IST-Africa

IST-Africa 2012 Conference Proceedings
Paul Cunningham and Miriam Cunningham (Eds)
IIMC International Information Management Corporation, 2012
ISBN: 978-1-905824-34-2

Digital Forensic Framework for a Cloud Environment

George SIBIYA[1], Hein S. VENTER[2] and Thomas FOGWILL[1]
[1]*CSIR, 627, Meiring Naude Road, Brummeria, Pretoria, 0001, South Africa*
Tel: +27 12 841 3976, Fax: + 27 86 587 2208,
Email: gsibiya@csir.co.za, tfogwill@csir.co.za
[2]*University of Pretoria, Lynwood road, Pretoria, 0002, South Africa*
Tel: +27 12 420 3654, Fax: +27 12 362 5188, Email: hventer@cs.up.ac.za

Abstract: The advent of cloud computing provides good opportunities for both good and malicious use. Cloud computing is at its infancy stage and its security is still an open research issue. Malicious users take advantage of the current lack of advanced security mechanisms in the cloud. Cloud computing paradigm enables users to access computing resources without necessarily owning physical infrastructures. It is therefore easy for an attacker who intends to perform malicious activities in the cloud to create a remotely hosted desktop, perform their activities and then destroy the desktop later. With the remotely hosted desktop destroyed, there is very little evidence left that can be collected by forensic experts using traditional static digital forensic methods. A scenario such as this therefore requires live digital forensic processes as a large amount of evidence can be gathered while the system is running. Key issues in cloud forensics include, but are not limited to, identity, encryption, and jurisdiction and data distribution. Digital forensic investigators currently face a challenge when criminal incidences occur as there are no well developed tools and procedures for conducting digital forensic investigations in the cloud. This paper proposes a novel framework that addresses issues of digital forensics in the cloud computing environment.

Keywords: Cloud Computing, Digital Forensics and Security

1. Introduction

The concept of virtualisation in computing involves operating systems running on another operating system as if they were running on their own hardware [12]. Virtualization provided grounds for the birth of cloud computing [10]. Such developments in computing paradigms present more opportunities for cyber crimes.

These developments therefore also present new challenges to law enforcement agencies. Research efforts were at an advanced stage in addressing issues of digital forensics for traditional computing paradigms including virtual environments but these solutions may not be directly applicable in the cloud [4]. User data in a cloud environment is distributed and often resides beyond the jurisdiction of forensics investigators.

In Section 2, a brief background on digital forensics and cloud computing is presented. In Section 3, digital forensic challenges presented by the cloud paradigm are discussed. In Section 4, the authors present a framework that addresses the issues of digital forensics in a cloud environment. Section 5 presents the current environmental set-up of our proposed framework. Section 6 concludes the paper and also presents planned future work.

2. Background

In this section, the authors present background concepts on cloud computing and digital forensics.

2.1 Cloud Computing

Cloud computing can be defined as highly scalable computing resources provided as an external service via the Internet on a pay-as-you-go basis [10]. This means that service consumers in the cloud pay for services as they use them. A cloud model offers a solution to resource constrained small, micro and medium enterprises (SMMEs) in Africa. The cloud can be deployed in four different forms, i.e., private cloud, community cloud, public cloud and hybrid cloud [8]. A private cloud is deployed to be used within an organization and the entire cloud infrastructure including hardware resources are owned by the organization. A community cloud is a cloud shared by organizations with a common business interest. A public cloud is a cloud deployed for the purpose of being used by public organizations regardless of their business interests. A hybrid cloud is a combination of any of the previous cloud models. Services in a cloud are grouped into three layers, i.e., cloud application, cloud platform and cloud infrastructure [10].

These layers in a cloud are offered as services, where we have software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS). One of the services offered in the cloud are Hosted Desktops [10]. A Hosted Desktop is a virtual machine hosted in the cloud. In a hosted desktop, applications and data are hosted on a remote data centre and not in a local user machine as in traditional computers. Hosted desktop owners access their applications and data though ordinary desktops or thin clients. Such a hosted desktop can be used to commit cyber crime in the cloud in the same way as a criminal can use a physical desktop. In [14], the authors define cloud crime as "any crime that involves cloud computing where the cloud can be the object, subject or tool of crimes." It is when such crimes are committed in the cloud that the services of a forensic expert will be required.

