Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 29 (2012) 4254 – 4258

Procedia Engineering

www.elsevier.com/locate/procedia

2012 International Workshop on Information and Electronics Engineering (IWIEE)

Research on the Architecture Model of Volatile Data Forensics

Liang Hu, XiaoLu Zhang, Feng Wang, WenBo Wang, Kuo Zhao*

Department of Computer Science and Technology, Jilin University, Changchun 130012, P.R.China

Abstract

This paper proposed a new architecture model of volatile data forensic. The model applied to all the volatile data sources is a general model. It can rebuild the evidence data fragment to chains of evidence which contains the behavior characteristics, so as to assist investigators to do case analysis. With the accumulated experience, the model can help judicial officers to intelligently analyze the same type of computer crimes, and based on currently available information to predict the impending crimes.

Keywords: Computer forensic; Digital forensic; Volatile data; Forensic model; Live forensic; Memory forensic

1. Introduction

In order to effectively combat computer crime, how to obtain reliable and strong evidence has become a key factor in the detection of crime. But with the development of forensic science, the evidences of crime forensic not only came from the information in computers, but in various types of electronic devices contain digital information relevant to the case. This resulted in the concept of digital forensic.

In the early time, digital forensic is on permanent storage devices to extract and analyze data, one typical representative of such devices is hard disks. Until the summer of 2005, Digital Forensic Research Workshop sponsored a memory analyze challenge. As deeply analyzing from the original physical memory image (source file from windows 2000) can get a lot of information which can not be obtained

* Corresponding author. Tel.: +86-431-85168716; fax: +86-431-85166494.
E-mail address: zhaokuo@jlu.edu.cn.

doi:10.1016/j.proeng.2012.01.653

from stable storage medium such as hard disks. After that, the analysis, discussions, researches, and development tools on original image of physical memory became a new hot area of digital forensic.

Currently a well-known conception of volatile data is the information which will be lost when system is shutdown or neglect and have been recorded in main memory or temporary file on hard disk. RFC3227 [1] shows the order of volatility. For memory is typically volatile data carrier, we make the memory of electronic devices as an example of volatile data and give precise descriptions of this model in section 4.

2. Related work

Following the success of 2005 DFRW memory analyze challenge, researches of volatile data forensic has been further developed. Especially the research and analysis methods for memories from various electronic devices are improved rapidly and endless.

After six years of forensic science development, forensic practitioners have been capable of extracting large amounts of forensic information from the original memory dump. The related technology includes the way of analysis and extraction of structured information on the memory. For instance, [2] can extract process, configuration, and network activity information from any of the Windows NT operating systems. Similar researches have dug up a deeper level of content from memory dump. Such as [3] has a need for more memory forensic techniques to extract user-entered data retained in various Windows applications like the Windows command prompt. The contribution of [4] is a new technique for extracting sensitive information that can not be extracted by string matching-based techniques. [5] [6] provides the methods of finding Rootkits and registry information which have been released from windows memory dump. The memory research domain in Linux [7] [8] or mobile phone [9] also achieved similar achievements.

By analyzing the above related research results, we could easily find: the research results provided the reliable and effective realization of achieving the volatile data and analyzing the format of the original memory dump. These results also became important references and foundational works that we used to further identify volatile data and analysis criminal actions in the proposed model.

3. Volatile Data Forensic Model

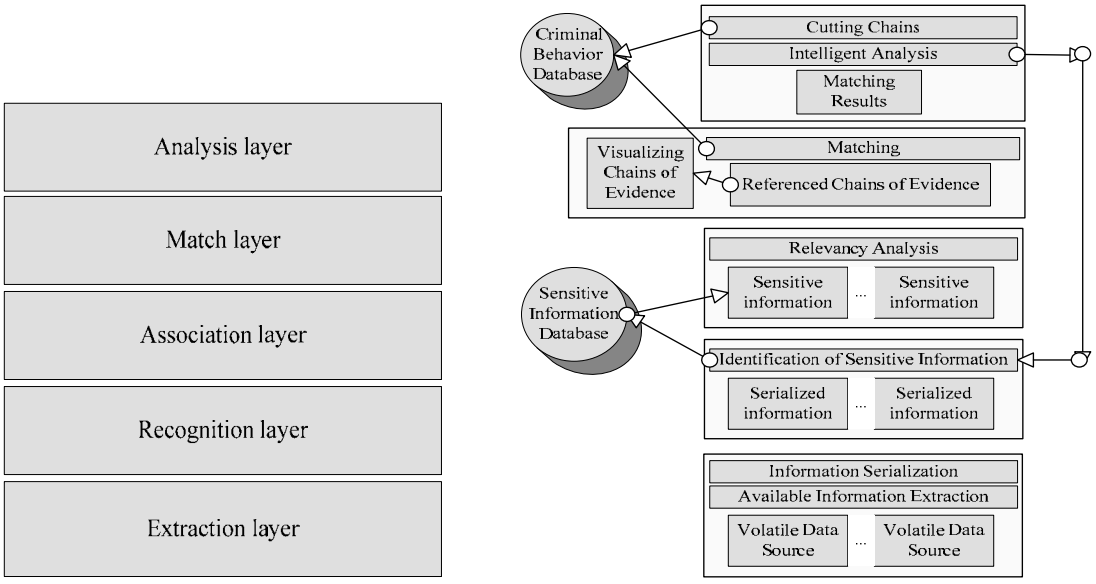

Fig. 1. (a) the abstract structure of volatile data forensic model; (b) the abstract structure of volatile data forensic model

