情報保護學會論文誌
第 21 卷 第 2 號, 2011. 4

VMware Workstation 가상 머신 이미지에 대한 디지털 포렌식 조사 절차 및 손상된 이미지 복구 방안*

임 성 수,† 유 병 영, 박 정 흠, 변 근 덕, 이 상 진‡
고려대학교 정보보호연구원

A study on an investigation procedure of digital forensics for VMware Workstation's virtual machine and a method for a corrupted image recovery*

Sungsu Lim,† Byeongyeong Yoo, Jungheum Park, KeunDuck Byun, Sangjin Lee‡
Center for Information Security Technologies, Korea University

요 약

가상화는 논리적인 가상환경으로 하드웨어의 물리적인 한계를 극복하기 위한 기술이다. 최근 비용 절감 및 그린 IT 정책의 일환으로 가상화 환경을 도입하는 기업들이 증가하는 추세이다. 특히 데스크톱 가상화는 한 대의 물리적 컴퓨터에서 다양한 운영체제를 효율적으로 사용할 수 있기 때문에 가장 활발하게 사용되는 기술유형 중 하나이다. 가상화 기술의 핵심 요소인 가상 머신 이미지는 하드 디스크 이미지와 구조적으로 다르기 때문에 조사에 어려움이 있다. 따라서 가상 머신에 대한 기술적 이해를 바탕으로 가상 머신 이미지에 적합한 조사 절차 및 방안에 대한 연구가 필요하다. 본 연구는 가장 많은 사용자를 가지고 있는 VMware Workstation 가상 머신 이미지에 대한 디지털 증거 조사 절차와 손상된 이미지에 대한 조사 방안을 제안한다.

ABSTRACT

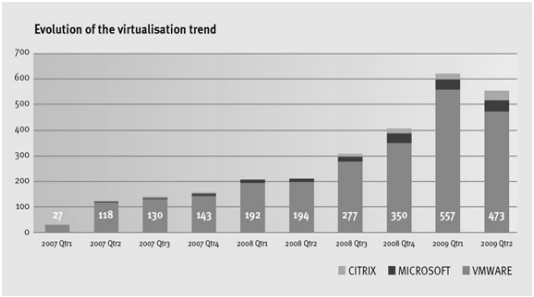
Virtualization is a technology that uses a logical environment to overcome physical limitations in hardware. As a part of cost savings and green IT policies, there is a tendency in which recent businesses increase the adoption of such virtualization. In particular, regarding the virtualization in desktop, it is one of the most widely used technology at the present time. Because it is able to efficiently use various types of operating systems in a physical computer. A virtual machine image that is a key component of virtualization is difficult to investigate. because the structure of virtual machine image is different from hard disk image. Therefore, we need researches about appropriate investigation procedure and method based on technical understanding of a virtual machine. In this research, we suggest a procedure of investigation on a virtual machine image and a method for a corrupted image of the VMware Workstation that has the largest number of users.

Keywords: Digital Forensics, Virtualization, VMware, Virtual Machine

I. 서 론

가상화는 IT 투자의 효율성을 높이기 위한 주요 기술로써 서버, 스토리지, 네트워크, 소프트웨어 영역에 이르기까지 전 세계적으로 활용이 증가하고 있다. 가상화란 물리적으로 다른 시스템을 논리적으로 통합하거나 하나의 시스템을 논리적으로 분할해 자원을 효율

접수일(2010년 7월 30일), 게재확정일(2010년 10월 22일)
* 본 연구는 한국연구재단을 통해 교육과학기술부의 바이오 연구개발사업으로부터 지원받아 수행되었습니다. (20100020634)
† 주저자, nemography@korea.ac.kr
‡ 교신저자, sangjin@korea.ac.kr



〔그림 1〕 VMware 가상화 솔루션 시장 점유율[15]

적으로 사용하게 하는 기술이라고 정의할 수 있다[6]. 즉 기존의 단일 하드웨어를 활용하여 다수의 운영체제와 응용프로그램 등을 서비스할 수 있으며 하드웨어의 한정된 리소스를 효율적으로 공유할 수 있는 것이다.

가상화 기술 유형 중 하나인 데스크톱 가상화는 호스트의 자원을 복수의 논리적 시스템 자원으로 할당받아 사용하는 것으로 가상 머신이라는 중간계층을 통하여 호스트 운영체제 위에서 게스트 운영체제를 동작시킬 수 있다[6]. 가상화 기술을 주도하는 대표적인 벤더는 VMware社로 x86서버 플랫폼의 가상화 제품을 시장에 가장 먼저 출시했으며 〔그림 1〕과 같이 높은 시장 점유율을 유지하고 있다.

VMware Workstation 가상 머신은 사용하는 용량에 따라 가상 머신 이미지의 크기가 동적으로 증가한다. 이는 가상 머신을 생성할 때 설정에 따른 고유의 파일 구조 때문이다. 이러한 파일 구조의 특성이 하드 디스크 이미지와 다르기 때문에 조사를 어렵게 한다. 특히 손상된 가상 머신 이미지의 경우 기존의 어떤 도구로도 조사 가능한 상태로 복구할 수 없기 때문에 이에 대한 대응이 시급하다. 하지만 디지털 포렌식 조사 시 가상화 환경을 접할 수 있는 기회가 많지 않았으며 조사 방법과 복구에 대한 구체적인 연구가 부족한 현실이다.

