Digital Investigation in Cloud Environments

Yue Zhang

Center for Education and Research in

Information Assurance and Security – CERIAS

Purdue University

West Lafayette, IN 47906

Tel: 765-714-9689

Email: zhan1210@purdue.edu

Abstract

Cloud computing brings many promising technological and economical opportunities and is believed to be one of the most transformative technologies in the history of computing (Ruan, et al., 2011). However, this promising technology also brings huge challenges for digital forensics. Due to the decentralized nature of data processing in the Cloud, traditional approaches to evidence collection and recovery are no longer practical. The lack of physical access to servers also constitutes a completely new and disruptive challenge for investigators (Birk, 2011). To make things more complicated, different types of services may have totally different issues for digital investigation. This paper first identifies general challenges for digital investigation in cloud environments as well as specific issues in XaaS service models that make traditional investigation process difficult. Then it discusses possible solutions to these challenges.

Keywords: cloud computing, digital investigation, XaaS, forensic capabilities

-

Digital Investigation in Cloud Environments

According to NIST definition (Mell, 2011), 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. This new technology enables small business to have the same platform to develop or plant their own services by transferring their IT infrastructures into the cloud environment. Many enterprises and customers, however, refused to take advantages of this technology and many of them due to security concerns. In fact, no matter how careful you are with your own personal data. Once you subscribe it to the cloud, you are giving some extra control to the external source. Therefore, cloud security becomes a big concern for cloud customers. Both the Cloud Service Provider (CSP) and customer call for forensically sound digital investigations to protect and assure their privacy and properties.

Digital investigations in a cloud environment should share common features in a traditional network environment. The general investigation processes may also apply in cloud environments such as forensic data collection, evidence examination, analysis and report. The difference is that new approaches and tools should be used in these processes. What’s challenging is mostly happening in the forensic data collection phase. Where is the data? How to get these data in a forensically sound manner? What are the laws or regulations should be concerned? Are there any forensic tools available for gathering data in a cloud environment? These questions illustrate the general concerns for digital investigations in cloud environments.

How to carry out digital investigation along with these challenges is also a focus of this paper. This paper recognizes that investigations need mutual effort of CSPs and investigators.

-

The CSP needs to provide forensic services so that it can better protect customer's privacy as well as the robustness and security of its own. The investigators on the other hand, cannot use traditional approaches and need specific procedures and tools to carry out these investigations in a cloud environment.

General Challenges

Both law enforcements and cloud organizations are facing great challenges in establishing forensic capabilities dealing with cybercrimes in cloud environments. In the technical sense, there are limited tools and procedures that are specifically designed for cloud environments that forensic investigators can use. The cloud organizations like CSPs, on the other hand, may yet lack the capability of segregate and collect forensic data from customers and provide these data to the law enforcements. In the legal sense, there is currently no agreement among cloud organizations on collaborative investigation (Ruan, et al., 2011). Things also become much more complicated when multi-jurisdiction is involved in most international criminal cases.

Difficulties in data collection

Prosecuting criminals must be based on proper forensic evidences. In a cyber crime, these evidences come from data collected from a suspect's computer, electronic devices, storage mediums, online accounts, etc. From a forensic point of view, data collection is paramount because it's the basis and the start point of digital investigations. However, in a cloud environment, data collection becomes difficult. In most of the cloud services, data are not completely stored on the local machines. Instead, many crucial data are stored on the remote storage center owned by the CSP which are out of customer's control. These data centers may be scattered all around the globe and are under the restrictions of different laws and regulations.

Access to forensic data varies dependent on the cloud model; IaaS customers enjoy relatively easy access to all data required for a forensic investigation, while SaaS customers may have little to no access to data required. However, the common feature is that people have decreased access to forensic data so that forensic investigators have to turn to the CSP to require more forensic data with proper warrants. The CSP on the other hand, may or may not store these crucial data required by law enforcements. For example, assuming an evil suspect was observed to be using an online chatting program which was a cloud application. The chatting history was considered very important evidence. When the investigators asked the CSP for these histories, however, they were told that no histories were stored on the database because of economic concerns. This imaginary but frustrating scenario illustrates that when the CSP lacks the capability of providing interfaces and services for forensic data collection, forensic investigation becomes difficult.

