7-2012

Forensic Acquisition and Analysis of VMware Virtual Hard Disks

Manish Hirwani
Rochester Institute of Technology

Yin Pan
Rochester Institute of Technology

Bill Stackpole
Rochester Institute of Technology

Daryl Johnson
Rochester Institute of Technology

Recommended Citation

Hirwani M., Pan Y. , Stackpole W., and Johnson D. Forensic Acquisition and Analysis of VMware Virtual Hard Disks. In SAM'12 - The 2012 International Conference on Security and Management (Las Vegas, NV, USA, July 2012)

Forensic Acquisition and Analysis of VMware Virtual Hard Disks

Manish Hirwani, Yin Pan, Bill Stackpole and Daryl Johnson
Networking, Security and Systems Administration
Rochester Institute of Technology
mhirwani@gmail.com; {yin.pan; bill.stackpole; daryl.johnson}@it.rit.edu

Abstract— With the advancement in virtualization technology, virtual machines (VMs) are becoming a common and integral part of datacenters. As the popularity and the use of VMs increases, incidents involving them are also on the rise. There is substantial research on using VMs and virtual appliances to aid forensic investigation, but research on the appropriate forensics procedures for collecting and analyzing evidence within a VM following is lacking.

This paper presents a forensically sound way to acquire and analyze VM hard disks. A forensics tool for analyzing VM snapshots and vmdk files is developed and has been proven to be forensically sound.

Keywords
Digital forensics, Virtual Machines, virtual hard disk, Sleuthkit.

1. INTRODUCTION

Traditionally, computer systems such as desktops and servers have been considered physical devices. With the introduction of virtualization in the IT industry, this may no longer be the case.

With the benefits of virtualization, virtualization is becoming a widely adopted practice across organizations of various sizes [2, 13]. VMware is a popular provider of virtual machine (VM) software and holds a large share of the market. Products offered by VMware include VMware Workstation, VMware Server, and VMware ESXi among others. With the growing trend in virtualization, more and more production systems, workstations and desktops are being virtualized. Being a popular provider, VMware virtual environments are likely to be encountered by forensic investigators. To address the increase in exposure to VMs, this research will focus on acquisition and analysis of VMware products.

VMware VMs are implemented using virtual adapters for devices such as network cards, memory, etc. The VM, however, is stored in a set of files. VMware Workstation creates files with extension like virtual machine configuration (.vmx), virtual hard drive (.vmdk), snapshot of the virtual machines' memory
(.vmem) etc [18]. The conventional methods for incident response and evidence acquisition - such as pulling the plug of a machine containing the VMs or suspending and resuming the VMs to gather evidence - may not be the best solution for forensic analysis of VMs since resuming a VM could potentially change both volatile and non-volatile evidence leaving the evidence to be NOT admissible to court.

In this paper, the authors propose a sound forensics methodology to acquire and analyze VMs hard disk evidence by utilizing the VMware artifact – the VM files. The rest of this paper is organized as follows. Section 2 reviews the existing work performed in the area of forensic analysis of VMs. A forensically sound procedure to acquire and analyze virtual hard disks is presented in Section 3. The authors also present a forensics tool for analyzing VM snapshots and vmdk files and prove it to be forensically-sound in section 3. In Section 4, the authors evaluate the proposed forensic analysis tool. Section 5 addresses the limitations of the proposed methodology. This paper is concluded in Section 6.

2. LITERATURE REVIEW

According to Kruse & Heiser [11], the conventional forensic process can be broadly classified into four main phases, namely Acquire, Preserve, Analyze and Report. The acquisition state of the process involves capturing as much volatile system data as possible, then powering down the system and creating a forensic image of all the remaining non-volatile storage devices that are found [5]. A forensic image of a device is a bit-by-bit copy of the drive. The bit stream copy can be either stored as a file on another device or can directly be copied to another drive of similar or greater capacity. The bit stream copy of the storage drive is generally acquired using a *dd* based tool [15]. This image is stored in a raw format supported by a *dd* or a propriety format which is typically based on *dd* [16]. The acquired image is either an identical copy of the storage device, for example dd image, or the data from the device which is stored in a format that can be used to evaluate the contents and presented in court as permissible evidence. Analysis can be conducted using any of the many open-source or propriety tools available such as Sleuthkit [17], Forensics ToolKit [1], EnCase [9] etc.

