Big Data and Digital Forensics

Rethinking Digital Forensics

Oluwasola Mary Adedayo
ICSA, Department of Computer Science
University of Pretoria
Pretoria, South Africa
madedayo@cs.up.ac.za

***Abstract* — One of the main challenges in digital forensics is the increasing volume of data that needs to be analyzed. This problem has become even more pronounced with the emergence of big data and calls for a rethink on the way digital forensics investigations have been handled over the past years. This paper briefly discusses the challenges and needs of digital forensics in the face of the current trends and requirements of different investigations. A digital forensics analysis framework that puts into consideration the existing techniques as well as the current challenges is proposed. The purpose of the framework is to reassess the various stages of the digital forensics examination process and introduce into each stage the required techniques to enhance better collection, analysis, preservation and presentation in the face of big data and other challenges facing digital forensics.**

Keywords — Digital forensics; Digital forensics examination process; Forensics analysis framework; Big data

I. INTRODUCTION

The field of digital forensics has received an increasing amount of attention in the past years as digital evidence found on different devices has become more and more valuable during investigations. Despite this fact, new challenges facing digital forensics are also emerging, calling for a need to reconsider some the ways in which digital forensics have been handled until now.

Some of the already known challenges in digital forensics include the use of encryption, the need to handle different types of devices, file formats and contents, the need to sometimes analyze incomplete and inconsistent data, the availability of anti-forensics tools, and the need for specialized ways of extracting information from some devices. Emerging challenges include the increasing number and size of storage capacity on many devices, the increasing volume of data that may be collected in relation to an investigation and the need to provide fast results during such analysis. The emerging challenges can be attributed to technological advances, the ability to interconnect various devices capable of generating volumes of data, the need to gather and investigate data found on databases (both relational and non-relational) as well as the need to conduct forensic analysis on information stored in the cloud. All of these emphasize a relationship between the field of data science and digital forensics and point out the need to analyze "big data" in digital forensics.

This work is based on the research supported by the National Research Foundation (NRF). Any opinion, finding and conclusion or recommendation expressed in this material is that of the author and the NRF does not accept any liability in this regard.

978-1-5090-6096-2/16/$31.00 ©2016 IEEE

Although there are existing standards and process models currently being followed in the digital forensics, researcher have identified the need to reevaluate the processes in a digital investigation [1], [2], [3], [4] and for research into new techniques and algorithms for handling the increasing volumes of data in an investigation [3], [4], [5] to enable efficient collection and preservation of evidence, rapid searching of data, analysis of data from different sources, speedy results from analysis and the reliability, scalability and automation of digital forensics tools.

Several researchers have also considered different ways of addressing the emerging challenges and solutions such as triage [4], [6], [7], [8], data mining [9], [10], [11], data reduction [5], [10], [12], distributed processing [3], [13], cross-drive analysis [12], the use of artificial intelligence [14], [15] and intelligent analytical techniques [5] have been proposed. However, many of these solutions focus on a subset of the digital forensics process in which it is relevant and do not reassess the complete digital forensics process. The purpose of this paper is to examine the solutions that have been proposed for the different challenges being faced by digital forensics in correlation with already established digital forensic process models. The paper reassesses each of the stages of the process and introduces additional techniques and steps that may be required to handle the current challenges into the framework.

The proposed framework contributes to existing frameworks by introducing more efficient ways of collecting, preserving, analyzing and presenting information during an investigation despite the current challenges faced by digital forensics.

The rest of the paper is organized as follows. Section II briefly discusses digital forensics and the existing process models. The concept of big data as well as the challenges faced in the different stages of the digital forensics process from the perspective of big data is also discussed in the section. In section III the proposed process model is described. Section IV discusses the framework and points out some of the advantages and disadvantages of the model. Section V gives the conclusion and future work.

II. BACKGROUND

This section briefly discusses the field of digital forensics and the processes involved in a forensic investigation. It also describes big data and highlights some of the challenges being faced in digital forensics from the perspective of big data.

A. Digital Forensics

Over the last four decades, the field of digital forensics has grown from being a minor part of criminal investigations to an important aspect of many investigations containing digital information [4], [5]. Digital forensics is defined by Palmer as: "the 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 facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations" [11].

