Cloud Computing and Digital Forensics:

What are the Challenges?

NIST Cloud Computing WG
December 20, 2012
Overview

- Introductions
- Definitions
- Architecture
- Challenges
- Cases
- Discussion
Background


- AIS: @bwtech, UMBC, Catonsville MD: Develops technologies to support effective operations within and across the entirety of the cyber domain. Other offices in Dayton Ohio, San Antonio, TX, Denver Colorado, and Beaverton, Oregon.
Founded in 2001 as outgrowth of AFRL Adversarial Science Laboratory

Currently 160 employees

Classified R&D for DoD, IC, and law enforcement

- CNO: CNA, CND, CNE
- Forensics (anti-, counter/anti-, …)
- Code analysis
- Embedded systems
- Cyber mission operations
- Secure architectures
What is Cloud Forensics?

- Is it doing the digital forensics examination using a cloud environment?
- Is it collecting information from a cloud?
- Is it both?
Nation’s first

- Edmond Locard (1877-1966) Sherlock Holmes of France
- Two attic rooms and two assistants, to start what became the first police laboratory
- Developed 12 matching points for fingerprint identification (circa 1918)
- What first computer forensics lab started in Washington, DC and later Maryland?
Locard’s Exchange Principle

*It is impossible for a criminal to act, especially considering the intensity of a crime, without leaving traces of this presence.*

- In the cyber world, the perpetrator may or may not come in physical contact with the crime scene. This brings a new facet to crime scene analysis.

- How does Locard’s exchange principle apply to cloud computing?
Cyber Exchange Principle

“Artifacts of electronic activity in digital devices are detectable through forensic examination, although such examination might require access to computer and network resources involving expanded scope that may involve more than one venue and geolocation.” (Zatyko and Bay, 2011)
Forensics and Cloud Computing: Defining Digital Forensics

“The application of computer science and investigative procedures for a legal purpose involving the analysis of digital evidence after proper search authority, chain of custody, validation with mathematics, use of validated tools, repeatability, reporting, and possible expert presentation.” (Zatyko)

Given this definition, this scientific process contains the following eight steps:

- Search authority
- Chain of custody
- Imaging/hashing function
- Validated tools
- Analysis
- Repeatability (Quality Assurance)
- Reporting
- Possible expert presentation

“Sherlock Holmes in the 21st Century”
Cloud Computing Defined

“Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.”

- NIST Tech Beat: October 25, 2011

Five essential characteristics

- on-demand self-service
- broad network access
- resource pooling
- rapid elasticity or expansion
- measured service
Cloud Computing Defined

**Service Models**

- **Infrastructure as a Service (IaaS)**
  - Provides consumers with a virtualized environment
  - Offers storage, networking, and servers
  - Allows consumers to host and develop their own applications
  - ex. Amazon AWS

- **Platform as a Service (PaaS)**
  - Provides customers with an environment to host and build applications using a solution stack consisting of a platform and services (Orlando, 2011)
  - Built upon IaaS
  - ex. Google App Engine

- **Software as a Service (SaaS)**
  - Provider hosts and serves applications to consumers
  - Built upon PaaS and IaaS
  - ex. TurboTax online
Cloud Service Models

- Application
- OS + App Server Stack
- Software Application
  - Platform
  - Infrastructure

(CISCOSCHOOL.NET)
A Generic Cloud Computing Architecture
Cloud Architecture Components

- **Cloud Client**
  - Consists of the hardware, operating system, and applications used to access cloud instances

- **Cloud Scheduler/Manager**
  - Management software that monitors and logs usage
  - May include load balancing and adjust to changing demand by managing the creation or de-allocating of cloud instances

- **Cloud Instance**
  - The virtual environment accessed by users
  - Hosts applications, services, operating systems, software stacks, etc.
Cloud Architecture Components

- **Hypervisor (Virtual Machine Monitor)**
  - Hypervisors allow multiple cloud instances, or virtual machines, to run on a single physical system
  - Responsible for resource provisioning of the hardware inventory (Orlando, 2011)
    - “Resource pooling is what makes virtualization possible, and virtualization is what makes multi-tenant computing possible” (Orlando, 2011)
    - Manages cloud instances and virtual images

- **Administrative Domain (aka “dom0”)**
  - A privileged virtual machine used to manage the creation and de-allocation of cloud instances
  - Capable of directly accessing physical hardware
  - Able to access, analyze, and modify cloud instances

- **Cloud Storage**
  - The database system used to store information in the cloud
  - ex. Microsoft’s SQL Azure
Forensics and Cloud Computing: Traces

- **Cloud Client**
  - Traditional forensics (hard disk, memory, etc.)
  - ISPs may also retain IP address allocations

