Chapter 2: Theoretical Background and Literature Review

Digital forensic techniques for cybercrimes investigation in cloud computing environment (i.e. cloud forensic) is comprising various areas such as digital forensics and cloud computing. The aim of this chapter is to gain a deeper understanding and identify challenges and opportunities in forensic science concerns in cloud computing as well as current research review which can help to develop strategies for measurements, standards and technology research to mitigate these challenges that cannot be handled with present technology and approaches. This chapter discusses the basic concepts of cloud computing, digital forensic, cybercrimes, and cloud forensic in addition to literature review about related work to digital forensics techniques in cloud computing environment.

The rest of this chapter is organized as follows: section 2.1 discusses an overview and principles of cloud computing, digital forensic, cybercrimes, and challenges, and opportunities of cloud forensic while a state-of-the-art in cloud forensic domain is presented in section 2.2. Finally, the chapter summary is provided in section 2.3.

2.1 Theoretical Background

This section introduces the principles of cloud computing, digital forensic, cybercrimes, and cloud forensic challenges, and opportunities.

2.1.1 Cloud Computing

This subsection discusses cloud computing definition, characteristics and models.

2.1.1.1 Cloud Computing Definition

There are many definitions for cloud computing such as:

- **National Institute of Standards and Technology (NIST) defined cloud computing as:** “Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable resources (e.g., networks, servers, storage, applications, and services) that can be rapidly
provisioned and released with minimal management effort or service provider interaction” [6].

- **Cloud Security Alliance (CSA) defined cloud computing as:** “Cloud computing is an evolving term that describes the development of many existing technologies and approaches to computing into something different. Cloud separates application and information resources from underlying infrastructure, and the mechanisms used to deliver them. Cloud enhances collaboration, agility, scaling, and availability, and provides the potential for cost reduction through optimized and efficient computing” [7].

**2.1.1.2 Cloud Computing Characteristics**
The cloud computing definition by the NIST identifies five crucial characteristics which are described below as shown in Figure 2.1 [6]:

- **On-demand self-service:** A user can automatically provision cloud capabilities without needing human interaction with every service request.

- **Broad network access:** It means can provide cloud capabilities which can be accessed over the network using standard devices.

- **Resource pooling:** The cloud resources are pooled to serve numerous users using a multi-tenant paradigm.

- **Rapid elasticity:** The cloud capabilities can be provisioned and released elastically.

- **Measured service:** Cloud resource use can be optimized and controlled automatically by leveraging a metering ability which is suitable to the service type.

**2.1.1.3 Cloud Computing Models**
Cloud computing models divide into two main types, Deployment and Service models as shown in Figure 2.1.
Deployment Models

The definition of cloud computing by the NIST identifies four deployment models which are described below in Figure 2.2 [6]:

1. **Private Cloud**: Private cloud services are used only by a single company or organization and are not exposed to the public.

2. **Public Cloud**: Services of public cloud are exposed to the public and can be used by anyone.

3. **Community Cloud**: Community cloud services are used by numerous organizations in order to lower the costs, as compared to the private cloud.

4. **Hybrid Cloud**: The hybrid cloud services can be distributed in multiple cloud types.

---

**Figure 2.1**: Cloud Computing Framework.
Service Models

The definition of cloud computing by the NIST identifies three service models which are described below in Figure 2.3 [6]:

1. *Software as a Service (SaaS):* In SaaS, the user is able to use applications running on the cloud which provides by the cloud provider. These applications are accessible from various machines over the Internet. The user does not control or manage the primary cloud infrastructure involving servers, storage, network and operating systems. Example of the SaaS is storage service such as SkyDrive, Google Drive, and Dropbox.

2. *Platform as a Service (PaaS):* In PaaS, the user is able to deploy on the cloud infrastructure applications produced using programming languages and tools supported by the cloud provider. The user does not control and manage the primary cloud infrastructure involving servers, network, or storage, but has control over the deployed applications. Example of the PaaS is Windows Azure from Microsoft and Google App Engine from Google.

3. *Infrastructure as a Service (IaaS):* In IaaS, the user is able to provision cloud resources such as network, processing, and storage. Also, the user is able to deploy and run software which can involve operating systems and applications. The user does not control or manage the principal cloud infrastructure but can control over applications, storage, and operating systems. Example of IaaS is Amazon EC2 service.
Figure 2.2: Deployment Models: (a) Private Cloud, (b) Public Cloud, (c) Community Cloud and (d) Hybrid Cloud.
Figure 2.3: Service Models: (a) Software as a Service (SaaS), (b) Platform as a Service (PaaS) and (c) Infrastructure as a Service (IaaS).
2.1.2 Digital Forensic

2.1.2.1 Digital Forensic Definition

One of the widely used definitions of digital forensic was introduced by Digital Forensic Research Workshop (DFRWS) as: “The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations” [8].