Since 2001, various researchers have proposed different frameworks and process models for digital investigations [11], [16], [17], [18]. From a general overview of these models, digital forensics consists of the following processes:

- Preparation: this involves the preparation of the tools, equipment and people required during an investigation together with the necessary authorizations or approval to collect data. The purpose of this stage is to ensure that digital evidence can be efficiently and correctly collected when the need arises.
- Preservation: this involves the isolation and securing of the state of the physical and digital evidence at the crime scene. Necessary actions are taken to maintain the integrity and authenticity of evidence during the investigation.
- Collection: in the collection stage, physical and digital evidence to support the investigation is recorded and collected using standardized techniques.
- Examination: during examination, the collected evidence is searched to identify and locate data relevant to the investigation.
- Analysis: based on the relevant data collected, analysis is performed to determine the significance of the data and make conclusions.
- Presentation: summarization and explanation of the findings and conclusions drawn during the examination and analysis stages.

In the rest of the paper, the challenges and the proposed framework will be addressed in relation to the processes described above. Although some of the existing models sometimes incorporate stages or steps not mentioned above, the key part of the process models are similar to the stages described above.

B. Big Data

The increasing volume of data that needs to be analyzed during a digital forensic investigation has been raised by researchers over many years [11], [19], [20]. The interesting fact is that the data considered to be large a decade ago is almost insignificant when compared to the current volume of data encountered today. This issue can be viewed from two perspectives when considering its impact on digital forensics: first is the increasing rate at which large volumes of data can be created on a variety of devices (big data) and second is the increasing size of storage media capacity that is encountered during investigations.

Big data can be defined as datasets that are too big, too fast and too difficult to be processed using relational databases or existing tools [21], [22], [23]. It is characterized by the volume, velocity, variety and variability of data [21], [22], [23], [24]:

- Volume: refers to the huge amount of data that can be generated. This calls for scalability in terms of data storage as well as the need for a distributed approach in processing data.
- Velocity: refers to the speed at which data is generated or moved around. The velocity of big data implies that data acquisition and analysis need to be conducted expeditiously in order to maximize the value of the data at any point.
- Variety: describes the different types of data; structured, semi-structured and unstructured that can be collected. In the past, the focus has been on analyzing structured data that can be easily stored in relational databases. With the rise of unstructured and semi-structured data, there is a need for better storage and analysis of data based on data type.
- Variability: refers to the inconsistency in the volume, velocity and variety of data over time. Data processes may generate peak data load at different times. It is necessary that such peak loads be handled when it comes to storage and processing.

Other characteristics of big data that have been described by researchers [23], [24] include: veracity (the accuracy and trustworthiness of data), value (potential value that can be generated from analyzing the data) and validity (suitability of the data for the planned use).

Although digital forensics faces challenges because of the volume, velocity, variety and variability of data being generated today, it seems the biggest challenge stems from the storage of all these data and the rapidly increasing capacity of storage media and this has been identified by researchers over many years [11], [19], [20].

According to Moore's law, computing power doubles every 18 – 24 months. However, the amount of data being stored as well as that of storage capacity is doubling every 9 – 12 months [25], [26] causing a wide gap between computing power available and the amount of data that can be collected for analysis. Although there are digital forensics tools and techniques that can be employed in an investigation, the time and effort required to perform an analysis is still a major challenge that we may never outpace [2]. The storage of data in the cloud further exacerbates this problem as it provides a means of storing very large amounts of data and presents its own challenges during an investigation [27].

C. Challenges in Digital Forensics

This section describes some of the challenges facing digital forensics when dealing with big data. The challenges are

described by focusing on the stages of the digital forensics process identified above.

