Cloud Forensics: An Overview - InfoSec Resources

Introduction
When discussing cloud forensics, we’re actually talking about the intersection between cloud computing and network forensic analysis. Cloud computing basically refers to a network service that we can interact with over the network; this usually means that all the work is done by a server somewhere on the Internet, which might be backed up by physical or virtual hardware. In recent years, there has been a significant increase on the use of virtualized environments, which makes it very probable that our cloud service is running somewhere in a virtualized environment. There are many benefits of virtualized servers, which we won’t go into now, but the most prominent ones are definitely low cost, ease of use, and the ability to move them around in seconds without service downtime. Basically, cloud computing is just a fancy term created by marketing people, but we’ve all been using it for years. A good example of cloud computing is an email service where we don’t have to install an email client on our local computer to access our new email and which serves as storage for all email. Instead, everything is already done by the cloud, the email messages are stored on the cloud and, even if we switch to a different computer, we only need to login with our web browser and everything is there. Therefore, we only need an interface with which we can access our cloud application, which in the previous example is simply a web browser. Cloud computing has many benefits, but the two most distinct disadvantages are definitely security and privacy. Since we store all data in our cloud somewhere on the Internet, the cloud provider has access to our data, and so does an attacker if a breach occurs in the provider’s network.

Network forensic analysis is part of the digital forensics branch, which monitors and analyzes computer network traffic for the purposes of gathering information, collecting legal evidence, or detecting intrusions [1]. When talking about network forensics, we’re actually talking about the data that has been transmitted over the network, which might serve as the only evidence of an intrusion or malicious activity. Obviously that’s not always the case, since an intruder often leaves evidence on the hard disk of the compromised host as well in the form of log files, uploaded malicious files, etc. But when the attacker is very careful not to leave any traces on the compromised computer, the only evidence that we might have is in the form of captured network traffic. When capturing network traffic, we most often want to separate the good data from the bad by extracting useful information from the traffic, such as transmitted files, communication messages, credentials, etc. If we have a lot of disk space available, we can also store all the traffic to disk and analyze it at a later time if needed, but obviously this requires a great amount of disk space. Usually we use network forensics to discover security attacks being conducted over the network. We can use a tool like tcpdump or Wireshark to perform network analysis on the network traffic.

Cloud Computing
Let’s talk a little bit about deployment models of cloud computing, which are described below
(summarized after [2]):

- **Private cloud**—The services of a private cloud are used only by a single organization and are not exposed to the public. A private cloud is hosted inside the organization and is behind a firewall, so the organization has full control of who has access to the cloud infrastructure. The virtual machines are then still assigned to a limited number of users.

- **Public cloud**—The services of a public cloud are exposed to the public and can be used by anyone. Usually the cloud provider offers a virtualized server with an assigned IP address to the customer. An example of a public cloud is Amazon Web Services (AWS).

- **Community cloud**—The services of a community cloud are used by several organizations to lower the costs, as compared to a private cloud.

- **Hybrid cloud**—The services of a hybrid cloud can be distributed in multiple cloud types. An example of such a deployment is when sensitive information is kept in private cloud services by an internal application. That application is then connected to the application on a public cloud to extend the application functionality.

- **Distributed cloud**—The services of a distributed cloud are distributed among several machines at different locations but connected to the same network.

The service models of cloud computing are the following (summarized after [2]):

- **IaaS** (infrastructure as a service) provides the entire infrastructure, including physical/virtual machines, firewalls, load balancers, hypervisors, etc. When using IaaS, we’re basically outsourcing a complete traditional IT environment where we’re renting a complete computer infrastructure that can be used as a service over the Internet.

- **PaaS** (platform as a service) provides a platform such as operating system, database, web server, etc. We’re renting a platform or an operating system from the cloud provider.

- **SaaS** (software as a service) provides access to the service, but you don’t have to manage it because it’s done by the service provider. When using SaaS, we’re basically renting the right to use an application over the Internet.

There are also other service models that we might encounter:

- **Desktop as a service**—We’re connecting to a desktop operating system over the Internet, which enables us to use it from anywhere. It’s also not affected if our own physical laptop gets stolen, because we can still use it.

- **Storage as a service**—We’re using storage that physically exists on the Internet as it is present locally. This is very often used in cloud computing and is the primary basis of a NAS (network attached storage) system.

- **Database as a service**—Here we’re using a database service installed in the cloud as if it was installed locally. One great benefit of using database as a service is that we can use highly configurable and scalable databases with ease.
- Information as a service—We can access any data in the cloud by using the defined API as if it was present locally.
- Security as a service—This enables the use of security services as if they were implemented locally.

There are other services that exist in the cloud, but we’ve presented just the most widespread ones that are used on a daily basis.

If we want to start using the cloud, we need to determine which service model we want to use. The decision largely depends on what we want to deploy to the cloud. If we would like to deploy a simple web application, we might want to choose an SaaS solution, where everything will be managed by the service provides and we only have to worry about writing the application code. An example of this is writing an application that can run on Heroku.

We can think of the service models in the term of layers, where the IaaS is the bottom layer, which gives us the most access to customize most of the needed infrastructure. The PaaS is the middle layer, which automates certain things, but is less configurable. The top layer is SaaS, which offers the least configuration, but automates a large part of the infrastructure that we need when deploying an application.

**Cloud Network Forensics**

The first thing that we need to talk about is defining why cloud network forensics is even necessary. The answer to that is rather simple: because of attackers trying to hack our cloud services. We need to be notified when hackers are trying to gain access to our cloud infrastructure, platform, or service. Let’s look at an example. Let’s imagine that company X is running a service Y in the cloud; the service is very important and has to be available 24/7. If the service is down for a few hours, it could mean a considerable financial loss for X’s site. When such an attack occurs, the company X must hire a cloud forensics expert to analyze the available information. The forensic analyzer must look through all the logs on the compromised service to look for forensic evidence. The forensics analyzer soon discovers that the attack was conducted from the cloud provider’s network, so he asks the cloud provider to give him the logs that he needs.

