# SECURITY ANALYSIS OF RECOVERED DATA IN CLOUD BASED HYPERVISOR VIRTUAL MACHINES


By

Syed Fawad Ali Shah

A thesis submitted to the faculty of Information Security Department Military College of Signals, National University of Sciences and Technology, Rawalpindi in partial fulfilment of the requirements for the degree of MS in Information Security

AUGUST 2017

# SUPERVISOR CERTIFICATE

It is to certify that final copy of thesis has been evaluated by me, found as per specific format and error free.

Dated:_____ _____

(Asst Prof Dr Rabia Latif)

## ABSTRACT

Virtualization is the foundational element of cloud computing. With the advancement in virtualization technology, virtual machines (VMs) are increasingly used by both; Data Centre and the end users, the data centres are increasingly dependent on the VMs [1]. The proliferation of virtualization environment provides a completely virtualized set of hardware to the operating system that results in increased number of illegal and inappropriate activities in the cloud environment. Virtual Machines can be both; major target of a cyber-attack or an attack vector, therefore they pertain an added level of risk; particularly corruption of data in storage and transit [2]. The focus of this research is on the acquisition as well as security and forensic analysis of the virtual machines related files from the host operating system. Further it focuses on the effects of the incidents, secure recovery and analysis of the data in hypervisor based virtual machines in cloud computing environment.

## DEDICATION

“In the name of Allah, the most Gracious, the most Compassionate".

I dedicate this thesis to my parents, wife, sisters and daughter for their support and patience; and to the faculty of Information Security Department at Military College of Signals for their valuable input and guidance.

## ACKNOWLEDGEMENT

First and the foremost, I would like to thank Allah Almighty for blessing me with good health, understanding of the subject, and fortitude for completing this thesis.

I would like to convey deepest gratitude to my supervisor Asst Prof Dr. Rabia Latif for her critical guidance, precious time, and total support.

I would also like to thank my committee members Asst Prof Mian M Waseem Iqbal and Lec Narmeen Shafqat for their valuable input, insightful analysis and precious time.

I would also like to show my deepest gratitude to Lt Col Sajid Iqbal (Retd) of Joint Conflict And Tactical Simulation (JCATS), Pakistan's Centre of Excellence for Nuclear Security (PCENS) for imposing confidence in me, and for sharing their critical analysis and ideas on this subject.

I am immensely thankful to my parents, wife, sisters and daughter who constantly encouraged and supported me throughout my MS Studies.

## Table of Contents

## LIST OF FIGURES

*Figure Number and Title* *Page*

## LIST OF TABLES

*Table Number and Title* *Page*

*Chapter 1*

# Introduction

## 1.1 Overview

The rapid emergence of cloud computing and expanding utilization of cloud technologies is driving the development of virtualization technology, like VMware for many years. With the rise in cloud computing the forensic investigators must be able to conduct forensic analysis on virtual machines or technologies exists in cloud computing [2]. There is a lot of research being done on using of VMs and virtual technologies to assist in forensic analysis, but research on collecting, recovering and analysing evidence from VMs is deficient [1]. Various researchers have highlighted that there is a need to have research in the field of cloud forensics and authenticate if conventional forensics method and tool is sufficient to conduct cloud forensics or not. [3][4][5]

## 1.2 Need for Research

According to 2016 Spiceworks report as shown in figure 1.1, more than 76% of organizations are benefiting from virtualization technology and this percentage is expected to increase up to 85% in 2017 [6]. The data proliferation, processing constraint on the existing resources, and the use of multiple operating systems on single hardware have increased the requirement of virtualization technology or VMs. However, virtualization has made the information assets vulnerable to cyber incidents. This research will therefore provide an efficient and reliable technique to recover the data efficiently from VMs for forensic analysis of virtual machines running in the cloud environment in case of any security breach.

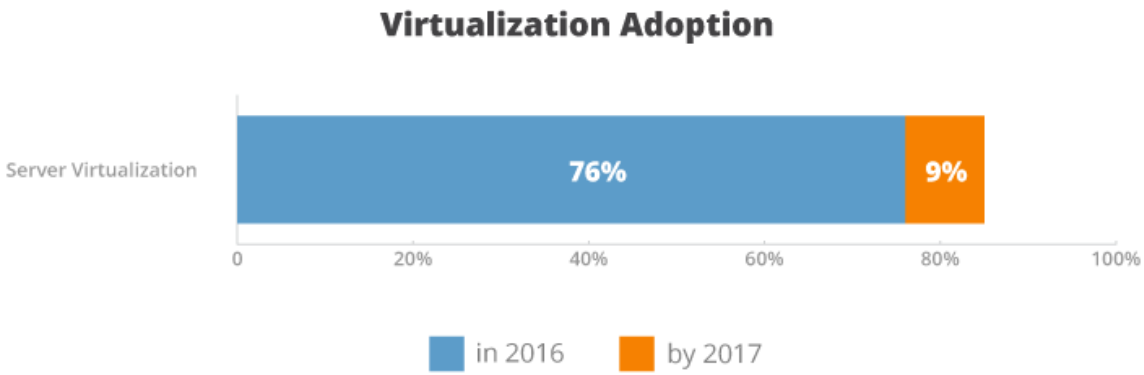

Figure 1.1 Server Virtualization and OS Trend.

## 1.3 Problem Statement

Data collection and recovery for a forensic examiner in a virtualized environment always poses challenges other than faced in physical computer forensics. With ease of use and rapid development capabilities of virtualization technology products also make it harder for forensic analysis. The literature available on virtualization technology exists on performance optimization, security consideration and disaster recovery, but there is little to no literature available on virtualization forensics with respect to being used in a host environment. The purpose of this research is to analysis and formulate a systematic procedure for data collection and recovery from a mostly used VMware vSphere Hypervisor (ESXi) virtual machine (VMs).

## 1.4 Objectives

The objectives of this research are as under:

- Use Identified tools for recovery and reconstruction of VMs.
- Formulate a systematic digital forensic procedure for recovering artefact from VMware ESXi version 6.5 Virtual Machines

- Comparative analysis with existing VMware vSphere Hypervisor ESXi version 5.5

## 1.5 Thesis Contribution

This thesis work will try to contribute to the research and development in the virtualization forensics of VMware ESXi as a base product. This will support digital forensics professional, researchers and first responders in the domain of virtualization forensics to have detailed understanding of how the process of data collection and analysis is carried out and how the artefacts are acquired. All work will be completed by developing a real time scenarios and practices on a physically developed platform. The research will also demonstrate a proper method for data acquisition and analysis tools. Moreover, this research work will highlight the comparison between the ESXi version 5.5 update 2 and version 6.5

## 1.6 Thesis Organization

The thesis is structured as follows:

- Chapter 1 starts with a brief overview of the topic, discusses details about the need for research in domain, define problem statement and research objectives; thesis contribution to the community and thesis organization.

- Chapter 2, literature review contains the introduction of hypervisor and its types, research on forensics analysis of virtualization and related work so far available to the research community.

- Chapter 3 discusses about the methodology used in this research, staring with brief introduction to data collection, analysis, procedure formulation and comparative phases. The chapter also discusses the environment created for the research along with tools and operating system used.

- Chapter 4 is about the data collection phase, as what type of data available for acquisition and in which state they are in. the chapter shows data acquisition for all the three type of data.

- Chapter 5, data analysis discusses about the analysis done on the acquired data, as what steps are required and followed to recover deleted virtual machine by identifying artefact required for recovery.

- Chapter 6 discusses about the procedure formulation, systematic steps required for data collection. The chapter also shows the steps required for analysis for recovery of deleted virtual machine files.

- Chapter 7, comparison and summary covers the research comparison by showing the steps can be performed on previous version and then discusses the limitation faced in this research and scope of future work.

*Chapter 2*

# Literature Review

## 2.1 Introduction

Virtualization has revolutionised the way the data centre operates around the world by allowing multiple virtual servers to run on single physical server by utilizing shared hardware resources. Companies have started to optimized their businesses operations by reducing the physical servers and infrastructure with the help of resource pooling. Virtualization has drastically reduced Information Technology (IT) cost which has immensely affected the cost benefit analysis (CBA) of a company. As virtualization technology is adopted by large enterprises; medium and small enterprises are also benefiting from it.

Virtualization concept first appeared in 1960's and 1970's in mainframe computing, and wasn't popular in modern computing till late 1990's when it was revived by VMware. For six years running, Gartner has ranked VMware as a "leader" in its annual "Magic Quadrant for x86 Server Virtualization Infrastructure" shown in figure 2.2 [15] [16] [17] [18]. With over 500,000 customers, presence in 100% of Fortune 100 companies, and with 2016 revenue of $7.09 billion, VMware is a massive supplier of virtualization software [2] [17]. Nowadays virtualization has matured, becoming a popular and almost necessary technology within IT operations. Gartner also estimates that "at least 80% of x86 server workloads are virtualized" [15] [16] [17] [18].


Figure 2.1 Gartner Magic Quadrant for x86 Server Virtualization Infrastructure

Even though IT community has benefited from virtualization, computer security incidents and data integrity have effected businesses and IT operations which are of main concern for the IT community. In past, few years there have been numerous high profile security breaches at companies such as Target, TJX, Living Social, and Sony resulting in the disclosure of personal information and intellectual property. For reacting to these type of incidents companies will look towards computer forensic experts to carryout forensic analysis in order to find out how attack was executed, what all was compromised and if they still have access or not.

There are numerous proprietary and open source tools available in the market for conducting digital forensics which should be tested for the ability to acquire and

analyse digital forensic evidence of a virtual machine running on top on a virtualized platform.

## 2.2 Overview

Nelson et. al. [9] defines that "computer forensics involves obtaining and analysing digital information for use as evidence in civil, criminal, or administrative cases." Research on virtual machine (VM) forensics is very limited as compared to forensics analysis on physical machines. The limited research available on virtual machines analysis exists from the perspective of hypervisor as a host system. Analysis of ESXi as a host system near to none, while virtual machine file system (VMFS) research analysis is almost non-existent.

## 2.3 Related Works

In a way, the growing use of cloud computing and increasing use of cloud technologies is driving the growth of virtualization technologies like VMware. With cloud computing on the rise, there is still a need to be able to conduct digital forensics investigations on virtual machines or appliances that exist in the cloud [7]. Several researchers have drawn attention to the fact that not much research has been done in the domain of cloud forensics and question arises that if traditional forensics tools and methods can be used to conduct forensics on the cloud [7][8][9]. Delport et. al. [3] focused on methods of isolating a cloud instance targeted for investigation in order to preserve potential evidence, much like a physical crime scene. Work was done to determine if existing digital forensics tools and acquisition methods could work to perform cloud forensics [3] [9]. Urias et. al [5] determined that many tools are not designed to deal with the complex and fluid structure of virtualization technologies

utilized in cloud environments such as the pooling of CPU, memory, and storage resources that could potentially be spread across many different physical sets of hardware. Atkison and Cruz [12] explained what tools could be used to acquire and analyse digital forensic images from virtual machines but pointed out new tools need to be created to fill the specific need of conducting digital forensics on virtual infrastructure. Martini and Choo [4] developed a six-step process to collect digital evidence from a cloud platform, utilizing VMware vCloud as a case study. In addition, a proof-of-concept program was created that made use of vClouds REST (Representational State Transfer) API (Application Programming Interface) to acquire digital forensics information following their proposed process [8].

Research has also been done to evaluate the use of virtual environment for conducting forensics analysis. After acquiring a digital forensics image from a suspect machines hard drive, it is converted into a VM allowing an investigator to boot the machine and perform digital forensics without affecting the original evidence [10].

Work specifically related to analysing a virtual machine has also been carried out in past [11]. Hirwani [1] securely acquired the virtual hard disk file and corresponding snapshots from a VMware virtual machine. After acquiring these digital forensics images, they were analysed by a program developed by the author that compared the snapshot files to determine what files had been created, deleted, or modified.

## 2.4 Hypervisors

A hypervisor is a computer software or hardware that creates and runs virtual machines, they are also called virtual machine monitor (VMM). The system on which

hypervisor runs is called *host machine* and virtual machine running on that host machine are called *guest machine*. There are two types of hypervisors:

- Type-1, native or bare-metal hypervisors
- Type-2 or hosted hypervisors

### 2.4.1 Type-1 hypervisor

Type -1 is a bare metal or native hypervisor directly runs on the hardware to control it and to run the host operating systems. Due to higher performance, security and availability type -1 hypervisor become more popular than type-2 hypervisors [10]. Some type-1 hypervisors are VMware ESXi, Oracle VM, Microsoft Hyper-V and Xen etc.

### 2.4.2 Type-2 hypervisors

Type-2 or hosted hypervisors are configured to run on a host operating system. There are number of type-2 hypervisors available in the market like VMware Workstation, VMware Fusion, Windows Virtual PC, Oracle Virtual Box etc. They are mostly used for personal use [10].

## 2.5 Hypervisors Forensic Analysis

Brett Shavers' [8], "Virtual Forensics: A Discussion of Virtual Machines Related to Forensic Analysis," provides some detailed forensic analysis information on VMware's type-2 hypervisor products, "in the context of VMware, unless otherwise noted, it is intended that VMware refers to the applications related to this paper to include VMware Workstation, VMware Player, and VMware Server". While Shavers'

paper did not focus on VMware ESXi, it only provides information on how virtual machine can be used for forensic analysis. Shavers [8] explained the challenges faced in recovering fragmented virtual machines as full recovery is not possible due to fragmentation of files, he further explained that due to large size of data files and even some of its fragmentation, full recovery becomes impossible. Table 2.1 shows the comparative analysis done the hypervisor technology and highlight the weakness/ limitation in the existing work.