- Preparation: The major challenge faced in the preparation stage revolves around the volume and variety of data, as well as the diversity of devices on which evidence may be found. Although ensuring that the necessary standards, policies and procedures to follow during an investigation are in place may be fairly easy; training the investigator and having the right tools for every situation is still a challenge. Ensuring that an investigator and/or the investigators tools are prepared to handle every application, operating system, mobile device, protocol, file format, encryption, as well as cloud data is almost an impossible mission.
- Preservation: Ensuring that the integrity and authenticity of evidence is preserved throughout an investigation also presents the challenge of knowing how to handle different devices. Although the techniques and standard practices to ensure that data is preserved may not change, having the appropriate tool for a device or media (e.g. a mobile phone) may still be a challenge. Also, with the increasing volume of data, the time involved in preservation becomes significantly larger leading to higher response times during an investigation.
- Collection: The collection, examination and analysis stages of the digital forensics process face the biggest challenge as they are affected by the growing volume, variety and variability of data. Although there has been a decrease in the price of storage devices over the years, storing huge volumes of data in an uncompressed manner still has significant cost involved. On the other hand, if the data is compressed, the cost involved in the collection is reduced but this implies that the data is not readily available for an analysis [10]. The main challenge in data collection however comes when the evidence to be collected is stored in the cloud. Whereas traditional investigations often involve making a full forensic image of the media, this is not possible when the data is stored in the cloud since it may be stored remotely in locations not accessible to the investigator (due to legal or jurisdictional issues) or it may be so large that it is stored on the cheapest (and least accessible) media possible. In the face of big data, collecting all possible evidence in an investigation is often impossible.
- Examination and Analysis: examining and analyzing data during an investigation involving large volumes and a wide variety of data is a challenging task. Although computing power is increasing according to Moore's law, data volume in increasing at a much faster rate leading to backlogs in examination and analysis of many investigations [27]. Another problem in this phase is the fact that many of the techniques used in traditional digital forensic analysis such as string search, pattern matching and text mining are not suitable for the current problem space of digital forensics mainly because the techniques do not scale and the available computing power is underutilized [4], [5]. The issue of false positives in the examination and analysis stage also results in longer processing times [5]. The need to analyze volumes of data in a short amount of time is becoming more and more important and new techniques or algorithms are required to meet this need.

  The wide variety of data that could be encountered in an investigation also poses a challenge during the examination and analysis phases as digital forensic tools are limited in the number of file formats and devices that they can handle. Digital forensic tools need to be able to process old, current and emerging data formats and devices, and this is a major challenge during an investigation.
- Presentation: The purpose of the presentation phase is to disseminate the finds and conclusions from an investigation in a way that is understandable to an audience or the court. The main challenge in this may be identifying the most appropriate way to describe the techniques and processes used in the examination and analysis of such large volumes of data. Justification of effectiveness and correctness (or validity) of such techniques and processes may also be required for evidence to be acceptable.

III. Rethinking The Digital Forensic Process Model

Considering the challenges identified above, it is essential to reevaluate each stage of the digital forensic process model to ensure that the challenges are dealt with. Researchers have emphasized the problem of increasing data volume in digital investigations and various techniques that can be applied at different phases of an investigation have been proposed [4], [5], [10], [13], [15].

The objective of this section is to present a holistic view of the digital forensic process with a focus on the steps and techniques that may be applied in each phase of the process in order to handle the current challenges. Figure 1 presents the proposed framework. It is important to note that even though we have referred to this as a framework, it is not intended to be a new stand alone model, but rather, one that explains the steps and techniques that may be followed at different stages of existing models.

A discussion of the steps outlined in the framework follows.

A. Step 1 - Preparation

The first phase of the framework is the preparation phase and it focuses on ensuring that the tools, equipment and investigators required for the investigation are available as in existing digital forensics process models.