2.1.2.2 Digital Forensic Investigation Life Cycle

From the digital forensic definition, digital forensic investigation process involves many steps as follow as shown in Figure 2.4:

- **Identification**: It is involved in two key phases: identification of crime and identification of digital evidence.

- **Collection**: In this phase, an examiner gathers digital evidence from the crime scene for using in the next examination phase.

- **Extraction**: In extraction phase, the digital investigator extracts digital evidence from various types of devices such as cell phone, hard disk, and e-mail.

- **Analysis**: In this phase, examiner interprets and correlates the extracted digital evidence to come to a summary, which can prove or disprove criminal accusations.

- **Examination**: In the examination phase, the investigator extracts and inspects the data and their characteristics.

- **Report**: In this process, the investigator and examiner make a prepared report to represent his/her findings from forensic analysis of crime evidence. This report should be suitable enough to present in the court of law.
2.1.2.3 Digital Evidence

Digital evidence is the source data that help and assist digital investigators for cybercrimes investigation and examination to bring the criminals to condemnation. The digital evidence may be in various forms such as text, audio, image, and video. In the court of law, the evidence used to prove and establish that cybercrime or incident has been committed or can deliver a relation between a crime and its victim. Figure 2.5 shows different types of digital evidence.

![Digital Evidence Diagram](image-url)
2.1.3 Cybercrimes

In the recent time, cybercrimes become more critical as threats for breaching system security due to innovative ideas that criminals have with regards to new ideas and ways to commits these crimes. The criminals exploit vulnerabilities of new technologies to commit their crimes in a way that make it is very difficult to discover and trace them back. The dynamic nature of cloud computing contributes to the problems met by investigators when extracting and preparing digital evidence for a court of law.

2.1.3.1 Cybercrime Definition

Cybercrime is defined as “Offences that are committed against individuals or groups of individuals with a criminal motive to intentionally harm the reputation of the victim or cause physical or mental harm to the victim directly or indirectly, using modern computing devices and telecommunication networks” [9].

2.1.3.2 Cybercrime Classification in Computer Forensic

Cybercrime is classified as the following [9]:

1. *The computer as a target*: The criminal seeks to deny the legitimate users or owners of the system access to their data or computers such as Denial of Service (DOS) attacks.

2. *The computer as a tool of the crime*: The computer is used to gain some other criminal objective. For example, a thief may use a computer to steal personal information.

3. *The computer as incidental to a crime*: The computer is not the primary instrument of the crime; it simply facilitates it.

4. *Crimes associated with the prevalence of computers*: This includes crimes against the computer industry, such as software piracy.

2.1.3.3 Cybercrime Classification in Cloud Forensic

In the cloud computing environment, cyber-crimes divided into two main types:

1. Crimes using the cloud infrastructure resources capacities to be performed the malicious attacks.

2. Crimes against the cloud infrastructure.
2.1.3.4 Classification of Attacks based on Cloud Components

The classification of attacks based on in terms of the network, virtual machine, storage and applications executing on a cloud platform as follows [53]:

1. **Network-based attacks**
   - **Botnets:** A botnet may be used to steal data from a host machine and communicate it to a bot-master. A command and control system is established with a bot-master and several machines can act as a stepping-stone to steal private information.
   - **Spoofing attacks:** Spoofing attacks in a network impersonate entities for malicious purposes. An IP spoofing attack may swap the IP address in a network packet with a fake source IP address. Similarly, DNS spoofing attack may cause DNS server to return an incorrect IP address thereby redirecting network traffic to an attacker's machine.
   - **Port scanning:** A port on a machine may be probed to check the status of a service executing on the target machine. The port scanning requires access to the network hosting the target machine.

2. **VM based attacks**
   - **Cross VM side-channel attacks:** These types of attacks can extract information regarding resource usage, cryptographic keys and other information from a target VM which is residing on the same physical machine as that of the attacker VM.
   - **VM scheduler based attacks:** A few vulnerabilities of the scheduler may result in resource stealing or theft-of-service.
   - **VM migration and rollback attacks:** When an active VM is being migrated from the host physical machine to another physical machine, the contents of the VM files become vulnerable to various attacks.
   - **VM creation attacks:** A malicious code can be placed inside a VM image which is then replicated during the creation of virtual machines.

3. **Storage-based attacks**
   - **Data deduplication:** It becomes possible to recognize the files and their contents due to the data deduplication.
Data scavenging: The process of erasing data from a storage system does not remove data entirely. Therefore, the removed data may be recovered by attackers which are referred to as data scavenging.

