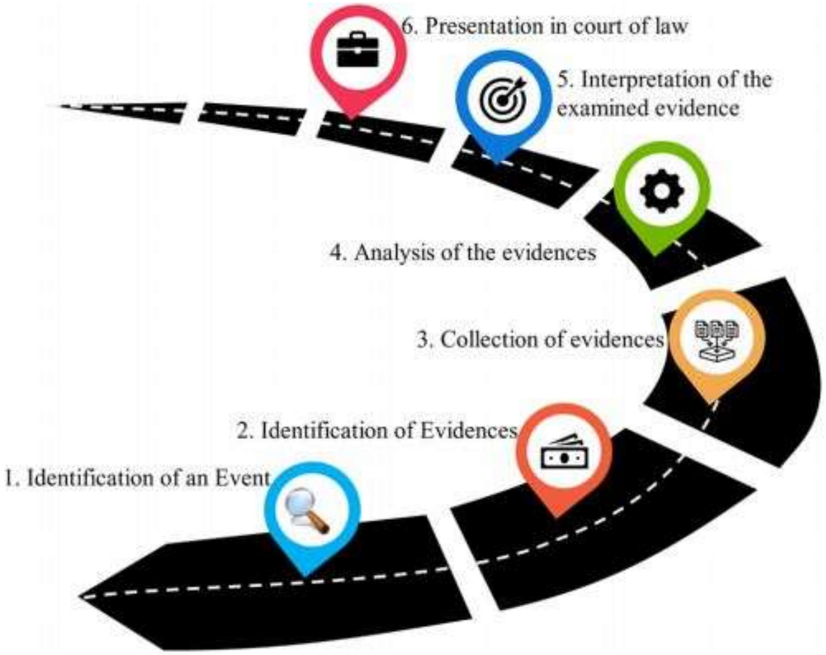
Cloud Forensics

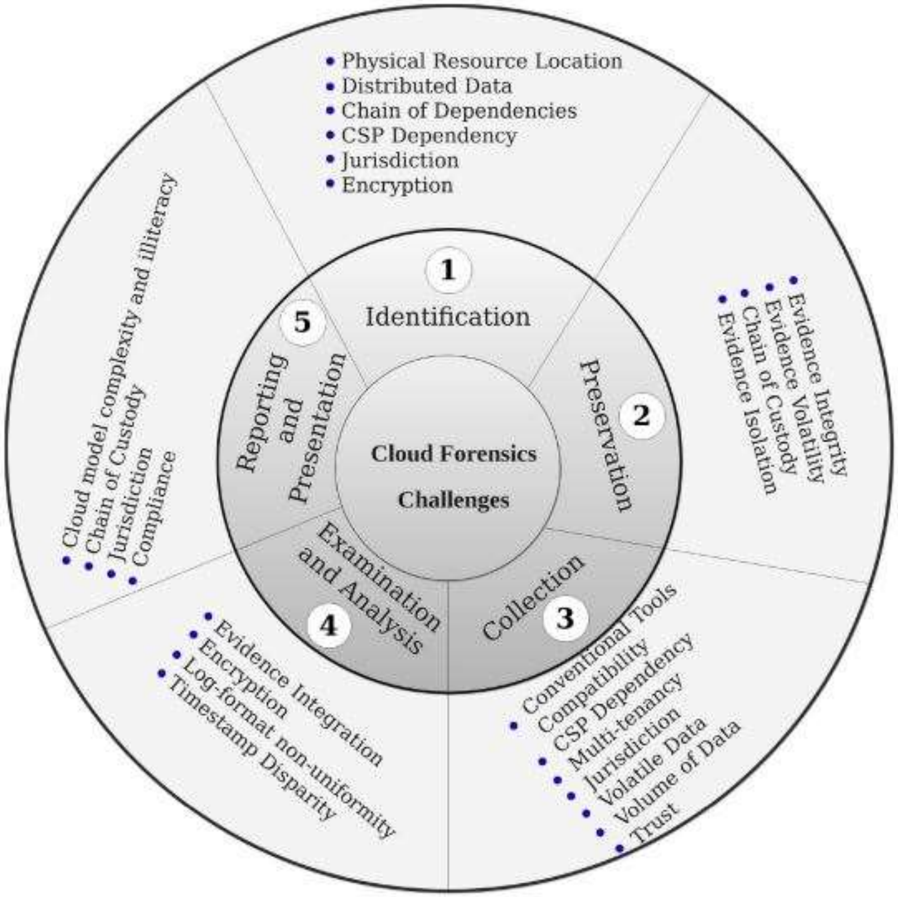
Digital Forensics


6. Presentation in court of law
5. Interpretation of the examined evidence
4. Analysis of the evidences
3. Collection of evidences
2. Identification of Evidences
1. Identification of an Event

Why Cloud Forensics?

- The cloud infrastructure and their services may be used in the commission of a conventional physical crime (e.g., storage of incriminating materials in the cloud, relating to a physical terrorist attack) and hence, subject to a forensic investigation.
- Cloud forensics is more complex than the typical computing device and mobile application (app) forensics, partly due to the architectural complexity of the cloud (an amalgamation of diverse technologies, such as virtualization of resources, utility computing based on remotely available resources, and distributed systems) .
- There have been a number of challenges noted in the literature, such as lack of compatibility and reliability of existing tools and mechanisms to facilitate cloud forensic investigations, and evidence sources can be at the client-side, network-side (e.g., for data-in-transit), and server-side (cloud service provider (CSP)) .

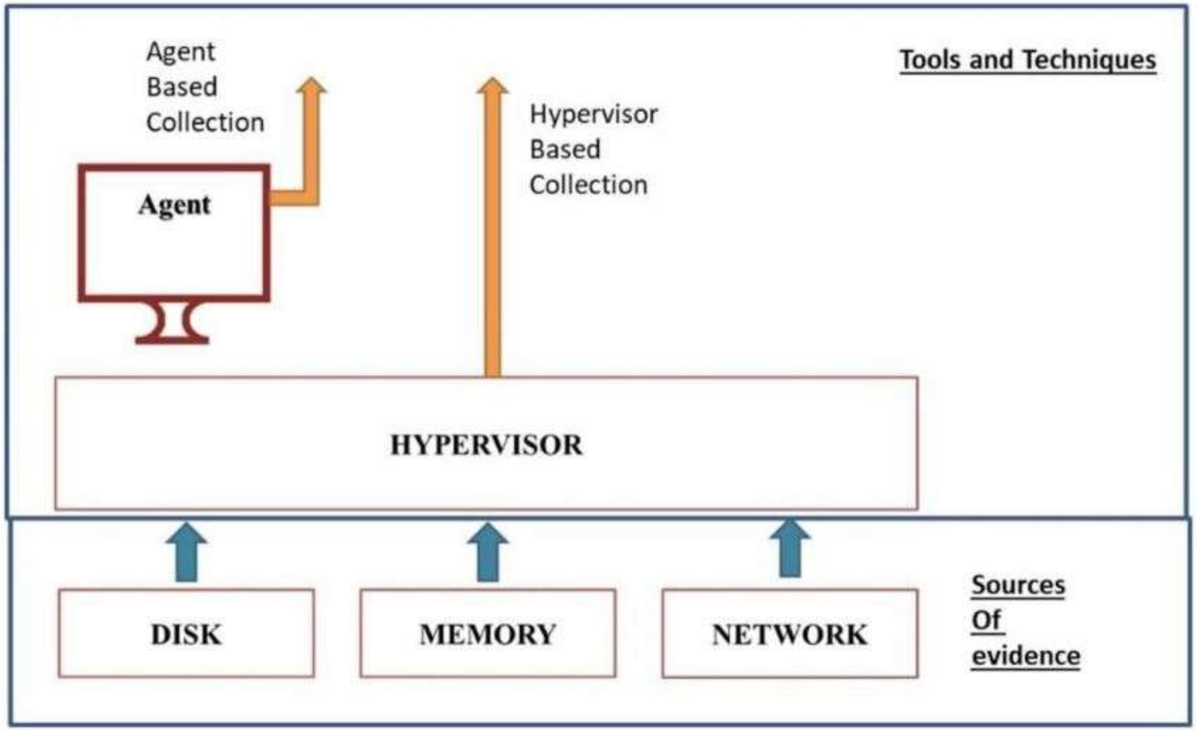
Why Cloud Forensics?



