A New Triage Process Model for Digital Investigations

Bo Yang[1,2], Ning Li[1,2], Jianguo Jiang[1],
1. Institute of Information Engineering, Chinese Academy of Sciences, China
2. Beijing Key Laboratory of Network Security Technology, Beijing, 100093 China
yangbo32@iie.ac.cn, lining6@iie.ac.cn, Jiangjianguo@iie.ac.cn,

***Abstract*—As the amount of crimes involving the digital dimension grows, an ever increasing number of digital artifacts collected from a forensic investigation needs plenty of resources to process in a forensically sound manner. Digital forensic triage provides a way to deal with this scalability problem, as it is tailored to maximize the utilization of resources based on a priority system. Unfortunately, the paucity of definite solutions limits efforts to triage implementation. In this article, we propose a Dual-Triage Digital Forensic Process Model, termed DTDFPM, which increases the effectiveness and efficiency of examinations. The DTDFPM simultaneously enjoys the following properties: i) background information is utilized to prioritize cases and specific features are determined which media contain information relevant to the investigation, ii) a Priority Sorting with Artificial Neuron algorithm (PSAN) is designed, which is the first application of neural network to sorting solution in digital triage, iii) efficient integration, the proposed model implemented based on the Python programming language can be easy to integrate into existing forensic tools. Thoroughly theoretical analysis and performance evaluation indicate the advantage of our proposed process model.**

Keywords—Digital forensics; digital triage; process model; case study

I. INTRODUCTION

With the advance of science and technology, most data are processed, transmitted and stored in digital form. Digital forensics faces the difficulty in increasing investigation efficiency of data processing. This inclination can be shown in the Regional Computer Forensics Laboratory (RCFL) Program Annual Reports [1]. Not only the number of cases but also the processed data keep growing rapidly on the basis of their past years statistics (as shown in Fig 1, the latest data available is from FY13). For example, the numbers in FY12 have gone up by 12 percent and 40 percent respectively comparing with FY11 (FY13 is an exception that occurs due to substantial budget cuts faced by the RCFL program in this year). The growing volumes of data suggest that law-enforcement organizations are under-resourced and understaffed.

Besides that, digital forensic investigators need to obtain information quickly by virtue of the forensic discipline nature. The goal of investigators is to uncover the truth to serve justice and protect victims or organizations from further damage. In many cases, such as any emergency threatening life or in which digital evidence is probably to be tampered or destroyed, time is a critical factor. Crucial incriminating information may be lost due to the volatility of digital evidence, therefore the investigation loses its value.

One way to attack these problems is to perform forensic triage. It provides a way to accelerate collecting and processing artifacts. Triage is applied in the medical field first. Recently digital forensics researchers start to focus on triage. Triage introduces a solution tailored to make decisions about allocating scarce medical resources when an emergency incident has happened. It must satisfy three conditions before performing this method. (1) There are insufficient resources to meet requires. (2) An assessment is made by efficient examinations. (3) There has been an established priority system [2]. In the digital forensic field, the research on triage ought to conduct on the basis of these commonly held conditions as well.

Most recent studies in the digital triage field are used to address specific needs of investigators. Moser et al. [3] introduce the use of GRR Rapid Response system to prioritize evidence in a corporate context. Martin et al. [4] propose a method to conduct triage tasks over the network from a central triage server. Roussev et al. [5] present a target acquisition method which is in a file-centric way. There exists no concrete mechanism to implement a digital triage process model on general purpose. This is the motivation of our research.

The pertinent contributions of this paper are as follows:

●The DTDFPM enables investigators to sort investigation requests and distribute investigative resources very quickly. Background information is used to prioritize cases and specific features are served to find which media contain information relevant to the investigation on the basis of our proposed model,. Digital triage is conducted in two stages of our workflow.

●A sorting algorithm (PSAN) which optimizes resources is designed. To best of our knowledge, this is the first application of neural network to the sorting solution in digital triage. It is easy for a neural network to address the growing number of features by increasing input neurons. In a neural network the computer learns from training data, figuring out our solution to the sorting problem at hand.

●We present an implementation of our scheme based on the Python programming language. It can be considered as a digital triage modular and is relatively easy for investigators to integrate our implementation into digital forensic tools for different investigation needs.

978-1-4673-9613-4/16/$31.00 ©2016 IEEE

Our article is organized as follows. The related works is reviewed in section II. We present the proposed triage process model, and describe each stage of the workflow and the sorting algorithm in section III. Section IV contains the case study, Section V concludes the article and describes the future work.

