Design Framework for Digital Evidence Analysis Using the Virtual Machine Forensic Analysis & Recovery (VMFAR) Method

Juhartini[1]
Universitas Teknologi Mataram, Indonesia
Juhartini8815@gmail.com

Erfan Wahyudi[2,*]
Universitas Teknologi Mataram, Indonesia
*erfan.wahyudie@gmail.com**

Bahtiar Imran[3]
Universitas Teknologi Mataram, Indonesia
bahtiarimranlombok@gmail.com

Zaenudin[4]
Universitas Teknologi Mataram, Indonesia
Zen3d.itb@gmail.com

Abstract*—Virtual Machine is a virtualization technology which is most widely used today to simplify work and save hardware resources. In addition to standard use, this virtual machine is also widely used as a tool for conducting research on malware, network installations and more. The increasing use of virtualization technology is a new challenge for digital forensics experts to conduct further research related to the restoration of evidence of deleted virtual machine image. Because this Virtual Machine (VM) is also widely used by cybercrime actors to commit crimes in cyberspace, and then delete digital traces by destroying the virtual machine image that has been used or returning it to a snapshot, this technique is known as anti-forensic. Many previous studies have discussed about this VM forensics, such as VM memory dumps and snapshots. But no one has discussed the process model or flow used to perform the analysis to digital evidence in the form of a virtual machine. This study tires to identify the Virtual Machine Forensic Analysis & Recovery (VMFAR) which the researchers design as a framework for analyzing digital evidence. After implementing this framework in the process of handling digital evidence, the results of the analysis show that the experimental process was successfully carried out*Index***

Keywords— Virtual; Machine; Forensics; Recovery; Framework.

I. INTRODUCTION

Digital forensics is a sequence of process of identifying, obtaining, analyzing and presenting evidences to the court to resolve a criminal case by observing and maintaining the integrity and authenticity of the evidence[1]. The applying of digital forensics in a virtual machine is by and large called as virtual machine forensics. However, this case cannot be separated from the existence of various techniques or other methods to remove evidences, this technique is commonly called anti-forensics. From such anti-forensic techniques, removing and restoring the VM to the system's initial snapshot are categorized into the tapping of artifact and trace removal [2].

When the attacker has finished carrying out the action, the attacker immediately destroys the evidence by deleting or downloading all files on the virtual machine which is used to carry out the crime. This will certainly make it is difficult for the forensic investigator to return the file and the data or evidence stored by the perpetrator inside the VM. Because what is acquired is a drive in the operating system, that is in the operating system (virtual machine). Even though the VM has been destroyed by the perpetrator, it is possible that the file can still be returned and evidence can be found[3].

Forensic investigations on virtual machines has brought out a challenge to investigators due to their systems are different from physical computer in common. It is not corresponding with the ease of usage and the rapidity development of this technology today. Most of literature discuss about file recovery, performance optimization, and security enhancements, only a few which is deal with virtual machine forensics[4].

Based on the background above that has been described, the purpose in this research is to extract information and perform forensic analysis on the virtual machine to the files that have been downloaded by the perpetrator to retrieve the virtual operating system files and all the information inside that supports investigators to find digital evidence and solve related cases using the VMFAR framework method that the researchers has designed.

II. RELATED WORKS

Previous research duplicated a server into two or more virtual machine servers, in which each virtual machine image as the result of the duplication was run in a different VM. Various usage of VMs depended on the computing power, availability and cost. As a result, they presented a new optimization model to determine the number and type of VM required for each server that could minimize costs and ensured the availability of the SLR (Service Level Agreement). It also showed that the use of duplicate on several different VMs could be more cost-effective to run multiple servers in virtual machine rather limited the server copy to run in single VM [5].

