Virtual Forensics in Evidence Investigations

吳欣儒　　王旭正*
中央警察大學資管所　　中央警察大學資管系
sjwang@mail.cpu.edu.tw

摘要

使用者若想在硬體資源有限情況下，滿足多種作業系統的需求，虛擬化技術是多數使用者的選擇。雲端時代的來臨，虛擬化的技術更是雲端發展的一大利器。在電腦犯罪中，非法者亦有可能利用虛擬機器軟體建置多個虛擬主機，進行詐騙及電腦犯罪使用。而在犯罪後，虛擬機器內的犯罪證據也可快速的銷毀。對於鑑識人員而言，因虛擬機器的發展，使得電腦鑑識工作比起以往更加複雜。為了解決上述所提及的困境，本文研究如何進行虛擬機器的鑑識工作以及如果解決映像檔遭到破壞的困境。首先虛擬機器的鑑識工作需要先瞭解虛擬機器的特性。當對虛擬機器有一定的瞭解，鑑識人員也可得知該從何處發現證據或線索，以及面對損壞的映像檔，知道如何獲取相關的犯罪跡證。

關鍵詞：虛擬機器，電腦鑑識，映像檔損壞。

1. 前言

在目前科技進步的時代，硬體性能每隔一小段時間就提升許多，許多個人電腦及公司行號的主機都具有極佳的硬體配備。對於企業而言，為了充分利用大部分使用時的閒置資源或者是減少硬體方面的花費，虛擬化技術被廣泛的應用於建置多個作業系統提供服務。對於使用者而言，透過虛擬化技術所設計的虛擬機器，就可滿足使用多個作業系統的需求。在現今的電腦犯罪當中，虛擬機器的使用讓非法者可以節省硬體方面的開銷，只要由少數幾台主機便可達到數十台主機的效果。在進行犯罪後，由於數位證據的特性，數位跡證可以快速的被銷毀。對於偵查人員而言，由於非法者可能向業者租用主機，使得偵查人員不易追蹤非法者外。在採證方面亦須業者配合，造成取證的困難性。

為了解決上述所提及的困境，本文研究如何進行虛擬機器的鑑識工作以及如果解決映像檔遭到破壞的困境。首先虛擬機器的鑑識工作需要先瞭解虛擬機器的特性以及虛擬機器所屬檔案的檔案特性，例如虛擬化技術的兩種架構以及虛擬機器中不同檔案扮演著硬碟以及記憶體的角色。當對虛擬機器有一定的瞭解後，鑑識人員也可得知該從何處發現證據或線索，以及面對損壞的映像檔，該如何獲取相關的犯罪跡證。

本文將在第二節介紹虛擬機器的相關背景。第三節為我們對使用虛擬機器的主機資料取證和復原損壞映像檔的研究。第四節會討論以及分析有關虛擬機器的犯罪以及鑑識現況及方法。第五節做本文結論。

2. 背景知識

2.1 虛擬機器

虛擬機器是虛擬化技術的其中一種軟體，它可以在主機上或是終端伺服器與終端使用者之間建立一種環境，創造出一台虛擬的硬體機器。簡而言之，可以將虛擬機器看成一種模擬器。虛擬機器在軟體和實體硬體之間，軟體可以透過虛擬機器和實體硬體構通，所以虛擬機器也可以改善實體機器相容性的限制。提供更好的可攜性及適應性。因此對於硬體效能較好的主機而言，可以模擬一個或多個作業系統來進行作業，減少硬體成本。目前市面上有許多不同的虛擬機器軟體能選用，例如VMware、Microsoft、Citrix…等。根據 iThome 2011 年調查的數據虛擬機器軟體的平台採用，以 VMware 市占率最高[6]，如圖 1 所示。且 VMware 也是最早發展虛擬機器平台的公司，因此本文以 VMware Workstation 為例進行虛擬機器的介紹。