본 논문에서는 가상 머신 이미지에 대한 디지털 포렌식 조사를 위해 수집한 VMware Work- station 가상 머신 이미지에 대한 조사 절차 및 손상된 이미지에 대한 조사 방안을 제안한다.

II. 가상화 기술

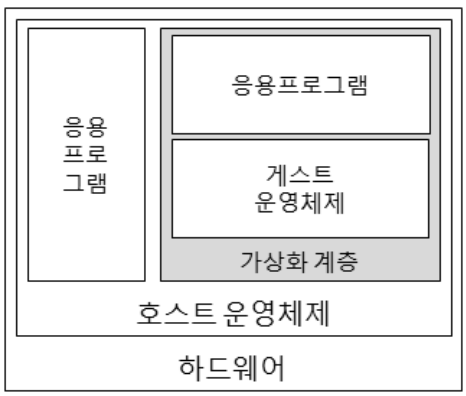
기업 및 개인 사용자들의 가상화 솔루션 사용이 증가함에 따라 가상화 환경에 대한 조사 필요성이 높아지고 있다. 가상화된 환경에 대한 디지털 포렌식 조사를 하기 위해서는 가상화 환경 구성 방법과 데이터 저장위치, 주요 파일의 종류에 대한 지식이 필요하다.

2.1 VMware Workstation의 가상화 기술

VMware社는 데스크톱에서 데이터센터에 이르는 모든 컴퓨팅 시스템, 스토리지, 네트워크 등을 가상화하는 다양한 제품을 가지고 있으며, 베어 메탈 하이퍼바이저형(Bare-metal Hypervisor)과 호스트형(Hosted) 가상화 방식을 사용한다[14]. 본 논문에서는 호스트형 가상화 방식을 사용하는 VMware Workstation을 대상으로 한다.

2.1.1 호스트형 가상화 (Hosted Architecture)

호스트형 가상화는 〔그림 2〕과 같이 물리적 하드웨어 상에서 호스트 운영체제가 존재하며 응용프로그램과 같이 호스트 운영체제 안에서 3계층 수준의 하이퍼바이저를 통해 게스트 운영체제를 동작시킨다. 물리적인 하드웨어를 에뮬레이트(emulate)하기 때문에 오버헤드가 크지만 사용할 수 있는 게스트 운영체제에 제한이 없기 때문에 x86 데스크톱에서 가장 많이 사용되는 방식이다[14].



〔그림 2〕 호스트형 가상화 구조

2.2 VMware Workstation의 가상 머신 파일 구성 및 가상 머신 이미지의 구조

가상화 환경은 실제 환경과의 구조적 차이를 가지고 있기 때문에 물리적 호스트 시스템 자체를 대상으로 압수, 수색이 이루어지게 될 경우, 다른 사용자의 호스트 자원에 대한 가용성과 프라이버시를 침해할 수 있다[5]. 따라서 가상화 환경에 대한 조사를 수행할

경우 관리자 권한을 통하여 특정 가상 머신의 이미지 및 관련된 파일만을 획득하는 최소한의 증거 수집이 이루어져야 한다. 하지만 가상화 환경에 대한 이해와 가상 머신 이미지의 구조에 대한 지식이 부족하다면 중요한 파일을 수집하지 못하거나 이미지로부터 증거로써 의미 있는 데이터를 획득하지 못할 것이다. 이러한 이유로 포렌식 조사관은 가상 머신을 구성하는 파일에 대한 지식이 있어야 한다.

2.2.1 VMware Workstation의 가상 머신 파일 구성

〔표 1〕의 파일 목록에서 가상 머신을 생성할 때 기본적으로 구성되는 파일은 .vmx, .vmdk, .nvram, .vmsd, .log 이며, 포렌식 관점에서 중요한 파일은 .vmx, .vmdk, .vmem, .log 파일이다.

VMX 파일은 텍스트 형식의 데이터이며 가상 머신을 생성할 때 선택한 설정 정보를 저장한다. 설정 정보에는 가상 디스크의 연결 방식, 인코딩 방식, 설정된 전체 가상 디스크 크기 정보, 연결된 가상 장치 정보 등이 포함된다. 이와 같은 대부분의 설정 정보를 가지고 있기 때문에 가상 머신에 대한 전반적인 정보를 확인할 수 있다.

VMDK 파일은 가상 머신의 가상 디스크 역할을 하는 바이너리 파일로 VMware Workstation의 가상 머신 이미지를 구성한다. 가상 머신의 설정 크기만큼 VMDK 파일 크기를 할당하거나 동적으로 증가시킬 수 있으며, 2GB 단위의 복수 VMDK 파일로 분할하여 관리할 수 있다. VMDK 파일은 가상 머신에서 접근할 수 있는 파티션에 대한 정보를 저장하고 있으며, MBR(Master Boot Record)과 기타 파일시스템의 구성요소가 존재한다. 포렌식 관점에서 의미 있는 데이터를 가장 많이 포함하고 있다.

