Volatile Internet Evidence Extraction from Windows Systems

Neethu Joseph, Sherina Sunny
ER&DC Institute of Technology,
Centre for Development of Advanced Computing
Thiruvananthapuram, India

Dija S, Thomas K L
Resource Centre for Cyber Forensics
Centre for Development of Advanced Computing
Thiruvananthapuram, India
dija@cdac.in

***Abstract*—Internet users are increasing day by day and hence browser related evidence provides crucial information regarding a cyber crime. The rate of possible cyber crimes are increased unimaginably with this high usage of popular social networking websites and online internet services for banking, shopping etc. Thus the need for collecting internet browsing related information through a Browser Forensics Analysis is inevitable in a cyber crime investigation. Browser Forensics can be done as part of offline forensics by analyzing browser related files containing cookies, cache and other history information available in the hard disk. But, these files usually stores limited information and its content varies based on user settings. On the other hand, when a live forensics approach is adopted, the prime source of forensically relevant information is physical memory. So, in an internet related cyber crime, the chance of getting crucial information by analyzing physical memory content collected from the Suspect's machine is very high. This paper presents a methodology for extracting user credentials of popular web applications by analyzing a Windows system's physical memory content. It helps cyber crime investigators to retrieve usernames and associated passwords used in various web based mail accounts, online banking and shopping sites etc. Another important methodology the paper presents is for the retrieval of high profile browser forensics information related to the suspect's internet activity by memory dump analysis.**

Keywords—Digital Evidence; Digital Forensics; Live Acquisition; Live Forensics; User Credentials;

I. INTRODUCTION

As the development of computer technology, cyber crimes have become very common. In such a circumstance, more and more attention has been paid on cyber forensics. Cyber forensics is a branch of forensic science encompassing the recovery and investigation of material found in digital devices, often in relation to computer crime [1]. This involves obtaining and analyzing digital information for use as evidence in civil, criminal, or administrative cases [2]. This digital evidence can be either static or live. Static data is stored in storage devices like hard disks and other removable storage media of a computer. On the other hand live data is stored in physical memory, which is highly volatile. Also, the live data stored in a system changes continuously as the state of the system changes. In live forensics, digital evidences are collected from a running system itself because of its volatile nature. Here, physical memory is taken as an important source of digital evidence. Live forensics involves acquiring and analyzing physical memory content of the system. All the information available in the physical memory is lost forever once the machine is turned off. This is due to the volatile nature of physical memory. Thus, live forensics collects crucial information that may be lost by powering down a system. Web browser forensics is an increasingly important field in cyber forensics which deals with the extraction of internet related evidence. Browser related details such as visited sites, user credentials, information about the sent or received emails, searched queries etc can be extracted from by analyzing physical memory dump file. None of this information may be available in the hard disk. Thus the evidence collected from the physical memory dump provides crucial information especially in case of internet related crimes.

II. TRADITIONAL VS LIVE FORENSICS

There are two different approaches in cyber forensics. First one is the widely accepted traditional offline forensics. Traditional offline forensics is performed through static analysis of data preserved on permanent storage media [3]. It attempts to preserve all storage media content in an unchanging state. Here, concentration is mostly on the content present in hard disks and other removable storage media. In this case, even if the suspect's system is in the running mode at the scene of crime, the investigator pulls the power plug and then images the hard disks to a new file in a new storage media. This bit stream image file is then analyzed in an analysis lab. This is according to the widely accepted cyber forensics procedure 'never work on the original evidence'. But, this approach has the following drawbacks.

- The disk capacity keeps increasing and now terabyte hard disks are available and are very common. Mirroring, indexing and searching of these disks are time consuming.
- Large corporate can't bear losses when pulling the power plug of a critical server even with a court order.
- Some crucial evidence may be sometimes available only in physical memory and there may not be any foot prints related to the suspected cyber crime in the hard disk.