As figure 1, the volatile data forensic model in this paper comprises five layers. In order to accomplish the functions of each layer, we have established a sensitive information database and another criminal behavior database.

3.1. The first layer: extraction layer

In the evidence collection process of the actual criminal cases scene, the investigators search and retrieve possible evidence of importance with their experience and forensic science techniques. But the process requires investigators protect the scene and save the original form before sorting and classifying the criminal environment to extract further evidence.

Similarly, the main task of getting the first layer of the model in this paper is to remove the noise data, which is unstable and unable to produce significance information, only extract the available information in the volatile data source (usually as a binary data source) by means of the specific treatment. Due to the difference of data sources (such as different operating systems have different memory management), the extracted available information is also different in the aspects of natural types, orders and so on. In order to analyze the available volatile data source effectively, we need to package all the available volatile data going through this layer into a unified format. The reasonable definition of serialization format will help the disordered volatile data to become effective evidence on the court in the future.

3.2. The second layer: recognition layer

We still refer to the on-site work of getting evidence in actual criminal cases. In the process of forensic investigation, investigators do not extract and bring back everything as evidence. On the contrary, investigators only save items those that may have a role in detecting cases as evidence for the next further investigation.

In the course of the investigation of digital evidence, we should also follow this rule, particularly in the evidence gathering of volatile data in order to reconstruct the evidence chain. Because the amount of forensic data source after aggregating itself is massive, if we analyze one by one, the overheads is unthinkable. It is the same as real crime scene investigation, investigators need to explore many times, we need the iteration process to find those which is helpful for us to solve the case, or useful as evidence of sensitive information.

3.3. Sensitive information database

As the second layer we have completed to identify the sequence of sensitive information in the volatile data that requires the existence of a knowledge database as the basis for identification of sensitive information, with the serialized data to match. With the identification of sensitive information related to the contents of the definition of 1 in the following description.

Definition 1: Sequence of any information contained n characteristic descriptors; Characterization of any item has a finite number of discrete values. With T_i stand for the i characteristic descriptor set of values ($0 < i \le n$) and $\{T_n\}$ stands for all the characteristics that describe a collection of items of value. Then any sequence information can be expressed as $x = \{\{t_{11}, t_{12}, ..., t_{1\kappa}\}, \{t_{21}, t_{22}, ..., t_{2\lambda}\}, ..., \{t_{n1}, t_{n2}, ..., t_{n\eta}\}\}$ ($t_{11}, t_{12}, ..., t_{1\kappa} \in T_1$, $t_{21}, t_{22}, ..., t_{2\lambda} \in T_2$,…, $t_{n1}, t_{n2}, ..., t_{n\eta} \in T_n$). All sensitive information in the database gives the sensitive information must meet a number of characteristic values, as $Y = \{y_1, y_2, ..., y_s\}$. For any $y_i \in Y$, there is a feature descriptor set of m

values $\{T_m\}$, y_i include m characteristic items, $\{T_m\} \subseteq \{T_n\}, 0 < m \le n$. Define the mapping f : $\{T_m\} \to \{T_n\}$ for the same features characterize the mapping entry. As $\{T_m\}$ is the subset of $\{T_n\}$, then any $T_j \in \{T_m\}$ can be found only $T_i \in \{T_n\}$ in the mapping. If there is $y_k \in Y$, for any $T_j \in \{T_m\}$, T_j : $y_k = \{t_{j1}, t_{j2}, ..., t_{j\phi}\}$, $T_j \in \{T_n\}$, sequence information x in the T_l , which values are $\{t_{l1}, t_{l2}, ..., t_{l\varphi}\}, \{t_{j1}, t_{j2}, ..., t_{j\phi}\}$ is subset of $\{t_{l1}, t_{l2}, ..., t_{l\varphi}\}$. If those conditions were satisfied, we consider that x is the sensitive information.

3.4. The third layer: association layer

The association analysis layer is responsible for the reconstruction of the evidence chain. With the use of time, location, maps, and other factors, as measured factors on the second layer to obtain sensitive data for conducting an analysis of the association between each other. The known relevance rules of the evidence will help analysis' organize specific information together, which constitute the behavioral characteristics of chain of evidence. The following definition 2 is the association between the sensitive events.