Table 2.1 Comparative analysis of hypervisor technology

| Year | Authors | Paper | Description | Weakness |
|---|---|---|---|---|
| 2016 | Hu Bo, Li Nan, Liu Zhiyong, Li Min, Liu Chao | A Proactive Forensics Approach for Virtual Machines via Dynamic and Static Analysis | VM Forensics in IaaS (Infrastructure as a Service) cloud services | VM as a cloud service |
| 2016 | Joshua Sablatura, Umit Karabiyik | The Forensic Effectiveness of Virtual Disk Sanitization | Disk sanitisation in type 1 and type 2 VM hypervisors | Scope of research is disk sanitisation |
| 2016 | Sameena Naaz, Faizan Ahmad Siddiqui | Comparative Study of Cloud Forensics Tools | Discusses Cloud Forensic Tools | Does not discusses Virtualization Forensics |
| 2016 | Jidong Xiao, Lei Lu, Haining Wang, Xiaoyun Zhu | VM Introspection and Memory Forensic Analysis without Kernel Source Code | Virtual Machine Introspection | Virtual Machine Memory Forensics |
| 2013 | Meera V, Meera Mary Isaac, Balan C | Forensic Acquisition and Analysis of VMware VM Artifacts | Type 2 VMware hypervisor forensics | Only discusses type 2 VMware Hypervisor |
| 2013 | M Graziano, A Lanzi, D Balzarotti | Hypervisor Memory Forensics | Virtual Machine Introspection | Only Discusses Memory Forensics of hypervisor |
| 2012 | Manish Hirwani, Yin Pan, Bill Stackpole and Daryl Johnson | Forensic Acquisition and Analysis of VMware Virtual Hard Disks | Type 2 VMware hypervisor forensics | VM Forensics of type 2 VMware hypervisor |

## 2.6 Conclusion

This chapter contains that what are the type of hypervisors, what are their basic functionality and what are they used for, it also discusses about the research available on forensics analysis of virtualization. In this chapter it is discussed that there is lot of research available on how the virtualization can be helpful in forensics analysis but there is little to no material available on how hypervisor forensics analysis can be carried. Being type-1 hypervisor propriety to the companies, the material available on their products which can help in forensics analysis is also scarce.

*Chapter 3*

# Methodology

## 3.1 Introduction

This research methodology follows a process model described in next section. Concluding the existing literature on ESXi, VMFS file system and other VMware hypervisor forensic analysis, the focus of this research is to perform four tasks, including collection of virtual machine files from a VMFS volume, recover deleted data files from acquired virtual machine files, recover deleted virtual machine files from collected data and recover forensic artefacts from those recovered VMs.

## 3.2 Process Model

A generic process was adopted already being implemented in research community which comprises of Data Collection, Analysis, Procedure Formulation and Comparative Analysis.

### 3.2.1 Data Collection Phase.

This phase defines the potential data source identification and acquisition method of those identified data. It follows a plan which takes likelihood value of the potential data source, the resources required for acquisition of each data, process to validate volatility and integrity of each acquired data.

### 3.2.2 Analysis Phase

Assessment and extraction of relevant piece of information from acquired data and reduction of data to be analysed are done in this phase. Also, conclusions are drawn

from analysis of the extracted data. The analysis phase use methods to analyse acquired data to determine that the extracted/ acquired data can be conclusive or not.

### 3.2.3 Procedure Formulation

This section concludes the research and formulates a procedure for collection and analysis phase.

### 3.2.4 Comparative Analysis

This section analyses the effects of research on the previous versions. As the time passes so the updated versions emerge and technologies used within changes, therefore the research effectiveness is to be tested on previous version also.

## 3.3 Environment

In order to perform the steps data acquisition, analysis, procedure formulation, and comparison a test environment was created to simulate a small business using ESXi server and number of tests were performed. The test environment included VMware ESXi 6.5 running on Dell PowerEdge T620, a NAS (Network Attached Storage) on physical machine running FreeNAS 9.3 based on BSD (Berkeley Software Distribution), three Windows 7 Ultimate 64 bit machines, one running on physical machine hosted on Dell OptiPlex 9020 while other two were target machines hosted on VMware ESXi. Logically all the machines are connected to the same Ethernet network, Table 3.1, 3.2, 3.3, 3.4 and 3.5 shows the detailed specifications of the machines. Similarly, figure 3.1 & 3.2 shows the logical and physical topology of the test environment.

### 3.3.1 Environment Hardware

Each physical and virtual machine hardware specifications are as following

Table 3.1 Dell PowerEdge T620

| | |
|---|---|
| Name | ESXi |
| Operating System | VMware ESXi 6.5 |
| Physical / Virtual | Physical : Dell PowerEdge T620 |
| CPU | Intel Xeon CPU E5-2643 @ 3.3GHz |
| RAM | 16 GB |
| Storage | 300 GB SAS |
| Access Controller | iDRAC 7 (Integrated Dell Remote Access Controller) |

Table 3.2 HP Proliant DL 120 G7

| | |
|---|---|
| Name | FreeNAS |
| Operating System | FreeNAS 9.3 |
| Physical / Virtual | Physical : HP ProLiant DL120 G7 |
| CPU | Intel(R) Core(TM) i3-2100 CPU @ 3.10GH |
| RAM | 8 GB |
| Storage | 4 TB SATA |
| Access Controller | HP iLO 3 (Integrated Lights-Out 3) |

Table 3.3 Dell OptiPlex 9020

| | |
|---|---|
| Name | VMWARE MACHINE |
| Operating System | Windows 7 Ultimate 64 bit |
| Physical / Virtual | Physical : Dell OptiPlex 9020 |
| CPU | Intel(R) Core(TM) i7-4770 CPU @ 3.40GH |
| RAM | 4 GB |
| Storage | 500 GB & 4 TB SATA |
| Access Controller | Not Present |

Table 3.4 Virtual Machine Win_7_Alpha

| | |
|---|---|
| Name | Win_7_Aplha |
| Operating System | Windows 7 Ultimate 64 bit |
| Physical / Virtual | Virtual Machine |
| CPU | Intel Xeon vCPU E5-2643 @ 3.3GHz |
| RAM | 1 GB |
| Storage | 20 GB |
| Access Controller | Not Present |

Table 3.5 Virtual Machine Win_7_Bravo

| | |
|---|---|
| Name | Win_7_Bravo |
| Operating System | Windows 7 Ultimate 64 bit |
| Physical / Virtual | Virtual |
| CPU | Intel Xeon vCPU E5-2643 @ 3.3GHz |
| RAM | 1 GB |
| Storage | 20 GB |
| Access Controller | Not Present |

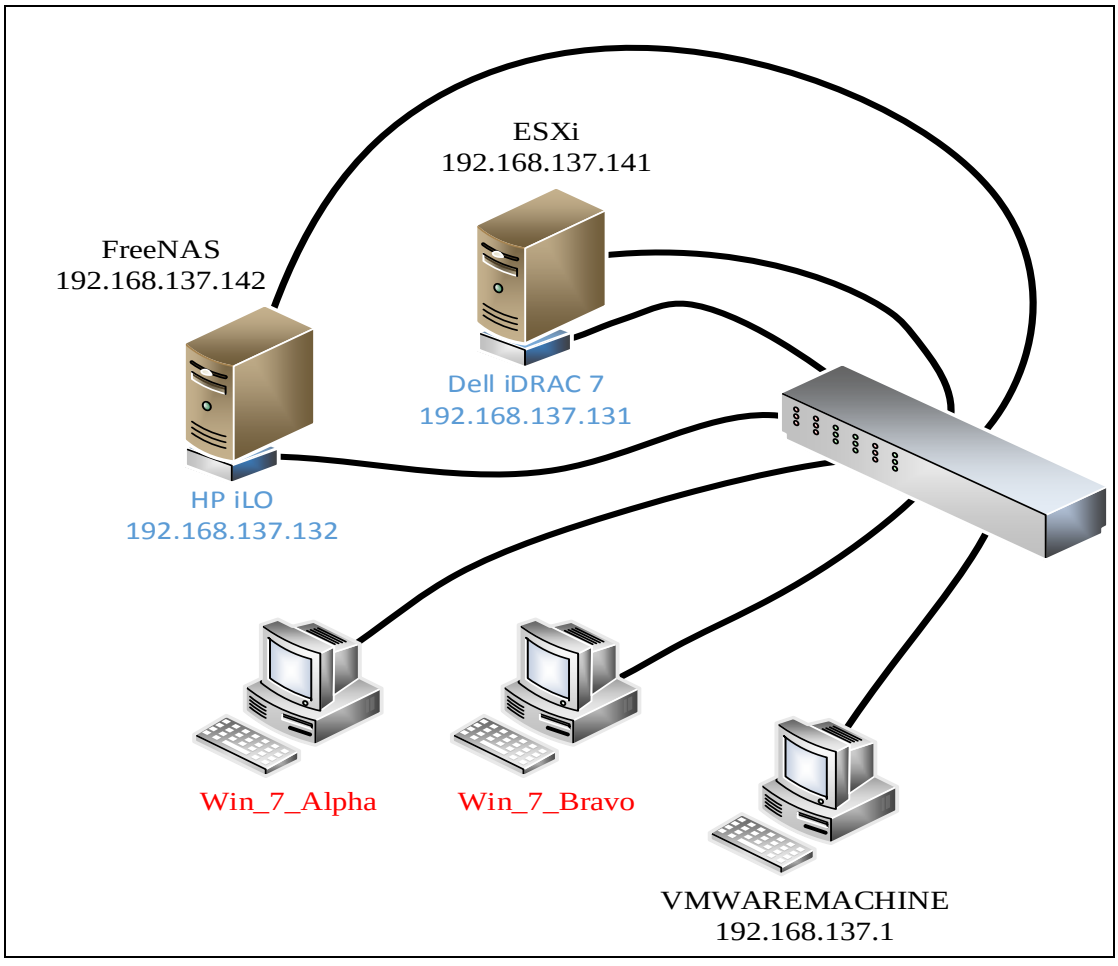


Figure 3.1 Logical Topology of Test Environment

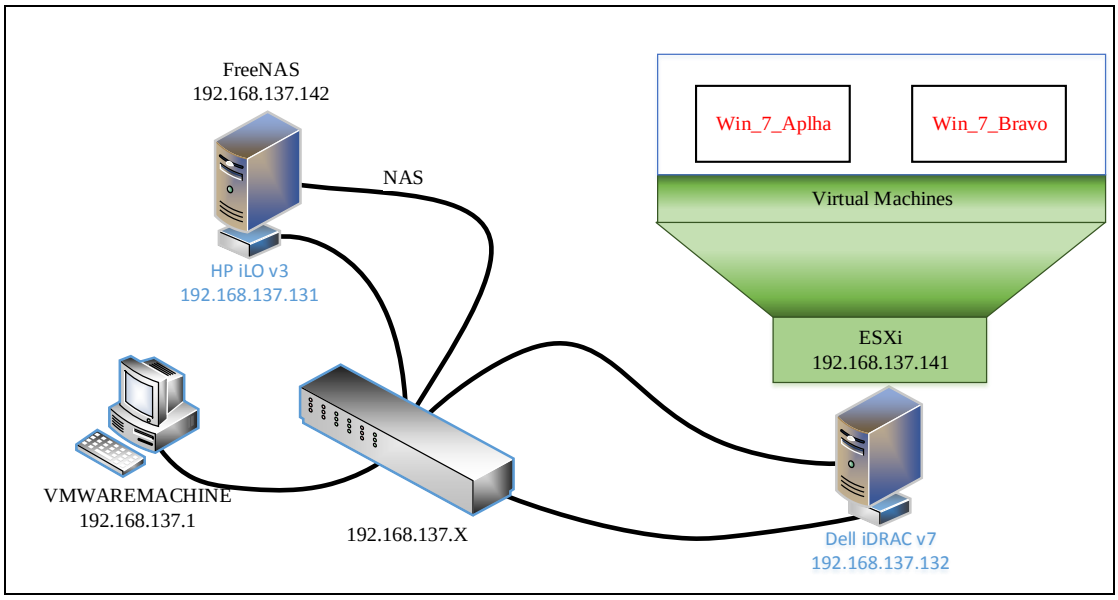


Figure 3.2 Physical Topology of Test Environment

## 3.4 Data Collection and Analysis Tools

Tools used in data acquisition and analysis phase are described in this section. All tools used in data collection, examination and analysis were open sourced and freely available. Windows 7 ultimate 64-bit machine was used as base machine on which multiple VMware machines were loaded and VMware software was used to handle other VMware machines. Kali Linux version 2016.2 was primarily used for collection, examination and analysis of data. Following are the tools used in research phase: -

- **dd v11.2**. dd also known as GNU dd is the oldest imaging tool. Apart from being very useful and requires only minimal resources, it lacks some useful features mostly used in modern imaging tools like hashing, error correction metadata gathering and a user friendly interface. It is a command line program that uses obscure input arguments for imaging purpose and if confused can destroy source data that the examiner wants to acquire. It generates raw file images which can be read by many programs. The examination and analysis phase made use of this imaging

tool for acquiring target files, partition imaging, disk imaging and sector imaging from disk images [19] [20] [21] [22] [23].

