情報保護學會論文誌
第 21 卷 第 6 號, 2011. 12

IaaS 유형의 클라우드 컴퓨팅 서비스에 대한 디지털 포렌식 연구*

정 일 훈†, 오 정 훈, 박 정 흠, 이 상 진‡
고려대학교 정보보호연구원

Digital Forensic Methodology of IaaS Cloud Computing Service*

IlHoon Jeong†, JungHoon Oh, JungHum Park, Sangjin Lee‡
Center for Information Security Technologies, Korea University

요 약

최근 유무선 통신 네트워크의 확산 및 고속화에 따라 인터넷 기술을 활용한 높은 수준의 확장성을 제공하는 클라우드 컴퓨팅 서비스(Cloud Computing Service) 이용이 증가하고 있다.

클라우드 컴퓨팅 서비스란 네트워크, 서버, 스토리지, 응용프로그램 등 다양한 컴퓨팅 자원들의 공유된 풀에 네트워크로 접근하여 언제든지 편리하게 사용가능한 컴퓨팅 방식으로써 컴퓨팅 환경의 가상화라는 클라우드 컴퓨팅 서비스의 본질적인 특성으로 인해 디지털 포렌식 관점에서 사건 수사 시 데이터를 확보하는 일 자체가 어려운 현실에 직면했다.

본 논문에서는 클라우드 컴퓨팅 서비스에 대한 디지털 포렌식 관점의 연구와 IaaS 형태의 클라우드 컴퓨팅서비스 중 시장 점유율의 대부분을 차지하고 있는 AWS(Amazon Web Service)와 Rackspace에 대한 증거데이터 수집 및 분석방안을 제시한다.

ABSTRACT

Recently, use of cloud computing service is dramatically increasing due to wired and wireless communications network diffusion in a field of high performance Internet technique. Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. In a view of digital forensic investigation, it is difficult to obtain data from cloud computing service environments. therefore, this paper suggests analysis method of AWS(Amazon Web Service) and Rackspace which take most part in cloud computing service where IaaS formats presented for data acquisition in order to get an evidence.

Keywords: Cloud Computing Service, Digital Forensic, IaaS, AWS, Rackspace

접수일(2011년 7월 5일), 게재확정일(2011년 9월 5일)
* 본 연구는 지식경제부 및 한국산업기술평가관리원의 산업원천기술개발사업의 일환으로 수행되었습니다.
〔10035157, 실시간 분석을 위한 디지털 포렌식 기술 개발〕

† 주저자, ihjeong@korea.ac.kr
‡ 교신저자, sangjin@korea.ac.kr

I. 서 론

유무선 통신 네트워크의 고속화와 무료 소프트웨어의 보급 확대, 그리고 가상화 기술의 발전 등 IT 인프라의 급속한 발전은 IT 시장의 패러다임 자체에 영향을 미쳤고 그에 따라 최근 IT 산업의 최대 화두인 클라우트 컴퓨팅 서비스 시장의 발전과 이용자의 급증으로 이어졌다.

IT 시장조사기관인 가트너(Gartner)는 클라우드 컴퓨팅을, 인터넷 기술을 활용하여 다수의 고객들에게 높은 수준의 확장성을 제공하는 컴퓨팅 환경이라고 정의하고 있으며, 2010년 클라우드 컴퓨팅 서비스의 수익은 약 683억 달러로 전년도 대비 16.6% 증가했고, 2014년에는 서비스 수익이 무려 1,488억 달러에 이를 것으로 전망했다[1].

컴퓨팅 환경의 가상화라는 클라우드 컴퓨팅의 본질적 특성상 현재까지의 사건 처리방법과 디지털 포렌식 조사방법은 클라우드 컴퓨팅의 발전과 함께 변화가 요구된다.

본 논문에서는 클라우드 컴퓨팅 서비스의 특징에 대해 소개하고, 클라우드 컴퓨팅 환경에서의 디지털 포렌식 문제점과 증거 데이터의 구성요소에 대해 설명한다. 마지막으로 클라우드 컴퓨팅 서비스의 핵심적인 모델인 IaaS(Infra as a Service)에 대한 디지털 포렌식 관점의 조사방법과 IaaS 모델 중 시장의 대부분을 점유하고 있는 Amazon Web Service(AWS)와 Rackspace에 대해 데이터 수집 및 분석 방안에 대해 제시한다.

II. 클라우드 컴퓨팅 서비스의 특성

클라우드 컴퓨팅 서비스는 강력한 컴퓨팅 리소스를 바탕으로 원격지의 클라우드 환경에서 서비스를 제공하고 클라이언트는 단순 컴퓨팅 파워만 제공하는 단말에서 웹 브라우저를 통해 이용하는 형태를 갖는다. 이는 최소한 의 관리 노력으로 빠르게 서비스를 공급받을 수 있으며, 쉽고 편리하게 서비스를 이용할 수 있음을 의미한다.

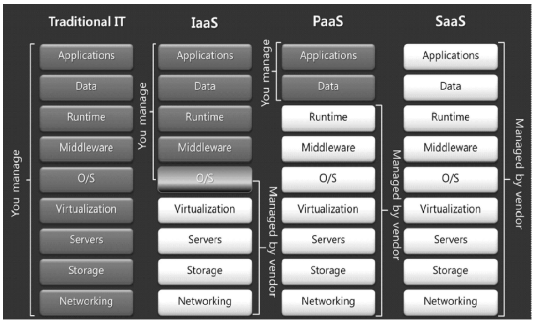
〔그림 1〕 클라우드 5-4-3 모델

클라우드 컴퓨팅의 개념은 〔그림 1〕과 같이 5가지의 주요 특징과 4가지의 배치모델, 그리고 3가지의 서비스 모델로 구성된 5-4-3 모델에 기초한다[2].