Another challenge is collecting *data in execution* (Birk, 2011). Traditionally, forensic data can be categorized into two different states, *data in transit* and *stored data*. *Stored data* is statically stored on allocated hard disks. When *stored data* is "deleted", it is de-allocated but it won't disappear on the hard disk unless it's overlapped by newly allocated data. This fact is often exploited by investigators which explore these "deleted" files on hard disks. *Data in transit* is transferred from one party to another e.g. a typical online instant messaging (IM) over a network can be seen as a *data in transit*. The investigators can use sniffing or network monitoring tools to collect these forensic data in real-time. However, there are data that can be loaded into memory and executed as a process. In this case, the data is neither stored or in transit but in execution. One important implementation of *data in execution* in the cloud environment is snapshot. A snapshot is the state of a system at a particular point in time, process information, machine instruction and allocated/de-allocated data can be analyzed by creating a snapshot of the current system state. This snapshot technology is further discussed later.

Multi-jurisdiction issues

In a cloud environment, resources are shared among people across the world. This implies that people's data may be stored in a shared database but in the physical world people and network services are under restrictions of different jurisdictions. In fact, to ensure service availability and cost-effectiveness, major CSPs, such as Amazon, Salesforce.com, and Google all have data centers around the world in different jurisdictions providing cloud services (Ruan, et al., 2011). Data stored in one data center is replicated to multiple locations to ensure abundance and reduce the risk of single point of failure. Of course these points of replication are potentially under many separate jurisdictions.

The legal challenges of multi-jurisdiction concern the differences among legislations in countries and states where the cloud and its customers reside in. These differences between jurisdictions affects on issues such as what kind of data can be accessed and retrieved in the jurisdiction, how to conduct evidence retrieval without breaching privacy or legal rights of cloud customers according to the privacy policies and regulations in the organizations and specific jurisdiction where multiple customers' data is located, what kind of evidence is admissible to the court in the specific jurisdiction, what kind of chain of custody is needed in the evidence preservation in the jurisdiction.

Sophisticated collaborations between the CSP and the customers or among international law enforcement are required in most of the cloud forensics cases and these collaborations.

Lack of forensic tools

Digital investigation is tool-centric. Forensic practitioners depend on forensic tools to

-

collect and examine evidences. Also, cloud organizations and law enforcements need well-trained staffs to carry out investigations and analysis. However, most cloud organizations in today are dealing with investigations with traditional network forensic tools and staffing, or are simply neglecting the issue (Ruan, et al., 2011).

Traditional forensic tools fall into two categories: static analysis forensic tools analyze stationary data obtained through a formalized acquisition process e.g. disk imaging; live forensic tools collect and analyze “live data” in a real-time manner. However, cloud models break this paradigm because information is difficult to locate, acquisition is impossible when location is questionable, and analysis is nonexistent without acquisition. Cloud forensic tools need to be a hybrid of the current static and live collection and analysis methods, and they need intelligence to note and predict artifacts based on forensic heuristics. The next generation forensic tools must visualize the physical and logical data locations. The visualization must indicate obtainable and unobtainable artifacts, easing the collection burden and preservation estimates. Unobtainable artifacts should be annotated as such in an automated fashion (Tyler, 2011).

The deep-rooted problem of these challenges lies in the fact that it’s hard for relatively “young” cyber forensics to catch up with rapidly emerging technologies and devices. When a new technology appears and may affect the field of cyber forensics, the forensic practitioners must be trained and renew their knowledge related to this technology. Regulations and laws also need to be established to catch up.

Admissibility of evidence

Due to the multi-tenant nature of cloud computing, it’s challenging to prove that the forensic evidences haven’t been contaminated by other users who share the same database. In the IaaS environments, investigators and customers cannot gain control over the hypervisor level.