II. PRIOR AND RELATED WORK

Digital investigators follow a certain process model when conducting a digital investigation (see, e.g., [6~16]). These process models are designed to be related to none of specific existing technology or to none of known criminal cases. Moreover, these models contribute to recognize the essence of digital forensics. Therefore, process models can direct research, benchmark practice performance and avoid mistakes [17]. Although the digital forensic process model up to date has not been formally standardized, there is widespread agreement on the abstract level about the digital forensics process [18~20]. Details of each model aside, which is depicted as a linear progression frequently, each model is constituted by several steps. The next step cannot start until the previous step has completed [21]. The latency comes into being in the interval. As the amount of data subject to examination grows, case backlogs exist. Consequently digital forensics needs a promising solution to perform a fast and accurate examination.

Motivated by triage techniques practiced in the medical field, Rogers et al. [22] introduced the Cyber Forensic Field Triage Process Model (CFFTPM). This model is designed to acquire timely clues from on-site examination. Commonly held forensic principles are observed in this process. Meanwhile, two constraints must be paid attention by authors in an on-site examination, namely, time and resource limitations. In emergency cases, a quickly examination must get desired results based on their model. The Semi-Automated Crime-Specific Digital Triage Process Model proposed by Cantrell and Dampier is another example of only few triage process models [8]. It is suitable for the investigators without investigation experience. When the computer profile stage and the crime potential stage are accomplished, investigators can determine with confidence in the presentation stage whether a detailed investigation is needed and how to establish the sorting solution.

Due to the fragility, the acquisition of digital artifacts should be accomplished according to the volatility sequence which is presented in RFC 3227 [23]. Compared with the volatility sequence, different investigators with different opinions about the triage execution may apply different schemes in the triage step. At first, three elementary factors described above have to be satisfied before the triage solution is executed, the first factor is the resource deficiency; the second one is the evaluation according to a swift inspection; the third one is an established triage solution. Furthermore there are different types of digital triage. In the medical field, by contrast, ED triage, incident triage and military triage are all common categories [2]. In the digital forensic, the three factors still are applied. However, the triage execution faces many problems in the new domain. The types of digital triage can be categorized as administrative triage, technical triage, content

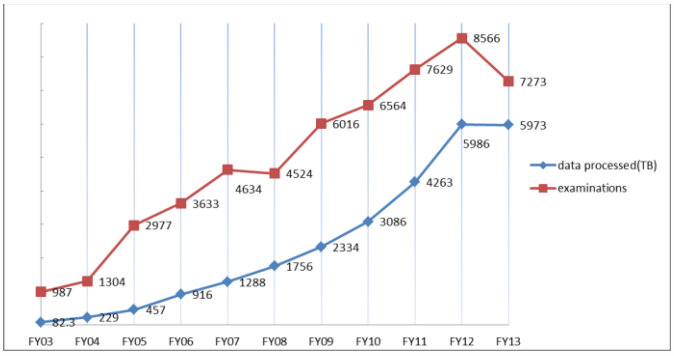

Fig. 1. RCFL digital forensic services annual statistics: FY 2003-2013.

triage [24], [25]. They are described as follows:

● Administrative triage: Each month, hundreds of crimes cases are processed by forensic practitioners. Thus, each case cannot be processed in a complete way due to the time and resources limitation. Administrative triage is implemented at the beginning of the investigation process. Generally, investigators may make a response and perform an investigation when receive a request. Previously, investigators have already acquired some case background information, such as the type of crime, the level of severity or urgency and so on. Investigators should make an examination scheme according to the background information of each case in the light of the ever-increasing cases, and decide which one is high in the priority list. Therefore, it helps to modify hopeless cases listed at the bottom of the priority queue and allocate urgent examinations with a high priority. However, the generally accepted standard to the administrative triage has not been established yet because of the complexity of each investigation and the distinction among investigators.

● Technical triage: The examination stage in a process model can implement technical triage technique. As computer technology advance, investigators have started to use commercial tools for digital triage, such as Access Data's AD Triage. These tools enable investigators to identify apparent evidence in a safe way before a full examination. Their search results make a decision on whether needs to a full examination. Investigators require careful observation and correct assessment in the triage step. Moreover they need valid and reliable triage products. The Computer Forensic Tool Testing (CFTT) project, which is funded by the National Institute of Standards and Technology (NIST), has tested most forensic commercial products for evaluation. However, the project has never made an assessment of digital triage products.