Cloud Forensics : Evidence

- Disk
- Network
- Memory

Tools and Techniques for Collection of the Evidence



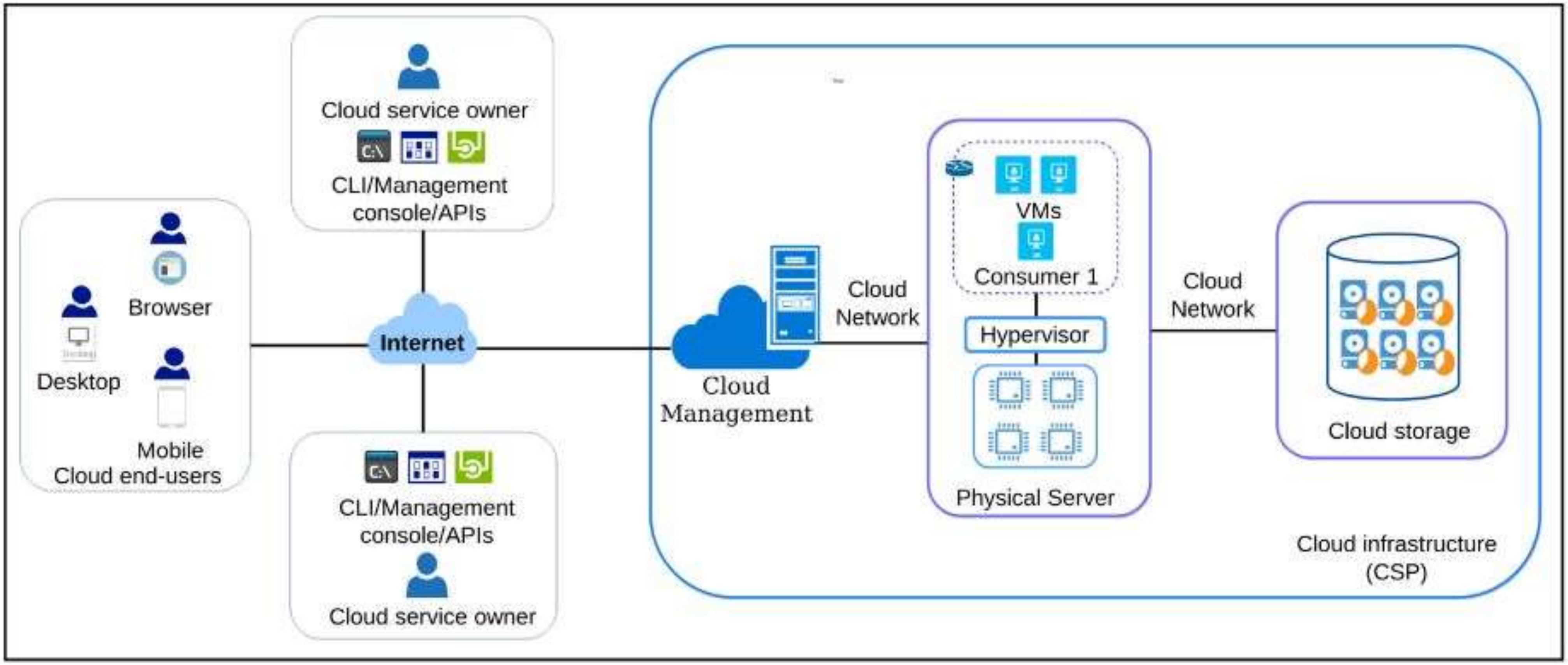
Tools and Techniques for Collection of the Evidence

- Hypervisor Based Collection:
  - The hypervisor based evidence collection goes one step further than the IDS) at the hypervisor . IDS can be used to trigger the collection mechanism. The hypervisor can provide the functionalities of gathering information and data about the virtualized environment. This is a powerful approach as the control provided has the highest level privileges. At the same time, this approach requires additional functionality to be installed at the kernel level hence a kernel-level programming approach is needed.

Tools and Techniques for Collection of the Evidence

- Third-Party Agent-Based Collection
  - An evidence collection agent is deployed on the resource to collect the data. This agent does not have high-level privileges and hence requires the permission of the owner of the resource and services on which it is deployed. The agent-based collection can use adaptive cloud deployment methods to be robust. Implementation of the third-party agent is easier than the Hypervisor based collection technique. However, an adversary using the cloud resources for exploiting any vulnerability can tamper with the agent and compromise the evidence acquisition.

Forensic stakeholders in cloud eco-system.



Forensic stakeholders in cloud eco-system.

- **Cloud End-user.** Cloud end-users (service users) utilize cloud applications, which include all the web-services hosted in a cloud. Examples include cloud-hosted websites, e-commerce portals, email, cloud storage, and content sharing applications. End-users utilize the services over the Internet through a browser, desktop application, or a phone-based app/browser. Each access method leaves data artifacts/remnants on the end-user machine due to the various interactions with cloud-based services. Identification of these artifacts may provide forensic investigators a potential evidence on an end-user machine.
- **Cloud Service Owner.** Cloud service owners are VM owners that host particular server/service and has complete control over the execution environment. Cloud service owners usually manage their cloud services through three CSP-provided mechanisms: CLI (command line interface), management console, and APIs. In the scenario of non-cooperative service owner, forensic investigators may have to rely upon CSPs to acquire the artifacts.
- **CSP (Hypervisor).** Many of the artifacts that are part of traditional OS environments are also available in hypervisors. Event logs/Syslog, audit logs, and messages are the fundamental artifacts available in a hypervisor. These logs provide various information about individual events of the VM instances. In the cloud, hypervisor access is not available to cloud consumers and restricted to the CSP only. To collect low-level system information, forensic investigators must depend on the cooperation of CSP.
- **Virtual Machine.** Virtual machines provide a rich source of artifacts to the investigators as it has a complete operating system and all the related services. VM owners can access these artifacts using three methods: CLI, dashboard, and APIs. There are guest OS files that provide all the information about the user’s activities and data. There are numerous logs associated with various applications and background processes, including databases and transaction logs. Logs associated with different security measures such as firewall, anti-virus software, and IDS/IPS may provide information about malicious incoming and outgoing activities of a virtual machine.

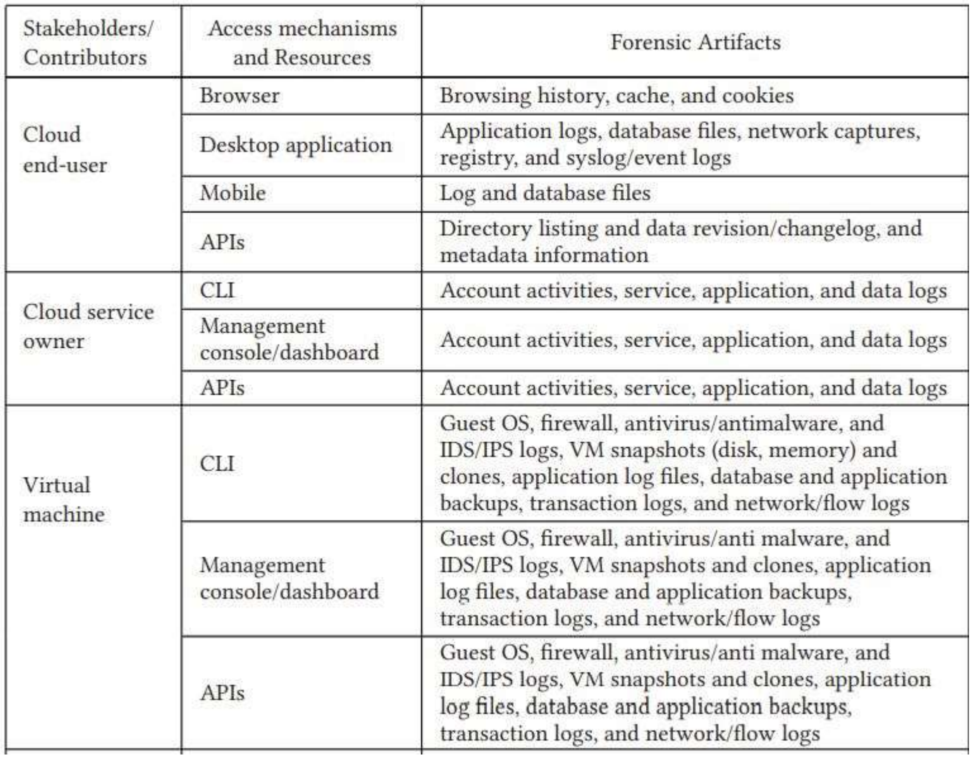
Forensic stakeholders in cloud eco-system.

- **CSP (Cloud Management Software).** Examples of cloud management software include OpenStack and Eucalyptus. These software suites help in establishing cloud platforms that manage computing, networking, and storage resources in a data center along with features such as accounting, elasticity, and resource allocation. All these software are part of the cloud infrastructure and has a control of CSP, which results in their CSP dependency.

- **CSP (Cloud Storage Management).** Logs associated with various storage management software may aid the investigators with access information of various storage devices. As part of storage devices, backups, archives, VM snapshots, and replicas/clones may provide information about the residing content for each user. Database information is another important source of artifacts in cloud storage. Forensic investigators must acquire these artifacts from the CSP, which marks the CSP dependency for having artifact acquisition for this category.