Definition 2: For any two-sensitive events A, B , α is the feature descriptor for the characterization of A , β is the feature descriptor for the characterization of B . Setting maps (f : A $\to$ B) are the two interrelated features of the mapping descriptor, if there $\alpha \in$ A , $\beta \in$ B is enough to meet β is on the map $\beta = f(\alpha)$. Then we say that sensitive events A 、 B are associated.

3.5. The fourth layer: match layer

Matching layer can achieve two aspects of the work. The first is to match the evidence chain that obtained through dealing with the association layer and the other one in the criminal behavior database. The other is to visual the remaining evidence chain that is not in the behavior database, but may be helpful to investigate the case. As the information in this layer is presented in the format of evidence chain, it is with a certain logic relationship, which is helpful for investigators to reason, reconstruct the case, detect the case finally, and gather evidence in court.

3.6. Criminal behavior database

Establishment of criminal behavior database (i.e., chain of evidence database) is the purpose of summing up the experience by cracked criminals. Creating such database not only avoid those available data repositories become "data tombs" (data archives that are seldom visited) after closing a case but also every time collect criminal actions as "chain of evidence" which used to help an analysis of evidence and case reconstruction.

3.7. The fifth layer: analysis layer

As the final layer of the whole architecture model, analysis layer would continue to accomplish the work came from association layer. In this layer, by intelligent analysis, it was confirmed that whether the part of evidence chains that would not be recognized through criminal behavior database can be forensic evidence. If the results obtained from extraction layer were not satisfactory. Our work should go back to the second layer and reordered the range of the sensitive information until finding a reasonable set of sensitive information.

After one case closed, all valid chains of evidence on further abstraction and cutting will be uploaded to the criminal behavior database, which could be the foundation for uncovering other similar cases in

future. With the expansion of information and the constant improvement of the criminal behavior database, we will be more knowledgeable of criminal behavior models and characteristics in similar cases. In the near future, we will uncover more methods to restrain crimes and even prevent crimes by further analysis of the criminal behavior database.

4. Conclusions

The significance of this model is that we fully utilized the research results of volatile data analysis and extraction which are more close to the physical level, and those results may be abstracted to be one layer function of a general model which adapt to any type of sources. Thereby, the previous situation of digital evidence extraction is avoided, which only use a specific analysis and extraction methods to analyze and extract the available information from data sources. In addition, we also proposed: the chain of evidence from criminal process will be restored and matched with features of criminal acts in the database, which can be accumulated past experience of handling similar cases as the basis for the future, to solve the investigators simply by their own experience. Our model not only has a strong universal but it will aim at intelligent analysis to face the growing and massive information in digital evidence, which will dramatically change the extraction and analysis work of existing forensic patterns.

Acknowledgments

This work was supported in part by the National Grand Fundamental Research 973 Program of China under Grant No. 2009CB320706, the National High Technology Research and Development Program of China under Grant No. 2011AA010101, the National Natural Science Foundation of China under Grant No. 61103197 and 61073009, the Youth Foundation of Jilin Province of China under Grant No. 201101035, and the Fundamental Research Funds for the Central Universities of China under Grant NO.200903179.

References

[1] D. Brezinski, Guidelines for evidence collection and archiving, http://tools.ietf.org/html/rfc3227, RFC3227, 2002.

[2] James Okolica, Gilbert L. Peterson, Windows operating systems agnostic memory analysis, Digital Investigation, vol.7, 2010, pp.48-56.

[3] Richard M. Stevens, Eoghan Casey, Extracting Windows command line details from physical memory, Digital Investigation, vol.7, 2010, pp.57-63.

[4] S. M. Hejazi, C. Talhi, M. Debbabi, Extraction of forensically sensitive information from windows physical memory, Digital Investigation, vol.6, 2009, pp.121-131.

[5] Jesse D. Kornblum, Exploiting the Rootkit Paradox with Windows Memory Analysis, International Journal of Digital Evidence, vol.5, Issue 1, 2006, pp.1-5.

[6] Brendan Dolan-Gavitt, Forensic analysis of the Windows registry in memory, Digital Investigation, vol.5, 2008, pp.26-32.

[7] Treasure and tragedy in kmem_cache mining for live forensics investigation, Andrew Case, Lodovico Marziale, Cris Neckar, Golden G. Richard, Digital Investigation, vol.7, pp.41-47, 2010.

[8] Lodovico Marziale, Andrew Case, Golden G. Richard, Dynamic recreation of kernel data structures for live forensics, Digital Investigation, vol.7, 2010, pp.32-40.

[9] Vrizlynn L. L. Thing, Kian-Yong Ng, Ee-Chien Chang, Live memory forensics of mobile phones, Digital Investigation, vol.7, 2010, pp.74-82.