- **dc3dd v7.1.646**. dc3dd is an open source tool patched version of GNU dd command with added features for computer forensics. It was first released on 1st Feb 2008 and is developed at the Cyber Crime Centre, Department of Defense by Jesse Kornblum. Following are the features of this tool [24] [25]:
    a. On the fly multi algorithms (MD5, SHA-1. SHA-256 and SHA-512) piecewise variable sized hashing.
    b. Error write provision to file directly
    c. Error log conjunction/ grouping.
    d. Verification mode for data verification
    e. Progress bar that shows the operation progress while running
    f. Ability to split output files into fixed size chunks

- **dcfldd v1.3.4-1**. An open source tool developed by Nicolas Harbour at Department of Defense Computer Forensics Lab (DCFL). It is also an enhanced version of GNU dd command which includes security and forensics features. The collection phase uses dcfldd for acquiring whole disk image.
- **foremost v1.5.7**. foremost is an open source data recovery tool for Linux used to recover files by using file carving method through known headers, footers and data structures. The analysis phase used foremost for carving of files from disk and partition images. [19] [20]

- **vmfs-tools v0.2.5**. vmfs-tools developed by Christophe Fillot and Mike Hommey, is an open sourced program developed from vmfs code by fluidOps. It handles more feature from VMFS and allows access through standard Linux VFS using fuse framework. The analysis phase used extensively vmfs-tools to mount vmfs drives. [26]
- **xxd v1.7**. xxd written by Juergen Weigret is a Linux command which creates hexadecimal dump from standard input or input file. It can also be used to convert hex dump to back to binary form. It was used in analysis phase for showing of possible strings on drive image which helped in identifying different string files like .log and .vmdk in general.
- **hexdump**. hexdump is a command line tool used to show the raw bytes of a file in various ways including hexadecimal. Each byte is represented as two-digit hexadecimal number. It can be used with hexadecimal memory address at the beginning of each line. In analysis phase hexdump was used for showing of raw data bits.
- **GPT fdisk (gdisk) v1.0.1**. GPT gdisk or fdisk an open sourced command line utility used to modify, create and list GUID Partition Table (GPT) information. In analysis phase, it was used to gather information about the partition of disk and partition of image.
- **wxHexEditor v0.23**. wxHexEditor is an opened source cross-platform hex editor written in C++. It is faster, can work on low level disk and can handle huge files up to $2^{64}$ bytes of data, due to nature of not copying files to RAM. In analysis phase, it was used for examining partition and disk images for file and data searching.

- **fsck.** fsck is open sourced Unix utility used to check file system consistency in Unix and Linux operating system. It is similar to CHKDSK used in windows platforms. This tool was used in analysis phase to check the consistency of the disk image acquired in collection phase.
- **losetup**. A mechanism known as loopback device is used to read files as real devices. Tools used on real devices can also be used on loopback devices as an advantage to loopback device. losetup is a tool used to associate regular files with loopback devices, to detach files from loopback devices and to query the loop device [28].
- **GNU md5sum v6.4**. md5sum is an open sourced program used in GNU core utilities in Linux distributions. It calculates and verifies 128-bit MD5 hashes. It is used to verify file integrity, any change in the file or any bit change will cause change in MD5 hash. Most commonly, it is used for file integrity as a result of data transfer or disk error. During every major step of data collection, data extraction and some data analysis md5sum hash was used to verify file integrity
- **GNU sha1sum v6.5**. sha1sum is a computer program which calculates SHA-1 hashes. It is mostly used for file or data integrity. During every major step of data collection, data extraction and data analysis sha1 hash was used to verify file integrity. [19] [20]
- **debian nfs-common v1.2.6-4**. Originally developed by Sun Microsystems in 1984 Network File System (NFS) is protocol for distributed file system, allowing system to access storage on network like accessing it as a local storage. nfs-common is an open sourced tool

used in Linux distributions for file sharing. NFS is Request for Comments (RFC) open standard for anyone to use openly. NFS was used to share data bank for collection and analysis phase. [19] [20]

- **CIFS.** An enhanced version of Microsoft's open cross-platform Server Message Block (SMB) protocol. Common Internet File System (CIFS) is used to provide access to files, serial ports and printers. In analysis phase, CIFS was used to share data bank with windows platforms [27] [28].
- **Tera Term v4.94.** Tera Term is an open sourced tool used for terminal emulation (communication). It emulates different types of computer terminals and also supports telnet, SSH 1 &2 and serial port communication. This research used Tera Term for terminal connection in data collection phase.

## 3.5 Operating Systems Used

Operating systems used in this research are described below:

- **VMware ESXi (vSphere Hypervisor) v6.5**. ESXi is the name for VMware's vSphere Hypervisor. The acronym ESX stands for Elastic Sky and X was added to sound more technical and after release of version 3.5 'i' was added to signify integrity. It is a type-1 enterprise class hypervisor meaning it is not a software which can be loaded on an operating system rather it is an operating system on which multiple operating systems can be hosted.
- **Kali Linux v2016.2**. Kali Linux is derived from Debian Linux distribution specially designed for penetration testing and digital

forensics, it is funded and maintained by Offensive Security Ltd. Kali Linux has over preinstalled programs for penetration testing and forensic analysis. It can be installed on system hard disk, can be booted from live USB or CD and it can run from virtual machine. Kali Linux has evolved from BackTrack which is developed from Knoppix Linux distribution for information security testing. It is developed using secure environment and only small number of trusted people can make changes by digitally signing the packages themselves. Kali Linux was mostly used in data collection and analysis phase.

- **SANS SIFT.** The SANS (System Administration, Networking, and Security institute) Investigative Forensic Toolkit (SIFT) created by international team of forensics experts led by SANS faculty fellow Rob lee. SIFT is a computer forensics VMware machine developed on Ubuntu distribution and preconfigured with digital forensics examination necessary tools. It is used to demonstrate that advanced incident investigations and reporting to intrusions can be achieved through open source tools that are widely and freely available. The analysis phase used SIFT for initial analysis but some of the tools were not updated therefor analysis was shifted to Kali Linux platform.
- **Windows 7.** Windows developed by Microsoft is widely used as a personal operating system which makes it more prone to attacks. For this research, Windows 7 ultimate 64-bit was made a targeted platform for analysis purposes.
- **CentOS 7**. Community Enterprise Operating System (CentOS) is a Linux distribution derived from Red Hat Enterprise Linux (RHEL),

which provides an enterprise- class free computing platform. The thesis used CentOS version 7 to confirm the usability of data acquisition source code developed by Matthew Joseph Tentilucci [13].

- **FreeNAS v9.3.** FreeNAS is a Network Attached Storage (NAS) operating system which is free and open sourced that can be installed on any hardware platform to share resources on network. It is based on FreeBSD distribution and OpenZFS file system. FreeNAS supports platforms like Windows, OS X and Unix like systems and virtualization platforms such as XenServer and VMware using protocols like NFS, SMB, iSCSI, SSH and AFP etc. in this research, FreeNAS was base platform for storage sharing for this research.

## 3.6 Hardware Tools

Apart from hardware and tools mentioned above Dell iDRAC and HP iLO were used for remote access to simulate physical access to the system.

### 3.6.1 Dell iDRAC

Integrated Dell Remote Access Controller (iDRAC) is a dell out-of-band management platform embedded in every PowerEdge server with Lifecycle Controller. Due to its own resources and network connection, user can login through browser-based or command-line utility, manage the server and can reboot the server even if the core operating system has crashed. As it is installed on motherboard of the server with Lifecycle Controller therefore it requires no operation system of hypervisor to manage, configure and run the server. The remote access control gives control to an administrator as if sitting in front of the system, unlike other remote consoles it can

work even if the system is shutdown. The DRAC can remotely share disk-images as if they were physically connected to the system, which can be used to install, manage and run the operating system or hypervisor remotely. The collection phase used iDRAC v7 with enterprise license to demonstrate that no physical presence is required on the server for data acquisition.

### 3.6.2 HP iLO

Integrated Lights-Out (iLO) is a HP proprietary server management technology embedded in HP servers. It also provides out-of-band management facilities like Dell iDRAC. iLO has similar features of any lights out management (LOM) technology. Remote access is possible from remote location of HP server through iLO. As it has its own network connection to which administrator can connect to manage, configure and run sever. Some of the features are: -

- Server power resetting (In case the server doesn't respond or it crashed)
- Powering up server (power up can be possible though remote console even it is in shutdown state)
- Remote access control (remote console access is possible through separate IP to iLO)
- Virtual media connection (can mount remote media images)

iLO can also work without the presence of an operating system or hypervisor software. The collection phase used iLO v3 with advance license to simulate physical connection as physical presence is not required on the server for data acquisition.

## 3.7 Test Scenario

Test scenario was created for this research in which three types of file formats .jpg, .pdf and .doc were created.

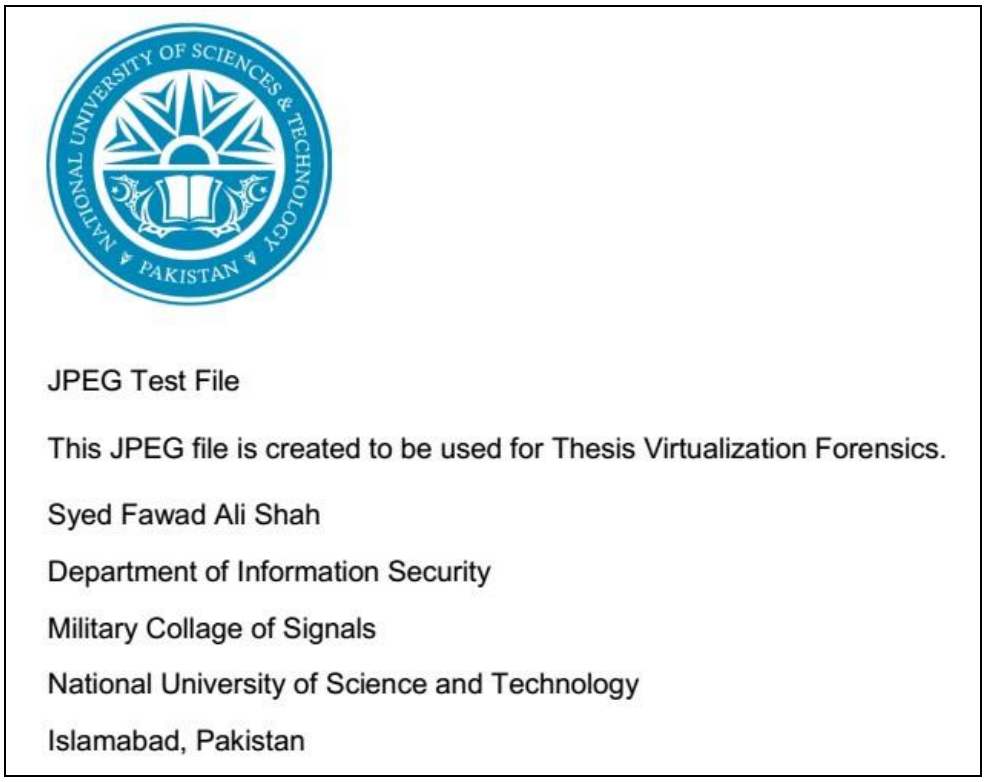


Figure 3.3 JPG Test File

Figure 3.3 shows the sample image created in jpg, the same was created in word and as well as in pdf format. The name to jpg, word and pdf are JPEG Test File.jpg, Word Test File.docx and PDF Test File.pdf respectively. The files were copied in Win 7 Alpha and Win_7_Bravo virtual machines the files were duplicated and then were deleted through recycle bin and also through shift delete method. For the test scenario the Win_7_Bravo virtual machine was deleted from VMware ESXi server. The target was to show how a file deleted from a virtual machine can be recovered without booting that image or by tools outside to that virtual machine and how a virtual machine containing deleted files can be recovered.

## 3.8 Conclusion

This section contains the conclusion of this chapter by defining the methodology used in this research for data collection, analysis and comparative phases, also this chapter contains the environment created for the research along with software tools, operating system and hardware tools used in this research. Also the test scenario created for this research was also discussed in this chapter

*Chapter 4*

# Data Collection

## 4.1 Introduction

The first step in any forensics is the data collection also known as data acquisition. Bit-stream copy is used to obtain digital evidence, also known as image of the target device [3]. Exact replica of the original data is created through bit-stream. Hashing algorithm is used to verify image integrity, it creates unique value of the input file. If any bit of the file deleted, altered or added, hash will be created differently by the hashing algorithm. Image integrity of the copied data is verified through hashing as both the target device and copied image will be of same hashing value. Message Digest (MD5) and Secure Hash Algorithm (SHA1) are commonly used to verify the integrity of data.

Many tools are available which can automatically acquire target system data from any platform apart from virtualized one, so there is a need to formulate a systematic procedure for data acquisition from virtualized platform especially from VMware ESXi. Matthew Joseph Tentilucci [ref] carried out study by the title of "Secure Acquisition of Digital Evidence from VMware ESXi Hypervisor", in which he had acquired the target data by placing it on the same hard disk drive which is the worst practise for any evidence recovery, because by placing the acquired data will replace with the data already present there but deleted or not allocated, therefore a method is needed which is systematic and does not store the data on the same hard disk drive.