4. Application-based attacks

- Malware injection and steganography attacks: A malicious code may be injected in an application if a cloud platform permits for an insecure interface for application development.
- Shared architectures: On a shared architecture, the execution path of a victim's application can be traced. It can be further used to identify victim's actions and hijack his account.
- Web services and protocol-based attacks: The web services use several protocols such as Simple Object Access Protocol (SOAP) whose message header can be modified to contain invalid requests.

2.1.4 Cloud Forensics

2.1.4.1 Cloud Forensics Definition

The term of cloud forensics was introduced by Ruan et al. [12] to identify the rapidly emerging need for digital forensics in the cloud. She defined cloud forensic as a cross-discipline of cloud computing and digital forensics. Also, mentioned that in “Cloud forensics is a subset of network forensics. Network forensics deals with forensic investigations of networks. Cloud computing is based on broad network access. Therefore, cloud forensics follows the main phases of network forensics with techniques tailored to cloud computing environments”.

The Ruan’s working definition of cloud forensics is: "Cloud forensics is the application of digital forensic science in cloud computing environments. Technically, it consists of a hybrid forensic approach (e.g., remote, virtual, network, live, large-scale, thin-client, thick-client) towards the generation of digital evidence. Organizationally it involves interactions among cloud actors (i.e., cloud provider, cloud consumer, cloud broker, cloud carrier, cloud auditor) for the purpose of facilitating both internal and external investigations. Legally it often implies multi-jurisdictional and multi-tenant situations" [10].
2.1.4.2 Cloud Forensics Challenges

To understand the cloud challenges, the NIST [11] developed a formula for a normalized sentence syntax that allows expression of all cloud forensics challenges in a format as follows:

Normalized challenge [formula]:

  For an [actor/stakeholder], [action/operation] applicable to [object of this action] is challenging because [reason]

- **Actor/Stakeholder:** This variable [a noun] identifies the stakeholder(s) who are affected by the challenge that has been identified. Examples of stakeholders include cloud consumers, investigators, first responders, etc.

- **Action/Operation:** This variable [a verb] identifies the activity that the stakeholder would like to perform. Examples of actions include decrypting, imaging, gaining access, etc.

- **Object of This Action:** This variable identifies the specific item upon which the action is to be performed. Examples of objects include data, audit logs, timestamps, evidence, etc.

- **Reason:** This variable identifies the primary challenges that the stakeholder faces in order to perform the specified action on the object.

The normalized description of some challenges are [11]:

- For **forensic examiners**, identifying and attributing data that is deleted in the cloud to a specific user is a challenge because the sheer volume of data and users constantly operating in a cloud environment limits a number of backups that the cloud Provider will retain.

- For **investigators**, correlation of activity is a challenge because there is no interoperability between cloud Providers

- For **all investigators** and courts, reconstruction of virtual images or storage is challenging because these reconstruction algorithms need to be validated or developed.
For investigators/law enforcement/analysts, the collection and preservation of forensic evidence is challenging because there is a lack of interoperability among providers and there is lack of control from the customer's perspective into the proprietary architecture and/or the technology used.

For law enforcement, ensuring proper chain of custody and security of data, metadata, and possibly hardware is a challenge because it may be difficult to determine ownership, custody, or accurate location.

For law enforcement and courts, ensuring proper chain of custody of data is a challenge because the distributed, shared infrastructure of cloud computing makes identifying and validating a chain of custody difficult.

To assist in a meaningful analysis, the NIST [11] categorized the challenges into the following nine major groups as shown in Figure 2.6 while the description of some cloud forensic challenges is tabulated in Table 2.1. The categories and associated descriptions are summarized below as follows:

1. Architecture: In cloud forensics, the architecture challenges include dealing with variability in cloud architectures between providers; tenant data compartmentalization and isolation during resource provisioning; proliferation of systems, locations and endpoints that can store data; accurate and secure provenance for maintaining and preserving chain of custody; infrastructure to support seizure of cloud resources without disrupting other tenants; etc.

2. Data collection: Challenges of data collection include locating forensic artifacts in large, distributed and dynamic systems; locating and collecting volatile data; data collection from virtual machines; data integrity in a multi-tenant environment where data is shared among multiple computers in multiple locations and accessible by multiple parties; inability to image all the forensic artifacts in the cloud; accessing the data of one tenant without breaching the confidentiality of other tenants; recovery of deleted data in a shared and distributed virtual environment; etc.

3. Analysis: Analysis challenges in cloud forensics include correlation of forensic artifacts across and within cloud providers; reconstruction of events from virtual images or storage; integrity of metadata; timeline analysis of log data including synchronization of timestamps; etc.
4. **Anti-forensics**: Anti-forensics is a set of techniques used specifically to prevent or mislead forensic analysis. Challenges in cloud forensics include the use of obfuscation, malware, data hiding, or other techniques to compromise the integrity of evidence; malware may circumvent virtual machine isolation methods; etc.