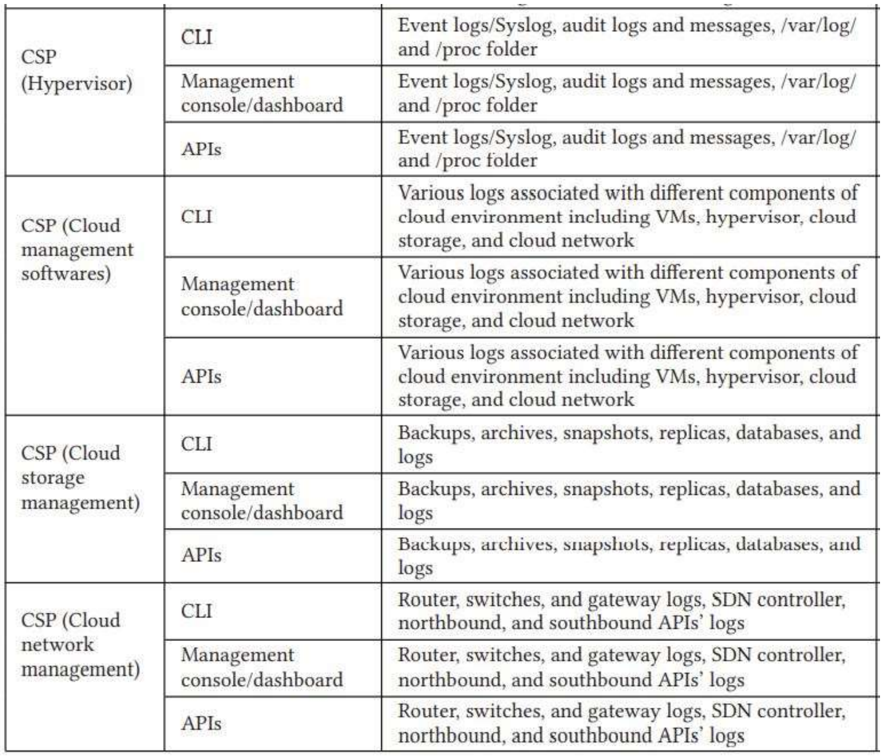
- **CSP (Cloud Network Management).** In today’s time, CSPs’ infrastructure spans to tens of data centers consisting of hundreds of devices including compute, storage, and network devices. From the forensics’ point of view, logs associated with each network device are fundamental artifacts in the CSP network. In the case of SDN, SDN controller logs are important as SDN manages complete networking infrastructure from a single location including northbound and southbound APIs. Logs associated with these APIs may also provide a detailed history of communication between various devices. As for the dependency, again it is CSP dependent as CSP owns and controls the complete cloud side network infrastructure.

Location of Forensic Artifacts in Each Forensic Stakeholder



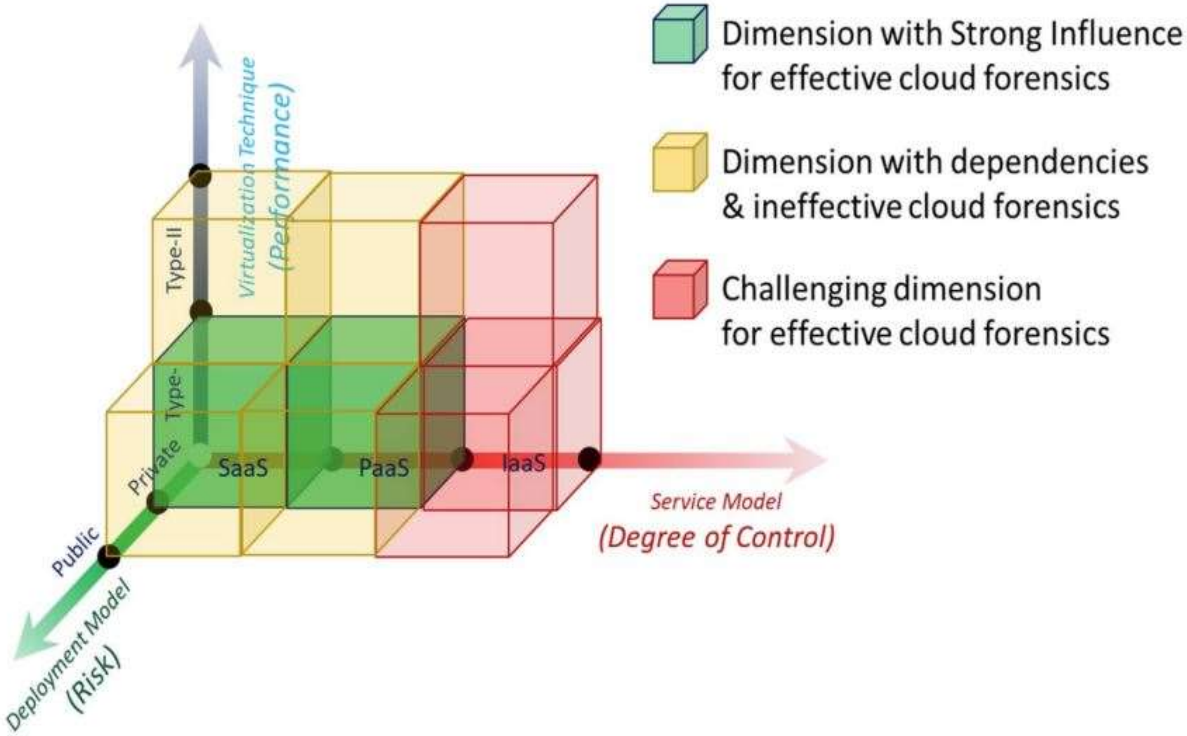
Stakeholders/ Contributors	Access mechanisms and Resources	Forensic Artifacts
Cloud end-user	Browser	Browsing history, cache, and cookies
	Desktop application	Application logs, database files, network captures, registry, and syslog/event logs
	Mobile	Log and database files
	APIs	Directory listing and data revision/changelog, and metadata information
Cloud service owner	CLI	Account activities, service, application, and data logs
	Management console/dashboard	Account activities, service, application, and data logs
	APIs	Account activities, service, application, and data logs
Virtual machine	CLI	Guest OS, firewall, antivirus/antimalware, and IDS/IPS logs, VM snapshots (disk, memory) and clones, application log files, database and application backups, transaction logs, and network/flow logs
	Management console/dashboard	Guest OS, firewall, antivirus/anti malware, and IDS/IPS logs, VM snapshots and clones, application log files, database and application backups, transaction logs, and network/flow logs
	APIs	Guest OS, firewall, antivirus/anti malware, and IDS/IPS logs, VM snapshots and clones, application log files, database and application backups, transaction logs, and network/flow logs

Location of Forensic Artifacts in Each Forensic Stakeholder



CSP (Hypervisor)	CLI	Event logs/Syslog, audit logs and messages, /var/log/ and /proc folder
	Management console/dashboard	Event logs/Syslog, audit logs and messages, /var/log/ and /proc folder
	APIs	Event logs/Syslog, audit logs and messages, /var/log/ and /proc folder
CSP (Cloud management softwares)	CLI	Various logs associated with different components of cloud environment including VMs, hypervisor, cloud storage, and cloud network
	Management console/dashboard	Various logs associated with different components of cloud environment including VMs, hypervisor, cloud storage, and cloud network
	APIs	Various logs associated with different components of cloud environment including VMs, hypervisor, cloud storage, and cloud network
CSP (Cloud storage management)	CLI	Backups, archives, snapshots, replicas, databases, and logs
	Management console/dashboard	Backups, archives, snapshots, replicas, databases, and logs
	APIs	Backups, archives, snapshots, replicas, databases, and logs
CSP (Cloud network management)	CLI	Router, switches, and gateway logs, SDN controller, northbound, and southbound APIs' logs
	Management console/dashboard	Router, switches, and gateway logs, SDN controller, northbound, and southbound APIs' logs
	APIs	Router, switches, and gateway logs, SDN controller, northbound, and southbound APIs' logs

Dimensions of cloud forensics


Dimension with Strong Influence for effective cloud forensics
Dimension with dependencies & ineffective cloud forensics
Challenging dimension for effective cloud forensics
Virtualization Technique
(Performance)
Type-II
Type-
Private
Public
Deployment Model
(Risk)
SaaS
PaaS
IaaS
Service Model
(Degree of Control)