## 4.2 Data and file acquisition types

Before data acquisition is performed a data acquirer should know what type of data is to be expected or what files to look for. Table 4.1 shows how many files are there and what are their types and what are they used for in VMware ESXi.

Table 4.1 Virtual Machine file descriptions

| Extension | Description |
|---|---|
| .log | Virtual machine logs are stored in this file which helps in troubleshooting and its located where configuration file is stored |
| .nvram | Virtual machine's BIOS settings are stored in NVRAM file |
| .vmdk | VMDK file contains the information about the hard disk of a virtual machine. Virtual machine can be made of single or multiple virtual disks. |
| .vmem | Paging file of virtual machine, guest main memory is backed up on host file system. It only exists when the VM is running / powered on or the VM fails / crashes. |
| .vmsd | Snapshot's metadata and allied information is stored in a centralized file. |
| .vmsn | Running state of a VM is saved in this file |
| .vmss | Stores the suspended state of a VM in suspension. |
| .vmx | Important configuration files of VM are saved in this extension |

In this chapter, three different methods are discussed for data acquisition, they are as following: -

1. Data Acquisition from Web GUI
2. Data Acquisition through terminal using VMware kernel
3. Data Acquisition using Live OS (Kali Linux)

In the above mentioned method number 1 and 2 it is assumed that the data acquirer has the root privileges which are essential, as both uses the VMware kernel to access the data and only root is allowed to access it. As for the third method is not required for to have root privileges or any type of privileges as the acquisition is done through live operating system and the system is powered off, most of the Live OS has root or admin privileges so it is assumed that acquirer also has them.

Table 4.2 Data Acquisition Types & Methods

| **Ser** | **Method** | **VMware Kernel** | **Guest OS Status** | **Remarks** |
|---|---|---|---|---|
| 1. | VMware GUI | Yes | Live | |
| 2. | VMware SSH / telnet (remote login) | Yes | Online | |
| 3. | Physical | No | Offline | HDD is removed or physical access is required to run live OS |
| 4. | Physical through iDRAC | Yes & No | Online & Offline | No physical access is required to run live OS |
| 5. | ESXi Imager | Yes | Offline | |

Table 4.2 describes the possibility of data acquisition through different methods. As there are five methods mentioned but actually they are sub parts of three methods previously mentioned. Physical and Physical through iDRAC method are one in the same as the iDRAC simulates.

Table 4.3 Target Data Types Vs Acquisition Methods

| **Target Data** | **VMware GUI** | **VMware Shell** | **Live OS** | **Remarks** |
|---|---|---|---|---|
| VMware Target Folder | Yes | Yes | Yes | |
| Whole Data store Partition | No | Yes* | Yes | *Data is being written |
| Whole Disk Drive | No | Yes* | Yes | *Data is being written |

Table 4.3 shows the comparison of target data types vs data acquisition methods, as the target data location is important for forensics examiner in a sense that the deleted data can be of same value as undeleted data. In the first two methods, data on the disk is being controlled by VMware kernel and the disk is in constant change which can affect the data that will create difficulty for the examiner but in third method the Live OS acquisition, the Host OS is dormant therefore hard disk is not being affected.

## 4.3 Data Acquisition from Web GUI

VMware data storage is visible via Web GUI as shown in figure 4.1. Acquisition can be performed by simple mouse clicks, but the disadvantage is that the whole data bank is not shown and only available data is download as to perform bit by bit data copy which is normally done by dd command in software or hardware. Data duplicators are used in hardware

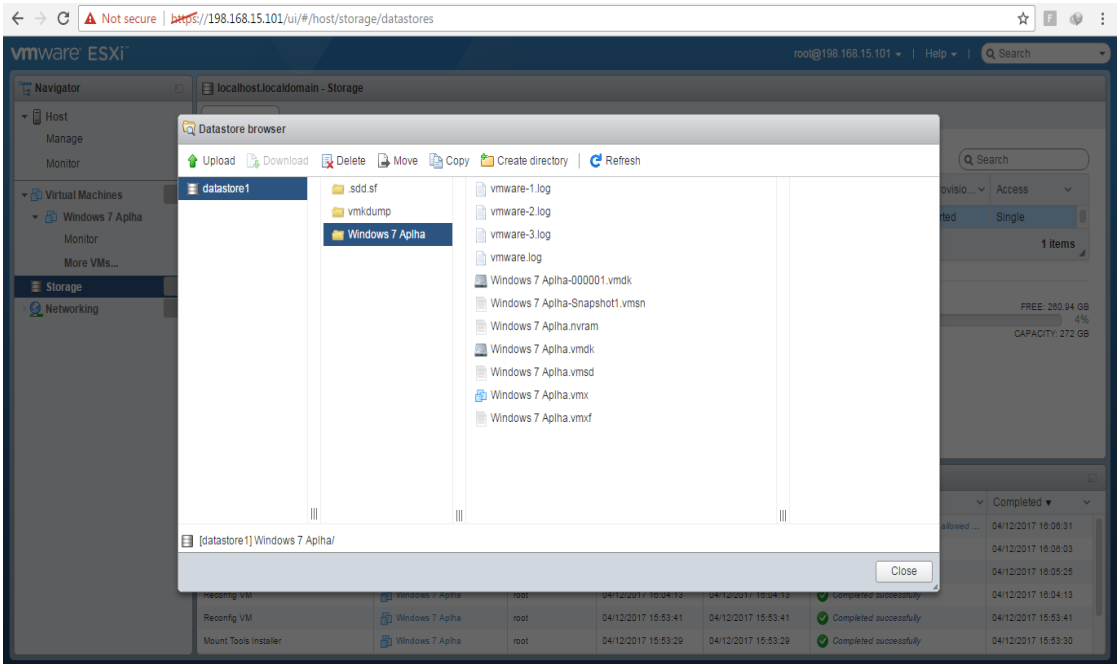


Figure 4.1 VMware vSphere Web Interface

## 4.4 Data Acquisition through terminal using VMware kernel

A platform was created to simulate the data acquisition through terminal connection using VMware Kernel. Terminal connection was accessed using SSH protocol from windows system using Tera Term using IP 192.168.137.141. The connection was made to issued dd command to make partition and disk image of the targeted drive. Before imaging the disk drive, data bank shared by FreeNAS was mounted on VMware Machine using command “esxcli storage nfs add 192.168.137.141:/NAS NAS_Mount”

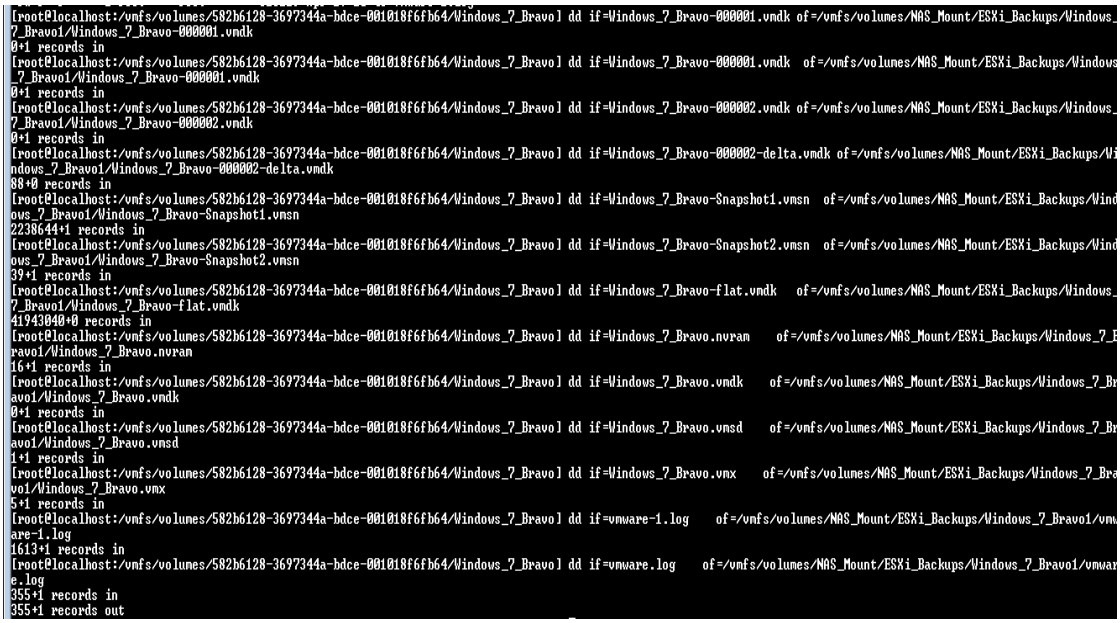


Figure 4.2 iDRAC Remote Connection

"dd if='/vmfs/volumes/datastore1/Windows_7_Bravo/files' of=/vmfs/volumes/NAS_Mount/ESXi_Backup/Window_7_Bravo/'files'" was to create dd image of each file and similarly a same dd command was used to image the disk with disk parameters. MD5 and SHA hashes were taken before and after the files and disk images to check and keep the integrity of the images. Before images was taken it was made sure that all the resident VMware machines are turned as to keep the data integrity of the drive. Also, partition image was taken using dd command along with MD5 and SHA hashes before and after partition imaging. First image was identified by issuing "df –h" command

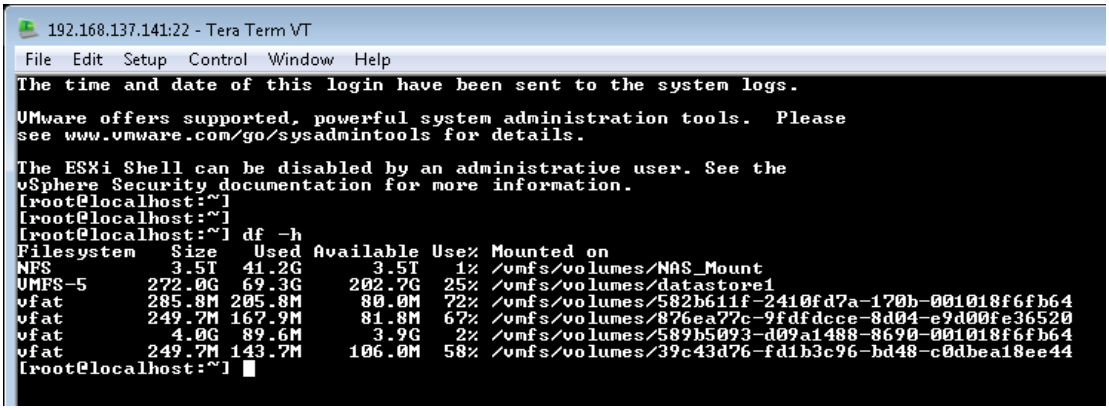


Figure 4.3 SSH Remote Connection

## 4.5 Data Acquisition using Live OS (Kali Linux)

The previous two collections methods are prone to data change which can lead to data integrity issue therefore in this section a third method is discussed which eliminates the possibility of data integrity issue which can be caused through VMware kernel. In this method, VMware ESXi server is powered off and then booted through Kali Linux Live USB mount virtually through iDRAC to simulate remote connection. Kali Linux was booted in forensic mode for the purpose of data integrity. After booting drive was searched though "fdsik –l" command results as shown in figure 4.4.


Figure 4.4 fdisk –l Command

After disks and partitions were discovered, NAS was mount through mount command to /NFS and then command “dcfldd if=/dev/sda3 hash=md5,sha256 hashwindow=1G md5log=/NFS/Drive_Backup/md5.txt sha256log=/NFS/Drive_Backup/sha256.txt hashconv=after conv=noerror,sync of=/NFS/Drive_Backup/driveimage.dd” as shown in figure 4.5 and same command was used to take drive image.

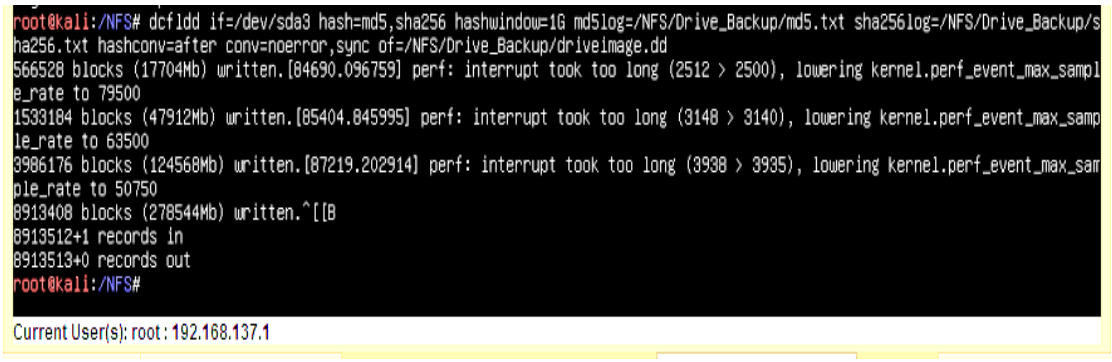


Figure 4.5 dcfldd Command

## 4.6 Conclusion

In this chapter file types used by VMware virtual machine are discussed and what type of data and files are available for data acquisition and what are the methods for their acquisition along with partition and disk imaging methods.