Most of the research conducted in the area of virtualization and forensics, makes use of VMs as forensics tools. VMs can be used to conduct analysis of evidence. VMs provide the examiner the ability to have a clean operating system without having to wipe a drive and install a fresh operating system on it for every new case. Helix [7] & Penguin Sleuth Kit Virtual Computer Forensics and Security Platforms [6] are popular Linux-based operating systems tailored for forensics acquisition and analysis. Both platforms are readily available as virtual appliances that can be used with VMware products. Similarly other virtual appliances are available which use virtualization to assist in conducting a forensic investigation.

Mrdovic et al. used a tool called Live View, which creates a VMware VM from a raw image of a drive or a physical drive [14]. Guo et al. use a similar process of using Live View to boot an image acquired by dd and use that to augment their static forensic methods [10]. This enables the investigator to boot up the disk in a virtual environment and gain an interactive, user-level perspective of the suspect’s environment. All this is done without modifying the image or the physical drive and is considered to be forensically sound.

As incidents related to VMs are on the rise, they have caught the attention of forensics experts. New methods to collect evidence from VMs are needed. Fiterman and Durick in their article titled “Ghost in the Machine: Forensic Evidence Collection in the Virtual Environment” point out that tools and options that enable an examiner to investigate virtual data are currently limited [8]. Similar concerns regarding the absence of forensics tools and procedures for VM analysis are raised and methodologies are proposed by Beek [4]. He also suggests a tool which compares the memory files (.vmem) of snapshots created by VMware products for any new files or processes.

Bares [2] in his research studied the amount of data that could be recovered from VMs in a NTFS partition. His research also showed that lesser amounts of data were recoverable if the VM was incorrectly shut down as compared to when it was shutdown gracefully. In this paper, we propose a solution that is able to acquire and analyze VM hard disk evidence on a running VM, without shutting down, in order to preserve the most evidence with the least number of modifications.

In conclusion, there is abundant research on using VMs and virtual appliances to aid forensic investigation, but research on collecting and analyzing evidence from VMs hard disk is lacking. The proposed forensics methodology provides forensics examiners a forensics sound procedure and tool to acquire and analyze VMs hard disk images using only the existing VM files.

3. FORENSICALLY-SOUND PROCEDURE FOR VIRTUAL DISK ACQUISITION AND ANALYSIS

The use of VMs in corporate and personal environments is rapidly increasing; 18% of servers were virtualized in 2009 and that grew to 25% in 2010. It is expected that by 2012, about half of the servers hosted will be virtualized and hosted virtual desktops will reach 49 million units by 2013 [13]. With the growing number of virtual systems it becomes imperative that a methodology to analyze virtual systems is developed. Many systems carry out critical tasks that cannot be stopped. If such systems are compromised, or are suspected of being compromised, they cannot be taken offline for analysis. In such a scenario it becomes important to conduct a live analysis of the system. However, when a live analysis is carried out, the investigator may change information that resides in memory and any remaining open network connections will be terminated. According to Kurse & Heiser [11], the common practice to conduct forensic analysis of a physical machine is to take the machine offline at some point. The machine’s hard disk is then imaged, and its data acquired for analysis.

When VMs are involved in an incident, the VM is usually suspended or a snapshot of the machine is created to preserve the processes and network status for forensics analysis. There are two paths that a forensics investigator can follow: a) resume the suspended VM and use the normal procedure to acquire and analyze the live machine, and b) analyze the VM files without resuming the suspended machine. However, the method of resuming a suspend VM before acquisition may potentially change the evidence, leading to the possibility that the evidence will NOT be legally admissible.

In this section, the authors describe a forensics solution to acquire, preserve, and analyze snapshots of disk images using VM suspend or snapshot of VM files created by the VMware utility without resuming or shutting down VMs.