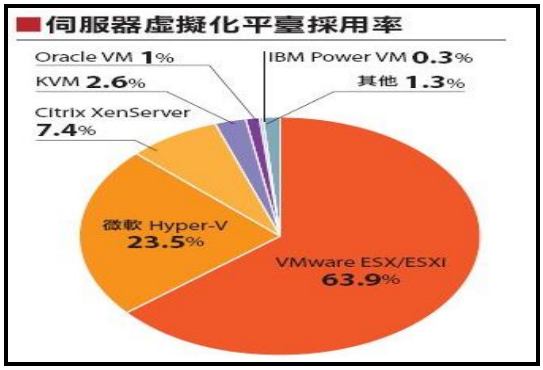

圖 1. 虛擬化平台採用比例

2.1.1 虛擬化技術

我們在瞭解虛擬化技術之前，必須先瞭解何謂 Host OS 以及 Guest OS。所謂的 Host OS 就是安裝虛擬機器時運行虛擬機器軟體的作業系統。Guest OS 則是在虛擬機器軟體上安裝的作業系統。例如在原本的 Windows XP 系統下安裝了虛擬機器軟體，並且在虛擬機器軟體內安裝了 Linux 作業系統，此時 Windows XP 即為 Host OS，Linux 則為 Guest OS。但目前的虛擬化技術並不一定要具備 Host OS 方能架設虛擬機器，因此虛擬機器的虛擬化技術主要分為寄宿架構與裸金屬架構[10]：

1. 寄宿架構(Hosted architecture)

所謂的寄宿架構也稱為初期架構，使用虛擬化技術模擬虛擬的硬體和軟體，並將模擬出的軟、硬體架構於 Host OS 之上，形成一個在作業系統中存在另一個作業系統的架構。如圖 2 左圖所示，虛擬機器軟體讓 Host OS 認為模擬出的硬體是一個應用程式，而虛擬機器層透過 Host OS 與實體硬體溝通，進而使用硬體的資源。此種架構最大優點是硬體相容性高，只需要具備 Host OS，就可以使用大多數的作業系統。但此種架構的缺點是效能低，各虛擬機器層沒有獨立的硬體資源，一旦 Host OS 被攻擊的話也會導致所有虛擬機器無法運作。VMware Workstation 及 Microsoft Virtual PC 即是採用此架構。

2. 裸金屬架構(Bare-metal architecture)

裸金屬架構比起寄宿架構為更進階的一種技術，此種架構不再需要透過 Host OS 與硬體對話，虛擬機器層直接運行在硬體上。如圖 2 右圖所示，虛擬機器層直接接管所有硬體資源，每一個虛擬機器都直接使用硬體資源。在這樣的架構下，因為直接使用硬體資源，效能也會提升。當任何一個 Guest OS 遭受攻擊或損壞皆不會影響到其他的 Guest OS。不過此架構的缺點是硬體相容性低，對於作業系統有一定的限制。Mware ESX / ESXi、Microsoft Hyper-V 以及 Citrix Xen Server 即是採用此種架構。

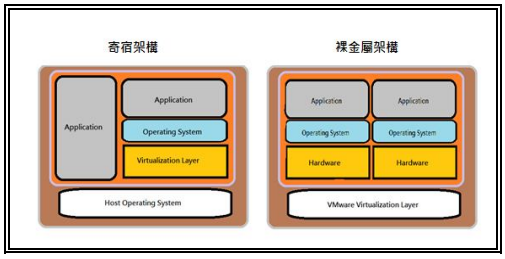

圖 2. 虛擬機器目前採行的兩種架構，寄宿架構及裸金屬架構

2.2 VMware Workstation 的檔案類型

運行 VMware Workstation 後，會在電腦中留下相關操作、設定之檔案[9]。以下將介紹主要的幾種檔案。

1. Vmx 檔

虛擬機器的主要配置檔，記錄使用者對於虛擬機器的設定值，例如作業系統版本、系統資訊等。Vmx 檔是一個可以容易被編輯的文字檔。

2. Log 檔

Log 檔記錄著 VMware Workstation 的主要活動，當虛擬機器運作上發生問題時，可以查看先前的 Log 檔發生什麼問題。而 Log 檔也存放一些 Vmx 檔中沒有提及的訊息。

3. Vmsd 檔、Vmsn 檔及 Vmss 檔

Vmsd 檔、Vmsn 檔及 Vmss 檔則為存儲快照資訊和資料的檔案。「快照」會將拍照時間點的資料紀錄下來，若以後系統有所變動的時候，可以當成還原的依據。當使用者對虛擬機器進行快照時，會將當時虛擬機器運行狀態記錄在 Vmsn 檔。因此 Vmsd 與 Vmsn 兩者在虛擬機器中扮演著復原點的角色。而 Vmss 檔則記錄虛擬機器暫停時的狀態。