*Chapter 5*

# Analysis

## 5.1 Introduction

One of the objective of this thesis was to identify tool which can perform automated analysis of acquired raw data or perform data recovery of already acquired vmdk (virtual machine disk) file without booting it up. To perform analysis two operating systems platforms namely windows and Linux were used for test purpose.

## 5.2 Analysis on Windows Environment

Forensics tools are mostly available for windows platform as it is widely used at the client end. Following available tools available were tested for analysis purpose.

### 5.2.1 OSFMount

OSFMount v1.5.1015 is a tool developed by PassMark used for mounting of disk image to be used by PassMark OSForensics for forensic analysis. OSFMount was unable to read the acquired disk image, as shown in figure 5.1.

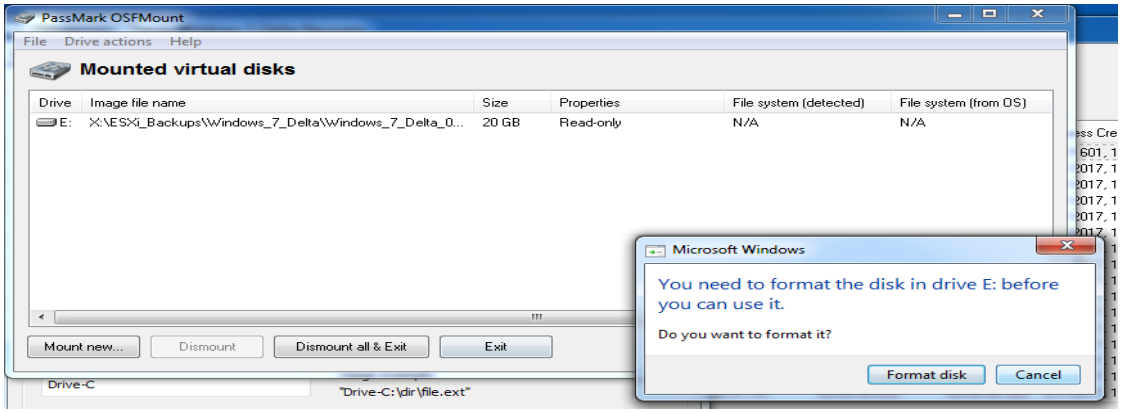


Figure 5.1 OSFMount unable to mount image

### 5.2.2 FTK

FTK v1.81.2 developed by Access Data was tested to check whether it can read the acquired disk image. Figure 5.2 shows that it was not able to open the image

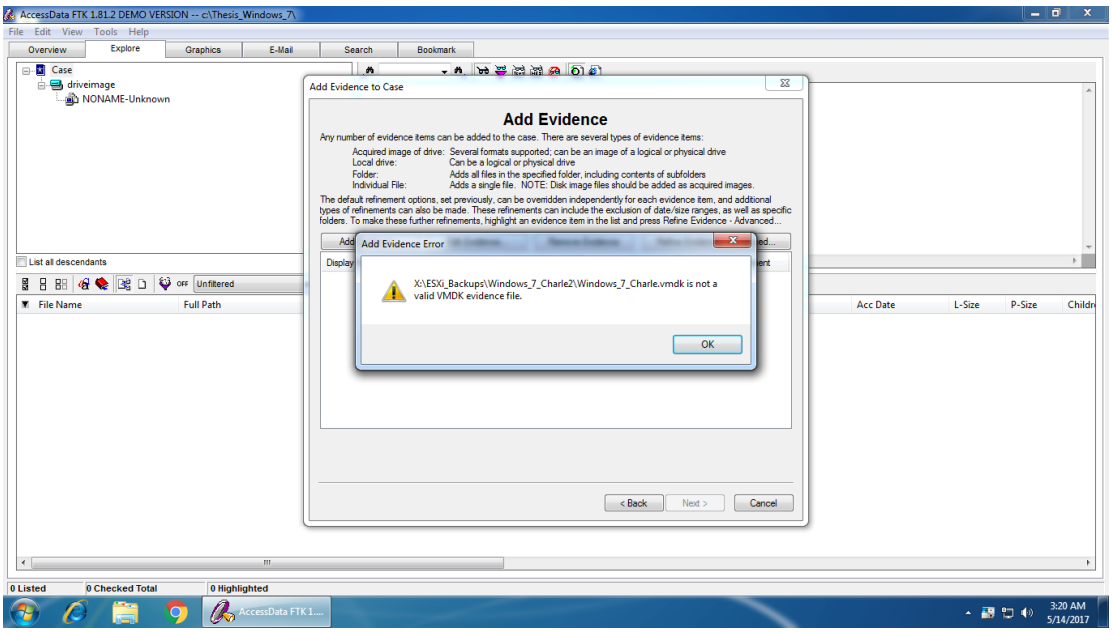


Figure 5.2 Access Data FTK

### 5.2.3 Autopsy for windows

Autopsy v4.3.0-64-bit software maintained by Basis Technology Corp uses open source programs and plugins used in The Sleuth Kit which is library and collection of Unix and Windows based utilities helping in forensics analysis. It was written by Brain Carrier and is now maintained by Basis Technology Corp. It was used in analysis Sleuth Kit read vmdk file only and was not able to read the raw images neither the partition and now the disk image. It could identify and recover the deleted sample image from the virtual machine as shown in figure 5.3.

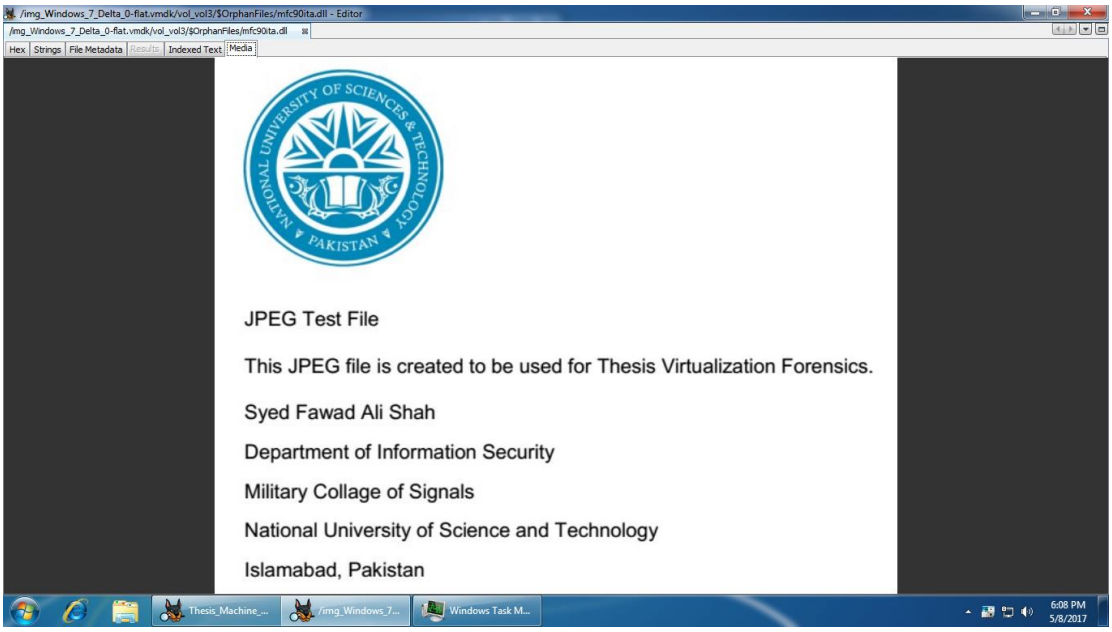


Figure 5.3 Autopsy for Windows

### 5.2.4 VMFS Recovery

A tool was identified by the name of VMFS Recovery has version 3.3 developed by DiskInternals Research. An evaluation copy was tested which was able to mount the disk image and then within disk image vmdk files were shown and recovery analysis were carried and it was able to show deleted files for recovery. For recovery, full version is required and standard version was available at the analysis for $699 [28], but this software was also unable to show deleted vmdk for recovery. Figure 5.5 and 5.6 show the images of VMFS Recovery.

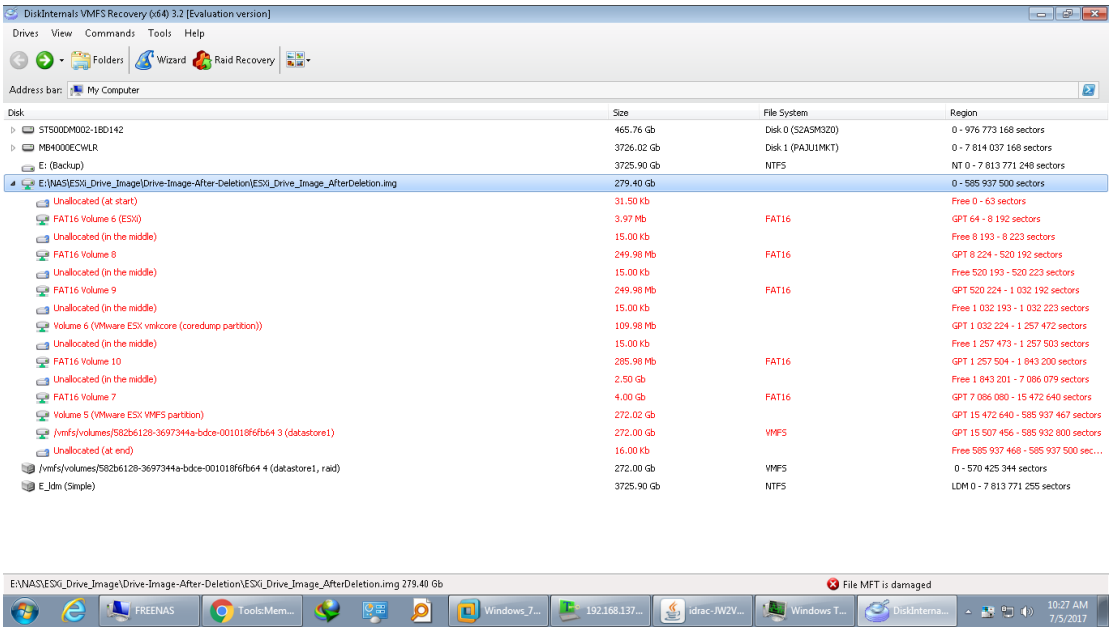


Figure 5.4 Image loaded by VMFS Recovery

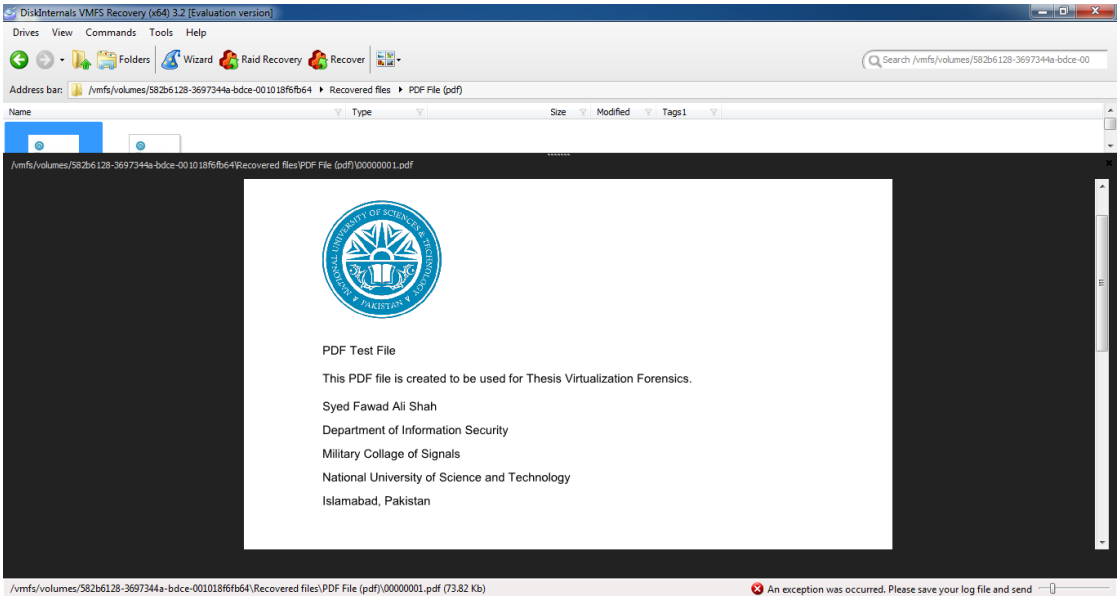


Figure 5.5 Sample file located by VMFS Recovery

## 5.3 Analysis on Linux Environment

As there is huge library for open source tools for Linux and some of them are for forensics purpose so analysis was carried on Linux environment.

### 5.3.1 Autopsy for Linux

Autopsy was also tested on Kali Linux for analysis, it was only able to mount partition image and has error with drive image. Initial analysis was able to recover the targeted file from raw partition created in Win 7 virtual. Analysis on autopsy for Linux concludes that it can search partition and able to recover know deleted files but cannot find and recover vmdk files. Results are as shown in figure 5.6.

