Digital Forensics through Application Behavior Analysis

Shuaibur Rahman  
Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan  
Email: akhunzada.shaib@gmail.com

M. N. A. Khan  
Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan  
Email: mnak2010@gmail.com

Abstract—The field of digital forensic analysis has emerged in the past two decades to counter the digital crimes and investigate the modus operandi of the culprits to secure the computer systems. With the advances in technologies and pervasive nature of the computing devices, the digital forensic analysis is becoming a challenging task. Due to ease of digital equipment and popularity of Internet, criminals have been enticed to carry out digital crimes. Digital forensic is aimed to investigate the criminal activity and bring the culprits to justice. Traditionally the static analysis is used to investigate about an incident but due to a lot of issues related the accuracy and authenticity of the static analysis, the live digital forensic analysis shows an investigator a more complete picture of memory dump. In this paper, we introduce a module for profiling behavior of application programs. Profiling of application is helpful in forensic analysis as one can easily analyze the compromised system. Profiling is also helpful to the investigator in conducting malware analysis as well as debugging a system. The concept of our model is to trace the unique process name, loaded services and called modules of the target system and store it in a database for future forensic and malware analysis. We used VMware workstation version 9.0 on Windows 7 platform so that we can get the detailed and clean image of the current state of the system. The profile of the target application includes the process name, modules and services which are specific to an application program.

Index Terms—Digital Forensic Analysis, Digital Crime Investigation, Live Forensic Analysis, Memory-based analysis, Exculpatory Evidence.

I. INTRODUCTION

Computer is an important part of our professional life now, with continuously increasing of unauthorized criminal activities digital forensic caught attention. But alongside advance technologies have prolonged criminal base which is regularly desire for discovering new way to commit crimes in more and more complicated ways. Since computer is fundamental part of our professional life now therefore, digital evidence is officially conventional in civil or criminal matters. Digital field is usually clear approximately the identification, securing, documentation and analysis of evidence for final presentation in a law of court.

Digital evidences can be in the figure of a fragment which is recovered from diverse storage devices such as history of browsing, email, and application document. The digital evidence might be also found in the shape of deleted files and different techniques are applied to recover the deleted data. Also, digital evidences can be recovered from different Storage and other medium installed in digital equipment such as cameras computers hardest, cell phones. Therefore, forensic investigation should be complete watchfully by documenting to exactly exhibit the chain of custody to put together the evidence adequate in court of law.

Traditionally the static analysis been used to examine equipment, where static analysis only used to analyze non volatile data store in permanently storage devices. What if the non volatile data or memory is so large it may be time consuming? On other hand the live analysis show more clean picture of the attack system and examiner are analyses the volatile memory and the running process, memory allocation and open network port of the attack system.

Some authorized issues about live forensics consist is inconsistency of end result when compared to diverse analysis techniques the indomitable state of rare data is inconsistent. If one investigator produces end result after analysis, the other examined end result is not similar to the first one. Another type issue is how do we auto organize a system to identify the attached devices [1].

In this paper we present a module for profiling the behavior of program application and trace down the process name what services it load and what modules it call when a specific application program launch. Before to creating the profile of specific application we must have data about that application program.

The computer crimes through Internet are increasing nowadays and digital forensics plays an important role to reduce their rampant and unchecked usage. Forensic investigation is used to find out the digital proof or evidence using different tools and it is inherently a difficult and complex process. Digital investigations take
place in three main phases. The first phase is acquisition. In this phase, the investigator takes a snapshot or images of digital device and replicates these images from the target device to some other device for in-depth analysis. The second phase is called analysis. In this phase, the investigator identifies the digital evidence using different types of techniques such as recovering the deleted files, acquiring information of user accounts, identifying information about the attached devices like USB, CD/DVD drives, external hard disks etc. The third phase is called reporting in which the investigator reconstructs the actual scenario based on the sequence of activities happened on the target system. Digital forensic analysis is divided into two main categories. The first one is dead or the static forensic analysis. During this analysis, all the target devices that are required in the analysis are shutdown. The second category is live analysis. During this type of analysis, the system stays in the boot mode and is kept alive to acquire pertinent information from the physical memory content.