-

However, attackers may compromise these hypervisors and gain full control of the virtual machines and snapshots. Another concern is that there is currently no standard procedures for collecting evidences in a cloud environment, therefore, how to make sure the investigation process is forensically sound and can later be admissible to the courtroom is difficult.

Specific Challenges in XaaS Environments

In different services types, the extent of customer's control over cloud resources is different. In Software as a Service (SaaS) model, customer can hardly get any control over the database and infrastructures behind the cloud. In Platform as a Service (PaaS), the CSP provides virtual development environment like database, operation systems and development tools for customers to develop their own software and application packets. The investigators and customers may gain some control over how they interact with other dependencies (database, operation system, etc.). In the Infrastructure as a service (IaaS) model, the customers have almost full control of the virtual machine. Customers have the ability to install and set up the image for forensic purposes. Figure 1 illustrates customers' control over resources in different service models.

-

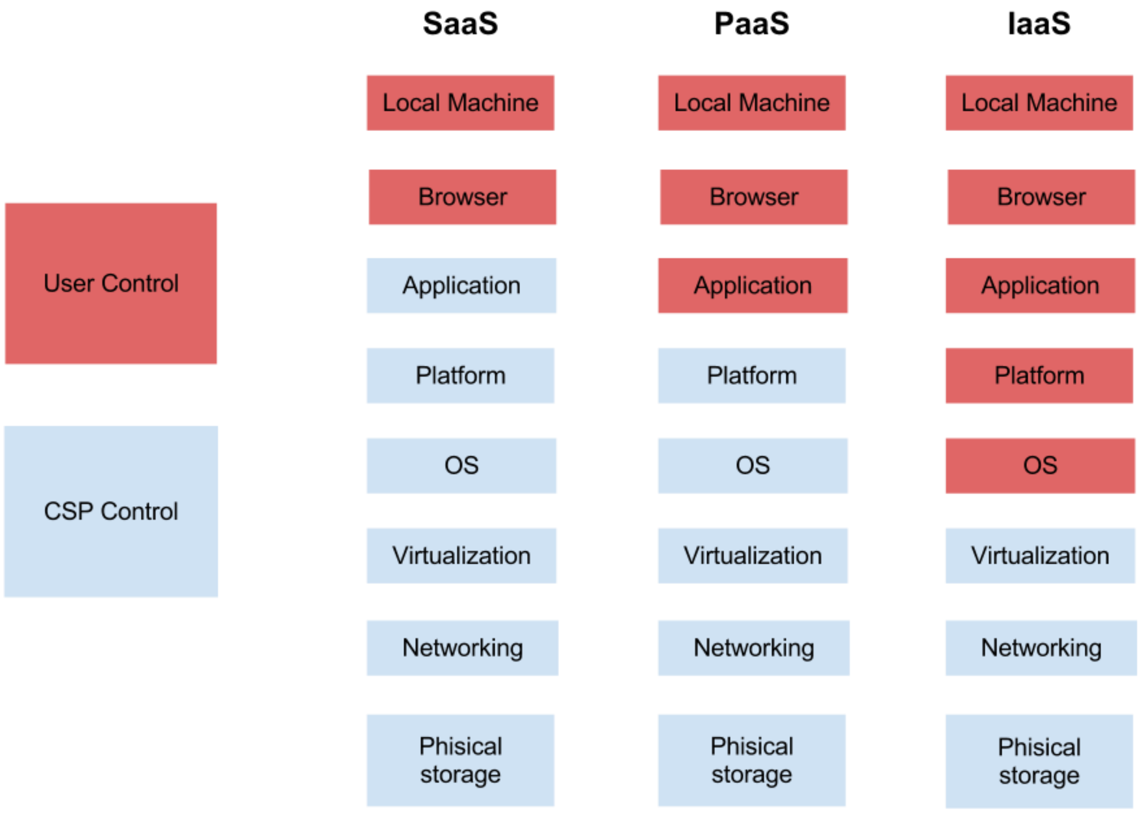

Figure 1. User's Control in Different Service Models

The amount of potential evidence available to the investigator strongly diverges between the different Cloud service and deployment models in which different concerns and strategies should apply.