2.2 Digital Forensics

Digital forensics can be defined as a discipline that combines elements of law and computer science to collect and analyse data from computer systems, networks, wireless communications, and storage devices in a way that is admissible as evidence in a court of law [13]. According to [19], a digital forensic process can be broken into four distinct phases:

1. Collection of artefacts (both digital evidence and supporting material) that are considered of potential value are collected
2. Preservation of original artefacts in a way that is reliable, complete, accurate, and verifiable
3. Filtering analysis of artefacts for the removal or inclusion of items that are considered of value
4. Presentation phase in which evidence is presented to support investigation.

Traditionally, two categories of digital forensics existed i.e., static digital and live forensics, and in [19] the authors argue that the two categories existed as a result of forensic evolution to recreate and document sophisticated incidences. Static forensics involves analysis of static data such as hard drives obtained using traditional formalized acquisition procedures. Live forensics involves the analysis of the system memory and any other relevant data while the system being analysed is running. More digital forensic challenges in the cloud are presented in the next section.

3. Cloud Forensics Challenges

Both live forensic and static forensic approaches face challenges in the cloud. The static forensic process involves, for example shutting down the system so that the hard disk can be cloned. This cannot be carried out in the cloud as a number of virtual machines share the same physical infrastructure. Shutting down the host machines disrupts all the hosted virtual machines wherein some of them may be running mission critical systems. Live forensics, on the other hand, only involves taking snapshots of running virtual machines and crime scenes cannot be recreated as in the case of static forensics. Digital forensic procedures could be performed easily in traditional settings where data storage centres are within physical reach.

Digital forensics has to catch up with the continuous changes in computing paradigms. A cloud model poses challenges to digital forensics as information is difficult to locate, acquisition is impossible if it cannot be located, and there would be no analysis without acquisition. Challenges relating to digital forensics that hinder data acquisition in the cloud include distributed storage, protected data, identity, etc. [18]. In a cloud environment, data is likely to be partitioned and stored in distributed systems usually spanning different jurisdictions; hence, it would be extremely difficult to locate. Cloud users may also store their data in the cloud in an encrypted format such that even if data is located and acquisitions are performed successfully, it cannot be useful to the digital investigator. This is still a challenge in traditional settings but it is worse in the cloud as data is encrypted further by the storage service provider over another encryption layer. Obtaining decryption keys from the data owner will not be sufficient. Other further decryption keys need to be obtained from the cloud storage service provider. In the cloud, there is also an issue of identity, wherein it is difficult to associate a cloud user with data stored in the cloud. In traditional settings, a physical machine owner is by default assumed to be the owner of all data stored in that machine. In a cloud environment, data is stored in remote locations and accessed using thin clients. It is a challenge in a cloud environment to single out a user from a large number of cloud users distributed globally and assume them to be the owner of the data. Cloud users may also use aliases as owners of data and not their real names.

According to Birk in [4] there are three sources from which evidence can be extracted in a cloud, i.e., the client side, the network layer and the cloud service provider (CSP). Of the three sources the most difficult to gather evidence from is the cloud service provider side. What makes it difficult on the cloud provider side is that the provider is usually outside the jurisdiction of the investigators. International laws and international collaborations have to be taken into consideration, which may be costly and time consuming. Gathering evidence from the client and the network layer, existing tools and digital forensic frameworks can still be used for both static and live digital evidence acquisition. Data acquisition from the client and the network layer is easier especially if the suspect is within the same organization that may be requiring the services of a forensic investigator. If the victim is from outside the organization that hosts the suspect, the organization may be reluctant to release their data or allow investigators to monitor their networks. Gathering evidence from the service provider side can only be easier in private clouds.

Due to the infancy of the cloud there are few research contributions on the state of the tools and procedures that can be used to acquire evidence from the cloud. We share the same sentiment with Zimmerman and Glavach in [19] on their argument that new forensic tools in the cloud need to be offered as services. As acquisition is a challenge in the cloud, they argue that new forensic tools must be able to visualize data locations. In the visualized data, data that can be obtained and data that cannot be obtained needs to be tagged as such and be presented as such in court proceedings. Other solutions include digital forensics as a cloud service [9].

Marty in [11] presents a logging framework and guidelines that provide a proactive approach to logging to ensure that the data needed for forensic investigations has been generated and collected. The standardized framework eliminates the need for logging stake-holders to reinvent their own standards. This framework is a step forward in the right direction towards addressing digital forensic issues in the cloud. The approach though merely focuses on the logs which give details on past events. The framework overlooks critical and volatile evidence that can be obtained while the system is running. Table 1 gives a summary of digital forensics challenges in cloud computing.