먼저 클라우드 컴퓨팅의 5대 주요 특징은 첫째, 주문형 셀프서비스 형태이다. 컴퓨팅 자원이 필요한 경우 서비스 제공자의 개입 없이 언제든지 네트워크 스토리지, 서버 등과 같은 자원에 접근하여 사용할 수 있는 특징이 있다. 둘째, 광대역망 액세스 형태의 서비스이다. 사용자는 여러 종류로 이루어진 클라이언트 플랫폼에 의해 서비스에 접근할 수 있으며, 이는 네트워크를 통해 사용되는 특징이 있다. 셋째, 자원 공동관리의 형태이다. 이는 다중 임대 모델(Multi-Tenant Model)을 통한 자원 할당을 가능하게 한다. 넷째, 빠른 요구 탄력성을 가진다. 사용자는 비즈니스 상황에 따라 탄력적으로 자원 활용이 가능하다. 다섯번째, 사용자는 서비스를 사용한 만큼 비용을 지불하는 도수제 특성을 갖는다.

다음으로 클라우드 컴퓨팅은 4가지 배치모델로 구분된다. 먼저 주로 다수의 대중이나 거대한 산업 그룹을 위해 운영되는 공공 클라우드 모델(Public Cloud)과 폐쇄적인 조직을 위해 운영되며 주로 보안 기능이 강화된 방화벽 내에서만 배타적으로 이용이 가능한 형태인 사설 클라우드 모델(Private Cloud), 특별한 보안요구나 공통적인 관심사를 공유한 조직의 그룹이 제어하고 이용하는 커뮤니티 클라우드 모델(Community Cloud), 그리고 공공 클라우드와 사설 클라우드, 커뮤니티 클라우드 중 두 개 이상의 클라우드로 구성된 형태로써, 비즈니스에 중요하고 보안이 필요한 서비스와 데이터는 사설 클라우드로 관리하고 상대적으로 덜 중요한 정보는 공공 클라우드로 관리하는 특징을 갖는 하이브리드 클라우드(Hybrid Cloud)로 구성된다.

마지막으로 클라우트 컴퓨팅을 3가지의 서비스 형태로 구분 가능하다. 즉, 서버나 스토리지를 제공하는 형태인 IaaS(Intra as a Service) 모델과 플랫폼을 제공하는 PaaS(Platform as a Service) 모델, 그리고 소프트웨어를 제공하는 SaaS(Software as a Service) 모델 이렇게 세 가지의 형태로 구분 가능하다.



〔그림 2〕 클라우드 컴퓨팅 서비스 모델 형태

클라우드 컴퓨팅 서비스는 이용자 관점에서 간단하고 유연하게 컴퓨팅 자원을 활용할 수 있다는 큰 장점이 있지만 정전 혹은 데이터 자체의 훼손 및 유출 등 보안관점의 위험 또한 존재한다.

III. 클라우드 컴퓨팅 포렌식의 문제점

클라우드 컴퓨팅 서비스에서 데이터는 다양한 물리적인 위치와 논리적인 위치를 갖는다.

클라이언트 측에서는 기술적인 제어나 모니터링을 클라이언트 또는 소유자의 컴퓨터와 네트워크에서 실행하는데, 침입탐지시스템, 웹 컨텐츠 엔진 로깅, 방화벽, 접근 로그 등의 데이터를 확인할 수 있다. 클라이언트와 공급자 양쪽의 결합된 측면에서는 기술적인 제어나 모니터링을 클라우드 고객에게 할당한 컴퓨터와 네트워크에서 실행가능하고 접근 로그, 트랜잭션 로그, 사용 로그 등의 데이터를 확인 가능하다. 공급자 측의 경우, 기술적인 제어나 모니터링을 클라우드 서비스를 구성하는 컴퓨터나 네트워크에서 실행하게 되고 방화벽, 로드 밸런서, 어드민 접근 로그, 침입탐지시스템, 넷 플로우 데이터 등을 확인할 수 있다.

클라우드 컴퓨팅 환경에서 디지털 포렌식의 문제점은 관련 사건 발생시, 클라우드 컴퓨팅 서비스로부터 포렌식 조사 데이터를 얻기 어려운 부분이다. 기술적 관점으로 위치문제와 시간문제로 정리하면, 국제적으로 물리적인 드라이브들이 흩어져서 존재하기 때문에 데이터가 관할권이 다른 클라우드 벤더의 스토리지에 존재할 경우 데이터 확보의 문제점이 발생한다. 또한 클라이언트단의 로그 파일과 서비스공급자단의 로그 파일의 타임 스탬프가 다른 시간을 가지고 있다면 이는 증거로 다루기 어려울 수 있다.

클라우드 컴퓨팅 배치 모델 중 실질적으로 디지털 포렌식 관점에서 문제를 야기시키는 유형은 공공 클라우드 모델이다. 사설 클라우드 혹은 하이브리드 클라우드 모델의 경우 기업 환경에서 구성하는 형식으로써 수사 입장에서는 일반적인 IT 인프라 환경과 동일한 형태이기 때문에 데이터 압수 수색 시, 조사 대상 데이터에 물리적 접근이 가능하므로 문제가 발생하지 않는다.

클라우드 컴퓨팅의 등장으로 인해서 기존의 디지털 포렌식 조사 방법은 변화가 필요한 상황이다. 현재 디스크, 메모리, 네트워크 등을 공유하는 클라우드 환경에서 개체를 물리적으로 수집하기 어렵고, 전통적인 소유권 경계가 흐려지고 있는 국외 클라우드 서비스 업체에 대한 조사 관할권 문제가 존재한다. 따라서 법정에서 효력이 유지될 수 있는 조사방법론과 현재의 법적 한계점 내에서 상용 서비스 중인 클라우드 컴퓨팅 서비스에 대한 절차적 조사방안과 구체적 분석방법이 필요한 상황이다.

IV. 클라우드 컴퓨팅 포렌식 조사 방법

4.1 클라우드 컴퓨팅 환경의 증거데이터 구성 요소

클라우드 컴퓨팅 서비스에서는 증거 데이터로 활용 가능한 세 가지 형태의 요소가 존재한다[3].