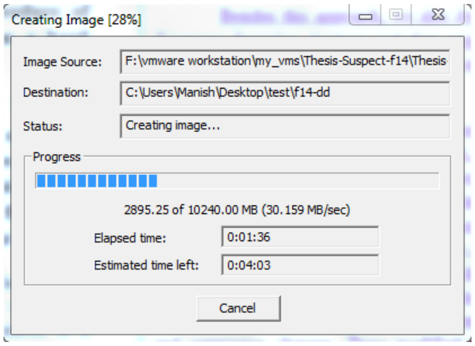
3.1 Virtual Disk Acquisition

The aim of forensic image acquisition is to minimize contamination and ensure legally-admissible evidence. To accomplish this, the digital evidence acquisition process has to follow an appropriate procedure. Acquiring non-volatile data from a physical hard disk entails many steps [11]. A machine is first powered off by disconnecting the power supply from the machine (i.e. pulling the plug). The hard disk is then removed from the suspect machine and connected to a forensic analysis machine. The hard disk is then imaged using any of the many tools available for imaging a disk such as dd, FTK Imager, EnCase, etc. This image is then used by a forensics investigator to conduct an analysis of the events the machine may have experienced.

When working with suspended VMware images, there are two options for acquiring the virtual disks: resuming the suspended system, then use bit-by-bit copy or to directly work with the VMware .vmdk and snapshots files. The problem with resuming a VM is that during the resume process, many files stored on the hard disk are changed, which may destroy evidence. Another disadvantage of resuming the suspended VM is the loss of information stored in the memory as the state of the VM changes. Such information could be vital to the investigation being carried out.

To overcome the shortcomings of resuming the suspended VM, the better solution is to create a snapshot of the VM and then work directly with the VM files that are stored on the host system. Upon taking a snapshot, the state of the hard disk is preserved and any changes to the disk are stored in a separate file. The virtual memory of the VM is stored in a file and any state changes are written to another virtual memory file. Following this procedure, we ensure that the evidence is preserved and can be presented to court as forensically sound and admissible evidence.

Both EnCase and FTK support conversion of .vmdk files to raw (dd) format. When FTK is pointed to a snapshot for converting it to a raw image, it converts the snapshot along with any previous snapshots and the base vmdk files (see Figure 1).



Creating Image [28%]

Image Source: F:\vmware workstation\my_vms\Thesis-Suspect-f14\Thesis

Destination: C:\Users\Manish\Desktop\test\f14-dd

Status: Creating image...

Progress

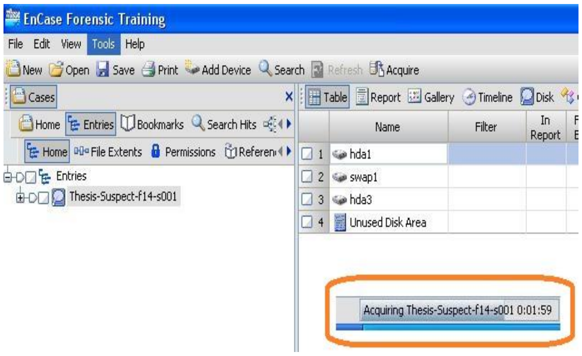
2895.25 of 10240.00 MB (30.159 MB/sec)

Elapsed time: 0:01:36

Estimated time left: 0:04:03

Figure 1: FTK acquiring .vmdk in raw (dd) format.

When EnCase is pointed to the base clean vmdk files, it successfully converts them to raw/dd format, but when EnCase is pointed to a later *snapshot*, it only converts the delta that is created after snapshoting a VM. As a result, to include the previous snapshots and the base images as well, one has to assimilate a flat vmdk file by using utilities that are packaged with VMware Workstation, namely vmware-vdiskmanager.exe. Vmware-vdiskmanager.exe creates one vmdk file that includes: the selected snapshot, any previous snapshots and the base vmdk files. The resulting single vmdk file can then be converted to raw/dd format using EnCase (see Figure 2).



