

CLOUD FORENSICS – LAB SYSTEMS’ SOLUTIONS

The term Cloud Computing can be defined as a pay-per-use model for enabling convenient, on-demand network access to a shared pool of configurable and reliable computing resources (e.g., networks, servers, storage, applications, services) that can be rapidly provisioned and released with minimal consumer management effort or service provider interaction. A Cloud model generally comprises of five key characteristics, three delivery models and four deployment models.

It is a model, or concept where infrastructure, platforms, applications and services are offered up over the internet like any webpage where data and the modifications that can be done to the aforesaid data, is stored on line. Existing examples of such web pages include webmail, online backups, hosted services, etc.

The term 'cloud computing' thus covers everything from smart data centers, managed hosting, infrastructure, applications and services offered over a network with the main storage and processing being completed at the cloud service provider(CSP) end, with cheap, thin clients at the other end accessing the service via a browser.

Five Key and Essential Characteristics of Cloud Computing are mentioned below:

- 1. On-demand self-service.**
- 2. Broad network access.**
- 3. Resource pooling.**
- 4. Rapid elasticity.**
- 5. Measured Service. (Pay-Per-Use in other terms)**

There are further Three Service Models:

- 1. Cloud Software as a Service (SaaS):** Using provider’s applications over a network
- 2. Cloud Platform as a Service (PaaS):** Deploy customer-created applications to a cloud.
- 3. Cloud Infrastructure as a Service (IaaS):** Rent processing, storage, network capacity and other fundamental computing resources.

It is important to note that in order to be considered as a tool of “cloud computing” the above must be deployed on top of cloud infrastructure that has the key characteristics.

There are four Cloud deployment models as follows:

- **Public Clouds:** These are sold to the public or mega-scale infrastructure companies.
- **Community Cloud:** These are shared infrastructure for specific communities.
- **Private Cloud:** These are enterprise owned or leased
- **Hybrid Cloud:** These are composition of two or more cloud computing systems.

One of the advantages about cloud computing is that you basically exist in an on demand system, so if you are served with a preservation letter, or other legal reasons to preserve an environment, you can easily backup your environment and put it onto the cloud for the investigators to use, while the normal course of business happens. This also means that all the data stores or other information that investigators will need will also be cryptographically hashed much easier and much quicker using the on demand resources in the cloud.

Amazon web services is a good example of this. Amazon Web Services (AWS) can automatically provide a MD5 of every file that is on the system; so when you do a bit by bit copy of the file, everything is carried over with it. Add to that the Meta data that goes along with every file in Microsoft Office, you have a fairly good unimpeachable record of the file that the courts will need. Email stores and exact backups of a person's computing system are also available as well using this same kind of process.

The forensics tools can also be in their own off shoot of the environment allowing for very tight control over who has access to those tools and how they will be used. There are definite advantages to having a separate investigation environment for all the resources that are on the same cloud. Costs can be contained by making direct DVD copies from the investigation environment as needed or when needed making the process much more portable as well if information has to be turned over to the legal department or other investigators'

concept of using multiple instances for data mining - kind of essence of cloud computing. That opens a lot of new cool possibilities and potentially not so cool liabilities. It may be very interesting experiment in fact.

If we take out forensic limitations that come out of the cloud environment specifics and slightly modify the acquisition procedure (if needed for the case), before documenting the process to acquire evidence, then it can be explained in the court in a logical way. Will it stand in actuality is something that's to be tested, but assuming that a logical approach was used to gather it in the first place, one can only hope that it wouldn't be much of problem.

Finally, before we dive into how "*the Cloud*" will monkey wrench the modern-day computer investigator mindset, we need to reflect on a few of the basics of investigation and forensics:

- Data has to be collected in a manner that maximizes its integrity.
- Preserving chain of custody for the “*best evidence*” is critical to admissibility in a court of law.
- Conclusions that are derived from evidence should be reproducible by peers through well-accepted methods, within a controlled and similar environment.

If we take these tenets to a cloud context, many questions immediately come to mind...

1. **In the heavily virtualized/abstracted world of cloud computing, how can one identify and obtain the data that one needs?**
2. **In the distributed cloud model, what collateral data can one identify and collect to help one prove (or) disprove of a hypothesis?**
3. **What data does my provider log? How long do they keep it?**
4. **What data will my provider give to me?**
5. **What knobs can one turn up to get the data that one needs?**
6. **Does my provider expect me to do this in a self-serve fashion so they are not involved in the interpretation of data? Do they expect me to use an API that I can use to gather it?**
7. **If I need to ask my provider for data, how long will it take them to produce it?**
8. **How/will they vouch for the integrity of the data?**
9. **How/will they transfer it to me in a way that preserves integrity?**
10. **What/where exactly is the “best evidence”?**
11. **What methods and procedures are accepted?**

The answers to these questions are complex and are not based upon “The Cloud” itself. That is so because “The Cloud” by itself is not uniform. Each provider will have their unique approach to their cloud offerings and each in turn will enable a different form and depth of investigation.

Thus in order to comprehend the pro and con of the Cloud system, let us try and understand each service in detail.

