Incident Response in Amazon EC2: First Responders Guide to Security Incidents in the Cloud

Tom Arnold
Incident Response in Amazon EC2:  
First Responders Guide to Security Incidents in the Cloud

GIAC (GCFE) Gold Certification

Author: Tom Arnold, tom@paysw.com  
Advisor: Adam Kliarsky  
Accepted: April 19, 2016

Abstract

The Amazon Elastic Compute Cloud ("EC2") is one of the richest and most robust cloud environments. The full list of services gives the cottage entrepreneur the computing power and Internet presence of a Fortune 500 company. Although Amazon’s environment is very robust, humans are still a part of building and fielding the application, running on EC2; as such, a fully secure environment can hardly be assured. The likelihood of significant flaws in the applications, or configurations of the systems, opens the risk of a security breach or compromise that will require a security incident response. This paper examines the steps that a first responder should take in response to a detected security incident within Amazon EC2. Forensic examination as covered in FOR408 begins with a trusted, scientific acquisition of evidence to support the analysis and examination process. If a first responder blunders the impact can destroy important evidence; drive the attack to ground; and, leave their environment exposed. This paper is NOT a full discussion on the steps a forensic investigator should take in analyzing the incident; rather the focus is on the immediate action that an Amazon EC2 subscriber should prepare to take in advance of the forensic cavalry arriving on scene.
1. Introduction

As Head of Digital Forensics for Payment Software Company Inc. (“PSC”), a company that focuses exclusively on Clients that accept or process payments, we’ve responded to sites operating within cloud environments, most notably Amazon EC2. From the earliest case, it was identified that the Amazon environment held several unique services that could actually enhance an incident responder’s ability to rapidly locate, triage, and react to a security breach. Although most of the breaches encountered involved poorly configured applications (like the use of WebDAV, without suitable authentication) or lacking correctly configured network security and access controls over ssh, there was a stark contrast between companies that had an incident response plan crafted for EC2 and those that didn’t. Invariably, the companies that didn’t have a plan fitting for EC2 lost valuable data and risked not being able to fully contain the incident.

Establishing and maintaining the capability to respond to any form of a security incident is a daunting task for most organizations. The vast number of attack vectors range from Web-based application attacks, remote access account compromises, to spear phishing. Even anti-virus alerts demand different responses to issues as they are detected. Gone are the days of just pulling the network cable or disconnecting the power to a system. Each of these demand the organization detect the attack, and the first responder react appropriately. Several organizations work diligently to customize their plans and respond accordingly. The use of cloud-processing services for applications supporting an organization further complicate the task of planning and responding to a security incident. When critical information technology assets run in the Cloud, an IT department must consider how this impacts the basic response to a security incident.

Building on the basic principles of an incident response plan, the paper includes specific actions and steps a first responder should take that are inclusive of the Amazon EC2 environment.

1.1. Acknowledgements

The author expresses great appreciation to the following people for their help and thoughtful proofreading and peer review: Margarita Atlasova, William

1 For more information, please see website at: http://www.paysw.com/company
2. Refresh on what is an incident response plan?

Establishing a capability to respond to a security incident is necessary to rapidly detect security events, minimizing loss and destruction, mitigating the weaknesses that were exploited, and restoring information technology services (Computer Security Incident Handling Guide, August 2012).

This is a well-accepted objective for an incident response policy and plan. Let’s consider some of the key components of this before pushing forward. The first component is the rapid detection of a security incident. It is essential to highlight that this is the single most important fact confronting organizations today and has been the area of greatest weakness. In one recent case, the organization was shown to have been under the control of an outside bad-actor for 1,296 days. Detection of this major security event was only achieved by accident, when a system administrator noticed unusual user account on the system. For the remainder of this paper, an assumption will be made that the organization is capable of detecting a security incident on their electronic commerce servers operating in the Amazon EC2 environment.

The next concept is minimizing loss and destruction. For a moment, consider whether or not it’s possible for a volunteer fireman, responding to a house fire, to actually make the situation worse? What if the fireman sprays water on the chemical fire burning in a structure that the fireman believes is just a single-family dwelling? The obvious answer here is that some level of threat intelligence is needed about the security incident, so that a first responder can take appropriate action. There are many variables that need to be considered during this stage of an incident response before concluding what first steps a responder should take. Suffice it to say that there are several papers and discussions on appropriate response action and that probability of an inappropriate action resulting in increase loss or destruction is very high. Consider a case in this area, where the CIO screams at his tech-guys to power down the compromised system; thus starting the adventures of whack-a-mole as the malicious software pops up on a different machine. Before long, the entire...
environment was powered off and everyone in the company was looking at one another in an effort to post blame.

