Journal of Engineering and Applied Sciences 11 (3): 390-394, 2016
ISSN: 1816-949X

Mobile Block Hash Filtering Based on Irrelevant Data

Yusoof Mohammed Hasheem, Kamaruddi Malik Mohamad,
Ahmad Nur Elmi and Rashid Naseem
Department of Information Security, Faculty of Computer
Science and Information Technology,
University Tun Hussein Onn Malaysia, Parit Raja, 86400 Johor, Malaysia

Abstract: Mobile forensics is an exciting new field of research. An increasing number of open source and commercial digital forensics tools are focusing on extracting accurate results. There is a major issues affecting some mobile forensics tools that enable the tools to extract high number of false positive result during triage examination. This research is focusing on reducing the high number of false positive result generated by Decode. The 66MB data set has been used in the experimenting which is obtained from DFRWS 2010 and five mobile forensics tool have been used for the experiments by comparing their precision, recall and fmeasure performance. The proposed M_Triage tool has successfully scaled down 75% of performing accurately on the 66MB file obtained from DFRWS 2010 as compared to Decode, Lifter, XRY and Xaver. Thus, M_Triage tool is more accurate than Decode, Lifter, XRY and Xaver in avoiding the extraction of high number of false positive result.

Key words: Mobile forensics, triage examination, Decode, Lifter, M_Triage

INTRODUCTION

The methods that show how evidence are taken from mobile telephones are known as mobile phone forensic. It is the science of recovering digital evidence from a mobile phone under forensically sound conditions using accepted methods. It comprises an examination of both SIM and phone memory. Mobile phones have the similar potential of holding evidence as any other digital media can (Brothers, 2009). Any data which confirms or refutes the incident hypothesis in a crime are considered as digital evidence (Umale and Nilav, 2014). However, press to obtain information quickly from digital devices has progressed to a critical point due to the rising number of investigations that involve an ever-increasing amount and diversity of mobile phones containing large amounts and varieties of data.

Nevertheless, introducing efficiencies without careful consideration of the ramifications can cause significant problems and delays in a digital investigation (Casey, 2013). Also, Akkaladevi *et al.* (2011) mention that forensic tool kit lacks performance speed during the investigation process, the traditional approach utilized a single workstation to perform digital investigation against a single source media, which is time-consuming. And the software applications for mobile forensics available today are not 100% forensically sound. The understanding is that they use command and response protocols that offer indirect access to memory. This lead to a drawback to most of the today's mobile forensics tools for them not to manage in handling devices in minimal time and extraction of accurate information. For instance, a mobile f orensics triage tool known as Decode returned over 6.2 million results during mobile forensics triage extraction of which only about 12 thousand were relevant. However, Beebe *et al.* (2011) still mention that current industry standard digital forensics tools and search processes are incapable of handling massive data sets (i.e., Gigabytes and terabytes) in an efficient way. As an answer, relevant evidence can be neglected. Nevertheless, little research exists on how to improve information retrieval effectiveness by adapting the triage method while conducting digital forensic on any devices found. The researchers move forwarder and stated, the focus should be on decreasing human analysis time by transferring part of the analytical burden to the computer, thereby improving query processing and boiling down the systematic impact of non-relevant search hits (Beeb *et al.*, 2011).

In this study, an improved block hash filtering based on irrelevant data is introduced which was formerly recognized as "block hash filtering" developed by Decode and Liter. The purpose of this block hash filtering based on irrelevant data is devloped to thin out the percentage of the false positive result and also to reduce the sum of the file that are examined during triage extraction by M_Triage Module 2.

Corrresponding Author: Yusoof Mohammed Hasheem, Department of Information Security, Faculty of Computer Science and Information Technology, University Tun HusseinOnn Malaysia, Parit Raja, 86400 Johor, Malaysia

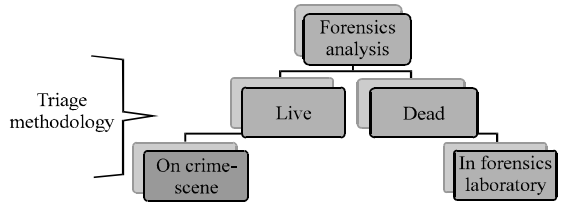

Fig. 1: Triage methodology

Literature review

Triaging: The method of speedily in obtaining critical evidence on-scene or first matter in the lab whiten a partial but perfect and smartness is named forensic triage (Walls *et al*., 2011). According to guidelines on mobile device forensics 2014, "triaging involves performing a data source (i.e., Manual or Logical) on-scene followed at once by a preliminary analysis of the data extracted" (Garfinkel, 2013). Also, Bashir and Khan (2013) elaborated more research, he mentions that triaging in mobile forensic comprises of three key stages, firstly, the evidence is accumulated from the victim's mobile phone. Secondly, the data related to forensic analysis are controlled so that they should not be any alteration on the evidence; this is archived using a hash function. The third and the most critical, the step is called data classification and tagging. In this measure, evidence classification is done by ordering and predicting similar data by employing knowledge management classification algorithms so that amount of data could be scaled down for more dependable performance in the mobile forensic analysis (Bashir and Khan, 2013). Figure 1 illustrate more about where and when triage examination should take place.

