International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 3 Issue: 8 5204 - 5208

A Conceptual Framework on Digital Forensics Readiness for Criminals Tracking: Data Reduction Modalities

Onyemauche U.C.
Department of Computer Science
Nnamdi Azikiwe University
Awka Anambra State, Nigeria.
Osigwe.uchenna@yahoo.com

Nwosu Q.N .
Department Of Physical Health Sciences
University Of Nigeria
Nsukka Enugu State, Nigeria.
goodluckokwuchi@yahoo.com

Mbanusi, C.E.
Department of Computer Science
Nnamdi Azikiwe University Awka
Anambra State, Nigeria.
Chary4sam@yahoo.com

Abstract:- The ever-growing threats of fraud and security incidents present many challenges to law enforcement and organizations across the globe. The volume of digital forensic evidence is rapidly increasing, leading to large backlogs. However, Digital Forensic Data Reduction and Data Mining Framework is proposed. The framework outlined is not suggested to replace full analysis, but serves to provide a rapid triage, collection, intelligence analysis, and review and storage methodology to support the various stages of digital forensic examinations. This study contributes to the greater body of knowledge on the design and implementation of a digital forensic readiness programme, aimed at maximizing the use of digital evidence in an organization.

Keywords*: digital forensic readiness, Digital Forensics, Data Ming, data integrity.*

*****

1. INTRODUCTION

The increase in digital evidence presented for analysis to digital forensic laboratories has been an issue for many years, leading to lengthy backlogs of work. This is compounded with the growing size of storage devices. Theincreasing volume of data has been discussed by various digital forensic scholars and practitioners such as McKemmish (2013). While many of the challenges posed by the volume of data are addressed in part by new developments in technology, the underlying issue has not been adequately resolved. Over many years, there have been a variety of different ideas put forward in relation to addressing the increasing volume of data, such as data mining. This paper investigates recent challenges that technology presents with regard to the reliance and admissibility of electronic evidence in a court of law. A systematic literature review was used to gather relevant information and this data is critically analyzed in order to identify gaps and to improve upon them. A section dedicated to explaining the scientific research method adopted in this paper is presented next. This is followed by a section on the application of the said research method, in reviewing existing literature relating to digital forensics. Preceding the conclusion is a section that presents the conceptual model for DFR. The value of extracting or using intelligence from digital forensic data has not been discussed, nor has there been any research regarding the use of open, closed and confidential source information during digital forensic analysis. The escalation in volume and number of devices impacts forensic examinations in many ways, including increasing lengths of time to create forensic copies and conduct analysis, which contributes to the increase in the backlog of requests. Digital forensic practitioners, especially those in government and law enforcement agencies, will continue to be under pressure to deliver more with less especially in today's economic landscape. This gives rise to a variety of needs, including:

- A more efficient method of collecting and preserving evidence. A capacity to triage evidence prior to conducting full analysis. Reduced data storage requirements.
- An ability to conduct a review of information in a timely manner for intelligence, research and evidential purposes.
- An ability to archive important data.
- An ability to quickly retrieve and review archived data.
- A source of data to enable a review of current and historical cases
- (intelligence, research and knowledge management).

2. RELATED LITERATURE

The issue of the volume of data required to be analyzed in a digital forensic examination has been raised over many years. McKemmish (2013) stated that the rapid increase in the size of storage media is probably the greatest single

