed in this paper are drawn from twenty-one researchers and practitioners, many of whom are leaders in the field. Researchers were asked to critically examine the collective contribution of the research community to the current body of knowledge and to identify the most pressing research questions in digital forensics. Practitioners were queried about current and previous research activities, the contribution of these activities to their real-world experiences, and pressing research needs.

The remainder of the paper is organized as follows. First, we discuss what is collectively seen as "good" in the field, outlining the notable research contributions over the past decade. Next, we discuss the "bad" – what needs to be improved as far as digital forensic research is concerned. Finally, we attempt to set a course for future digital forensic research by discussing four major themes and several individual research topics that demand investigation. A brief literature review of the research themes and topics is presented to assist individuals who are interested in embarking on research in these areas.

2. The Good

All the respondents felt that there was unequivocal improvement in the prominence and value of digital evidence in investigations. It is now mainstream knowledge 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 – criminal, civil, military and corporate. Even popular crime shows and novels regularly incorporate digital evidence in their story lines.

The increased public awareness of digital evidence says nothing about the state of digital forensics as a science. Indeed, the awareness of the need to collect and analyze digital evidence does not necessarily translate to scientific theory, scientific processes and scientifically derived knowledge. The traditional forensic sciences (e.g., serology, toxicology and ballistics) emerged out of academic research, enabling science to precede forensic science applications, as it should. Digital forensics, however,

emerged out of the practitioner community – computer crime investigators and digital forensic tool developers seeking solutions to real-world problems [47]. While these efforts have produced a great amount of factual knowledge and several commonly accepted processes and hardware and software tools, many experts concede that the scientific method did not underlie much of early digital forensic research.

The call for a more scientific approach to digital forensic research is not new. The first Digital Forensic Research Workshop (DFRWS 2001) convened more than 50 researchers, investigators and analysts "...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" [46].

The prevailing sentiment is that the scientific foundation of digital forensics has been strengthened. A few examples highlight this point. The Scientific Working Group on Digital Evidence (SWGDE) has released several documents since 1999 concerning digital forensic standards, best practices and testing and validation processes. In 2001, the U.S. National Institute of Standards and Technology (NIST) started the Computer Forensic Tool Testing (CFTT) Project and has since established and executed validation test protocols for several digital forensic tools. DFRWS 2002 focused on scientific standards and methods. DFRWS 2003 hosted a plenary session emphasizing the need for a scientific foundation for digital forensic research. DFRWS 2004 featured several presentations defining the digital forensic process. In short, the field has experienced considerable progress in formalizing, standardizing and formulating digital forensic processes and approaches. The respondents also felt that the research community was showing signs of a stronger scientific underpinning as evidenced by the publication of research in mainstream computer science, information systems and engineering journals.

To date, research questions have largely centered on the "archaeology" of digital artifacts. Carrier [12] observes that digital forensic artifacts are a function of the physical media, operating system, file system and user-level applications – that each impacts what digital evidence is created and left behind. Like archaeologists who seek to understand past human behavior by studying artifacts, digital forensic investigators seek to understand past behavior in the digital realm by studying digital artifacts. Because digital forensic research during the past decade has focused on the identification, excavation and examination of digital artifacts, there is now a relatively solid understanding of what digital artifacts exist, where they exist, why they exist and how to recover them (relative to

commonly-used operating/file systems and software applications). To its credit, the digital forensic research community has done a good job sharing this knowledge with other academic disciplines (e.g., computer science, information systems, engineering and criminal justice) as well as with the practitioner community (law enforcement, private-sector practitioners and e-discovery specialists).

Digital forensic research has profited from "digital forensic challenges" designed to stimulate scientific inquiry and the development of innovative tools and analytical methodologies. Examples are the annual digital forensic challenges sponsored by DFRWS (since 2005) and the U.S. DoD Cyber Crime Center (since 2006). Two research areas that have experienced significant growth as a result of the challenges are data carving and memory analysis. Other areas that have benefited include steganography, encryption and image identification (especially distinguishing between real and computer-generated or computer-altered images).

However, the digital forensic challenges may have shifted research attention away from the response and data collection phases to the analysis phase. Many of the respondents opined that digital forensic research initially focused its attention on response and data collection. As a result, robust hardware write blockers became widely available; live response processes, tools and methodologies that minimized the digital evidence footprint were developed; and commonly accepted acquisition policies and procedures emerged. One might argue that the research community has marched along the digital forensic process: Preparation → Response → Collection → Analysis → Presentation → Incident Closure [4]. Many research questions pertaining to the response and collection phases are still unanswered, it is just that the research community now sees the most pressing questions as residing in the analysis phase.