Table 1: Digital forensic challenges in the cloud

CHALLENGE	**DESCRIPTION**
Identity	It is hard to link data stored in the cloud to an individual cloud user
Encryption	Data encrypted by the client before sending it for storage and further by the cloud service provider before storing it
Jurisdiction	Accessing data stored in computers beyond local borders may violate laws in other countries
Distribution	A cloud user may distribute data in several countries hence collaborating with each of these countries may be costly

4. Proposed Framework

In this section, the authors propose and present a Live Digital Forensic Framework for a Cloud (LDF2C) environment. The proposed framework addresses issues such as:

1. What live forensic technique can be used to collect forensically sound evidence in a cloud environment?
2. How can data stored in the cloud be associated with an individual user?

In Section 4.1 typical criminal scenarios in the cloud are presented. Section 4.2 presents basic requirements for a digital forensic solution in a cloud. Section 4.3 presents features of LDF2C and Section 4.4 presents LDF2C architecture.

4.1 - Digital Forensic Service Application Scenario

A virtual machine hosted in a cloud environment can be used to commit crime as it used to be done with physical computers. Cyber crimes that can be committed in the cloud include unauthorized access to resources in the cloud, money laundering, distributed denial of services attacks, storage of pirated software, music, movies, etc. In this section, two scenarios of criminal activities carried out in the cloud are considered, i.e., DDoS attacks and unauthorized file sharing.

A DDoS attack in a cloud can be performed a guest virtual machine could take control of the physical host infrastructure [7]. Having taken control of the host machine, the attacking virtual machine will then exploit computing resources, hence limiting usage by other guest virtual machines.

In the same way as a criminal uses the bot-net to deny services to clients, the bot-net can be used to share and distribute pirated software, movies, child pornography, etc. As the bot-herder takes control of an infected host machine, they may use it to store pirated files.

For all criminal activities that are carried out in the cloud, the criminal has to send commands over the Internet. One of the challenges in the cloud is accessing data hosted in remote cloud service provider outside the jurisdiction of the investigators. The requirement

to communicate with a remote desktop over the Internet provides an opportunity for investigators to gather evidence. Section 4.2 presents requirements for a framework that exploits this opportunity in addressing digital forensic issues in the cloud

4.2 - Requirements for a Digital Forensics Framework in a Cloud environment

In the light of the raised digital forensic issues in this paper, we propose that approaches that address digital forensic issues in the cloud need to have the following features:

1. Live forensics - In virtual machines in cloud environments, valuable evidence can only be extracted while the system is running. Data that can be extracted include memory dumps and cached Internet files. Once the virtual machine is shut down such data is lost.
2. User and data association - In a cloud environment user data is hosted remotely. Even if data is discovered, it is still a challenge to associate the data with a suspect.
3. Log data mining - based on available information, the framework need to be able to extract data from log files from locations of interest.
4. Intelligence - Also based on the available information, the framework needs to be able to apply intelligence to infer relationships between data and a suspected user.

4.3 - LDF2C features

LDF2C meets the live forensics, user and data association, log data mining and intelligence requirements raised in Section 4.2 as follows:

1. Deployed as a service.
   LDF2C monitors live network traffic and activities on an identified live virtual machine. If monitoring would be carried out on a full-time basis in an enterprise, the need for storage would grow exponentially with time. Deploying LDF2C as a service allows it to be discovered and utilized when needed. Calling the service when a need arises mitigates the need for extra storage in an enterprise.
2. Uses data-mining techniques.
   When the framework is invoked, it utilizes the data-mining techniques to retrieve information from log files in locations which were determined to have been accessed by the suspect. This also enables the framework to utilize storage resources in the cloud due to a rise in demand for storage as the investigation progresses.
3. It is based at network level.
   Interaction between a cloud user and any cloud resource is always via a network. This provides an opportunity for an investigator to monitor communication between a client and any cloud service. The network traffic source and destination are then used to locate user data in the cloud and also to associate data in the cloud with a user. Preliminary evidence can therefore be gathered before acquiring data form a cloud service provider who often may be beyond the jurisdiction of an investigator.
4. Employs artificial intelligence techniques.
   In the process of associating a user with data in the cloud this framework employs semantic reasoning. Cloud users use light weight applications installed in thin clients or ordinary desktops to access data in the cloud. The semantic reasoning process utilizes the domain knowledge to associate tools installed in those clients to the activities taking place at their remote desktop [17].

4.4 - LDF2C Architecture

LDF2C is implemented on a dedicated virtual server as shown in Figure 1. When invoked, LDF2C service performs forensic analysis between two virtual organizations. The LDF2C system running on a virtual server is divided into three layers, i.e., the back-end, middle

layer and the front-end. Below, the authors give a brief description of each component within the respective layers.

