International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 3, Issue 7, July 2016

Survey on “Triage-based” Digital Forensics Models

[1] Shital Gade [2] Vanita Mane
[1][2] Department of Computer Engineering Ramrao Adik Institute of Technology
Navi Mumbai, India.
[1]gade.sheetal@gmail.com [2] vanitamane1@gmail.com

Abstract: **Over the past few years, the rapid advancements in information and communication technology in the world has contributed to the increased number of crimes involving digital devices. A large amount of information is produced, accumulated and distributed via electronic means. Traditional techniques of forensic investigation are not appropriate for such growing amount of digital data which require large amount of efforts for analysis. It is indeed very crucial for digital forensics investigators to timely identify, analyze and interpret the digital evidence. Triage is a system or process frequently used at medical facilities for ranking injured or ill patients, it’s aim in digital forensics is to speed up investigation process. In the paper, we reviewed few recently developed triage-based digital forensic models. Some investigation cases such as child pornography, murder, financial crimes, missing/exploited persons are very time sensitive and there is a need for timely identification and analysis of digital evidence. Triage-based models are best suitable to such aforementioned time-sensitive crime investigation cases.**

***Index Terms*— Data mining, digital evidence, digital forensics, triage**

I. INTRODUCTION

Digital forensics is the use of scientific methods for identification, preservation, extraction, analysis and documentation of digital evidence derived from digital sources. With the rapid advancements in information and communication technology in the world, the number of crimes related to the digital devices with huge storage space and broadband network connections has increased dramatically and these crimes are becoming technically intensive. It is indeed very crucial for digital forensics investigators to timely identify, analyze and interpret the digital evidence. The digital forensics investigations are carried out to investigate a wide variety of crimes including child pornography, murder, child abductions, missing or exploited persons. In such types of cases, there is a need for timely identification and analysis of digital evidences found at the crime scene. The forensic experts dealing with such crime investigations need quick investigative leads. The traditional, manually intensive and time consuming procedures indeed, may no longer be appropriate in such cases. There is a need of advanced investigative techniques which can speed up investigation process. Recently developed one of such advanced techniques is ’Triage’, which combines the principles of data mining and machine learning. Triage is a system or process frequently used at medical facilities for ranking injured or ill patients, it’s aim in digital forensics is to speed up investigation process. In recent years, various. digital forensics models based on ’triage’ principle have been proposed by many authors. The paper reviews few of such triage-based digital forensics models with detailed discussion of each model. Triage-based models are needed in the crime investigation cases where timely identification and analysis of the digital evidence is a priority.

II. BACKGROUND

With the rapid advancements in communication and technology, the need for timely identification and analysis of digital evidence is becoming crucial. The traditional forensic techniques of seizing a system, transporting to the lab, making a forensic image(s) and then searching the entire system for evidences is no longer appropriate in some investigation cases where time is a crucial factor. Recently developed triage-based models are best suitable to such investigation cases where time is of essence. Triage is a system or process frequently used at medical facilities for ranking injured or ill patients according to the severity of their injuries or sickness. Triage is used in medical emergencies, telephoning medical advice system, at disaster sites and on battlegrounds as a way of efficient allocation of limited medical resources. The aim of triage in digital forensic investigation is to speed up the investigation process by ranking the seized digital devices and quickly identifying the crime-related evidences. The next section reviews few selected triage-based digital forensic models.

III. TRIAGE-BASED PROCESS MODELS

A new “live” forensics methodology called Computer Forensic Field Triage Process Model (CFFTPM) has been proposed by Rogers M. K. et al. [3], a field or on-site approach for providing timely identification and analysis of digital evidence(s). The computer forensics field triage process model involves the field analysis of the computer system i.e. the model can be applied at the crime scene without the need of transporting the digital device(s) to the lab for acquiring the forensically sound image or for the detailed analysis of the seized digital device(s). Their proposed model is shown in Fig.1. The CFFTPM model involves six stages: Planning, Triage, User/Usage profiles, Chronology/Timeline, Internet activity and Case specific evidence. The CFFTPM's on-site approach is best suitable for the investigation cases where time is a crucial factor such cases are missing or exploited persons, murder, child pornography as in these cases human life or safety is at stake. The forensic experts dealing with such investigations need quick investigative leads. The computer forensics field triage process model supports to the traditional forensic principles and its on-site/field approach provides the benefit of minimizing the contamination or tampering of the original evidence and scene, maintaining the integrity of digital evidence. The Computer Forensic Field Triage Process Model has been used in various real world cases successfully and furthermore, the derived evidence from these cases has not been challenged in the proceedings of court where it has been introduced. The CFFTPM has its applications in cases such as child pornography, financial fraud, identity theft, cyber stalking and murder.