SaaS Environments

In SaaS environments, the customer does not obtain any control of the underlying operating infrastructure such as database, network, servers, operating systems etc. or even the application that is used. The user side may contain some log information and configuration information of the application, however, investigators have to turn to the CSP for more detailed and high-level logs and data in a lot of cases. Another problem is that data may change after the communication between the user side and CSP. This means what investigators can get from CSP may be different from the original data. For example, the server will store the hash value of user's sensitive information like passwords rather than the original plaintext for security concerns. Things become even worse if the CSP lacks the ability to provide logs that tracing back the activities of the user. In this scenario, investigators may get very limited help from the server side.

PaaS Environments

In PaaS environments, challenges are similar to those in SaaS environments. But one of the main advantages of this model is that the core application is under the control of the customer. Given this circumstance, the customer obtains theoretically the power to dictate how the application interacts with other dependencies (databases, storage entities etc.) Since the customer has the ability to interact with the platform over a prepared API, system states and specific application logs can be extracted.

IaaS Environments

In IaaS environments, the capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. Therefore, IaaS provides much more information that may become forensic evidence than SaaS and PaaS do. Traditional data collection may apply in a IaaS environment which potentially makes investigation much easier (Zimmerman et al., 2011).

Currently, the most common service in IaaS environment is virtual machine (VM). VMs enable users to create snapshot to preserve the current state of the virtual system. Therefore, investigators can take advantage of this technology to use snapshots as a forensic image for further analysis. However, the snapshot is not completely identical to a physical image of hard drives. One difference that investigators may concern about is that the storage of virtual system is logical and based on allocated spaces. Therefore, unallocated spaces and remnants may not be included in the snapshot. Another concern is that criminals may use virtual machines as tools for cyber crimes, but later they may cancel the contracts with CSP, therefore all the history of the VMs may be deleted.

Besides, sharing storage for multi-tenants many cause problems such as the data may be intentionally or accidentally contaminated by other users.

Carrying out Digital Investigations in Cloud Environments

Carrying out digital investigations in a cloud environment should follow basic rules of cyber forensics. In a high level view, the traditional models and procedures also apply in cloud environments. According to DCSA (Rogers, 2006), the practical approach to crime scene analysis can be divided into two parts: *Corpus Delicti* and *Lab*. The *Corpus Delicti* consists of evidence identification, evidence collection and transportation. The *Lab* consists of examination, transportation and report. This model also applies in cloud investigations. The only difference lies in the evidence identification and evidence collection phases while others remain the same in both cloud environments and non-cloud environments.

Evidence Identification & Collection

In the evidence identification phase, investigators have to identify the service type (s) of the specific crime. They then have to identify the types of technologies and devices behind the service. Another important thing is to decide what kind of warrant they would need to compel information from CSP, this may involve crime scene investigation of the suspect, and collecting account and configure information for the cloud service.

-

Evidence collection however, should be based on types of service models. In SaaS environments, investigators should collect local information about which services and accounts the suspect was using and then require a proper warrant to compel more logging information from the CSP. In a PaaS scenario, investigators may be available to more information and interfaces from the client's side. Unfortunately, the customer has no direct control of the underlying runtime environment which strongly depends on the configuration done by the CSP. The Microsoft Azure platform, the environment is made of an virtualized OS (Microsoft Windows), a web server and the runtime environment (.NET). These environments may contain crucial evidence but yet cannot be fully controlled form the client side. Therefore, in many cases, investigators still need to turn to the CSP for more information. In an IaaS environment, client's local machine may gain full control over the forensic data. Evidence collection becomes easier as the local machine may contain full information needed as forensic evidences such as snapshots. Traditional disk imaging applies in all of these models, the internet histories, local files, registry files, caches, cookies etc. are also important evidence in a cloud environment.