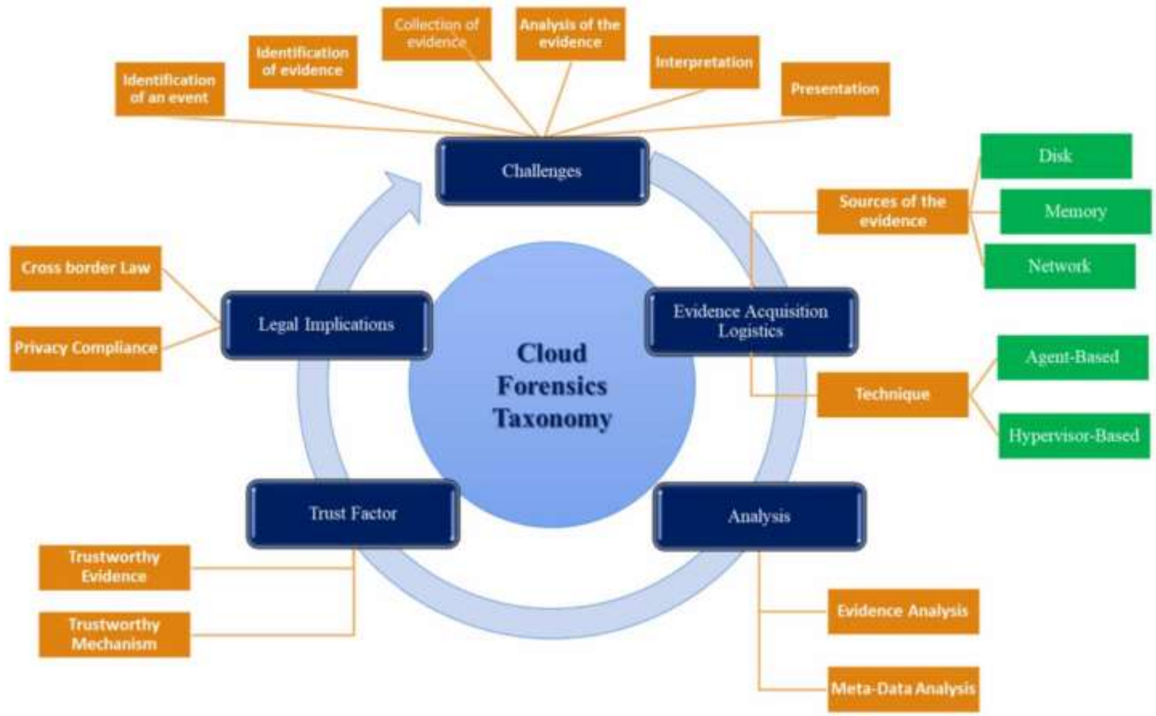
Dimensions of cloud forensics

- **Service Model Dimension** Most of the researchers have presented their studies considering the degree of control issue which arises in different service models. The degree of control is maximum for the CSP in SaaS model and minimum in IaaS model. SaaS model allows CSP to manage up to the application-layer, PaaS model allows CSP to manage platform-layer and IaaS model allows CSP to manage just up to the virtualization layer. Degree of control is directly associated with evidences that can be collected.

- **Deployment Model Dimension** The deployment models also have an impact on cloud forensics. Public deployment model services are exposed to the world wide web which make them more vulnerable than private deployment service models. The nature of private cloud is single tenant compared to public cloud which could be multi-tenant. Multi-tenant environment is more challenging for the cloud forensics investigation as it involves segregating the naïve user's data. Physical location of public cloud is not likely to be physically accessible unlike private cloud. Considering the popularity of public cloud it is evident that remote acquisition solutions will be in demand over solution that works only with physical confiscation of evidence.

- **Virtualization Model Dimension** The virtualization implementation has an impact on the performance of the cloud. Vulnerabilities and security risks of the host operating system in a hardware assisted virtualization affects the entire system including the hypervisor. Private SaaS and PaaS model deployed with Type-I or bare metal hypervisor is an optimal environment for cloud forensics, Whereas IaaS model for same public cloud with bare metal have challenges of evidence access. Public cloud model and hardware assisted virtualization makes cloud forensics dependent on CSP and inefficient respectively. These dimensions of cloud are good way to get a perspective of the scope of cloud forensic solution.

Cloud Forensic



Cloud Forensic Challenges

- **Identification of an Occurrence of an Event** In the traditional forensics process, incidence detection is a straight forward process as the resource (target) is a simple system. Cloud services operate as collaborative services; cloud implementation is complex. Identification of malicious activity may go unnoticed for a duration long enough for the adversary to hide or go away. Cloud forensic comes after the event has already happened. However, if an Intrusion Detection Mechanism (IDS) is provided for incident identification in the architecture of the cloud, it can make the forensics easier and mitigate the challenges of the identification phase.

- **Identification of Evidence Related to the Event** Evidence can be defined as electronic evidence (EE) if it is a piece of information either stored in a standalone device or transmitted through network resources. The information EE has is sensitive and can be easily altered, damaged or destroyed by improper handling or examination making it very fragile. Due to such nature of EE, it poses a huge challenge for its admissibility in court of law. For this purpose, ISO/IEC 27,037:2012 provides guidelines for handling of the digital evidence. EE should satisfy the legal criteria and it should be identifiable in the cloud environment. Not every event-related data is evidence. The evidence should be identified from the pool of data that is associated with the event. Such identification of the evidence in a cloud platform requires special architectural support

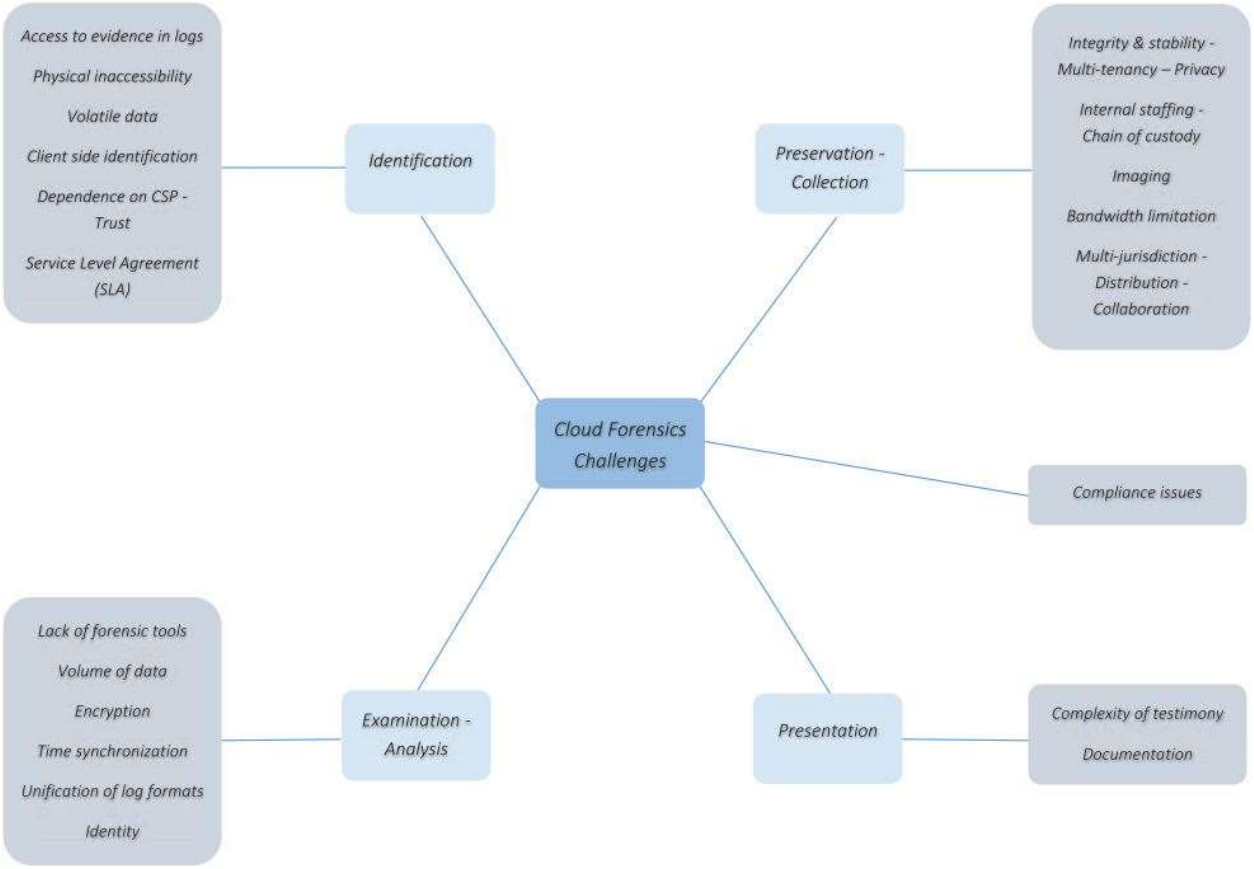
- **Collection of Identified Evidence** Once the evidence is identified the next stage in a forensics process is the collection of that evidence. Cloud forensics faces many challenges in the collection of evidence. These challenges can be classified but not restricted to physical inaccessibility, remote acquisition, access to cloud network devices, volatile data, access to data, evidence, logs on the cloud. Details of log forensics are presented thoroughly in. These details also suggest an automated log analysis tool to overcome some challenges and vulnerabilities in log forensics. CSPs might not have a mechanism to store the logs or the logs that are stored by the CSPs might not contribute as evidence. Data that is logged on a cloud platform has no standard format . In case evidence is identified and needs to be collected, the volume of the evidence can pose a challenge. High volume data needs high bandwidth for data collection from the cloud .