● Content triage: Content triage is usually implemented for lab examination. Due to inefficiency of commonly file centric methods, it employs substitute methods. Generally there are many hashing and indexing operations in traditional investigations, and these operations give rise to reduce efficiency. The similarity digests means, which is proposed by Roussev and Quates [24], is used for content triage to increase investigation efficiency.

III. PROPOSED APPROACH DESCRIPTION

In this section, we first introduce the background of digital forensics. Digital forensics, compare to other research areas, is a case-driven field from the very beginning and its advance is always derived from the response to a specific case [26]. Since the initial investigators came from law enforcement agencies, which were most computer hobbyists or with a background in computer science, digital forensics does not originate from laboratories [17]. Therefore digital forensics has not developed a comparatively complete theoretical system yet. It is a branch of traditional forensics that involves the application of computer science to present digital evidence in court. Likewise, when facing a legal judgment involved in a person's liberty, digital investigators must be careful with their conducts.

Given all of this, we require digital investigators must be qualified for this job with a wealth of practical experience. Compare to the medical world, an intern engaged in the clinical medical treatment is adequately supervised. Moreover, experienced investigators are clearly aware of what their targets are, how they deal with, which method is applied and so on [27].

A. Dual-Triage Digital Forensic Process Model

This subsection introduces the proposed DTDFPM. The process model is depicted in Fig.2. We designed DTDFPM by referring to Digital Forensics Research Conference (DFRWS) Investigative Process Model (DIP Model), which has been widely accepted as a good digital forensic practice [28].

The first step of the process, which is preparation and background checks, is tailored to generate a plan of action and acquire background information. The aim of preparation is to guarantee the necessary tools and personnel to accomplish investigation when a request is received. If necessary, a search warrant is required to obtain before investigation. All the preparation is on the basis of the case background checks. We believe that the collected background information should contain the feature of the case as much as possible, for example investigators should keep detailed records of the type of case, the extent of severity, the level of urgency, etc. All the background information contributes to sort the investigation sequence and allocate resource, when their tasks do not confront a few investigations.

The second step of our investigation process is administrative triage, which is responsible for sorting the investigation sequence so that the limited investigation resource may be made reasonable use. A combination of values needs to be assigned to the corresponding background information. Then, cases can be sorted according to this set of values. Prior experience and a sorting algorithm can guarantee a good sequence formulated for investigations. We will introduce the corresponding sorting algorithm in the next subsection.

The third step of the process, which is preservation and collection, is designed to acquire integrated evidence. The collected evidence can be verified and the implemented technique can be performed repeatedly on the basis of forensic principles such as Daubert principles. According to the first principle in the Association of Chief Police Officers

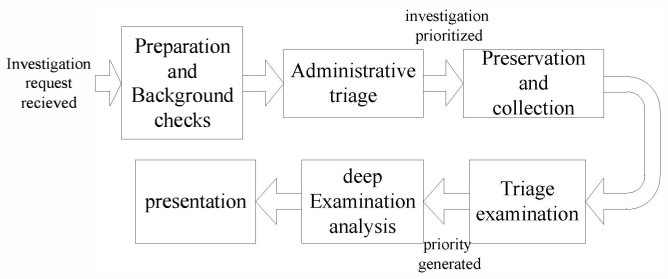

Fig. 2. Dual-Triage Digital Forensic Process Model

(ACPO) Good Practice Guide for Computer based Electronic Evidence[29] mentions that "No action taken by law enforcement agencies or their agents should change data held on a computer or storage media which may subsequently be relied upon in court." Therefore, the collection stage defines conductions after the preservation to ensure that digital artifacts are integral in the whole investigation process [28]. Meanwhile, documentation is one of the key factors to forensic soundness. Documenting the detailed information in the investigation will enable others to evaluate whether evidence is admissible. Generally, the preservation and collection sequence is according to the volatility and importance of the data.

The fourth step of the process is a triage examination stage. The object in this step is to search for the media that most probably suggest admissible evidence according a certain sequence. As a rule of thumb, investigators are inclined to acquire evidence in a gradual manner, due to the analysis performance bottleneck. The next stage profits from the straightforward results directly. The technical triage and content triage are all implemented in this step at the same time. The second principle in the ACPO guidelines [29] asserts that investigators "must be competent to do so and be able to give evidence explaining the relevance and the implications of their actions." Therefore, experienced investigators are needed to fulfill the accurate assignment according to each feature of a case. The corresponding sorting algorithm is summarized in the next subsection.