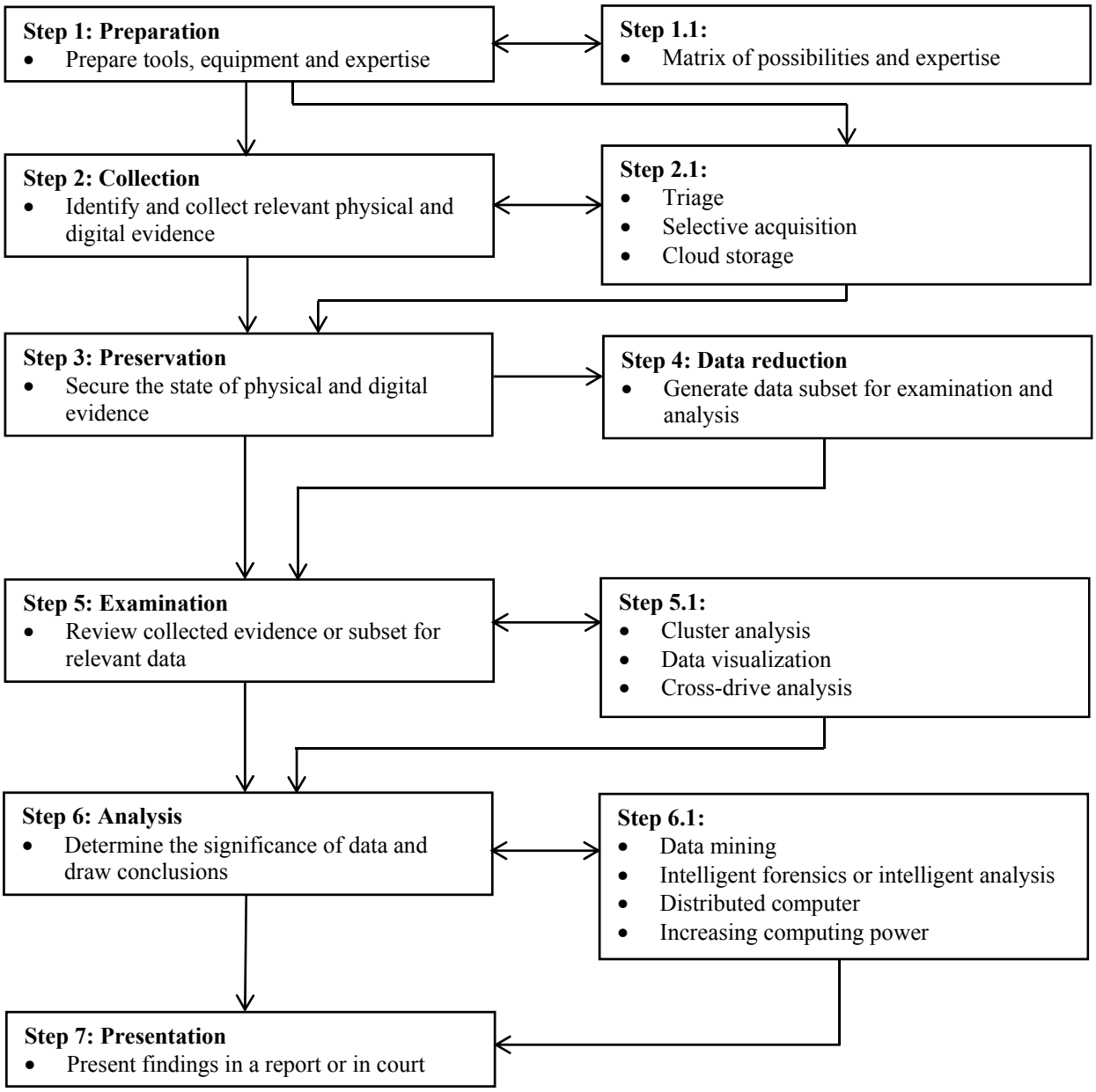

Fig. 1. Proposed Digital Forensic Framework

An approach that has been suggested before starting an investigation is the use of a matrix that lays out the possible scenario of the crime scene, digital evidence and the suspect, and specifies the required skills for the investigators on the investigation team [6]. The purpose of the matrix is to determine the aim of the investigation by identifying what is known or unknown in the investigation. Understanding the aim of the investigation and expertise available will assist in preparing for additional expertise where necessary. Understanding the unknowns in an investigation will also assist in identifying who may be approached to aid the investigation.

From a proactive perspective, it is important that the laws protecting information in a particular jurisdiction be understood prior to the selection of a cloud storage to ensure that such information can be retrieved if required.

B. Step 2 - Collection

The second step of the framework involves the collection of potential evidence for the investigation. To handle the problem of large data volumes that needs to be collected during an investigation. Techniques such as triage [4], [28], [6], [29] sampling [2], [4], [30], selective and intelligent acquisition [5],

[31], [32], [33] have been proposed to limit the amount of data that has to be collected.

