ISSN:0975-9646

Weight-Based-Priority Approach for Analyzing Data Using Traces and Filters

Mahendra S. Pandit
Computer Science and Engineering
K.J. Somaiya College of engineering, Mumbai

***Abstract*- The main goal of digital forensics is the extraction of suspected files from the target devices that can be defined as digital evidence. The digital world is developing at a very fast pace. The size of hard disks made available to the users is also increasing rapidly. The tools used now-a-days explore and analyze the whole digital device which takes more time and resources. Taking into consideration the ever increasing size of hard disk and the data, it is very time consuming and complex task to analyze whole device in time. To make the process of investigation easier, the property of traceability and filters is used. The traceability process has become a key or an important element of the digital investigation process, as it is capable to map the events of an incident from difference sources in obtaining evidence of an incident to be used for other auxiliary investigation aspects. Filter are used by the investigator to remove unwanted data. But the traceability and filter have not been explored to its limits. Because of these, little manipulation in the data on the digital device makes the use of specific trace or filter useless. These work like a loophole, which can be used by criminals to divert the investigation away from evidence. The loophole can be made less harmful by creating a priority based investigation using traces and filters. Priority is given to files which may be or may not be evidence, by assigning them the weight on the basis of the results of traces and filters. By assigning the weight, all files will be taken into consideration and files can be arranged based on the weight. So with weight-based-priority in use, the little or more manipulation to data will be taken into consideration.**

Index Terms- Digital Forensic, traceability, filters, weight , priority, data analysis.

I. INTRODUCTION

Computers and other digital devices are becoming ubiquitous in our modern society. It was inevitable that they would begin to feature as heavily in crime and law. Since the late 1970s the amount of crime involving computers has been growing very quickly, creating a need for constantly developing forensic tools and practices. Almost 99 percent of criminals leave evidence which could be captured and analyzed through proper computer forensic procedure. At one end as the technology is getting advanced, the space of the digital devices are increasing rapidly. The global data supply reached 2.8 zettabytes (ZB) in 2012 - or 2.8 trillion GB - but just 0.5% of this is used for analysis, according to the Digital Universe Study. Volumes of data are projected to reach 40ZB by 2020, or 5,247 GB per person, with emerging economies accounting for an increasingly large proportion of the world's total[1]. Thus the data to be analyzed becomes huge and a challenge for the forensic investigator to perform the forensic investigation in time. One of the important factor in analyzing the data is traceability.

Whenever any operation is done on a device, it makes a lot of entries for security and auditing purpose. These entries are often called as traces. With the help of these traces, it will be known what should be traced and what location and what data. This process is known as traceability. Traceability is the means to identify and follow real or imaginary objects through a process chain. It gives the opportunity to back-track a chain of events, or to predict process outcomes given in the origin of an object. In digital forensic investigation process, tracing is described as a process of finding or discovering the origin or cause of certain scenario. The tracing activities are able to discover the traces left in digital devices. In the computer crime perspective, trace can be found in any digital devices. These traces consist of activities such as login and logout of the system, visit of pages, accesses documents, create items and affiliation groups found in records of data [2].

II. BACKGROUND

Digital forensic have solved many crimes committed with the help of computers where evidence may reside on a computer. From its start in 1970 till today, the field of digital forensic have came a long way and have made many developments. Digital forensic started in early 1970. At that time forensic techniques were developed primarily for data recovery. By the late 1980s utilities were being widely advertised that could perform a variety of data recovering, including "Unformat, Undelete, Diagnose & Remedy. In these early days forensics was largely performed by computer professionals who worked with law enforcement. The years from 1999 to 2007 were a kind of "Golden Age" for digital forensics. During this time digital forensics became a kind of magic window that could see into the past (through the recovery of residual data that was thought to have been deleted) and into the criminal mind (through the recovery of email and instant messages). The Golden Age was also marked by a rapid growth in digital forensics research and professionalization. Universities around the world started offering courses in digital forensic[3]. On the other hand many companies developed softwares specialized for forensic investigation. Open source platform also contributed towards the development of digital forensic.

A. Size of digital devices - A problem

Today much of the last decade's progress is quickly becoming irrelevant. Digital Forensics is facing a crisis. Hard-won capabilities are in jeopardy of being diminished or even lost as the result of advances and fundamental changes in the computer industry:[4]

- The growing size of storage devices means that there is frequently insufficient time to create a forensic image of a subject device, or to process all of the data once it is found.
- The increasing prevalence of embedded flash storage and the proliferation of hardware interfaces means that storage devices can no longer be readily removed or imaged.
- The proliferation of operating systems and file formats is dramatically increasing the requirements and complexity of data exploitation tools and the cost of tool development.
- Whereas cases were previously limited to the analysis of a single device, increasingly cases require the analysis of multiple devices followed by the correlation of the found evidence.