3. The Bad

Interestingly, several of the key successes of digital forensic research follow directly into the discussion of "The Bad." Take, for example, acquisition process standardization and formalization. Several of the respondents suggested that the digital forensic community has almost hyper-formalized processes and approaches, especially with respect to the response and data collection phases. Some would argue this point, citing the fact that there is no single, universal standard for digital evidence collection. Many organizations have their own standards and guidelines for data collection. Further, it can be argued that these documents constitute high-level, work-flow guidance rather than proscriptive

checklists. Thus, at first glance, one can easily make an argument against the allegation of hyper-formalization.

The argument that digital forensic processes are hyper-formalized centers on the fact that the evidentiary principles established over the years cannot be attained under certain circumstances. Consider the evidentiary principles of integrity and completeness. The digital forensic community has worked hard to get the judiciary to understand that the right way to respond and collect digital evidence does not alter the evidence in any way and obtains all the evidence. The problem is that the changing technological landscape often necessitates a different approach – one where evidence will be altered (albeit minimally and in a deterministic manner) and where not all the evidence can be seized. Modern digital crime scenes frequently involve multi-terabyte data stores, mission-critical systems that cannot be taken offline for imaging, ubiquitous sources of volatile data, and enterprise-level and/or complex incidents in which the scope and location of digital evidence are difficult to ascertain. Many organizational standards and guidelines fail to address response and data acquisition in such circumstances; they often fail to facilitate proper decision-making in the light of unexpected digital circumstances; and they often present evidentiary principles as "rules," leaving little room for improvisation.

One of the successes identified in the previous section was the collective ability to archaeologically identify, excavate and examine digital artifacts. The problem, however, is that knowledge and expertise are heavily biased toward Windows, and to a lesser extent, standard Linux distributions. The FAT12/16/32, NTFS and EXT2/3 file systems, the operating systems that implement them (Windows9X/ME/NT/XP/Vista and various Linux distributions), and common user applications installed on them (Microsoft Internet Explorer and Outlook, Mozilla Firefox and Thunderbird, etc.) have been well studied. Researchers have paid insufficient attention to other operating systems, file systems and user applications, especially in the light of current market trends.

The market share enjoyed by Microsoft operating systems has decreased from 91.8% in May 2008 to 88.1% in March 2009 [43], due in large part to Apple's Mac OS X and its portable device operating systems. "Mac forensics" (e.g., HFS+ file system forensics) has received increased research attention, but more efforts are needed. New file systems, such as ZFS by Sun Microsystems, have received little attention. UFS and ReiserFS are also examples of file systems that deserve to be the focus of more research.

The digital forensic research community must also challenge itself by raising the standards for rigor and relevance of research in digital foren-

sics. In years past, the challenge for the community was limited knowledge, skills and research experience. Despite backgrounds in computer science and information systems, and in some cases, real-world digital forensic training and case experience, members of the research community still had to overcome a steep learning curve with respect to the core body of knowledge in digital forensics (if, in fact, such a core ever existed). This unfortunately led to lower standards for scientific rigor and relevance in digital forensic research compared with other traditional fields of research.

The problem of rigor and relevance in digital forensic research is exacerbated by two publication dilemmas. First, the receptivity of mainstream scientific journals to digital forensic research is nascent. Second, peer reviewed, scientific journals dedicated entirely to digital forensics are relatively new. Journals typically take time to achieve high quality standards (a function of increasing readership and decreasing acceptance rates over time), and the discipline of digital forensics is simply not there yet. A survey of journals dedicated to digital forensics reveals that the ISI impact factors, circulation rates and acceptance rates are below par (but they are improving). This problem will subside as time passes and the discipline grows and matures.

Regardless of the mitigating circumstances noted above, the research community must regulate itself and "raise the bar." It should ensure that every research publication makes a clear (even if only incremental) contribution to the body of knowledge. This necessitates exhaustive literature reviews. More importantly, the work should be scientifically sound and comparable in rigor to research efforts in more established scientific disciplines.

It is also important to ensure that digital forensic research is relevant to the practitioner community. One strategy is to better address the problem phrased by Pollitt as "data glut, knowledge famine." Investigators usually do not lack data, but they often struggle with transforming the data into investigative knowledge. Research should strive to minimize noise and maximize contextual information, thereby converting data to investigative knowledge (Figure 1). The second strategy is for digital forensic research to facilitate tool development (at least to some degree). While some believe it is the responsibility of the digital forensic research community to develop usable tools, many of the respondents disagreed with this assertion. Nonetheless, it is important to ensure that the research is accessible and communicated to digital forensic tool developers, so that key contributions to the body of knowledge are placed in the hands of practitioners.