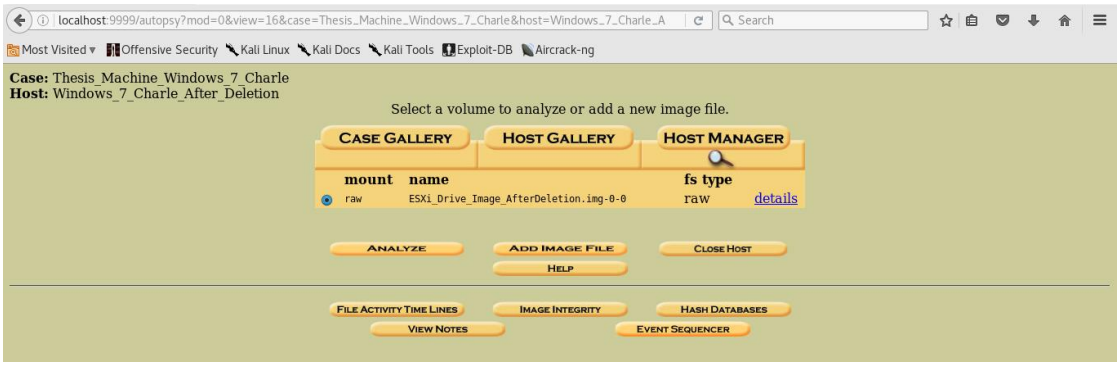


Figure 5.6 Autopsy for Linux

### 5.3.2 Manual Analysis

First in manual analysis commands were identify to mount the partition image. Though vmfs-fuse partition was mounted to /tmp/NAS_Mount/DriveImage as shown in figure 5.7.

```
root@kali:/tmp/NAS_Mount# vmfs-fuse /mnt/NAS/Drive_Backup/driveimage.dd DriveImage/
root@kali:/tmp/NAS_Mount# df -h
Filesystem                    Size  Used Avail Use% Mounted on
udev                          2.0G     0  2.0G   0% /dev
tmpfs                         394M   17M  378M   5% /run
/dev/sda1                      38G  9.5G   26G  27% /
tmpfs                         2.0G  424K  2.0G   1% /dev/shm
tmpfs                         5.0M     0  5.0M   0% /run/lock
tmpfs                         2.0G     0  2.0G   0% /sys/fs/cgroup
tmpfs                         394M   16K  394M   1% /run/user/132
tmpfs                         394M   36K  394M   1% /run/user/0
192.168.137.142:/mnt/NAS/NFS  3.6T   47G  3.5T   2% /mnt/NAS
/dev/fuse                     272G   11G  262G   4% /tmp/NAS_Mount/DriveImage
root@kali:/tmp/NAS_Mount#
```

Figure 5.7 vmfs-tool usage

After partition, first the mount was identified then the steps were identified to mount the disk image. The steps involved to first identify the sector size and starting sector of the partition where the data resides as shown in figure 5.6 by issuing parted command.

```
root@kali:/tmp/NAS_Mount# parted /mnt/NAS/ESXi_Drive_Image/Drive-Image-After-Deletion/ESXi_Drive_Image_AfterDeletion.img
GNU Parted 3.2
Using /mnt/NAS/ESXi_Drive_Image/Drive-Image-After-Deletion/ESXi_Drive_Image_AfterDeletion.img
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted) unit
Unit?  [compact]? b
(parted) print
Model:  (file)
Disk /mnt/NAS/ESXi_Drive_Image/Drive-Image-After-Deletion/ESXi_Drive_Image_AfterDeletion.img: 300000000000B
Sector size (logical/physical): 512B/512B
Partition Table: gpt
Disk Flags:

Number  Start        End            Size           File system  Name  Flags
 1      32768B       4194303B       4161536B       fat16              boot, esp
 5      4210688B     266338303B     262127616B     fat16              msftdata
 6      266354688B   528482303B     262127616B     fat16              msftdata
 7      528498688B   643825663B     115326976B
 8      643842048B   943718399B     299876352B     fat16              msftdata
 2      3628072960B  7921991679B    4293918720B    fat16              msftdata
 3      7921991680B  299999983103B  292077991424B
```

Figure 5.8 parted tool usage

Another command “fdisk –lu” was identified and used to list the partition sizes and starting sector. Figure 5.9 revealed that the targeted partition is of 272Gb and starting sector is 15472620. By multiplying the sector size with the sector location revealed that the bit address of the starting of the partition which was 7921991680.

```
bash: fdisk: command not found
root@kali:/mnt/NAS/ESXi_Drive_Image/Drive-Image-After-Deletion# fdisk -lu ESXi_Drive_Image_AfterDeletion.img
Disk ESXi_Drive_Image_AfterDeletion.img: 279.4 GiB, 300000000000 bytes, 585937500 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: gpt
Disk identifier: FBE0F571-52AC-4961-9126-7677F11BC00E

Device                                  Start       End   Sectors  Size Type
ESXi_Drive_Image_AfterDeletion.img1        64      8191      8128    4M EFI System
ESXi_Drive_Image_AfterDeletion.img2   7086080  15472639   8386560    4G Microsoft basic data
ESXi_Drive_Image_AfterDeletion.img3  15472640 585937466 570464827  272G unknown
ESXi_Drive_Image_AfterDeletion.img5      8224    520191    511968  250M Microsoft basic data
ESXi_Drive_Image_AfterDeletion.img6    520224   1032191    511968  250M Microsoft basic data
ESXi_Drive_Image_AfterDeletion.img7   1032224   1257471    225248  110M unknown
ESXi_Drive_Image_AfterDeletion.img8   1257504   1843199    585696  286M Microsoft basic data

Partition table entries are not in disk order.
```

Figure 5.9 Results of fdisk tool

```
root@kali:/mnt/NAS/ESXi_Drive_Image/Drive-Image-After-Deletion# sudo losetup -o 7921991680 /dev/loop0 ESXi_Drive_Image_AfterDeletion.img
root@kali:/mnt/NAS/ESXi_Drive_Image/Drive-Image-After-Deletion# sudo fsck -fv /dev/loop0
fsck from util-linux 2.29.2
root@kali:/mnt/NAS/ESXi_Drive_Image/Drive-Image-After-Deletion# sudo mount /dev/loop0 /tmp/NAS_Mount/
mount: unknown filesystem type 'VMFS_volume_member'
root@kali:/mnt/NAS/ESXi_Drive_Image/Drive-Image-After-Deletion# sudo mount /dev/loop0 /tmp/NAS_Mount/DriveImage/
mount: unknown filesystem type 'VMFS_volume_member'
root@kali:/mnt/NAS/ESXi_Drive_Image/Drive-Image-After-Deletion# sudo vmfs-fuse /dev/loop0 /tmp/NAS_Mount/DriveImage/
VMFS: Warning: Lun ID mismatch on /dev/loop0
ioctl: Invalid argument
ioctl: Invalid argument
root@kali:/mnt/NAS/ESXi_Drive_Image/Drive-Image-After-Deletion# ls -l /dev/loop0
brw-rw---- 1 root disk 7, 0 May 14 05:39 /dev/loop0
root@kali:/mnt/NAS/ESXi_Drive_Image/Drive-Image-After-Deletion# df -h
Filesystem                     Size  Used Avail Use% Mounted on
udev                           2.0G     0  2.0G   0% /dev
tmpfs                          394M   17M  378M   5% /run
/dev/sda1                       38G  9.5G   26G  27% /
tmpfs                          2.0G  476K  2.0G   1% /dev/shm
tmpfs                          5.0M     0  5.0M   0% /run/lock
tmpfs                          2.0G     0  2.0G   0% /sys/fs/cgroup
tmpfs                          394M   16K  394M   1% /run/user/132
tmpfs                          394M   44K  394M   1% /run/user/0
192.168.137.142:/mnt/NAS/NFS   3.6T   47G  3.5T   2% /mnt/NAS
/dev/fuse                      272G   11G  262G   4% /tmp/NAS_Mount/DriveImage
root@kali:/mnt/NAS/ESXi_Drive_Image/Drive-Image-After-Deletion# ls -l /tmp/NAS_Mount/DriveImage/
total 16
d-w-r-xr-T 2 root root  420 Apr  5 15:20 vmkdump
drwxr-xr-x 2 root root 2100 Apr 28 01:11 Windows 7 Aplha
root@kali:/mnt/NAS/ESXi_Drive_Image/Drive-Image-After-Deletion#
```

Figure 5.10 Mount process of image

Then the partition was mount from the disk image to loop device which is shown in figure 5.10. The figure shows that the disk image after mounting on the loop device is checked for alignment with “fsck -fv” command and the was mounted with command “vmfs-fuse”, “ls –l” command was used to list the partition contents which in showed one folder Windows 7 Alpha showing the files of its virtual machine. The virtual machine Windows 7 Bravo were not found as the files were deleted before the disk image was created.

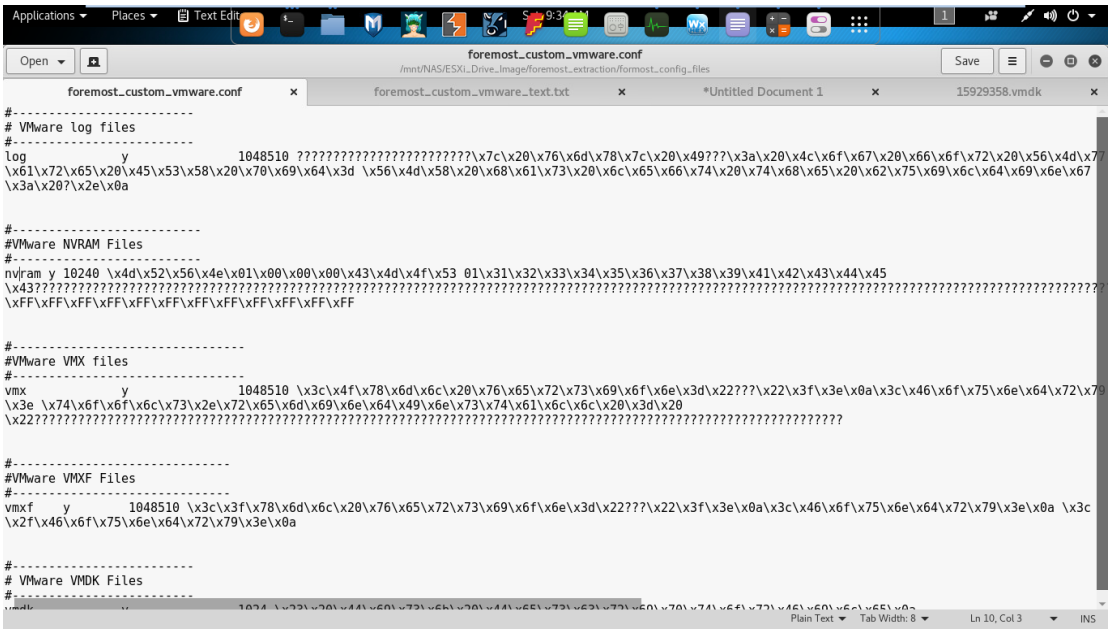


Figure 5.11 foremost custom configuration file

To recover the files of deleted virtual machine foremost tool was used with custom configuration file with file foremost_custom_vmware.conf. To create custom file available VMware logs file revealed that the file starts with specific words like "# Disk DescriptorFile" and ends on "ddb.virtualHWVersion = "13"", which help in creating the header and footer for the recovery of log file to be used for configuration of foremost configuration file. Similar method was and nvram, vmx, vmxf and vmdk files header and footers were recovered and used in foremost configuration file as "foremost –c formost_config_vmware.conf /mnt/NAS/Drive_Image.img" command,

```
root@kali:/mnt/NAS/ESXi_Drive_Image/foremost_extraction/output# ls -l
total 31
-rw-r--r-- 1 root root 2465 Jun  9 10:21 audit.txt
drwxr-xr-- 2 root root    8 Jun  9 09:19 log
drwxr-xr-- 2 root root    7 Jun  9 09:19 nvram
drwxr-xr-- 2 root root    7 Jun  9 09:08 vmdk
drwxr-xr-- 2 root root   18 Jun  9 09:19 vmx
```

Figure 5.12 Result of foremost tool

Figure 5.12 shows the output of the foremost command run with the custom configuration file.

```
root@kali:/mnt/NAS/ESXi_Drive_Image/foremost_extraction/output# cat vmdk/15929358.vmdk
# Disk DescriptorFile
version=1
encoding="UTF-8"
CID=408fa4d4
parentCID=ffffffff
isNativeSnapshot="no"
createType="vmfs"

# Extent description
RW 41943040 VMFS "Windows 7 Aplha-flat.vmdk"

# The Disk Data Base
#DDB

ddb.adapterType = "lsilogic"
ddb.geometry.cylinders = "2610"
ddb.geometry.heads = "255"
ddb.geometry.sectors = "63"
ddb.longContentID = "836e867c4e7f4cccfc382fa4408fa4d4"
ddb.thinProvisioned = "1"
ddb.toolsInstallType = "1"
ddb.toolsVersion = "10272"
ddb.uuid = "60 00 C2 9e 52 c9 cc bd-06 f8 7c 6e 52 5f 45 bb"
ddb.virtualHWVersion = "13"
```