첫 번째 요소인 Virtual Cloud Instance는 클라우드 내에 존재하는 가상 인스턴스로써 데이터가 저장된 위치나 프로세스가 구동된 위치와 같은 잠재적인 증거를 제공할 수 있다. 이 인스턴스는 CSP(Cloud Service Provider)와 인스턴스를 구동하고 있는 사용자 모두 접근이 가능하지만, SaaS와 PaaS 에서는 가상 인스턴스에 접근하는 것이 매우 제한적이거나 불가능하다.

두 번째 요소는 Network Layer로써 클라우드의 외부 인스턴스와 내부 인스턴스 사이의 통신 정보와 프로토콜의 여러 정보를 제공한다. 하지만 현재 일반적인 CSP는 네트워크 구성들의 로그 데이터를 제공하지 않는다.

세 번째 요소는 Client System으로써 시스템 내의 브라우저는 클라우드 내의 서비스와 통신하는 응용 프로그램 역할을 수행하기 때문에 브라우저를 분석하여 얻은 데이터는 중요한 증거데이터로 활용 가능하다.

본 연구에서는 Network Layer를 제외한 Virual Cloud Instance와 Client System의 데이터 관점으로 분석을 진행한다.

4.2 클라우드 컴퓨팅 포렌식 조사 절차

본 연구에서는 〔그림 3〕과 같이 클라우드 컴퓨팅 서비스 환경에 대한 디지털 포렌식 조사 절차를 수립하였다.

기존의 디지털 포렌식 절차에서 추가적으로 고려해야 할 사항은 디지털 증거의 최초 분석 시 가장 먼저 클라우드 컴퓨팅 서비스의 사용유무를 파악하는 절차이다. 만약 클라우드 상에 중요 증거 데이터가 존재하는 상황에서 디지털 증거 분석 시 클라우드 컴퓨팅 서비스의 사용 여부를 고려하지 않는다면, 아무리 정밀한 분석 절차를 진행하더라도 정작 핵심적인 증거 데이터들은 획득할 수 없을 것이다. 또한 최초 클라우드 컴퓨팅 서비스의 사용여부를 확인하지 않은 채 기존의 디지털 포렌식 분석 절차를 수행한 후 뒤늦게 클라우드 컴퓨팅 서비스 사용을 확인했다면, 중요한 증거가 사라졌을 가능성이 존재한다. 따라서 디지털 증거 분석의 최초 과정에서 가장 먼저 클라우드 컴퓨팅 사용유무를 확인한 후 본 논문에서 제시하는 클라우드 컴퓨팅 포렌식 조사 절차와 기존의 디지털 포렌식 조사 절차를 병렬로 진행하는 방안이 필요하다.

수많은 클라우드 컴퓨팅 서비스들이 존재하는 현실에서 용의자가 클라우드 컴퓨팅 서비스를 이용했는지, 더 나아가 어떤 상용 서비스를 이용했는지를 파악하는 일 자체가 쉽지 않다. 따라서 본 연구에서는 클라우드 시그니처(Cloud Signature)라는 개념을 제시한다.

클라우드 시그니처(Cloud Singature)란, 각각의 클라우드 컴퓨팅 서비스 별로 특정 아이덴티티(identity)를 지정하고 디지털 증거 분석 과정 중 조사 대상 클라이언트에서 해당 시그니처를 발견했다면 사용자가 어떤 클라우드 컴퓨팅 서비스를 사용했는지를 간단히 확인할 수 있는 개념이다. 따라서 디지털 증거 분석과정에서 최초 클라우드 시그니처 확인을 통해 용의자의 클라우드 컴퓨팅 서비스 사용유무와 사용한 서비스의 종류를 파악할 수 있다.

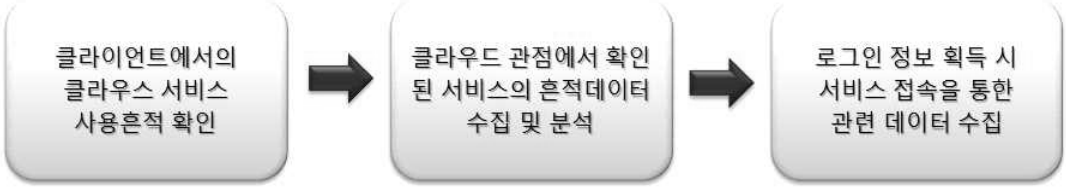
〔그림 4〕는 기존의 디지털 포렌식 상세 분석 절차에서 최초 분석 시 클라우드 시그니처를 통해 클라우드 컴퓨팅 서비스의 사용 유무를 확인해야 하는 내용과 클라우드 컴퓨팅 서비스의 유형별 특징을 고려한 클라우드 컴퓨팅 포렌식 관점의 분석 절차를 추가한 내용이다.

디지털 포렌식 상세 분석 절차는 대상 시스템 분석 시에 활성데이터 수집이 가능한 상황이라면 먼저 물리 메모리 및 가상 메모리 수집을 통해 활성데이터를 수집한 후 인터넷 사용 흔적분석이나, 레지스트리 분석, 파일분석 등 일반적인 디지털 포렌식 관점의 데이터를 분석하는 절차를 제시하는 가이드라인이다〔8〕.