〔표 1〕 VMware Workstation 가상 머신의 구성 목록

파일 확장자	파일 요약
.VMX	가상 머신의 설정 정보 저장
.VMTM	가상 머신의 지정된 그룹 데이터를 포함하는 설정 정보
.VMXF	지정된 그룹이 삭제되었을 때 남는 설정 정보
.VMDK	가상 디스크를 구성하는 파일
.LOG	가상 머신의 로그 정보
.NVRAM	가상 머신의 BIOS 상태 정보
.VMSS	Suspended 상태의 가상 머신 정보
.VMSN	가상 머신의 Snapshot 정보
.VMSD	가상 머신의 Snapshot 메타 데이터
.VMEM	가상 머신의 paging 파일 또는 전체 메모리 저장

VMEM 파일은 가상 머신의 메모리 데이터가 저장되는 바이너리 파일로써 가상 머신이 동작중이거나 서스펜드(Suspended) 상태, 스냅샷(Snapshot)을 생성한 경우에 존재한다. VMEM 파일은 동작 중인 가상 머신의 메모리를 백업하며 가상 머신이 종료되면 자동적으로 삭제된다. 또한 가상 머신이 서스펜드 상태로 종료될 경우에는 삭제되지 않은 상태로 남는다. 두 상태에 따라 남겨지는 VMEM 파일의 이름이 다르게 설정되는데, 가상 머신이 동작중인 상태에서 VMEM 파일의 이름은 '〈uuid〉.vmem' 형식을 사용하며 서스펜드 상태일 때는 '〈vmname〉.vmem' 형식, 스냅샷을 생성한 경우에는 '〈vmname-snapshot〉.vmem' 형식을 사용한다.

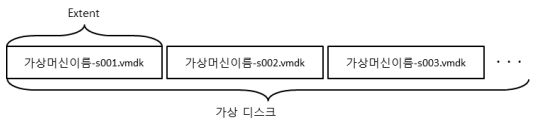
LOG 파일은 가상 머신을 동작시키는 VMware Workstation의 활동 정보를 텍스트 형식으로 저장한다. 가상 머신의 동작에 문제가 발생했을 경우 LOG 파일을 통해서 문제의 원인을 확인할 수 있으며, VMX 파일에 저장되지 않는 호스트 시스템의 이름, 네트워크 IP 주소 등의 설정 정보도 저장되어 있다. 따라서 가상 머신에 대한 정보를 확인하는 과정에서 VMX 파일과 함께 확인해야 한다.

이외에도 가상 머신의 복구지점 역할을 하는 VMSN, VMSD 파일과 서스펜드된 가상 머신의 상태를 저장하는 VMSS 파일은 모두 가상 머신의 스냅샷 정보를 포함하고 있기 때문에 추가적인 조사가 필요하다.

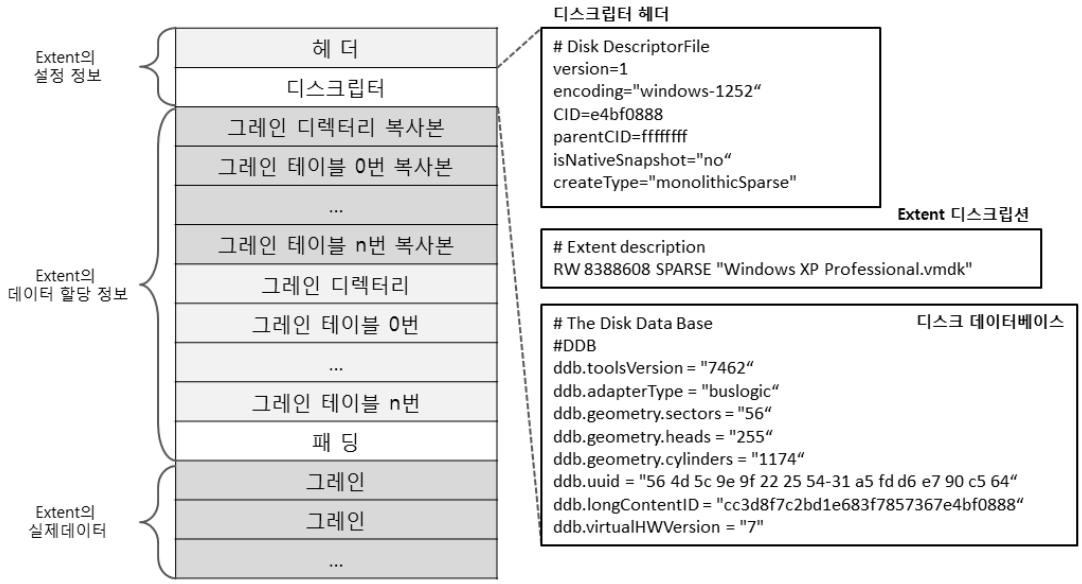
2.2.2 VMware Workstation의 가상 머신 이미지 구성

〔그림 3〕와 같이 가상 머신은 하나 이상의 Extent로 이루어진 가상 디스크를 사용한다. Extent란 가상 머신이 사용하는 물리적인 저장 공간인 VMDK 파일을 의미하는 논리적인 요소이다.

가상 머신은 생성할 때 설정에 의해 Extent를 전체 크기만큼 한번에 할당하는 방식(FLAT)이나 필요



〔그림 3〕 가상 머신 이미지 구조



〔그림 4〕 SPARSE Extent 구조

한 크기만큼 할당한 후 점차 증가하는 방식(SPARSE)으로 데이터를 저장한다. 또한 가상 디스크를 하나의 Extent로 구성(monolithic)할 것인지 여러 개의 Extent로 분할하여 구성(twoGbMaxExtent)할 것인지를 설정할 수 있다. 이와 같은 가상 머신의 설정 정보는 텍스트 형식의 디스크립터(descriptor) 파일에 저장하며, 디스크립터 파일은 Extent내에 삽입되거나 독립된 파일로 존재한다.