Live mobile forensics: Live mobile forensics, sometimes referred as a live incident response on digital devices, is a technique to evoke memory, system processes on to power devices such as MP in addition, live, mobile forensics plays a critical role during mobile forensics examinations due to the possible availability of the digital evidence in the volatile memory such as running processes. Live mobile forensic investigation mainly targets the volatile data that can only be evoked from a running OS; thus, the term "live" is created for such type of examinations else, such information cannot be excerpted from a "dead" OS whose power is down. Conducting live, mobile forensics has become compulsory in the modern era (Bashir and Khan, 2013).

Dead mobile forensics: Al-Zarouni (2006) mention in his research that digital investigations can involve dead and live analysis techniques. In the dead forensic analysis, the target device is powered off/ none damaged or damaged where the mobile phone cannot be powered along. Also, Crisalis stated that the approach that data are being evoked from a powered down system is known as dead forensics. Furthermore, Gary (2012) mention that "dead" examination is established on the evidence that has previously remained powered off either because the mobile housing it has been booted into a digital triage environment or it has been seized and powered down for proper analysis (Cantrell and Dampier, 2012).

Triage forensics tool: In the year 2011, a novel mobile forensics and triage tool known as "decode" was produced by Robert J. Walls and his team. The tool focuses on a data-driven approach to telephone triage. Their end is to allow detectives to remove evidence swiftly in 20 min from a phone, irrespective of whether that exact phone model has been come across previously. The researcherss highlighted that their assessment emphases on past phones for instance, mobile phones with a lesser amount of ability than smart phones. The 20 min processing time claimed by decode was archived on 48 MB mobile phones. Nevertheless, the researcherss developed an algorithm called "Block Hash Filtering" a module whitin their too to reduce the amount of file to be processed during triage examination. Block hash filter essentially is to splits the input byte stream into minor subsequences of bytes. They refer to each of these subsequences as a block. The tool filters out a block if its hash value ties a value in a library of hashes figured from other phones. Blocks may repeat within the same phone but merely the first happening of each block remains after filtering. The tool uses block hashes as an alternative during triage examination than a direct byte comparison, to improve system performance (Walls, 2014). One of the major rezones for the Block Hash Filtering algorithm is to filter out an operating system from the mobile phone during triage extraction. The researcherss did not manage to conceder filtering irrelevant data at the earlier stage; rather they focus on the phone OS and the next coming phones of the same model. These lead to a high number of the false positive result. Which cause "Decode", reverted over 6.2 million results during mobile forensics triage extraction of which only about 12 thousand remained relevant (Varma *et al*., 2014). Furthermore, the same researcherss in the year 2014 extend Decode to support smart phones they came out with another version of a triage tool called "Lifter." The researcher recommend the use of importance feedback to

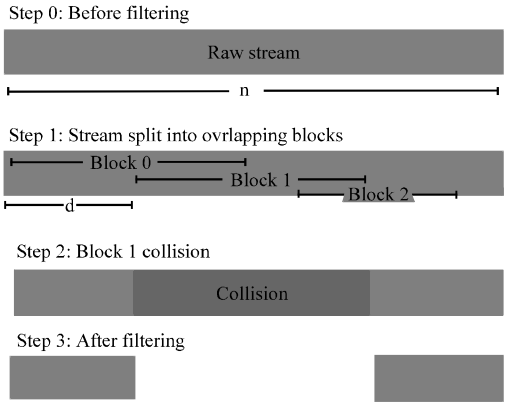

Fig. 2: Block hash filtering

solve this problem known as false positive result. Minor quantity of detective feedback can powerfully and correctly rank in order of significance, the results of a forensic triage tool. And again the toll Lifter failed to conceder filtering irrelevant data at the earlier stage.

This research adapted the same algorithm and improved it to be called "Block Hash Filtering based on irrelevance data" (Fig. 2) illustrate more about the original Block Hash Filtering while (Fig. 3) illustrate about the improved block Hash Filtering based on irrelevance data. Block hash filtering takes a stream of n bytes and produces a series of overlapping blocks of length b. The commencement of each block differs by d≤b bytes. Any collision of the hash of a block with a block on another phone (or the same phone)is filtered out.

