CS 798: Digital Forensics and Incident Response

Lecture 20 - Cloud Forensics

Diogo Barradas

Winter 2025

University of Waterloo

Digital forensics in the cloud

- The usage of cloud platforms by cybercriminals is on the rise
  - While providing valuable sources of evidence

A popular definition of “the cloud”

A popular definition of “the cloud”

Why is that an issue for forensics?

Introduction to cloud forensics

- What do we intend with cloud computing?
- What are the specific forensic challenges in cloud forensics?
- What are the main techniques for cloud forensics?

Outline

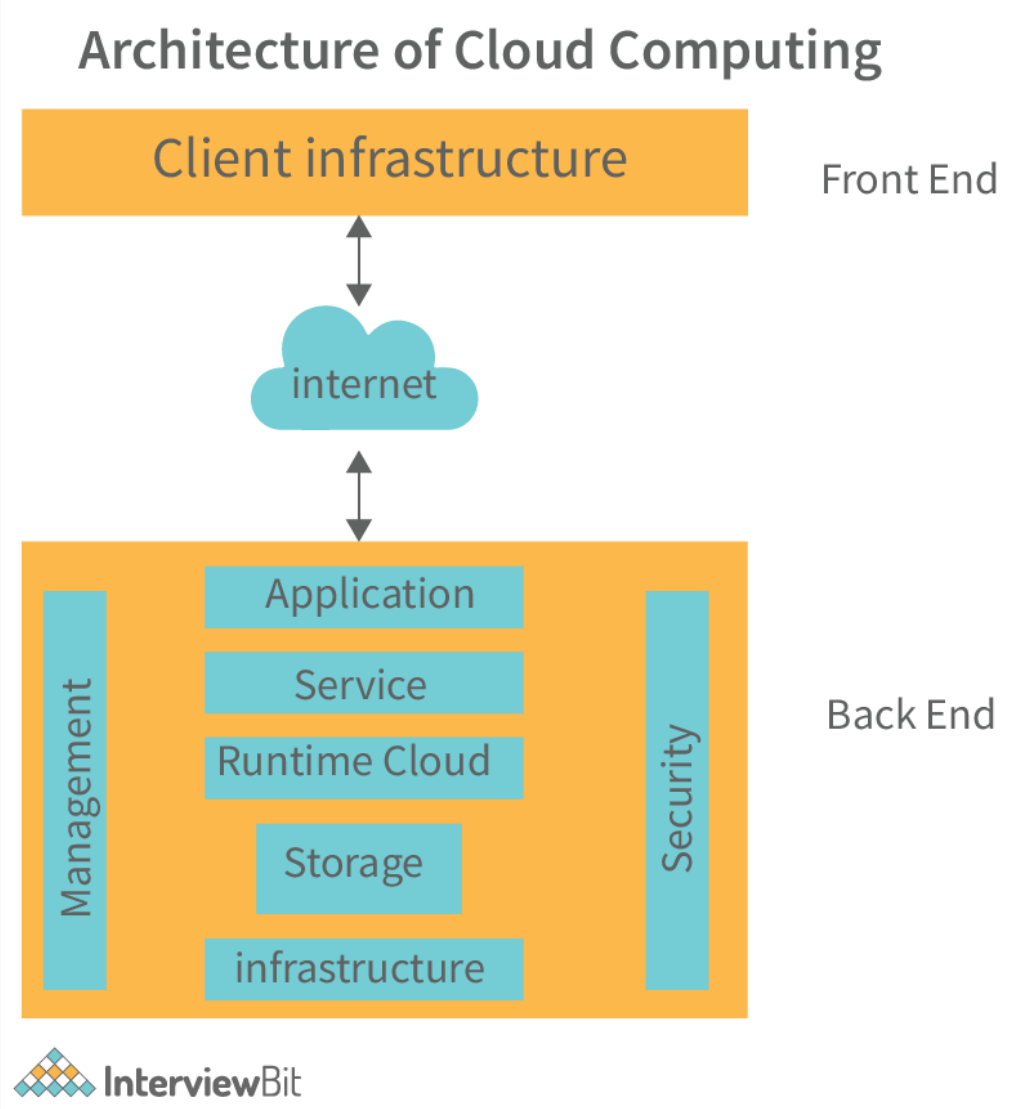
1. The cloud
2. Cloud forensics challenges
3. Forensic investigation of cloud machines
4. Investigating cloud customers
5. Tools for cloud forensics

The cloud

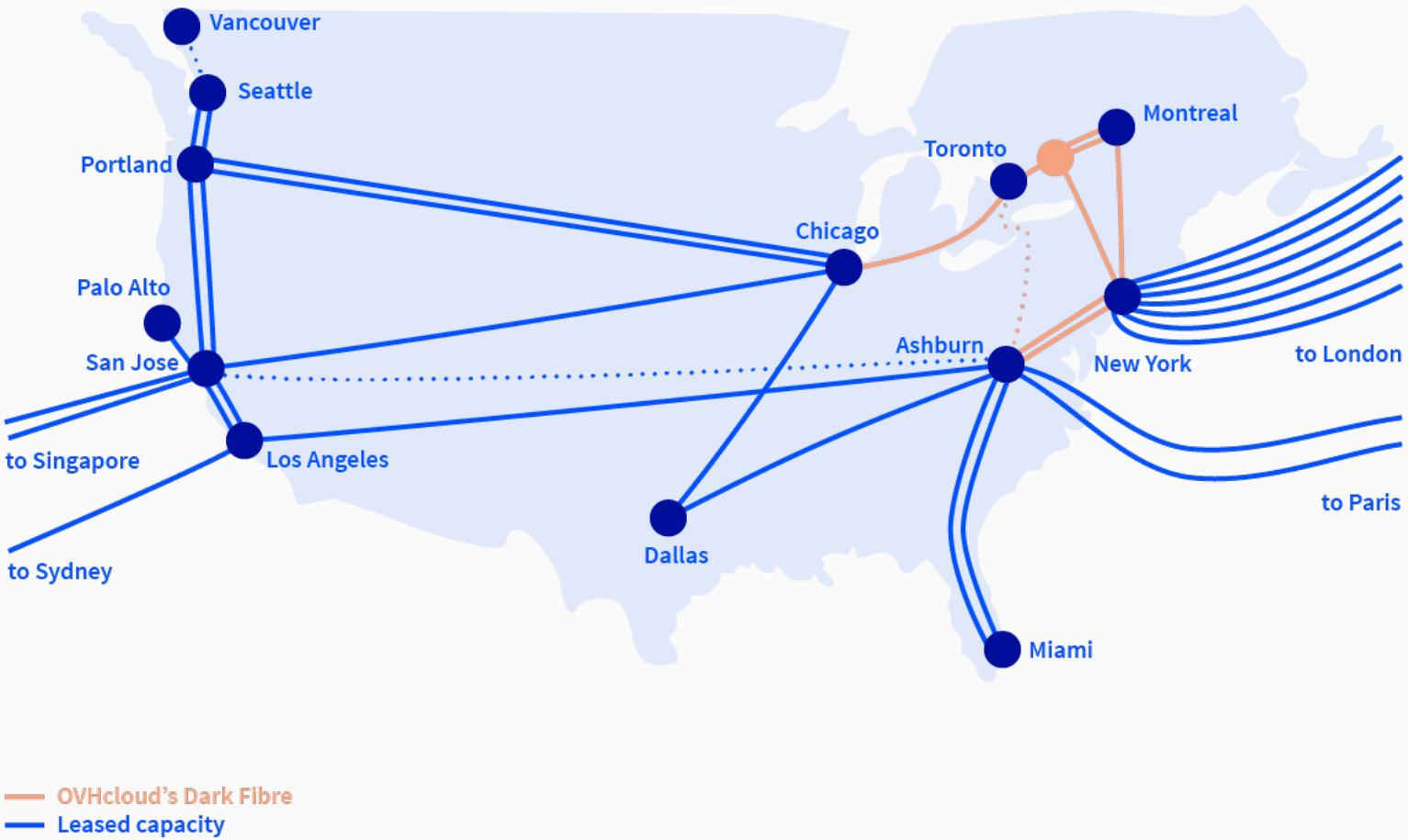
The cloud as defined by NIST

"Cloud computing is a model for enabling 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."