Figure 2 illustrates evidence identification and collection procedure in a cloud environment.

-

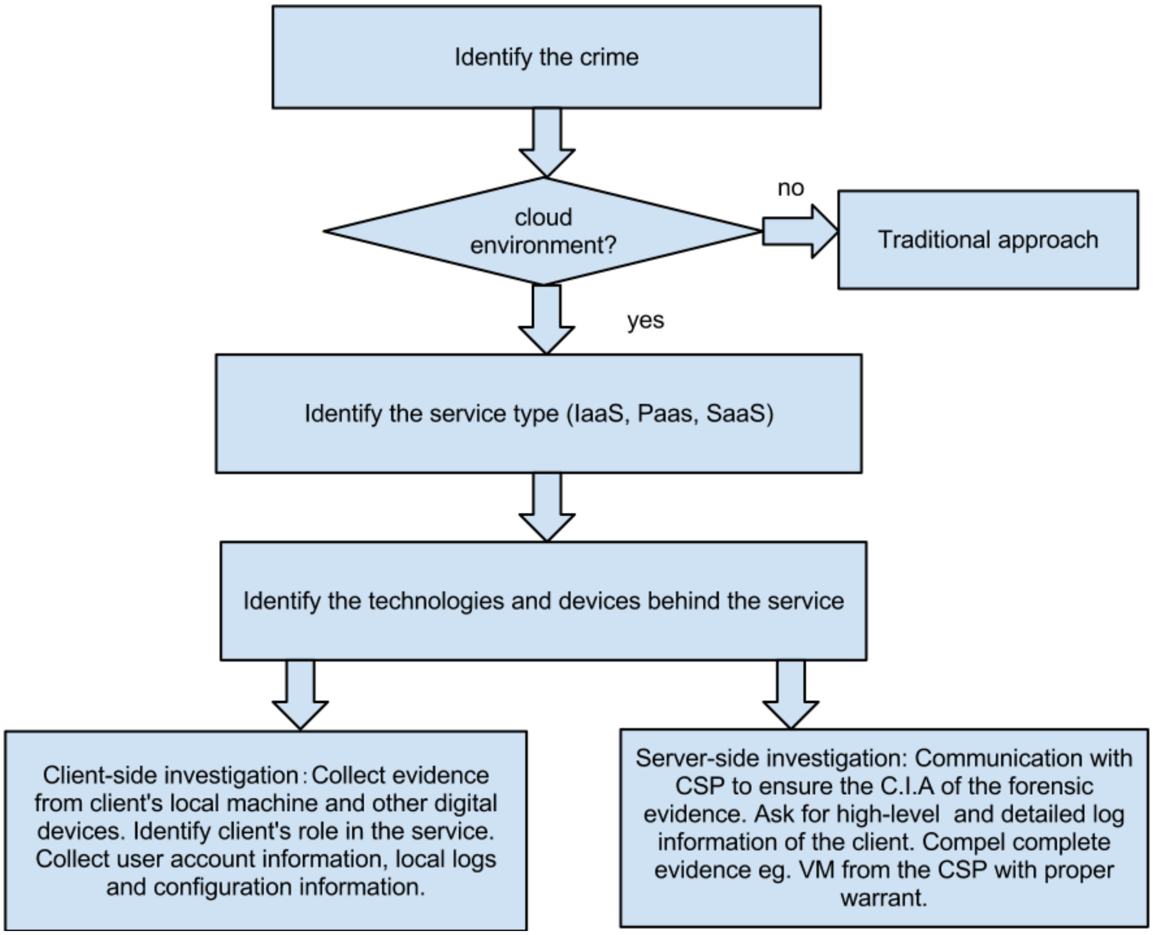
Identify the crime
cloud environment?
no
Traditional approach
yes
Identify the service type (IaaS, Paas, SaaS)
Identify the technologies and devices behind the service
Client-side investigation : Collect evidence from client's local machine and other digital devices. Identify client's role in the service. Collect user account information, local logs and configuration information.
Server-side investigation: Communication with CSP to ensure the C.I.A of the forensic evidence. Ask for high-level and detailed log information of the client. Compel complete evidence eg. VM from the CSP with proper warrant.