Earlier file carving tools for mobile phones: Many file carvers have been prepared to date for mobile telephones. Few improvements are necessary to enhance the efficiency and accuracy of mobile forensic tools. According to Walls *et al*. (2011) there are two conditions need to be strongly considered in developing any mobile forensic tool that is "higher carving recall" and "higher precision".

"Higher carving recall" is about detecting as much useful information as potential and do not only flip away any interesting data from any mobile telephone set that contain possible evidence. Also, Kloet (2007) highlighted some of the goals that all mobile forensic tools should be able to do in society to execute the high caving recall criteria. The goals are as follows.

Support many file types to decrease the number of unsupported false negative. Consider the partial result and brand them as known false positive since they might contain possible evidence. Carving corrupted files as are known false positive and carry on to carve a file, even corruption detected to recover as much as possible.

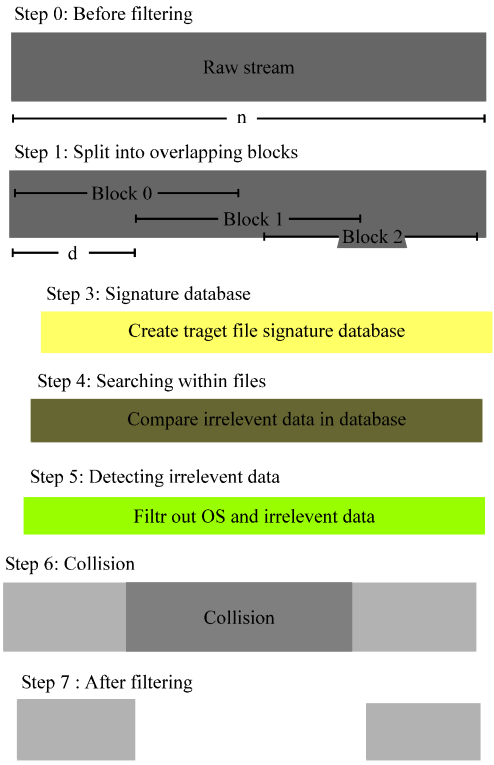

Fig. 3: Block Hash Filtering Based on Irrelevance Data

Higher carving precision is about carving known false positive outcome by any mobile forensic tool which likewise causes a goal as follows: detect false positive and mark them as known false positive to reduce their negative impact.

MATERIALS AND METHODS

This portion discusses the method used in designing Block Hash Filtering based on irrelevance data for M_Triage as shown in Fig. 3.

Block hash filtering based on irrelevant data takes a stream of n bytes and produces a series of overlapping blocks of length b. The commencement of each block differs by d≤b bytes. Then it will build the signature database of all the target file of interest. Next stage is to search for any data that is not defined in the database by comparing the irrelevant data with the specified signature in the database and detect or mark their location when it's done then filter the irrelevant data including the mobile OS. The next stage is to create a hash library and then generate the SHA-1 value of each block if completed then captured and compare to the hash library. Any collision of the hash of a block with a block on another phone (or the same phone) is filtered out.

Experiment: During the experiment block hash filtering based on irrelevant data permanently removes those blocks from the consideration that correspond to the phone's operating system and other redundant data, since these portions where they are stored would not contain any useful information for triage. It boosts the performance of the system by cutting roughly 75% of the original blocks with no effect on performance.

Run_Block_Hash_Filter_Based_On_Irrelevant_Data () of WorkerThread is the next method called by Run () during the filtering process. This process automatically builds the file signature database to compare the signature of the irrelevant file with the once initially define in the signature database and call the Filter () method of the Run_ Block_Hash_Filter_Based_On_Irrelevant_Data class to validate the signature. The Filter () method uses the method HashGetUnfilteredBlocks2 () of the Run_Block_Hash_Filter_Based_On_Irrelevant_Data class to get a list of those block hashes of this phone from the database that did not match the block hashes of other phones in the database. Using the returned hashes, it determines their corresponding memory blocks.

It does so again by comparing the returned hashes with the hashes of the original blocks. The blocks whose hash matches with a returned hash are kept for further processing whereas the rest are discarded. Run_ Block_Hash_Filter_Based_On_Irrelevant_Data () returns an object of the Filter Result class that contains the list of unfiltered blocks which could be used as the input to the inference procedures.