Software as a Service (SaaS): The SaaS model is the easiest to understand. A SaaS provider invokes an instance of an application for an organization. There are knobs that can be turned up and down, and basic configuration can be applied. The consumer may be able to interface with the application via an API, but there won’t be deeper programmatic control that will modify the core application of your system. Examples of this model: *Google Gmail*, *Microsoft Online BPOs*, the popular less packages [<http://lesseverything.com/solutions>], and the Zoho suite [<http://www.zoho.com>].

- **Pro:** You might be able to get high level application logs. This data might log success and failures, and might reflect the actual activities within the environment. It will depend on what the provider decides to log, and how long they store it. In some cases, some services – such as federated authentication – may be handled by a separate service. This separate service (which may be operated by a completely separate provider) might have “*collateral*” data that you may find of interest, but you’ll have to figure out how to get it.

The best news of this scenario is that you'll be able to recreate the environment with a high level of accuracy (of course, many disclaimers apply here). For example, in the SaaS email deployment, many providers of messaging solutions have an option for "message journaling." This feature tells the provider to transparently forward a "carbon-copy" of all messages to archive service. This service can be used to gather evidence that sits in the organization's email activity. However, to use it, you need to ensure that (a) it's been purchased, (b) it is configured right (i.e. retention policy), and, in the case that you choose a different provider and (c) your email service is configured to send data to it.

- **Con:** Low level disk imaging is very unlikely as is installing forensic tools to obtain system state information. Interpreting information is also difficult. You'll need to know a lot about the application, and all of the scenarios that it can be used. For example, many SaaS accounting applications have (a) a web interface, (b) a client application, and (c) an web service API (that's usually invoked with an API key (aka "shared secret authentication")). Will you be able to interpret log entries properly across these code paths? And, by the way, you may not have a version of the server-side software that you can deploy in your lab for low-level analysis.
- **Platform as a Service (PaaS):** In this model, the consumer deploys application packages to a runtime environment that is hosted by a cloud provider. For example, if you use Microsoft Windows Azure, you build applications in Visual Studio, compile and publish a package through the Azure developer portal. This package (and a respective configuration file) is uploaded to Azure, where it is executed within a uniform Windows/.NET runtime environment. Similarly, with Google applications, you can build Python or Java applications which can be uploaded to Google's Application engine for execution. In this model, you **own the core application**, and programmatically dictate how it will interact with other dependencies (such as calling Database resources).
- **Pro:** Your development organization controls the core application. Thus, you can log information as you desire (to both a location, such as blob storage, an external database (even at another Cloud provider), or even SYSLOG to your SIEM solution). Since you have programmatic control of the platform, you can likely invoke custom code that interrogates system state and pulls logs. You just need to invest the time to configure it, and convince your development team that this feature is worth their time.
- **Con:** You may or may not be able to get logging information from the underlying runtime environment. This will depend on how the provider has it configured, and what they'll allow you to query. In this model, there are two important elements of the platform that need to be considered. In the Microsoft stack, it's (a) the virtualized OS, (b) IIS and (c) the .NET runtime environment. The Google model is similar. There's an OS, a web server, and a runtime environment (either Java or Python).

Infrastructure as a Service (IaaS): The consumer deploys virtual machines, in which they have administrative access. In some cases (*GoGrid*), the virtual machines use persistent storage (if a VM is rebooted, bits written to the disk will remain). Others, such as Amazon ECS, do not have persistent storage – when a VM is rebooted, it is "reset"

back to the base VM image. In Amazon AWS, persistent storage is derived from writing bits to Amazon EBS (Elastic Block Storage), or other data repository (such as an Amazon SimpleDB).

- **Pro:** You also have the ability to connect to the underlying VM (i.e. Linux console or Windows Terminal Services) to perform deep interrogation of the machine. Therefore, many of the “*traditional*” processes associated with forensics can be observed (such as querying system state). Also, many of the IaaS providers support snapshotting a running VM. So, you can also capture the state of a running host quickly (via API that shoots a host after system monitoring detects an abnormality).
- **Con:** To do this level of investigation, you’ll need some robust connectivity to the Internet (or provider’s network) and therefore it’s plausible that you can have a similarly hosted “*security cloud*” that has loose coupling (and robust connectivity) to your “*production cloud*” where you can pull disk images, system state, etc.

BY Emma Webb Hobson (Senior Digital Forensic Investigator, QinetiQ):

Jurisdiction is definitely a problem. I attended a conference recently where a techie chap from Microsoft was presenting, and he was asked the question "Is Microsoft planning to build data centres in the UK, on the basis that for legal reasons in some cases, and also for preference, UK companies need or want to store their data in the UK?" (Microsoft's nearest data centre is in Ireland).

The answer was no, Microsoft is not planning on building UK data centres; in usual MS fashion, they are talking to the EU to try to get the laws changed! It's easy to roll your eyes at MS plans to dominate the world, but I think good luck to them on this issue. If they can get some international agreement in this area, that has to be a good thing, and I imagine / hope that agreement would only come with law enforcement needs, forensic readiness requirements, FOI compliance, e-discovery capability and so on built in as part of the deal. I think MS are one of the very few who have the resources and desire to push something like this.

In the meantime, we're in the old boat of establishing SPOCs with the big providers, and hassling the ACPO and other powers to improve and speed up the process for requesting and receiving data.