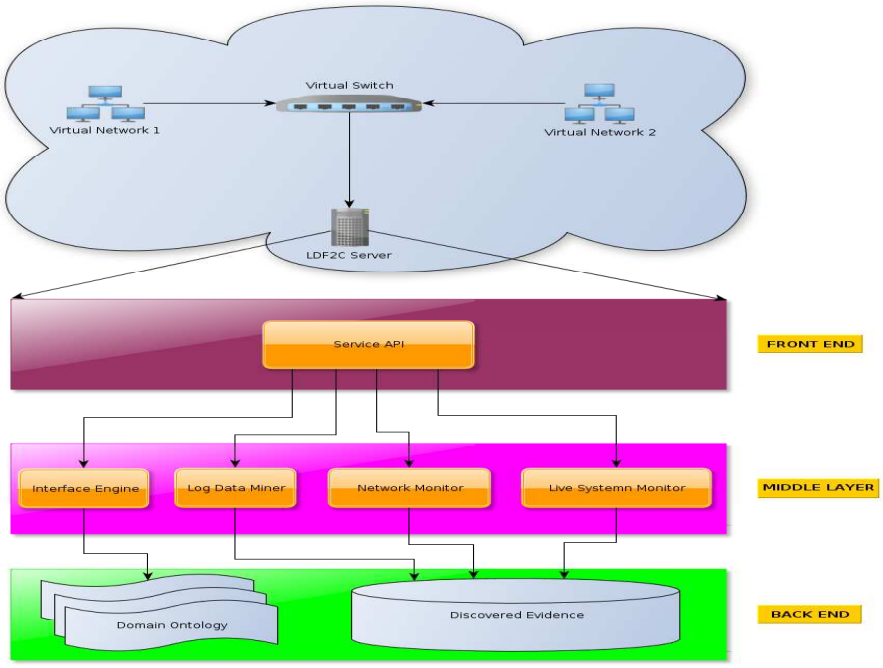

Figure 1: LDF2C Conceptual Architecture.

Back End

This layer manages persistent storage of data discovered by the forensic service. Data managed at this layer are log files from locations determined to be relevant by the service and domain ontologies. These ontologies are used by the inference engine to reason on association between data stored in the cloud and an individual user. This layer also consists of a database management system (DBMS), which stores data that can be stored in a relational database. The DBMS is used by both the network monitor and the live remote system monitor components in the middle layer. Lastly, the layer consists repositories that manage snapshots and memory dumps captured by the Live remote system monitor.

Middle Layer

The middle layer consists of four components, i.e., Inference Engine, Log Data Miner, Network Monitor and the Remote Live System Monitor. The Inference Engine uses data discovered in the client device used to access the cloud and remote desktop to determine the relationship. Examples of such data are software installed in the client cloud accessing device and the activities that take place in the remote desktop or victim. The Log Data Miner will implement an algorithm that is used to discover log files from locations which have been determined to be relevant such as the hosted remote desktop and the accessing client device. The Network Monitor intercepts communication between the host of a suspected malicious user and their remote desktop or between the malicious host and the victim. The Remote Live System Monitor component monitors and caches activities in a remote virtual machine. It captures information such as system snapshots and memory dumps.

Front End

The front end consists of only one component i.e., the Service API. This component exposes the functionalities of the forensic service to external clients. The service is manually triggered by the forensic investigators using the API.

The next section presents environmental set-up and the experimental procedures that will be used when evaluating the various aspects of LDF2C.

5. LDF2C environmental set-up

For preliminary testing, LDF2C is deployed in the private cloud. The environment is set up using two Desktop PC's, each running Ubuntu 11.04. In this scenario, open source cloud manager, OpenNebula [5] is used. One of the hosts runs OpenNebula which hosts virtual machines. Three virtual machines are deployed in OpenNebula where one virtual machine is used to launch attacks and one of the Virtual machines is used to monitor communications between the attacker and the victim virtual machine. The attacker uses the second physical host to access their virtual machine hosted in OpenNebula. Currently the monitoring virtual machine uses WireShark and nmap [2] to monitor activities occurring in the victim machine.

5.1 - Network monitoring tests

This test will be conducted to determine the ability of the service to isolate suspected packets and to determine the source and the destination of these packets. This is important because the destination may be a virtual machine, which differs from traditional settings where sources and destinations are physical machines. Physical machines connect directly to a physical network. A virtual machine on the other hand connects via a virtual network bridge.

5.2- Log data mining test

After the service has successfully determined the source and destination of the packets, the service needs to remotely install digital forensic tools in those locations. These tools implement the data mining algorithm that discovers log files in those locations. This test will be conducted to determine latency introduced by these tools in these remote machines. This is important because if the tools introduce latency in these locations, it may raise suspicions to the perpetrator that they are being monitored.