EnCase Forensic Training
File Edit View Tools Help
New Open Save Print Add Device Search Refresh Acquire
Cases
Table Report Gallery Timeline Disk
Home Entries Bookmarks Search Hits
Home File Extents Permissions Referen
Entries
Thesis-Suspect-f14-s001

	Name	Filter	In Report
1	hda1		
2	swap1		
3	hda3		
4	Unused Disk Area		

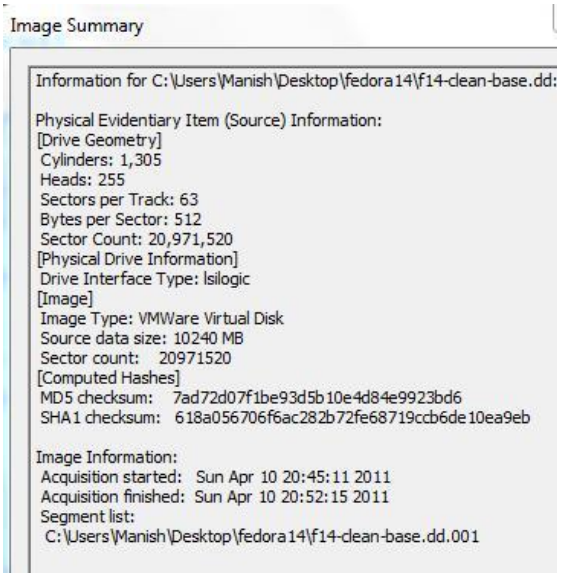
Acquiring Thesis-Suspect-f14-s001 0:01:59

Figure 2: EnCase acquiring .vmdk in raw (dd) format.

3.1.1 Forensically Sound Virtual Disk Acquisition

It is now clear that both FTK and EnCase can acquire the contents of the virtual hard disk without shutting down the VM if a snapshot of this machine is created at the time of incident response. The next question is whether this solution is forensically sound. In other words, do FTK Imager and EnCase change the original VM disk image?

In general, both FTK imager and EnCase require a write blocker device to image a live physical drive. However, since VMware virtual disks are implemented as files, can they be acquired without the use of a write blocker device using FTK and EnCase? In this experiment, the authors studied the functionality of both tools when applied to VM file conversion. In particular, the authors used both tools to create raw images for VM hard disks and calculated hashes of the raw images. We found that both tools produced the matching MD5 and SHA1 hashes, as can be seen from figures 3a & 3b. Therefore, we conclude that VM hard disk files can be safely converted to raw/dd images using either tool without relying on a write block device.



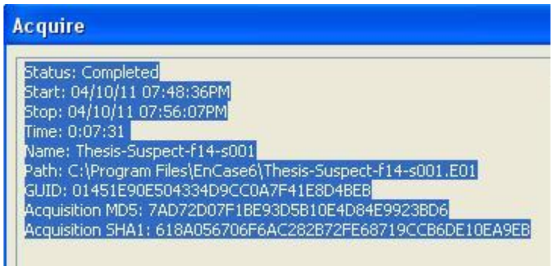
```
Image Summary

Information for C:\Users\Manish\Desktop\fedora14\f14-clean-base.dd:

Physical Evidentiary Item (Source) Information:
[Drive Geometry]
 Cylinders: 1,305
 Heads: 255
 Sectors per Track: 63
 Bytes per Sector: 512
 Sector Count: 20,971,520
[Physical Drive Information]
 Drive Interface Type: lsilogic
[Image]
 Image Type: VMWare Virtual Disk
 Source data size: 10240 MB
 Sector count:    20971520
[Computed Hashes]
 MD5 checksum:    7ad72d07f1be93d5b10e4d84e9923bd6
 SHA1 checksum:   618a056706f6ac282b72fe68719ccb6de10ea9eb

Image Information:
 Acquisition started:   Sun Apr 10 20:45:11 2011
 Acquisition finished:  Sun Apr 10 20:52:15 2011
 Segment list:
  C:\Users\Manish\Desktop\fedora14\f14-clean-base.dd.001
```