978-1-4799-3975-6/14/$31.00

The biggest limitation of traditional forensics is that it cannot provide a complete picture of events [4] happening in the system. But, live forensics is a relatively new area of cyber forensics where an investigator collects physical memory content to a file and performs analysis of this memory dump file in an analysis lab. Live forensics considers the value of volatile data and collects bit stream copy of the physical memory if the system is in the running mode at the scene of crime. By analyzing a physical memory dump file, information such as the list of currently running processes, open ports and listening applications, system information, system users, network connections etc. can be collected[5]. This is done by exploring data structures present in the physical memory. Other than all these forensically valuable information, physical memory dump file may contain information like user passwords in clear text, browser related information, encryption keys[6], internet protocol (IP) address, executed console commands, foot prints of malwares, raw form of encrypted data in hard disk, instant message data etc. These information may not be available anywhere in the hard disk.

Browser forensics is collection of evidence related to the internet usage of the suspect. This can be done as part of traditional offline forensics or live forensics. There are few browser related files saved in the hard disk in some predefined location of the operating system drive in a Windows system. In case of traditional forensics, these browsers related files are analyzed to extract internet related evidence. But the content available in these files varies depending on user settings. So evidence that can be obtained from this type of an analysis is usually limited. On the other hand, if a live forensics approach is adopted, physical memory content of the suspect's machine is collected to a file for analysis. Crucial internet related evidence can be retrieved by analyzing this memory dump file at an analysis lab. Evidence is often saved in the memory for later use in the belief that it can be accessed anytime in the future [7]. Thus analyzing a memory dump file reveals crucial evidence pointing to the cyber crime. This paper presents methodologies for retrieving internet browsing related evidence by this type of a memory analysis. Any of this evidence may not be available in the hard disk and thus a traditional forensics approach is not recommendable in internet related crimes.

III. MEMORY FORENSICS

Memory Forensics is the main area under Live Forensics which deals with physical memory acquisition and analysis. The first step here is to obtain physical memory content to a file. This is to be done when the suspect's machine is in running mode. The second step is to analyze this memory dump file to retrieve forensically crucial information [5]. This analysis is usually done at the Investigator's site or analysis lab because the basic principle in live forensics is to minimize the tampering made in the Suspect's machine because of running an cyber forensics tool. The following section describes the acquisition and analysis process in memory forensics.

A. *Phsical Memory Acquisition*

Acquisition is the process of creating the forensic duplicate [3]. Most of the live forensics acquisition tool supports acquisition of physical memory to a file. Here, for the research purpose, DumpIt tool is used for acquiring the physical memory content. DumpIt [8] developed by Matthieu Suiche is a freeware tool that supports memory acquisition from both 32-bit and 64-bit versions of Windows Operating System. A screenshot of the tool is given in Fig. 1. DumpIt saves the contents of physical memory to a file with .raw extension. This memory dump file is generated in the current directory and file size is as same as that of the system's physical memory.

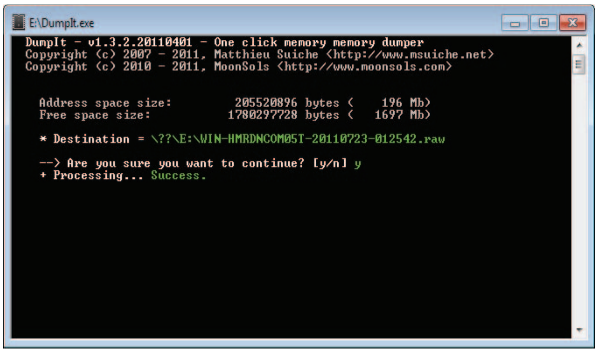

Fig. 1. Screenshot of DumpIt tool.

B. *Phsical Memory Analysis*

Memory Analysis is analyzing memory dump file for extracting crime related evidence. Usually this file may be typically of gigabytes (GB) of size and contains normal ASCII and junk characters. For analyzing the memory dump file, hexadecimal editors like HxD-Hex-Editor [9] can be used. While opening a memory dump file using hexadecimal editor, it is possible to view several browser related fragments in plain text from. But manually looking into and analyzing the memory dump file of huge size is not practical and is very time consuming. The next section presents the important patterns that are to be searched in order to obtain crucial internet related evidence from the memory dump. It explains methodologies for retrieving user credentials and other internet related evidence like email fragments, Facebook messages, searched urls etc.