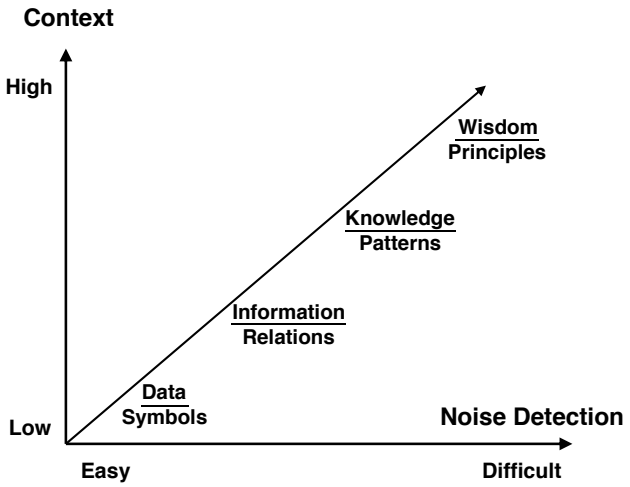

Figure 1. Knowledge management understanding hierarchy [44].

All things considered, "The Bad" is not that horrible. Rather, it calls for a reorientation and introspection by the digital forensic research community. As the research community approaches the apex of the learning curve, it should return to its scientific roots. It should reach deep into the relevant reference disciplines, build theory via strong analytics and methodological rigor, conduct scientifically sound experiments, and demand that all publications make clear contributions to the body of knowledge in digital forensics. The digital forensic community as a whole must achieve an awareness of the changing technological landscape that goes beyond practice. Indeed, researchers should be the ones advising practitioners about technologies on the horizon, not vice versa. Researchers should be the ones driving tool development based on their studies of archaeological artifacts, human analytical approaches and computational search and extraction algorithms, rather than having the tools drive digital forensic analytical approaches, processes and outcomes. Overall, digital forensic research must become more rigorous, relevant and forward thinking.

4. The Unaddressed

Having reviewed the current state of digital forensics, it is important to identify strategic directions for research in the field. Our study reveals four major themes: (i) volume and scalability, (ii) intelligent analytical approaches, (iii) digital forensics in and of non-standard computing environments, and (iv) forensic tool development. The following sections

discuss the four themes and list several pressing research topics after the discussion of each theme.

4.1 Volume and Scalability

Data storage needs and data storage capacities are ever increasing. Ten years ago, it was common to acquire hard disks in 700 MB image segments in order to burn an entire image to a handful of CD-ROMs. Now, "small" cases often involve several hundred gigabytes of data and multi-terabyte corporate cases are commonplace. Two years ago, the size of Wal-Mart's data warehouse exceeded the petabyte mark [75].

One solution to the volume and scalability challenge is selective digital forensic acquisition. Instead of acquiring bit-stream images of entire physical devices, subsets of data are strategically selected for imaging. Typically, the result is a logical subset of the stored data and not all logical data at that. We contend that selective acquisition can and should include certain portions of allocated and unallocated space, but admittedly, research is needed to facilitate such acquisitions (especially related to the decision making process that would identify the data to be selectively acquired).

Research on selective, intelligent acquisition includes using digital evidence bags [71, 72] and risk sensitive digital evidence collection [32]. Digital evidence bags are designed to store provenance information related to the data collected via selective acquisition. This type of approach is important because, when acquiring subsets of data from disparate sources, the source and contextual data (i.e., the physical device and the subset of data that is not acquired) are no longer implicitly available and must be explicitly retained. Furthermore, any explicitly retained information can and should be managed in order to contribute knowledge to the analytical process. Risk sensitive collection provides a framework for allowing cost-benefit considerations to drive the selection process, considering costs and benefits to the investigating and data-owning entities.

Another solution to the volume and scalability challenge is to utilize more effective and efficient computational and analytical approaches. Research focused on improving the efficiency and effectiveness of digital forensic analyses include distributed analytical processing [56], data-mining-based search processes [5], file classification to aid analysis [60], self-organizing neural networks for thematically clustering string search results [6], and massive threading via graphical processing units (GPUs) [38]. Researchers have also investigated network-based architectures and virtualization infrastructures to facilitate large evidence stores, case and

digital asset management systems, and collaborative, geographically distributed analysis [19, 20].

Clearly, more research is needed to address volume and scalability issues. The following research questions are proffered for further inquiry by the research community.

- How can decision support systems be extended to the digital forensics realm to aid selective and intelligent acquisition?
- What are the dimensions of the selective acquisition decision making process and how do they differ from other processes where decision support systems are applied?
- How can data warehousing and associated information retrieval and data mining research be extended to digital forensics? Beebe and Clark [5] discuss how data mining can be extended to digital forensics, but more research is needed. Which information retrieval and data mining approaches and algorithms can be extended to address specific analytical problems?
- Since information retrieval and data mining approaches are typically designed to handle logical and relatively homogeneous data sets, what adaptations are necessary to deal with physical, highly heterogeneous data sets?
- Link analysis research has been extended to non-digital investigations in the form of crime network identification and analysis, but how can similar research be extended to digital investigations and data sets? Can this research help characterize the relationships between digital events (chronological or otherwise) and digital data?
- How do investigators search and analyze data? What cognitive processing models are involved? How do these compare with other human information processing, knowledge generation and decision making processes?
- How does and/or should Simon's concept of "satisficing" [66] extend to digital forensic investigations?
- Why is digital forensic software development lagging hardware advances in the areas of large-scale multi-threading and massive parallelism? Are there unique characteristics about data and information processing tasks in a digital forensic environment that make it more difficult or necessitate different development or engineering approaches?