가상 디스크의 크기 전체를 할당하는 FLAT 형태의 Extent는 실제 하드 디스크의 이미지와 동일한 구조이다. FLAT 형태의 Extent를 사용할 경우 디스크립터 파일은 Extent에 포함되지 않고 독립적인 파일로 존재하며, 각각의 파일 이름은 '〈vmname-flat.vmdk〉', '〈vmname.vmdk〉'의 형식을 갖는다. 또한 복수의 Extent를 사용하는 'twoGbMaxExtentFlat' 형태의 가상 디스크의 Extent 파일은 '〈vmname-f순번.vmdk〉' 형식의 이름을 갖는다.

SPARSE 형태의 Extent로 구성된 이미지는 가상 디스크의 크기를 필요한 만큼만 할당한 후 점차 크기를 증가시킨다. VMware Workstation에서 'monolithicSparse' 형태의 가상 머신 이미지는 디스크립터 파일을 포함하는 〔그림 4〕의 구조로 이루어져 있으며, '〈vmname.vmdk〉' 형식의 파일 이름을 갖는다.

한편 'twoGbMaxExtentSparse' 형태의 가상 디스크로 이루어진 경우에는 '〈vmname.vmdk〉' 형식의 독립적인 파일로 디스크립터가 생성되며, '〈vmname-s순번. vmdk〉' 형식으로 가상 머신 이미지 파일들이 존재하게 된다. 이때의 Extent는 〔그림 4〕의 구조에서 디스크립터 영역을 제외한 형태이다.

SPARSE Extent 헤더는 파일의 식별자(Signature)를 가지고 있으며 SPARSE 포맷(format)의 버전, 그레인(grain) 크기, Extent의 전체 크기, 디스크립터 파일의 시작 오프셋과 크기, 그레인의 압축여부 등의 정보를 저장한다. 그레인이란 가상 디스크의 데이터를 저장하는 섹터들의 블록이다. 그레인의 기본 크기는 128섹터로 하나의 그레인에는 64KB의 데이터를 저장한다. Extent 헤더의 다음 위치에는 디스크립터 파일이 삽입되어 있고, 이후 가상 머신 이미지의 데이터 할당 정보를 기록하는 그레인 디렉터리와 그레인 테이블, 가상 머신의 실제 데이터를 저장하는 그레인 순서로 위치한다〔7〕.

디스크립터 파일은 가상 머신의 설정 정보를 텍스트 형식으로 저장하며, 〔그림 4〕와 같이 디스크립터 헤더와 Extent 디스크립션(description), 디스크 데이터베이스(Disk Database)로 구성된다. 헤더에는 인코딩 방식과 가상 디스크의 종류 등에 대한 정보가 저장되고, Extent 디스크립션 영역은 가상 디스크를 구성하는 모든 Extent에 대한 정보를 저장한다. 주요 항목으로 Extent의 종류, 전체 섹터 수, 이름,

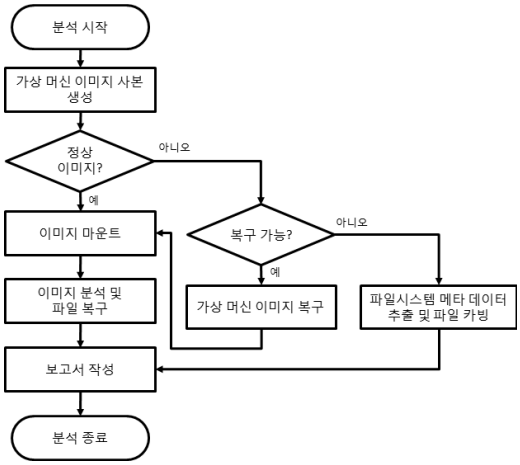
접근권한, 시작 오프셋 등이 있다. 디스크 데이터베이스 영역은 가상 머신 이미지에 대한 추가 정보를 저장하는 영역이다.

가상 머신 이미지에 대한 조사는 각 Extent의 이름, 전체 크기, 종류, 구조를 확인해야 한다. Extent 파일 이름을 통해 조사해야 할 가상 머신 파일을 선별할 수 있으며, 각 Extent의 전체 크기를 통해 가상 머신의 최대 크기를 확인할 수 있다. 또한 Extent의 종류는 가상 머신 이미지의 할당 방식과 파일 형식을 짐작할 수 있는 단서이다. 참고로 Extent의 종류에는 앞서 설명한 FLAT, SPARSE가 일반적으로 널리 사용된다. 이외에 ZERO, VMFS, VMFS SPARSE, VMFSRDM, VMRSRAW 가 있으며, 이와 같은 Extent들은 일반적으로 사용되지 않으므로 자세한 설명을 생략한다.

III. 가상 머신 이미지에 대한 조사 절차

가상 머신은 실제 시스템과 동일하게 동작하기 때문에 기존의 디지털 포렌식 조사 관점을 적용하여 디스크 및 메모리에 대한 조사를 수행할 수 있다. 호스트 시스템으로부터 가상 머신의 이미지 파일, 가상 머신 메모리 파일과 설정 파일들을 수집한다. 하지만 경우에 따라 삭제된 가상 머신 이미지를 복구하는 과정과 사용자의 의도적인 증거 인멸 행위에 의해 손상된 가상 머신 이미지를 획득하게 된다. 손상된 가상 머신 이미지는 가상 머신의 동작 상태에 대한 분석이 불가능하며 가상 머신 이미지에 대한 직접적인 분석에 어려움이 따른다. 따라서 수집한 가상 머신 이미지의 상태에 따라서 조사 절차 및 방법을 다르게 할 필요가 있다. 〔그림 5〕은 수집한 가상 머신 이미지를 분석하기 위해 제안한 조사 절차이다.