IV. DETAILED PHYSICAL MEMORY ANALYSIS

Whenever an application is executed in a computer, it loads fully or partially into the physical memory. So, the physical memory contains details about all running applications including internet activities. This includes all browser related evidence like searched urls, email related data, Facebook related evidence, user credentials etc. Usually in a cyber crime investigation, suspect may not be willing to reveal the user credentials of personal accounts in different internet applications. If these user credentials are available, investigators can collect more information like emails, chat contents, uploaded files and other important evidence, by login into these web accounts, if necessary, as part of the investigation. The following sections describe a pattern searching methodology to be done to retrieve user credentials and other important internet related evidence.

A. Extracting User Credentials

Usually passwords are stored in encrypted form or as hash values inside the hard disk. System logged-in passwords of windows user accounts are stored as hash values [10]. Similarly, passwords corresponding to internet accounts are also stored in encrypted formats inside the hard disk. Our research shows that these passwords are stored in physical memory in plain text form. This is usually saved into the memory when 'remember password' option is enabled in that page or from the browser. This research is done in Windows operating system by login into sixty five commonly used internet websites using three popular web browsers such as Google Chrome, Internet Explorer and Mozilla Firefox. While logging in 'remember password' option is checked and physical memory is collected after this. Then this memory dump file is analyzed to retrieve the username and password. And all these user credentials are successfully retrieved using a pattern searching made in the collected memory dump. In physical memory user credentials of web applications are preceded by some constant string patterns. So, these patterns are used for searching the user credentials. These patterns are different from one web site to another depending upon the control names used for username and password edit boxes at the time of developing that web site. After this research, we have successfully identified the keywords to be searched for finding user credentials of all sixty five web sites under the experiment.

TABLE I. PASSWORD SEARCHING PATTERNS OF USER ACCOUNTS

No	Website	Username Keyword	Password Keyword
1	amazon.com	&email=	&password=
2	ebay.in	&userid=	&pass=
3	facebook.com	&email=	&pass=
4	flipkart.com	&email=	&password=
5	gmail.com	&Email=	&Passwd=
6	hotmail.com	&login=	&passwd=
7	irctc.co.in	&userName=	&password=
8	linkedin.com	&session_key=	&session_password=
9	myntra.com	&email=	&password=
10	pepperfry.com	&email=	&password=
11	rediff.com	&id=	&num=
12	skype.com	&username=	&password=
13	twitter.com	&session%5Busername_or_email%5D=	&session%5Bpassword%5D=
14	yahoo.com	&login=	&passwd=
15	federalbank.co.in	&CorporateSignonCorpId=	&CorporateSignonPassword=

Table I shows the patterns used for extracting usernames and passwords of fifteen popular internet websites. For example, consider the fifth row in table I that shows the keywords to be searched for obtaining username and password for a gmail account. Here, the username is found after a keyword '&Email=' and password is after '&Passwd=' keyword. Fig. 2 shows a screenshot of such a hit in physical memory dump. The user entered credentials are stored in plain text form in physical memory except special characters in the password. The special characters are converted into corresponding ASCII hex value before storing. This is the only conversion that is used before passwords are stored in physical memory. So, this conversion is to be done in order to get the actual password at the time of analysis. Thus, it is possible to extract the usernames and passwords of web application accounts by analyzing memory dump. Fig. 3-10 shows snapshots of user credential searching done in physical memory dump file based on the details furnished in Table I. These are corresponding to Yahoo Mail, Facebook, Rediff Mail, Twitter, Hotmail, LinkedIn, Federal bank and Oriental Bank of Commerce websites respectively.

