Cloud Computing Architecture and Forensic Investigation Challenges

Ghania Al Sadi
Sohar University, Computing Department
Sohar, University Rd, 311
Sultanate of Oman

ABSTRACT
Contrasting to traditional digital forensic investigations, investigating cloud crimes is considered as more difficult and complex process. The architecture of cloud computing is behind the complexity of conducting forensic investigation on cloud where data are synchronized and accessed using multiple and different devices in different places that reduce the chance to find a real device to seize for forensics investigation. There are a number of challenges in forensic investigation field faced by investigators which may complicate the way of conducting cloud forensic investigations to extract evidences. This research is studying the complexity of cloud architecture and how it affects digital investigations by addressing a number of challenges on conducting cloud forensic investigation.

General Terms
Cloud Computing, Cloud Forensic, Forensic Challenges

Keywords
Cloud computing, cloud criminals, digital forensic, forensic challenges, virtualization, synchronization

1. INTRODUCTION
As defined by NIST, cloud computing refers to a model of providing available, convenient, on demand access to a shared pool of resources configured on network like servers, storage, applications and other type of services that can be released and managed by less efforts (Mell and Grance 2011). Cloud Computing utilizes both hardware and software stored on provider's datacenters to be delivered as services over internet for users (Huo et al. 2011). Moreover, cloud computing can be defined as a distributed computing model with large scale which contains a pool of virtualized and scalable computing resources on its infrastructure that is delivered on demand for users (Foster et al. 2008).

Cloud computing has a number of characteristics that make it more flexible to use when compared to traditional computing services. Cloud computing provide on demand services where users can request for a service based on their needs and pay as they use without the need to an actual interaction with the cloud service provider. A high number of users who use multi-tenant model can access data on a shared pool of resources provided by cloud provider based on their demand. Moreover, cloud computing has the capability to access data over internet using different type of computing devices like PCs, tablets, mobile phones or workstations (Mell and Grance 2011).

2. CLOUD DEPLOYMENT AND SERVICE MODELS
Cloud computing introduced a number of service models to meet all types of customer requirements. The available service models are Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). In SaaS model, the cloud enables users to access applications running on a cloud infrastructure via web browsers or client applications. In IaaS model, cloud provide users with basic computing resources like processing, storage and network to run and deploy arbitrary software which may consists of operating systems and applications. Users can get virtual servers hosted by cloud providers and they pay for the resources they use only. In this model, users have no control over cloud resource but may be provided with limited privileges to control some network components like firewalls. On the other hand, PaaS model enable users to deploy their application on cloud created by supported programming languages, libraries and tools. Also, it provides a service that contains a complete set of software development lifecycle management. Users utilizing PaaS can control and configure only the deployed application on cloud infrastructure (Mell and Grance 2011). One more cloud service model is Storage as a Service (StaaS) that has been grown to accommodate the capability to store data in cloud that is accessed over a wide range of devices. StaaS is owned and managed by cloud provider and provided as a service that is accessed through web based applications or Application Program Interfaces (APIs) like desktop storage applications (Martini and Choo 2013). Cloud services can be deployed over a number of introduced cloud deployment models that can be public, private, community or hybrid cloud models to accommodate customer requirements (Mell and Grance 2011).

3. CLOUD COMPUTING ARCHITECTURE
Cloud computing employs virtualization technology on its infrastructure along with other computing services and resources provided over internet. Many and different virtual machines are hosted on cloud servers that are monitored and controlled by hypervisors (Sabahi 2011). By using virtualization in cloud servers, data can be highly available for users anytime they request. Moreover, virtualization helps to reduce maintenance costs of cloud servers by improving resource utilization. Also, it provides a quick way of disaster recovery (Maguire 2013).

Virtualization is considered as a main layer in cloud infrastructure that sets between physical layer and abstraction layer as shown in Figure 1. It isolates hardware from software to simplify the way of reassigning application on servers based on user's demands. It is managed and controlled by

hypervisor. Above virtualization layer, abstraction layer is set up that consist of software resources. Abstraction layer provide and manage user interface to access and use cloud services and thus reduce the complexity of cloud infrastructure which is hidden from end users (Maguire 2013). Users can use web services like web browsers, web-based applications or APIs to access cloud services that are designed and managed by cloud providers.