3.1 정상적인 가상 머신 이미지 조사 방법



〔그림 5〕 가상 머신 이미지 조사 절차

정상적인 가상 머신 이미지는 기존의 디지털 포렌식 조사 관점을 활용한 정적 분석 방법을 적용할 수 있다. 또한 많은 연구가 이루어지고 있는 가상화 포렌식 조사 플랫폼을 이용한 동적 분석 방법을 적용할 수 있다.

완전한 가상 머신 이미지에 대한 조사는 증거의 손상 및 변조를 방지하기 위해 수집한 가상 머신 이미지 파일의 사본을 생성하고, 마운트(Mount)하여 분석 시스템의 서브 디렉터리(Sub Directory)로 인식시키는 방법을 이용한다.

마운트 방법으로 VMware Workstation 제품의 가상 머신 이미지를 논리적인 디스크로 매핑(Mapping)시키는 기능을 이용하는 방법이 있으며, GetData社의 Mount Image Pro〔19〕와 ASR Data社의 SmartMount〔20〕를 이용하는 방법이 있다. Mount Image Pro와 SmartMount 도구는 실제 하드 디스크 이미지와 동일한 FLAT 형태의 Extent로 구성된 가상 머신 이미지뿐만 아니라 SPARSE 형태의 Extent로 구성된 가상 머신 이미지까지 마운트가 가능하다.

0x	0 1 2 3	4 5 6 7	8 9 A B	C D E F
0	매직 넘버	버 전	SPARSE 형태 버전	(a) Extent의 전체 크기 (섹터 단위)
10		그레인 크기		디스크립터 파일 위치 (섹터 단위)
20		디스크립터 파일 크기		그레인 테이블 엔트리 수
30	그레인 디렉터리 복사본 위치 (섹터 단위)		(b) 그레인 디렉터리 위치 (섹터 단위)	
40	(c) 메타 데이터가 할당된 섹터 수			

〔그림 6〕 SPARSE Extent 헤더 구조

마운트 된 이미지는 Guidance Software社의 EnCase와 같은 포렌식 조사 도구를 이용하여 키워드 검색 및 삭제된 파일 복구를 수행하고, 레지스트리 하이브 파일 및 웹 브라우저에 의해 생성된 파일, 프리패치 파일 등을 조사하여 사용자의 행위를 분석할 수 있다.

다른 조사 방법인 동적 분석은 가상 머신 이미지의 사본을 이용하여 가상 머신을 동작시킬 경우 원본에 대한 훼손 없이도 활성 상태를 조사할 수 있는 장점이 있다. 이와 같은 동적 분석은 가상 머신의 하이퍼바이저를 통하여 가상 머신 이미지의 활성 상태 정보를 수집하는 방법으로 이와 같은 디지털 포렌식 조사 플랫폼(platform)에 대한 가상화 연구가 활발하게 진행되고 있다[16].

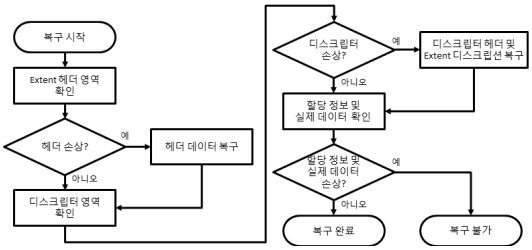
3.2 손상된 가상 머신 이미지 조사 방법

가상 머신 이미지는 수집 과정에서 손상될 수 있으며 사용자가 의도적으로 손상시켰을 수도 있다. SPARSE 형태의 Extent로 이루어진 가상 머신 이미지는 특정 부분이 손상되었을 경우 정상적인 파일시스템을 구성할 수 없다. 하지만 손상된 가상 머신 이미지에도 의미 있는 데이터가 남아있을 수 있기 때문에 조사가 필요하다.

3.2.1 복구 가능한 가상 머신 이미지

실제 하드 디스크의 이미지와 동일한 구조를 갖는 FLAT Extent 형태의 가상 머신 이미지와는 달리 SPARSE Extent 형태의 가상 머신 이미지는 VMware Workstation만의 고유한 파일 구조를 사용하기 때문에 손상되었을 경우 분석이 어렵다. 하지만 SPARSE Extent 형태의 가상 머신 이미지는 손상 범위에 따라 복구가 가능하다.

SPARSE Extent 복구는 [그림 7]과 같이 Extent 헤더와 디스크립터 영역에 대한 데이터 복구 과정으로 이루어진다. 이와 같은 과정에서 앞서 설명한 VMX와 LOG 파일의 정보를 이용해야 한다.



[그림 7] SPARSE Extent 복구 절차

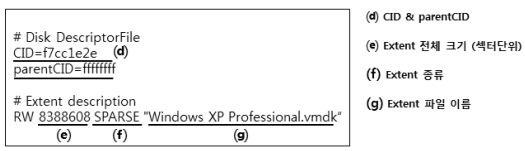
3.2.1.1 SPARSE Extent 헤더 영역 복구

433바이트 패딩(padding)을 포함하여 총 512바이트 크기로 이루어진 SPARSE Extent의 헤더를 복구하기 위해서는 [그림 6]에서 Extent의 전체 크기(a), 그레인 디렉터리 위치(b), 메타 데이터가 할당된 섹터 수(c) 항목을 각 Extent에 맞게 설정해야 한다. (a) 항목은 Extent 전체 크기를 섹터 단위로 저장하며, 가상 머신의 LOG 파일을 통해서 설정된 값을 확인할 수 있다. (b) 항목은 Extent의 데이터 할당 정보를 가지고 있는 그레인 디렉터리의 위치를 섹터 단위로 저장한다. (c)는 헤더, 디스크립터, 그레인 디렉터리, 그레인 테이블 등의 Extent 메타 데이터들이 할당된 전체 크기를 섹터 단위로 저장하는 항목이다. 3가지 이외의 항목들은 공통적으로 사용되는 기본 값을 이용하여 설정한다.