- **Cloud Scheduler/Manager**
  - Logs of inbound connections, cloud instances and physical hardware used to service clients
  - Consumer account information, etc.
  - Internal cloud service provider audit logs
  - Authentication and access logs
    - Control granted to customers for use of applications and services

- **Cloud Instances**
  - Traditional forensics (hard disk, memory, etc.)
  - May require remote acquisition and credentials
Forensics and Cloud Computing: Trace Locations

- **Hypervisor**
  - Dependent on type of hypervisor (bare metal vs. hosted, etc.)
  - Log files detailing cloud instance behavior
  - Cloud instance memory and disk state
  - VM introspection data (if available)

- **Administrative Domain**
  - Virtual disk images
  - Cloud instance memory

- **Cloud Storage**
  - Data stored by a cloud instance

- **Physical Systems**
  - Traditional acquisition of disks and memory
Forensics and Cloud Computing: Attack Vectors

- Traditional attacks against cloud instances
- Supply chain attacks against firmware and hardware of physical systems
- Virtualization break-out attacks
- Traditional insider threats within the consumer’s organization
- Malicious insiders at the cloud provider
- Malicious cloud providers
- Foreign espionage facilitated by offshore hosting and data storage
Forensics Challenges in the Cloud

- **Malware can hinder forensics**
  - Firmware rootkits and hardware attacks are highly effective because they are hard to detect and can be pervasive
  - If acquisition occurs at a higher level, such as the cloud instance, malware operating through the hypervisor can go undetected and hide evidence

- **Clouds are usually multi-tenant systems**
  - Searching through a constantly changing system with many simultaneous users
    - Potential legal issues for accessing information that is not the suspects’
      - Unclear definition of “in plain sight”
    - Potentially causing service interruptions may violate agreements by the cloud provider with other non-suspect customers
Cloud architectures vary between providers, affecting where evidence exists and how collection occurs

- Traditionally effective techniques can sometimes be used in cloud forensics (Zimmerman & Glavach, 2011)
- The level of access an examiner is granted can make techniques impossible that would otherwise work if they had full system access
- Data may be distributed across multiple jurisdictions

Locating evidence in a large and constantly changing system

- Static analysis on stationary data will need to take place in a formal data acquisition process (Zimmerman & Glavach, 2011)
- Performing static analysis of hardware and searching for virtual disk images will become exponentially more challenging as time passes from the incident to the start of an investigation
Cloud facts and issues

- Attack on Sony compromised more than 100 million customer accounts, the largest data breach in the United States
- Hackers hide their tracks beneath several layers of proxy servers that can span the globe. A recent attack against computers in South Korea was controlled from servers in more than 20 countries
- “The biggest issue is centralizing everything in the iCloud around an Apple ID,”
- Where's the data?
- Types of attacks (DOS, side channel, authentication, man-in-the-middle)
- iCloud secures your content by encrypting it when sent over the Internet, storing it in an encrypted format, and using secure tokens for authentication
- Secure Socket Layer (SSL exploits)
- In June 2011, Dropbox, which provides online storage, glitch allowed people to log into any Dropbox account for several hours by typing in any password
- Researchers at the Ruhr University in Bochum in western Germany managed to hack into private Clouds, using a weakness in the Eucalyptus software platform, which is widely used in Cloud services.
- As it becomes more common for law enforcement to raid online facilities like Megaupload, it is incumbent on law enforcement to respond to the needs of innocent users
The Ninth Circuit Court of Appeals, in its October 3, 2011 decision in *Suzlon Energy Ltd v. Microsoft Corporation*, has taken another step in defining the rights of people to protect their emails from being disclosed in civil court proceedings.

The question before the Suzlon court was whether a party can require a U.S. electronic communication service provider to produce emails stored on a U.S. server for the account of a non-U.S. national without regard to the safeguards and restrictions imposed by the *Electronic Communications Privacy Act* of 1986 (ECPA).

The court answered with a clear “no,” stating that the protections of the ECPA against unrestricted disclosure of emails by an electronic communication service provider apply to non-U.S. nationals as well as to U.S. citizens.
Discussion

- Does Locard’s Principle apply to cloud computing?
- What other evidence locations may exist in the cloud?
- How can forensically sound acquisition occur in the cloud?
  - Particularly when evidence may disappear as the system deallocates instances.
- Can forensic analysts effectively acquire data from cloud instances, when the implementations vary between providers?
- What standards should exist to guide forensic analysis of cloud environments?
- Others?


Questions

- Ken Zatyko, VP, Maryland Operations
  - zatykok@ainfosec.com or 410-570-0980
  - 5523 Research Park Drive, Suite 130, Baltimore, MD 21228