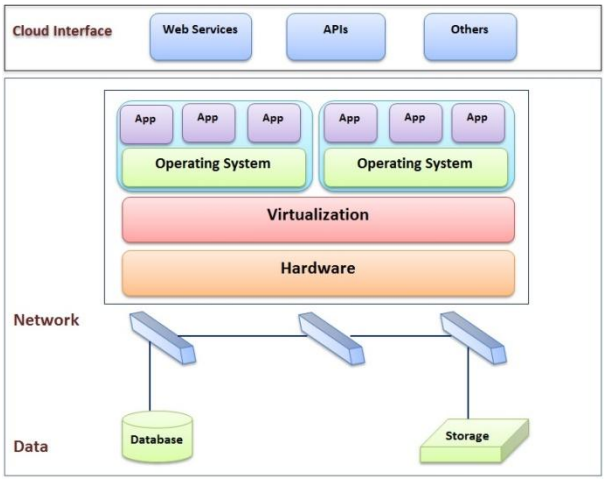

Figure 1: Cloud Computing Architecture

Cloud provides a repository for data considered as Cloud Storage that is managed and maintained to store users' data and provide it over internet. A high amount of data and resources are stored on public cloud providers' servers that can be accessed by a high number of users around the world. This feature raises the security and privacy issues on cloud environment while data are stored on a shared pool of resources. Cloud data are encrypted while stored on cloud systems to ensure security but separating each user data is not managed in cloud where data are stored on a shared pool system. On the other hand, cloud storage system use different techniques to ensure authorization of accessing data in cloud shared pool system that will protect user data from being leaked by unauthorized access (Mulazzani et al. 2011). Although the highly distribution of Cloud Computing and its development to eliminate security and privacy issues, it still seems to be susceptible and vulnerable by attacks while data are logically stored on remote servers with invisible structure. Cloud users have no rights to technically manage their data on cloud storage in such way used to manage physical storage disks. In criminal cases, traditional forensic techniques can be applied to investigate cloud criminals but the accuracy of evidence cannot be maintained due to the absence of physical storage devices.

4. CLOUD COMPUTING FORENSIC

As defined by NIST, digital forensic is "the identification, collection, examination, and analysis of data while preserving the integrity of information and maintaining the strict chain of custody for data" (Kent et al. 2006). Therefore, cloud forensic can be defined as a process of applying digital forensic on cloud environment. Figure 2 shows a number of interrelated steps that must be followed respectively to go thorough forensic investigation process.

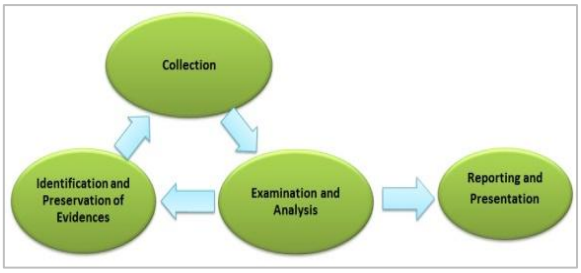

Figure 2: Digital Forensic Investigation Process

While cloud is based on network access, then cloud forensic is considered as network forensic where forensic investigation can be conducted on both public and private networks. As well as traditional forensic, cloud forensic may require investigating memory processing, registry files, network logs, histories and file systems to collect evidences (Haggerty 2013). However, finding a suitable forensic technique for cloud is essential to extract accurate evidence from cloud environment that must be accepted by law enforcement agencies.

Cloud forensic is connected to legal, technical and organizational implications that are depending on each other to complete investigation process. Legal aspect is responsible for regulations that monitor forensic activities and ensure that will not breach the law during investigation. On the other hand, technical aspect is required to provide the proper forensic toolkit to conduct the investigation on cloud environment while organizational aspect defines the internal structure of the organization and the external assistant involved in a specific incident (Marturana et al. 2012).