Figure 2. Evidence Identification and Collection in a Cloud Environment

Establishing Forensic Capabilities for CSP

Successful digital investigations in a cloud environment often need sophisticated collaboration between the CSP and the forensic investigators. Once the infrastructures or services are under malicious attacks, they may need to provide proper and enough forensic evidences to the law enforcements. On the other hand, law enforcements and customers may ask the CSP for detailed evidence to identify the criminal. The mutual effort of the law enforcements, CSPs and cloud customers should always be a win-win game. In this sense, the CSP should establish standards and regulations to ensure some forensic capabilities to provide better security and robustness of its services.

-

Forensic Capabilities	Objectives
Data segregation	Segregate multi-tenants' data within a shared database.
C.I.A	The cloud server should ensure the basic principles of security to assure confidentiality, integrity and availability of data and services.
Auditing	Logging the activities of both the clients and the servers for future reconstruction.
Secure Defaults	When users stop or close a service in the client side, or when users cancel contracts with a CSP, the server should respond in a secure way without damaging the forensic evidences.
Forensic interfaces	The CSP should provide interfaces for investigators or customers to get forensic evidence or a proof of an activity.
Backups	The CSP should store important volatile evidence for forensic investigation.

Table 1. Defining forensic capabilities for CSP

Data segregation

Due to the fact that multi-tenants' data be shared in a single data base. It's the CSP's responsibility to segregate these data for different users when giving these evidences to law enforcements. The biggest concern here is privacy, the CSP must be able to extract all the related forensic data without contaminating or leaking other user's private data.

C.I.A

-

Confidentiality, integrity and availability are the basic principles of information security (Perrin, 2008). Confidentiality requires the CSP should leak information of the unrelated data to the crime or other user's data. This means that extracting forensic data should be based on a "need to know" principle. Integrity requires the CSP to protect the data from malicious or accidental contamination. This may be an important basis for the admissibility of the evidence. Availability requires that the CSP shouldn't be vulnerable to Denial of Service (DoS) attack, which may interrupt the investigation process. These requirements may involve technologies like intrusion detection, encryption, authentication, access control, etc.

Auditing (logging)

Auditing is crucial for the CSP or the investigators to reconstruct the crime activities. The CSP should keep detailed logs for the activities between the users and servers. The CSP also should have the ability to reconstruct the incidents based on these logs.

Secure defaults

Secure defaults aim at protecting volatile data. When end users close the web browser or log off their cloud accounts, some volatile data like caches, temporary files, un-saved data would be lost. Another scenario that has been mentioned above is that an attacker may use virtual machines to carry out the attacks and then cancel the contract with the CSP. In this scenario, the CSP shouldn't completely wipe out all the information from the virtual machine; otherwise the investigators would have little chance to identify the criminal. Therefore, when defaults happen, the CSP should preserve important information rather than completely ignore or delete the volatile data. The CSP should also preserve snapshots of VM so that the activities of the criminals use that machine can be traced back.

Forensic interfaces

-

The CSP should provide interfaces for investigators or customers to collect evidences. It would make investigation easier and faster. For example, the CSP can provide API for customers to query basic log information of activities related to their account at specific time, which can be used as a proof of activity. For investigators, the CSP can provide easy interfaces for them to collect or look up the forensic data needed.

Backups for important forensic data

Backups require that the CSP store all the important live data when the customer is communicating with the server. From a forensic point of view, any data in transit may become potential evidences. For CSPs however, storing all these data may become huge burden for database, but at least some of the important data should be stored like chatting histories, subscribe information, ip addresses, etc.