4. Vmdk 檔

Vmdk 檔在虛擬機器中的角色如同電腦中的硬碟。一個虛擬硬碟主要是由多個 Vmdk 檔組成，因此 Vmdk 檔內容都存放虛擬機器的資料。

5. Vmem 檔

相較於 Vmdk 為扮演虛擬硬碟的角色，Vmem 則扮演虛擬機器中的記憶體角色。它記錄虛擬機器存放於記憶體的資料，並且只有當虛擬機器運行或是暫停狀態才會存在，同時它也擔任一個備份資料的角色，不過當虛擬機器被終止運行時，就如同電腦被關機後揮發性記憶體資料會消失，此檔案也會被虛擬機器自動刪除。以上虛擬機器的各種檔案，整理如表 1 所示。

表 1.虛擬機器的各種檔案

副檔名	檔案特色	檔案名稱
.LOG	紀錄 VM 的主要活動	<vmname>.log/vmware.log
.VMDK	擔任 VM 的硬碟	<vmname>.vmdk
	紀錄虛擬硬碟的變化	<diskname>-<###>.vmdk
.VMSD	紀錄快照的資訊及 metadata	<vmname>.vmsd
.VMSN	紀錄進行快照時 VM 的狀態	<vmname>-snapshot.vmsn
	紀錄 snapshot 的狀態	<vmname>-snapshot<###>.vmsn
.VMSS	紀錄 VM 暫停時的狀態	<vmname>.vmss
.VMX	紀錄 VM 的設定值	<vmname>.vmx
.VMEM	擔任 VM 的記憶體	<uuid>.vmem/<vmname>.vmem

3. Our Scheme in Virtual Forensics

虛擬機器與實體系統的運作方是沒有太大的差異，因此若是在一個正在運行的虛擬機器而言，對於虛擬機器的硬碟資料與記憶體資料可以如同普通的電腦鑑識進行採集。但是在虛擬化環境的調查中，為了取得完整的資料必須透過管理者的權限執行虛擬機器[7]。

在電腦鑑識中，許多被刪除的資料都可以被還原。一般的檔案被刪除時大部分都會被丟進資源回收筒中，而虛擬機器相關的檔案在被刪除時，由於檔案容量過大，會直接被系統刪除。雖然被系統直接刪除的檔案還是可以還原，但是還原的檔案並不如原本的檔案完整。對於虛擬機器而言，不完整的檔案可能會導致無法啟動，鑑識人員也無法取得虛擬機器中的資料。

由於不完整的映像檔可能導致虛擬機器無法啟動，鑑識人員在進行虛擬機器的數位鑑識時，為了避免虛擬機器的資料遭受損害，應盡量採用「現場蒐證鑑識法」。也就是在主機仍在運行的情況下，進行資料的取證以及分析。如此一來除了可取得揮發性記憶體中的資料外，也避免因為關機或系統重新啟動後失去了一些重要的資料。

3.1 複製虛擬機器資料

傳統的電腦鑑識會將主機的整顆硬碟資料複製成一個映像檔，所有相關虛擬機器的檔案也會被複製到映像檔中。在第二節時有提到虛擬機器有許多檔案類型，檔案如圖 3 所示。鑑識人員可能可以從這些檔案中找到一些線索，這些檔案也可以進行虛擬機器的重新啟動。若要啟動虛擬機器，就必須要有完整的檔案，因此若複製的資料不完整的話，有可能無法啟動虛擬機器，因而無法從中獲取資料。除此之外，虛擬機器可能與主機會有共享資料夾。若是單純複製虛擬機器的檔案，而沒有啟動虛擬機器，便無法取得相關資訊。

圖 3.VMware 檔案及其檔案類型

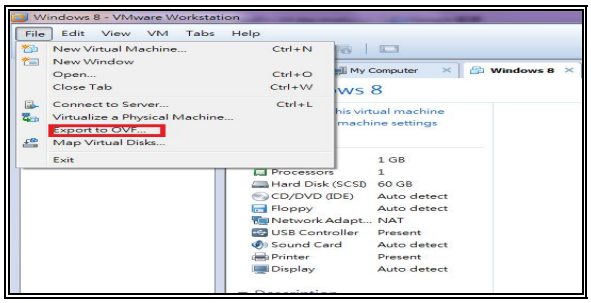

圖 4.將虛擬機器輸出為映像檔

現場蒐證鑑識法可以在主機尚在運行的狀態下採取資料，因此此時除了可以複製硬碟資料外，還可以採取揮發性記憶體的資料。當主機正在運行時，虛擬機器軟體也可能正在運行，因此除了可以取得虛擬機器的記憶體資料也可以將虛擬機器輸出成一個映像檔。將虛擬機器輸出成映像檔可以利用虛擬機器本身相關的程式，如圖 4 所示。