Investigating cloud criminals vary depending on cloud model either deployment or service models. For instance, the accurate application logs can be obtained from Cloud Service Provider while investigating SaaS. On the other hand, collecting evidences from IaaS is more reasonable by acquiring virtual machines image from customer device because users of this model can get virtual servers and control over some network components. Furthermore, accessing private cloud servers for forensic investigation is available for investigators while it is not possible in public cloud.

Technically, conducting forensic investigation on cloud computing requires both client artefact and cloud server artefact depending on the cloud model. Regardless of accessing cloud servers, a number of potential evidences can be collected from client devices by applying network forensic techniques. For example, investigators can analyze network logs and client's browser to obtain evidence. Depending on each cloud model, the ability to conduct investigation varies based on the available area to search for evidence. For example, the available investigation area in private cloud is wider than area provided in public cloud. Also, the available evidence in cloud server is more than that available in client artefact. However it is still challenging to conduct forensic investigation on cloud and extract accurate evidences that must be accepted by law enforcement agencies.

5. CLOUD FORENSIC CHALLENGES

In most criminal cases, forensic investigators deal with big challenges on collecting evidences from cloud systems due to the complexity and invisibility of cloud architecture. Actually, the accurate evidence can be collected from cloud servers but unfortunately it is difficult to seize cloud servers due to privacy policies followed by cloud providers. In this regard,

most of available forensic tools are inconsistent with the nature of the cloud (Haggerty 2013). Thus, it is a big challenge for digital forensic investigators to conduct such investigation in cloud environment. A number of cloud forensic investigation challenges are discussed to cover technical and legal dimensions on cloud computing. Some of these challenges are listed below as following:

5.1 Forensic Data Collection

In traditional digital forensic, the investigator has a suspected device to seize like computers, routers and storage medium. Therefore, collecting forensic data is considered as a simple process by collecting network logs, process logs and examining file systems. On the other hand, collecting forensic data from cloud is considered as a complex process because of its invisible and complex infrastructure where in most cases it seems difficult to suspect a specific device to seize or even access cloud servers to obtain forensic evidences. Actually, forensic investigators have limited rights to collect and examine evidence from the entire cloud system that consists of cloud client and cloud server. During investigation process, examining client device is only feasible for investigators where accessing cloud servers requires high-level of privileges and permission from cloud providers. In fact, there is no access provided to public cloud servers even to investigate criminals because granting access to cloud servers may break privacy of cloud data while it is stored on a shared pool on cloud servers. Therefore, it will be a big challenge for forensic investigators to collect evidence from client devices only that miss the accurate evidence required for law enforcement agencies. Moreover, the difficulty of finding a real device to examine during investigation is one concern faced in cloud forensic field. As stated before, cloud enables users to access cloud service from anywhere via internet using any device that is synchronized with cloud account. Synchronization complicate investigation process where some types of network logs like IP address of synchronized devices will be required during investigation process to collect accurate evidences. Some cloud systems provide such evidences on client accounts but still not accurate to make a decision on the criminal. In some cases, Law enforcement agencies may issues a subpoena to cloud providers to get network logs that contains IP addresses of devices used to access a compromised cloud account (Haggerty 2013). Even though, the obtained logs will not be as accurate as needed because collecting digital evidences is best done by experts like digital forensic investigators because most attackers are considered as sophisticated attackers who clear all evidence left by criminals. So, it is still a challenging for cloud forensic investigators to collect forensic data from public cloud over the complexity of cloud infrastructure.

However, collecting forensic evidences from cloud varies depending on cloud service model where in IaaS model, users have high level of control and access to digital forensic while other models don't provide such facilities to their users (Ruan et al. 2011). In most cloud models, accurate forensic data are collected from cloud servers however investigator has no rights to access provider's datacenters to collect forensic evidences. Even though private cloud has a wide area to collect digital evidences where investigating both server and client is reasonable but it is still a challenging to collect accurate evidences. Experienced attacks can perform anti-forensic actions after attacking attempts like modifying access log files to convince forensic investigators that no illegal action has been performed on the compromised cloud account. Absolutely, it is a big challenge to conduct such forensic investigation when the cloud is compromised by internal experienced attacks with authorized access to cloud system that will make examining process more difficult to be performed. Nevertheless, Cloud providers need to meet legal regulation to deal with compliance provided by cloud customers where in criminal cases the provider will be subject to such regulation more than the client (Dam and Chen 2011).