Conclusions

Cloud computing is a new battlefield of cyber crime, as well as a new ground for novel investigative approaches. The traditional digital investigation approaches and tools are now facing great challenges in cloud environments caused by loss control of forensic data. The multi-jurisdiction and multi-tenant feature also brings about legal and privacy concerns about forensic data collection. Also, in different service models, they extent to which the customer gain control over the data are different. The technologies and devices behind these services also varies. Therefore, different investigation strategies should apply in different service models. In the first part of the paper, I discussed some general problems caused by the common features of the cloud services and then go into specific issues in SaaS, PaaS and IaaS environments.

The rise of cloud computing is pushing digital forensics into a new horizon with so many challenges. However, from a forensic point of view, the basic procedures and model of

-

traditional cybercrime investigation may also apply in cloud environments. The only difference lies the evidence identification and evidence collection phases. In the identification phase, investigators should decide which service model the suspect was using and what kind of technologies and devices are behind the service. They also have to identify the role of the cloud service in the crime and to identify the suspect's account and specific cloud applications he/she was using so that they can apply for proper warrants and compel evidence from the CSP. In the evidence collection phase, traditional tools and procedures like disk imaging are also useful and necessary. But the types of evidences can be collected and the strategies to collect evidence may vary according to different service models.

We realized that the digital investigation must involve sophisticated collaboration between forensic investigators and CSPs. Therefore, several forensic capabilities were defined for the CSP to help both the efficiency of investigation and the robustness of its own services.

Cyber forensics is a young and maturing field. The deep-rooted problem is the relative slow progress of forensic research and relevant laws and international regulations compare to the rapidly evolving technology. Techniques need to be developed, regulations need to catch up, forensic practitioners need to be trained and need to be equipped with new knowledge and skills to deal with the new grounds of cloud computing.

-

References

Anderson, N. (2012). *Confirmed: US and Israel created Stuxnet, lost control of it.* Retrieved from http://arstechnica.com/tech-policy/2012/06/confirmed-us-israel-created-stuxnet-lost-control-of-it/

Barrett, D. & Kipper, G.(2010). *Virtualization and Forensics: A Digital Forensic Investigator's Guide to Virtual Environments*. Syngress, 6 2010.

Beebe, N. (2009). *Digital forensic research: The good, the bad and the unaddressed.* Advances in Digital Forensics

Birk, D. (2011). *Technical challenges of forensic investigations in cloud computing environments.* Retrieve from: http://www.zurich.ibm.com/~cca/csc2011/submissions/

Fowen, A. (2012). *Cloud computing and computer forensics.* Retrieved from http://www.intaforensics.com/Blog/Cloud-Computing-And-Computer-Forensics.aspx

Perrin, C. (2008). *The CIA Triad.* Retrieved from http://www.techrepublic.com/blog/security/the-cia-triad

Ruan, K., Carthy, J.,Kechadi, T., Crosbie, M. (2011). *Cloud forensics: An overview*. Advances in Digital Forensics VII, Springer.

Guan,Y. (2007). *Digital Forensics: Research Challenges and Open Problems.* Iowa State University. Retrieved from itsecurity.uiowa.edu/securityday/documents/guan.pdf

Mell, P. & Grance. T. (2011). *The NIST Definition of Cloud Computing.* Retrieved from: http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

Meyers, M. & Rogers, M. (2004). *Computer forensics: The need for standardization and certification within the U.S. court systems*. International Journal of Digital Evidence, Fall 2004, 3(2).

-

Rogers, M. & Seigfried, K. (2004). The future of computer forensics: A needs analysis survey. *Computers and Security, 23 (1)*, 12-16.

Rogers, M. (2006). *DCSA: A Practical Approach to Digital Crime Scene Analysis.* West Lafayette, Purdue University.

Saferstein, R. (2004). *Criminalistics: An introduction to forensic science* (8th ed.). Upper Saddle River: Pearson Education.

Zimmerman, S. & Glavach, D. (2011). *Cyber Forensics in the Cloud.* IAnewsletter, Volume 14, November 1, Winter 2011.

-