High level view of the cloud architecture

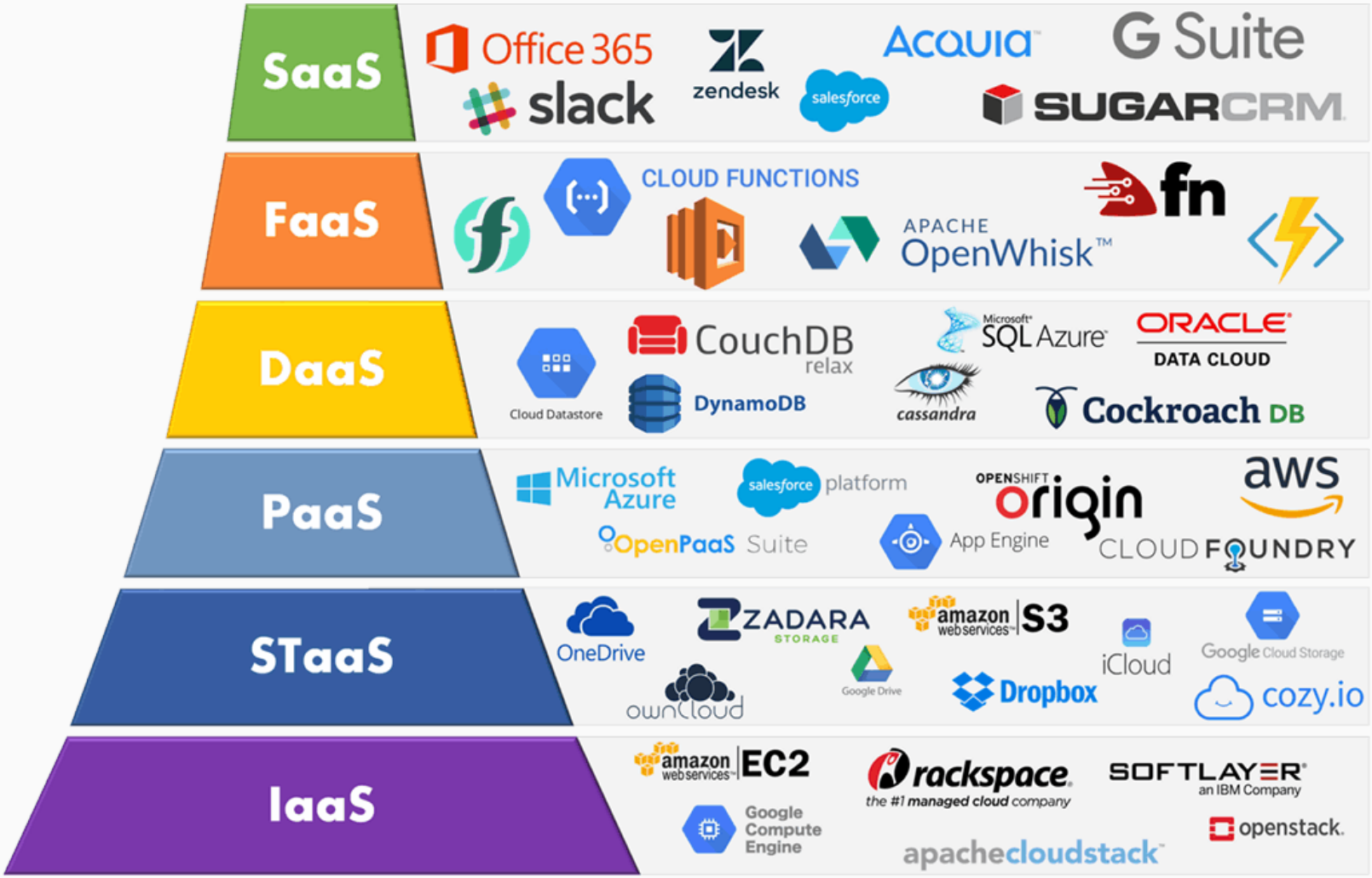


The cloud backbone: Datacenters

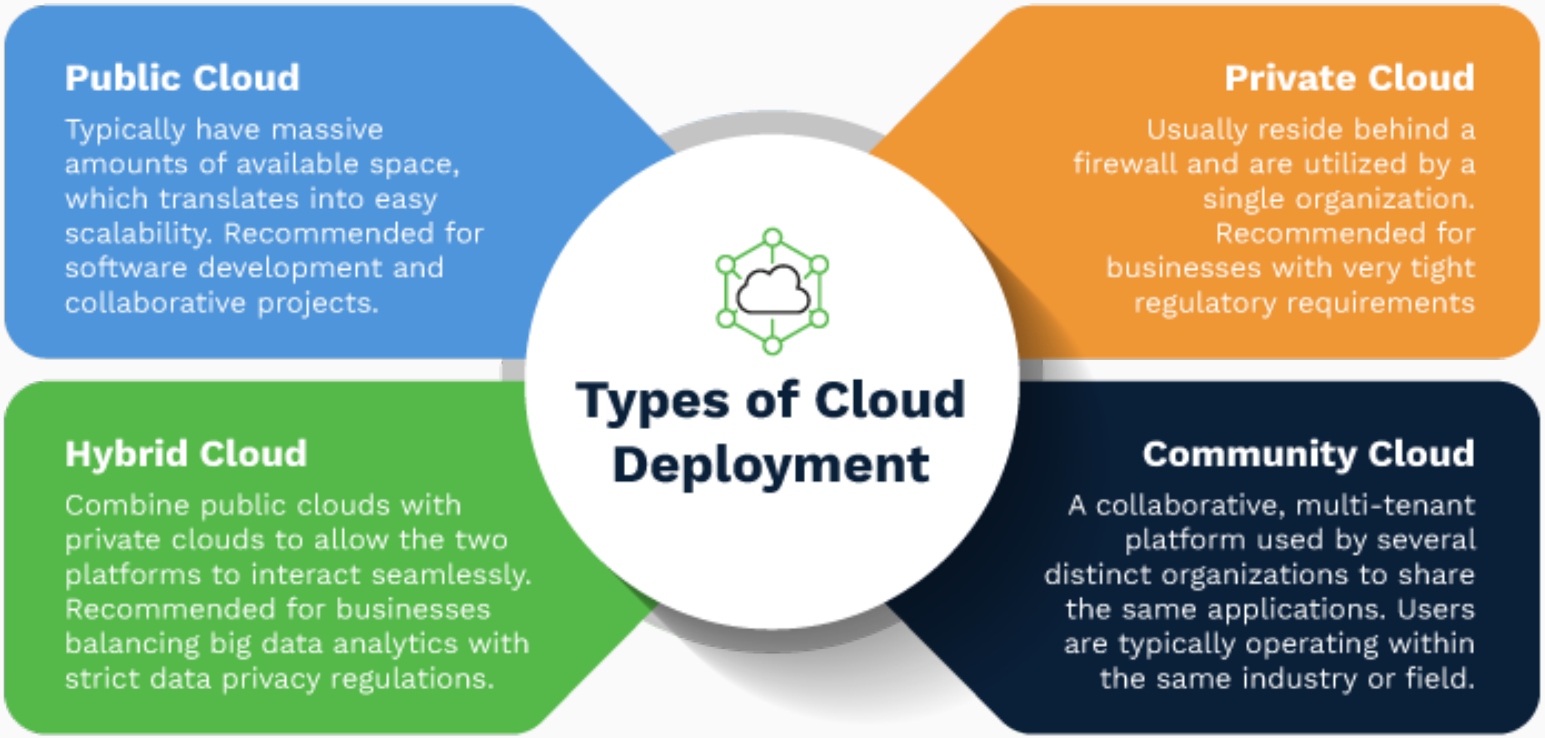


OVH backbone in North America

Cloud service models



Cloud deployment models



Cloud forensics challenges

Cybercrime in the cloud

Amazon's EC2 Cloud Service Fueled PlayStation Network Attack

By David Murphy | May 14, 2011 03:23pm EST | 7 Comments | Email | Print

+1 0 | Share 123 | Tweet | Share 16 | 1 Digg

If you're looking for the source of the network attacks that brought down Sony's PlayStation Network—yes, it's still down—look no further than Amazon. The online retail giant didn't bring down the PlayStation Network per se, but an undisclosed source speaking to Bloomberg News has indicated that hackers used Amazon's cloud services to fuel the break-in.

According to the source, the hackers posed as a normal business and signed up for a legitimate server rental through Amazon's EC2 service—otherwise known as Amazon Elastic Compute Cloud. It's unclear how the hackers specifically used EC2 to push the attack out, which is almost as unknown a figure as the exact treasure trove of data the attackers were able to access within Sony's network.

Cloud forensics as defined by NIST

"Cloud forensics is the application of digital forensics 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."

Some challenges

- Storage system is no longer local
- Each cloud server contains files from multiple users
- Separating data from different users is difficult
- Other than cloud service providers (CSPs), there is usually no evidence that links a given data file to a particular suspect
- Healthcare, business or national security related data!

Example: Malware for cryptocurrency mining

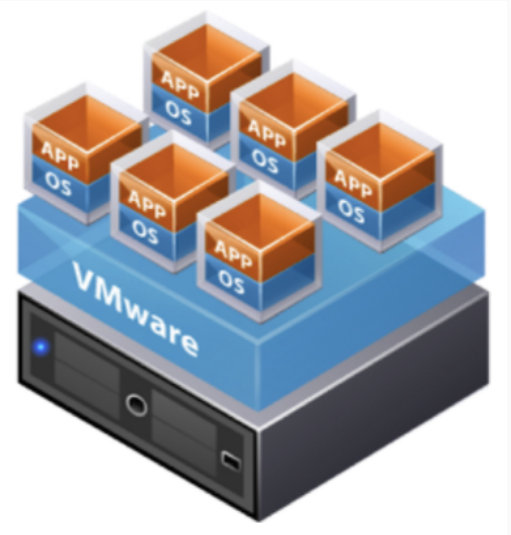
- A cloud container was compromised and started to mine cryptocurencies
- To investigate this case, forensics examiners would need a bit-for-bit duplication of the malware to prove the occurrence of cryptocurrency mining
- In the cloud, however, investigators cannot collect this data themselves

Obtaining a search warrant

- There are some problems with the search warrant in respect of cloud environment, for example:
  - Warrant must specify a location, but in cloud the data may not be located at a precise location or a particular storage server
  - The data can not be seized by confiscating the storage server as the same disk may contain data from unrelated users
- Almost in all aspects, it depends on the transparency and cooperation of the cloud provider

Virtual Machines and volatile data

- When we turn off a Virtual Machine (VM), all the data will be lost if we do not have the image of the instance