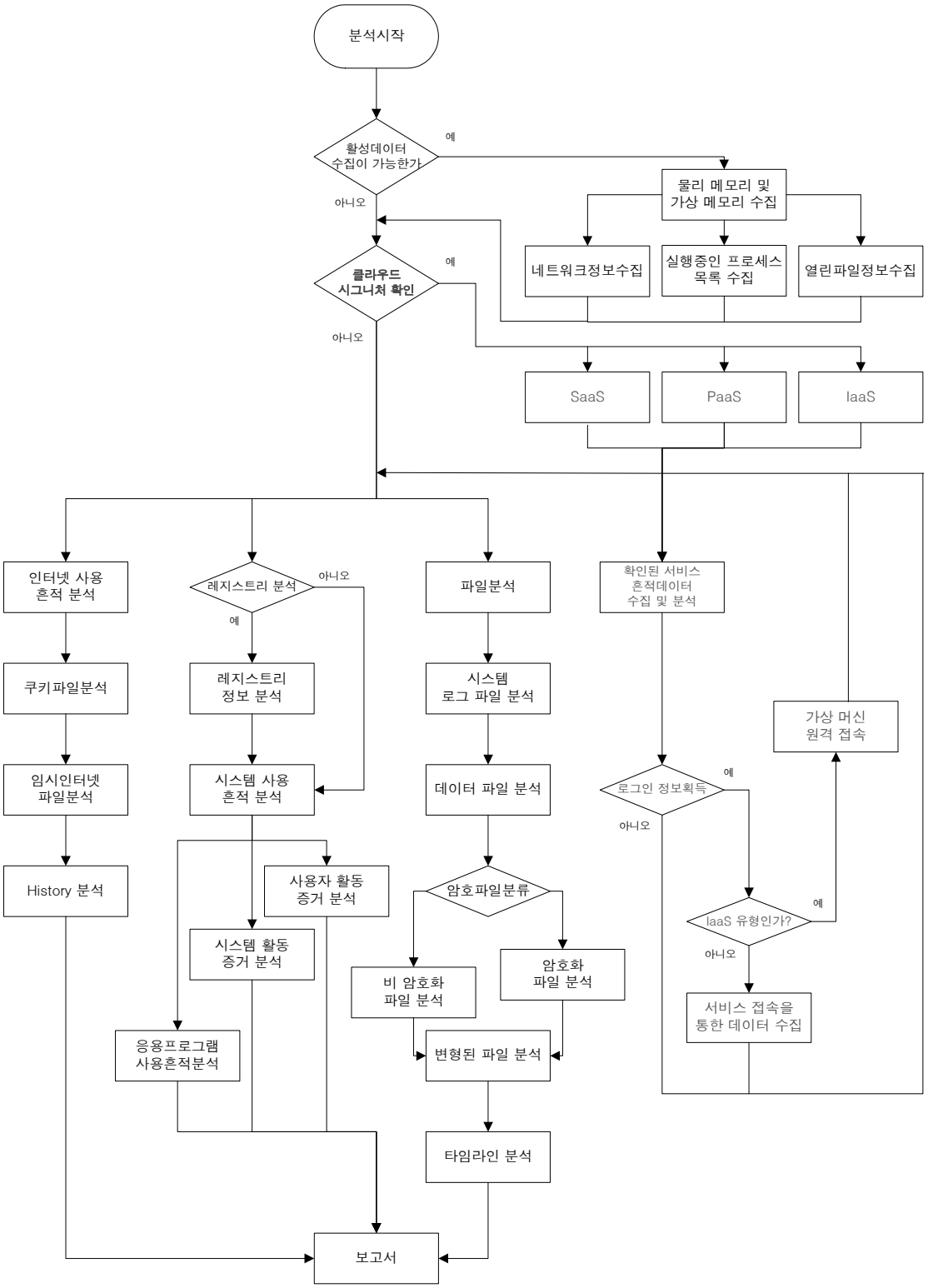
〔그림 4〕의 클라우드 컴퓨팅 서비스관점에서의 분석절차는 시스템 분석 시 먼저 클라우드 시그니처를 통해 시스템에서의 클라우트 컴퓨팅 서비스 사용 유무와 서비스 종류를 파악하고 확인된 해당 서비스의 흔적 데이터를 일반적인 디지털 포렌식 분석방법에 따라 수집 및 분석을 진행한다. 만약 IaaS 유형의 클라우드 컴퓨팅 서비스 사용 흔적이 발견되었다면 IaaS 유형의 경우 원격지의 데이터센터에 가상머신을 생성하여 원격접속을 통해 컴퓨팅 환경을 이용하는 특징을 가지고 있기 때문에 추가적으로 대상 분석 시스템에서 관련 로그인 정보 수집을 통해 원격 가상머신에 접근하여 관련데이터를 수집해야 한다.

〔표 1〕은 앞서 설명한 여러 클라우드 컴퓨팅 서비스별로 특정 아이덴티티를 지정하여 클라우드 시그니처를 정리한 목록이다.

결론적으로 앞으로의 디지털 포렌식 분석 과정에서는 〔그림 4〕에서 제시한 클라우드 컴퓨팅 서비스 디지털 포렌식 조사 절차에 따라 첫 번째 클라이언트에서 클라우드 시그니처를 이용하여 클라우드 서비스 유형별 사용흔적을 확인하고, 두 번째 확인된 서비스에 대한 클라이언트 관점에서 관련 데이터 흔적을 수집 및 분석을 진행한다. 클라우드 컴퓨팅 서비스는 기본적으로 웹브라우저를 통해 이용하는 형태를 가지기 때문에 클라이언트에 관련 흔적이 남을 수 있고, 그 외에 다양한 서비스 종속적 흔적 데이터가 존재할 수 있다.



〔그림 3〕 클라우드 컴퓨팅 서비스 디지털 포렌식 조사 절차



〔그림 4〕 클라우드 컴퓨팅 포렌식 관점의 분석 절차

〔표 1〕 클라우드 시그니처

종 류	서비스	시그니처	
IaaS	AWS	웹브라우저 흔적	aws
		원격데스크톱주소	amazonaws
		관련 소프트웨어	S3Fox, JungleDisk, Transmit, Mac, Backup Manager, S3 Backup
	Rackspcae	웹브라우저 흔적	rackspacecloud,rackcdn
		원격데스크톱주소	50.56.x.x
	Ucloud	웹브라우저 흔적	ucloud
		파일	sc1.log
	Tcloud	웹브라우저 흔적	Tcloudbiz
PaaS	Azure	웹브라우저 흔적	cloudpowersolutions
		관련 소프트웨어	Windows Azure AppFabric SDK
SaaS	ThinkFree	웹브라우저 흔적	thinkfree
	Zoho Office	웹브라우저 흔적	zoho
	MS Live Online	웹브라우저 흔적	office.microsoft
	Google Docs	웹브라우저 흔적	docs.google
	Glice OS	웹브라우저 흔적	glidedigital
	Photoshop Express Online	웹브라우저 흔적	photoshop.com/tools

그리고 마지막으로 수집된 데이터의 분석과정에서 해당 서비스의 로그인 정보를 획득했다면 직접 서비스에 접속하여 기존의 디지털 포렌식 분석절차를 진행해야 한다.