5. **Incident first responders**: Incident first responder challenges in cloud forensics include confidence, competence, and trustworthiness of the cloud providers to act as first-responders and perform data collection; difficulty in performing initial triage; processing a large volume of forensic artifacts collected; etc.

6. **Role management**: Role management challenges in cloud forensics include uniquely identifying the owner of an account; decoupling between cloud user credentials and physical users; ease of anonymity and creating fictitious identities online; determining exact ownership of data; authentication and access control; etc.

7. **Legal**: Legal challenges in cloud forensics include identifying and addressing issues of jurisdictions for legal access to data; lack of effective channels for international communication and cooperation during an investigation; data acquisition that relies on the cooperation of cloud providers, as well as their competence and trustworthiness; missing terms in contracts and service level agreements; issuing subpoenas without knowledge of the physical location of data; seizure and confiscation of cloud resources may interrupt business continuity of other tenants; etc.

8. **Standards**: Standards challenges in cloud forensics include lack of even minimum/basic SOPs, practices, and tools; lack of interoperability among cloud providers; lack of test and validation procedures; etc.

9. **Training**: Training challenges in cloud forensics include misuse of digital forensic training materials that are not applicable to cloud forensics; lack of cloud forensic training and expertise for both investigators and instructors; limited knowledge by record-keeping personnel in cloud providers about evidence; etc.
Table 2.1: Description of Some Cloud Forensic Challenges.

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Deletion in the cloud</em></td>
<td>Deletion in the cloud is frequently based on the deletion of nodes pointing to data in virtual machines.</td>
</tr>
<tr>
<td><em>Recovering overwritten data</em></td>
<td>Recovery of data which deleted in the cloud is difficult because it gets overwritten by another user.</td>
</tr>
<tr>
<td><em>Log format unification</em></td>
<td>Unification of log formats challenges in the cloud because it is extremely difficult to unify log formats from the enormous resources available in the cloud.</td>
</tr>
<tr>
<td><em>Use of metadata</em></td>
<td>The use of metadata (as an authentication method) may be in danger as common fields such as creation date, last modified date, and last accessed date may be altered as the data is migrated to and within the cloud.</td>
</tr>
<tr>
<td><em>Multiple venues and geo-location</em></td>
<td>Geo-location unknowns can impact the chain of custody in finding evidence and identifying resources that are required for access to the system</td>
</tr>
<tr>
<td><em>Lack of transparency</em></td>
<td>The cloud's operational details aren't transparent enough to users</td>
</tr>
<tr>
<td><em>Cloud confiscation and resource seizure</em></td>
<td>Confiscation of cloud resources can often affect the business continuity of co-tenants</td>
</tr>
<tr>
<td><em>Potential evidence segregation</em></td>
<td>Segregation of forensic data in an infrastructure shared by multiple users is required. Tools used for provisioning and de-provisioning resources are being improved. It is a challenge to segregate resources during investigations without breaching the confidence of other users who share the cloud infrastructure.</td>
</tr>
</tbody>
</table>
2.1.4.3 Cloud Forensics Opportunities
The digital forensic investigation has various opportunities to be applied in cloud computing environment as follow [12]-[13]:

- **Cost effective:** Implement forensic service in cloud computing environment that allow utilizing the huge capacities of cloud computing without transfer the digital evidence from the cloud to the other side to perform the investigation process which needs high bandwidth.

- **Data abundance:** Replication of data in cloud environment introduces the essential opportunity for cloud forensic for recover the lost and deleted data from the cloud to proof the crime.

- **Scalability and flexibility:** Cloud forensic services can utilize the facilities of scalability and flexibility resource use, for example, providing unlimited storage, compute and network resources with the pay-per-use method.

- **Policies and standards:** Develop new standards and policies for cloud forensic science due to rapid change of the technology of cloud computing and cybercrimes against it.

- **Forensics as a Service (FaaS):** Cloud computing provides one powerful option for digital investigators and examiners called Forensic as a Service (FaaS). The forensic investigator can deliver the FaaS through utilizing the enormous cloud capabilities. This service makes digital forensics as an “on-demand” service for allowing massive storage and processor power as necessary to conduct a digital investigation of crimes. Forensic servers will reside on the cloud side, offline, until require arises for them. Documents could be backed up into the cloud for the digital investigators to use without having to disrupt normal business. Indeed the cloud resources could be used for sorting, searching, and hashing the evidence data. There are many benefits of the Forensic as a Service as follows [5]:

  - **Reduce evidence acquisition time:** If a server in the cloud is compromised, it can be cloned and made immediately available to a cloud forensics server.
- **Reduce service downtime:** Due to the hardware abstraction in the cloud, specialized hardware will not have to be obtained in order to continue the acquisition of the evidence in some situations.