The fifth step of the process is deep examination and analysis. Deep examination and analysis is executed according to the sequence, when the assignment is finished. The deep examination and analysis stage starts until part of the sorting algorithm's results are enough to satisfy available resources, rather than when the whole sorting is finished. The feedback result from this stage serves to modify search goals and keep the investigation in the right direction. The outcome created in this step also introduces a feedback process that can modify the priority standard on the basis of a further understanding of the case. For the second point, in order to decrease the manpower and increase automation, practitioners can set a threshold standard to initialize examinations. If specific evidence to be analyzed is above the standard, our model starts deep analysis. Furthermore, this step applies to perform in a parallel way. Multiple items which are above the value can be executed simultaneously.

The sixth step of the process is named presentation. Digital artifacts collected in our model are generalized and presented to law enforcement agencies.

In this article, our contribution is twofold. First we generalize the administrative triage step and the triage examination step in the proposed DTDFPM. Second a double screening effectiveness can be realized according to them. The proposed DTDFPM can allocate scarce investigation resources rationally and increase efficiency.

B. Our algorithm

In this part, we introduce a sorting algorithm, called Priority Sorting with Artificial Neuron algorithm (PSAN). (see Appendix for code)

Our algorithm implements sorting based on a three-layer neural network. In contrast to other learning algorithms, we modify the design of the output layer which aims for prioritizing. The neuron model used is the sigmoid neuron [30], it has weights for each input, $w_1, w_2, \cdots$, and an overall bias, b, and is denoted $\sigma(wx + b)$, where σ is called the sigmoid function. the output of a sigmoid neuron is defined by

$$\frac{1}{1 + exp(-\sum_j w_j x_j)}$$

To describe in details, a sigmoid neuron consists of:

1. A linear transformation by weights $w_1, w_2, \cdots,$
2. A translation by bias b
3. Point-wise application of sigmoid.

As a nonlinear activation function, a sigmoid neuron gives neural network its nonlinear capability. With sigmoid neurons, the network transforms the input, creating a new representation. It converts the input by an affine transformation followed by point-wise application of a sigmoid function. As a result, the sigmoid function maps input values to a unique value between 0 and 1.

Inspired by the sigmoid neuron, we make use of neural network sort evidence to be examined and it and can perfectly satisfy the requirement for a sorting system. In our algorithm the network can be thought about a device that prioritizes cases or items by weighing up specific features. Let some sort of background information or factor characterizing an item be a feature. The inputs of network are regard as these features and take on any values between 0 and 1. We can specify the value in this range denote the probability of the presence of a feature. This can be useful because there are too many uncertainties at the initial stage. Moreover, we define inputs which are taken on value 1 if features are present and 0 otherwise. The hidden layer, in the middle network, will be adjusted appropriately as the number of features grows.

In our process model, there are two stages need to perform priority processing. Because different types of features are required in these two stages, we need to establish two networks with different parameters on the same principle.

TABLE I. EXAMPLE OF DATASET

Feature1			*Feature2*	*Output*
0	0	1	0	0.1
0	0	1	1	0.2
0	1	0	0	0.4
0	1	0	1	0.5
1	0	0	0	0.7
1	0	0	1	0.8

Having defined our neural network, let us have a look at an example to illustrate our algorithm. Suppose we have a dataset with two features as input, as shown in Table I. The feature1 can be seen as three levels of severity of an offence to flag a case or the "Order of Volatility" to rank the types of evidence. The feature2 can represent a time constraint whether there is a deadline or an allusion that relates to a key piece of evidence or indicates whether an item of evidence involves the use of encryption. We can consider these two features as prototypes for various factors in an investigation. As for the output, we assign a set of values to reflect the examination sequence, and the bigger the value the higher the location listed in the sequence. In practice, parameters are based on experience accumulation from digital investigations. Experience contributes to represents the significance of the corresponding features to the output. Moreover, research on existing investigations is a good practice to understand property which makes collected digital artifacts uncertainly suggest evidence, so applying these characteristic is suitable to low the level in listed in the sequence. Involved to digital forensic triage process, it may contribute to miss latent greater incriminating evidence. However, we do not need to concern that greater incriminating artifacts will be omitted in our proposed model, they are analyzed subsequently. The algorithm efficiency is controlled by network parameters.