Figure 3a: MD5 & SHA1 sums obtained from FTK Imager



```
Acquire

Status: Completed
Start: 04/10/11 07:48:36PM
Stop: 04/10/11 07:56:07PM
Time: 0:07:31
Name: Thesis-Suspect-f14-s001
Path: C:\Program Files\EnCase6\Thesis-Suspect-f14-s001.E01
GUID: 01451E90E504334D9CC0A7F41E8D4BEB
Acquisition MD5: 7AD72D07F1BE93D5B10E4D84E9923BD6
Acquisition SHA1: 618A056706F6AC282B72FE68719CCB6DE10EA9EB
```

Figure 3b: MD5 & SHA1 sums obtained from EnCase

3.1.2 Guest System Time Skew

When the guest system (the VM) is being acquired it is critical for the incident response professional to record the time of the guest operating system as well as that of the host operating system. The time of the guest operating system could be skewed and if this is recorded the forensic examiner can make more definitive and accurate statements about activities that may have taken place. If the guest system clock is synchronized with the host system clock, the incident response professional should make sure to check if the host system time is correct else he should record the skew. If the guest is using an external source besides the host system clock to synchronize the time, any skew that exists should be recorded. The Forensics Snapshot Tool takes into account the time skew of the guest operating system to conduct analysis.

3.2 Forensics Snapshot Analysis Tool

With the acquired image from section 3.1, one can use open-source or commercial forensics tools, for example, EnCase, FTK, and Sleuthkit to conduct a forensics analysis. However, with the additional VM files created from the VM and features supported by VMware, are there other efficient techniques that could assist forensics investigation?

VMware offers the Snapshot feature that allows one to freeze the state of a VM at a given point of time [18]. In this research, the authors developed a *Forensics Snapshot Analysis* tool that compares the snapshot with a pre-staged (recorded at an earlier time) snapshot to identify possible malicious activities.

This tool is written as a Bash script and incorporates existing tools, such as Sleuthkit and md5sum, to verify the integrity of the evidence and also automate parts of the forensic analysis. The script extracts files from both the clean and the compromised image snapshots. Then a comparison is made to determine the changes detected - such as files created, changed or deleted. The tool is also capable of identifying MAC time changes, content changes and permission changes. These modified files are reported and can then be further investigated by a forensics examiner.

To prove that the Forensics Snapshot Analysis tool is a forensically sound, the authors computed the MD5 checksum of the image before and after applying this tool and found that the hash results are identical. This validates that the tool does not modify the evidence files or their contents.

4. EXPERIMENTS AND ANALYSIS

4.1 Experiment Setup

Suspect virtual Machine:

For our experiments, VMware Workstation 7.0.1 was installed on a Windows 7 Home Premium operating system with a New Technology File System (NTFS) partition. A Fedora 14 operating system was installed in VMware Workstation using an extended file system (ext), viz. ext3, with 10GB of disk space. Once the system was installed a snapshot was taken to establish a clean base system. Various packages were installed and changes to various files were made. Changes to file permissions such as execute bit, set user ID, etc. were also made to emulate a real world scenario where a malicious user might change file permissions to gain access to restricted parts of a system. Once these actions were performed, another snapshot of the VM was made. Since FTK Imager is a better solution to convert .vmdk and snapshot files to a completed raw image, we used FTK Imager to convert both the snapshots and create a raw disk images.

Forensics Analysis Machine:

Another Fedora 14 operating system was installed as a VM which was used for forensics analysis with the *Forensics Snapshot Analysis Tool* and its dependencies such as Sleuthkit and md5sum installed.

4.2 Analysis and Results

The experiment follows the processes shown in Figure 4.