Triage is a process in which pieces of evidence or potential sources of evidence are ranked in the order of importance or priority so that forensic tasks are sequenced properly during an investigation [4], [6]. Sources of evidence such as storage media and devices are prioritized based on the relevant data stored in them which are retrievable within a short time or that have a short life span (e.g. volatile data). Mislan et al. [29] specify some guidelines that may be followed for a triage inspection.

Another approach that can be used to handle huge volumes of data during collection is sampling of the media so that only the data in certain sections have to be collected. Using statistical techniques, it is possible to gain knowledge about the content of the media and reducing the volume of data collected allows collection and subsequently examination and analysis of the data to be done faster [4], [30]. Although there is a probability that small pieces of data relevant to an investigation may be missed when using sampling techniques, the probability can be further reduced by collecting more sectors or increasing the size of the blocks collected [2].

The use of selective and intelligent acquisition is another way of reducing data volume which has proposed by various researchers [5], [31], [32], [33]. Selective acquisition involves the collection of logically selected file types or categories of data based on the specific investigation. Kenneally and Brown [33] discuss the costs associated with collecting evidence completely and selective acquisition. They posit that completeness of collected should not be measured by the volume of data collected but rather by the legal standards of reasonableness and relevance. Using this technique, Turner [31] proposed the use of a digital evidence bag.

One obviously beneficial application of these techniques is in the retrieval of evidence from cloud storage. Triage, sampling and selective and intelligent acquisition can be applied in the collection of cloud data to reduce collection time, storage space and time required for analysis. In situations where some data resides in a jurisdiction different from that of the investigator, the techniques may also be useful in identifying important data that must be collected.

Lastly, even though there are concerns about the use of cloud storage for digital evidence, and very little research has been conducted to investigate how cloud storage affects digital evidence, the use of cloud storage for digital evidence is a possible way of handling large data volumes in an investigation. However, research to aid a better understanding of the associated legal requirements and acquisition techniques to be followed is still required in this regard.

C. Step 3 - Preservation

The preservation of collected data is similar to the steps and techniques involved in existing frameworks. However, in order to handle large volumes of data and diversity of devices in an investigation, this step may be preceded by step 3.1, which includes the techniques that may be applied in the collection phase described above.

D. Step 4 – Data Reduction

Many of the techniques that have been proposed for reducing data volume in an investigation are often applied as the initial phase of data collection. The implication of these techniques is that a full forensic image of the device or media is not collected. In investigations that need to obtain a full forensic image, the data reduction step can be used to reduce the amount of data that has to be analyzed. Quick and Choo [10] proposed a data reduction framework in which a full forensic image is collected before the data volume is reduced.

The reduction process ensures that the evidence is preserved and is rendered usable by filtering out a data subset with the greatest potential for the investigation, using write blockers and other forensic tools. The advantage of the reduction step is that an investigator can reassess the full forensic image in the event that trace evidence is missed in the reduction process. Another advantage of the reduction step is that separate subsets can be obtained for different information that needs to be examined [10] and the examination and analysis of data subsets can be faster than in a full forensic image.

E. Step 5 - Examination

The fifth step of the framework involves searching collected evidence or a subset to locate relevant data. The need here is for intelligent algorithms that work better than techniques currently used in investigation (e.g. string search and pattern matching) and which reduces retrieval overhead so that examination can be completed fast enough despite increasing data volumes. Some of the techniques that have been proposed to achieve this objective include cluster analysis [34], data visualization [4], outlier analysis [4], [34], [35] and cross-drive analysis [12].

Cluster analysis involves the use of neural networks for clustering search results so that the number of hits that has to be analyzed further is reduced [34]. Data visualization assists an investigator to visually interpret data and aids the process of anomaly detection [4], [36], [37]. This can assist in prioritizing analysis and reduce the time taken to search for relevant evidence. Garfinkel [4] pointed out the need for forensic tools to integrate interactive visualization in analysis techniques.

Another approach that can be applied to quickly identify data that may be of relevance to an investigation is outlier analysis. Carrier and Spafford [35] describe an automated technique in which collected data is searched against target evidence. The target evidence could be determined based on existing evidence or through the use of outlier analysis which determine data that are hidden or which are different from those in their surroundings. The technique employs several data mining techniques and can also be applied in the data reduction step to identify data subsets relevant to a particular question in an investigation.