Live analysis aims at gathering evidence from systems using different operations and techniques related to primary memory content. Live forensic is the most challenging kind of digital forensic investigations. To perform the live forensics, it is vital to understand the basic techniques and tools used in digital forensics. The investigator needs to acquire the complete image of a computer usage history as well as the current state through live forensic analysis tools. Though static analysis is kind of a developed part of digital forensics, but other techniques related to live analysis need to be developed to mitigate its weaknesses.

Due to rapid increase in memory size, the forensic investigators strongly recommend the live response approach for acquisition of volatile evidence. Through this technique the investigator can collect not only the information about live processes but also about the terminated and cache processes. Volatile memory analysis becomes an important piece of investigation because the physical memory could have potential evidence which investigator cannot find on the disk storage. To get hold of the incident, the volatile data acquisition is the initial step in digital investigation. Usually the investigator gathers the volatile data through live response, while the attacker might use different libraries to make the system calls to connect to the kernel and alter the volatile data.

This paper consists of six sections. Related work along with our motivation to conduct this research is provided in section II. Our proposed model is discussed in the section III, followed by describing the implementation of the proposed methodology through experimentation and experimental results in section IV. The discussion is provided in Section V. Conclusion along with future dimensions to this research are described in the last section.

II. RELATED WORK

In [1], the authors described a few live analysis techniques such as Standard user interface technique, command shell or a Telnet connection. In [2], the authors explain over all process of digital forensic investigation. Forensic analysis is a process which is used to identify the digital evidence through different forensic tools. In [3], the authors alert on the countermeasure tools to deal with rootkits, so many countermeasure tools available, for application level rootkits the examiner use CD one of trusted tool to counter the change. In [4], the authors paying attention on taking out memory page file, because the page file have data which directly related to the RAM dump. In [5], the authors propose a model for digital analysis by contravention it into different stages such as collecting evidence, examining it, and then generate a report. The present live forensic approaches suffer from some issues such as credibility, fidelity and integrity which are difficult to verify. On the other hand, anti-forensic techniques might change the static data when acquired by the investigators using different tools. In live forensic analysis, both the evidence gathering process and the analysis itself take place at the same time, so it might be difficult to recognize whether the acquired data values are legal or otherwise. Through this technique the investigator can collect not only the information about live processes but also about the terminated and cache processes. Volatile memory analysis becomes an important piece of investigation because the physical memory could have potential evidence which investigator cannot find on the disk storage. To get hold of the incident, the volatile data acquisition is the initial step in digital investigation. Usually the investigator gathers the volatile data through live response, while the attacker might use different libraries to make the system calls to connect to the kernel and alter the volatile data. In [6], the authors state the volatile memory techniques, due to quick increase in memory size, the forensic examiner stoutly advise the live forensic approach for acquire of volatile data. Through this approach the examiner collect information not only the live process but also about the executed and cache process.

In [7], the authors alert on virtualization forensic approach, virtualization approach are also used in digital forensics which after acquiring memory image and copy of hard desk boot that to some virtual machine and collect the useful evidence which examiner may not entrée in the real environment. In [8], the authors conducted live forensic analysis on two virtual machine installed two diverse operating system window XP and windows 7. In [9], the authors paying attention on virtualization forensic technique, standard computer forensic is accepted on a target system’s RAM dump. In [10] the authors propose a technique how to identify encryption keys from the memory dump, in Linux using TruCrypt software of AES.
encryption algorithms with XTS mode. In [11], the authors through an attempt to identify the important information from memory dump about the latest browser sessions of logged in users. In [12], the authors focused on how to acquire live files from the operating system that are linked with the virtual machine. References [16-46] reviewed different techniques in different domains and reported their critical evaluations.

III. PROPOSED TECHNIQUE

We map different processes, services and modules to an individual application program which are specific to it. We identify how many processes, services and modules are loaded when we launch an application program. Then we list down the processes, services and modules which are unique to the application program by separating the processes, services and modules which are directly linked to the operating system. We then store the unique processes, services and modules which are indigenous to an application program into a database table for future reference. We call this activity as profiling the application program. The saved profile of an application program can be used later on to determine the footprint of the execution of the application program for digital analysis purpose.