challenge to forensic analysis. In the interim years, there have been many publications stating the increasing volume of data is a major issue for forensic analysis. However, there have been no overall solutions proposed and the problem is still discussed. Alzaabi et al.(2013) discuss the ongoing trend of storage capacity increasing and the prices of devices decreasing, and while there are tools and techniques to assist an investigator, the time and effort to undertake analysis remains a serious challenge. For example, Raghavan (2013) stated that the 'exponential growth of technology has also brought with it some serious challenges for digital forensic research' and he suggests that this is the 'single largest challenge to conquer' .When discussing the challenges posed to the field of digital forensics, Spafford (2014) stated that digital technology continues to change rapidly. Terabyte disks and decreasing time to market are but two symptoms that cause investigators difficulty in applying currently available analytical tools. Moore's Law is the observation that the number of transistors on an integrated circuit doubles every eighteen to twenty four months and that this assists in predicting the development of technology as cited in Wiles et al. (2007). Kryder (2005) observed that in the space of under 15 years, the storage density of hard disks had increased 1,000 fold, from 100 million bits per square inch in 1990, to 2005 when 110 gigabit drives were released by Seagate. Kryder's Law can equate to the storage density doubling every 12 months, holding true since 1995. This is about twice the pace of Moore's Law . While storage capacity is doubling every year, the capacity to process data is only doubling every 18 to twenty four months, leading to an ever- growing gap in the capability to process the volume of data seized using processing power alone. Over the past decade, well-understood procedures and methodologies have evolved within computer forensics digital evidence collection. Kenneally et al.(2004) further noted that "Computer forensic autopsies are no longer performed on single machines with small data storage capacities. Rather, the scope for potential evidence has expanded to networks of interconnected computers, each with vast storage capacities containing potential artifacts of legal relevance". Available literature relating to digital forensic readiness (DFR) addresses various technical components of this concept, but none brings all the components into one framework. The need for a consolidation of research efforts in creating frameworks and models that help to address recent threats was recently identified by Garfinkel (2009), who opined that "without a clear strategy for enabling research efforts that build upon one another, forensic research will fall behind the market, tools will become increasingly obsolete, and law enforcement, military and other users of computer forensics products will be unable to rely on the results of forensic analysis".

The need for a more cost effective approach

There is an opportunity to consider methods to reduce the volume of data at each stage of the forensic analysis process in relation to the seven needs listed in the introduction, namely faster collection, reduced storage, timely review, intelligence, research, knowledge management, archive and retrieval. Consideration can be given to the type of data collected, stored and reviewed, with a focus on data that will provide the greatest information. Keneally et al.(2005) outlined a process for selective imaging to address the risks associated with collecting full forensic images for large drives, primarily the cost in time and resources, by selecting which data to image at the collection stage. The legal standards of reasonableness and relevance are raised to address concerns in relation to not undertaking analysis of a full forensic image. However, it could be argued that as the difference relates to hours or days, in a criminal or civil arena, it could be deemed reasonable to take a full bit-for-bit image and conduct analysis with all available and potentially relevant data. Hence, the proposed framework retains full imaging and analysis steps, with the reduced collection and review steps included to assist and support full analysis, rather than replace it. Turner (2005) introduced the concept of Digital Evidence Bags as a method to store a variety of digital evidence while retaining information relating to the source and location of the data subset. Schatz (2006) introduced the concept of a Sealed Digital Evidence Bag, providing for referencing between evidence bags. Commercial forensic software now provides the capability of selectively imaging files to support the collection of subset data into logical evidence files. Garfinkel (2006) discusses Forensic Feature Extraction (FFE) and Cross Drive Analysis methods. FFE is outlined as a scan of a disk image for email addresses, message information, date and time information, cookies, social security and credit card numbers. The information from the data scan is stored as XML for analysis and comparison. However, as the original data is interpreted, there may be instances where new techniques are difficult to apply to the original or historical data. There have been many developments in recent years whereby additional information is able to be extracted from data holdings that were previously unknown.

For example, Windows Registry analysis methodologies include newly discovered areas for locating information as stated by Carvey (2011).