- **Reduce evidence transfer time:** The clouds distributed file system allows for making fast bit-for-bit copies.

- **Reduce forensic image verification time:** Some cloud environments use a cryptographic checksum or hash that can drastically reduce the time required to hash files offline.

- **Decrease time to access protected documents:** The pooling of CPU power available in the cloud can make decryption much faster.

- **Virtually unlimited log storage:** Cloud storage solutions will make the need for estimating how much disk space is needed for logging unnecessary, allowing for a considerable amount of log data to be kept and used during an investigation.

- **Improve log indexing and searches:** Along with unlimited storage, logs can be indexed and searched effectively in real time with cloud resources.

### 2.2 Literature Review

Despite significant research in digital forensics, little has been written about the applicability of forensics to cloud computing environments. This section provides State-of-the-Art of research work related to cloud forensic to understand current challenges and proposed solutions.

#### 2.2.1 Current Research Problems in Cloud Forensics

In this division, a study of open problems and impediments in cloud forensics will discuss as follows: Collection and Acquisition of Forensic Data, Log Information, Service Level Agreement (SLA), Virtual Machine Introspection (VMI), Data Provenance in Cloud, Trusted Platform Module (TPM), Isolating a Cloud Instance and Forensic Analysis for Cloud Storage Services.

- **Collection and Acquisition of Forensic Data:** Data collection and acquisition from cloud infrastructure is a challenging step in cloud forensics. Cloud Service Providers
(CSPs) can play a dynamic role in this stage by providing a web-based management console such as AWS management console. Dykstra et al. [14] suggested a cloud management plane for use in IaaS model. From the console panel, users, as well as investigators, can collect virtual machine image, network, process, logs, and other digital evidence, which cannot be collected in other methods. The only problem with this solution is that it needs an extra level of trust – trust in the management plane. In classical evidence collection process, where they have physical access to the system, so this level of trust is not necessary. Dykstra et al. [14] introduced a six layers trust model which are Guest application/data, Guest OS, Virtualization, Host OS, Physical hardware, and Network. The extra down the stack is, the less cumulative trust is needed. There are many issues that make the data collection in cloud forensic is challenging. These issues such as physical inaccessibility of digital evidence, dependence on the CSP, volatile data in the cloud, less control in cloud infrastructure, legal and trust issues, multi-tenancy, and large bandwidth.

- **Log Information:** One of the significant digital evidence for forensic investigation is log information which for forensic investigations. Some researchers have identified logging related to the cloud. There are many challenges related to log information in cloud forensics such as decentralization, volatility of logs, multiple tiers and layers, accessibility of logs, dependence on the CSP and absence of critical information in logs.

Marty et al. [15] proposed a log management solution, which can solve many logging challenges. The proposed approach consists of three steps. In the first step, logging must be enabled on all infrastructure components to gather logs. The second step is for creating a synchronized, reliable, bandwidth, and encrypted transport layer to transfer log from the source to a central log collector. The last step deals with ensuring the presence of the desired information in the logs.

To facilitate logging in the cloud, Zafarullah et al. [16] introduced logging provided by operating system and security logs. So as to explore digital forensics in the cloud, they set up cloud computing environment by using Eucalyptus. For their experiment, they launched a DDoS attack from two virtual machines and analyzed bandwidth
usage log and processor usage log to detect the DDoS attack. From the logs in
/var/eucalyptus/jetty-request-05-09-xx file on Cloud Controller (CC) machine, it is
possible to recognize the attacking machine IP, browser type, and content requested.
From these logs, it is also possible to determine the total number of VMs, controlled
by single Eucalyptus user and VMs communication patterns.

To get essential logs from all the three cloud service models, Birk et al. [17] proposed
that the CSP could provide process, network and access logs to the user by read-only
Application Programming Interface (APIs). Using these APIs, the user can provide
appreciated information to digital investigators. In PaaS, users have full control on
their application and can log a variety of access information. They proposed a central
log server, where the user can store the log information. The users can encrypt and
sign the log data before sending it to the central server In order to protect log data
from possible eavesdropping and modifying the action.

- **Robust Service Level Agreement (SLA):** In the recent time, there is an immense gap
in the current Service Level Agreement (SLA), which neither describes the cloud
service provider (CSP) responsibility at the time of some malicious crimes nor their
role in the digital forensic investigation. Researchers gave emphasis on sound and
robust SLA between cloud service providers and users [18]-[19]. To resolve the
transparency issues, the CSP should build a long-term trust relationship with users
or customers. A robust SLA should state how the providers deal with the cloud-based
crimes, i.e., how and to which extent they help in the forensic investigation process.
In this context, another question can come – how we can be sure of the robustness of
an SLA. To ensure the quality of SLA, it can by taking assistance from a Trusted
Third Party (TPA).