**Process:** An instant of computer program that is executed is called process. A process is a running program and having a specific set of data associated with it just like a task. A process is started when a program is launched.

**Service:** A service is associated to the application program. The services usually run in the background and do not have any user interface. The services may be initiated manually or starts automatically when the system is booted or restarted. Each service provides some unique functionality to an application program, therefore, there could be several services linked to an application program.

**Module:** Module is a DLL (Dynamic Link Library) or executable file which has multiple procedures and code. DLL files are coded in such a way that multiple processes can load it simultaneously. A single process can load multiple modules.

The key steps involved in our proposed technique are described below:

**A. System Preparation**

In this phase, we prepare the system to acquire clean image of the current state of the system. For this purpose, we installed VMware Workstation version 11.0 on the machine running under Windows operating system and created a new user account on Microsoft Windows 7. The purpose of installing VMware Workstation and creating this account was to get clean state of the system. In this way, we only get the necessary operation system related running processes on the machine without the fear of third party application programs that are set to run in the auto mode. After running the VMware Workstation, we launch different application programs such as MS-Word, MS-Excel etc. This account will also have some permission to access some of applications which are installed on administrator account such as “What’s Running” software and Microsoft Office.

**B. Image Acquisition**

There are several software applications which can be used for file system image acquisition such as FTK imager, “what’s running”. For experimentation in this study, we used the “what’s running” software to acquire the image and analyze it. For acquisition of images of different application programs, we perform the following steps:

- Launch “what’s running” application.
- Take image or snapshot of the current state of the system and save it on the local hard drive.
- Run the target application with and without opening a file in it.
- Take image or snapshot of the current state of the system and save it on the local hard drive.
- Close the application program.
- Take image or snapshot of the current state of the system and save it on the local hard drive.
- Close “what’s running” application.
- Log out from the user account.

The dump of images is stored in XML file on a local drive. The dump of the images is obtained in such a way that it can be easily loaded into a database and compared with other images so that an investigator can easily analyze it.

**C. Image/ Snapshot Comparison**

In this phase, we compare the saved images with the current state of the system. After comparison we will get the process name, associated services and modules which are idiosyncratic to that application program. For example, if we are going to compare the image which we had taken when MS-Word application was running on the machine and compared it with the image or current state of a system on which the MS-Word was not running, then we can effectively identify the processes, services and modules eccentric to the MS-Word. We also used the “what’s running” tool for comparing the running processes, services and modules on a system in either state i.e., by launching an application program and by not executing the application program at all.

We can also compare the two images which have been taken after and before the specific activity, by acquiring system snapshot after the running the specific application and before.

**D. Report Generation**

After comparison of the system activity state before and after execution of the specific application program, we extract all the unique services, process and module intrinsic to the application program. We repeat the same process for different application programs. We verify stored names of these unique processes, services and
modules in a database which are distinctive to different specific application programs respectively. The purpose of this is to use these application profiles in future forensic analysis of a system. By using the proposed profiling approach, the digital forensic investigator can easily identify and search specific process name, services and module in a database and determine which of the application programs were used to compromise the target system.

IV. EXPERIMENT AND RESULTS

We have performed a series of experiments to validate our proposed technique. We mainly used “what’s running” tool which was launched on a virtual machine on Window 7 platform. The “what’s running” tool provides us list of all the process, services and modules associated to each application as well takes snapshot of all the running processes and compares it with the other snapshots. Our repeated experiments observed the unique process name, services and module of Microsoft Word 2007, Excel 2007, PowerPoint 2007, Adobe Reader 10.13, Internet explorer and google chrome we saved them in a database for future digital forensic analysis use. Following is detail of our experimental results.

To get unique processes, services and modules we have to compare first four experiments very meticulously because as we know that first four experimented applications used in this study are products of Microsoft Corporation and are part of Microsoft Office, so there would be some shared modules and services. In this way, we obtain trace of the processes, modules and services which specific to each application. The results for MS-Word, MS-Excel, MS-PowerPoint, Adobe Reader and Google Chrome profiles traced through our experiments is shown in Table 1, Table 2, Table 3, Table 4 and Table 5 respectively.