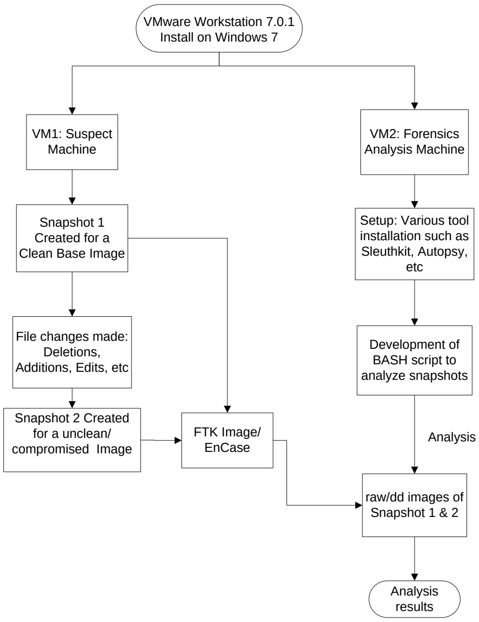
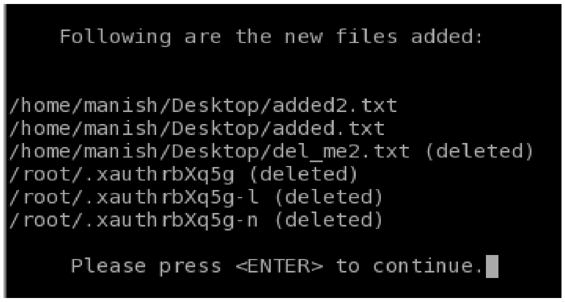

VMware Workstation 7.0.1
Install on Windows 7
VM1: Suspect Machine
VM2: Forensics Analysis Machine
Snapshot 1 Created for a Clean Base Image
Setup: Various tool installation such as Sleuthkit, Autopsy, etc
File changes made: Deletions, Additions, Edits, etc
Development of BASH script to analyze snapshots
Snapshot 2 Created for a unclean/ compromised Image
FTK Image/ EnCase
Analysis
raw/dd images of Snapshot 1 & 2
Analysis results

Figure 4: Flow Chart of the processes

As shown in Figure 5 below, the Forensics Snapshot Analysis tool successfully analyzes and compares snapshots of the same virtual machine taken at different points in time. Using the tool, a forensics examiner can generate a list of files that have been added, deleted, modified and changed by comparing the snapshots. The tool produces formatted reports of each analysis procedure it carries out. The forensics examiner can then decide on areas of further investigation based on the items of interest generated by the Forensics Snapshot Analysis tool.

The Forensics Snapshot Analysis tool is forensically sound and does not modify the raw files in any way. This can be proven by computing the hashes of the raw files after analysis is complete. The MD5 & SHA1 hashed computed for the raw files after the tool has analyzed the raw files match the hashes computed before analysis was run.



```
    Following are the new files added:


/home/manish/Desktop/added2.txt
/home/manish/Desktop/added.txt
/home/manish/Desktop/del_me2.txt (deleted)
/root/.xauthrbXq5g (deleted)
/root/.xauthrbXq5g-l (deleted)
/root/.xauthrbXq5g-n (deleted)

     Please press <ENTER> to continue.
```

Figure 5: Output of *Forensics Snapshot Analysis* tool showing a list of recently deleted files

4.3 Other usage of this tool

This tool can also be used to study OS behavior. One example would be inode reallocation. In some OSes, if a new file is created soon after deleting an existing file, the inode used by the deleted file is marked as "not in use" and could be assigned to a new file. This result can be seen from Figure 5 above when the file del_me.txt is deleted and added2.txt is created. The inode of file del_me2.txt is assigned to added2.txt.

5. LIMITATIONS: POSSIBLE METHODS OF OBFUSCATION

The Forensics Snapshot Analysis tool relies heavily on the MAC time of files to generate files of interest. A possible method for obfuscation could be the use of a tool that does not modify MAC times of files. TrueCrypt is one such tool. Depending on how the preferences are set, when a TrueCrypt volume is modified it can be configured to update only the change time, leaving the modification (mtime) and access (atime) times unchanged. The authors used TrueCrypt with their script to study this behavior of TrueCrypt and added features to the script which now checks for change time (ctime) difference between the files found in both snapshots.

The Forensics Snapshot Analysis tool cannot view the contents of files that have been encrypted. The tool will still list files which have differing modification & change times but the examiner will not be able to view the contents of the file, if it is encrypted. A method to analyze encrypted files and volumes can be developed and incorporated in the tool.