Maintaining the integrity of original evidence is essential for the forensic examination process since only changing one bit between the gigabits will change the data and can not be undone and doubt the evidence being extracted. In traditional write-blockers, virtual machine forensics are used to maintain the integrity of the evidence and prevent the OS from altering, but it presents a more difficult challenge to be handled.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Accessing digital storage is less likely to be done, usually, the only storage that can be accessed is a virtual hard drive. It certainly has the same integrity issues as real devices and with additional complication. In this case, it is not possible to use hardwarebased write-blockers to prevent changes to the data. In his paper entitled "*A Lightweight Software Write-blocker for Virtual Machine Forensics* presented an implementation of their own write-blocker software and demonstrated how to use it in order to be conformable to ACPO principles in digital evidence [2].

In the same year, Hu Bo proposed and developed a new forensic analysis approach model named VM Forensics, which uses dynamic and static analysis to obtain the digital evidence needed in a virtual machine. The results of the experiments conducted show that the system can detect SSDT hooks effectively. Valid active and static evidence can be obtained efficiently. In addition, the system performance overhead for scanning the target VM SSDT is low enough not to affect normal VM usage [3].

Meanwhile, Rani and Geethakumari's previous research also developed an efficient approach to conducting forensic investigations in the cloud using virtual machine (VM) snapshots. This approach combines the Intrusion Detection System in VM and VMM to identify malicious VMs and improve cloud performance in terms of size and time by saving photos of malicious VMs [4].

Ajay Kumara (2015) explains that one way to view malicious activity from virtual machines is to view live virtual machines that use Lib VMI. An alternative way is to analyze the RAM dump of the virtual machine using the MFA tool. In that study, the Volatility execution speed was measured and compared with Rekall. This is done to see that Rekall's execution speed is slower for most plug-in compared to Volatility. However, both are able to overcome the semantic gap by providing information that can be read from a memory dump [5]. Meanwhile, Almulla (2016) performs memory dump analysis without changing the evidence and proposes and tests a forensic investigation model based on the NIST model to examine private cloud-based VM snapshots such as XenServer [6].

Meanwhile, Ruuhwan in 2016 conducted a forensic investigation on a Smartphone using the Integrated Digital Forensic Investigation framework version 2 (IDFIF v2) and successfully conducting experiments on case studies that have been developed [7].

III. BASIC THEORY

A. Static Forensic

Static forensics is the most method of acquisition used today by extracting, analyzing and obtaining electronic evidence which is conducted after the incident occurred. Static forensics technology is well developed, especially in aspects of digital evidence extraction, analysis, assessment, submission and conformity with applicable legal procedures (Sakhamuri, 2017).

Static analysis methods are often more effective in the process of recovering data from storage. There are some advantages of this method such as: accessing and identifying the file system; recovering deleted files that have not been overwritten by other files; specifying the file type, using the file by keywords and appropriate pattern or MAC (Modify, Access, Creation) times, and carving relevant data from a larger portion of the raw data. This static analysis method forms the basis of most digital evidence recovery processes and is widely used by legal practitioners [6].

Static Acquisition is performed on electronic evidence confiscated by officers at the scene of a crime or submitted by the suspect. Generally, this method is preferred by the investigator in collecting digital evidence because the process of data acquisition will not change the existing data on electronic evidence during the acquisition process. Before performing the acquisition on the analytics computer, the write blocker is turned on first to prevent any data changes such as hash on the drive when connected to computer analysis[7].

The challenge of the static acquisition is when it is in certain situations where the drive or the data-set is encrypted and read if only the computer is switched on and logged in with the owner's username and password, or if only the computer can only be accessed over remote network from the investigator. So the right solution for such case is to use Live Acquisition digital evidence collection method[8].

B. Digital Forensic Framework

Framework is a basic conceptual structure that used to solve or handle a complex problem. This term is often used, among others, in the software sector to describe a reusable software system design, as well as in the management field to describe a concept that allows the handling of various types or business entities homogeneously [9].

Meanwhile, in the Oxford Dictionary defines Frameworks as "a supporting structure or underlying". The computer forensic framework can be defined as a structure to support the success of a forensic investigation. It can be concluded that the goal the forensic expert is trying to achieve that the result must be the same as that other people who carrying out the same investigation. A framework also depends on a number of structures [10].