利用映像檔啟動虛擬機器後，可以使用鑑識軟體建置鑑識環境，其他步驟與傳統數位鑑識無太大差別。除了使用虛擬機器本身相關程式外，也可以利用 FTK Imager[4]等鑑識軟體將虛擬機器輸出為映像檔。安裝映像檔方面除了虛擬機器軟體外也可以利用 Mount Image Pro[5]進行映像檔的安裝。FTK Imager 與 Mount Image Pro 如圖 5 及圖 6 所示。

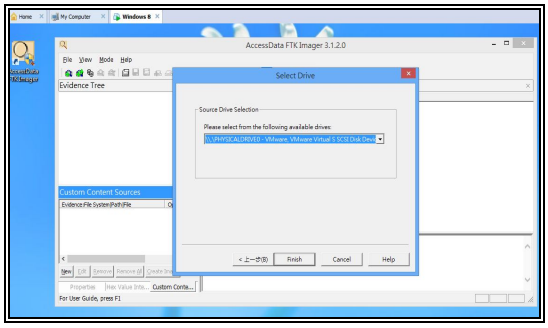

圖 5.FTK Imager 於虛擬機器中的操作畫面

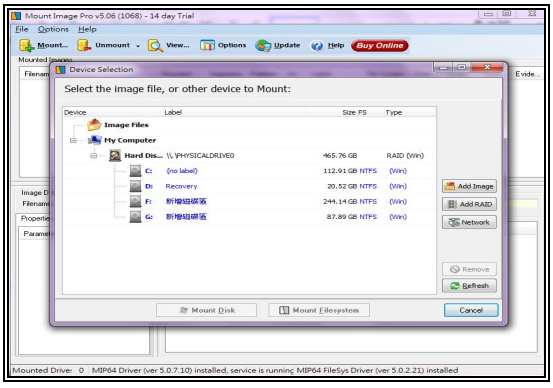

圖 6. Mount Image Pro 操作畫面

3.2 虛擬機器映像檔損壞的鑑識方法

虛擬機器的映像檔可能會在採集的過程中或是人為的操作下損壞。例如映像檔的 VMDK 檔案損壞，導致資料表頭不完整，無法啟動虛擬機器。為了找尋虛擬機器中的證據，鑑識人員必須復原損壞的映像檔。若映像檔的檔案格式是以 SPARSE 資料型別，在一定的損壞程度下，可以透過下列步驟進行復原，SPARCE 資料結構如圖 7 所示[8]。

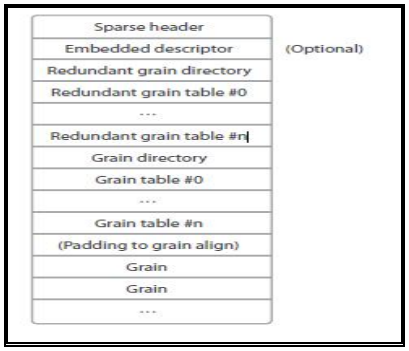

圖 7.SPARSE 文件結構

3.2.1 SPARSE 資料表頭的復原

在映像檔的還原中，首先要復原的是資料表頭(Header)的部分，表頭的內容如圖 8[11]所示。我們可以根據 Vmx 檔以及 Log 檔的內容進行表頭的復原。因為表頭中主要都存放一些虛擬機器的設定值，前面有提到 Vmx 檔主要記錄的就是虛擬軟體的設定值，而 Log 檔則記錄一些主要的活動。因此若是設定值有所變更，也可以從 Log 檔中發現。若要復原表頭必須復原以下三個欄位:

1. Extent 的大小(capacity)，
2. 存放 Metadata 的 sector 數量(numGTEsperGT)
3. 指出 Grain 資料夾的扇區位置。(gdOffset)
4. 以上三個欄位的數值可以從 LOG 檔中找到，而其他欄位的數值只要填入預設值即可，LOG 檔中 capacity 的數值如圖 9 所示。

```
typedef uint64 SectorType;
typedef uint8 Bool;
typedef struct SparseExtentHeader {
   uint32        magicNumber;
   uint32        version;
   uint32        flags;
   SectorType    capacity;
   SectorType    grainSize;
   SectorType    descriptorOffset;
   SectorType    descriptorSize;
   uint32        numGTEsPerGT;
   SectorType    rgdOffset;
   SectorType    gdOffset;
   SectorType    overHead;
   Bool          uncleanShutdown;
   char          singleEndLineChar;
   char          nonEndLineChar;
   char          doubleEndLineChar1;
   char          doubleEndLineChar2;
   uint16        compressAlgorithm;
   uint8         pad[433];
} SparseExtentHeader;
```