C. Our testing

The objective of this subsection is to study the performance of our algorithm. In light of uncertainties in digital evidence, it is difficult to determine priority directly. Our algorithm, taking into account uncertainties, prioritizes examinations by introducing the probability of the presence of an item of evidence. The establishment of probability is most commonly used in forensic sciences [31], [32].

In the test, we create a network with 4 neurons in the input layer, 12 neurons in the hidden layer, and 1 neuron in the third output layer. Firstly we initialized randomly all the neurons weights in the network by using the Numpy np.random.randn function. We perform a "full batch" training due to a small amount of features in this test. Training is to find appropriate weights in the network. Several items representing a case or a piece of evidence have been tested, as shown in Table II. The results from these tests demonstrate we prioritize all of thembased on their features, as shown in Fig3. transformation function and related definitions in classical control theory, as shown in Fig. 3.

TABLE II. TESTING DATA

Item	*Feature1*			*Feature2*	*Result*
a	0	0	0.9	0	0.10786939
b	0	0.4	0.6	0	0.17352385
c	0	0.4	0.6	1	0.37661903
d	0.4	0.6	0	0	0.55824447
e	0.4	0.6	0	1	0.67795899

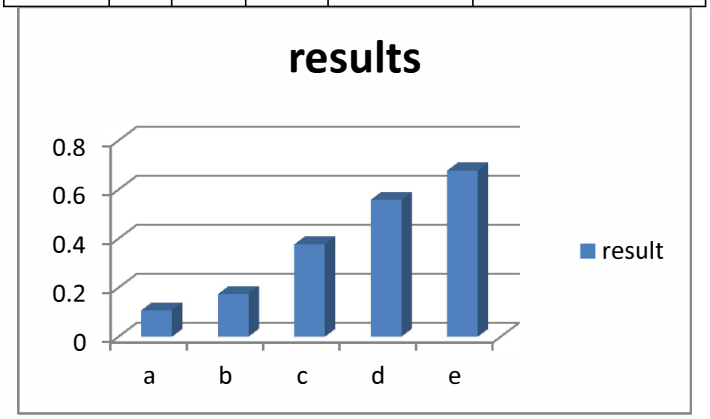

Fig. 3. Comparison of results.

IV. CASE STUDY

This section introduces how the model is applied to an actual case. Although digital investigations vary with different forensic areas such as law enforcement, civil, military, etc., theories and methodologies, there exits commonalities in the investigation process. With the introduction of our process model, it helps researchers understand triage in investigations, and it lets practitioners deal with cases in a timely fashion.

Due to privacy constraints, actual cases and their related data are always not available for academic research. Meanwhile almost every openly available dataset is not suit for digital forensic research. Thankfully, the M57 Patents scenario is an exception provided by the Naval Postgraduate School [24]. The M57 data set contains digital records which is a 17-day length between November 16th, 2009 and December 11. The patents research company is an invented story whose employees are involved in plenty of misdeeds. To mimic real world interaction, different kinds of personas are created in contact with the employees. Thus, multiple scenarios provide the best available public platform for forensic research.

A. Investigative scenario

In 2009, a buyer on the secondary market purchased a workstation, which belonged to a company named m57.biz. He found illegal digital images and videos had not been removed by the previous owner. When a complaint was reported to the police, an investigation was launched. Police contacted the CEO of m57.biz, and made a forensic image of all computer equipment on site at m57.biz with authorization. When a prime suspect had been identified, police further obtained a warrant to seize his personal storage equipment.

This is a case where illegal activity originated from a company with plenty of digital equipment. It is necessary to perform forensic triage in order to avoid case backlogs. This part will make a brief description of the process that practitioners follow our proposed process model.

In the first stage, what is known is the detail information of the company. The M57.biz is a brand-new corporation, which is responsible for handling client patent information, and owns a lot of digital equipment for extending business. Four staff are Pat (CEO), Terry (IT administrator), Jo (patent researcher) and Charlie (patent researcher). The computer which has been sold to the outside of the company was used by Jo, so he is the suspect. Furthermore, we can define the investigation request by the case types, the severity and involved locations, etc.