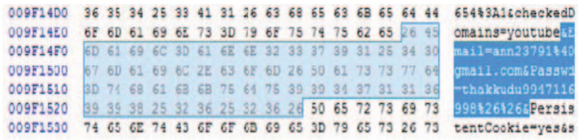

Fig. 2. User Credentials of Gmail Account

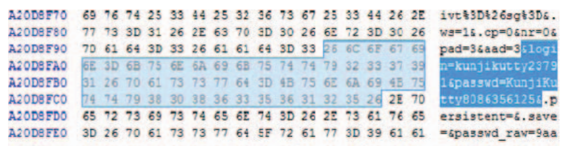

Fig. 3. User Credentials of Yahoo Account

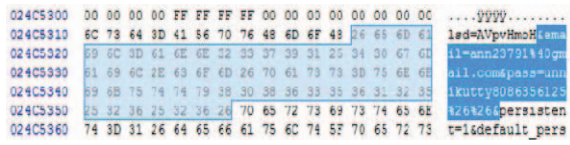

Fig. 4. User Credentials of Facebook Account

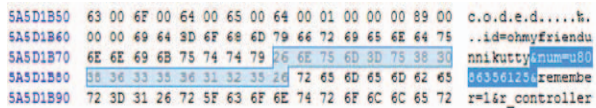

Fig. 5. User Credentials of Rediff mail Account

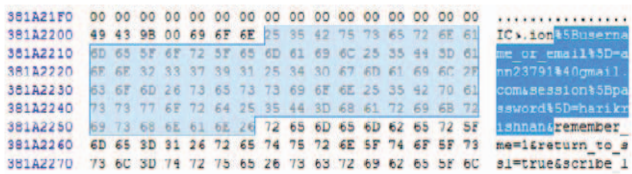

Fig. 6. User Credentials of Twitter Account

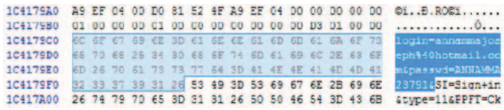

Fig. 7. User Credentials of Hotmail Account

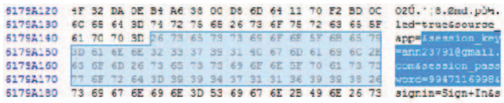

Fig. 8. User Credentials of LinkedIn Account

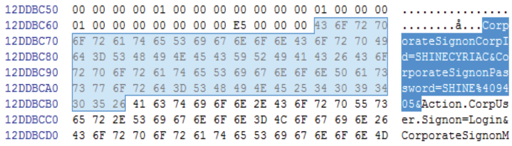

Fig. 9. User Credentials of Federal Bank Online Banking

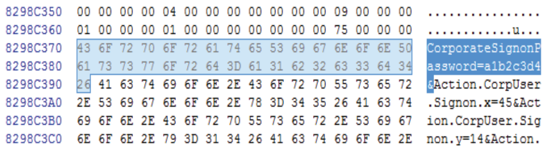

Fig. 10. User Credentials of Oriental Bank of Commerce

B. *Extracting Other Forensically Relevant Internet Evidence*

In a cyber crime investigation, obtaining user credentials from the memory dump file is crucial. But there are many other forensically sound internet related information which can be retrieved from physical memory content. The main information among these are list of visited sites, email information, search queries, browser details, cookies information and Facebook fragments. Here, a detailed research for retrieving this information is conducted with three popular internet browsers Internet Explorer, Google Chrome and Mozilla Firefox. Even though the memory dump contains junk characters, it is possible to extract crucial browser related data. And the importance of this information is that it may not be available anywhere in the hard disk.

Visited sites can be retrieved by searching for the pattern 'http://------.com', 'https://------.com' and '<url>-----</url>'. Search engines are very popular nowadays. So, the search strings used by the Suspect gives a picture of the possible ways in which the suspected crime was committed. These search strings can be extracted by locating the patterns 'search=--+--+--' and 'search_query=--+--'. Other than visited sites and searched terms, many other crucial information such as browser information, cookies details, email fragments, user ids, Facebook details etc can also extracted in the same way. Table II shows patterns to be searched for retrieving various crucial evidence from different browsers. Fig. 11-22 shows memory dump fragments obtained while searching the memory dump against the details given in table II. Fig. 11 and 12 shows details of visited sites. Fig. 13 and 14 shows mail related data and 15 and 16 shows user ids. Fig. 17 shows a search query and 18 shows browser information. Fig. 19 shows cookies details and 20-22 shows Facebook information.