6. CONCLUSIONS AND FUTURE WORK

The authors studied the solutions for acquiring and analyzing live virtual machines based on VM files. A forensically sound procedure to acquire virtual disk images is provided. In addition, the snapshot analysis tool, Forensics Snapshot Analysis, was developed. It is a powerful tool for forensics investigators to analyze VM hard disk images and present pertinent evidence in court. This tool can also be useful in incident response to confirm a breach and to carry out further forensic analysis.

The Forensics Snapshot Analysis tool can also be used for academic and training purposes. Students are taught that booting a suspect machine or VM that has been shutdown or suspended can modify the contents, changing potential evidence. Snapshots of a shutdown or suspended VM can be acquired before and after booting and these snapshots can be analyzed using the tool developed which will practically demonstrate why booting is not a forensically sound procedure.

Several avenues can be pursued as an extension to this research. The effect of encrypted files and volumes on the analysis conducted by this tool can be studied further. Changes can be made to the developed tool to handle other file systems such as FAT, NTFS and other file systems.

7. REFERENCES

[1] Access Data Forensics ToolKit, Retrieved from http://www.accessdata.com/forensictoolkit.html, 2010

[2] Bares, R., “Hiding in a virtual world using unconventionally installed operating systems”, IEEE International Conference on Intelligence and Security Informatics. Dallas, TX, 2009.

[3] Barrett, D., and Kipper, G., “Investigating Dead Virtual Environments, Virtualization and Forensics”, Syngress, Boston, 2010, Pages 83-107.

[4] Beek, C., “Virtual Forensics”. Retrieved from: http://securitybananas.com/wp-content/uploads/2010/04/Virtual-Forensics_BlackHatEurope2010_CB.pdf, 2010.

[5] Brown, C. L. T., “Computer Evidence: Collection & Preservation”, Hingham, MA: Charles River Media, 2005

[6] Ebaca, “Penguin Sleuth Kit Virtual Computer Forensics and Security Platform”. Retrieved from http://www.vmware.com/appliances/directory/249, 2010.

[7] E-fense, Cyber Security & Computer Forensics Software Page. Retrieved from http://www.e-fense.com/helix/, 2010.

[8] Fiterman, E. M., & Durick, J. D., “Ghost in the machine: Forensic evidence collection in the virtual environment”, Digital Forensics Magazine, 2, 2010, pp. 73–77.

[9] Guidance Software, EnCase Forensic. Retrieved from: http://www.guidancesoftware.com/forensic.htm, 2010.

[10] Guo, H., Huang, D., & Zhang, Y. (2012). Implication of Virtualization Technologies in Computer Forensic. Energy Procedia, Volume 13. Pages 4133-4137, ISSN 1876-6102

[11] Kruse W. G., & Heiser, J. G., “Computer Forensics: Incident Response Essentials (1st ed.)”, Addison Wesley Professional, 2002.

[12] Metasploit, Metasploit Anti-Forensics Project. Retrieved from http://www.metasploit.com/research/projects/antiforensics/, 2010.

[13] Messmer, E., “Gartner predicts nearly half of server workloads will be virtualized”, Network World. Retrieved from http://www.networkworld.com/news/2009/102009-gartner-server-virtualization.html, 2010.

[14] Mrdovic S., Huseinovic, A., and Zajko, E., “Combining static and live digital forensic analysis in virtual environment. Information”, Communication and Automation Technologies, 2009. ICAT 2009. XXII International Symposium on, vol., no., pp.1-6, 29-31 Oct.2009.

[15] Nelson B., Phillips A., Enfinger F., and Steuart, C., “Guide to Computer Forensics and Investigations”,

Second Edition. Boston, MA: Thomson Course Technology, 2006

[16] Rude, T., “DD and Computer Forensics”, Retrieved http://www.crazytrain.com/dd.html, 2010.

[17] sleuthkit.org, Sleuthkit. Retrieved from http://www.sleuthkit.org/sleuthkit/, 2010.

[18] VMware, “What Files Make Up a Virtual Machine?” http://www.vmware.com/support/ws55/doc/ws_learning_files_in_a_vm.html, 2010.