4.2 Intelligent Analytical Approaches

The second major research theme is intelligent analytical approaches. Several respondents felt that computational approaches for searching, retrieving and analyzing digital evidence are unnecessarily simplistic. Current approaches largely rely on: (i) literal string searching (i.e., non-`grep` string searches for text and file signatures), (ii) simple pattern matching (i.e., `grep` searches), (iii) indexing data to speed up searching and matching, (iv) hash analyses, and (v) logical level file reviews (i.e., log analysis, registry analysis, Internet browser file parsing, viewing allocated files, etc.). There are two problems associated with these approaches: underutilization of available computational power and high information retrieval overhead.

Current information retrieval and analytical approaches underutilize available computational power. Many forensic search processes require large amounts of processing time and researchers continue to seek ways to conduct searches and analyze data more quickly [38, 56]. However, the amount of time required to conduct byte-by-byte matching or full-text indexing is not the issue. The point is that high-end, user-class computing platforms (akin to typical digital forensic workstations) can handle intelligent search, retrieval and analytical algorithms that are much more advanced than literal string searches and simple pattern matching. Advanced algorithms already exist and are the result of longstanding research efforts in artificial intelligence, information science, data mining and information retrieval.

Current search and analysis approaches also have significant information retrieval overhead. In addition to the computational time required to execute a search, the overhead includes the human information processing time spent to review hits that are not relevant to the investigative objectives (i.e., false positives in the investigative sense).

The classic precision-recall trade-off dilemma is that as recall increases, the query precision decreases. Since digital forensics seeks recall rates at or near 100%, query precision is usually low. The heterogeneity of data sources during physical-level forensic analysis intensifies the problem. Note that traditional query recall targets could be reconsidered in the light of legal sufficiency and Simon's notion of satisficing [66].

The cost of human analytical time spent sifting through non-relevant search hits is a significant issue. Skilled investigators are in limited quantity, highly paid and often face large case backlogs. Anything that can be done to reduce the human burden should be seriously considered, even if it means increasing computational time. Trading equal human analytical time for computer processing time is a worthwhile proposition in and

of itself, but we believe that the trade will seldom be equal. Extending intelligent search, retrieval and analytical algorithms will not require a one-for-one trade between human and computer time. Computational processing time will indeed increase, but it will pale in comparison with the amount and cost of human analytical time savings.

Research in intelligent analytical approaches is relatively scant. Much of the work was discussed in the context of volume and scalability challenges. "Smarter" analytical algorithms would clearly reduce information retrieval overhead. They should help investigators get to relevant data more quickly, reduce the noise investigators must wade through, and help transform data into information and investigative knowledge.

In addition to improving analytical efficiency, intelligent analytical approaches would enhance analytical effectiveness. Research has shown that data mining algorithms can reveal data trends and information otherwise undetectable by human observation and analysis. Indeed, the increased application of artificial intelligence, information science, data mining and information retrieval algorithms to digital forensics will enable investigators to obtain unprecedented investigative knowledge.

The following topics in the area of intelligent analytical approaches should be of interest to the research community:

- Advances in the use and implementation of hashing, including the use of bloom filters to improve the efficiency of hash analyses [55], and the use of hashes as probabilistic, similarity measures instead of binary measures of identicalness [34, 57, 58].
- Using self-organizing neural networks to thematically cluster string search results and provide relevant search hits significantly faster than otherwise possible [6].
- Automated, large-scale file categorization within homogeneous file classes or file types [60].
- Statistically assessing "Trojan defense" claims [11].
- Feature-based data classification without the aid of file signatures or file metadata [64].
- Applying artificial intelligence techniques (e.g., support vector machines and neural networks) to analyze offline intrusion detection data and detect malicious network events [42].
- Using association rule mining for log data analysis and anomaly detection [1, 2].
- Using support vector machines for email attribution [21].

4.3 Non-Standard Computing Environments

The standard computing environment has long been the personal computer (desktops and laptops). Accordingly, digital forensic research has focused on acquiring and analyzing evidence from hard disk drives and memory. However, the technological landscape is changing rapidly – there is no longer a single "standard" computing environment. Small, mobile devices are ubiquitous and vary greatly (e.g., mobile phones, PDAs, multimedia players, GPS devices, gaming systems, USB thumb drives, etc.). Virtualization is widespread in personal and organizational computing infrastructures. Cloud computing is rapidly relocating digital evidence and commingling it with data from other organizations, which introduces new legal challenges. Operating systems and file systems are no longer 90% Microsoft Windows based. Investigators are encountering large numbers of custom-built digital devices. How will the digital forensic community deal with these non-standard computing environments and devices?