- **Virtual Machine Introspection (VMI):** Virtual Machine Introspection (VMI) is the
process of externally monitoring and auditing the runtime state of the virtual machine
(VM) from either the Virtual Machine Monitor (VMM) or from some virtual
machine other than the one being examined. By runtime state, it is referring to
memory, registers, disk, processor, network, and other hardware-level events.
Through this process, can execute a live forensic analysis of the system, while
keeping the target system unaffected. Hay et al. [20] presented that if a VM instance is compromised by installing some rootkit to hide the malicious events, it is still possible to identify those malicious events by performing the VMI process. They used an open source VMI library, Xen (VIX) suite to perform their experiment. However, this tool is no longer maintained under this name, it is now known as LibVMI [21].

- **Data Provenance in the Cloud:** Provenance plays an important role in the forensic investigation in the cloud because provenance provides the history of an object. By implementing secure provenance, we can get some important forensic information, such as, who owns the data at a given time, who accesses the data, and when. Some researchers have applied the principles of provenance to cloud forensics. Secure provenance can ensure the chain of custody in cloud forensics as it can provide the chronological access history of evidence, how it was analyzed, and preserved. There have been several projects for secure provenance in cloud computing [22], [23], but no CSP has practically implemented any of the mechanisms yet.

- **Trusted Platform Module (TPM):** To preserve the integrity and confidentiality of the data, several researchers proposed Trusted Platform Module (TPM) as a solution. The TPM for cloud computing proposed by several researchers for ensuring trust in cloud computing [24], [25]. By using TPM, can get machine authentication, signing, hardware encryption, secure key storage, and attestation. It can provide the integrity of the running virtual instance, trusted log files, and trusted deletion of data to customers. However, Dykstra et al. [14] declared that TPM is not totally secure and it is possible to change a running process without being discovered by the TPM. Furthermore, at current, CSPs have heterogeneous hardware and few of them have TPM. Hence, the CSPs cannot ensure a homogeneous hardware environment with TPM in the future.

- **Isolating a Cloud Instance:** Isolation is crucial due to it supports to protect evidence from contamination so the cloud instance has to be isolated if any crime happens in that instance. However, as multiple instances can be located in one node, this task becomes challenging. Delport et al. [26] provided some possible techniques for cloud
isolation such as Instance Relocation, Server Farming, Failover, Address Relocation, Sandboxing, and Man in the Middle (MITM). These techniques are as follows:

- **Instance Relocation Technique:** To move an instance, data on the secondary storage, the content of the virtual memory, and the running processes need to be moved. Relocation can be done in two manners – manual and automatic. In the manual mode, the administrator has all the power to move the instance. In automatic mode, CSP moves the instance from one node to another. While moving, the challenge is to ensure confidentiality, integrity, and availability of other users’ data.

- **Server Farming Technique:** In which can be used to re-routing request between user and node.

- **Failover Technique:** Where there is at least one server that is replicating another. There are three ways of failover – Client-based failover, DNS-based failover, and IP-address take over. Address relocation is another technique, which is actually a special case of DNS-based failover. When it is detected that the main computer has failed, the traffic is rerouted to the backup server. However, this technique depends on the success of replication.

- **Sandbox Technique:** It can also isolate an instance by placing it in a sandbox. One approach of creating a sandbox is installing a sandboxing application in cloud operating system. Another approach is creating a virtual box around an instance and observes all the communication channel.

- **Man in the Middle (MITM) Technique:** It can be by placing a MITM between cloud instance and hardware. In that manner, can get log information from the network, RAM, CPU, and hard drive. To get advantage from this technique, the CSP should embrace this technique for implementation in its cloud infrastructure.

- **Forensic Analysis for Cloud Storage Services:** Criminals can keep their secret files such as terrorist documents and child pornography in cloud storage and can destroy all evidence from the local storage to be clean. Cloud storage services are part of
cloud computing services that are subject to exploit by attackers to steal or modify data in cloud storage. Store data in the cloud storages that are remotely distributed on cloud servers in overseas jurisdictions rather than in local machines make a new challenge for forensic practitioners and law enforcement agencies to acquire digital evidence for analysis and examination in a forensically manner to be admissible in the court of law. Quick and Choo [27]-[30] make study on data remnants on client devices and found that there is information in cloud storage accounts(i.e. Dropbox, Microsoft SkyDrive and Google Drive) which is not available on user machine which may either accessed an account through web browser or is synchronized to an account using the client software[30]. This information includes previous and historical versions of files and information that identify the cloud storage user such as computer name, IP address, times and dates associated with the modification made in his account’s contents.