圖 8.SPARSE Extent Header 內容

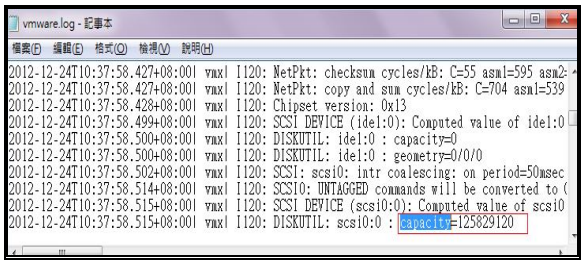

圖 9.LOG 檔中的 Capacity 數值

3.2.2 SPARSE 資料描述的復原

完成表頭復原工作後則檢查 SPARCE 文件結構內的資料描述(descriptor)是否有所損壞，若有損壞則對以下欄位進行復原動作，如下所述：

1. CID 以及 Parent CID：CID 為虛擬機器第一次啟動時隨機產生的 32 位元的數值，每次重啟都會改變。此數值可以從 LOG 檔中的 longContentID 下找到數值，如圖 10 所示。ParentID 此欄位值可以從虛擬機器的快照中找到，也就是本文前段所提到的 Vmsd 檔、Vmsn 檔及 Vmss 檔，若是沒有虛擬機器的快照，或是無法找尋到此欄位的內容，可以將他設為此欄位本身的預設值"ffffffff"，在圖 10 中也可看到 parentCID 的數值即為預設值。
2. Extent 的大小：此欄位值跟表頭中第一個要復原的欄位是相同的值。
3. Extent 的格式：格式可以分為 sparse 以及 flat，此欄位可以由 Vmx 以及 Log 檔的檔名進行辨識。
4. Extense 的儲存檔名，此欄位亦可由 Vmx 檔以及 Log 檔中發現。

進行完這兩個復原動作後，映像檔大致上已經復原，可以進行分析的動作。上述的欄位都可在正常的 description 看到，如圖 11 所示。

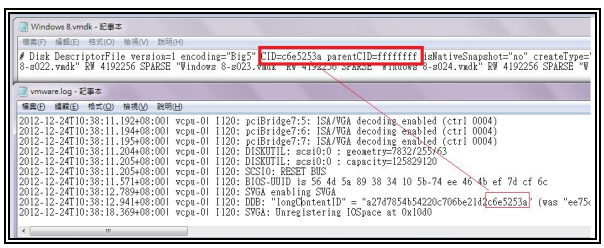

圖 10.descriptor file 中的 CID

```
Disk DescriptorFile version=1 encoding="Big5" CID=c6e5253a parentCID=ffffffff
sNativeSnapshot="no" createType="twoGbMaxExtentSparse"
Extent description RW 4192256 SPARSE "Windows 8-s001.vmdk" RW 4192256 SPARSE "Win
```

Extent的大小　格式　檔案名稱

圖 11.虛擬機器 description

3.2.3 無法復原以及其它鑑識手法

若採用上述的方法，映像檔仍然尚未恢復，鑑識人員還是必須對於映像檔進行 metadata 的採取以及分析。此外若虛擬機器是建置於 Windows 作業系統環境下，可以從註冊檔中找到使用者的帳號以及線索。而除了對於映像檔做分析外，鑑識人員也必須對於 Vmem 檔進行分析。因為虛擬機器的運作方式與時體硬體並無太大差別，因此鑑識人員還是必須針對虛擬機器中記憶體進行採證[1]。在分析 Vmem 檔時可以利用「Compare VMware snapshots 工具」或是「Memparser 工具[2]」，從記憶體中取得有用的資料，Memparser 工具如圖 12 所示。