In the case of computer forensics, or forensics in general, the laws under it must be strong. Forensic investigations must be carried out scientifically and must comply with all legal requirements. Evidence must be collected in this manner to achieve the desired objectives in internal investigations, disciplinary hearings or court cases [11].

Several digital forensics that have been developed previously are Integrated Digital Forensic Investigation (IDFIF) and Integrated Digital Forensic Investigation version 2 (IDFIFv2) [12].

IV. RESEARCH METHOD

In this paper, the researchers propose a methodology for conducting acquisition and analysis. It is expected to be able to obtain information relating to existing digital evidence in accordance with the case, the method can be seen in figure 1 below.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

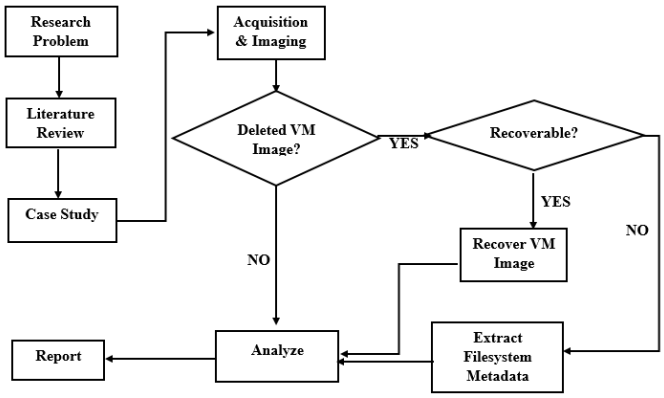

Figure 1. The Proposed Methodology

To test the methodology and support the research which is conducted, hardware and software are needed as shown in Table 1.

Tabel 1. Hardware and Software

Hardware	Software
Notebook	Oracle VirtualBox
Flashdrive 16Gb	Autopsy v4
SSD WD 60 GB	Forensic Toolkit 1.71
	FTK Imager
	Regshot
	USB Writeblocker

Case scenarios used to test the proposed method is to delete files on storage and hacking to analyze the browser history. The details can be seen in table 2 below.

Table 2 Test Scenario

No	Scenario 1	Scenario 2	Information
1	Backbox.vmdk	-	*Destroy*
2	-	Backbox 2.vmdk	*Remove From Library*
3	Recovery Deleted File in Backbox		*Delete*

V. ANALYSIS AND RESULT

A. Acquisition and Imaging

The method which applied was static acquisition method where the acquisition process is performed when the machine or device is switched off. In order to guarantee the authenticity of the result of imaging, it is necessary to record information from the acquisition process. Such information is begun and ended by acquisition process, the hash value, and the size of the imaging result file.

The MD5 hash obtained is f1a7025eb94a2d358b3e648323d61ee8 and SHA1 hash e76a5fcf62670c9a66ff27f4ed1d92c19c588622 after verifying the results match.

B. Recover VM Image

The forensic investigation process does not always run smoothly, there are various obstacles that must be faced. One of the problems that some digital evidence artifacts have been deleted by criminals, whether they are deleted in the usual way such as deleting files in general in Windows, some are deleting them on purpose to cover their tracks. The VM image recovery process needs to be done to recover files that have been completely or partially deleted so that they can be analyzed further. This is certainly a challenge for forensics investigators how to restore the files that have been deleted, either in their own VM or deleted VMI files.

To perform recovery on deleted virtual machine image files, researchers used the Forensics Toolkit and EnCase forensics. After the acquisition was carried out, the researcher succeeded in performing recovery on the deleted backbox.vmdk using the remove from library method. Then extract the backbox2.vmdk file using autopsy to analyze the files in it, and hashing to adjust the original hash value and the extraction results. As a result, the MD5 Hash value was b319377fb10b71cfa65d12d2ba4f13fd which turned out to be Match with the original MD5 Hash.

C. Filesystem Metadata Extraction

In some cases, not all recovery processes can be carried out and complicate the investigation process. Steps that can be taken if this happens is to extract metadata on the virtual machine image file system through EnCase to verify the file's metadata which contains information on file names, properties, allocated cluster numbers, and MAC (Modified, Accessed, and Created).) ever in virtual machine image that can be identified. Using this data allows investigators to investigate the virtual machine image and find out the main reason the file cannot be recovered.