Shams and Hassan [13], they summarize some of the challenges in three cloud service models in Table 2.2.

**Table 2.2: Some of the Challenges in Cloud Forensics [13].**

<table>
<thead>
<tr>
<th>Challenges of Cloud Forensics</th>
<th>Exists in IaaS</th>
<th>Exists in PaaS</th>
<th>Exists in SaaS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical inaccessibility</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Dependence on CSP</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Volatile Data</td>
<td>√</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Trust Issue</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Large bandwidth</td>
<td>√</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Multi-tenancy</td>
<td>√</td>
<td>√</td>
<td>x</td>
</tr>
<tr>
<td>Decentralization of Logs</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Volatility of logs</td>
<td>√</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Logs in multiple tiers and layers</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Accessibility of logs and Depending on CSP for logs</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Chain of Custody</td>
<td>√</td>
<td>√</td>
<td>x</td>
</tr>
<tr>
<td>Problem of current forensic tools</td>
<td>√</td>
<td>√</td>
<td>x</td>
</tr>
<tr>
<td>Crime scene reconstruction</td>
<td>√</td>
<td>√</td>
<td>x</td>
</tr>
<tr>
<td>Cross-border law</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>
2.2.2 Existing Systems and Current Solutions

Shams Zawoad and Ragib Hasan [13] they mentioned and discussed some existing proposed solutions related to Cloud Forensics, which can alleviate some of the cloud forensics challenges as follow in Table 2.3.

Table 2.3: Analysis of Challenges and Proposed Solutions [13].

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Proposed Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trust issue for depending on CSP</td>
<td>Trust Model</td>
</tr>
<tr>
<td>Preserving integrity</td>
<td>Distributed signature detection framework</td>
</tr>
<tr>
<td>Decentralization of logs, logs in multiple tiers and layers, absence of</td>
<td>Log management solution</td>
</tr>
<tr>
<td>critical information in logs, Volatility of logs</td>
<td></td>
</tr>
<tr>
<td>Depending on CSP for logs</td>
<td>API provide by CSP for logs</td>
</tr>
<tr>
<td>Dependability on CSP for data acquisition</td>
<td>Cloud management plane</td>
</tr>
<tr>
<td>Compliance issue, dependability on CSP</td>
<td>Robust SLA</td>
</tr>
<tr>
<td>Compliance issue, Developing a robust SLA</td>
<td>Trusted third party</td>
</tr>
<tr>
<td>Cross-border law</td>
<td>Global unity</td>
</tr>
<tr>
<td>Live forensics issue</td>
<td>Virtual machine introspection</td>
</tr>
<tr>
<td>Volatile Data</td>
<td>Continuous synchronization</td>
</tr>
<tr>
<td>Trust issues of cloud computing</td>
<td>Trusted platform module (TPM)</td>
</tr>
<tr>
<td>Multi-tenancy issue</td>
<td>Isolating a cloud instance</td>
</tr>
<tr>
<td>Chain of custody</td>
<td>Data provenance in cloud</td>
</tr>
</tbody>
</table>

2.2.3 Research Tools and Test Environments

Researchers and investigators who work in digital forensic investigation field, they used many forensic tools and test environments to perform digital forensic investigation process. There are many forensic tools that are used in computers forensic such as Encase and FTK.

In cloud forensic area, there is research desired to either modify existing forensics tools or to present new tools tailored to meet cloud forensics requirements. For example, outsourcing infrastructure in most of the cloud service models raises the need for tools that are capable of performing analysis and examination using a secure remote connection. Furthermore, it needs digital forensics tools to acquire and process memory and network dumps. A conclusion of most of the tools used to perform digital evidence extraction or analysis are tabulated in Table 2.4.

To evaluate the theoretical methods, scientists and researchers need to recognize an appropriate cloud test environment for their experiment testbed. Table 2.5 summarizes test
environments that used that could suit a cloud computing projects in general and cloud forensics scenarios in particular. These environments are prepared and developed with the cloud computing structure and characteristics in mind.