RESULTS AND DISCUSSION

M_Triage uses 66MB dataset which is downloaded from DFRWS 2010 challenge. The dataset is generated specifically to test the carving performance of M_Triage application. However, the performance of M_Triage Application is measured based on valid Address-book, Call logs and SMS, Images and Videos files, Figure 4 illustrate the result obtained by M_Triage. The result is compared with the result obtained from Decode, Lifter, XRY and Xaver mobile forensics tools. Nevertheless, precision, recall and fmeasure are the criteria used in measuring the tools.Based on the experiment taken by M_Triage, Decode, Lifter, XRY and Xaver. The result obtained by the tools shows that all the tools are capable of generating high precision result while in the other round M_Triage and Lifter generate best result in recall. However in evaluating an accuracy result precision and recall evaluation are not good enough,

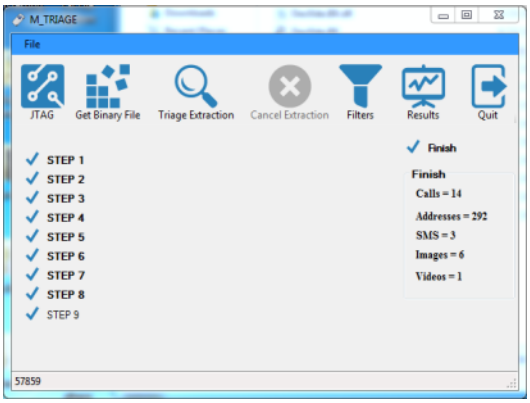

Fig. 4: 66MB result extracted by M_Triage

Table 1: Comparison table of the performance of block hash filtering based on irrelevant data

66 MB	M_Triage	Decode	Lifter	XRy	Xaver
Precision	0.985106	0.126884	1	1	1
Recall	1	1	0.982721	0.570194	0.25486
F-measure	0.992497	0.225195	0.991285	0.726272	0.406196

an f-measure need to be calculated as f-measure = 2* ((precision*recall)/(precision+recall)). The f-measure result shows that M_Triage Module 1 is much more stable for reduction of false positive result than Decode, Lifter, XRY and Xaver. Table 1 for more clarification about the result.

CONCLUTION

This study addresses the issue of generating high number of false positive result by some Triage mobile forensic tools for instance Decode. A novel Block Hash Filtering based on irrelevance data is developed in M_Triage Module 2 to address such problem. Experiments are conducted on 66 MB data sets obtained from DFRWS 2010. Comparisons are made between M_Triage Module 2 with Decode, Lifter, XRY and Xaver. The result shows that M_Triage Module 2 has reduced the number of false positive result by 75% as compared to Decode, Lifter, XRY and Xaver (Table 1). Thus, this shows that M_Triage Module 1 is much more stable for reduction of false positive result than Decode, Lifter, XRY and Xaver.

ACKNOWLEDGEMENT

Researchers would like to acknowledge University Tun Hussein Onn Malaysia (UTHM) for providing financial support (Vot U061) towards this research.

REFERENCES

Akkaladevi, S., H. Keesara and X. Luo, 2011. Efficient forensic tools for handheld device: A comprehensive perspective. Soft. Eng. Res. Manage. Applied Stud. Comput. Intell., 377: 349-359.

Al-Zarouni, M., 2006. Mobile handset forensic evidence?: A challenge for law enforcement. Proceedings of the 4th Australian Digital Forensics Conference, December 4, 2006, Australia.

Bashir, M.S. and M.N.A. Khan, 2013. Triage in live digital forensic analysis. Int. J. Forensic Comput. Sci., 1: 35-44.

Beebe, N.L., J.G. Clark, G.B. Dietrich, M.S. Ko and D. Ko, 2011. Post-retrieval search hit clustering to improve information retrieval effectiveness: Two digital forensics case studies. Decision Support Syst., 51: 732-744.

Brothers, S., 2009. Cell phone and GPS forensic tool classification system. Digital Forensics, May 2009.

Cantrell, G. and D.A. Dampier, 2012. Implementing the automated phases of the partially-automated digital triage process model. J. Digital Forensics Secur. Law, 7: 99-116.

Casey, E., 2013. Triage in digital forensics. Digit. Invest., 10: 85-86.

Garfinkel, S.L., 2013. Digital media triage with bulk data analysis and bulk-extractor. Comput. Secur., 32: 56-72.

Kloet, S.J.J., 2007. Measuring and improving the quality of file carving methods. Master's Thesis, Eindhoven University of Technology, Almere.

Umale, B. and M. Nilav, 2014. Survey on document clustering approach for forensics analysis. Int. J. Comput. Sci. Inform. Technol., 5: 3335-3338.

Varma, S., R.J. Walls, B. Lynn and B.N. Levine, 2014. Efficient smart phone forensics based on relevance feedback. Proceedings of the ACM Workshop on Security and Privacy in Smartphones and Mobile Devices, November 2014, USA.

Walls, R.J., 2014. Inference-based forensics for extracting information from diverse sources. Ph.D. Thesis, University of Massachusetts Amherst, USA.

Walls, R.J., E. Learned-Miller and B.N. Levine, 2011. Forensic triage for mobile phones with Decode. http://forensics.umass.edu/pubs/Walls.usenixSecurity.2011.pdf.