  - e.g., if we restart a VM instance in IaaS (e.g., Amazon EC2), we lose all data
- Some owner of a cloud instance can fraudulently claim that their instance was compromised by someone else
  - Later, it will be difficult to prove this claim as false



Trust issues

- After issuing a search warrant, the examiner needs a technician of the cloud provider to collect data
  - However, the employee of the cloud provider who collects data is most likely not a licensed forensics investigator and it is not possible to guarantee their integrity in a court of law

- Are there any restrictions on collecting evidence from remote cloud storage?
- If data is in a remote location, can the CSP provide a forensically sound connection to it?
- Data timestamps may be questionable if they come from multiple systems

Large bandwidth requirements

- In traditional forensic investigation, we collect the evidence from the suspect's computer hard disk
  - In cloud, we do not have physical access to the data

- One way of getting data from cloud VM is downloading the VM instance's image
  - The size of this image will increase with the increase of data in the VM instance
- We will require adequate bandwidth and incur expenses to download this image

Data location

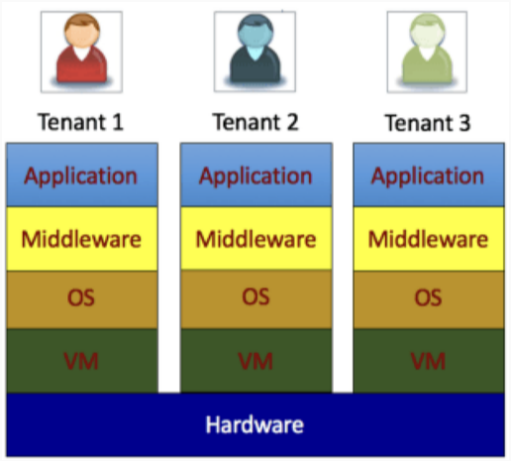
AWS Regions

Q: Where is my data stored?

You specify an AWS Region when you create your Amazon S3 bucket. For S3 Standard, S3 Standard-IA, S3 Intelligent-Tiering, S3 Glacier Instant Retrieval, S3 Glacier Flexible Retrieval, and S3 Glacier Deep Archive storage classes, your objects are automatically stored across multiple devices spanning a minimum of three Availability Zones (AZs). AZs are physically separated by a meaningful distance, many kilometers, from any other AZ, although all are within 100 km (60 miles) of each other. Objects stored in the S3 One Zone-IA storage class are stored redundantly within a single Availability Zone in the AWS Region you select. For S3 on Outposts, your data is stored in your Outpost on-premises environment, unless you manually choose to transfer it to an AWS Region. Refer to AWS regional services list for details of Amazon S3 service availability by AWS Region.

Multi-tenancy issues

- Multiple VM can share the same physical infrastructure, i.e., data for multiple customers may be co-located
  - This nature of clouds is different from the typical single owner computer
- How to prove that a piece of data was not tangled with other users' data?
- How to preserve the privacy of other tenants while performing an investigation?
- How to ensure that VM isolation has not been violated through side-channels?