3.2.1.2 SPARSE Extent 디스크립터 영역 복구

Extent에 삽입된 14섹터의 크기의 디스크립터 영역이 손상된 경우에는 [그림 8]과 같이 디스크립터 파일 헤더의 CID와 parentCID, Extent 디스크립션 영역의 Extent 전체 크기, 종류, 파일 이름 데이터를 복구해야 한다.

(d) 항목은 Content ID와 Parent Content ID를 의미한다. CID는 32비트 임의의 값으로 처음 가상 머신이 동작할 때 생성되며, 이후 가상 머신이 재시동할 때마다 변경된다. ParentCID는 가상 머신의 스냅샷을 생성했을 경우 원본 Extent의 CID를 저장하는 항목으로, 스냅샷이 없는 경우 기본 값 'ffffffff'로 설정된다. (e)는 [그림 6]의 (a) 항목과 같은 값으로 (a) 항목을 10진수로 표현하여 저장한다. (f) 항목은 SPARSE, FLAT 등의 Extent 종류를 기록한다. Extent의 종류는 가상 머신을 구성하는 파일 이름과 VMX, LOG 파일을 통해서 확인할



[그림 8] 복구된 디스크립터 데이터

수 있다. (g) 항목 또한 VMX와 LOG 파일을 통해서 확인할 수 있는 Extent 파일의 이름을 저장한다.

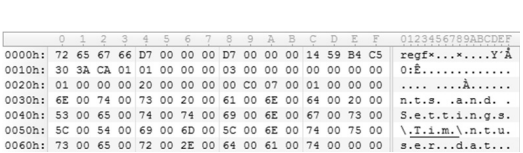
Extent 헤더, 디스크립터 복구가 완료된 가상 머신 이미지는 완전한 가상 머신 이미지를 분석하는 방법을 적용하여 조사할 수 있다. 만약 할당 정보 영역과 실제 데이터 영역이 손상되었을 경우에는 복구가 불가능하기 때문에 다른 조사 방법을 적용해야 한다.

3.2.2 복구 불가능한 가상 머신 이미지

SPARSE Extent 형태로 이루어진 가상 머신 이미지의 경우 그레인 디렉터리 및 그레인 테이블까지 손상되었다면 가상 머신 이미지 마운트를 통한 정적 분석이 불가능하다. 또한 동작되지 않기 때문에 동적 분석도 수행할 수 없다. 따라서 수집된 가상 머신 이미지가 복구 불가능한 상태라면 이미지 파일에 대한 직접적인 조사를 해야 한다. 이미지 파일에 대한 조사는 남아있는 데이터 복구 방법과 파일 시스템의 메타 데이터를 조사하는 방법을 이용할 수 있다.

가상 머신 이미지는 그레인 단위로 조각내어 RAW 데이터를 저장하지만 데이터가 연속된 그레인에 할당되었다면 파일 카빙(file carving)을 통해서 의미 있는 데이터를 복구할 수 있다. 문서 파일과 그림 파일 등의 사용자 행위의 단서가 될 수 있는 파일이 주 복구 대상이며, 웹 페이지 파일과 웹 브라우저(Web Browser)의 사용 흔적 파일을 복구하여 접근했던 사이트에 대한 정보를 획득할 수 있다. 특히 가상 머신이 윈도우(Windows) 시스템일 경우 카빙을 통해 〔그림 9〕과 같은 'regf' (\x72656766) 식별자를 갖는 레지스트리 파일 등을 획득한다면 사용자 계정과 사용 흔적에 대한 정보를 획득할 수 있다〔17〕.

파일 카빙과 더불어 파일 시스템의 메타 데이터를 추출하여 파일들의 시간관계를 확인하는 조사도 유용하다. 가상 머신의 파일시스템이 NTFS인 경우 가상 머신 이미지에 'FILE' 식별자를 갖는 1024바이트 크기의 MFT 엔트리(Entry)가 존재한다. 16개의 파일 시스템 메타 데이터를 가상 머신 이미지를 대상으로 추출하여 파일시스템의 할당 정보를 확인할 수 있으며, 각 파일의 MFT 엔트리들을 통해서 가상 머신에 존재했던 파일의 이름, 속성, 할당된 클러스터, MACE (Modified, Accessed, Created, Entry Modified) 시간 등의 정보를 확인할 수 있다. FAT 파일시스템을 사용하는 가상 머신의 경우에도 구조 검증을 통해서 디렉터리 엔트리(Direc- tory Entry)를 추출할 수 있다. 추출한 디렉터리 엔트리를 통해 가상 머신에 존재한 파일들의 이름과 MAC 시간 등의 정보를 확인할 수 있다. 이와 같이 복구가 불가능한 가상 머신의 경우에도 많은 정보를 획득할 수 있다.