Cloud Forensic Challenges

- Analysis of the Collected Evidence Data

- Analysis is a process of examining the evidence for proving or disproving an event. Traditional digital forensics makes use of forensics tools such as Forensics Tool Kit (FTK), sleuth kit, volatility, CAINE, Autopsy, etc. However, using these traditional forensics tools in a cloud forensics investigation to analyze the virtual environment of cloud computing is a challenging task. A multi-tenant environment of the cloud allows users to share the services with some degree of isolation of their data. Undertaking a forensic investigation in a multi-tenant environment is complicated. Anti-forensics techniques are being used by criminals on the cloud.

Cloud Forensic Challenges



Cloud Forensics Challenges

Identification

- Access to evidence in logs
- Physical inaccessibility
- Volatile data
- Client side identification
- Dependence on CSP - Trust
- Service Level Agreement (SLA)

Preservation - Collection

- Integrity & stability - Multi-tenancy – Privacy
- Internal staffing - Chain of custody
- Imaging
- Bandwidth limitation
- Multi-jurisdiction - Distribution - Collaboration

Compliance issues

Examination - Analysis

- Lack of forensic tools
- Volume of data
- Encryption
- Time synchronization
- Unification of log formats
- Identity

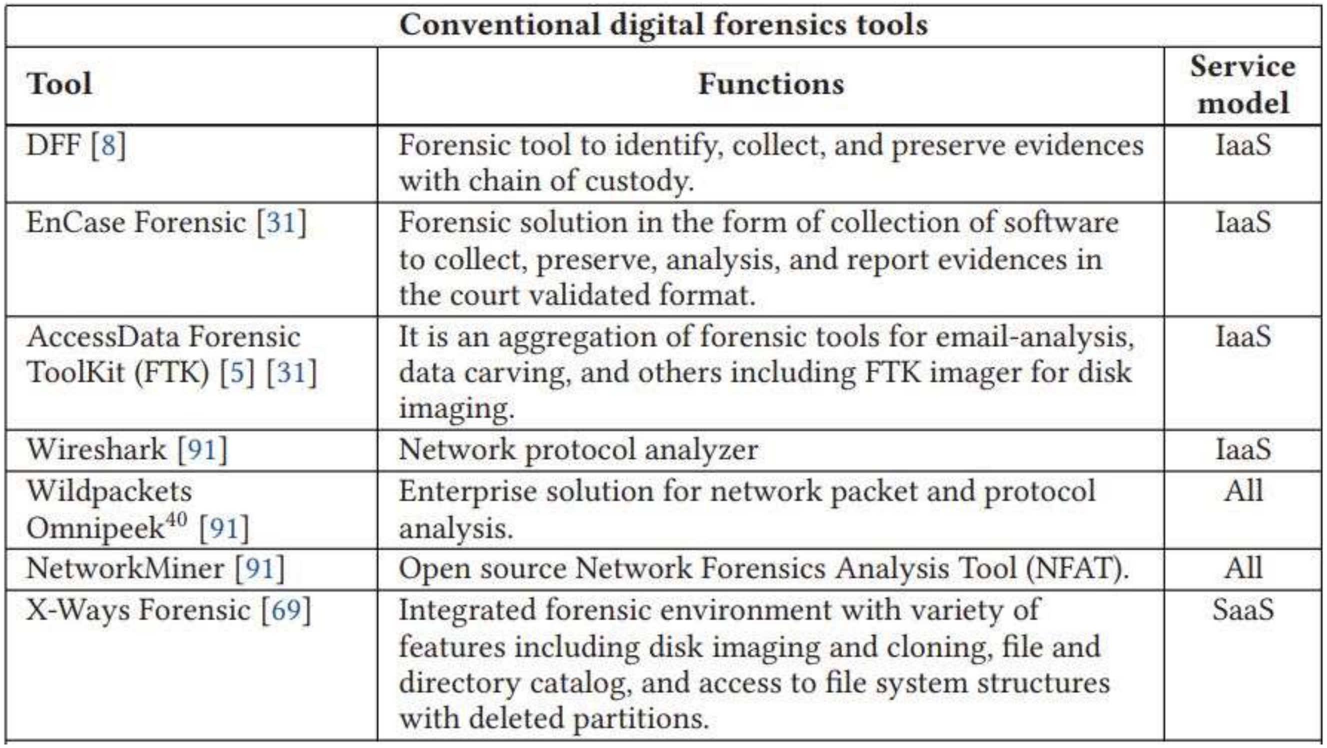
Presentation

- Complexity of testimony
- Documentation

Cloud Forensic Tools

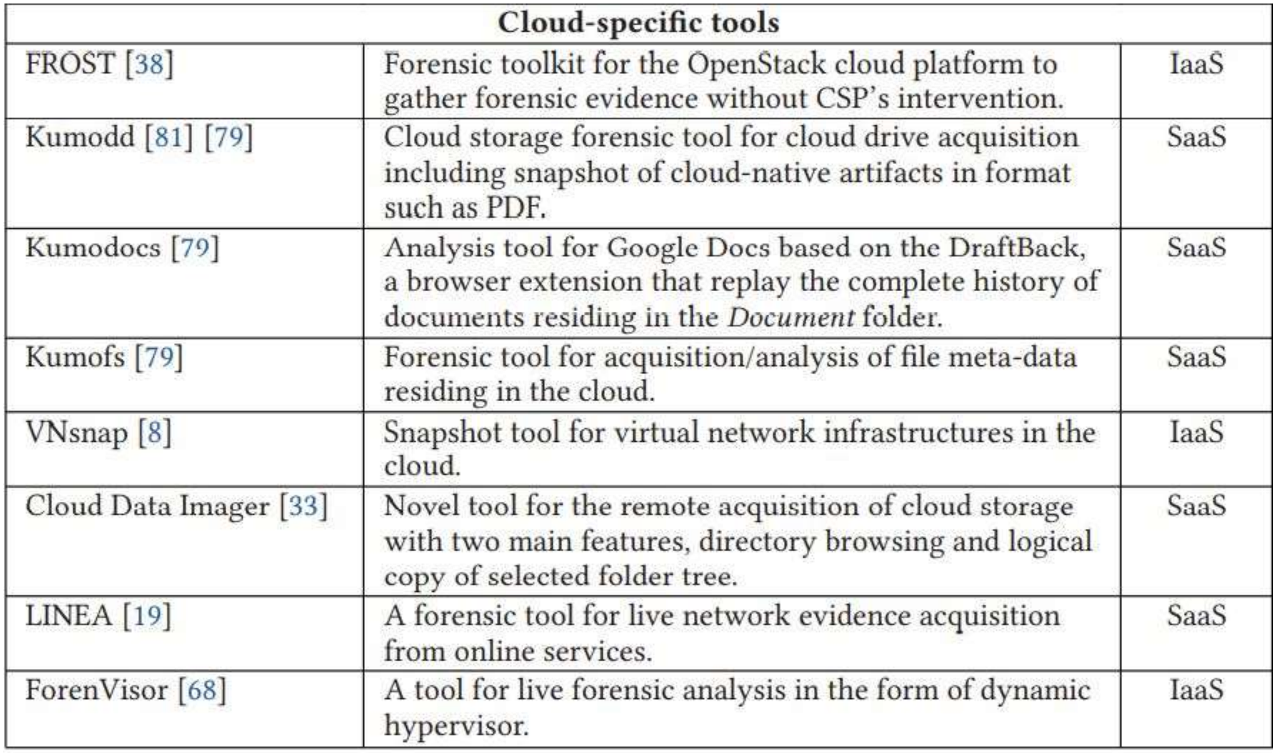
- En-Case and Accessdata FTK tools can be used to acquire evidence, and the results were successful, but too much trust is required.
- On the other hand, tools such as Internet Evidence Finder, and F-Response make use of relevant extensions to recover various cloud and social network related artifacts
- We can design and implement a forensic toolkit in a private instantiation of the OpenStack cloud platform (IaaS), which is called Forensic Open-Stack Tools—It can provide trustworthy forensic acquisition of virtual disks, API logs, and guest firewall logs.