5.2 Cloud Virtualization and Data Segregation

Considering cloud architecture, cloud provides a shared pool of resource that contains high amount of data related to a high number of users where separating users' data seems to be a difficult process. Cloud utilizes virtualization to distribute cloud resources among clients where different virtual machines are enabled to share same physical infrastructure. Technically, cloud may use one physical machine to run different client instances that are separated using virtualization (Haggerty 2013). As shown in Figure 3, one physical infrastructure is used to store all virtual machines' data where virtualization is monitored and controlled by hypervisors. Although all virtual instances are reside on one physical location, each instance in cloud has control over own virtualized disk only and it is not possible to access other instances' data or access raw disk devices. However, all users' data can be easily accessed from server side by network administrators or any authorized person since all instances are stored on one physical place.

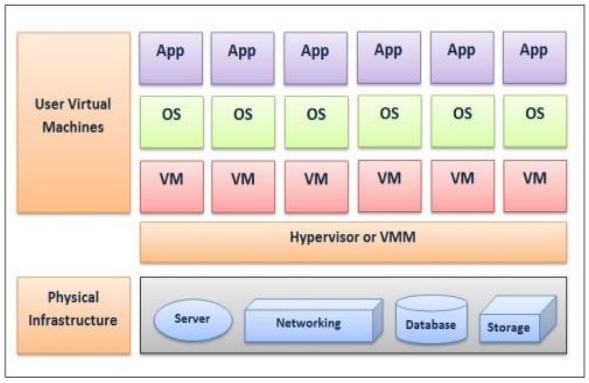

Figure 3: Cloud Physical and Virtualized Infrastructure

Figure 4 shows an example of cloud server configuration on a Linux platform. As shown, user's data can be easily displayed from server side where all data are stored in one directory on server storage and each account data is stored on a sub-directory to separate data from each other. However, during investigation, it is a challenge for investigators to separate resources of a compromised instance from others without affecting confidentiality and privacy of other instances since all instances share same physical infrastructure and stored in one location (Mulazzani et al. 2011). For example, when collecting logs as part of evidence, log files will display information related to all instances stored on that physical machine which will breach privacy of other instances. Figure 5, shows an example of access log collected from cloud web server that list all activities occurred on cloud server that have been done by all cloud users. As shown, cloud accounts and IP addresses of machines used to access cloud server are displayed in a clear text which affects users' privacy. Technically, it seems difficult to isolate instance's details from each other while are sharing same physical machine. Thus, when providing evidence to courts, investigators should prove that the provided evidence related to a suspected user.