V. IaaS에 대한 디지털 포렌식 분석

5.1 IaaS 환경의 특징

IaaS는 클라우드 컴퓨팅 서비스 중 가장 높은 이용율을 기록하는 대표적인 클라우드 컴퓨팅 서비스 모델이다. IaaS는 고객에게 가상 머신(Virtual Machine)과 스토리지를 제공하는 특징을 가진 모델로써 클라이언트 시스템에서 원격접속을 통해 클라우드에 존재하는 가상 머신에 접속하여 컴퓨팅 자원을 활용하고 추가적으로 웹 하드 형태의 스토리지 자원을 제공하는 유형으로 디지털 포렌식 관점에서 SaaS와 PaaS 모델보다 많은 증거 데이터가 존재한다〔4〕.

IaaS 환경에서는 Snapshot, Volatile Data, Virtual Introspection과 같은 특징이 존재한다〔4〕.

Snapshot은 한 번의 클릭으로 실행 중인 시스템의 메모리도 함께 포함하여 복제할 수 있는 강력한 기능으로써 조사 과정에서 Snapshot을 얻을 수 있다면, 로그인된 사용자, 열린 포트, 구동중인 프로세스, 시스템 그리고 레지스트리 정보 등 해당 시점의 활성 및 비활성 데이터를 분석할 수 있다.

Volatile Data란, 가상 머신 인스턴스의 휘발성 데이터로써 지속적으로 보관할 데이터는 웹 하드와 같은 형식의 환경에 저장해야 한다. 현재 클라우드 환경에서는 가상환경의 휘발성 데이터가 삭제되었을 경우 확인이 불가능하다.

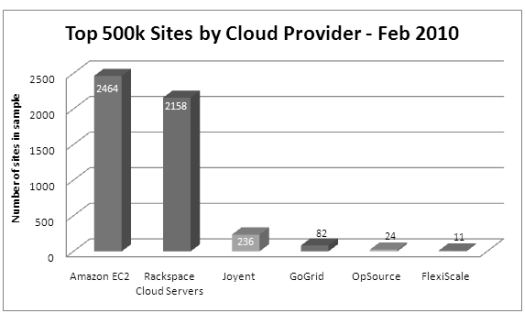
Virtual Introspection은 조사받을 대상이 아닌 다른 가상 머신이나 VMM(Virtual Machine Monitor)으로부터 조사받을 가상 머신의 상태를 관찰하는 과정이다. VMM은 모든 가상 머신의 자원에 대한 접근 권한을 가지므로 고객의 프라이버시에 위협이 될 수 있다. 하지만 이러한 기능은 포렌식 조사에서 유용하게 활용 가능하다.

〔그림 5〕는 IaaS 유형의 클라우드 서비스 제공사의 시장 점유율을 나타낸 그림으로써 AWS와 Rackspace가 점유율의 대부분을 차지하고 있음을 알 수 있다〔5〕. 국내의 경우 현재 AWS나 Rackspace와 같이 원격지에 사용자가 원하는 형태의 가상머신을 생성하여 컴퓨팅 환경을 제공하는 서비스형태가 존재하지 않기 때문에 국내외 사용자가 더욱 증가 것으로 판단되어 해당 서비스를 대상으로 분석을 진행하였다.

5.2 AWS(Amazon Web Service) 분석

앞서 수립한 클라우드 컴퓨팅 서비스 환경에서의 디지털 포렌식 조사 절차에 따라 IaaS 유형의 AWS (Amazon Web Service)를 분석한다.

구체적인 분석 절차는 먼저 대상 시스템에서 클라우드 시그니처를 통해 해당 클라우드 서비스의 사용유무를 확인하고, 분석 대상 시스템 사용자의 클라우드 컴퓨팅 서비스 사용여부가 확인되었다면 해당 서비스를 고려하여 웹 히스토리 분석이나 레지스트리 분석, 파일분석 등 일반적인 디지털 포렌식 분석과정에서의 클라우드 컴퓨팅 서비스 관련 정보를 수집 분석한다.



〔그림 5〕 클라우드 프로바이더 시장 점유율

추가적으로 IaaS 유형의 클라우드 컴퓨팅 서비스의 경우 앞서 설명했듯이 원격지에 가상머신을 생성하여 원격접속을 통해 서비스를 이용하는 형태를 갖기 때문에 확인된 해당 서비스의 특징적인 원격접속 로그인 정보를 확인한다.

AWS는 미국 아마존사의 클라우드 컴퓨팅 솔루션으로써 현재 전체 클라우드 컴퓨팅 서비스 점유율 1위의 서비스이자 IaaS 시장 점유율 1위의 서비스로써 해외는 물론 국내에서도 널리 이용되고 있다.

AWS에서 제공하는 서비스 중 IaaS 모델 관점에 해당하는 서비스인 EC2(Elastic Compute Cloud)와 S3(Simple Storage Service), Cloud Front에 대해 분석을 진행한다.

Amazon EC2는 AWS의 대표 서비스로써 가상화된 하드웨어 자원을 사용자에게 제공하고 사용자는 그 위에 OS와 소프트웨어를 설치하여 클라우드 서비스를 사용하는 개념으로 클라우드 내의 다양한 서버군을 가상머신 이미지로 생성하여 클라이언트 단에서 원격접속을 통해 컴퓨팅 자원을 이용하는 형태를 가진다. Amazon S3는 우리가 일반적으로 알고 있는 웹하드의 개념과 동일한 형태로써 사용자가 대용량의 데이터를 간편하게 저장하고 검색할 수 있도록 지원하는 확장성 높은 고속 인터넷 데이터 스토리지 시스템이다. Amazon S3는 웹이나 관련 소프트웨어를 통해 이용가능하다. Amazon Cloud Front는 AWS에서 제공하는 CDN(Content Delivery Network) 서비스로써 S3에 올려둔 파일들을 CDN 등록을 통해 여러 지역에서 지정된 URL 주소 입력을 통해 언제든지 접속하여 확인 가능하다.