Cloud Forensic Tools



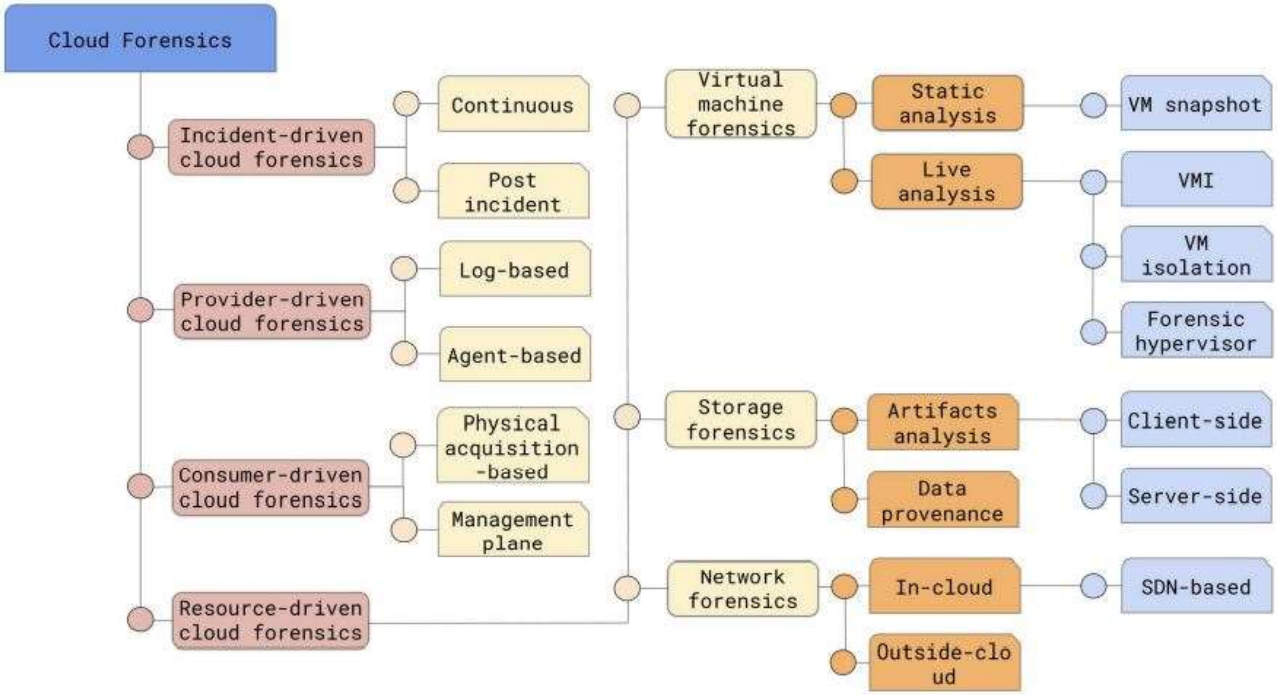
Conventional digital forensics tools		
Tool	**Functions**	**Service model**
DFF [8]	Forensic tool to identify, collect, and preserve evidences with chain of custody.	IaaS
EnCase Forensic [31]	Forensic solution in the form of collection of software to collect, preserve, analysis, and report evidences in the court validated format.	IaaS
AccessData Forensic ToolKit (FTK) [5] [31]	It is an aggregation of forensic tools for email-analysis, data carving, and others including FTK imager for disk imaging.	IaaS
Wireshark [91]	Network protocol analyzer	IaaS
Wildpackets Omnipeek[40] [91]	Enterprise solution for network packet and protocol analysis.	All
NetworkMiner [91]	Open source Network Forensics Analysis Tool (NFAT).	All
X-Ways Forensic [69]	Integrated forensic environment with variety of features including disk imaging and cloning, file and directory catalog, and access to file system structures with deleted partitions.	SaaS

Cloud Forensic Tools



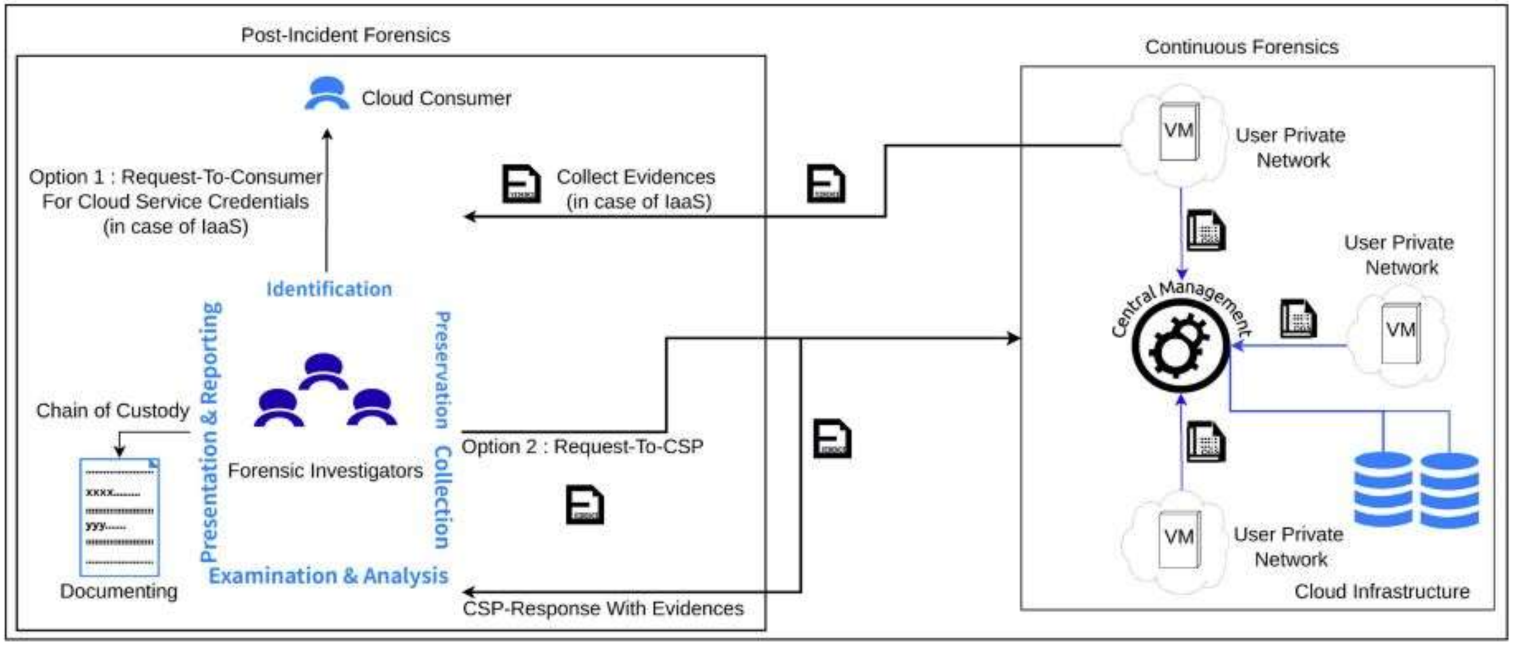
Cloud-specific tools		
FROST [38]	Forensic toolkit for the OpenStack cloud platform to gather forensic evidence without CSP's intervention.	IaaS
Kumodd [81] [79]	Cloud storage forensic tool for cloud drive acquisition including snapshot of cloud-native artifacts in format such as PDF.	SaaS
Kumodocs [79]	Analysis tool for Google Docs based on the DraftBack, a browser extension that replay the complete history of documents residing in the *Document* folder.	SaaS
Kumofs [79]	Forensic tool for acquisition/analysis of file meta-data residing in the cloud.	SaaS
VNsnap [8]	Snapshot tool for virtual network infrastructures in the cloud.	IaaS
Cloud Data Imager [33]	Novel tool for the remote acquisition of cloud storage with two main features, directory browsing and logical copy of selected folder tree.	SaaS
LINEA [19]	A forensic tool for live network evidence acquisition from online services.	SaaS
ForenVisor [68]	A tool for live forensic analysis in the form of dynamic hypervisor.	IaaS

Cloud Forensic Architecture



Cloud Forensic Architecture

- **Incident-driven cloud forensics**: This subcategory includes methods following continuous forensics through central logging. The other subcategory (post-incident forensics) has two options for artifact collection, namely: (1) cloud consumer credentials and (2) by request to the CSP.