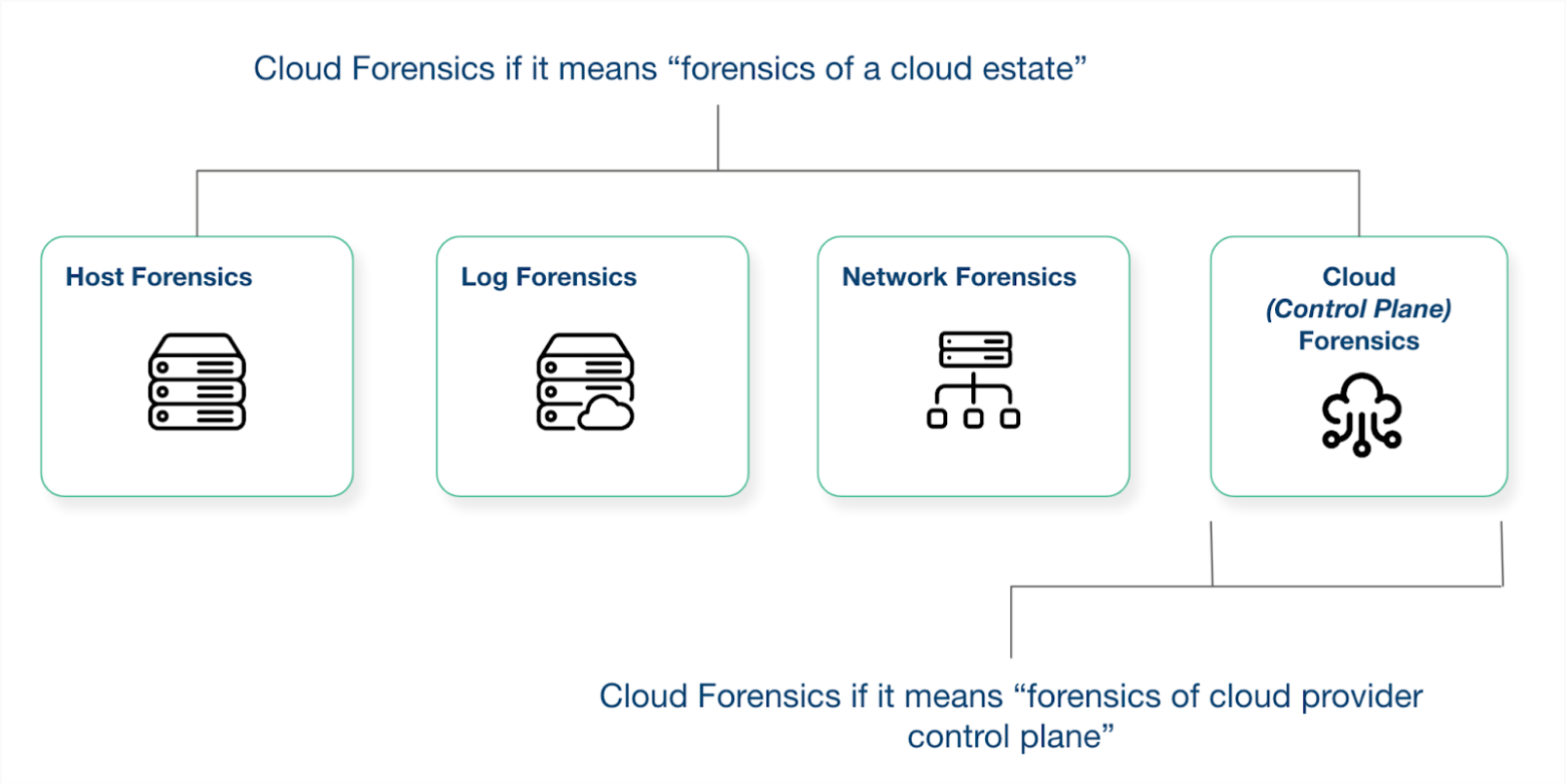
Logging

- Process logs, network logs, and application logs are really useful to identify a malicious user
- Not as simple as it is in privately owned computer system:
  - Decentralization
  - Volatility of logs
  - Multiple tiers and layers
  - Accessibility of logs
  - Dependence on the cloud provider

Forensic investigation of cloud machines

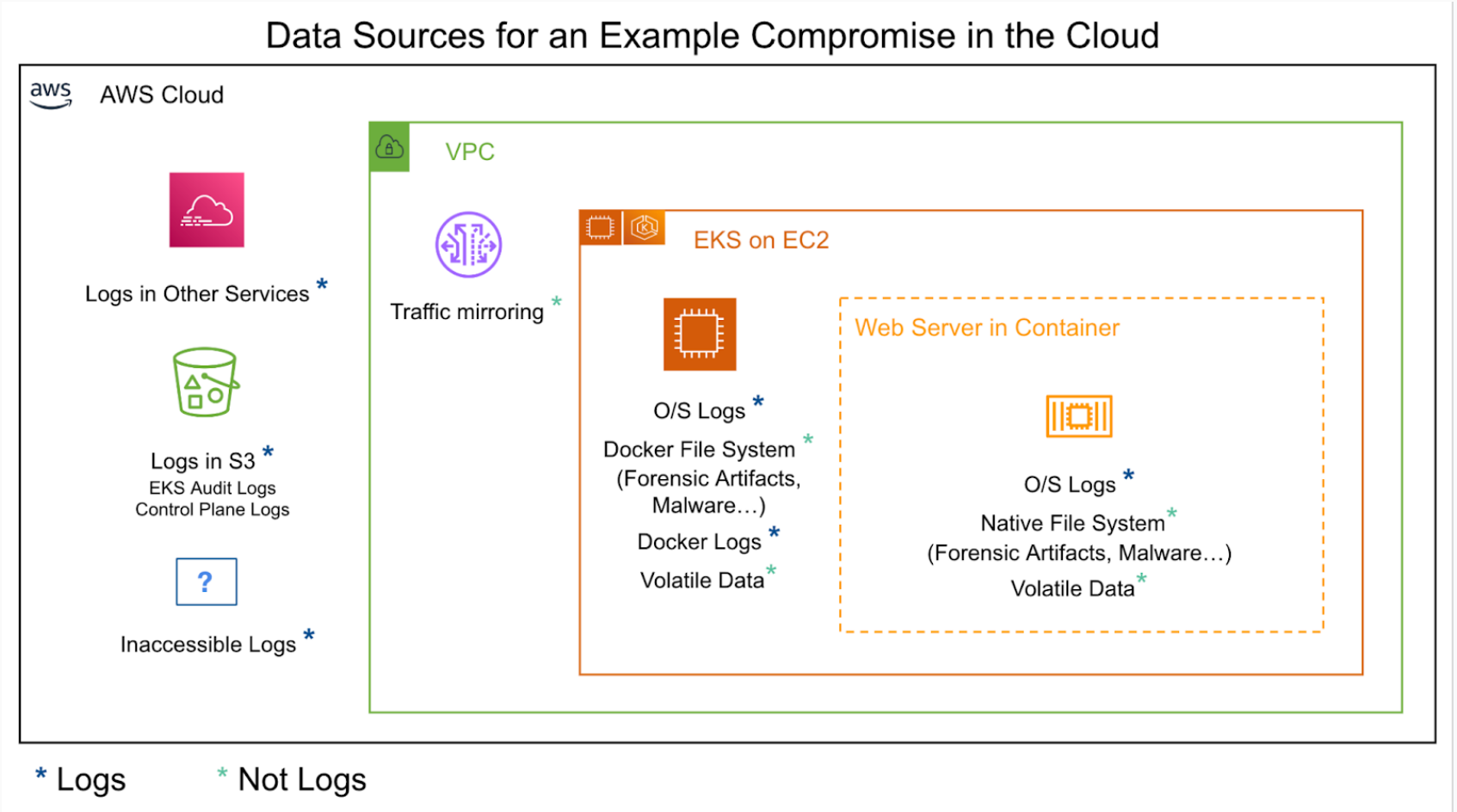
Is cloud forensics just log analysis?