5.3 - Remote system monitoring test

After the log information on the remote systems have been gathered, live system activities in the remote machines need to be captured. Captured information such as memory dumps need to be sent periodically to the central digital forensics server. This test therefore needs to be conducted to determine network traffic congestion, if any, that may be introduced by the the forensic data transmission.

5.4 - User identification test

This test will be conducted to determine the ability of the semantic matchmaker to compute association between a cloud user to data or remote desktop in the cloud. By varying the number of virtual guest machines to match a client against, scalability will be measured.

6. Conclusion and Future Work

This paper presented LDF2C, a framework aimed at addressing digital forensics challenges in a cloud environment. The framework addresses the issue of data acquisition in the cloud that may be beyond the jurisdiction of investigators. It makes use of accessible information to build up a case before the costly data acquisition from foreign countries could be carried out. It also addresses the issue of identity in the cloud where it is difficult to associate data in the cloud, with a cloud user. Currently LDF2C environment is set-up as explained in Section 4. The virtual machine that monitors the network uses an existing network monitoring tool, WireShark [1]. The next step in this research involves development of an algorithm that will extract log information from accessible locations in the crime scene. The domain knowledge representation that is used in reasoning while associating an attacker with data in the cloud is also going to be developed. The last step that will be carried out in this research is the development of a guideline that will be used to validate evidence collected using LDF2C service.

References

[1] Wireshark: The world's foremost network protocol analyser, [Online] Available at: http://www.wireshark.org/ [Accessed 30 November 2011].

[2] http://nmap.org/, [Accessed 10 December 2011].

[3] https://cloudsecurityalliance.org/, [Accessed 20 November 2011].

[4] D. Birk, Technical challenges of forensic investigations in cloud computing environments. [Online] Available at: http://www.zurich.ibm.com/~cca/csc2011/submissions/ [Accessed 15 November 2011].

[5] C. V. Blanco, The opennebula virtual infrastructure engine, [Online] Available at: http://www.xen.org/files/xensummit_germany09/OpenNebula.pdf [Accessed 20 November 2011].

[6] R. Buyya, S. Pandey, and C. Vecchiola, Cloudbus toolkit for market-oriented cloud computing. 2009. CloudCom, Springer-Verlag Berlin Heidelberg, pp 24–44.

[7] S. Castro, Virtual machine trojan: A new type of threat. [Online] Available at: http://www.infosegura.net/VMTthreat.html [Accessed 15 November 2011].

[8] E. W. Hobson, Qinetiq white paper: Digital investigations in the cloud, QinetiQ Digital Investigations Service, Farnborough, UK, 2010.

[9] R, Ieong, Challenges to digital forensics from cloud computing, [Online] Available at: http://www.china-forensic.com/downloads/2011 [Accessed 15 November 2011].

[10] R. Lovell, White paper: Introduction to cloud computing. [Online] Available at: www.thinkgrid.com/docs/computing-whitepaper.pdf - United States [Accessed 20 November 2011].

[11] R. Marty, Cloud application logging for forensics. Taichung, Taiwan, March 2011. SAC, ACM.

[12] M.A. Monroe, *Virtualization and Cloud Computing Draft Version.* Digital Reality Trust, drivers of data growth, part 2 edition, 2010.

[13] I. Resendez, P. Martinez, and J. Abraham, An introduction to digital forensics. [Online] Available at: http://acetweb.org/journal/ACETJournal_Vol6/ [Accessed 01 November 2011].

[14] K. Ruan, J. Carthy, T. Kechadi, and M. Crosbie, Cloud forensics: An overview. [Online] Available at: http://cloudforensicsresearch.org/publication/ [Accessed 10 November 2011].

[15] C. A. Schiller, J. Binkley, D. Harley, G. Evron, T. Bradley, C. Willems, and M. Cross. *Botnets: The killer web application*. Syngress, Inc Publishing, 2007.

[16] J. Shende. Digital forensic challenges within cloud computing. [Online] Available at: http://jonshende.blogspot.com/2010/10/digital-forensic-challenges-within.html [30 October 2011].

[17] T. Stallard and K. Levitt, Automated analysis for digital forensic science: Semantic integrity checking, University of California, One Shield Avenue, Davis, CA, 95616 USA, 2003.

[18] I.Walden, Law enforcement access in a cloud environment, Queen Mary University of Londin, School of Law , 2011.

[19] S. Zimmerman and D. Glavach. Cyber forensics in the cloud, IANewsletter, Volume 14 Number 1, Winter 2011, p 4.