D. Log Virtualbox Extraction

VBoxSVC.log is a VirtualBox log file and found in the metadata file in the form of ID: 876 which was created on 18-03-2020 at 19:28:36 and was last accessed and modified at 22:45:19 on the same date.

```
available (VERR_NOT_SUPPORTED)}, preserve=false
00:09:17.949290          ERROR [COM]: aRC=E_FAIL (0x80004005) aIID={05f2bbb6-a3a6-4fb9-9b4
9-6d0dda7142ac} aComponent={Medium} aText={UUID {dc782d1b-6804-4b24-916a-7a7485abe050} of
the medium 'C:\Users\tesis\VirtualBox VMs\Backbox\Backbox.vmdk' does not match the value {
55c42aab-e036-4cbe-b588-c4e64cdb5cbf} stored in the media registry ('C:\Users\tesis/.Virtu
alBox\VirtualBox.xml')}, preserve=false
00:21:26.196252          ERROR [COM]: aRC=VBOX_E_OBJECT_NOT_FOUND (0x80bb0001) aIID={fafa4
e17-1ee2-4905-a10e-fe7c18bf5554} aComponent={VirtualBox} aText={Could not find a registere
```

Figure 2. Backbox.vmdk information log

In the log above there is information that VirtualBox cannot find virtual machines that are in the C: / Users / theses / VirtualBox VMs / Backbox / Backbox.vmdk directory. The information printed in Figure 2 above provides information that the Backbox folder and the files in it have been deleted from the virtualbox system when it is destroyed via the VirtualBox app. After that, a search and analysis of the VirtualBox.xml registry file in the directory 'C: \ Users \ thesis / .VirtualBox' was carried out and the authors found information such as Figure 3 below.

```
    <ExtraDataItem name="GUI/LastItemSelected" value="m=Backbox 2"/>
    <ExtraDataItem name="GUI/LastWindowPosition" value="558,98,770,550"/>
    <ExtraDataItem name="GUI/RecentFolderHD" value="C:/Users/tesis/VirtualBox VMs/Backbo
x"/>
    <ExtraDataItem name="GUI/RecentListHD" value="C:\Users\tesis\VirtualBox VMs\Backbox\
Backbox.vmdk;"/>
    <ExtraDataItem name="GUI/SplitterSizes" value="156,609"/>
  </ExtraData>
```

Figure 3. VirtualBox Registry Information Log

In VirtualBox.xml, it found information stating that the backbox folder and the backbox.vmdk file are in the Recent

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Folder and Recent List which indicates that these folders and files have been accessed by the user. From this log, no footprints pointing to the storage path are found and we conclude that the backbox.vmdk file cannot be found.