5.2.1 클라이언트에서 AWS 사용흔적 확인

가장 우선적으로 클라이언트에서 클라우드 컴퓨팅 서비스의 흔적을 확인한다. 앞서 설명했듯이 클라우드 컴퓨팅 서비스는 기본적으로 웹을 기반하고 있기 때문에 브라우저의 히스토리 정보를 통해 방문했던 사이트의 URL과 시간정보 확인이 가능하다. 앞서 설명했던 클라우드 시그니처라는 개념을 이용하여 〔그림 6〕과 같이 클라이언트에서의 AWS의 사용흔적을 확인할 수 있다.

5.2.2 AWS 흔적 데이터 수집 및 분석

AWS의 사용유무를 확인했다면 다음 단계로 클라이언트 단에서 해당 서비스와 관련된 흔적 데이터를 확인한다.

〔그림 6〕의 내용과 같이 AWS URL내의 aws & aToken 값은 아마존 웹 서비스에 로그인할 때마다 새로 갱신되는 값이다. 해당 값을 통해 사용자가 AWS에 몇 회 로그인 세션을 맺었는지에 대한 정보를 확인할 수 있다. 또한 URL 정보를 통해 아마존 웹 서비스의 종류별 사용 내역을 확인할 수 있으며, 해당 서비스의 접속시간 또한 확인 가능하다.

클라이언트 시스템에 설치되어 있는 응용프로그램 중 〔표 2〕와 같은 AWS의 S3 서비스와 연동을 위해 사용되는 응용프로그램의 설치내역을 확인했다면 시스템에서의 클라우드 컴퓨팅 서비스의 사용여부를 판단할 수 있다.

AWS S3 서비스에 저장되어 있는 데이터를 웹 브라우저를 통해 다운로드 시 S3는 웹 브라우저 다운로더에 종속적으로 동작하기 때문에 웹브라우저 별 다운로더의 로그 기록 확인을 통해 용의자가 어떤 데이터를 다운로드 받았는지 확인 가능하다.

또한, CloudFront 서비스의 경우 AWS에서 제공하는 CDN(Content Delivery Network) 서비스로써 사용자의 웹 브라우저 URL 정보를 확인한다면 어떤 등록 데이터를 열람했는지 확인 가능하다.

〔그림 6〕 AWS 웹 히스토리 정보

〔표 2〕 AWS S3 연동 응용프로그램

응용프로그램	설 명
S3Fox	Plugin을 탑재한 Firefox 브라우저
Transmit	Mac용 FTP/SFTP 응용프로그램
Backup Manager	리눅스 용 명령형 도구
S3 Backup	윈도우 데스크톱 응용프로그램
jets3t	오픈소스 자바 툴킷 및 응용프로그램
JungleDisk	온라인 백업 프로그램
Sync2S3	아마존 S3 백업 응용프로그램
SME Storage	안드로이드 클라우드 파일 매니저
S3Sync	아마존 S3 백업 응용프로그램

5.2.3 로그인 정보 획득과 서비스 접속을 통한 관련 데이터 수집

AWS의 EC2 서비스의 경우 앞서 설명했듯이 서버군을 클라우드상에서 가상머신으로 생성한 후 원격접속을 통해 해당 서비스를 사용하는 형태이기 때문에 〔그림 7〕과 같이 클라이언트 시스템에서 원격접속 정보관련 레지스트리 값(HKEY_CURRENT_USER\Software\Microsoft\ Terminal Server Client\Default)의 확인을 통해서 사용자의 AWS EC2 서버의 원격 접속 주소를 확인할 수 있다.

원격 접속 주소를 확인했을 경우 Sam Spade나 IP2LOCATION 사이트를 통해 해당 주소에 대한 ISP 정보와 도메인 주소를 확인하면 그 내용들을 통해 AWS 서버인 것을 확인할 수 있다.

원격 접속 주소 확보 후 AWS의 EC2 가상머신 서버에 접속하기 위해서는 인증 패스워드가 필요하다.

〔그림 7〕 AWS 원격접속 주소 흔적

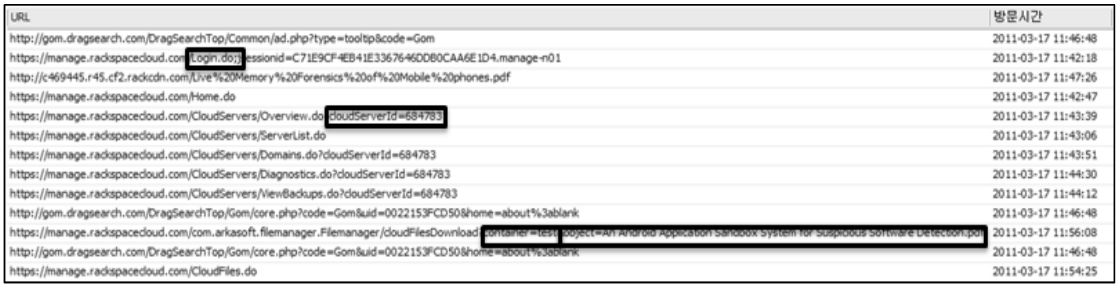
EC2 서비스의 패스워드는 〔그림 8〕과 같이 최초 가상머신 생성 시 AWS 가입절차에서 생성한 Private Key를 이용한다. 이 Private Key를 이용해 랜덤한 Decrypted Password를 생성하게 되고 해당 패스워드를 이용해 EC2 가상 머신에 접속하게 된다. 이 패스워드의 경우 랜덤한 임의의 값으로 생성되기 때문에 사용자가 기억하기 어려운면이 존재하고 회사 환경에서 다양한 계정 서비스를 이용할 경우 일반적으로 파일 형태로 저장 관리할 수 밖에 없는 특징이 있다.