Researchers have made great strides in the area of small device forensics (see, e.g., [10, 13, 14, 22, 30, 33, 37, 41, 61, 67, 69, 70, 74, 76]). But much more work remains to be done given the rapid pace with which new models and devices enter the market.

Virtualization is also an area that deserves attention. Most research efforts have focused on leveraging virtualization in digital forensic analytic environments [8, 19, 49] or in educational environments [51] rather than conducting digital forensics of virtual environments. Dorn and coworkers [25] recently examined the digital forensic impact of virtual machines on their hosts. This work falls in the important area of "analysis of virtual environments" [51], which includes forensic data acquisition, virtual platform forensics and virtual introspection.

Cloud computing is another area that has received little attention. Most research is geared towards leveraging cloud computing (data center CPUs) to conduct digital forensic investigations more efficiently (see, e.g., [59]). To our knowledge, no research has been published on how cloud computing environments affect digital artifacts, and on acquisition logistics and legal issues related to cloud computing environments.

Clearly, there are benefits to be gained from researching virtualization and cloud computing as resources for completing digital forensic investigations as well as for their impact on digital forensic artifacts. To date, however, there has been little research in either direction.

Digital forensic research is also playing catch-up with non-Microsoft-based operating systems and file systems. Mac forensics is a growing field, and necessarily so, but more research is required in this area. Other

important areas of research are HFS+ and ZFS file system forensics [7, 9, 18, 31, 40] and Linux and Unix system forensics [17, 26, 50].

4.4 Forensic Tool Development

The fourth and final unaddressed research theme is the design and implementation of digital forensic tools. Many of the respondents believed that current tools were somewhat limited in terms of their ease of use and software engineering.

Ease of use is a major issue. Tools must not be too technical and must have intuitive interfaces, but, at the same time, they should be customizable for use by skilled practitioners. Furthermore, the goal should be to provide information and knowledge, not merely data. This might be accomplished through data visualization, automated link analysis, cross-correlation and features for "zooming in" on information to reduce information overhead. Another approach is to shift from the tradition of presenting data hierarchically based on file system relationships to presenting data temporally. The digital forensic research community should consider, extend and adapt approaches devised by graphics and visualization and human-computer interaction researchers.

The respondents also suggested several improvements with respect software engineering. Software development must take advantage of hardware advances, including massive parallelism and streaming. Increased interoperability via standardized data (i.e., tool input/output) and API formats is needed. More operating-system-independent tools (e.g., PyFlag [16]) are required. All-in-one tools with respect to data types are needed; such tools would intelligently leverage data from static media, volatile memory/devices, network dumps, etc. It is also important to increase the automation of forensic processes.

These suggestions beg the fundamental question: Why are digital forensic tools not there yet? Is it a symptom of the relative nascence of the field and, thus, the tools? Or, are digital forensic analytical tasks fundamentally different from tasks in other domains where similar technologies work? If so, are the differences regarding the human analytical sub-task, the computational sub-task, or both? Is this even a valid research stream for digital forensic researchers, or should it be left to commercial software developers? Do these questions necessitate research, or simply awareness of the problem on the part of tool developers? In any case, these questions must be considered and collectively answered by the digital forensic community as soon as possible.

4.5 Other Important Research Topics

In addition to the four major themes, several important research topics emerged during our discussions with digital forensic researchers and practitioners. These include:

- Detection, extraction and analysis of steganographically-inserted data, particularly data inserted by non-standard stego applications (see, e.g., [3, 27, 29, 39, 52–54, 65]).
- Database forensics (see, e.g., [45]).
- Forensic processes on live and volatile sources of digital evidence, evidentiary disturbance caused by memory acquisition and live forensic analysis, and evidentiary integrity processes and standards (see, e.g., [23, 24, 28, 35, 62, 63, 68, 73]).
- Emergent metadata standards and trends (e.g., new XML Office document standards) (see, e.g., [48]).
- Investigations involving multiple, distributed systems.
- Increased insight into error rates as they pertain to digital forensic tools, processes, algorithms and approaches (especially related to Daubert standards). It is not clear if the traditional concept of error rates is appropriate or if the paradigm should be shifted to the determination and quantification of the confidence of conclusions (see, e.g., [15]).
- Formalization of the hypothesis generation and testing process in examinations.
- Experimental repeatability and comparability of scientific research findings (including the creation of common test corpora).

5. Conclusions

Digital forensic research has experienced many successes during the past decade. The importance of digital evidence is now widely recognized and the digital forensic research community has made great strides in ensuring that "science" is emphasized in "digital forensic science." Excellent work has been accomplished with respect to identifying, excavating and examining archaeological artifacts in the digital realm, especially for common computing platforms. Also, good results have been obtained in the areas of static data acquisition, live forensics, memory acquisition and analysis, and file carving.