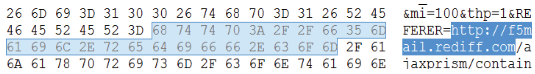

Fig. 11. Memory dump fragment of a visited site

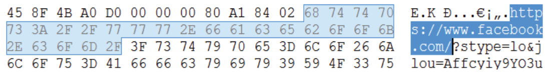

Fig. 12. Memory dump fragment of a visited site.

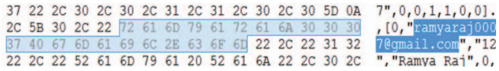

Fig. 13. Memory dump fragment of a Mail ID.

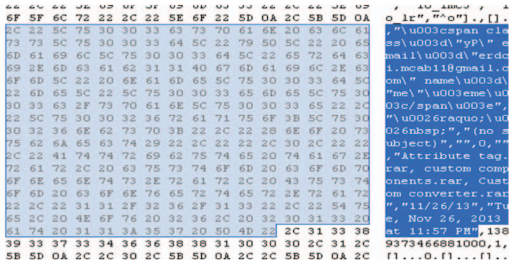

Fig. 14. Memory dump fragment of a Mail fragment.

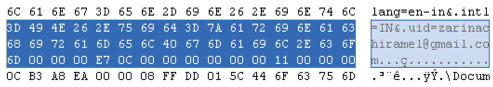

Fig. 15. Memory dump fragment showing User ID.

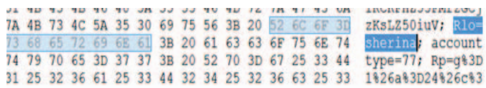

Fig. 16. Memory dump fragment showing User ID.

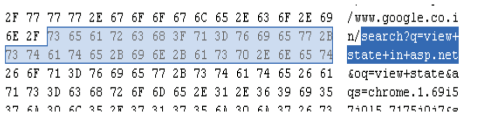

Fig. 17. Memory dump fragment of a Search query.

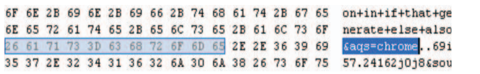

Fig. 18. Memory dump fragment of Browser information.

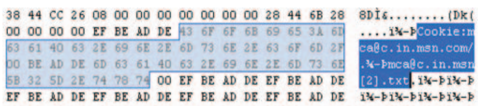

Fig. 19. Memory dump fragment of Cookie details.

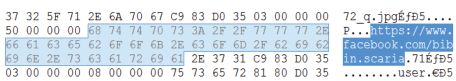

Fig. 20. Memory dump fragment of a visited pages in Facebook.

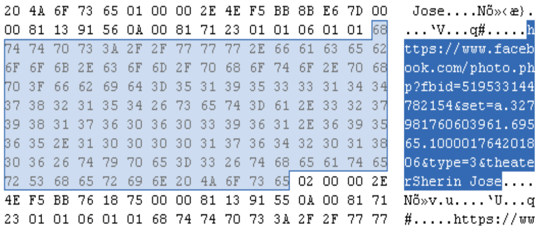

Fig. 21. Memory dump fragment of a Facebook photo.

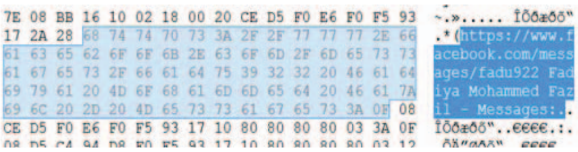

Fig. 22. Memory dump fragment of Facebook message recipient.