Third effort: mitigation of the weakness. It's important to know what it is you’re fighting before taking this step. The PCI Data Security Standard (Payment Card Industry Data Security Standard, 2015) is one of several standards for protecting sensitive data. Does this only mean that in response to an incident all you have to do is make certain that all components of a security standard are met? Now, reflect on the words of Professor Ian Angel: “It is sheer madness to impose the values of accountancy in so-called Information Audits, and then to imply that all decisions can be reduced to a form of algorithmic book keeping.” (Angel, I., 2005) In response to an incident, “If you know the enemy and know yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat. If you know neither the enemy nor yourself, you will succumb in every battle.” (Tsu, S & Giles, L, 2014) Prior to engaging in this step, it is crucial to have enough information to understand how the weakness is manifested and as much information as possible about the technical mechanics of the attack and incident. Given the fact that most incidents go un-discovered for a median of 205 days, (Mandiant/Fire Eye, 2015) there is no immediacy. This phase must be done correctly in order to lead to long-term containment. (Kral, P, 2011)

The final step is restoration of information technology services. All too often, this step is pushed to the top of the action list and systems restored too soon. A stark example of this was an e-commerce site. Customers were complaining that their credit card data had been stolen after shopping on the merchant’s site. The merchant went into the EC2 console, shut down the elastic beanstalks, terminated all servers, and then brought up new servers; with a declaration that the problem had been solved. Needless to say what the result of this approach was. Without clear knowledge of the nature of the attack and methods, the company spent a huge amount of money rebuilding the environment, only to have the same complaints flowing in after a few weeks.
3. What is an “Incident”?

This section is essential for any paper about digital forensic incident response. There are many conflicting views about what constitutes a computer security incident and this is a very hard topic to address.

After seeing over a hundred private clients each year and reviewing the work of 20+ consultants, some of the larger and highly complex organizations proclaim that they’ve never experienced a single computer security incident over the past year. They accuse our consultants of being too cynical; they’re all playing a very dangerous game by ignoring reality. In some cases, AV alerts have gone without response and active C^{2} channels are beaconing on their networks.

So what exactly is an incident? The strict definition of an incident is any violation of policy, law, or unacceptable act that involves information assets, such as computers, networks, smartphones, tablets, voice recording systems, cameras, and etc. (Bejtlich, R, 2005)

Consider the statement “any violation of policy, law or unacceptable act” draws a pretty clear distinction; so what does this mean in systemic terms? An activity on a computer system, like downloading and installing a piece of software, involves a sequence of events. (Yes; “event”, yet another hard-to-define term). In this case, refer to a “computing event”. These are individual actions or occurrences recognized by software that may be handled by software. Consider the following events, described in pseudo-program steps:

1. There are 15,000 unsuccessful login attempts; each 6 seconds apart
2. End user types name and password in login field, then hits enter
3. Login is successful
4. At the home screen, end user types in a search URL
5. Browser opens to a http site
6. End user clicks on a download link on the site
7. End user opens Windows Explorer
8. End user launches a software installer
9. Software installer opens a php script that exists on the local system

---

2 This term relates to a systemic event on a computer system that can be logged to a file. The most appropriate reference – without listing each operating system vendor – appears in: https://en.wikipedia.org/wiki/Event_(computing)
10. End user logs out
11. Each time a customer clicks on the “buy it” button, a connection to http://authorizet.com/ opens

Given this sequence of events, at which point have the events combined into a security incident? Actually, the first question should be: Do you believe a combination of these events may have resulted in a security incident? (Please answer yes) Is event number 1 the beginning of the security incident, even though all of the 15,000 logins failed?

This example is a clear case of a security incident that is hopefully detected quickly. Unfortunately, most of these actions go undetected for months and the log records describing these actions roll over and are overwritten. Frequently, the only evidence or facts is the presence of “funny” code in a php file found by a developer. And, the thousands of credit cards lost and the reputational damage are massive.

Unfortunately, some believe that the combination of these events alone do not comprise a security incident (because it doesn’t fit their strict definition that an incident involves proof data was exfiltrated). As we move forward in this paper, this example will be used as our sample event, and the EC2 environment shall consider that an incident response is warranted.