Cross-drive analysis [12] involves the use of statistical techniques for the correlation of data on multiple disk images. In investigations involving multiple drives, the techniques can be applied for several reasons, such as to identify which drive has the most relevant information for the investigation, identify associations based on the information on two or more drives,

aid in the discovery of previously unknown associations and to collate information from different drives into one containing the most relevant information to an investigation [12].

One advantage of these techniques during data collection is that they provide a way of improving the automation of digital forensics tools and enhancing the examination process.

F. Step 6 - Analysis

The analysis stage of the digital forensics process is possibly the stage most affected by the increasing variety, variability, device diversity and data volume, as the need to complete investigations in an expeditious manner becomes more pressing. In order to handle these challenges, techniques such as data mining, artificial intelligence or intelligent forensics, distributed computing, use of graphical processing unit (GPU) or other ways of improving available processing power during an analysis have been proposed by various researchers [5], [9], [11].

Data mining is a combination of techniques such as artificial intelligence, statistical modeling, machine learning, pattern recognition, data visualization and database processes that can be used to explore large amounts of data with the aim of extracting useful information from the dataset. Beebe and Clark [9] describe the use of data mining in digital investigation analysis and show the application of data mining techniques in different investigations. Data mining can be applied to digital forensics to improve the response time in an analysis as well as reduce associated cost and the level of manual analysis that needs to be completed by an investigator. Since the underlying concept in data mining is well understood and the field is well developed, techniques applied to digital forensic investigations can be better presented or explained in a court of law during a case. Various techniques in data mining that have been applied to analysis include discriminant analysis, rule mining and cluster analysis [9].

A similar technique that can be employed in the analysis phase is often referred to as intelligent forensics or intelligent analysis. It involves the use of tools and techniques such as artificial intelligence, computational modeling and social network analysis in digital investigations with the aim of reducing the amount of time involved in analysis [15], [5], [38]. Although there is relatively little research on intelligent analysis, there is potential for applying this technique.

The use of distributed computing or graphical processing unit (GPU) to improve the performance of digital forensic tools is another way in which the amount of time required for a digital forensic analysis can be reduced [13], [39]. Although there is very little amount of research has been done of this approach as well, the benefits of distributed computing in handling processes requiring large amounts of computer power as well as the amount of power in today's GPUs is well known and there is potential of gaining such benefits and computer power for digital forensic tools.

G. Step 7 - Presentation

The presentation step involves the communication of findings during an investigation. Although this step is similar to the presentation step in existing digital forensic frameworks, an additional step proposed herein is the explanation of the validity and acceptability of the techniques used during the investigation. Detailed information on the techniques and algorithms employed as well as their implication on the data will be of benefit during presentation.

IV. Discussion of the Framework

The aim of this section is to position the proposed framework in relation to existing digital forensic process models.

As there are already quite a number of digital forensic process models, our aim is not to create yet another one as most of the existing models already describe the major steps involved in an investigation. However, the framework in this paper goes a step further by mapping various techniques that have been proposed for use or applied during a digital forensic investigation to the appropriate step where such a technique can be used in existing models. To a large extent, the main steps identified in the framework are based on an amalgamation of most of the existing frameworks [11], [16], [18], [17]. Although our aim in this paper is not to show how the steps in the framework match the steps in the different existing frameworks, we posit that the main steps in the framework can be easily correlated with those in most of the existing digital forensic models. The purpose of the steps identified in the framework is to generalize the framework as much as possible. However, the sub-steps identified can be incorporated into the appropriate stage in any existing model.

In addition, the framework provides a general mapping of the digital forensic process in the big data environment. Although a practical demonstration of this is not provided, a discussion of how each of the 4 V's of big data is handled is given in the description of the framework in section III. The framework also provides an overview of the processes and techniques that can be applied especially when dealing with huge data volumes and is intended to complement existing models.

It is important to note that the exact techniques employed at any sub-step of the framework will be dependent on the unique situation in an investigation. Although we have highlighted different techniques that may be used each step, the applicability of each in a particular situation is a decision that has to be made by the investigator. Depending on the investigation, it is also possible that there may be other applicable techniques at each stage of the framework.