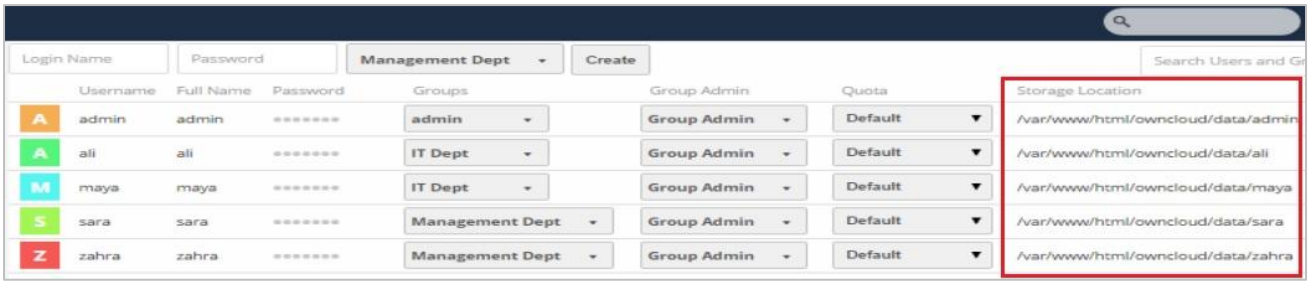

Figure 4: Example of Cloud Server Configuration on Linux

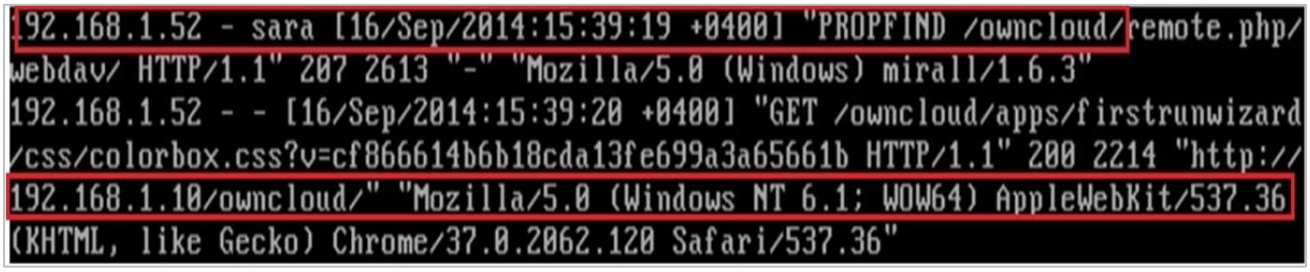

Figure 5: Example of Cloud server access log

Along with separation issues, registering to cloud account is considered as a weak procedure where anonymous accounts can be registered for cloud service to be used for criminal purposes. In complex cases, even if investigators obtain accurate data of a suspect by a successful data separation technique, it is still a challenge to prove the identity of the criminal.

As mentioned in this research, cloud computing apply replication and resource distribution to provide redundancy using virtualization. In most cybercrimes, the attackers target the hypervisor because it is responsible of monitoring the virtual instances in the cloud. Therefore, investigators need to utilize the hypervisor to access the runtime state of the virtual machine to collect most evidence (Birk and Wegener 2011). Technically, amount of evidence can be left on the hypervisor; however it is not accessible from client accounts. Thus, cloud virtualization maximizes cloud forensic investigation challenges. The main concern in this regard is the lack of forensic investigation procedures and policies over hypervisors (Ruan et al. 2011).

5.3 Data Recovery

Cloud computing is using Distributed File System (DFS) to manage processes on files which is considered as a transparent system where the user is not required to be aware of where data is actually stored while it is accessed similarly to accessing local files. Location transparency is one feature provided by DFS that provide a constant namespace where file name does not contain its location (Akarsu et al. 2013). This feature can complicate the process of recovering deleted data from cloud accounts whereas file location cannot be recovered from the file name. Generally, recovering deleted data is an essential evidence source in digital forensic, thus it is a required process in cloud forensic. In cloud, the user has a full right to create, modify, retrieve and delete own data but the right of deleting the original snapshot of data is reserved for cloud provider. In case of deleting cloud items and its attributes within the domain, the process of mapping removal in the domain is instantly completed which eliminate remote access to deleted data. Technically, the space reserved by the deleted data may be assigned to new data which will limit the ability to recover the deleted data. Based on the experimental conducted on cloud server that configured on Linux platform, each user account has a trash bin that contains temporarily deleted data. Data stored in trash bin can be accessed and restored from any synchronized device. Unfortunately, when deleting cloud files from the trash bin it can be only recovered from the device used to delete that file by using some powerful recovery tools. This feature limit the ability to recover files when deleted remotely using any other synchronized device. Moreover, files deleted from cloud server are permanently deleted and are not sent to the trash bin and cannot be recovered as illustrated in Figure 6 that show an example of the deleting files from cloud server. The figure illustrate the difficulty of recovering deleted data while it is not sent to the trash bin therefore, it cannot be restored either from client account or from the server.