Is Cloud forensics just log analysis?



- IAM logs, application logs, infrastructure logs, OS logs...
- “Inaccessible logs” that only the cloud providers have access to
- Undocumented logs

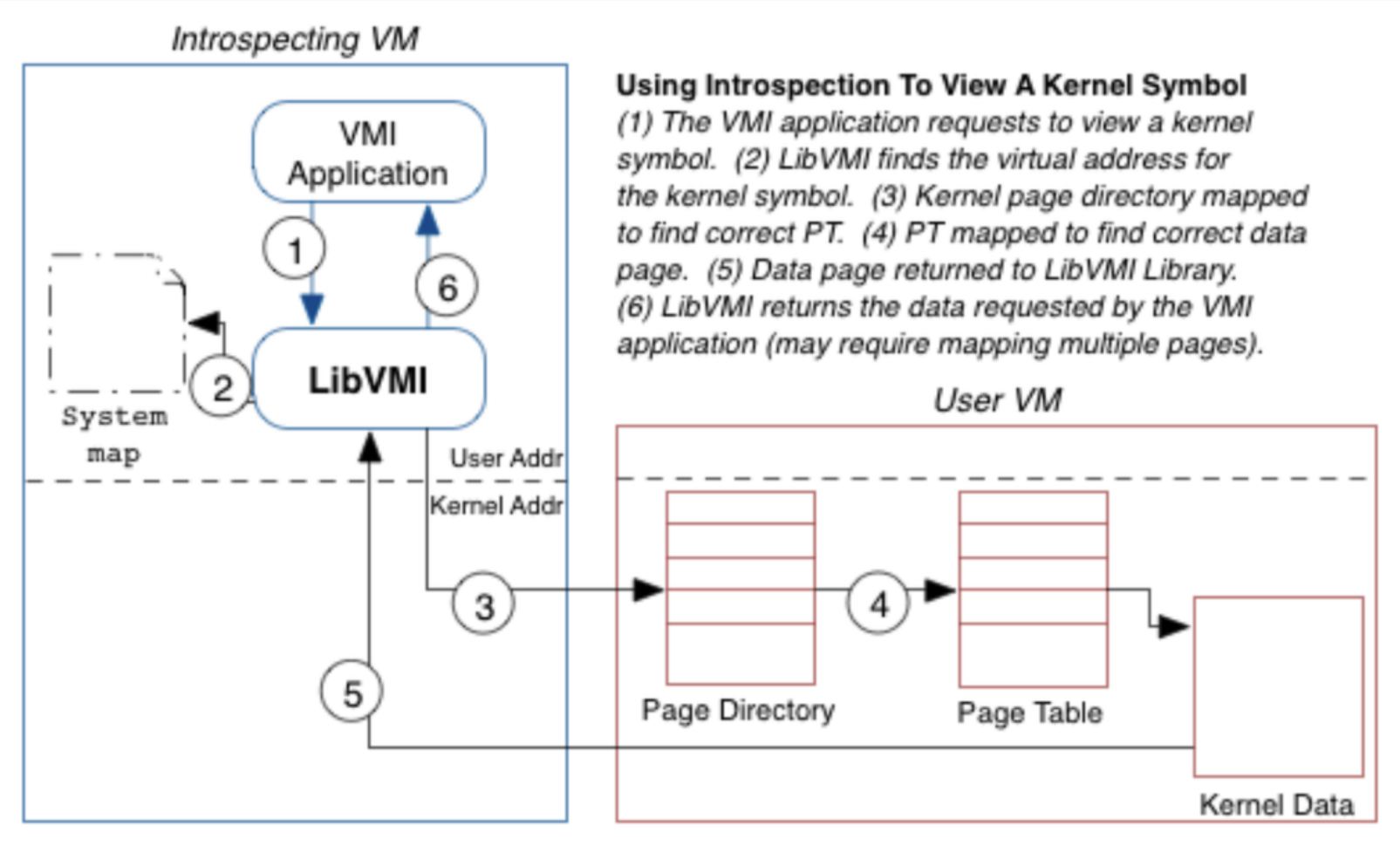
Putting it all together



Virtual Machine Introspection

- Virtual Machine Introspection (VMI) is the process of externally monitoring the runtime state of VM from either:
  - the Virtual Machine Monitor (VMM)
  - some VM other than the one being examined
- Acquire runtime state: processor registers, memory, disk, network, and other hardware-level events
- Through this process, we can execute a live forensic analysis of the system, while keeping the target system unchanged