Cloud Forensic Architecture

- **Incident-driven cloud forensics:**

1. **Continuous Forensics**

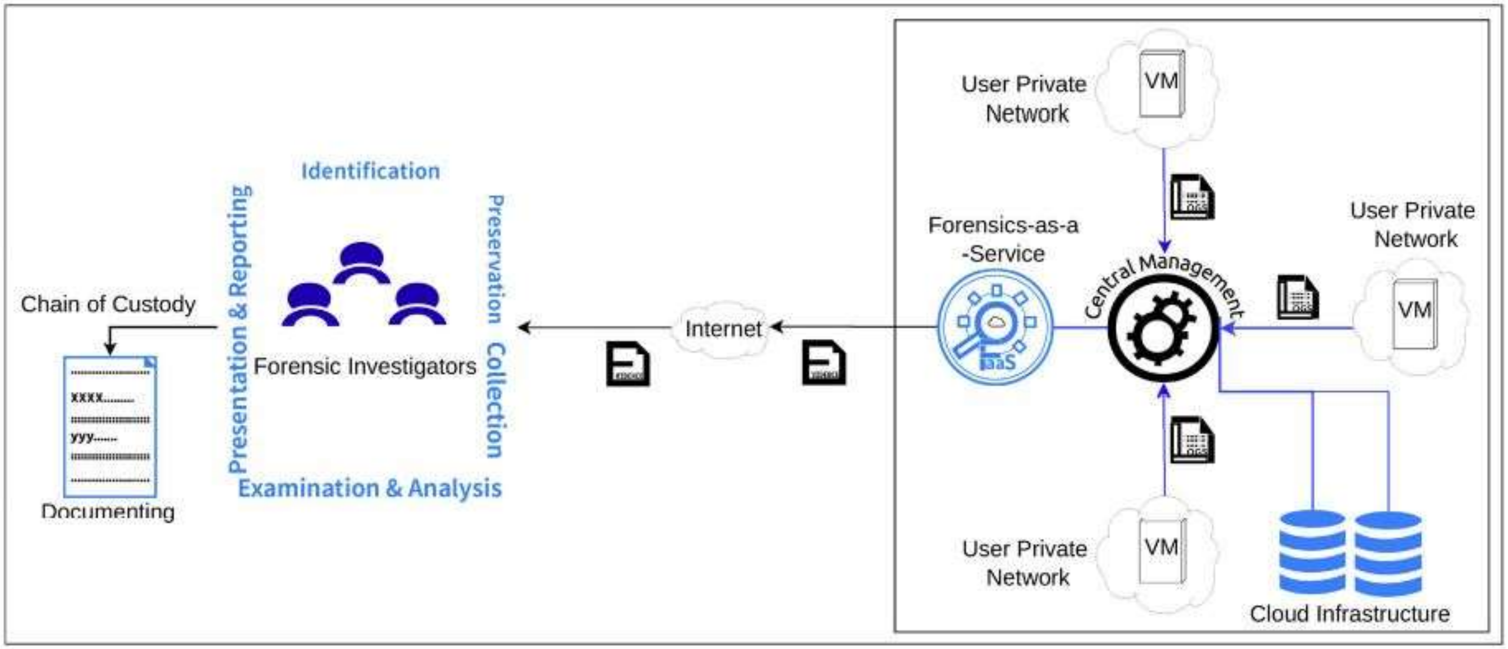
- Continuous cloud forensics includes the evaluation of forensic capabilities for an organization’s cloud infrastructure. Normally, the forensic process is a post-incident evaluation process. However, due to the popularity of cloud services and the constant evolving technological and cyber threat landscape, it is pivotal that forensically friendly cloud are in place for continuous evidence collection, aggregation, and storage

2. **Post Incident Forensics**

- One can use forensic toolkit, which focuses on evidenced acquisition for virtual disks, API logs, and guest firewall logs.
- A forensic hypervisor to perform reliable collection and preservation of evidential data from a compromised system, including malicious guest OS.

Cloud Forensic Architecture

- **Provider-driven cloud forensics**, such as solutions based on forensic-as-a-service through central log management. CSP dependent solutions, which can be categorized into log-based solutions and agent-based solutions



Cloud Forensic Architecture

- **Provider-driven cloud forensics:**

1. Log-based Solutions

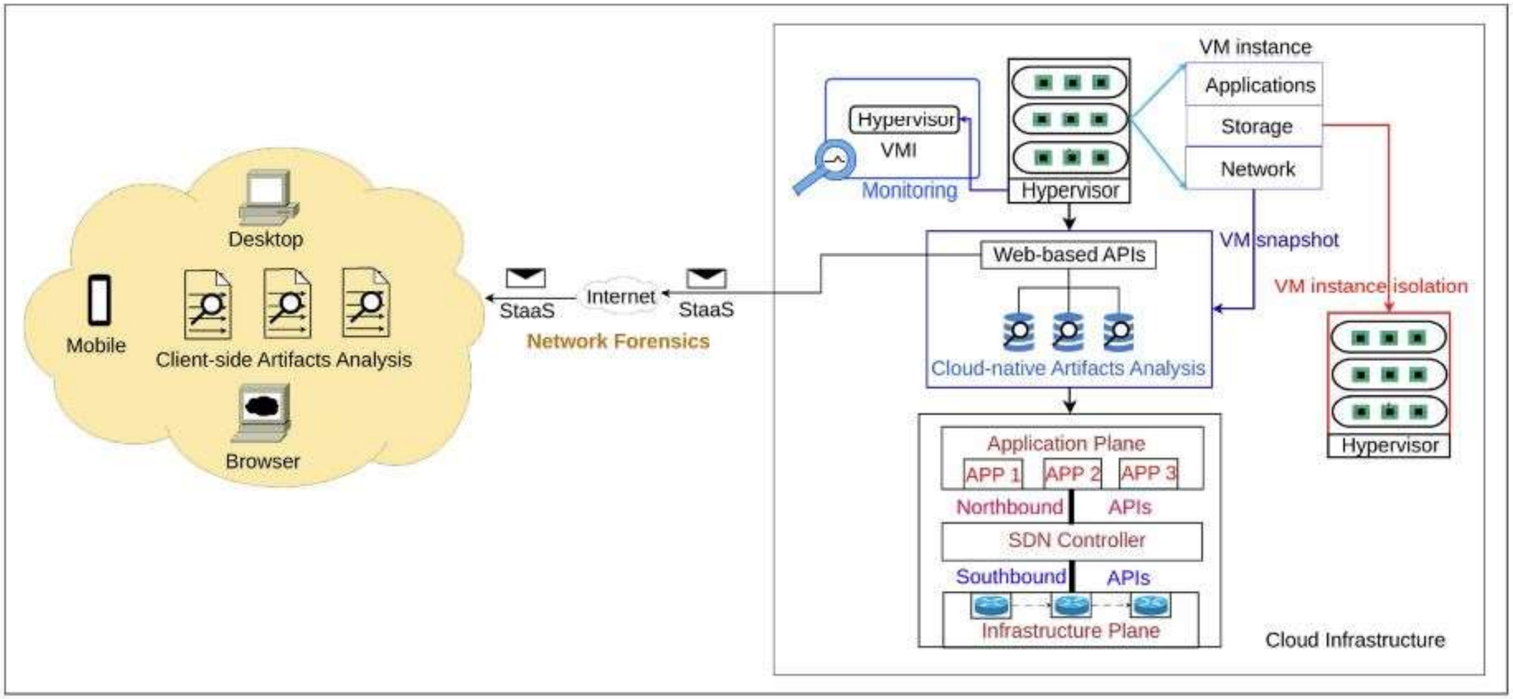
Cloud architecture generates and stores various types of logs including system, network, application, virtual machine, setup, security, web-server, and audit logs at multiple levels (VM, VMM, cloud) and access to these logs depends upon the particular service model.

2. **Agent-based solutions**

- forensic agent in VM that gathers necessary information and send them to the forensic center for storing and further processing
- use of bots and botnets as a forensic agent
- The compromised machines can then form a botnet, a network of (malicious) agents, which perform a set of predefined function in predefined times or on the order of its Command and Control (C&C) server/master
- gathered both volatile and non-volatile data from infected VMs, including network, and sent them to the centralized evidence preservation system. Later, investigators may access the information as Botnet-as-a-Service (BaaS) or as Agent-Based-Solution-as-a-Service (ABSaaS)
- The forensic center is responsible for processing raw forensic data and converting them into a series of tuples for the database storage. Forensic query server provided an interface to query, and analyze the evidential data such as logged users, opened files, network connections, process information, running and auto-run services, and system logs

Cloud Forensic Architecture

- Resource-driven cloud forensics: Marks solution contributions, which include client and server-side artifact analysis, VM instance isolation, VM snapshot, VMI, and potential forensic locations in SDN architecture.



Cloud Forensic Architecture

- Resource-driven cloud forensics: Marks solution contributions, which include client and server-side artifact analysis, VM instance isolation, VM snapshot, VMI, and potential forensic locations in SDN architecture.

- There are three fundamental services most CSPs deliver, compute, storage, and networking. In the cloud, “compute” does not refer to the bare physical hardware, rather virtual hardware in the form of virtual machines having vCPUs, and “storage” and “networking” relates to virtual disks and virtual network. \
- For forensics, focus would be on: SDN (software-defined networking) forensics, second part focuses on the storage forensics with client-side artifacts, and third part focused on VM forensics including VM isolation, VMI (Virtual Machine Introspection), and VM snapshot.