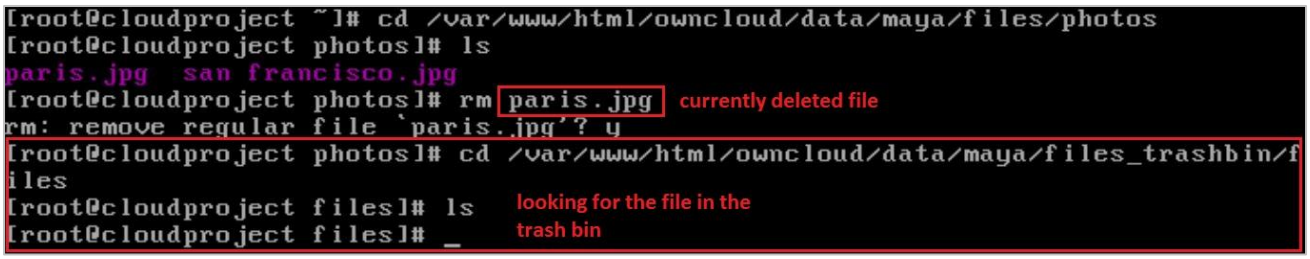

Figure 6: Example of the difficulty of recovering files deleted from cloud server

However, deleted data may still be accessible in the memory capture but it is difficult to identify the owner of these deleted data while it is available in the snapshot. Location transparency eliminates the ability to identify files' owner while file location cannot be extracted from the file name and as stated before all cloud data are stored in a shared pool, so finding owners of each file seems a challenge. Therefore accessing files in memory capture may be inaccurate evidence to be provided for courts. Moreover, the ability of recovering deleted data from client devices is considered as a complex process as data can be remotely accessed and deleted from any synchronized device.

5.4 Cloud Forensic Investigation Experience

Forensic investigators conduct investigations on cloud using traditional forensic techniques by analyzing network components. In fact, traditional forensic investigation is useless when conducted on most cloud crimes and the available tools are not capable to collect and analyze forensic evidences (Haggerty 2013). The lack of cloud investigation expertise and forensic tools is a major challenge in conducting such investigation. The rapid evolvement of cloud computing makes it a challenge for forensic researchers to develop the required tools and procedure to conduct cloud forensic investigations (Ruan et al. 2011). Therefore, organization should ensure having the required experts to address technical issues in cloud forensic investigations.

5.5 Cloud External Dependencies

Cloud service provider may depend on a third party to manage a type of applications provided to users like email applications and external storage drives that will create a chain of dependencies. Such dependencies may maximize the difficulty of conducting forensic investigation on cloud because a separate investigation process may be required to conduct on each dependency in the chain. In some scenarios, cloud clients use extensions provided by third parties that may contain potential threats like malicious codes. Thus, investigators should use advanced techniques to find such extensions. Investigators can search for the visited websites, accessed web-mails, downloaded files, login details, data entered in web forms and browser cookies. Moreover, they may consider malicious Java Scripts that will affect and modify the evidences. Investigating all components in this chain is mandatory to get accurate evidences. However, in some cases some components may be missed due to some technical or legal aspects provided by the third parties that in reverse will lead to a problem in coordinating investigation processes. In this regard, cloud organizations need to establish a communication channel by organizational policies with third parties to facilitate forensic activities during investigating cloud crimes. This channel should be legally linked to Service Level Agreements (SLAs) and determine the required chain of dependencies which will work together on investigating cloud crimes.

5.6 Live Forensic Investigation on Cloud

Conducting forensic investigation on cloud computing relay on network forensic, therefore live forensic is more suitable investigation mode to be used in such investigations. Live forensic extracts evidences form a running machine at the analysis time while mortem forensic based on extracting evidence from power off machines and capturing images of disk or memory. Live forensic is more reasonable on cloud computing because it is difficult to capture an image of the entire cloud service. Live forensic is conducted to capture network and memory data that could not be captured from hard disk images. However, live forensic rises challenges for investigators in term of extracting and preserving data. In some cases the investigators cannot ensure that all available data have been collected and preserved during the collection process. Preserving data during collection process is mandatory to ensure that evidence will be accepted by courts. Therefore the investigators should use a powerful forensic methods and tools to collect and preserve data from being changed. Nevertheless, in most cases, live forensic is the only provided option for law enforcement agencies to conduct forensic investigation on cloud (Martini and Choo 2012). Some researches stated that only live forensic method can be used to conduct analysis on virtual machines hosting cloud which do not have persistent storage (Birk and Wegener 2011).

While conducting live forensic, time synchronization is required to collect audit logs that are considered as a source of evidences. In cloud forensic, finding accurate time synchronization is more difficult where timestamp is synchronized among a number of different devices in different locations to provide remote access to cloud data.