4. Amazon EC2

Amazon Elastic Compute Cloud (EC2) is an environment that allows users to rent computing resources to run their own applications. Use of the term environment, is actually a reference to a virtual machine that the user configures and is called an Instance. Instances can be deployed in zones, which are like geographic regions that provide closer proximity to end customers across the Internet. For instance, an Instance can be configured in the Western United States and then that instance also deployed in Western Europe. These zones are referred to as Availability Zones and are intended to get content and applications closer to the end customers who will leverage them. (Amazon Elastic Compute Cloud, n.d.) Although this is not an exhaustive list, there are a few interesting EC2 features that may impact any incident response to an AWS site:

- Elastic IP addresses. These allow the binding of an external IP address to an Instance without having to wait for DNS binding or
have a network engineer perform this. These IP addresses belong to the entire account and are not statically bound to a given Instance. In essence, they can be moved. As an example, this allows for a different mechanism for promoting Web applications to production.

- Operating systems supported. EC2 supports Linux, Open Solaris, MS Windows, BSD, and Amazon’s own Fedora distro called Amazon Linux.

- Persistent storage. In many cloud computing environments, the root volume (drive) is defined and related to the Instance, as the instance is created. As such, an instance device will be destroyed if the Instance is terminated or removed. EC2 allows for definition of a Elastic Block Store (EBS) device to be assigned as a root device. These EBS volumes can survive past the termination of the instance from where they are mounted.

- Elastic Beanstalk (EBS). This is a deployment framework that allows users to define applications and platforms that can be pushed out to any of a set of Amazon cloud services. The EBS handles capacity provisioning, load balancing, auto-scaling, and application health monitoring (AWS Elastic Beanstalk n.d.). Additionally, the EBS can create and destroy instances dynamically. This latter point is extremely important for any incident response condition.

4.1. What’s different about the EC2 cloud?

The Amazon Elastic Compute Cloud (EC2) is a remarkable computing environment that allows entrepreneurs of all size the access and opportunity to build their electronic commerce dream.

Consider a “traditional” e-commerce system stack and how it may be implemented in a hosting environment, colocation data center and the EC2 cloud. Here are the components:

Web server

Connected to the public Internet (hopefully through a firewall); responsible for displaying the site images and interacting with your customers.
Incident Response in Amazon EC2

Application server  The software where the major business logic takes place. This is the system that opens connections with your payment processor to accept payments. It’s also the system that runs the shopping cart.

Database server  This is where all the data resides about the products you offer, data about your customers, data about orders.

Other servers  This might be a WordPress server or another application server. This is where social media sites interface with your site. It’s also where you have your customer support forum.

Now consider how the systems are implemented and the differences between the implementation.

<table>
<thead>
<tr>
<th>Facility</th>
<th>Entry cost</th>
<th>Expansion impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional colocation</td>
<td>The company owns a firewall, switch and 4 servers. From the facility, rack</td>
<td>Company purchases more servers and infrastructure to keep up with expanding</td>
</tr>
<tr>
<td></td>
<td>space, network and air conditioning are rented to host the systems.</td>
<td>workload.</td>
</tr>
<tr>
<td>Hosting environment</td>
<td>A fixed e-commerce stack, using whatever shopping cart the hosting company</td>
<td>If your company expands to such a size, you have to leave the hosting company and</td>
</tr>
<tr>
<td></td>
<td>provisions, is rented and shared. A company has little or no option for</td>
<td>go for a traditional colocation</td>
</tr>
<tr>
<td></td>
<td>additional servers or different software as the hosting company may not</td>
<td></td>
</tr>
<tr>
<td></td>
<td>provision such a thing in their usual menu of services</td>
<td></td>
</tr>
<tr>
<td>Amazon EC2</td>
<td>The company rents time and obtains 4 server instances using block storage</td>
<td>Although Amazon claims it can do everything and scale massively, it comes with a</td>
</tr>
<tr>
<td></td>
<td>that is appropriate for your services.</td>
<td>cost. That said, load balancing and expandable instances are one approach</td>
</tr>
<tr>
<td></td>
<td></td>
<td>for the cottage entrepreneur to expand during that Christmas rush.</td>
</tr>
</tbody>
</table>

So, what’s really different about all of these? The big difference is that they are physical machine environments associated with the hosting and traditional colocation. In essence, a machine can be shutdown without any loss of data on the disk. (Of course this is a bad thing to do during incident response when you don’t know your enemy and what you’re fighting). In Amazon EC2, a system can be suspended, but once it’s terminated, it’s gone forever.
4.2. The incident

Let’s consider the incident against our e-commerce site running in EC2.