원격 접속 주소 또한 일반적인 IP 정보가 아닌 AWS 고유의 주소 형태를 가지기 때문에 해당 주소를 파일형태로 저장하여 관리할 수밖에 없는 조건을 가진다. 따라서 디지털 포렌식 분석 시 해당 파일을 찾아낸다면, 용의자의 클라우드 컴퓨팅 환경에 직접적인 접근이 가능하다. 또한 패스워드 저장 파일 열람 시 링크파일의 형태로 시스템에 저장될 수 있기 때문에 해당 데이터의 확인도 필요하다.

하나의 가상 시나리오로써 USB에 해당 원격 접속 주소와 패스워드를 파일형태로 저장한 후 임의의 시스템에서 사용할 경우 AWS URL정보의 aws & aToken 데이터 생성시간을 통해 용의자의 최초 AWS 사용시간을 확인하고, 그 다음 USB 접속내역과 시간정보를 통해 사용자의 USB 연결시간을 확인한 뒤 마지막으로 링크파일 시간정보를 확인한다면 용의자가 USB를 이용한 AWS 행위를 파악할 수 있다.

앞서 설명했던 AWS EC2 로그인 관련 데이터 확보 시에는 용의자 서버에 직접적인 접근을 통해 기존의 디지털 포렌식 상세 분석절차 수행이 가능하다.

〔그림 8〕 AWS Private Key를 이용한 EC2 가상머신 패스워드 생성



URL	방문시간
http://gom.dragsearch.com/DragSearchTop/Common/ad.php?type=tooltip&code=Gom	2011-03-17 11:46:48
https://manage.rackspacecloud.com/Login.do?jsessionid=C71E9CF4EB41E3367646DDB0CAA6E1D4.manage-n01	2011-03-17 11:42:18
http://c469445.r45.cf2.rackcdn.com/Live%20Memory%20Forensics%20of%20Mobile%20phones.pdf	2011-03-17 11:47:26
https://manage.rackspacecloud.com/Home.do	2011-03-17 11:42:47
https://manage.rackspacecloud.com/CloudServers/Overview.do?cloudServerId=684783	2011-03-17 11:43:39
https://manage.rackspacecloud.com/CloudServers/ServerList.do	2011-03-17 11:43:06
https://manage.rackspacecloud.com/CloudServers/Domains.do?cloudServerId=684783	2011-03-17 11:43:51
https://manage.rackspacecloud.com/CloudServers/Diagnostics.do?cloudServerId=684783	2011-03-17 11:44:30
https://manage.rackspacecloud.com/CloudServers/ViewBackups.do?cloudServerId=684783	2011-03-17 11:44:12
http://gom.dragsearch.com/DragSearchTop/Gom/core.php?code=Gom&uid=0022153FCD50&home=about%3ablank	2011-03-17 11:46:48
https://manage.rackspacecloud.com/com.arkasoft.filemanager.Filemanager/cloudFilesDownload?container=test&object=An Android Application Sandbox System for Suspicious Software Detection.pdf	2011-03-17 11:56:08
http://gom.dragsearch.com/DragSearchTop/Gom/core.php?code=Gom&uid=0022153FCD50&home=about%3ablank	2011-03-17 11:46:48
https://manage.rackspacecloud.com/CloudFiles.do	2011-03-17 11:54:25

〔그림 9〕 Rackspace 웹 히스토리 정보

5.3 Rackspace 분석

마찬가지로 클라우드 컴퓨팅 서비스 환경에서의 디지털 포렌식 조사 절차에 따라 IaaS 유형의 Rackspace를 분석한다.

Rackspace는 텍사스, 샌안토니오에 위치한 IT 호스팅 업체로써 IaaS 모델 클라우드 컴퓨팅 서비스 중 AWS에 이어 시장점유율 2위를 기록하고 있는 서비스이다.

AWS와 마찬가지로 IaaS 모델의 특징인 Server Instance(Virtual Machine)를 생성하고 생성된 Server Instance에 Shell 또는 원격으로 접속하는 형태를 가진다. 생성된 이미지는 Cloud Files라는 Rackspace에서 제공하는 웹 하드 형태의 서비스에 자동 저장되는 특징을 갖는다. 또한, AWS와 마찬가지로 CDN 서비스를 제공하여 특정 데이터에 대해 URL을 통해 어디서든 열람이 가능한 형태를 지원한다.

5.3.1 클라이언트에서의 Rackspace 사용 흔적 확인

대부분의 클라우드 컴퓨팅 서비스와 마찬가지로 Rackspace 역시 웹 브라우저를 통해 해당 서비스에 접속한다. 따라서 웹 브라우저 히스토리 정보를 통해 해당 클라우드 서비스 접속기록을 확인한다.

5.3.2 Rackspace 흔적 데이터 수집 및 분석

해당 웹브라우저 URL 내의 Cloud Server ID는 서비스 제공업체에 로그인시 요청하는 데이터로 이용되기 때문에 서비스에 대한 방문기록을 통해 사용자의 로그인 시간과 로그아웃 시간을 확인할 수 있다. 마찬가지로 CDN에 등록한 데이터를 URL을 통해 접속 시 브라우저 히스토리 정보를 통해 해당 파일의 열람 흔적을 확인할 수 있다. 또한, Cloud Files에서 데이터 다운로드 시 웹 브라우저 다운로더를 사용하기 때문에 웹 브라우저 다운로드 목록을 통해서 Rackspace를 통해 다운로드한 데이터의 내역을 확인할 수 있다.

원격 접속 흔적의 경우 원격 데스크톱 연결이나 〔그림 7〕의 내용과 같이 레지스트리 정보를 통해 해당 주소를 확인 가능하다. 또한, IP2LOCATION 사이트에서 제공하는 서비스를 이용하여 ISP 정보와 도메인 정보를 확인하고 이 정보들을 통해서 Rackspace 서버인 것을 확인할 수 있다.