V. Conclusion and Future Work

The need to rethink the steps and techniques used during a digital forensic investigation is important because of the current challenges facing digital forensics, especially with regards to increasing data volumes in an investigation.

Although various researchers have proposed different techniques that may be applied during an investigation, many of these techniques are discussed without much focus on the specific stage of digital forensics involved or how other stages fit in together with the techniques. In this paper we have described a framework that is intended to complement existing models by describing techniques that can be applied at various stages of an investigation to handle the current challenges in

digital forensic. The paper has described the notion of big data and its characteristics. The challenges faced at different stages of a digital forensic investigation with regards to handling big data have also been discussed. The proposed framework has brought together various techniques proposed by different researchers to handle different challenges, into a model that can be incorporated into existing models.

Although we have discussed the techniques mentioned in the framework, research into the application and implication of some of the techniques is still required even though their potential in an investigation are clearly visible.

REFERENCES

[1] A. Gaurino, "Digital Forensics as a Big Data Challenge," in *ISSE 2013 Securing Electronic Business Processes: Highlights of the Information Security Solutions Europe 2013 Conference*, Springer Fachmedien Wiesbaden, 2013, pp. 197-203.

[2] S. L. Garfinkel, "Digital Forensics Innovation: Searching A Terabyte of Data in 10 minutes," in *DCACM 2013*, Washington DC, 2013.

[3] G. G. Richard and V. Roussev, "Next generation digital forensics," *Communications of the ACM,* vol. 49, no. 2, pp. 76-80, 2006.

[4] S. L. Garfinkel, "Digital forensics research: The next 10 years," *Digital Investigation,* vol. 7, no. Supplement, pp. S64-S73, 2010.

[5] N. Beebe, "Digital Forensic Research: the good, the bad and the unaddressed," in *Advances in Digital Forensics*, 2009.

[6] M. K. Rogers, J. Goldman, R. Mislan, T. Wedge and S. Debrota, "Computer Forensics Field Triage Process Model," *Journal of Digital Forensics, Security and Law,* vol. 1, no. 2, pp. 19-38, 2006.

[7] E. Casey, M. Ferraro and L. Nguyen, "Investigation Delayed Is Justice Denied: Proposals for Expediting Forensic Examinations of Digital Evidence," *Journal of Forensic Sciences,* vol. 54, no. 6, pp. 1353-1364, 2009.

[8] R. P. Mislan, E. Casey and G. C. Kessler, "The growing need for on-scene triage of mobile devices," *Digital Investigation,* vol. 6, no. 3-4, pp. 112-124, 2010.

[9] N. Beebe and J. Clark, "Dealing with Terabyte Data Sets in Digital Investigations," in *Advances in Digital Forensics*.

[10] D. Quick and K.-K. R. Choo, "Data reduction and data mining framework for digital forensic evidence: Storage, intelligence, review and archive," *Trends & Issues in Crime and Crimial Justice,* vol. 480, pp. 1-11, 2014.

[11] G. Palmer, "A Road Map for Digital Forensic Research," Utica, New York, 2001.

[12] S. L. Garfinkel, "Forensic feature extraction and cross-drive analysis," *Digital Investigation,* vol. 3, no. Supplement, pp. 71-81, 2006.

[13] G. G. Richard and V. Roussev, "Breaking the performance wall: The case for distributed digital forensics," in *Proceedings of the 2004 Digital Forensics Research Workshop*, 2004.

[14] S. Mukkamala and A. H. Sung, "Identifying Significant Features for Network Forensic Analysis Using Artificial Intelligent Techniques," *International Journal of Digital Evidence,* vol. 1, no. 4, 2003.

[15] A. Irons and H. S. Lallie, "Digital Forensics to Intelligent Forensics," *Future Internet,* vol. 6, pp. 584-596, 2014.

[16] N. L. Beebe and J. G. Clark, "A hierarchical, objectives-based framework for the digital investigations process," *Digital Investigation,* vol. 2, no. 2, pp. 147-167, 2005.

[17] S. Ó. Ciardhuáin, "An Extended Model of Cybercrime Investigations," *Internation Journal of Digital Evidence,* vol. 3, no. 1, 2004.