The vast size of today's storage devices means that time honored and court-approved techniques for conducting investigations are becoming slower and more expensive. External hard disks of any size starting from 1tb are easily available in market at reasonable prices. This rapid increase in size is becoming a challenge for forensic investigators.

B. Tools used for forensic investigation

Currently there are many tools that can be used for forensic investigation. But the tools must be validated by the the appropriate authority before being used for investigation. Some of the reputed and globally validated forensic tools are FATkit, EnCase, autopsy. Even some OS are developed for the purpose of forensic investigation. Both of these scenarios are currently happening to some extent. Digital investigation software suits such as EnCase (Guidance 2010), Forensic Tool Kit (AccessData 2010), Autopsy Forensic Browser (Carrier 2010), and others allow an investigator to conduct preliminary, and even some complex investigation tasks simply by knowing which button to press[5]. Most notable OS are SIFT and CAINE. All the tools and software's are designed to perform the thorough examination and analysis of all the space available on all devices on bit by bit basis. The size of the hard disk thats being made available to users is increasing. In order to analyze the huge data on disk it takes very long time. All the software's that are used for forensics basically works in same manner. In short the process can be understood with following points -

1. Imaging of acquired digital devices.
2. Analyzing the digital devices and putting forward the results of analysis.
3. Get evidence from traces or apply filters to find evidences.
4. Generating reports and submitting.

Most important factor in retrieving the evidence are the traces or filters. So traces and filters play very important role in getting the evidence or the hint for evidence

III. TRACES AND FILTERS

A trace is any entry that the operating system makes on the device when a certain operation is executed. The operating system maintains many such traces when working. Some of the trace points are listed below[6].

1. Recent files
2. Prefecth files
3. Jumplist
4. Lnk files
5. Event log
6. System log
7. MFT
8. Memory dump
9. Registry
10. Previous version

Filters are the nothing but the user specified conditions. The filters can be the metadata or contents of the file. The investigator uses the filters to filter out the excess data and get the data that may be the evidence. For example - Type Filter, if we give .pdf as type filter then outcome will be only .pdf file and all other file types will be not considered.

IV. EXPLORING TRACES

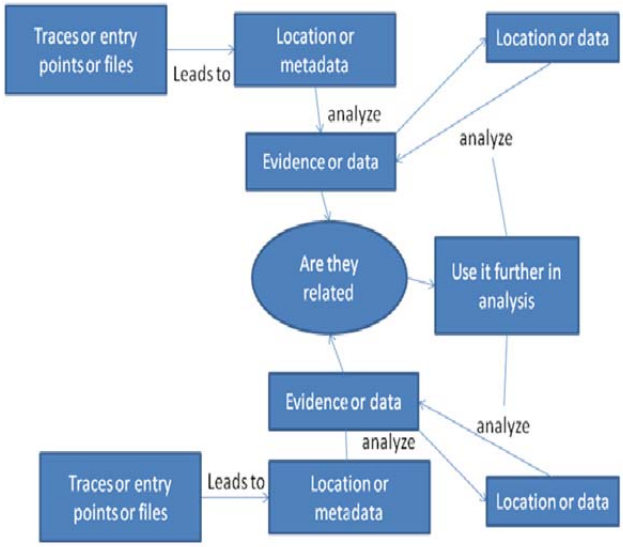
A trace is nothing but a entry of certain operation being executed on the OS. The trace when followed leads to certain file on the digital device. There are two possibilities - either the file may be present or file may not be present there. In current scenario, the trace that leads to a file that is not present in its location is ignored.

There are many possibilities of why the file may not be present at its location. For example - file may have been deleted, file may have been renamed, file may have been moved to some other location and so on. Instead of Ignoring such traces, it is possible to use the result of the traces in further exploring the devices. The next important point to consider is , they are many types of traces written ont the devices. Each trace leads to certain file or location. Each type of trace is evaluated independently of other type of trace. Instead of evaluating each trace independently, the results of one trace can be used further with the results of other traces. This process if continued will build up a relationship among the traces. This in turn will lead to pointing out to the evidence with higher possibilities. The following figure shows how the traces can be used in exploring the data:

In the figure below, only two entry point or traces to start with have been considered. In actual there are be more entry points. The weighting system that can be used can be complex or simple. For now, simple weighting system will be used. Every time a trace leads to same file, the weight is incremented by one.