E. Analyze

1.) Virtual Machine Analysis

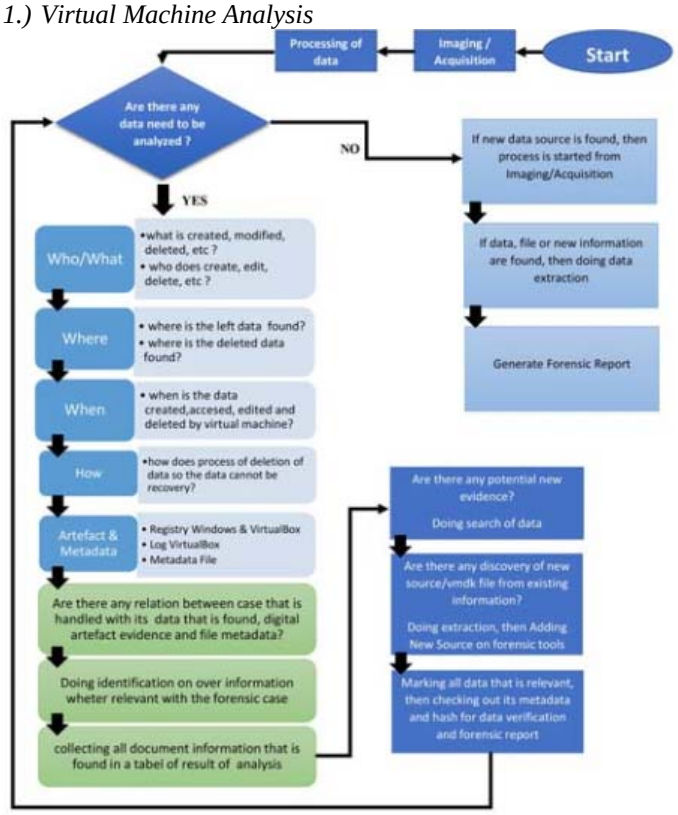

Figure 4. Stages of Analysis in VMFAR Framework

Figure 4 shows the analysis stages which are used in the Virtual Machine Forensic Analysis & Recovery Framework. The analysis of digital evidence in this study uses two different tools, that is the Forensics ToolKit and Autopsy to look for digital evidence and get more information about the evidence which can be obtained from the VMDK. The main focus is on analyzing the data in virtual machine files in the form of documents (pdf, xls, and doc), images, virtual machine images, virtual machine logs, and other data that can be used as evidence in accordance with existing scenarios. The result, in the linux operating system directory, found a "secret" folder in the path "root / home / Desktop / secret" and 2 files that have been deleted can still be recovered, one of which is the danau.jpg image.

Figure 5. Recovery Results of danau.jpg file which deleted

After analyzing the hex danau.jpg file, the researchers found no additional information. From these results the researchers concluded that Danau.jpg is an ordinary image and is not a steganography (anti-forensic) file. The results of verification of the md5 hash value can be seen in table 3 below.

Table 3. MD5 Hash Value Verification

Name	*Hash* Recovery	*Hash* Analysis	Note
Backbox 2.vmdk	b319377fb1 0b71cfa65d1 2d2ba4f13fd	b319377fb10b 71cfa65d12d2 ba4f13fd	Match
Backbox. vmdk	4ab3d1e813 405b8670c0f 7bb4ed5528 d	-	Not Match
dataku. docx	3fcc6527a00 35d8e39c05 296e07c710 d	3fcc6527a003 5d8e39c05296 e07c710d	Match
danau. jpg	1d4d4bebf0c 58dd3841be 36e7cfd04ac	1d4d4bebf0c5 8dd3841be36e 7cfd04ac	Match

From the four files found, MD5 Hash value was verified, the result was 3 files matched the original as in table 3 above, while 1 file named Backbox.vmdk could not be verified because the file could not be recovered.

2.) Registry Analysis

Registry analysis which uses regshot captures two conditions, the first is the condition when VirtualBox is installed, and the second when the virtual machine is destroyed from VirtualBox. After getting registries, then both the registry are compared to know the difference whether there is a change to the registry which is done by VirtualBox application on the operating system. After tracing, there are 3 activities that occur as follows;

1. The first activity is to add a new virtual machine, Backbox.vmdk, to the registry.
2. Then 1 value above is a virtual machine cloning process that is done in virtualbox. Backbox cloning 2.vmdk
3. Values deleted, that is registry created when destroying Backbox.vmdk

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

From the results of this analysis, it was found that the registry when the virtual machine was created and deleted, but cannot be used to recover virtual machine files that have been permanently deleted such as Backbox.vmdk.

F. Report

This research aims to apply the VMFAR Framework to search for information and potential digital evidence in deleted virtual machines and to perform recovery and analysis on virtual machines that have been destroyed.

After performing the forensic stages such as imaging & acquisition, recover VM image, file system metadata extraction, Virtualbox logs extraction, virtual machine analyze and registry analyze, it can be concluded that the VMFAR framework has been successfully tested and implemented in virtual machine forensics. The vmdk file that was destroyed from Virtualbox failed to return, this is certainly not a drawback of the VMFAR framework but an advantage of the anti-forensic method Destroy VM Image, which until now has not found a handling solution.