However, strong efforts should be directed towards four key research themes and several individual research topics. The four key themes are: (i) volume and scalability challenges, (ii) intelligent analytical approaches, (iii) digital forensics in and of non-standard computing environments, and (iv) forensic tool development. In addition to these larger themes, pressing research topics include steganography detection and analysis, database forensics, live file system acquisition and analysis, memory analysis, and solid state storage acquisition and analysis.

Acknowledgements

The following individuals have contributed to this assessment of the discipline of digital forensics: Frank Adelstein, Dave Baker, Florian Bucholz, Ovie Carroll, Eoghan Casey, DeWayne Duff, Drew Fahey, Simson Garfinkel, John Garris, Rod Gregg, Gary Kessler, Gary King, Jesse Kornblum, Russell McWhorter, Mark Pollitt, Marc Rogers, Vassil Roussev, Sujeet Shenoi, Eric Thompson, Randy Stone and Wietse Venema. Their assistance is gratefully acknowledged.

References

[1] T. Abraham and O. de Vel, Investigative profiling with computer forensic log data and association rules, *Proceedings of the IEEE International Conference on Data Mining*, pp. 11–18, 2002.

[2] T. Abraham, R. Kling and O. de Vel, Investigative profile analysis with computer forensic log data using attribute generalization, *Proceedings of the Fifteenth Australian Joint Conference on Artificial Intelligence*, 2002.

[3] K. Bailey and K. Curran, An evaluation of image based steganography methods, *International Journal of Digital Evidence*, vol. 2(2), 2003.

[4] N. Beebe and J. Clark, A hierarchical, objectives-based framework for the digital investigations process, *Digital Investigation*, vol. 2(2), pp. 147–167, 2005.

[5] N. Beebe and J. Clark, Dealing with terabyte data sets in digital investigations, in *Advances in Digital Forensics*, M. Pollitt and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 3–16, 2005.

[6] N. Beebe and J. Clark, Digital forensic text string searching: Improving information retrieval effectiveness by thematically clustering search results, *Digital Investigation*, vol. 4(S1), pp. 49–54, 2007.

[7] N. Beebe, S. Stacy and D. Stuckey, Digital forensic implications of ZFS, to appear in *Digital Investigation*, 2009.

[8] D. Bem and E. Huebner, Computer forensic analysis in a virtual environment, *International Journal of Digital Evidence*, vol. 6(2), 2007.

[9] A. Burghardt and A. Feldman, Using the HFS+ journal for deleted file recovery, *Digital Investigation*, vol. 5(S1), pp. 76–82, 2008.

[10] P. Burke and P. Craiger, Forensic analysis of Xbox consoles, in *Advances in Digital Forensics III*, P. Craiger and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 269–280, 2007.

[11] M. Carney and M. Rogers, The Trojan made me do it: A first step in statistical based computer forensics event reconstruction, *International Journal of Digital Evidence*, vol. 2(4), 2004.

[12] B. Carrier, *File System Forensic Analysis*, Addison-Wesley, Boston, Massachusetts, 2005.

[13] H. Carvey, Tracking USB storage: Analysis of Windows artifacts generated by USB storage devices, *Digital Investigation*, vol. 2(2), pp. 94–100, 2005.

[14] F. Casadei, A. Savoldi and P. Gubian, Forensics and SIM cards: An overview, *International Journal of Digital Evidence*, vol. 5(1), 2006.

[15] E. Casey, Error, uncertainty and loss in digital evidence, *International Journal of Digital Evidence*, vol. 1(2), 2002.

[16] M. Cohen, PyFlag – An advanced network forensic framework, *Digital Investigation*, vol. 5(S1), pp. 112–120, 2008.

[17] P. Craiger, Recovering digital evidence from Linux systems, in *Advances in Digital Forensics*, M. Pollitt and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 233–244, 2005.

[18] P. Craiger and P. Burke, Mac OS X forensics, in *Advances in Digital Forensics II*, M. Olivier and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 159–170, 2006.

[19] P. Craiger, P. Burke, C. Marberry and M. Pollitt, A virtual digital forensics laboratory, in *Advances in Digital Forensics IV*, I. Ray and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 357–365, 2008.

[20] M. Davis, G. Manes and S. Shenoi, A network-based architecture for storing digital evidence, in *Advances in Digital Forensics*, M. Pollitt and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 33–43, 2005.

[21] O. de Vel, A. Anderson, M. Corney and G. Mohay, Mining email content for author identification forensics, *ACM SIGMOD Record*, vol. 30(4), pp. 55–64, 2001.

[22] A. Distefano and G. Me, An overall assessment of mobile internal acquisition tool, *Digital Investigation*, vol. 5(S1), pp. 121–127, 2008.