At this point, we must evaluate what logs the forensics investigator needs in order to find our who was behind the attack. This is where cloud network forensics comes into play. Basically, we need to take the digital forensics process and apply it to the cloud, where we need to analyze the information we have about filesystems, processes, registry, network traffic, etc. When collecting the information that we can analyze, we must know which service model is in use, because collecting the right information depends on it.

When using different service models, we can access different types of information, as is shown in the table below [3,4]. If we need additional information from the service model that we’re using, which are not specified in the table below, we need to contact the cloud service provider and they can send us the
required information. The table below presents different columns, where the first column contains different layers that we might have access to when using cloud services. The SaaS, PaaS, and IaaS columns show the access rights we have when using various service models and the last column presents the information we have available when using a local computer that we have physical access to.

<table>
<thead>
<tr>
<th>Information</th>
<th>SaaS</th>
<th>PaaS</th>
<th>IaaS</th>
<th>Local</th>
</tr>
</thead>
<tbody>
<tr>
<td>Networking</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Storage</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Servers</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Virtualization</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>OS</td>
<td>X</td>
<td>X</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Middleware</td>
<td>X</td>
<td>X</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Runtime</td>
<td>X</td>
<td>X</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
It’s evident from the table that, when using a local computer, we have maximum access, which is why the analysis of a local machine is the most complete. I intentionally didn’t use the term “easiest,” because that’s not true, since when we have maximum access to the computer, there are multiple evidences that we can collect and analyze. The problem with cloud services is that the evidence needs to be provided by the CSP (cloud service provider): If we want to get application logs, database logs, or network logs when using the SaaS service model, we need to contact the service provider in order to get it, because we can’t access it by ourselves. Another problem is that the user’s data is kept together with the data of other users on the same storage system, so it’s hard to separate just the data that we need to conduct an analysis. If two users are using the same web server for hosting a web page, how can we prove that the server’s log contains the data of the user that we’re after? This is quite a problem when doing a forensic analysis of the cloud service.

Let’s describe every entry from the table above, so it will make more sense.

- **Networking**—In a local environment, we have access to the network machines, such as switches, routers, IDS/IPS systems, etc. We can access all of the traffic passing through the network and analyze it as a part of gathering as much data as we possibly can. When using the cloud, even the CSP doesn’t have that kind of data, because it mustn’t log all the traffic passing through the network, since users’ data is confidential and CSP can’t record, store, and analyze it. The CSP might only apply the IDS/IPS solution to the network, which is only analyzing traffic for malicious behavior and alerting the provider of such activity.

- **Storage**—When we have hardware access to the machine, we know exactly where the data is located but, when using a cloud service, the data could be anywhere, even in different states, countries, or even continents.
• Servers—In a traditional system, we have physical access to the machine, which is why we can actually go to the machine and analyze the data on it; all the data is local to the machine. This isn’t possible when using the cloud, because the data is dispersed through multiple data centers and it’s hard to confirm that we’ve actually collected all the needed data.

• Virtualization—In a local environment, we have access to the virtualization environment, where we can access the hypervisor, manage existing virtual machines, delete a virtual machine, or create a new virtual machine. In the public cloud, we normally don’t have access to the hypervisor, but if we absolutely must have access, we can run a private cloud.

• OS—In a local environment, we have complete access to the operating system as we do in the IaaS model, but not in the PaaS and SaaS models. If we want access to the operating system, we could connect to the SSH service running on the server and issue OS commands, which we can’t do when using Heroku, for example.

• Middleware—The middleware connects two separate endpoints, which together form a whole application. For example, we might have a database running on a backend systems and the web application connects to those databases by using different techniques.

• Runtime—When using the IaaS model, we can influence how the application is started and stopped, so we have access to its runtime.

• Data/application—In PaaS and IaaS models, we have access to all of the data and applications, which we can manage by using search, delete, add, etc. We can’t do that directly when using the SaaS model.

• Access control—In all service models, we have access to the access control because, without it, we wouldn’t have been able to access the service. We can control how access is granted to different users of the application.

When conducting forensic analysis in the traditional way, we can simply hire a forensics expert to collect all the data and analyze it from the local machine. In a cloud service, we can do the same, but we must also cooperate with the cloud service provider, which might not have the forensics experts available or simply might not care and therefore won’t provide us with all the data that we need.

**Conclusion**

In this article, we’ve seen that, when conducting a cloud network forensic analysis, we do not have access to the same information as we do when conducting an analysis of a normal local computer system. We often do not have access to the information that we’re after and must ask the cloud service provider to furnish the information we need. The problem with such data is that we must trust the cloud service provider to give us the right information; they might give us false information or hold back some very important information. This is just another problem when trying to use the data in court, because we must prove without a doubt that the evidence from the collected data belongs to the user; the process of collecting the data, preserving it, and analyzing it must be documented and acceptable in the court of law.
When an attack has occurred on a cloud service, there are a lot of different problems we need to address, but the most important of them is communication with our cloud service provider. Because the services are located in the cloud, there is a lot of information that could serve as evidence which can only be provided by the CSP, since only the cloud provider has access to it. Therefore, there are also other problems with gathering the data when working with cloud environments, such as data being located in multiple data centers located around the globe, data of different users being located in the same storage device, etc.

There is still a lot of research that must be done in order to improve forensic examination of cloud services. There is also a lack of professional cloud forensic experts, which are expected to increase in the next couple of years.

References


• 21