<table>
<thead>
<tr>
<th>Sr #</th>
<th>Process Name</th>
<th>Services</th>
<th>Modules</th>
<th>Module Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WINWORD.EXE</td>
<td>-</td>
<td>Wwb.dll</td>
<td>Wwb.dll is part of Microsoft office word it used to run the application stably and properly on PC.</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>Wwintl.dll</td>
<td>Wwintl.dll is part of Microsoft office word, it uses by window to load lot of setting including registry setting, color setting etc of application.</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>Msostyle.dll</td>
<td>Msostyle.dll is part of Microsoft office suit; it is providing the functionality to window platform and Microsoft application. As part of office suit it create IME (input method editor) a feature which is allow user to input symbol and character which is not on the ordinary keyboard, also encoding languages.</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>Msgr3en.dll</td>
<td>Msgr3en.dll is part of Microsoft office word, it responsible for Microsoft English Natural Language Server.</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Nlsdata0013.dll</td>
<td>Nlsdata0013.dll is the part of Microsoft windows operating system having Microsoft Neutral Natural Language Server Data and Code.</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Winword.exe</td>
<td>Winword.exe is part of Microsoft office WinWord application which loads the WinWord application on PC.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sr #</th>
<th>Process Name</th>
<th>Services</th>
<th>Modules</th>
<th>Module Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EXCEL.EXE</td>
<td>-</td>
<td>Nlsdata0013.dll</td>
<td>Nlsdata0013.dll is the part of Microsoft windows operating system having Microsoft Neutral Natural Language Server Data and Code.</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Photometadatahandler.dll</td>
<td>Photometadatahandler.dll is a part of Microsoft windows operating system it is responsible for photo thumbnail extraction.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Excel.exe</td>
<td>Excel.exe is part of Microsoft office Excel application which loads the Excel application on PC.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Ppts.dll</td>
<td>Ppts.dll is belong to Microsoft windows operating system and responsible for PlugPlay troubleshooting.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Rasadhlp.dll</td>
<td>Rasadhlp.dll is belong to Microsoft windows operating system and responsible for component Remote Access AutoDial Helper.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sr #</th>
<th>Process Name</th>
<th>Services</th>
<th>Modules</th>
<th>Module Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>POWERPNT.EXE</td>
<td>-</td>
<td>Ppml.dll</td>
<td>Ppml.dll is belong to Microsoft office PowerPoint, it uses by window to load lot of setting including registry setting, color setting etc of application.</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Ppcore.dll</td>
<td>Ppcore.dll is one of window common file which is applied to PowerPoint and responsible to process and load the application.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Powerpnt.exe</td>
<td>Powerpnt.exe is part of Microsoft office PowerPoint application which loads the PowerPoint application on PC.</td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Adobe Reader Profile

<table>
<thead>
<tr>
<th>Sr #</th>
<th>Process Name</th>
<th>Services</th>
<th>Modules</th>
<th>Module Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AcroRd32.exe</td>
<td></td>
<td>Acro32.dll</td>
<td>Acro32.dll is part of Adobe Reader application used to setting and configuration.</td>
</tr>
<tr>
<td>2</td>
<td>Acro32.exe</td>
<td></td>
<td>Acro32.exe</td>
<td>Acro32.exe is part of Adobe Reader application which loads the Adobe Reader application on PC.</td>
</tr>
<tr>
<td>3</td>
<td>AcroRd32.dll</td>
<td></td>
<td>Acroform.dll</td>
<td>Acroform.dll is part of Adobe Reader and using interaction with PDF and server.</td>
</tr>
<tr>
<td>4</td>
<td>Acroform.dll</td>
<td></td>
<td>Acroform.dll</td>
<td>Acroform.dll is part of Adobe Reader and using interaction with PDF and server.</td>
</tr>
<tr>
<td>5</td>
<td>AcroRd32.exe</td>
<td></td>
<td>AcroRd32.dll</td>
<td>AcroRd32.dll is part of Adobe Reader application which loads the Adobe Reader application on PC.</td>
</tr>
<tr>
<td>6</td>
<td>AcroRd32.dll</td>
<td></td>
<td>AcroRd32.dll</td>
<td>AcroRd32.dll is part of Adobe Reader application which loads the Adobe Reader application on PC.</td>
</tr>
</tbody>
</table>