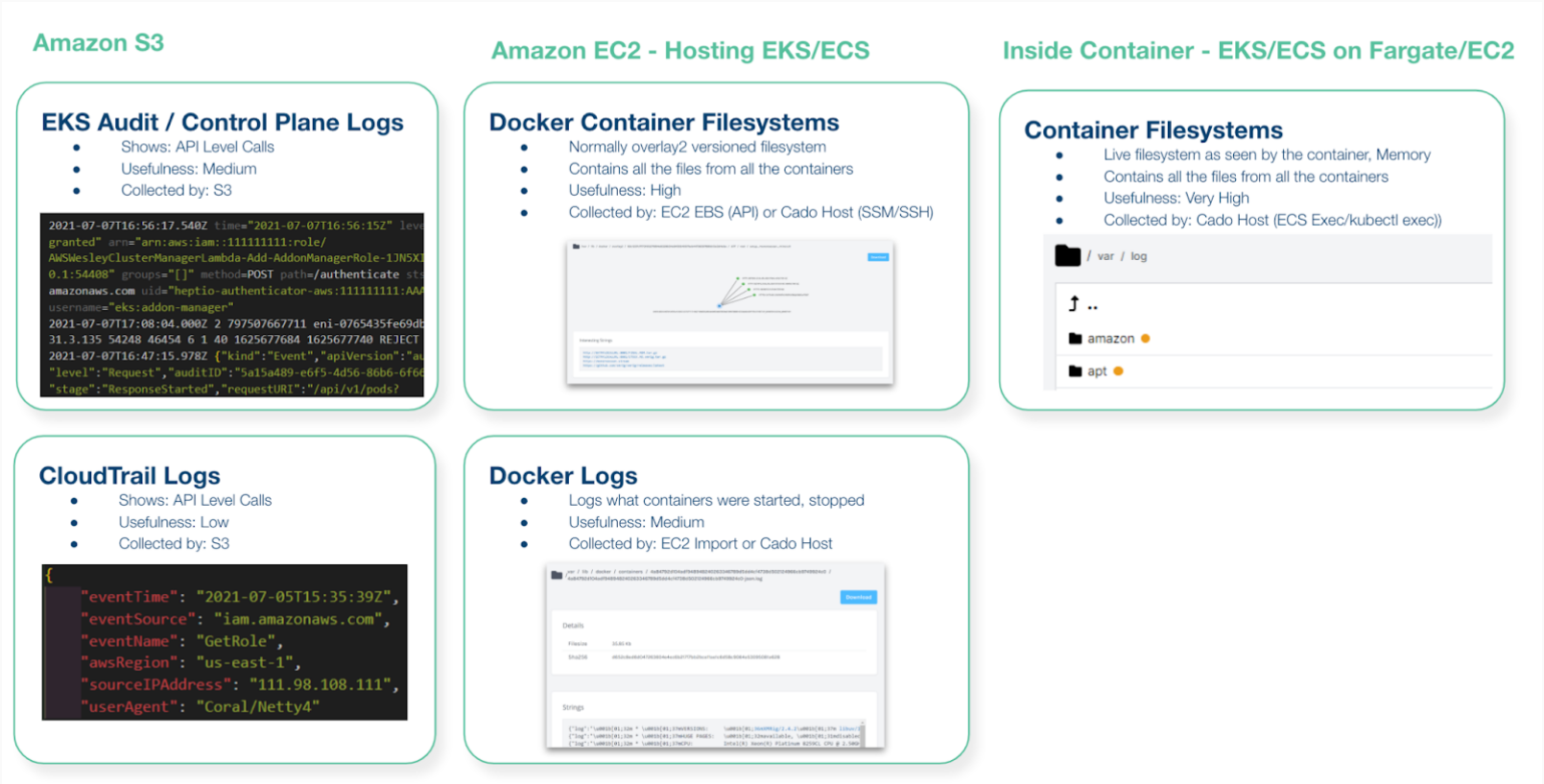
How does VMI work



http://libvmi.com/docs/gcode-intro.html

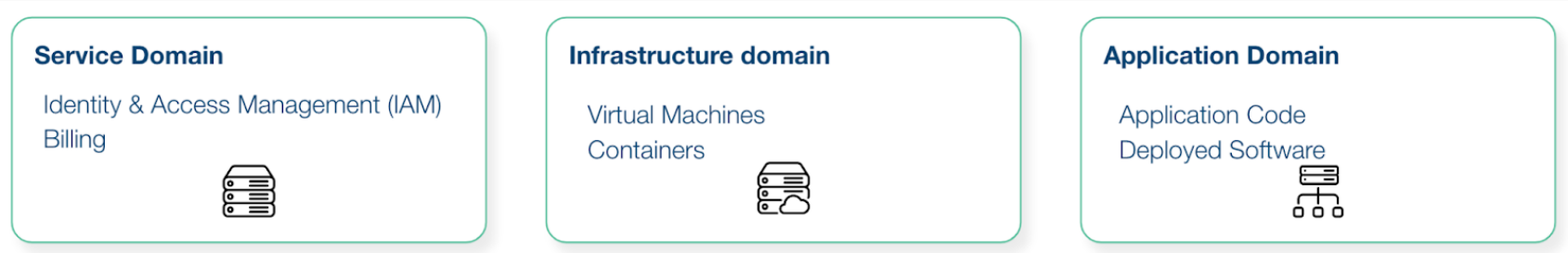
Other key data sources

Other key data sources when performing cloud forensics include access logs, network data and volatile data.



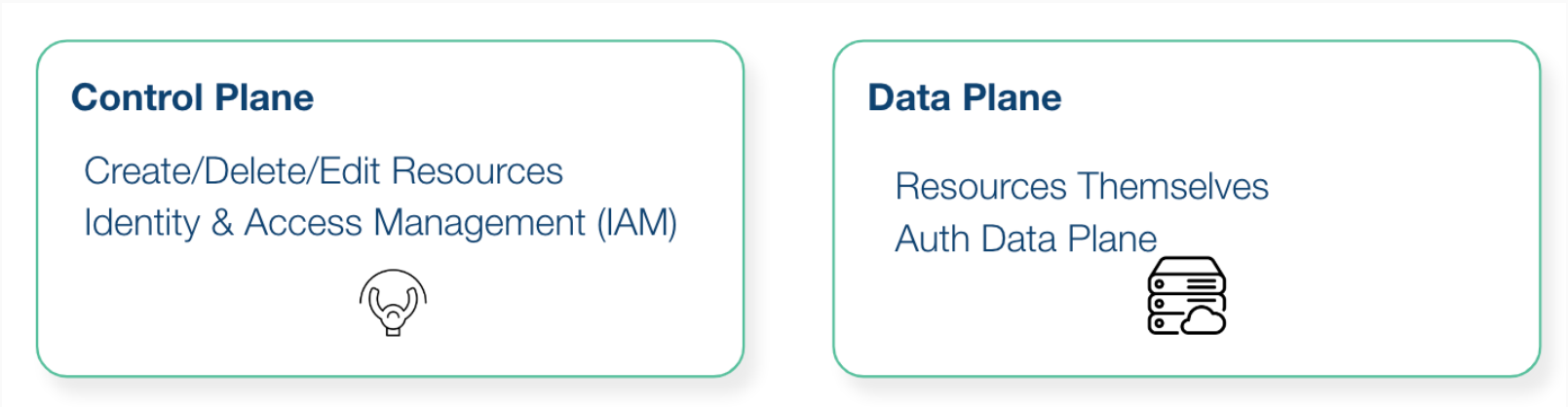
Incident domains

- AWS Cloud security incident domains:
  - Under this model, logs live across all three domains.
  - Most of the "non-log" data-sources exist in the infrastructure domain.



Incident planes

- Logs exist across both planes,
- Most of the "non-log" data can be found in the Data Plane.



Investigating cloud customers

Manual evidence collection

- Investigators can also manually collect and preserve data from cloud storage services via a browser or using client software
- If a cloud customer doesn't have the client software
  - Cloud-related evidence may be found in a Web browser's cache
  - Browser may have saved account credentials
- If the client software is installed
  - File transfers may be listed in the application's folder

Provenance between physical media and the cloud

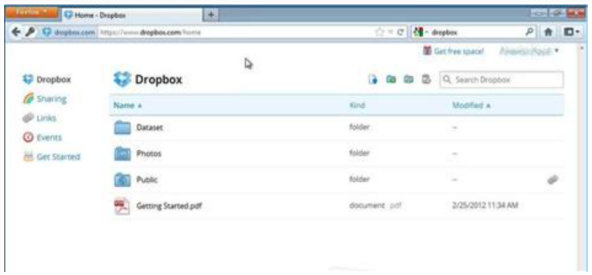
R v Paul James [2011] District Court of New South Wales

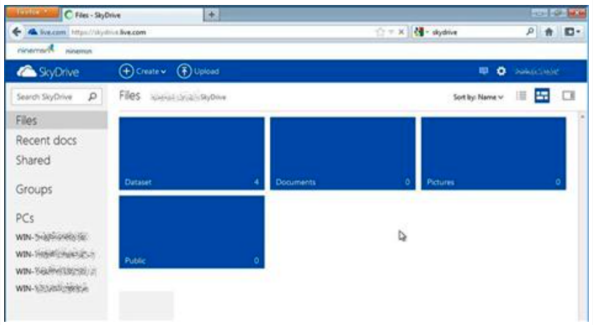
"police found that Mr James had a gmail account, a hotmail account and a yahoo account. Each of these accounts contained child pornography and in addition two computers which he possessed were found to have child pornography on them as well."