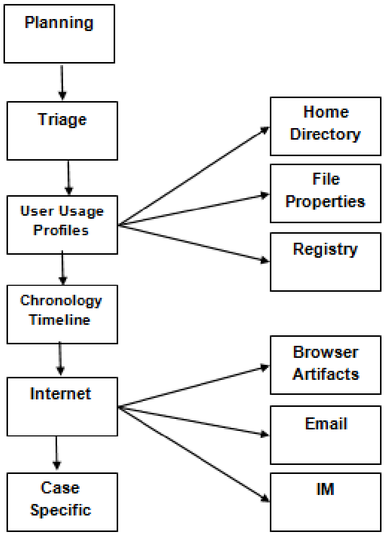

Fig. 1: CFFTPM Phases [3]

Veena H.B. et al. [4] proposed a new approach for data extraction, storage and analysis of data retrieved from the seized digital device(s) that can be used as a evidence in digital forensic investigation. A data mining approach has been used for data generation and analysis and a machine learning statistical approach is used in validating the reliability of the pre-processed data. Authors have focused on proposing an alternate framework for investigation process of physical storage devices, which builds on the models already proposed and chalks out the implementation process for extraction and preprocessing of data extracted from a flash drive. The framework is easy to implement and scientifically practical in approach. It involves six stages: Preparation, Collection and preservation of digital device, Data extraction and preprocessing, Data examination and analysis, Reporting and documentation, Presentation in the court of law. The data extraction and preprocessing phase has been tested out for effectiveness and discussed in detail in their paper. The architecture of the model is shown in Fig.2 which constitutes phases: Extraction, Transformation, Data mining server, Clustering, Classification and Cross validation.

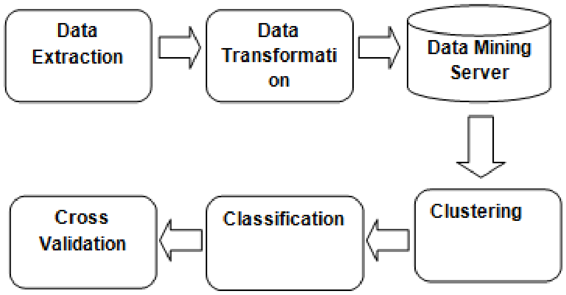

Fig. 2: Architecture of the model [4]

Rosamaria B. et al. [5] proposed a new approach to digital investigations, a 'post-mortem' computer forensics triage, based on the application of knowledge management theory and data mining. It has been realized that the classical forensics analysis process implies that, in each case, terabytes of seized data must be searched to isolate a single evidence. The proposed approach redefines the aforementioned classical forensics workflow, introducing the concept of 'post-mortem' triage which aims to build a priority list among the seized computers, highlighting their relative relevance according to a crime-dependent three-dimensional model of categorization concerning timeline, crime's features and suspect's private sphere. The proposed Post-mortem triaging model is shown in Fig.3. The first stage of the process is forensic acquisition, is in charge of making a disk image in order to preserve digital evidence integrity and guarantee the analysis repeatability. The second stage of the workflow, called feature extraction and normalization, is in charge of extracting relevant data from disk image like system configuration file, browser history, system event log, file statistics etc. The next stage of the process is called context and priority definition. This stage of the model adds the timeline of events and the presence of crime-related features. The datasets, complete matrix and reduced matrix are created at this stage. The fourth stage of the workflow is data classification and triaging and it is in charge of elaborating the reduced matrix in order to provide the final classification of the input data by calculating the class variable by means of classifiers, assigning a relative score to each analyzed exhibit.

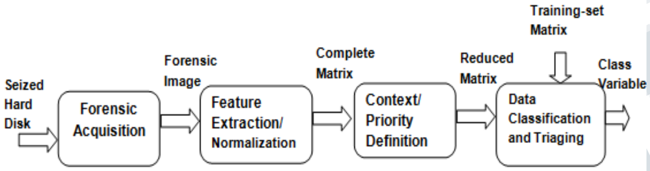

Fig. 3: 'Post-mortem' triaging model [5]

Marturana et al. [7] have proposed a Triage model for content based classification of digital media and presented the results of a case study in which the methodology was tested against forensic data from court cases of copyright infringement. Their research aims to add new pieces of information to the automated analysis of evidence according to Machine Learning-based "post mortem" triage and the research draws the guidelines for drive-under-triage classification (e.g. hard disk drive, thumb drive, solid state drive etc.), based on a list of crime-related features. The model is able to classify evidential exhibits by predicting the class variable according to the aforementioned crime-dependent features. Their proposed model shown in Fig.4 involves three stages. The first stage of the model called forensic acquisition is in charge of creating a forensic image of the seized device. The next stage, feature extraction and normalization is tasked of extracting the crime's dependent features e.g. the number of installed software or the media files average size etc., from each seized drive and creating the data-set. The third stage of the process called data classification and triaging stage provides a classification of the dataset by processing it with one or more machine learning schemes (Bayesian networks, Decision trees, Locally weighted learning and Support vector machines). A case study on Copyright Infringement has been implemented using the model.