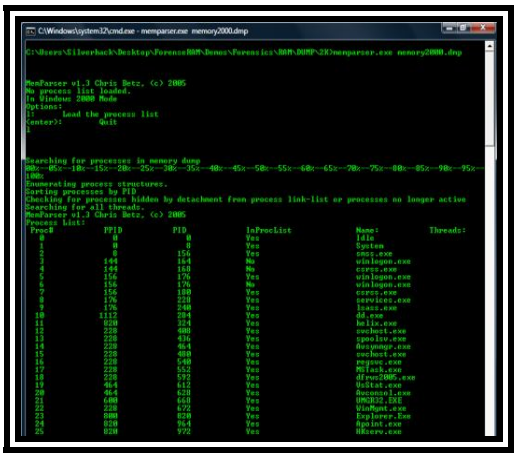

圖 12.Memparser 工具

4.討論與分析

虛擬機器可以在單一主機中創造出多個虛擬的主機，且當使用者透過虛擬機器進行非法行為時的記錄只會存在於虛擬機器中。在傳統電腦鑑識中，採取硬碟中資料，有可能忽略使用者在虛擬機器中的行為記錄或者揮發性資料。為了獲得更多證據，鑑識人員若發現使用者有使用虛擬機器的情形，需要針對虛擬機器進行數位鑑識。

虛擬機器的數位鑑識可以利用鑑識軟體蒐集相關檔案或是虛擬機器應用軟體將虛擬機器輸出成映像檔。利用映像檔啟動虛擬機器後，可在虛擬機器中安裝鑑識軟體，便可如傳統電腦鑑識一般取證。當映像檔損壞時，可以利用本文提出還原映像檔的方法進行還原，進一步從虛擬機器中採取到更多證據。

除了啟動虛擬機器以及還原映像檔外，在本文中探討到一些檔案的特性，如 Vmem 檔為虛擬機器的記憶體。單純針對檔案進行分析的情況下，可能可以從 Vmem 檔中找尋到一些使用者名稱或密碼等資訊。但要注意的是 Vmem 檔只能在虛擬機器尚在運行或暫停的狀態下才存在。另外如 Vmsn 和 Vmsd 扮演虛擬機器的還原點角色，而 Vmss 儲存虛擬機器暫停時的狀態。除了針對映像檔的掛載進行鑑識動作以及損壞映像檔的還原外，在虛擬機器的數位鑑識上也必須針對這些檔案進行調查及分析[3]。

5.結論

使用者在虛擬機器上的行為雖然會被虛擬機器記錄下來，而且這些行為可能會成為犯罪偵查中的重要線索。為了從虛擬機器中找到相關的線索或者犯罪證據，鑑識人員必須對虛擬機器的背景知識有相當的瞭解。例如虛擬機器的檔案類型以及虛擬化技術的架構等。在實務上，損壞的虛擬機器映像檔無法在啟動虛擬機器，也因此無法從虛擬機器中進行採證的動作。因此本文除了介紹如何進行虛擬機器的數位鑑識外，也提及復原映像檔的一種方法。除此之外，也介紹了一些工具可以在進行虛擬機器的數位鑑識上派上用場。

參考文獻

[1] Beek, C. (2010). Virtual forensics, in: BlackHat Europe 2010.

[2] Betz, C. Memparser. http://www.dfrws.org/2005/challenge/memparser.shtml

[3] Dorn, G., Marberry, C., Conrad, S. and Craiger, P. (2009). Analyzing the impact of a virtual machine on a host machine, International Federation for Information Processing, Advances in Digital Forensics V, IFIP AICT 306, pp. 69–81.

[4] FTK Imager is the registered trademark of Accessdata. http://www.accessdata.com/

[5] GetData, Mount Image Pro V4. http://www.mountimage.com/

[6] *iThome* 2012 年 CIO 大調查－IT 應用篇, Retrieved December 30, 2012 from http://www.ithome.com.tw/itadm/article.php?c=71808&s=5

[7] Kwon, T., Bang, J., Lim, K.S., and Lee, S. (2009). Study on digital forensics in virtualization environment, Korean

Institute of Information Technology, Journal of Korean Institute of Information Technology 7 (2) (2009) 159–167.
[8] Lim, S., Yoo, B., Park, J., Byun, K., and Lee, S. (2012) A research on the investigation method of digital forensics for a VMware Workstation's virtual machine, Mathematical and Computer Modeling 55 (2012) 151–160.
[9] Shavers, B. (2008) A discussion of virtual machines related to forensics analysis.
[10] Vitualization basics, Retrieved December 30, 2012 from http://www.vmware.com/virtualization/what-is-virtualization.html
[11] Virtual Disk Format 5.0, Retrieved December 30, 2012 from http://www.vmware.com/support/developer/vddk/vmdk_50_technote.pdf