In the administrative triage stage, the main work is to assign investigative resources in a rational manner. If multiple cases come at the same time, practitioners can assign the corresponding percentage of resources for each case or design a case investigation queue according to our PSAN algorithms. When the investigation enters the preservation and collection stage, investigators must take proper actions guarantee the integral evidence. In the M57 case, the duration of investigation is 17-day. Firstly the RAM contents are captured daily, secondly network traffic are captured by using the gateway's interface, finally all of the computer hard drives and storage devices equipment at M57 are imaged daily.

In the triage examination stage, investigators prioritize acquired digital evidence and create an evidence examination queue based on the prior information, expertise and PSAN algorithm. In the M57 case, practitioners need to initialize PSAN priority parameters based on the existing hypothesis and experience, then execute algorithm, obtain a proper examination sequence at last. For example, the result sets the evidential media belong to Jo as the first priority, network data involving Jo are classified as the first priority as well, the media that Jo probably used as the second one, and the media that Jo unlikely used as the third, and so on. As further investigation continues, more and more evidence can be found and require to reassess the target. The feedback result serves to adjust current investigation.

In the next two stages, the primary investigative considerations focus on what evidence is to support or refute hypotheses as follows, whether Jo is the owner of those illegal digital data, why the computer came to be sold, who else in the company was involved, how those illegal data was distributed.

V. CONCLUSIONS AND FUTURE WORK

The DTDFPM combining with PSAN algorithm applies to digital investigations on a large scale. Compare to existing digital forensic process models, it is competent for conducting rapid forensic investigation. This article gives a detail approach to perform digital triage techniques. The DTDFPM observes commonly held forensic principles, and it does not omit latent incriminating evidence.

There are several working directions for future work, which benefit the DTDFPM. Firstly, it is critical that the success of investigation based on digital triage owe to experienced practitioners. But so far, commonly accepted certification programs or lists of qualifications for investigators have not come into being in the forensic region. The development on

certifications and training programs is our next target. Secondly, a forensic triage system based on the DTDFPM, which comprises conceptual and technological capabilities, will be established to deal with practical tasks. Third, in order to become accepted, the triage approach must be tested and evaluated. The research will evaluate the approach by different examination scenarios and improve the effectiveness. Hopefully our work will serve as an inspiration to digital triage.

REFERENCES

[1] FBI, "Regional computer forensics laboratory (rcfl) program annual report for fiscal year 2013," Regional Computer Forensics Laboratory (RCFL) Program Annual Report, 2013.

[2] K. V. Iserson and J. C. Moskop, "Triage in medicine, part i: concept, history, and types," Annals of emergency medicine, vol. 49, no. 3, pp. 275–281, 2007.

[3] A. Moser and M. I. Cohen, "Hunting in the enterprise: Forensic triage and incident response," Digital Investigation, vol. 10, no. 2, pp. 89–98,2013.

[4] M. B. Koopmans and J. I. James, "Automated network triage," Digital Investigation, vol. 10, no. 2, pp. 129–137, 2013.

[5] V. Roussev, C. Quates, and R. Martell, "Real-time digital forensics and triage," Digital Investigation, vol. 10, no. 2, pp. 158–167, 2013.

[6] A. Agarwal, M. Gupta, S. Gupta, and S. Gupta, "Systematic digital forensic investigation model," International Journal of Computer Science and Security (IJCSS), vol. 5, no. 1, pp. 118–131, 2011.

[7] V. Baryamureeba and F. Tushabe, "The enhanced digital investigation process model," in Proceedings of the Fourth Digital Forensic Research Workshop. Citeseer, 2004, pp. 1–9.

[8] G. Cantrell and D. Dampier, "Evaluation of the semi-automated crime-specific digital triage process model," in Advances in Digital Forensics IX. Springer, 2013, pp. 83–98.

[9] B. Carrier, E. H. Spafford et al., "Getting physical with the digital investigation process," International Journal of digital evidence, vol. 2, no. 2, pp. 1–20, 2003.

[10] B. Carrier and E. H. Spafford, "An event-based digital forensic investigation framework," in Digital forensic research workshop, 2004, pp.11–13.

[11] R. F. Erbacher, K. Christiansen, and A. Sundberg, "Visual network forensic techniques and processes," in 1st Annual Symposium on Information Assurance: Intrusion Detection and Prevention, 2006, p. 72.

[12] K. Kent, S. Chevalier, T. Grance, and H. Dang, "Guide to integrating forensic techniques into incident response," NIST Special Publication, pp. 800–86, 2006.

[13] G. Palmer et al., "A road map for digital forensic research," in First Digital Forensic Research Workshop, Utica, New York, 2001, pp. 1–48.