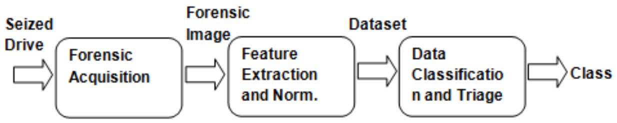

Fig. 4: Triage Process model [7]

Fabio Marturana a,Simone Tacconi [8] describe a Triage-based model for crime-related and content-based classification of digital media and present the results of two case studies in which the methodology was tested against forensic data from court cases of copyright infringement and child pornography exchange. Their proposed model is applicable for both "live" and "dead" digital forensics investigations. The case studies on child pornography exchange and copyright infringement have been implemented as an application of their model. The stepwise procedure for their proposed model is as follows: i) defining a list of crime-related features, (ii) identification and extraction of these features from available devices and forensic copies, (iii) populating an input matrix (iv) processing it with different machine learning mining schemes to come up with a device classification. The methodology aims at processing the digital media and identifying the most relevant data for investigation. Fig.5 shows the Triage-based model which consists of four phases: Forensic acquisition, Feature extraction and normalization, Context/priority definition and Data classification. The popular mining algorithms like Bayes Networks, Decision Trees, Locally Weighted Learning and Support Vector Machines have been used for device classification and the benchmark study about these algorithms has been performed to find out best performing ones.

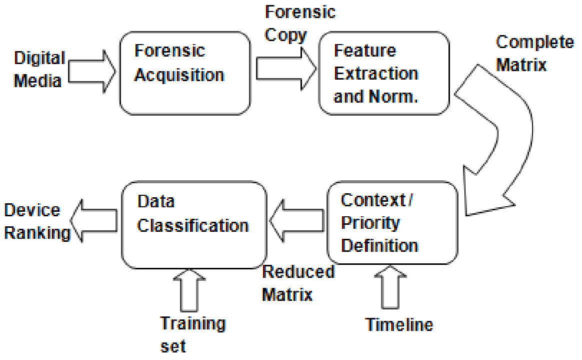

Fig. 5: Triage-based model [8]

As far as mobile forensics is considered, Fabio M. et al. [9] proposed a new methodological approach, 'Mobile Forensics triaging', which take advantage of the specific extraction tools with capabilities of machine learning theory and data mining with the aim of quick identification and analysis of extracted data from all the seized mobile phones. Fig.6 shows the 3-stage methodology. The first stage of the process called crime reports collection (forensics acquisition) is based on a classical forensics acquisition of a mobile phone memory with the appropriate extraction tool. The next stage of the workflow is called data normalization and feature extraction, is aimed at extracting the crime-related features and normalizing these extracted features to remove the misalignments if any and then creating the dataset. The data classification and triaging is the last stage of the workflow and it is tasked of classifying the retrieved evidences.

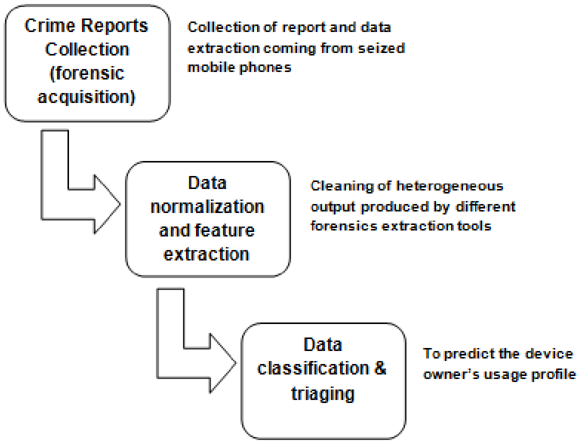
Crime Reports Collection (forensic acquisition)
Collection of report and data extraction coming from seized mobile phones
Data normalization and feature extraction
Cleaning of heterogeneous output produced by different forensics extraction tools
Data classification & triaging
To predict the device owner's usage profile

Fig. 6: Mobile Forensics Process Workflow [9]

The classification process can be either supervised or unsupervised. A supervised approach has been used in the workflow to create a training set, in order to train the classifier and a test set, in order to evaluate the accuracy of the method.

The supporting algorithms like Decision Trees, Bayesian Networks and Locally Weighted learning have been analyzed in order to select the best performing ones. The WEKA suit has been used to select the best performing classifier and the performance is evaluated by using the precision, recall and f-measure parameters. This 3-stage methodology has been applied on set of data, provided by the Italian Cybercrime Police Unit with the aim of helping cybercrime investigators to focus their search on the most relevant evidence first, in order to highlight the device owner's usage profile which can be categorized as base, medium or expert.