	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	0123456789ABCDEF
0000h:	72	65	67	66	D7	00	00	00	D7	00	00	00	14	59	B4	C5	regf×...×....Y´Å
0010h:	30	3A	CA	01	01	00	00	00	03	00	00	00	00	00	00	00	0:Ê..............
0020h:	01	00	00	00	20	00	00	00	00	C0	07	00	01	00	00	00	....À......
0030h:	6E	00	74	00	73	00	20	00	61	00	6E	00	64	00	20	00	n.t.s. .a.n.d. .
0040h:	53	00	65	00	74	00	74	00	69	00	6E	00	67	00	73	00	S.e.t.t.i.n.g.s.
0050h:	5C	00	54	00	69	00	6D	00	5C	00	6E	00	74	00	75	00	\.T.i.m.\.n.t.u.
0060h:	73	00	65	00	72	00	2E	00	64	00	61	00	74	00	00	00	s.e.r...d.a.t...

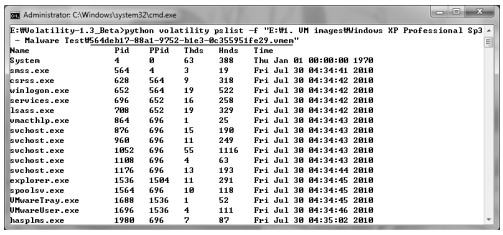
〔그림 9〕 복구된 레지스트리 파일 내의 계정 정보

3.3 가상 머신의 메모리 조사 방법

앞서 기술한 바와 같이 가상 머신은 완전한 시스템과 같이 동작하기 때문에 가상 머신 이미지에 대한 조사 이외에도 가상 머신에서 사용하는 메모리에 대한 조사가 필요하다. 메모리에는 활성 상태의 시스템에서 실행 중인 프로세스 내역, 네트워크 연결 상태, 실행 중인 서비스 내역, 아이디, 패스워드 등의 정보가 저장된다.

VMware Workstation의 가상 머신은 메모리를 VMEM파일로 백업하여 관리한다. VMEM파일은 가상 머신이 동작 중이거나 서스펜드 상태, 스냅샷을 생성한 경우에 존재하기 때문에 VMEM 파일에 대한 조사는 가상 머신에서 수행 중인 작업의 활성상태 정보로써 의미 있는 데이터를 획득할 수 있다.

VMEM 파일 분석에는 역공학 전문가 Zairon의 Compare VMware snapshots〔21〕 도구와 Chris Betz의 Memparser〔22〕를 사용할 수 있다. 또한 윈도우 XP 운영체제의 VMEM 파일은 파이썬 기반의 도구인 Volatility Framwork를 이용해 다양한 정보를 편리하게 분석할 수 있다〔1〕. Volatility Framework는 datetime, ident, dlllist, procdump, pslist, psscan 등의 명령어를 제공한다〔18〕. 위의 도구 사용방법 이외에도 수집된 모든 VMEM 파일은 문



〔그림 10〕 Volatility Framework를 통해 획득한 가상 머신의 프로세스 리스트

자열 검색 및 추출 방법과 커널 객체 구조 분석 방법, 물리 메모리 카빙 기법을 활용할 수 있다. 〔그림 10〕은 동작 중인 윈도우 XP 가상 머신의 VMEM 파일을 Volatility Framwork를 이용하여 확인한 프로세스 리스트이다. 실제 시스템과 같이 가상 머신의 메모리에서도 유용한 활성 데이터를 획득할 수 있으며 메모리 덤프와 같은 작업이 없어도 기본적으로 생성되는 파일을 이용할 수 있기 때문에 쉽고 빠르게 조사할 수 있다.

IV. 결 론

정부 및 기업의 정보화 시스템이 대형화되고 복잡해짐에 따라 가상화 기술의 적용이 가속화 되고 있다. 또한 개인 사용자들의 시스템 환경에도 영향을 미치고 있기 때문에 가상화 환경의 적용 범위는 점차 넓어지고 있다. 가상화 기술의 핵심 요소인 가상 머신은 하나의 물리적 시스템과 동일한 역할을 하기 때문에 가상화 환경에서의 주요 조사대상이다. VMware Workstation 가상 머신의 이미지는 기존 하드 디스크 이미지와 구조가 다르기 때문에 디지털 포렌식 조사를 어렵게 할 수 있다. 따라서 디지털 포렌식 관점에서 가상 머신 이미지에 대한 이해 및 조사 방법 연구가 이루어져야 한다. 특히 이미지에 대한 사용자의 고의적인 손상과 수집 과정에서 발생할 수 있는 손상에 대비하여 손상된 이미지에 대한 복구와 조사 방법이 필요하다.

본 논문에서는 VMware Workstation의 가상화 기술과 가상 머신 이미지에 대하여 설명하고, 가상 머신 이미지의 상태에 따라 조사 절차를 제안하였다. 이와 더불어 손상된 가상 머신 이미지에 대한 복구 기술과 복구 불가능한 가상 머신 이미지에 대한 조사 방법을 제안함으로써 VMware Workstation의 가상 머신 이미지에 대한 조사 방법을 구체화시켰다.

향후에는 이미 선점하고 있는 운영체제 시장을 기반으로 가상화 분야의 빠른 성장을 보이고 있는 마이크로소프트 데스크톱 가상화 기술과 VHD 가상 머신 이미지에 대한 조사 방법을 연구할 것이다.

참고문헌

〔1〕 Christiaan Beek, Virtual Forensics, Black-Hat Europe 2010, http://www.blackhat.com/html/bh-us-10/bh-us-10-briefings.html#Beek, Apr. 2010.