Figure 5.12 Result of carved vmdk file

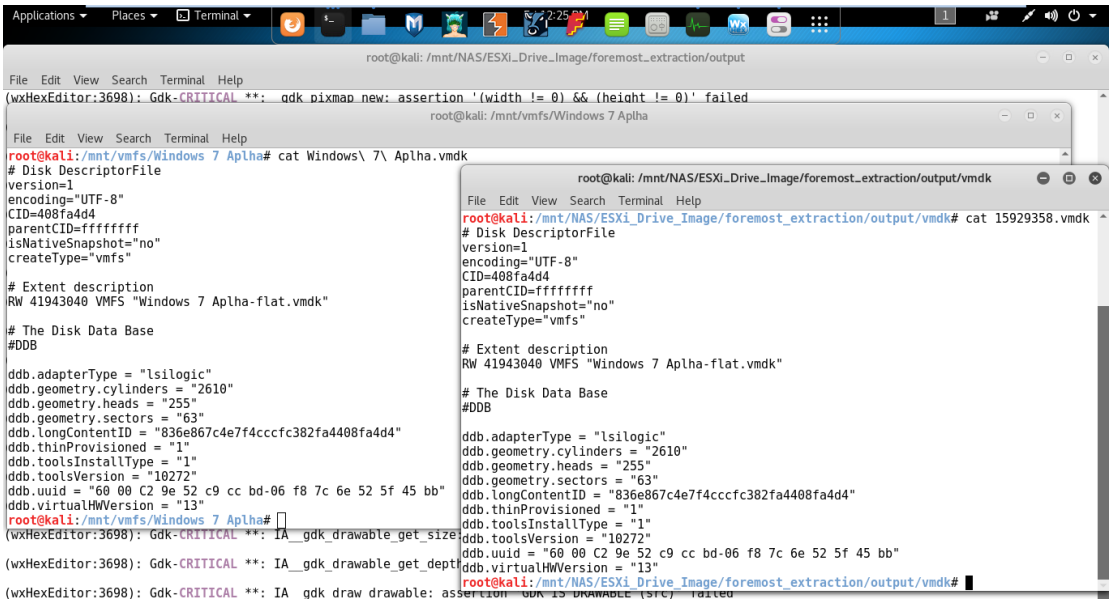


Figure 5.13 Comparison of vmdk files

Figure 5.13 shows the comparison of the recovered vmdk file with already held vmdk file. The vmdk file recovered shows information of the data vmdk file with name of “Windows 7 Alpha-flat.vmdk” and the sector size of the also shown in the file, other file contents information of the disk layout. The information acquired from vmdk file was to recover flat.vmdk of the virtual machine which is actual data file. But first the data location is located through searching the start bytes of the virtual machine data file in wxHexEditor as shown figure 5.14 which in this case are MBR of windows 7 as the virtual machine was a windows 7 operating system.

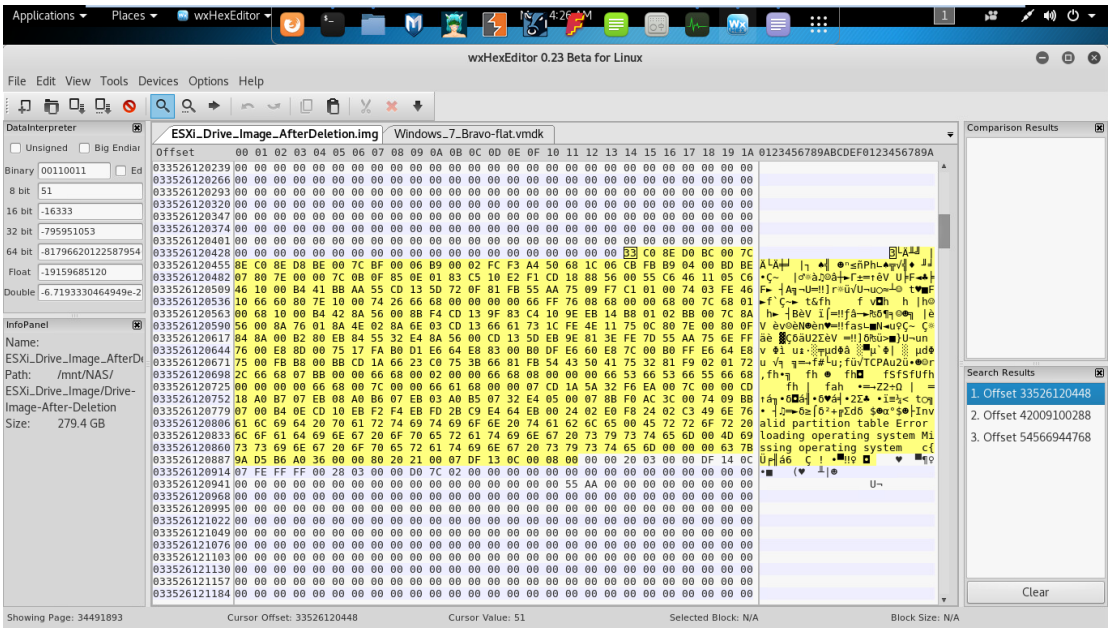


Figure 5.14 Location of start of flat.vmdk file

Figure 5.14 shows that there are three MBR matching partition exists on the disk image and that is true before taking the image there existed three images on the drive. After locating the starting byte location and data size from the recovered file the vmdk data file is extracted through dd command “dd ibs=1 status=progress skip 33536120488 count 21474836480 if=input_disk_image of=output_vmdk_file”as shown in figure 5.15

```
root@kali:/mnt/NAS2/ESXi_Drive_Image/Drive-Image-After-Deletion# dd ibs=1 status=progress skip=33526120448 count=21474836480 if=ESXi_Drive_Image_After
Deletion.img of=/mnt/NAS/Recontruction/Windows_7_Alpha/Windows_7_Aplha_flat.vmdk
167176704 bytes (167 MB, 159 MiB) copied, 53 s, 3.2 MB/s
```

Figure 5.15 dd command of file carving

After extracting the data vmdk file and combining the necessary virtual machine files in a single location the virtual machine was run in VMware Workstation. Figure 5.16 shows the successful run of the recovered virtual machine.

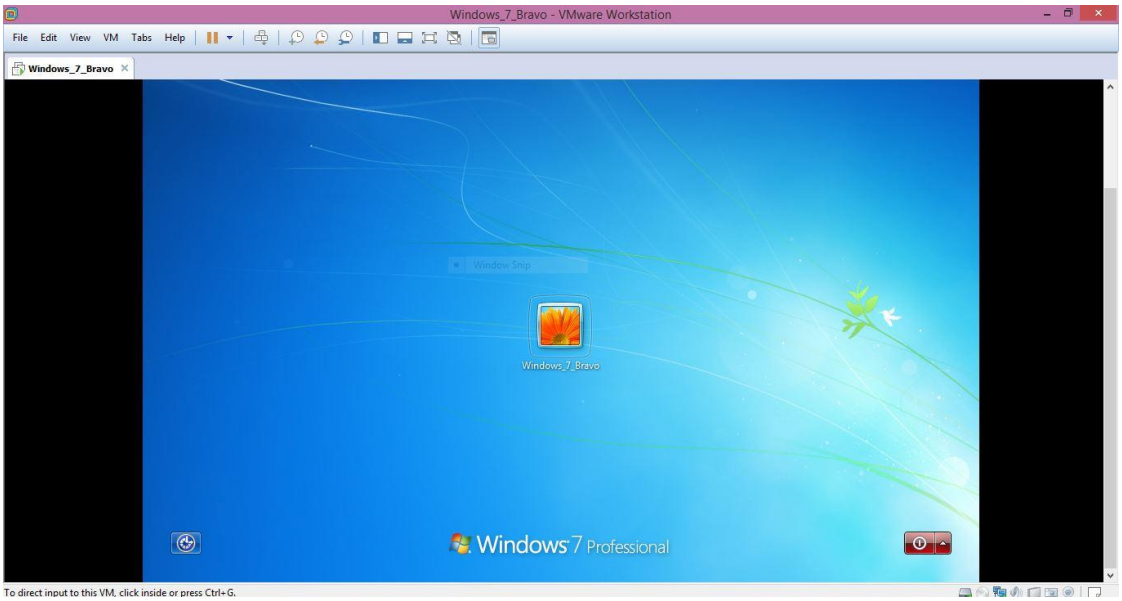


Figure 5.16 Resulted recovered virtual machine

## 5.4 Conclusion

Forensic examiner must have the ability for to work on multiple operating system platform for forensic analysis, this chapter discussed the possibility of automated forensics analysis tools available for Windows as well as Linux platform, then it also discusses the manual analysis procedure by showing the artefact required for data analysis.

*Chapter 6*

# Procedure Formulation

## 6.1 Introduction

In this chapter, procedure is formulated on the work discussed above. The procedure is divided into data collection and data analysis procedure for recovery.

## 6.2 Data Collection Procedure

For data collection, the first step is to determine the state of the data which is either online or in offline state. The next step is to identify that what type of access to the data is available as most the commands or collection requires root or administrator access so it is assumed that the root access is available and can easily be used.

The data collection procedure depends on type of data to be collected as shown in table 4.1, the result shows that the data collection form GUI isn't sufficient as all the files are not shown therefore alternate methods can be used for data collection. In other two methods dd command is used which is a core Unix and Linux utility and can easily be used. Before issuing of command the data it is important to identify either only files need to be copied and partition image is to be taken or the image of the whole drive needs to be acquired. For collection procedure, the target data location be given which can either be network attached storage (NAS) as used in this research or can be locally attached storage as in USB. Keeping above in view data collection can be defined in some of the sequential steps: -

- Access the data with root access.
- Identify the data location/ path.

- Connect the destination location for the collecting data.
- Calculate the hash of the data which is to be collected.
- Acquired the data through dd command.
- Calculate the hash of the acquired data
- Compare both the hashes and verify the integrity of the data.

In data collection procedure data integrity has utmost importance as of data collection. Data integrity can be performed through hashing, which not only ensures the integrity data but also ensures error free data.

## 6.3 Data Analysis Procedure

Data analysis procedure can be developed from the previously done research. First it identifies type of data to be recovered from acquired data, as the data is to be recovered from virtual machine or the virtual machine itself is to be recovered from vmdk files or partition or disk image. The research already done reveals that all three are data recovery either form vmdk file or virtual machines files from partition image or from the disk image, all are possible but the procedure is a bit different. The procedure to recover virtual machine from partition or disk image is same but in disk image the partition image is to be identified and then the procedure is same as of the data recovery from the partition image. The similarity in virtual machine recovery from disk and partition resemble in a way that the partition is first identifies from within the disk and the data recovery is applied on the partition.

Virtual machine data recovery from the partition occurs by first recovering the supporting files through running of foremost tool with the custom-built configuration file, then the vmdk file reveals the filename and the sector size of vmdk file. Then the starting location of the data vmdk file is searched and recovered through already known sector size that is converted into data size which is in bytes. After recovery of virtual machine files, they are placed in the same location and then virtual machine is run to verify that the machine run is correctly recovered.

As for the data recovery from within the virtual machine either VMFS Recovery tool or The Sleuth Kit can be used on windows platform. As the trail version of VMFS Recovery can only reveal the data which is to be recovered and for recovery full subscription is to be purchased, therefore The Sleuth Kit can be used to recover data from the virtual machine as it is an open source and freely available tool.

The procedure for virtual machine recovery from partition recovery can be laid down:

- Check the integrity of data
- Recover essential files with foremost with custom configuration file
- Search the location of the data file of the virtual machine
- Recover the data file from already located location with file size know the recovered vmdk file.
- Verify the recovered data

The above procedure was used to recover the deleted virtual machine files form the partition or disk image. After recovering the virtual machine then there is the recovery of the deleted files from within the virtual machine which can be perform through VMFS Recovery tool or through The Sleuth Kit tool, both can be used but VMFS Recovery is paid costing around $700 at the time of research and The Sleuth Kit is open sourced freely available. The recovery of the files also verifies the virtual machine recovery.

## 6.4 Conclusion

This chapter concludes the systematic procedures for data collection and analysis phase, the step required for data collection and then steps required for analysis on the acquired data.

*Chapter 7*

# Comparison and Summary

## 7.1 Introduction

In this chapter comparison between VMware ESXi version 5.5 and 6.5 with their results and research conclusions are discussed. The above reported research was carried out on the latest VMware platform available which was version 6.5 at time of research, so to show that the research carried is independent of the version change the same was don on previous version also.

## 7.2 Comparison between VMware ESXi versions

VMware tends to update its product regularly. This research was carried on VMware ESXi version 6.5 in order to check whether the data collection identified out previously can be applied or not. A platform was created shown in figure 7.1 and 7.2, as the purpose was to confirm the collection method therefore minimum resources were used by installing VMware version 5.5 on VMware Workstation 12 in a Windows 10 platform. Before carrying out the methods identified in this research it is to be kept in consideration that only the collection method is effected by change of the version from which the data is to be collected.