Table 5. Google Chrome Profile

<table>
<thead>
<tr>
<th>Sr #</th>
<th>Process Name</th>
<th>Services</th>
<th>Modules</th>
<th>Module Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CHROME.EXE</td>
<td></td>
<td>chrome.exe</td>
<td>Main executable of Google Chrome.</td>
</tr>
<tr>
<td>2</td>
<td>CHROME.EXE</td>
<td></td>
<td>chrome_elf.dll</td>
<td>Handle errors messages.</td>
</tr>
<tr>
<td>3</td>
<td>CHROME.EXE</td>
<td></td>
<td>chrome.dll</td>
<td>It is loaded within chrome.exe process to make the web efficient.</td>
</tr>
<tr>
<td>4</td>
<td>CHROME.EXE</td>
<td></td>
<td>ffmpegsumo.dll</td>
<td>Support video in browser.</td>
</tr>
</tbody>
</table>

V. DISCUSSION

Experiments were conducted on different applications such as Microsoft Word 2007, Microsoft Excel 2007, Microsoft PowerPoint 2007 Adobe Reader 10.13, internet explorer and Google Chrome. The aim of this study to find unique process name, services and modules related to running application, the configuration of system is we installed VMware Workstation version 11.0 on the machine running under Windows operating system and created a new user account on Microsoft Windows 7. The purpose of installing VMware Workstation and creating this account was to get clean state of the system. To trace down these specific process name, services and module we used “what’s running” tool there are plenty of tool which used for this purpose but the “what’s running” tool is used to acquire more clean and detail image of the running processes of the system as well as it provide the comparison functionality between two images. We did five to six experiments for each application because of running system there were other process relates to operating system so we get rid of that, we right down the common modules which mostly come in experiment. For confirmation of the module we also right down the description of specific modules. The first three experiments are on Microsoft application so we found some of module is overlapping so this will not be include in profile we only include those process name, services and modules which only specific to each application. The problem we faced in these experiments is we did not find the unique services of the target application. For the purpose of digital forensic and future use, we store each profile in database.

Internet Explorer executes two process when we run it for the first time one for the main application and the other is for child tab while Google chrome has Multi-process architecture it because why it open multi process when someone run this application, it puts the plug-ins and the web app in separate processes. Number of processes is about how much extension you have on browser and how much tab open.

VI. CONCLUSION AND FUTURE WORK

In this paper we have present a module profiling the behavior of the running application that what specific process name have a target application, what services it load when launch and what module it call. We performed experiments using MS-Word, MS-Excel, PowerPoint and Adobe Reader, we take image of the system after and before launch of target application save it on local drive, after that we compare both the image to find the unique process name, services and module which related to target application. We used “what’s running” tool for both image acquisition and comparing activity. The problem we faced in this experiment is we did not find the unique services yet.

The future direction of our work is, first we will conduct experiment on browsers and Skype application and our most focus in on services which uniquely loaded during execution of target application second how much memory allocated to a specific process and what is
impact on memory if we change the target document of specific application.

REFERENCES

and Computer Science (IJMECS), 5(2), 55.


Authors’ Profiles

Shuaibur Rahman obtained MS degree in Computer Science from Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan. His research areas include artificial intelligence, digital forensics and data mining techniques.

M.N.A. Khan obtained D.Phil. degree from the University of Sussex, Brighton. His research interests are in the fields of software engineering, cyber administration, digital forensic analysis and machine learning techniques.

How to cite this paper: Shuaibur Rahman, M. N. A. Khan,"Digital Forensics through Application Behavior Analysis”, International Journal of Modern Education and Computer Science(IJMECS), Vol.8, No.6, pp.50-56, 2016.DOI: 10.5815/ijmecs.2016.06.07