TABLE II. PATTERNS FOR BROWSER DATA RETRIEVAL

Data	Format
Visited sites	http://------.com https://------.com <url>-----</url>
Email	&email=---@gmail.com& RI=----%40rediffmail.com ------@yahoo.com
User id	Username= Uid= Yahoo id= &account id= RIo=…..;
Search strings	search=--+--+--; search_query=--+--;
Browser	&Source id=---; &aqs=-----; User agent:-----
cookies	Cookie:-------.txt; Cookie:cookie details
Email fragments	Forwarded message From:--mailid\username\message fragments\filename\date\time
Facebook	login_attempt=-- Email-----@----.com https ://www.facebook.com /photo.php?fbid=---- https://www.facebook.com/messages/receivername &theater

V. CHALLENGES

The recovery of user credentials and other internet related evidence from the physical memory content using a pattern based searching is a reliable method. But if the Suspect's machine is already in the switched off state at the scene of crime, live forensics cannot be adopted. In such cases, since physical memory dump is unavailable, this type of analysis cannot be conducted. Another challenge is, user credentials are usually added to the physical memory when the user enables the 'remember password' option available in the browser. If this option is not enabled, passwords may not be recoverable in this way. So, more research should be done to retrieve passwords in other cases.

VI. CONCLUSION

The methodology evolved in this research helps cyber forensics investigators to obtain the user credentials of popular internet applications used in the Suspect's machine. This is retrieved by a detailed analysis of the physical memory content collected from that machine. Since passwords of web applications are recoverable in this way, an investigator can login into the corresponding web applications and collect more crucial information about the Suspect, if it is necessary for a detailed crime investigation. The user credentials of web applications using secure version of Hyper Text Transfer Protocol (https) are also recoverable from physical memory dump file by adopting the same method. So, this paper reveals the security loopholes present in the internet while doing online banking transactions. The experimental results have an acceptable level of performance and can be adopted in a browser forensic analysis tool. Other than this, the paper presents an algorithm for retrieving forensically relevant information such as visited sites, email related information, search queries, Facebook fragments etc. from the physical memory dump file. Since internet usage is unavoidable nowadays, the chance of getting this type of information from the Suspect's machine is very high. Hence, methodologies explained in this paper may help cyber crime investigator to retrieve crucial evidence from the suspect's machine.

ACKNOWLEDGMENT

This research is supported by Resource Center for Cyber Forensics (RCCF) Department, Centre for Development of Advanced Computing (CDAC), Thiruvananthapuram, India.

REFERENCES

[1] G Thilagavathi and J Anitha, "Document Clustering in Forensic Investigation by Hybrid Approach," International Journal of Computer Applications, pp. 14-19, April 2014.

[2] Bill Nelson, Amelia Phillips, Frank Enfinger, and Christopher Steuart, Computer Forensics and Investigation, 2nd Indian Reprint, 2009.

[3] B. D. Carrier, "Digital Forensics Works," IEEE Security & Privacy, Vol.7, Issue. 2, pp. 26-29, 2009.

[4] Hay, M. Bishop, and K. Nance, "Live Analysis: Progress and Challenges," IEEE Security and Privacy, vol. 7, Mar. 2009, pp. 30- 37.

[5] Liming Cai, Jing Sha, and Wei Qian, "Study on Forensic Analysis of Physical Memory," Second International Symposium on Computer, Communication, Controland Automation. p221-224, 2013.

[6] Dija S, C Balan, Anoop V, and Ramani B, "Towards Successful Forensic Recovery of BitLocked Volumes," 6th IEEE International Conference System of Systems Engineering, pp 317-322, 2011.

[7] Lijuan Xu and Lianhai Wang, "Research on Extracting System Logged-in Password Forensically From Windows Memory Image File," Ninth International Conference on Computational Intelligence an d Security, pp 716-720, 2013.

[8] DumpIt, http://www.moonsols.com/windows-memory-toolkit/.

[9] Olajide. F, Savage. N, Akmayeva. G, and Trafford. R, "Forensic Memory Evidence of Windows Application," The 7th International Conference for Internet Technology and Secured Transactions , pp 715-718, 2012.

[10] HxD-Hex-Editor, http://download.cnet.com/HxD-Hex-Editor/3000-2352_4-10891068.html.