6. CONCLUSION

Actually, the rapid development of cloud system creates a gap between its infrastructure and finding a suitable forensic investigation technique to be conducted on cloud system crimes. In case of compromising cloud systems, it will be difficult to collect accurate evidences for forensic investigations while data are synchronized and accessed using different devices anywhere. Also, there is a difficulty on accessing public servers to obtain access logs stored on the server. Moreover, accidently deleting data stored on cloud servers raises the concerns of the possibility of recovering data again from cloud servers while it is not physically stored in a specific disk on client's device. In most cases the available digital forensic tools cannot provide the expected result when used to conduct investigation on cloud clients. More researches will be conducted in the future to find a suitable framework to conduct forensic investigation on cloud systems while it is going to be widely used in the coming years. Furthermore, an advanced research may be conducted to study cloud infrastructure and apply the required forensic techniques to collect some potential evidences from both cloud clients and servers by utilizing private cloud systems.

7. REFERENCES

[1] Akarsu B, Bayram K, Slisko J, Corona Cruz A. International Journal Of Scientific Research And Education. Ijsae.in [Internet]. 2013;6(3):221–32. Available from: http://ijsae.in/ijsaeems/index.php/ijsae/article/viewFile/157/137

[2] Birk D, Wegener C. Technical issues of forensic investigatinos in cloud computing environments. Syst Approaches to Digit Forensic Eng. 2011;

[3] Dam M, Chen K. On the Security of Cloud Storage Services. MartindamDk [Internet]. 2011; Available from: http://martindam.dk/wp-content/uploads/cloudstorage.pdf

[4] Foster I, Zhao Y, Raicu I, Lu S. Cloud Computing and Grid Computing 360-Degree Compared. 2008 Grid Comput Environ Work [Internet]. 2008;1–10. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4738445

[5] Haggerty J. Digital Forensics in the Organisation. 2013;(October):17–20.

[6] Huo Y, Wang H, Hu L. A Cloud Storage Architecture Model for Data- Intensive Applications. 2011 Int Conf Comput Manag. 2011;(61073009):26–9.

[7] Kent K, Chevalier S, Grance T, Dang H. Guide to Integrating Forensic Techiques into Incident Response: Recommendations of the National Institute of Standards and Technology. 2006;(August):121.

[8] Maguire JN. Cra Ries. 2013;

[9] Martini B, Choo KKR. An integrated conceptual digital forensic framework for cloud computing. Digit Investig [Internet]. Elsevier Ltd; 2012;9(2):71–80. Available from: http://dx.doi.org/10.1016/j.diin.2012.07.001

[10] Martini B, Choo KKR. Cloud storage forensics: OwnCloud as a case study. Digit Investig [Internet]. Elsevier Ltd; 2013;10(4):287–99. Available from: http://dx.doi.org/10.1016/j.diin.2013.08.005

[11] Marturana F, Me G, Tacconi S. A Case Study on Digital Forensics in the Cloud. 2012 Int Conf Cyber-Enabled Distrib Comput Knowl Discov [Internet]. 2012;111–6. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6384952

[12] Mell P, Grance T. The NIST Definition of Cloud Computing Recommendations of the National Institute of Standards and Technology. Nist Spec Publ [Internet]. 2011;145:7. Available from: http://www.mendeley.com/research/the-nist-definition-about-cloud-computing/

[13] Mulazzani M, Schrittwieser S, Weippl E, Leithner M, Huber M. Dark Clouds on the Horizon : Using Cloud Storage as Attack Vector and Online Slack Space. USENIX Secur [Internet]. 2011;8:11. Available from: http://research.securityresearch.at/wp-content/uploads/publications/dropboxUSENIX2011.pdf

[14] Ruan K, Carthy J, Kechadi T, Crosbie M. Cloud forensics : An overview. InformationWeek [Internet]. 2011;7(2009):16. Available from: http://cloudforensicsresearch.org/publication/Cloud_Forensics_An_Overview_7th_IFIP.pdf

[15] Sabahi F. Cloud computing security threats and responses. 2011 IEEE 3rd Int Conf Commun Softw Networks. 2011;245–9.