DIGITAL FORENSIC READINESS OVERVIEW

Rapid changes and advances in technology and related crimes have given rise to the need to review and improve on digital forensic models and processes. Gravetter et al.(2005) also make the observation that "unlike other forensic sciences, digital forensics subject matter continues to evolve, as do the techniques". Given recent advances in technology, Bell et al.(2013) argue that it would be imprudent and potentially reckless to rely on existing evidence collection processes and procedures. They add "conventional assumptions about the behavior of storage media are no longer valid". Unlike traditional storage media, modern storage devices can operate under their own volition in the absence of computer instructions. Such operations can be highly destructive of traditionally recoverable data. This process has the potential to contaminate evidence and can obfuscate and make validation of digital evidence difficult. For purposes of this study, the use of the term "traditional approaches" denotes forensic procedures undertaken from the dawn of the computer forensic practice.

2.1 Common Techniques

The techniques used to perform data mining for detecting criminals that commit digital fraud range from simple statistical averages to complex neural networks and cluster analyses.

2.2 Advanced Statistical Techniques

Work has been done in statistical and computer science field on advanced methods for fraud detection. These methods include Bayesian Networks, genetic algorithms, state transition analysis, rule matching, and cluster analysis.

3. PROPOSED DIGITAL FORENSIC READINESS DATA DIMINUTION FRAMEWORK

A detailed search of relevant databases was conducted. The relevance was determined by using the library's A-Z list of electronic resources [18]. From this, only seven databases containing the most relevant material were selected and analyzed further for articles and other publications. The databases were selected on the basis of being classified under the following categories:

i. Multidisciplinary;
ii. Computing;
iii. Bio Sciences;
iv. Library Science; and v. Engineering.

Furthermore, the databases that were used were the ones containing the majority of the search hit results. The search term used was "digital forensic". This keyword was used as the basis of the search as it relates directly to the topic under investigation.

Only English written material published in the last five years (2009-2014) was considered. The reasons for this were that, firstly, Unisa's online library is available in English and secondly, English is one of Nigerian's most commonly spoken language in business, politics and the media. As there was no law on digital crimes in South Africa prior to the Electronic Communications and Transactions Act in 2002, only articles written after promulgation of this law were taken into consideration. The decision for reviewing only articles was based on the logic that articles usually precede books, dissertations and theses. Therefore, by looking at articles, content from the latter is also covered. The next section deals with the methodology for screening articles for inclusion.

3.1. Screening of articles for inclusion according

Since the application of a systematic literature review was intended not only for publication purposes but also for instrumental utilization, an additional task to increase the reliability of the screening process was undertaken. Both the authors conducted the screening process on a subset of articles independently of each other and then met together to compare results. In order to ensure that this process was scientific, the Cohen's Kappa (K) interpreter was used in measuring reliability of this process. Interpreter reliability is the degree of agreement between two observers who have independently observed and recorded behaviors at the same time. The basic formula for Cohen's Kappa (K) used is as computed below:

$$\text{Cohen's Kappa} = \frac{PA\ (0.77) - PC\ (0.50)}{1 - PC\ (0.50)} = 0.54$$

Where PA is the observed percentage agreement and PC is the percentage agreement expected. The goal in this study was to produce a ***PA*** value above

75% from the total reviewed articles. This was done to ensure that all relevant articles were included for detailed review and to archive a kappa value above 0.50. The said kappa goal is generally considered to be satisfactory. Both authors met to calculate the inter rater reliability by calculating a percentage agreement. This process was repeated until the percentage agreement exceeded 75%. Abstracts of 459 articles were reviewed, resulting in the identification of 130 relevant articles for possible inclusion. The review process was refined further and the result was an agreement on the final 100 articles for inclusion.

4. RESULTS

Table 2 Data reduction applied to these data cases

Item	Multi disciplinary	Computing	Bio Sciences	Lib. Science	E01:HD ratio	L01:E01 ratio	L01: HD ratio
Smallest	1	40	4.5	.0415	11%	0.92%	0.10%
Largest	1	1000	121	.0143	12%	0.12%	0.01%
Total (all cases)	212	102396.5					
E01	107	45388	22040.68		51.1%		
L01	144	66438.5		62.98			0.196%
E01 & L01	37	9430	5197.9	22	55%	0.423%	0.233%
Average (across all)		461.4	136.79	0.44	58.7%	0.705%	0.196%

Evidence analysis

This step is common to digital forensic analysis and is well documented Evidence analysis is conducted as per standard methodology for files and data. In addition, the information gained from conducting the review and other source data can be used when conducting analysis of the full forensic image to locate data relating to an investigation, which may result in additional information being discovered. Evidential analysis can be undertaken to confirm the findings from the review of the subset data and to locate additional data of importance. Any additional data (not present in the subset files) can be preserved in a logical evidence container and included with the reduced subset store for archive or historical review.