**Table 2.4:** Summary of Digital Forensic Tools [31].

<table>
<thead>
<tr>
<th>Used Tool(s)</th>
<th>Possible Usability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wireshark [33]</td>
<td>Captures network traffic between VM and the CSP</td>
</tr>
<tr>
<td>Microsoft Expression Encoder4 [34]</td>
<td>VM windows video recorder</td>
</tr>
<tr>
<td>FTK Remote Agent [37]</td>
<td>Digital evidence acquisition remotely</td>
</tr>
<tr>
<td>Encase Remote Agent [37]</td>
<td></td>
</tr>
<tr>
<td>X-Ways [38]</td>
<td>Acquisition of Windows and Linux live system</td>
</tr>
<tr>
<td>Sleuth kit Hadoop [39]</td>
<td>Faster processing of video files for forensic acquisition and analysis</td>
</tr>
<tr>
<td>FROST [37]</td>
<td>Digital forensics tools for the OpenStack cloud platform</td>
</tr>
<tr>
<td>XenAccess [40]</td>
<td>Xen VM introspection library (Hypervisor level)</td>
</tr>
<tr>
<td>VMWatcher [41]</td>
<td>VMware VM introspection (Hypervisor level)</td>
</tr>
<tr>
<td>VMwall [42]</td>
<td>VMware VM introspection (VM level)</td>
</tr>
</tbody>
</table>

**Table 2.5:** Summary of Test Environments used for Cloud Forensics [31].

<table>
<thead>
<tr>
<th>Cloud Test Environment</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BonFire [43]</td>
<td>An EU project enables operating a multi-site cloud-based facility on top of different infrastructure testbeds such as Emulab.</td>
</tr>
<tr>
<td>Eucalyptus [44]</td>
<td>A software used to build Amazon Work Station (AWS) private and public cloud.</td>
</tr>
<tr>
<td>OpenNebulla [45]</td>
<td>An industry standard used to provide virtual data centres and IaaS.</td>
</tr>
<tr>
<td>CloudSim [46]</td>
<td>A solution to create large-scale cloud computing data center, virtual hosts, and capability of analysis for network traffic.</td>
</tr>
<tr>
<td>Emulab [47]</td>
<td>Public facility available for researchers to develop, debug and evaluate their systems.</td>
</tr>
<tr>
<td>OpenStack [48]</td>
<td>A project used to create various IaaS architectures such as storage, compute and network.</td>
</tr>
<tr>
<td>Rackspace [49]</td>
<td>Based on OpenStack and provides IaaS.</td>
</tr>
<tr>
<td>Amazon[50],[51],[52]</td>
<td>An appropriate solution provides a various flavour of IaaS, Amazon Simple Store Service (S3) as storage, Amazon Elastic Comput Cloud (EC2) as a computation required for AWS. Amazon Elastic Block Store (EBS) used for backups.</td>
</tr>
</tbody>
</table>
2.2.4 Evaluation of Current Forensic Tools in Cloud

Digital investigators use the most popular and proven digital forensics tools such as Accessdata FTK, Encase, and others. However, the data collection procedure is unlike in a cloud environment compared to conventional computer forensics. These tools can be used for data collection from cloud environment but with some modification to be suitable for the new environment (i.e. Cloud). So far, there has been a few work that evaluates the capability of some available forensic tools in a cloud environment such the work proposed by Dykstra et al. [14].

To evaluate the capability of these forensic tools such as Encase and Accessdata FTK, to be used in the cloud. Dykstra et al. [14] chose Amazon EC2 for their experiment and used EnCase and Accessdata FTK to remotely collect and acquire digital evidence in the cloud. They conducted three experiments to collect data from three different layers and got success in all the experiments as follow:

- **In the first experiment:** They remotely collected forensic data from the guest OS layer of cloud. They used the Encase Servlets and FTK Agents programs to communicate and collect data remotely.

- **For the second experiment:** They used Eucalyptus cloud platform and gathered the data from the virtualization layer.

- **In the third experiment:** They verified the data acquisition at the host operating system layer by Amazon’s export feature. They found that though it is possible to export data from S3, and it is not possible from EBS.

2.2.5 Comparison between Traditional Computer Forensic and Cloud Forensic

In Table 2.6, a comparison study between Classic Forensic and cloud forensic is tabulated to explain the differences between both above two types of forensics.
Table 2.6: Comparison between Computer Forensic and Cloud Forensic.

<table>
<thead>
<tr>
<th></th>
<th>Classic Forensic</th>
<th>SaaS</th>
<th>PaaS</th>
<th>IaaS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access Control</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Application</td>
<td>√</td>
<td>X</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Database</td>
<td>√</td>
<td>X</td>
<td>X</td>
<td>√</td>
</tr>
<tr>
<td>Operating System</td>
<td>√</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Compute</td>
<td>√</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Storage</td>
<td>√</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Network</td>
<td>√</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

2.3 Chapter Summary

In recent times, with the rapid adoption of cloud computing technology, a serious need has arisen for applying digital forensic science to this area. The validity and reliability of performing the digital forensic process are important in this new perspective and needs new procedures for investigation of crimes and incidents in multi-tenant cloud environments that deliver quick provisioning, universal elasticity, and broad-network accessibility. This chapter provided the basic concepts of cloud computing, digital forensic, cybercrimes and cloud forensic as well as introduced a literature review about research work included challenges and current solutions related digital forensics techniques in cloud computing environment.