〔2〕 Greg Dorn, Chris Marberry, Scott Conrad, and Philip Craiger, "Analyzing the impact of a virtual machine on a host machine," International Federation for Information Processing, Advances in Digital Forensics V, IFIP AICT 306, DOI: 10.1007/978-3-642-04155-6_5, pp. 69-81, 2009.

〔3〕 Richard Arthur Bares, "Hiding in a Virtual World Using Unconventionally Installed Operating Systems," ISI 2009, pp. 276-284, Jun. 2009.

〔4〕 권태석, 방제완, 임경수, 이상진, "가상화 환경에서의 디지털 포렌식 조사 방법론 연구," 한국정보기술학회, 한국정보기술학회논문지, 7(2)호, pp. 159-167, 2009년 4월.

〔5〕 김동희, 백승조, 심미나, 임종인, "서버 가상화 환경의 가상머신 이미지에 대한 법적 증거로서의 허용성에 관한 연구", 한국정보보호학회, 정보보호학회논문지, 18권 6(A)호, pp. 163-177, 2008년 12월.

〔6〕 탁정수, 가상화 기술현황과 공공기관 적용 시사점, 한국정보사회진흥원, 정보사회 현안 분석Ⅱ, pp. 1-21 , 2007년 12월.

〔7〕 VMware Virtual Disks Virtual Disk Format 1.1, www.vmware.com/app/ vmdk/?src=vmdk, vmware technical note

〔8〕 Brett Shavers, A Discussion of Virtual Machines Related to Forensics Analysis, http://www.forensicfocus.com/downloads/virtual-machines-forensics-analysis.pdf

〔9〕 Derek Bem, "Virtual Machine for Computer Forensics - the Open Source Perspective," Open Source Software for Digital Forensics, DOI 10.1007, pp. 25-42, Jan. 2010.

〔10〕 소프트웨어 시장 동향 및 전망, 소프트웨어 산업백서 2008, pp. 187-389, 2008년 12월.

〔11〕 Jeff Daniels, "Server Virtualization Architecture and Implementation," ACM Crossroads, Vol. 16 No. 1, Sep. 2009.

〔12〕 Derek Bem and Ewa Huebner, "Analysis of USB Flash Drives in a Virtual Environment," Small Scale Digital Device

Forensics Journal, Vol 1. No. 1, Jun. 2007.
〔13〕 Karl Ray, "Server Virtualization and Virtual Machine Operating Systems," http://anengineersperspective.com/wp-content/uploads/2010/03/VM.pdf, Mar. 2010.
〔14〕 가상화의 기본 개념, http://www.vmware.com/kr/technology/virtual-machine.html
〔15〕 Harry van der Lint, Michiel Alkemade, "Turbulentie betekent kansen, maar bent u up-to-date?," Computer Profile, pp 36-37, Sep. 2009.
〔16〕 Kara Nance, Matt Bishop, and Brian Hay, "Investigating the Implications of Virtual Machine Introspection for Digital Forensics", 2009 International Conference on Availability, Reliability and Security, pp. 1024-1029, Mar. 2009.
〔17〕 H. Carvey, "The Windows registry as a forensic resource," Digital Investigation, pp. 201-205, Sep. 2005.
〔18〕 Volatile memory extraction utility framework, http://www.volatilesystems. com/volatility/1.3/README.txt
〔19〕 GetData, Mount Image Pro V4, http://www.mountimage.com/download-computer-forensics-software.php?file=MIP-Setup.exe
〔20〕 ASR Data, SmartMount, http://www.asrdata.com/SmartMount/
〔21〕 Zairon, Compare VMware snapshots, http://zairon.wordpress.com/2007/08/31/find-out-hidden-files-comparing-vmwares-snapshots/
〔22〕 Chris Betz, memparser, http://www.dfrws.org/2005/challenge/memparser.shtml

〈著者紹介〉

임 성 수 (Sungsu Lim) 학생회원
2007년 8월: 연세대학교 공학사
2009년 3월~현재: 고려대학교 정보경영공학전문대학원 정보보호전공 석사과정
〈관심분야〉 디지털 포렌식, 가상화, 데이터 복구, 데이터베이스

유 병 영 (Byeongyeong Yoo) 학생회원
2009년 2월: 상명대학교 이학사
2009년 3월~현재: 고려대학교 정보경영공학전문대학원 정보보호전공 석사과정
〈관심분야〉 디지털 포렌식, 파일 시스템

박 정 흠 (Jungheum Park) 학생회원
2007년 2월: 한양대학교 공학사
2009년 2월: 고려대학교 공학석사
2009년 3월~현재: 고려대학교 정보경영공학전문대학원 정보보호전공 박사과정
〈관심분야〉 디지털 포렌식, 안티 포렌식

변 근 덕 (KeunDuck Byun) 학생회원
2004년 2월: 아주대학교 공학사
2006년 2월: 고려대학교 공학석사
2006년 3월~현재: 고려대학교 정보경영공학전문대학원 정보보호전공 박사과정
〈관심분야〉 임베디드 포렌식, 역공학

이 상 진 (Sangjin Lee) 종신회원
1987년 2월: 고려대학교 수학과 이학사
1989년 2월: 고려대학교 수학과 이학석사
1994년 8월: 고려대학교 수학과 이학박사
1989년 10월 ~ 1999년 2월: ETRI 선임 연구원
1999년 3월 ~ 2001년 8월: 고려대학교 자연과학대학 조교수
2001년 9월 ~ 현재: 고려대학교 정보경영공학전문대학원 정교수
〈관심분야〉 대칭키암호, 정보은닉, 디지털 포렌식