Cloud Forensic Architecture

- VM forensics : We further categorized virtual machine forensics into two sub-categories: static analysis and live analysis. Static analysis is a post-incident analysis process, performed by the forensic investigators once they gather all the relevant pieces of evidence and preserve them in persistent storage. In the cloud, live analysis is also a preferred forensic method as it works without shutting down the cloud/VMs

- VM snapshot. VM snapshot is a process of preserving the runtime state of a virtual machine at a particular point of time. It creates a copy of VM that records the state and the data to safeguard a runtime condition of VM for later restoration. The state of the VM includes the operational state of VM, including all the configuration parameters, and the data includes the disk, memory, and network files. All these contents are the fundamental elements of almost all the digital investigation cases that makes VM snapshot a viable forensic evidence. The available tools are Vnsnap, SNAPS (Snapshots-based Provenance Aware System), HyperShot

- VM isolation. VM isolation is a resource and environment separation process for a compromised or a malicious VM from its neighbored VMs as a security measure and forensic analysis. There's a possibility that if a single VM gets compromised and still is in running state, it may also lead to compromise of other co-hosted VMs, sharing the same hardware. A detachment of a virtual instance from its host environment to a more safe and controlled environment for later examination is a secure measure as it limits the security implications to the affected VM only.

- Virtual Machine Introspection : VMI is an introspection technique to monitor the run time state of a virtual machine at the hypervisor level outside the monitored VM. VMI gained ample attention in the field of computer security, including forensics and became an important method for virtual machine analysis. Forensic-as-a-service through VMI is a promising forensic aid, but access to the hypervisor (CSP-owned) is a practical concern.

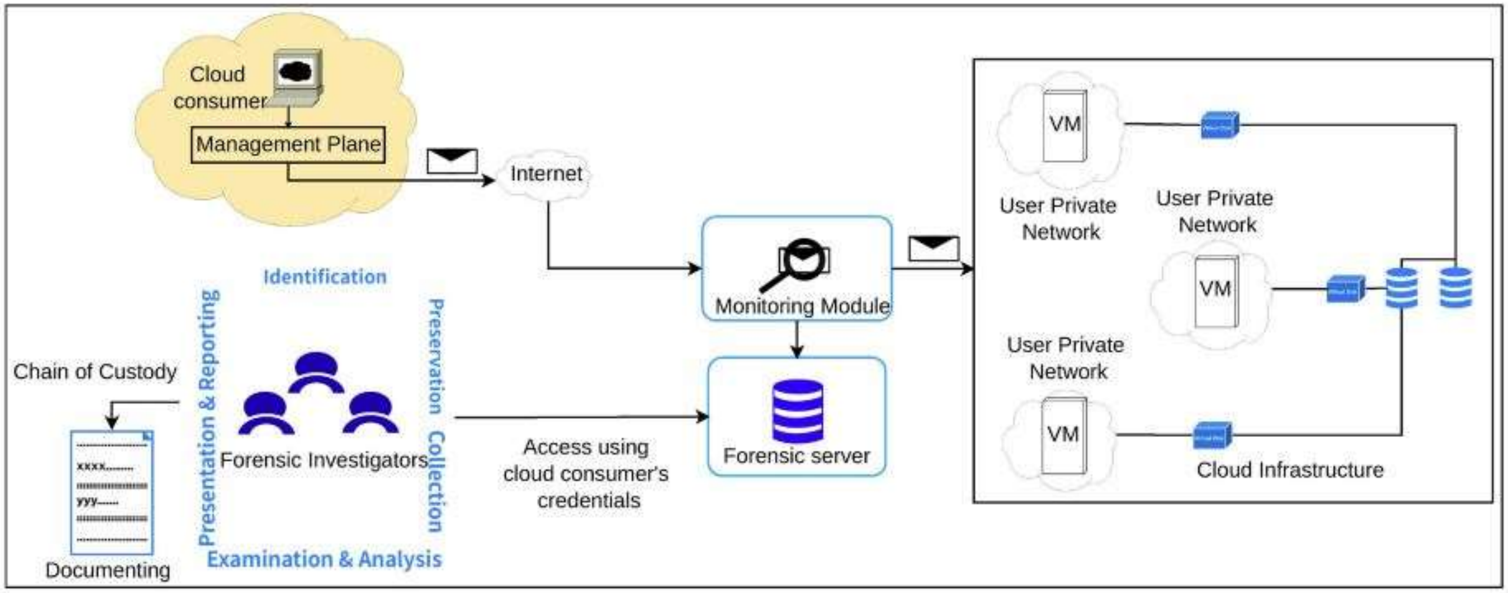
Cloud Forensic Architecture

- Storage forensics : Storage-as-a-service (StaaS) is a prominent cloud service model for cloud consumers. We further classify storage forensics in two categories, first, artifacts analysis that discusses the collection, preservation, and analysis of the evidential artifacts, and second, data provenance methods that emphasizes on the importance of provenance in the cloud forensics.

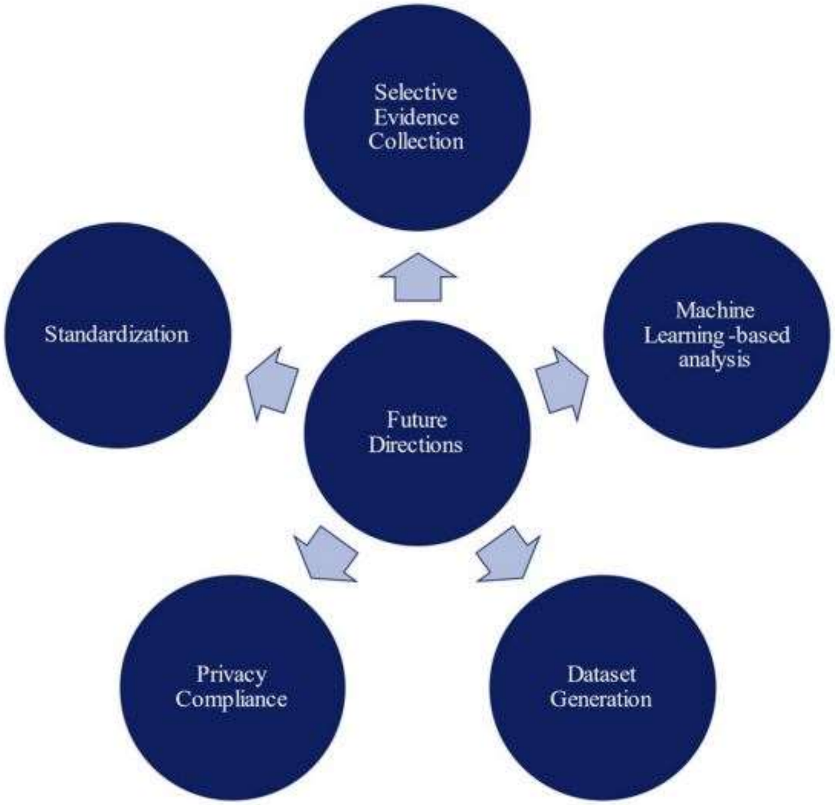
- Artifacts Analysis. Artifacts are the snippets of data left behind due to the interactions between the storage users such as client applications and storage services. These artifacts may comprise log files, registry entries, meta-data information, and many others, residing either on the client-side or server-side. Storage-as-a-service operates on a client-server model where consumers connect to the services facilitated on the cloud using different client devices, including desktop, laptop, and mobile devices. It is essential to identify and analyze the numerous digital artifacts residing on different client devices for proper investigation of cloud storage services.
    - Identified forensic remnants residing on client PC utilizing Amazon Cloud Drive during upload, download, and delete operations. use Perl-based scripts to automate the artifact collection
    - remote data acquisition tool, Cloud Data Imager (CDI) for cloud storage services with two main features: directory browsing and logical copy of selected folder to local repository.
    - artifacts identification on Android and iOS platform for cloud storage application, MEGA.
    - Marked the locations of distinct forensic artifacts on Windows machines and iOS devices with SpiderOak, JustCloud, and pCloud storage services
    - Determined client data remnants on client’s Windows 7 machine and Apple iPhone 3G amid various interactions between client devices and Dropbox cloud storage.

Cloud Forensic Architecture

- Consumer-driven cloud forensics: Showing two CSP independent solutions, management plane and central forensic server for evidence collection



Future directions of cloud forensics



Future directions of cloud forensics

- Segregating potential evidence from a huge amount of data can save a lot of time in the forensics investigation. Selective evidence collection techniques include collecting and preservation of evidence data from a pool of data in the cloud.
- The forensics process cannot be completely automated. However, a semi-automated approach for the analysis phase can aid the investigation.
- The use of machine learning approaches like the correlation of logs and meta-data analysis are the areas of focus that are picking up in cloud computing.
- Automation can be used in some phases of cloud forensics with a training approach that does not require collecting a user’s raw data, like federated learning can be used.
- Cloud forensics datasets which include attributes of the cloud-meta-data is not available publicly. Future direction includes a contribution to the dataset of cloud forensics.
- With upcoming privacy laws like GDPR and CCPA, a legal facet is also an emerging area of research for cloud forensics. Acquisition of the evidence data, preserving the privacy of the naïve user, and adhering to the new legislation requires up-to-date knowledge of the law.
- Researchers are leaning towards the ultimate aim of cloud forensics which is the standardization of the cloud. Standardization can be achieved through a holistic approach towards providing a cloud forensics solution.