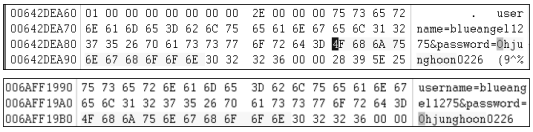
5.3.3 로그인 정보 획득과 서비스 접속을 통한 관련 데이터 확인

Rackspace의 클라우드 서버 IP와 계정 패스워드는 메일을 통해 전달되기 때문에, 시스템 조사 시 메일 계정을 획득한다면 Rackspace 서비스의 직접적인 접근이 가능하다. 또한, 클라이언트 시스템의 메모리 덤프 데이터를 이용해 메모리 내의 username이라는 시그니처를 갖는 Rackspace 아이디 정보와 password라는 시그니처를 갖는 패스워드 정보를 통해 Rackspace의 계정 데이터를 확인할 수 있다.

로그인 관련 데이터 확보 시에는 용의자 서버에 직접적인 접근을 통해 기존의 디지털 포렌식 상세 분석 절차를 수행할 수 있다.

5.3.4 Rackspace 가상머신 관점의 분석

Rackspace의 특징적인 사항은 서버 이미지가 Cloud Files에 자동 저장되고 해당 이미지를 클라이언트 시스템에서 다운로드가 가능하다는 부분이다. 해당 이미지는 .tar.gz.0의 형태로 다운로드 되며 파일명 뒷 부분의 .0을 제거한 후 압축 해제가 가능하다.



〔그림 10〕 메모리 내의 Rackspace 로그인 관련 데이터

Rackspace의 서버 이미지 파일 포맷은 마이크로소프트사의 가상하드 디스크 이미지 파일인 Virtual Hard Disk(.vhd) 형태로써 윈도우7에서 기본적으로 지원하는 가상 하드디스크 포맷이기 때문에 윈도우 7 운영체제에서 직접 마운트하면 해당 이미지의 분석이 가능하다.

또한 Rackspace 사용자가 해당 이미지를 Cloud Files에서 다운로드 받아 클라이언트 시스템에서 사용한 후 이미지를 삭제했을 경우, VHD 파일의 카빙 진행을 통해 해당 이미지의 복구가 가능하다.

VI. 결론 및 향후 연구방향

네트워크 통신의 고속화와 가상화 기술이 발전함에 따라 다수의 고객에게 높은 수준의 확장성을 제공하는 클라우드 컴퓨팅 서비스의 이용이 개인은 물론 기업 환경에서도 급증하고 있다.

클라우드 컴퓨팅 서비스는 컴퓨팅 환경의 가상화라는 특성으로 인해 사용자가 사용하는 리소스를 물리적으로 접근할 수 없는 문제가 존재하기 때문에 기존의 디지털 포렌식 기술을 활용하는데는 한계가 존재한다. 따라서 디지털 포렌식 관점에서 클라우드 컴퓨팅 서비스에 대한 이해와 조사 방법 그리고 이를 뒷받침할 수 있는 법률적 연구가 필요하다.

본 논문에서는 클라우드 컴퓨팅 서비스의 특성과 디지털 포렌식 관점의 문제점에 대해 설명하고, 디지털 포렌식 관점의 접근 절차를 제시했다. 또한, 대표적인 클라우드 컴퓨팅 서비스 IaaS 형태 중 높은 점유율을 보이는 AWS와 Rackspace에 대해 분석방안을 제시함으로써 클라우드 컴퓨팅 서비스에 대한 디지털 포렌식 조사 방법을 구체화하였다.

향후에는 IaaS에서 사용되는 다양한 가상머신 이미지에 대한 조사 방법과 IaaS 환경의 디지털 포렌식 수사에서 중요한 증거데이터로 활용될 수 있는 Snapshot 데이터에 대한 분석을 진행할 계획이다.

참고문헌

〔1〕 Gartner, "Gartner Says Worldwide Cloud Services Market to Surpass $68 Billion in 2010," Gartner Newsroom, June 2010.

〔2〕 Peter Mell and Timothy Grance, "The NIST Definition of Cloud Computing," NIST Special Publication 800-145(Draft), pp. 1-3 , September 2011.

〔3〕 Mark Taylor, John Haggerty, David Gresty, and David Lamb, "Forensic Investigation of Cloud Computing Systems," Network Security, March 2011.

〔4〕 Dominik Birk, Technical Challenges of Forensic Investigations in Cloud Computing Environments," Workshop on Cryptography and Security in Clouds, January 2011.

〔5〕 Jack of all Clouds, "State of the Cloud," Jack of all Clouds, January 2011.

〔6〕 Faith Shimba, "Cloud Computing : Strategies for Cloud Computing Adoption," Dublin Institute of Technology, pp. 1-117, September 2010.

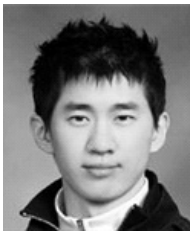
〔7〕 Cyril Onwubiko, "Security Issues to Cloud Computing," Computer Communications and Networks, pp. 271-288 , July 2009.

〔8〕 고려대학교 디지털포렌식연구센터, "디지털 증거처리 가이드라인", 고려대학교 디지털포렌식연구센터, http://forensic.korea.ac.kr/dfrc/sub_guideline/download/guideline_1.pdf.

〔9〕 Junghoon Oh, Seungbong Lee, and Sangjin Lee, "Advanced evidence collection and analysis of web browser activity," DFRWS, pp. 62-70 , August 2011.

〈著者紹介〉

정 일 훈 (Il-Hoon Jeong)
2010년 2월: 한세대학교 정보통신 공학과 졸업
2010년 3월~현재: 고려대 정보보호대학원 석사과정
〈관심분야〉 디지털 포렌식, 정보보호



오 정 훈 (Jung-Hoon Oh)
2010년 2월: 동국대학교 컴퓨터 공학과 졸업
2010년 3월~현재: 고려대 정보보호대학원 석사과정
〈관심분야〉 디지털 포렌식, 모바일 포렌식

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

이 상 진 (Sang-jin Lee)
1987년 2월: 고려대학교 수학과 졸업
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 교수
〈관심분야〉 디지털 포렌식, 모바일 포렌식, 심층 암호, 해쉬 함수