Steps:

1. Get the trace.
2. Analyze the trace.
3. If it's associated file/location is present.



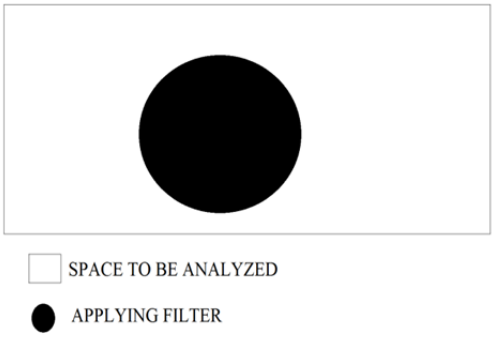
4. Analyze the file and location.(get the metadata of that location and files, get the types and number of files present at that location and its related information and so on) Store metadata and file information.
5. Add the file entry to evidence
6. Is the file and location associated with other traces present?
   a. If yes , associate the files with each other along with its relation. For example - 1) if the two traces lead to same file or location then two traces can be related on basic of the file and path. 2) the trace leads to different files and location but same properties or have maximum properties(name, type, MAC time, author etc) same then the traces can be associated on the basis of the matching properties.
   b. Increase the weight of associated files.
7. Are there any more traces left. If yes, get the next trace and go to step 2.
8. Follow this process until all traces have been analyzed.

V. EXPLORING FILTERS

Filters have always been of great help to the investigators in getting the particular file from many files on device. The filter works in very simple way. The following figure explains the use of filter.

In the diagram, the rectangle represents the digital device and the circle represents the result of filter. It can be clearly seen how some specified data is separated from all the data using the filter. More than one filter can be used to get more specific files.

Figure - Usage of filter



The figure below, explains the working of use of more than one filter. In the Diagram, 3 filters are being used. Lets represent the result of filter 1 as 'A', result of filter 2 as 'B' and result of filter 3 as 'C'. Let the final output of all all 3 filters be 'O'. So 'O' can be written as-

$$O = A \cap B \cap C$$

The red color represents the final output.

Figure - Usage of more filters(3)

With the current working of filters, result only contains the files that satisfy all filters. If any changes is made to the data, then the files affected by the change may or may not be the be the part of the results. In such case, the evidence itself may not be taken into consideration.

Implementing weight-based-priority can handle such changes effectively. Let's see how the weight-based-priority helps. Let's use simple weight system given below -

Weight(File) = No_Of_Filters_Satisfied(File)

According to above equation, weight of a file will be equal to number of filters satisfied by the file. So, files in red colored area will have weight 3, files in blue colored area will have weight 2, files in orange colored area will have weight 1 and files in white colored area will have weight 0. Files can be then giving priority based on this weights. There are two cases :

Some investigator would like to have the traditional way of working with filters. For them , the files of importance will be the files which satisfy the condition below -

Weight(File) = No_Of_Filters_Satisfied(File)
= Total_No_Of_Filters

This files would have the highest weight and based on priority will be placed at the top of all files. So the traditional way of working with filters have changed but

still the investigator can get the result of filters as if it were working in traditional way.

Other files with less weight will follow high priority files. The question is why this files are important. consider files whose weight will be equal to -

Weight(File) = Total_No_Of_Filters - 1

Example - Name, Author, Type, CreationTime, AccessTime, ModifyTime are the filters used by investigator for analysis. Suppose the suspect have renamed the file . So with traditional way of filters, all the filters will not be satisfied and the file will not be shown in results. With weight based priority, the file will be placed immediately following the top priority files(if any). The file which was changed by the suspect is also shown in results. Similarly if suspect changes the extension of file, the file will still be shown in results based on its priority.

VI. CONCLUSION

The use of traces and filters in forensic investigation for analysis and getting evidence have always been very important. Little changes to data associated with traces and filters can get the evidence away from investigation prompting the investigator to use more complex investigation process. But using the weight based priority approach to get the evidence by use of traces and filter can help the investigator in getting the all the data on their priority basis. Even if the evidence file is changed, weight-based-priority will help in placing it among the results(which would not be placed among results using traditional working) based on changes made to file and the weight assigned to it during process of analysis.

REFERENCE

1. Http://www.theguardian.com/news/datablog/2012/dec/19/bigdatastudydigitaluniverseglobalvolume
2. "Traceability in Digital Forensic Investigation Process" by Siti Rahayu Selamat, Robiah Yusof, Shahrin Sahib, Nor Hafeizah Hassan, Mohd Faizal Abdollah, Zaheera Zainal Abidin published in 2011.
3. Shelton Donald E. The 'CSI Effect': does it really exist? NIJ J March 2008;259, http://www.ojp.usdoj.gov/nij/journals/259/csieffect.htm.
4. Digital forensics research: The next 10 years
5. "Challenges with Automation in Digital Forensic Investigations" by Joshua I. James and Pavel Gladyshev published in 2013.
6. Issues in Computer Forensics
7. Research report on "Open Source Digital Forensics Tools: The Legal Argument" by Brian Carrier published in October 2002
8. "Automated Digital Evidence Target Definition Using Outlier Analysis and Existing Evidence " by Brian D. Carrier
9. Eugene H. Spafford published in 2005.