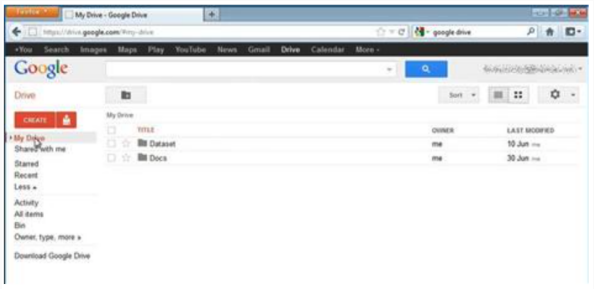
- Important to establish provenance between items on physical media and those preserved on a cloud storage account
- In this particular case, it was found that "the total number of images was much higher because there was a good amount of duplication over the email accounts and the two computers."

Changes on file contents

- The nature of cloud storage is that files can be modified while stored in the account
- Hash values (MD5 and SHA1) calculated for the files.
  - Hashes of downloaded, synchronized, and original files matched
- What if files were shared with someone?
  - May result in differences between the original uploaded file and the subsequent file downloaded from the account







How reliable is timestamp information?

EnCase: X-Ways and FTK:		Last Accessed (Accessed)	File Created (Created)	Last Written (Modified)	Entry Modified
Dropbox	browser	Last Written (UTC)	Last Written (UTC)	unZIP time	unZIP time
	sync	Download time	Download time	Same	Download time
Google	browser	Last Written (UTC)	Last Written (UTC)	unZIP time	unZIP time
Drive	sync	Last written	Download time	Same	Download time
SkyDrive	browser	Upload date/ time (UTC)	Upload date/ time (UTC)	unZIP time	unZIP time
	sync	Download time	Download time	Same	Download time

- Timestamp information varies depending on download method
  - Last Written time did not change when downloading a file using client software
  - All timestamps changed when downloading via a browser

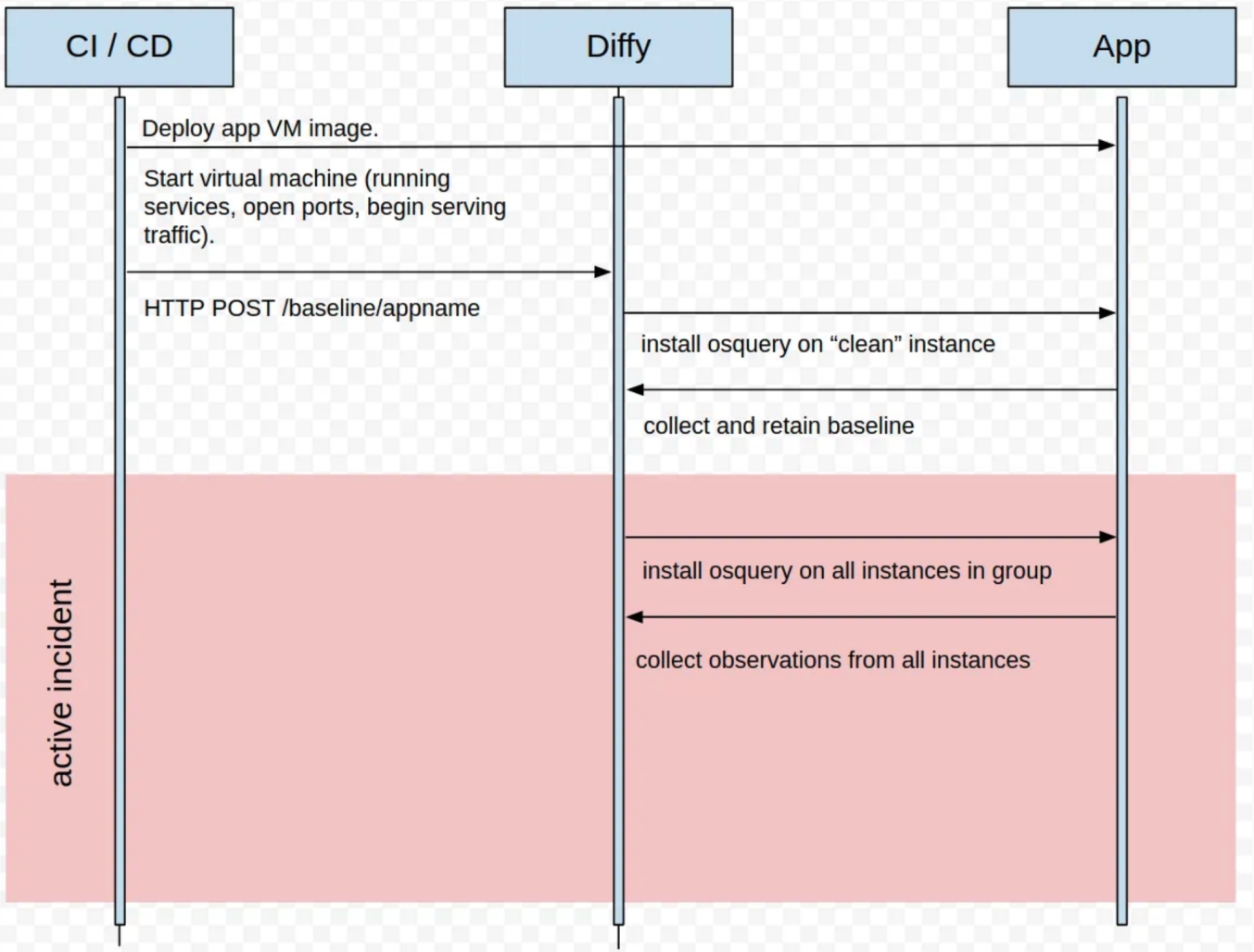
Tools for cloud forensics

Tools for cloud forensics

- Few tools designed for cloud forensics are available
- Many digital, network, and e-discovery tools can be combined to collect and analyze cloud data
- Vendors with integrated tools:
  - OpenText EnCase eDiscovery
  - Magnet AXIOM
  - eSentire Atlas XDR
- Open-source:
  - Diffy
  - Kuiper
  - Turbinia