[23] B. Dolan-Gavitt, The VAD tree: A process-eye view of physical memory, *Digital Investigation*, vol. 4(S1), pp. 62–64, 2007.

[24] B. Dolan-Gavitt, Forensic analysis of the Windows registry in memory, *Digital Investigation*, vol. 5(S1), pp. 26–32, 2008.

[25] G. Dorn, C. Marberry, S. Conrad and P. Craiger, Analyzing the impact of a virtual machine on a host machine, in *Advances in Digital Forensics V*, G. Peterson and S. Shenoi (Eds.), Springer, Heidelberg, Germany, pp. 69–81, 2009.

[26] K. Eckstein and M. Jahnke, Data hiding in journaling file systems, *Proceedings of the Fifth Digital Forensic Research Workshop*, 2005.

[27] C. Hosmer and C. Hyde, Discovering covert digital evidence, *Proceedings of the Third Digital Forensic Research Workshop*, 2003.

[28] E. Huebner, D. Bem, F. Henskens and M. Wallis, Persistent systems techniques in forensic acquisition of memory, *Digital Investigation*, vol. 4(3-4), pp. 129–137, 2007.

[29] J. Jackson, G. Gunsch, R. Claypoole and G. Lamont, Blind steganography detection using a computational immune system approach: A proposal, *Proceedings of the Second Digital Forensic Research Workshop*, 2002.

[30] W. Jansen and R. Ayers, An overview and analysis of PDA forensic tools, *Digital Investigation*, vol. 2(2), pp. 120–132, 2005.

[31] R. Joyce, J. Powers and F. Adelstein, MEGA: A tool for Mac OS X operating system and application forensics, *Digital Investigation*, vol. 5(S1), pp. 83–90, 2008.

[32] E. Kenneally and C. Brown, Risk sensitive digital evidence collection, *Digital Investigation*, vol. 2(2), pp. 101–119, 2005.

[33] M. Kiley, T. Shinbara and M. Rogers, iPod forensics update, *International Journal of Digital Evidence*, vol. 6(1), 2007.

[34] J. Kornblum, Identifying almost identical files using context triggered piecewise hashing, *Digital Investigation*, vol. 3(S1), pp. 91–97, 2006.

[35] J. Kornblum, Using every part of the buffalo in Windows memory analysis, *Digital Investigation*, vol. 4(1), pp. 24–29, 2007.

[36] G. Kowalski and M. Maybury, *Information Storage and Retrieval Systems: Theory and Implementation*, Kluwer, Norwell, Massachusetts, 2000.

[37] C. Marsico and M. Rogers, iPod forensics, *International Journal of Digital Evidence*, vol. 4(2), 2005.

[38] L. Marziale, G. Richard and V. Roussev, Massive threading: Using GPUs to increase the performance of digital forensic tools, *Digital Investigation*, vol. 4(S1), pp. 73–81, 2007.

[39] B. McBride, G. Peterson and S. Gustafson, A new blind method for detecting novel steganography, *Digital Investigation*, vol. 2(1), pp. 50–70, 2005.

[40] K. McDonald, To image a Macintosh, *Digital Investigation*, vol. 2(3), pp. 175–179, 2005.

[41] B. Mellars, Forensic examination of mobile phones, *Digital Investigation*, vol. 1(4), pp. 266–272, 2004.

[42] S. Mukkamala and A. Sung, Identifying significant features for network forensic analysis using artificial intelligence techniques, *International Journal of Digital Evidence*, vol. 1(4), 2003.

[43] Net Applications, Global Market Share Statistics, Aliso Viejo, California (marketshare.hitslink.com), April 9, 2009.

[44] J. Nunamaker, N. Romano and R. Briggs, A framework for collaboration and knowledge management, *Proceedings of the Thirty-Fourth Hawaii International Conference on System Sciences*, 2001.

[45] M. Olivier, On metadata context in database forensics, *Digital Investigation*, vol. 5(3-4), pp. 115–123, 2009.

[46] G. Palmer, A Road Map for Digital Forensic Research, DFRWS Technical Report, DTR-T001-01 Final, Air Force Research Laboratory, Rome, New York, 2001.

[47] G. Palmer, Forensic analysis in the digital world, *International Journal of Digital Evidence*, vol. 1(1), 2002.

[48] B. Park, J. Park and S. Lee, Data concealment and detection in Microsoft Office 2007 files, *Digital Investigation*, vol. 5(3-4), pp. 104–114, 2009.

[49] M. Penhallurick, Methodologies for the use of VMware to boot cloned/mounted subject hard disks, *Digital Investigation*, vol. 2(3), pp. 209–222, 2005.

[50] S. Piper, M. Davis, G. Manes and S. Shenoi, Detecting hidden data in EXT2/EXT3 file systems, in *Advances in Digital Forensics*, M. Pollitt and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 245–256, 2005.