[14] M. Pollitt, "Computer forensics: An approach to evidence in cyberspace," in Proceedings of the National Information Systems Security Conference, vol. 2, 1995, pp. 487–491.

[15] M. Reith, C. Carr, and G. Gunsch, "An examination of digital forensic models," International Journal of Digital Evidence, vol. 1, no. 3, pp. 1–12, 2002.

[16] G. Ruibin, T. Yun, and M. Gaertner, "Case-relevance information investigation: binding computer intelligence to the current computer forensic framework," International Journal of Digital Evidence, vol. 4, no. 1, pp. 1–13, 2005.

[17] E. Casey, Digital evidence and computer crime: Forensic science, computers, and the internet. Academic press, 2011.

[18] S. L. Garfinkel, "Digital forensics research: The next 10 years," digital investigation, vol. 7, pp. S64–S73, 2010.

[19] N. Beebe, "Digital forensic research: The good, the bad and the unaddressed," in Advances in digital forensics V. Springer, 2009, pp. 17–36.

[20] R. van Baar, H. van Beek, and E. van Eijk, "Digital forensics as a service: A game changer," Digital Investigation, vol. 11, pp. S54–S62, 2014.

[21] V. Roussev, C. Quates, and R. Martell, "Real-time digital forensics and triage," Digital Investigation, vol. 10, no. 2, pp. 158–167, 2013.

[22] 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.

[23] D. BREZINSKI and T. KILLALEA, "Rfc 3227: Guidelines for evidence collection and archiving. network working group. february," 2002.

[24] V. Roussev and C. Quates, "Content triage with similarity digests: the m57 case study," Digital Investigation, vol. 9, pp. S60–S68, 2012.

[25] A. Shaw and A. Browne, "A practical and robust approach to coping with large volumes of data submitted for digital forensic examination," Digital Investigation, vol. 10, no. 2, pp. 116–128, 2013.

[26] K. Nance, B. Hay, and M. Bishop, "Digital forensics: defining a research agenda," in System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on. IEEE, 2009, pp. 1–6.

[27] M. M. Pollitt, "Triage: A practical solution or admission of failure," Digital Investigation, vol. 10, no. 2, pp. 87–88, 2013.

[28] G. Grispos, T. Storer, and W. B. Glisson, "Calm before the storm: the challenges of cloud," Emerging Digital Forensics Applications for Crime Detection, Prevention, and Security, vol. 4, pp. 28–48, 2013.

[29] S. Wilkinson and D. Haagman, "Good practice guide for computer-based electronic evidence," Association of Chief Police Officers, 2010.

[30] M. A. Nielsen, Neural networks and deep learning. Determination Press, 2015.

[31] B. Jones, S. Pleno, and M. Wilkinson, "The use of random sampling in investigations involving child abuse material," Digital Investigation, vol. 9, pp. S99–S107, 2012.

[32] R. Saferstein, Criminalistics. 9th ed. Upper Saddle River, 2007.

APPENDIX

```
#Algorithm Implementation in Python
import numpy as np
import random

class Network(object):
    initialize parameters
    def __init__(self, sizes):
        self.num_layers = len(sizes)
        self.sizes = sizes
        self.weights = [np.random.randn(x, y)
                for x, y in zip(sizes[:-1], sizes[1:])]
    def sigmoid(slef,z):
        return 1.0/(1.0+np.exp(-z))
    def sigmoid_prime(self,z):
        return self.sigmoid(z)*(1-self.sigmoid(z))
    def training(self, training_data,target):
        for j in range(60000):
            inputdata = training_data
            hidden = self.sigmoid(np.dot(inputdata,self.weights[0]))
            output = self.sigmoid(np.dot(hidden,self.weights[1]))
            output_error = target – output
            output_delta = output_error * self.sigmoid_prime(output)
            hidden_error = output_delta.dot(self.weights[1].T)
            hidden_delta = hidden_error * self.sigmoid_prime(hidden)
            self.weights[1] += hidden.T.dot(output_delta)
            self.weights[0] += inputdata.T.dot(hidden_delta)
        print ("Output After Training:")
        print (output)
        print (self.weights)
    def testing(self, test_data):
        test = test_data
        test = self.sigmoid(np.dot(test,self.weights[0]))
        test = self.sigmoid(np.dot(test,self.weights[1]))
        print ("output with test data")
        print (self.weights)
        print (test)
```