FORENSICS OF CLOUD MACHINE LEARNING PLATFORMS – BASICS

Whenever I deal with the subject of the Cloud, no matter the context, it always reminds me "There is no other option; in a matter of a few years, everything will be in the Cloud." That came from a very experienced professional who, like many of our peers at the time, was not very comfortable with the idea of moving critical systems outside of our physical boundaries.

Nowadays, with most of our company's critical data moved to Cloud service providers, one of our major concerns is dealing with security matters. That includes being able to quickly respond to and report events that may lead to legal issues such as lawsuits and, in extreme instances, even involve law enforcement. This is by no means an easy task and matters are further complicated by the fact that we have to trust our Cloud provider's ability to deliver digital forensics data in the case of any legal dispute (either civil or criminal) during cyberattacks or even if a data breach occurs.

Well, maybe trusting provider capabilities is not actually the real problem, since a contract may include a hefty fine for such cases, but the fact remains that there is a huge difference between analysing locally stored data versus doing it at a Cloud service. Traditional computer forensics deals with collecting media at the crime scene or at least the location where the media was seized, the efforts for the preservation of that media, and subsequent validation, analysis, interpretation, documentation, and courtroom presentation of the results of the examination. For most situations, other than an internal investigation, any evidence contained within the media will be controlled by law enforcement from the moment of seizure. Now, in a Cloud scenario, the information may be anywhere around the globe, even outside your country boundaries. This can turn controlling the evidence (i.e., collection, preservation and validation) into quite a challenge.

To put it in simple terms, Cloud forensics combines Cloud computing and digital forensics, which mainly focuses on the gathering of digital forensic information from a Cloud infrastructure. This means working with a collection of computing resources, such as network assets, servers (both physical and virtual), storages, applications, and whatever service is provided. For most situations, this environment will remain (at least partially) live, and can be reconfigured quickly

with minimal effort. In the end, any sort of evidence collected must be suitable for presentation in a court of law.

WHAT ARE THE TYPES OF CLOUD ARCHITECTURES?

Before going any further into Cloud forensics, it is important to have a proper understanding of basic Cloud concepts:

There are three options of service models that define your Cloud architecture:

1. Infrastructure as a service (IaaS) delivers basic computer infrastructure (typically a platform virtualization environment) as a service, along with raw storage space and networking capabilities.
2. Platform as a service (PaaS) is the delivery of an entire computing platform and solution stack as a service, including all of the facilities required to support the complete life cycle of building and delivering web applications and services entirely available from the Internet. This allows the deployment of applications without the cost and complexity of buying and managing the underlying hardware and software and provisioning hosting capabilities.
3. Software as a service (SaaS) may be understood as "on-demand software." In this model, software and any associated data are hosted centrally and usually accessed by users using a thin client, such as a web browser over the Internet.

If you are using an external Cloud provider, it is important to understand that the lower down the stack your provider stops, the more you are directly responsible for implementing and managing security features. For instance, if you are using IaaS, it is expected that you are in charge of way more controls that can affect Cloud forensics than when using a PaaS or SaaS model.

WHAT ARE THE TYPES OF CLOUDS?

Once you have chosen your architecture, the next step is defining your deployment option. There are four basic Cloud types:

PUBLIC CLOUD

This is the most common type of Cloud offered by big players such as Amazon Web Services (AWS) and Google. In a public Cloud, the infrastructure is made available to the general public, so you will be sharing resources with other companies.

PRIVATE CLOUD

In this deployment option, the Cloud infrastructure is operated solely for a single organization. Think of it as your basic data center (located on-premise or off-premise) using Cloud technology and concepts. You may choose to manage it directly or even have a third party controlling it.

COMMUNITY CLOUD

A community deployment means that the Cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns. Either managed directly by the organizations or by a third party, it may be located on-premise or off-premise. For instance, this is a good option for highly regulated industry (e.g., healthcare) that does not want or need to build a private environment, but may not be able to use a public Cloud.

HYBRID CLOUD

As its name suggests, this delivery option combines two or more Clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., Cloud bursting for load-balancing between Clouds).

WHERE IS FORENSIC EVIDENCE MOST COMMONLY FOUND IN THE CLOUD

So, now that the basic concepts are clear, there is one very important question for Cloud forensics: Where is forensic evidence most commonly found in the Cloud? The first step is to know exactly where your data is and how much direct access you have to the infrastructure supporting it. As we stated before, it is

important to know what Cloud type and deployment option you are using. The lower down the Cloud stack your provider stops, the more direct control you have over data and evidence.

For instance, if you are using a private Cloud, it is more than likely that you have direct access to your hardware infrastructure and your Cloud forensics will not diverge too much from the usual digital forensics. On the other hand, if you are using a SaaS model over a public Cloud, initially direct evidence collection will be limited to whatever your provider offers in terms of logs or audit reports. Other than that, it all falls under what is covered on your contract, so special attention should be paid to your service level (SLA). If your agreement is not clear on what level of forensics information your service provider is bound to make available, and also how soon they are required to do it, you may find yourself in a very bad situation.

Also, if your data is not on-premise, you have to make sure to know where it is physically stored. This may affect your company from a legal standpoint, since laws and regulations may differ greatly depending on what state or country your information is stored in.

CONCLUSION

All in all, Cloud forensics is complex subject that demands a high level of experience. If you are with a company that wants to have its own digital/Cloud forensics team, you may be surprised by the lack of experienced professionals readily available on the market: The current cybersecurity skills gap means that the cost of an expert is on the rise, yet most businesses will put up quite a fight before letting a skilled professional go.