[18] B. D. Carrier and E. H. Spafford, "An event-based digital forensic investigation framework," in *Proceedings of the 2004 digital forensic research workshop (DFRWS)*, 2004.

[19] R. McKemmish, "What is Forensic Computing?," *Trends and Issues in crime and criminal justice,* no. 118, Australian Institute of Criminology, June 1999.

[20] S. Raghavan, "Digital forensic research: current state of the art," *CSI Transactions on ICT,* vol. 1, no. 1, pp. 91-114, 2013.

[21] S. Madden, "From Databases to Big Data," *IEEE Internet Computing,* vol. 16, no. 3, pp. 4-6, 2012.

[22] M. Chen, S. Mao and Y. Liu, "Big Data: A survey," *Mobile Networks and Applications,* vol. 19, no. 2, pp. 171-209, 2014.

[23] NIST Big Data Public Working Group, "NIST Big Data Interoperability Framework: Volume 1, Definitions," National Institute of Standards and Technology, 2015.

[24] B. Marr, "Why only one of the 5 Vs of big data really matters," [Online]. Available: http://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters. [Accessed 1 March 2016].

[25] T. Coughlin, "High Density Hard Disk Drive Trends in the USA," 2001.

[26] M. Ruback, B. Hoelz and C. Ralha, "A New Approach for Creating Forensic Hashsets," in *Advances in Digital Forensics VIII: 8th IFIP WG 11.9 International Conference on Digital Forensics*, Springer Berlin Heidelberg, 2012, pp. 83 - 97.

[27] G. Grispos, T. Storer and W. B. Glisson, "Calm Befor the Storm: The Challenges of Cloud Computing in Digital Forensics," *International Journal of Digital Crime n=and For,* vol. 4, no. 2, pp. 28-48, 2012.

[28] H. Parsonage, "Computer Forensics Case Assessment and Triage - Some ideas for discussion," 2009.

[29] R. P. Mislan, E. Casey and G. C. Kessler, "The Growing Need for On-scene Triage of Mobile Devices," *Digital Invetigation,* vol. 6, no. 3-4, pp. 112-124, 2010.

[30] J. Clayton, "Investigation Into A Digital Forensics Triage Tool Using Sampling, Hashes And Bloom Filters - SHAFT," 2012.

[31] P. Turner, "Unification of digital evidence from disparate sources (Digital Evidence Bags)," *Digital Investigation,* vol. 2, no. 3, pp. 223-228, 2005.

[32] P. Turner, "Selective and Intelligent Imaging Using Digital Evidence Bags," *Digital Investigation,* vol. 3, pp. 59-64, 2006.

[33] E. E. Kenneally and C. L. Brown, "Risk sensitive digital evidence collection," *Digital Investigation,* vol. 2, no. 2, pp. 101-119, 2005.

[34] N. L. Beebe and J. G. Clark, "Digital forensic text string searching: Improving information retrieval effectiveness by thematically clustering search results," *Digital Investigation,* vol. 4, pp. 49-54, 2007.

[35] B. D. Carrier and E. H. Spafford, "Automated Digital Evidence Target Definition Using Outlier Analysis and Existing Evidence," in *Proceeding of the Digital Forensics Research Workshop*, New Orleans, 2005.

[36] S. Teelink and R. F. Erbacher, "Improving the Computer Forensic Analysis Process Through Visualization," *Communications of the ACM,* vol. 49, no. 2, pp. 71-75, 2006.

[37] S. Chavhan and S. Nirkhi, "Visualization Techniques for Digital forensics: A Survey," *International Journal of Advanced Computer Research,* vol. 2, no. 4, pp. 74-78, 2012.

[38] B. W. P. Hoelz, C. G. Ralha and R. Geeverghese, "Artificial Intelligence Applied to Computer Forensics," in *Proceedings of the 2009 ACM Symposium on Applied Computing*, 2009.

[39] L. Marziale, G. G. Richard III and V. Roussev, " Massive threading: Using GPUs to increase the performance of digital forensics tools," *Digital Investigation,* vol. 4, no. Supplement, pp. 73-81, 2007.