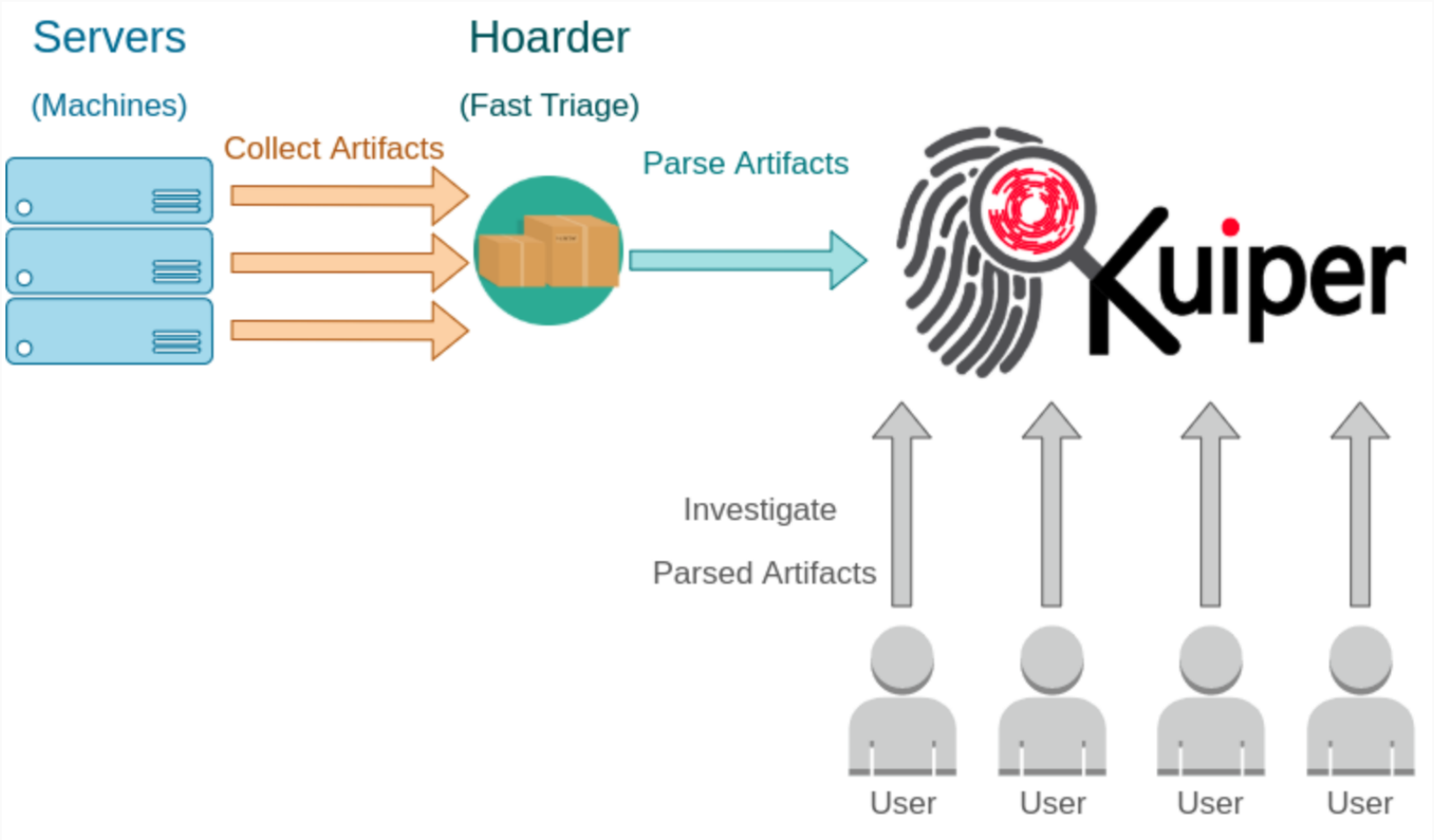
Diffy overview

- Highlights outliers in security-relevant instance behavior
  - Collects “clean” functional baseline and detects anomalies



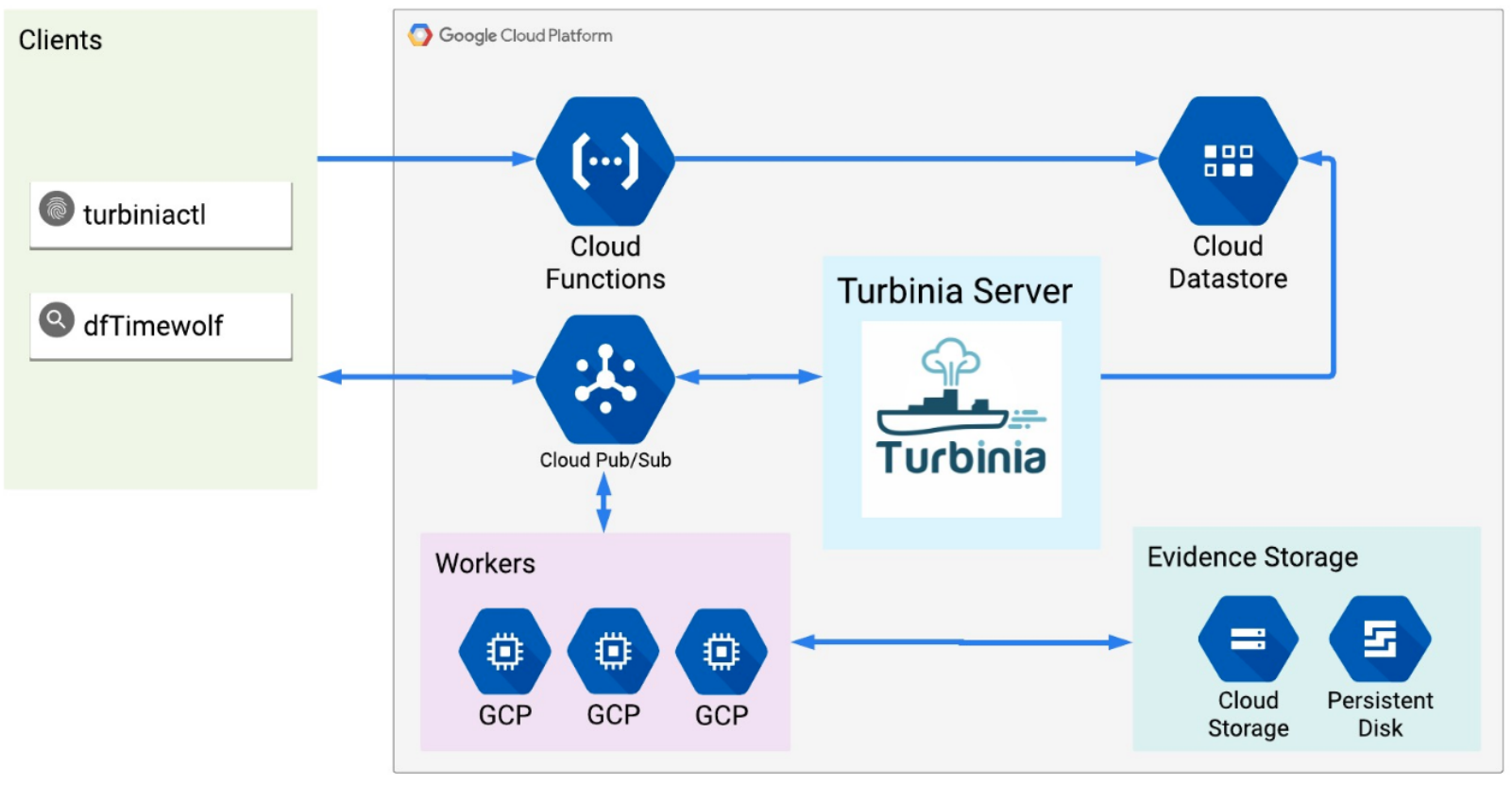
Kuiper overview

- Provides the capability for investigators to easily parse, search, and visualize collected evidence



Turbinia overview

- Helps deploying, managing, and running distributed forensic workloads to automate common forensic processing



Takeaways

- Cloud storage services are now popular options for users to store data across a range of devices. Cloud computing services are also an attractive alternative to meet the needs of enterprise customers
- Cloud services are subject to attacks which may facilitate online fraud and their resources might be abused for criminal purposes (e.g. CSAM storage)
- Cloud-focused investigations face several challenges due to the difficulties in accessing data, lacking legal support, and dearth of cloud-specific forensics tools

Pointers

- **Textbook:**
  - Quick – Chapters 1, 2, and 7
  - (If you are really curious, Quick – Case studies on chapters 3–6)
- **Other resources:**
  - Cloud forensics: Technical challenges, solutions and comparative analysis
- **Acknowledgements:**
  - Slides adapted from Nuno Santos's Forensics Cyber-Security course at Técnico Lisbon