IV. CONCLUSION

The paper reviewed few recently developed triage-based digital forensic models. In this world of technology, a large amount of data or information is generated, accumulated and distributed via electronics means. Computerized evidence requires special handling and analysis as the electronic data can be easily damaged, changed or erased, if handled improperly. Timely identification and analysis of the digital evidence is necessary. Traditional techniques of forensic investigation are not appropriate for such growing amount of digital data which require large amount of efforts for analysis. Triage is a system or process frequently used at medical facilities for ranking injured or ill patients, it's aim in digital forensics is to speed up investigation process. The research aim of the triage-based models reviewed in the paper is "a way to help cybercrime investigators to focus their search on the most relevant evidences first, i.e. by prioritizing the seized digital evidences and identifying the crime-related evidences in short time". Triage-based models are best suitable for the time-sensitive investigation cases such as child pornography, murder as in such situations human life or safety is at stake. Triage-based models have their applications in crime investigation cases such as child abductions, missing persons or death threats, murder, hacking, terrorism, financial crimes and many more where 'time' is a crucial factor.

REFERENCES

[1] Richard E. Overill, Jantje A.M. Silomona, Keith A. Roscoe, "Triage template pipelines in digital forensic investigations", Digital Investigation, Vol. 10, Sept. 2013.

[2] Vassil Roussev, Candice Quates, Robert Martell, "Real-time digital forensics and triage", Digital Investigation, Vol.10, Sept. 2013.

[3] Rogers, M. K., Goldman, J., Mislan R., Wedge T., " Computer Forensics Field Triage Process Model", Conference on Digital Forensics, Security and Law, 2006.

[4] Veena H Bhat, Abhilach R. V., P. Deepa Shenoy, L.M. Patnaik, Venugopal K.R., "A Data Mining Approach for Data Generation and Analysis for Digital Forensic Application", IACSIT International Journal of Engineering and Technology, Vol.2, No.3, ISSN: 1793-8236 , June 2010.

[5] Bertè, R., Marturana, F., Me, G., Tacconi S., "Data mining based crime dependent triage in digital forensics analysis", Proceedings of International Conference on Affective Computing and Intelligent Interaction (ICACII 2012) and IERI Lecture Notes in Information Technology, Vol.10, ISSN: 2070-1918, Feb. 2012.

[6] Fabio Marturana, Rosamaria Bertè, Simone Tacconi, Gianluigi Me, "Triage-based automated analysis of evidence in court cases of copyright infringement", First IEEE International Workshop on Security and Forensics in Communication Systems , June 2012.

[7] Fabio Marturana, Simone Tacconi, "A Machine Learning-based Triage methodology for automated categorization of digital media", Digital Investigation, Vol.10, Sept. 2013.

[8] Fabio Marturana, Rosamaria Bert, Gianluigi Me , Simone Tacconi, "Mobile Forensics "triaging": new directions for methodology", Springer ISBN: 978-88-6105-063-1, Proceedings of VIII Conference of the Italian Chapter of AIS (ITAIS 2011) Rome, Italy, 2011.

[9] Graeme Horsman, Christopher Laing, Paul Vickers, "A case-based reasoning method for locating evidence during digital forensic device triage", Decision Support Systems, Vol.61, May 2014.

[10] Robert J. Walls, Erik Learned Miller, Brian Neil Levine, "Forensic Triage for Mobile Phones with DEC0DE", Digital Investigation, 2012.

[11] Inikipi O. Ademu, Dr Chris O. Imafidon, Dr David S. Preston, "A new approach of digital forensic model for digital forensic investigation", International Journal of Advanced Computer Science and Applications, Vol.2, No.12, 2011.

[12] Yunus Yusoff, Roslan Ismail, Zainuddin Hassan, "Common phases of computer forensics investigation models", International Journal of Computer Science and Information Technology(IJCSIT), Vol.3, No.3, June 2011.

[13] P.A. Aguileraa, A. Fernández b, R. Fernández a, R. Rumí b, A.Salmeronb, "Bayesian networks in environmental modelling", Environmental Modelling and Software 26, July 2011.

[14] W.A. Awad, S.M. ELseuofi, "Machine Learning Methods For Spam E-Mail Classification", International Journal of Computer Science and Information Technology (IJCSIT), Vol 3, No.1, Feb. 2011.

[15] Ira Cohen, Nicu Sebe, Fabio G. Cozman, Marcelo C. Cirelo, Thomas S. Huang, "Learning Bayesian Network Classifiers for Facial Expression Recognition using both Labeled and Unlabeled Data", Computer Vision and Pattern Recognition, Proceedings, Vol.1 , June 2003.