The destroyed virtual machine recovery process failed because of the structure and characteristics of the virtual machine itself, as well as the data deletion method which was used by VirtualBox to delete files on each of its bit. It was in contrast to the common removal in the Windows operating system which deleted it by moving to recycle bin. The process of removal of VirtualBox can be almost same as doing erase or wipe of data to the document, more details need to understand the structure of the hard drive first. [13].

VI. CONCLUSION

From the results of this study it can be concluded that the Virtual Machine Forensic Analysis & Recovery Framework (VMFAR Framework) was successfully applied to Virtualbox. This is evidenced by the vmdk file and document files that were deleted by the Remove From library removal technique and the Delete can be restored using the proposed methodology.

Meanwhile, the backbox.vmdk file that was deleted using the destroy technique on Virtualbox cannot be recovered using the forensic toolkit, even though Virtualbox log analysis, file system analysis, and registry analysis have been performed to restore the file provided no result, because deletion is done using high-level deletion techniques, such as wiping data on a hard disk. Removing virtual machines with the destroy method is very effective for anti-forensic techniques or removing traces because it can make it difficult for investigators to recover and analyze evidence.

VII. FUTURE RESEARCH

This research is far from perfect, here are some suggestions that can be made in future research:

1. In VMFAR Framework it is necessary to add the Memory Analysis (RAM) process to find out how the data deletion process occurs.
2. This research only focuses on Destroy and Remove from Library, not yet discussing snapshots, further research can add to the discussion with virtual machine snapshots.

ACKNOWLEDGEMENT

We would like to thank *Direktorat Riset dan Pengabdian Masyarakat (DRPM), Direktorat Jenderal Penguatan Riset dan Pengembangan Kementerian Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia* for funding this research through the Beginner Lecturer Research (PDP) scheme.

REFERENCES

[1]. Karen Kent, Suzanne Chevalier, Tim Grance, and Hung Dang, “Guid to Integrating Forensic Techniques into Incident Response”, National Institute of Standard and Technlogy, US Departemen Commerce, Available at nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-86.pdf, 2006.

[2]. Tri Rochmadi, Imam Riadi and Yudi Prayudi. Live Forensics for Anti-Forensics Analysis on Private Portable Web Browser. International Journal of Computer Applications 164(8):31-37, April 2017

[3]. Alamsyah R, Digital Forensic, Security Day 2010, Inixindo, Yogyakarta. 2010

[4]. Neal, C. Forensic Recovery of Evidence From Deleted Oracle Virtualbox Virtual Machines. December 2013

[5]. Sakhamuri, Das. Acquisitio of Virtual Machines for Tiered Applications with Availability Contraints, 12-135. 2017

[6]. Tobil, P., & Kechadi, M. A Lightweb Software Write-blocker for Virtual Machine Forensic, 730-735.

[7]. Song, Y., Kwak. Electronics, Information Tehcnology and Intellectualization: Proceedings of the International Conference EITI 2014, Shenzhen, China, 16-17 August 2014. CRC Press.

[8]. Nelson, B. IT Forensics, Inc., 2(1) 2011.

[9]. Nicole L., Jan. Clark., A Hierarchial, Objectives-Based Framework for Digital Investigation Process. Pre-print version of paper copyrighted and published in Digital Investigation 2(2) 2005, pp 146-166.

[10]. M. Kohn, J. H. P. Eloff, and M. S. Olivier. Framework for a Digital Forensic Investigation. In Proceedings of the ISSA 2006 from insight to Foresight Conference. Sandton, South Africa, 2006

[11]. Siti Rahayu Selamat, Robiah Yusof, Shahrin Sahib. Mapping Process of Digital Forensic Investigation Framework. International Journal of Computer Science and Network Security 2008.

[12]. Sundresan Perumal. Digital Forensic Model Based on Malaysian Investigation process. International Journal of Computer Science and Network Security 2009.

[13]. Wahyudi, Erfan & Riadi, Imam & Prayudi, Yudi. (2018). Virtual Machine Forensic Analysis And Recovery Method For Recovery And Analysis Digital Evidence. International Journal of Computer Science and Information Security,. 16.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500