[51] M. Pollitt, K. Nance, B. Hay, R. Dodge, P. Craiger, P. Burke, C. Marberry and B. Brubaker, Virtualization and digital forensics: A research and teaching agenda, *Journal of Digital Forensic Practice*, vol. 2(2), pp. 62–73, 2008.

[52] B. Rodriguez and G. Peterson, Detecting steganography using multi-class classification, in *Advances in Digital Forensics III*, P. Craiger and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 193–204, 2007.

[53] B. Rodriguez, G. Peterson and K. Bauer, Fusion of steganalysis systems using Bayesian model averaging, in *Advances in Digital Forensics IV*, I. Ray and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 345–355, 2008.

[54] B. Rodriguez, G. Peterson, K. Bauer and S. Agaian, Steganalysis embedding percentage determination with learning vector quantization, *Proceedings of the IEEE International Conference on Systems Man and Cybernetics*, vol. 3, pp. 1861–1865, 2006.

[55] V. Roussev, Y. Chen, T. Bourg and G. Richard, `md5bloom`: Forensic file system hashing revisited, *Digital Investigation*, vol. 3(S1), pp. 82–90, 2006.

[56] V. Roussev and G. Richard, Breaking the performance wall: The case for distributed digital forensics, *Proceedings of the Fourth Digital Forensic Research Workshop*, 2004.

[57] V. Roussev, G. Richard and L. Marziale, Multi-resolution similarity hashing, *Digital Investigation*, vol. 4(S1), pp. 105–113, 2007.

[58] V. Roussev, G. Richard and L. Marziale, Class-aware similarity hashing for data classification, in *Advances in Digital Forensics IV*, I. Ray and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 101–113, 2008.

[59] V. Roussev, L. Wang, G. Richard and L. Marziale, A cloud computing platform for large-scale forensic computing, in *Advances in Digital Forensics V*, G. Peterson and S. Shenoi (Eds.), Springer, Heidelberg, Germany, pp. 201–214, 2009.

[60] P. Sanderson, Mass image classification, *Digital Investigation*, vol. 3(4), pp. 190–195, 2006.

[61] A. Savoldi and P. Gubian, Data recovery from Windows CE based handheld devices, in *Advances in Digital Forensics IV*, I. Ray and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 219–230, 2008.

[62] A. Schuster, Searching for processes and threads in Microsoft Windows memory dumps, *Digital Investigation*, vol. 3(S1), pp. 10–16, 2006.

[63] A. Schuster, The impact of Microsoft Windows pool allocation strategies on memory forensics, *Digital Investigation*, vol. 5(S1), pp. 58–64, 2008.

[64] M. Shannon, Forensic relative strength scoring: ASCII and entropy scoring, *International Journal of Digital Evidence*, vol. 2(4), 2004.

[65] M. Sieffert, R. Forbes, C. Green, L. Popyack and T. Blake, Stego intrusion detection system, *Proceedings of the Fourth Digital Forensic Research Workshop*, 2004.

[66] H. Simon, *Administrative Behavior*, Macmillan, New York, 1947.

[67] J. Slay and A. Przibilla, iPod forensics: Forensically sound examination of an Apple iPod, *Proceedings of the Fortieth Hawaii International Conference on System Sciences*, 2007.

[68] J. Solomon, E. Huebner, D. Bem and M. Szezynska, User data persistence in physical memory, *Digital Investigation*, vol. 4(2), pp. 68–72, 2007.

[69] A. Spruill and C. Pavan, Tackling the U3 trend with computer forensics, *Digital Investigation*, vol. 4(1), pp. 7–12, 2007.

[70] C. Swenson, G. Manes and S. Shenoi, Imaging and analysis of GSM SIM cards, in *Advances in Digital Forensics*, M. Pollitt and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 205–216, 2005.

[71] P. Turner, Unification of digital evidence from disparate sources (digital evidence bags), *Proceedings of the Fifth Digital Forensic Research Workshop*, 2005.

[72] P. Turner, Selective and intelligent imaging using digital evidence bags, *Digital Investigation*, vol. 3(S1), pp. 59–64, 2006.

[73] R. van Baar, W. Alink and A. van Ballegooij, Forensic memory analysis: Files mapped in memory, *Digital Investigation*, vol. 5(S1), pp. 52–57, 2008.

[74] C. Vaughan, Xbox security issues and forensic recovery methodology (utilizing Linux), *Digital Investigation*, vol. 1(3), pp. 165–172, 2004.

[75] M. Weier, Hewlett-Packard data warehouse lands in Wal-Mart's shopping cart, *InformationWeek*, August 4, 2007.

[76] S. Willassen, Forensic analysis of mobile phone internal memory, in *Advances in Digital Forensics*, M. Pollitt and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 191–204, 2005.