5. CONCLUSION

The growth in digital forensic data has been ongoing for many years and with the predicted ongoing growth in technology and storage, is estimated to become increasingly larger over the coming years. This has led to large backlogs of evidence awaiting analysis. By utilizing the Digital Forensic Data Reduction Framework and a reduced subset of data, a greater understanding of data can be made at a substantially reduced cost, by comparison with storing full forensic images. The data reduction subset process can be used to triage devices and media to quickly assess which devices may contain potential evidence and hence should be examined as a priority, and which devices have less potential evidence and can be given a lower priority for full analysis.

References

[1] Alzaabi, T. Mckemmish, I.Y.(2013). Event sequence mining to develop profiles for computer forensic investigation purposes. ACSW Frontiers '06: Proceedings of the 2006 Australasian workshops on Grid computing and e- research, Pp 145–153.

[2] Bell, M, Jones A., Martin, T.A (2013). An ontology-based forensic analysis tool. Journal of Digital Forensics, Security & Law vol. 2013 Conference Supplement, Pp 121–135

[3] Carvey H.T.(2011). Windows registry forensics: Advanced digital forensic analysis of the Windows registry, Elsevier Publishers Burlington, MA. Pp50.

[4] Garfinkel S.A., Casey E.(2005). Digital forensics research: The next 10 years. Journal of Digital Investigation. Vol. 2. Issue 3. Pp 64–73.

[5] Garfinkel S.A. (2006). Forensic feature extraction and cross-drive analysis. Journal of Digital Investigation. Vol. 3. Issue 4, Pp 71–81.

[6] Gravetter, D.S., Farrell P, Roussev V & Dinolt G (2005). Bringing science to digital forensics with standardized forensic corpora. DFRWS 2009. Montreal,Canada. Retrieved From http://simson.net/clips/academic/2009.DFRWS.Corpora.pdf

[7] Kenneally E .F.(2004). Risk sensitive digital evidence collection. A Journal of Digital Evidence, Vol. 2, Issue 2, Pp 101–119.

[8] Kryder, L.J.(2005). High-speed search using Tarari content processor in digital forensics. A Journal Of Digital Evidence, Vol.3, Issue 3, Pp 91–95. McKemmish, R.K., Richard G & Roussev, V.(2013). Massive threading: Using GPUs to increase the performance of digital forensics tools. A Journal of Digital Investigation, Vol.2, Issue 4, Pp 73–81.

[9] McKemmish R.K.(2013). What is forensic computing? Trends & Issues in Crime and Criminal Justice, Canberra: Australian Institute of Criminology. http://aic.gov.au/publications/current%20series/tandi/101-120/tandi118.html

[10] Raghavan, K.T.(2013). Digital forensics: Defining a research agenda. System Sciences, HICSS'13. 42nd Hawaii International Conference on IEEE: Pp 1–6.

[11] Schatz, S., Erbacher, R.(2006). Improving the computer forensic analysis process through visualization. Commun. ACM Publishing Press, Vol.49, Issue2, Pp 71–75.

[12] Spafford, M.K.(2014). Forensic relative strength scoring: ASCII and entropy scoring. International Journal of Digital Evidence, Vol. 2, Issue 4, Pp 151–169.

[13] Turner P 2005. Unification of digital evidence from disparate sources(digital evidence bags). A Journal Of Digital Investigation, Vol.2, Issue 3, Pp223–228.