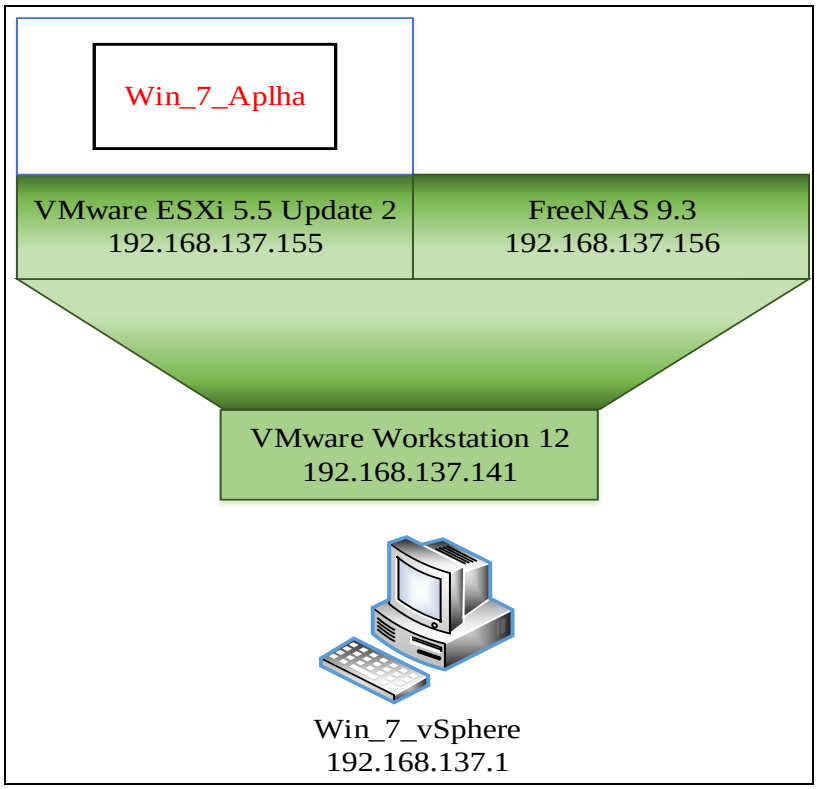


Figure 7.1 Physical Topology of Comparison Environment

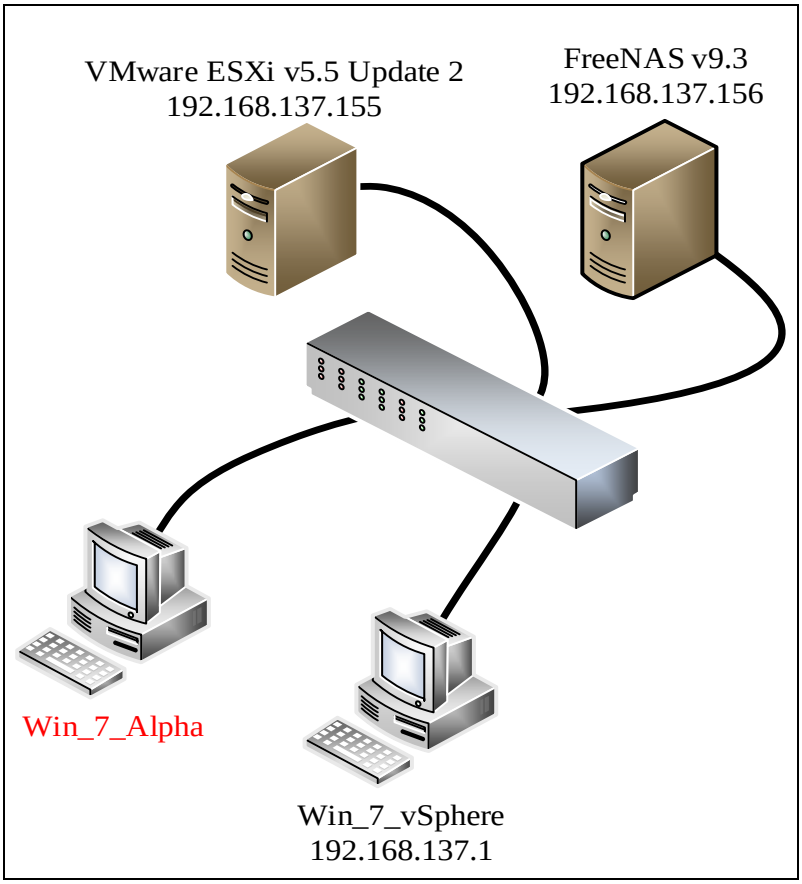


Figure 7.2 Logical Topology of Comparison Environment

The analysis conducted in this research is carried on another platform apart from the collection platform, so the version change will no effect the analysis and remains the same. The same is verified in this chapter by collecting data from a previous version and then the analysis is carried on that data which revealed that the version change has not effect.

## 7.3 Comparison Conclusion

For the data collection, it must be considered that the version change will have little to no effect as far the core utility of VMware remains which is based on Unix like kernel and tools used in the collection phase like dd and NFS in the Unix based kernels are rarely changes and the same is verified that he dd command and mount of the NAS through NFS is the same as used in the latest version. The targeted was created in same way and also easily acquired through procedure already defined. So, the version change concludes that there is little to no difference in data collection as far the core utilities used remains the same also shown in table 7.1. The data analysis is done independently from the collection platform.

Table 7.1 ESXi version comparison

| | **ESXi version 5.5** | **ESXi version 6.5** |
|---|---|---|
| VMFS version | 5 | 6 |
| VMFS v5 | Read- Write | Read- Write |
| VMFS v6 | Not Supported | Read- Write |
| dd command | Applicable | Applicable |
| NAS | Supported | Supported |
| Partition Acquisition | Possible | Possible |
| Disk Imaging | Possible | Possible |

## 7.4 Research Summary

This section consists of the limitation faced in the research phase along with future recommendation.

### 7.4.1 Limitation of the Study

The time and scope restraints limited the carried out research. There is need to perform forensics research in the field of VMware ESXi. Time restraints imposed during the research were to research further into forensics ability of the VMFS file system, disk provisioning of VMs, the VM BIOS setting file '.nvram' and also testing of methods to reduce virtual disk fragmentation. The VMFS file system still remains undiscovered by the academia. This research looked briefly into the VMFS file system some components. Specific research is required into VMFS file system understanding as how the file system works how the data is contained and how forensically it can beneficiate forensic community. Many file types are used to make a ESXi VM, this research focused on files that didn't contained any suspended or any snapshots information. This research carved deleted files from file system by reviewing each file type ".vmdk", ".vmx", ".log", "-flat.vmdk" and ".vmxf", the data within the files revealed the parent virtual machine of each file, the file name of each file and directory structure required for virtual machine reconstruction. The scope of this research did not included the analysis of the VM suspended state, snapshot, snapshot files and ".nvram" BIOS setting file. This research focused on what minimum requirements are there to require a deleted VM or deletion within VM. Expansion of this research can further lead into recovery and analysis of the remaining file types for virtual machine offline analysis being in suspended state, having snapshot files and swap files.

### 7.4.2 Future Recommendation

There is still a lot of research needed to be done on ESXi forensics by researchers. The scope of the research got restricted due to the virtual disk fragmented files, non-existence of VMFS file system documentation, lack of virtual machine supporting file documentation and lack of support of common forensics tools of VMDK file system. A major research is required into VM suspended state and snapshot files as they are the major functions/ features available with VMware ESXi. There is also little to no research material available for the analysis of VFAT which is being used by ESXi as its partitioning system. Another challenge which needs researcher's attention is the fragmentation of the virtual disk files which can be very helpful to forensics examiners. One major challenge faced during the research was the absence of the proper documentation on VMFS operations. Future research on proper documentation of the operation of VMFS file system will greatly help referencing forensics examiner with accurate and efficient evidence recovery from VMFS file system.

## 7.5 Conclusions

The research purpose was to analyse the potential of remote collection data and then analyse it to recover the deleted VMs and data within VMs. VMware ESXi version 6.5 and VMFS version 6 which also proprietary of VMware was the main focused of this research. In completion of the tasks, three questions are tried to answer in this research: How data can be acquired without loss of data integrity? What are the methods a forensic examiner can apply to recover deleted virtual machine or data within the virtual machine? What are the greatest challenges faced by an examiner in of ESXi virtual machine deleted file recovery and how can successful recovery probability be

increased? The best performing tools in this research were foremost for signature based carving of deleted files through known header and footer patterns, and dd for imaging the files, partition and disk images and also in extracting the specified files through start and end addresses of sector. The foremost carving method used already identified start and end patterns or also known as the header and footer patterns of the file, it helped in analysing the carved files of the same known types of file for comparison. The foremost custom configuration file after running against VMFS file system with having already identified header and footer revealed not only deleted but also active files present on the image. The start and end sector of the carved files were identified with wxHexEditor by searching the start and end file identifiers, the already identified start and end file addresses were carved with dd command. The dd tool was used to carve the complete sector which resided the identified file.

# BIBLIOGRAPHY


[1] Hirwani, Manish, et al. "Forensic acquisition and analysis of VMware virtual hard disks." Proceedings of the International Conference on Security and Management (SAM). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2012.

[2] Pfaff, Ben, et al. "Extending Networking into the Virtualization Layer." Hotnets. 2009.

[3] Delport, Waldo, Michael Köhn, and Martin S. Olivier. "Isolating a cloud instance for a digital forensic investigation." ISSA. 2011.

[4] Martini, Ben, and Kim-Kwang Raymond Choo. "Remote programmatic vCloud forensics: a six-step collection process and a proof of concept." 2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications. IEEE, 2014.

[5] Urias, Vincent, John Young, and Sherelle Hatcher. "Implications of Cloud Computing on Digital Forensics." GSTF Journal on Computing (JoC) 1.1 (2014).

[6] Peter, "Server Virtualization and OS Trends," https://community.spiceworks.com/networking/articles/2462-server-virtualization-and-os-trends

[7] Al Said, Taimur, and Omer F. Rana. "Analysing Virtual Machine Security in Cloud Systems." International Conference on Intelligent Cloud Computing. Springer International Publishing, 2014.

[8] Shavers, B. Virtual forensics: a discussion of virtual machines related to forensic analysis. Retrieved from http://www.forensicfocus.com/downloads/virtual-machinesforensics-analysis.pdf

[9] B. Nelson, A. Phillips, and C. Steuart, Guide to computer forensics and investigations, Cengage Learning, 2015.

[10] Virtual machine – Wikipedia, url: https://en.wikipedia.org/wiki/Virtual_machine [Accessed: 19-July-2017]

[11] Peter, “Server Virtualization and OS Trends,” https://community.spiceworks.com/networking/articles/2462-server-virtualization-and-os-trends

[12] T. Atkison and J. C. F. Cruz, “Digital Forensics on a Virtual Machine.”

[13] Matthew Joseph Tentilucci’s thesis report on “SECURE ACQUISITION OF DIGITAL EVIDENCE FROM VMWARE ESXI HYPERVISORS”

[14] Pfaff, Ben, et al. "Extending Networking into the Virtualization Layer." Hotnets. 2009.

[15] Gartner Says Worldwide Server Virtualization Market Is Reaching Its Peak, url:http://www.gartner.com/newsroom/id/3315817 [Accessed: 19-July-2017]

[16] Magic Quadrant for x86 Server Virtualization Infrastructure, published on: 03 August 2016, url:https://www.gartner.com/doc/reprints?id=1-3E2WESI&ct=160804&st=sb [Accessed: 19-July-2017]

[17] VMware and Microsoft are the top virtualization leaders, according to Gartner, url:http://www.techrepublic.com/article/vmware-and-microsoft-are-the-top-virtualization-leaders-according-to-gartner/ [Accessed: 19-July-2017]

[18] VMware Named a Leader in 2016 Magic Quadrant for x86 Server Virtualization Infrastructure, url:http://ir.vmware.com/overview/press-releases/press-release-details/2016/VMware-Named-a-Leader-in-2016-Magic-Quadrant-for-x86-Server-Virtualization-Infrastructure/default.aspx [Accessed: 19-July-2017]

[19] List of Unix commands - Wikipedia, url:https://en.wikipedia.org/wiki/List_of_Unix_commands [Accessed: 19-July-2017]

[20] List of GNU Core Utilities commands – Wikipedia, url:https://en.wikipedia.org/wiki/List_of_GNU_Core_Utilities_commands [Accessed: 19-July-2017]

[21] List of GNU packages – Wikipedia, url:https://en.wikipedia.org/wiki/List_of_GNU_packages [Accessed: 19-July-2017]

[22] GNU core utilities, url:http://info2html.sourceforge.net/cgi-bin/info2html-demo/info2html?(coreutils.info.gz)Top [Accessed: 19-July-2017]

[23] Dd – ForensicsWiki, url:http://www.forensicswiki.org/wiki/Dd [Accessed: 19-July-2017]

[24] Dc3dd - ForensicsWikihttp://www.forensicswiki.org/wiki/Dc3dd [Accessed: 19-July-2017]

[25] Dc3dd source code at sourceforge.net, url:https://sourceforge.net/projects/dc3dd/files/dc3dd/7.2.646/dc3dd%207.2.646/ [Accessed: 19-July-2017]

[26] glanium.org – vmfs-tools, url:https://glandium.org/projects/vmfs-tools/ [Accessed: 19-July-2017]

[27] Losetup (8) – Linux man page, url:https://linux.die.net/man/8/losetup [Accessed: 19-July-2017]

[28] Common Internet File System, url:https://technet.microsoft.com/en-us/library/cc939973.aspx

[29] Server Message Block – Wikipedia, url: https://en.wikipedia.org/wiki/Server_Message_Block

[30] DiskInternals Vmfs Recovery, url:https://www.diskinternals.com/order/vmfs/ [Accessed: 19-July-2017]