The site has a stable and persistent IP address associated with it. During setup, the site was configured with a Linux, Apache, MySQL, and PHP (“LAMP”) environment. All four systems run Amazon Linux. For remote administration, a secret SSH key was generated to authenticate and protect the communication channel. That said, the propagation of the secret key and other restrictions were cumbersome for software developers working in the Far East when they needed to perform software pushes or fix the shopping cart, so the company opened accounts using default SSH mechanisms for access. There was such turnover in managing individual accounts for the developers and they constantly complained about using the “sudo” command, so IT management decided that the default account for developers would be “root”.

One day, the newly-hired system administrator comes to IT Management after running a “last” command and reports that she sees something unusual. The output from the last command shows that there was a momentary successful SSH connection (less than a second) for the “root” account. The session terminated in less than a second and was from some other site.

Management discounts the system administrator’s observation as being some “technical anomaly” and “meaningless”. Our system administrator, not being satisfied, examines the syslogs and security logs during the same time period and discovers over 15,000 failed login attempts to the root account on the target system; where each attempt is exactly 6 seconds apart. After the successful connection, all the failed login activity ceases. She takes her findings to management.

Management again discounts the system administrator’s claim, claiming that nothing bad has happened yet. But during the 6 hours that management takes to ponder this situation, a login to the root account occurs again. Within seconds, the root user executes a “curl” command to reach an outside website. After that, the root user executes a wget command to download a small application. The user modifies code in the production application server, then logs out of the system.

After deliberating, IT Management comes back and orders the system administrator to change the root account password. She completes this request.
and sends the new password out to the developers and other authorized parties. Strangely, after changing the password, the failed login attempts discovered in the log start showing up again; a new password being tried every 6 seconds. Each attempt is from a variety of different IP addresses throughout the US and overseas.

Within the actual time it took to read the fourth paragraph above, the crime has occurred and a web shell installed on the server. Typing the commands for curl and wget is very easy. Editing the php script on the site is also very easy. This company is stunned when they receive a note from their bank advising them that their Web site has been compromised.

4.3. Incident response

The company’s incident response plan calls for the containment of the incident and then mitigation of the incident.

Given the notice from the bank, the company is in a full state of panic and very concerned about the potential fines levied by the bank for losing cardholder data; they’re also very concerned about their liability related to the loss of PII data and the need to notify consumers. The chairman of the company demands that action be taken to fix the situation now!

The system administrator proposes rebuilding the system instances from a known-good backup of the software. The action taken by the administrator is to login to the Amazon EC2 console and order that the volumes and instances for the four Linux machines be terminated. She creates four new machine instances with fresh volumes, and the back up is restored on the new machines. This process takes about 3 hours to complete.

During this time, the bank advises the company that they must have a forensic investigation to identify how the bad actors attacked the system and what took place.

At the end of the day, the forensic investigation is inconclusive and describes a situation where the evidence of the crime was lost because of the actions of the system administrator.

The company feels justified in the steps that were taken because they “cleaned” the problem up. The obvious issue here is that restoration of the...
software from a backup just re-introduced the modified PHP code and the bad-actor is the one back in business.

4.4. First responder

This section outlines the steps that a first responder should take when reacting to a security incident in the Amazon EC2 cloud. As used here, a “first responder” is the first technical person logging into either a system or the console in response to an identified security incident.

In our example, this is the system administrator, who is responding to the alarm. Unlike the failed incident response described earlier, this responder has now established a plan for how to respond. Remember, the bad-actor has been in the system for some period time. Do not panic! It’s important to take time planning what needs to done and how to proceed. Consider the impacts of all actions taken.

The first responder’s mission is to gather threat intelligence on what is happening and how it’s happening on the systems. Their job is to describe the enemy and what activities are going on. After collecting this information, a short-term containment plan is formulated and actions begin to limit the situation.

Here is the minimum plan of action for a EC2 first response:

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open a response journal</td>
<td>This is a written journal that is more like a folder on the local first responder’s workstation. This is where all of the known evidence and notes about the evidence are placed.</td>
</tr>
<tr>
<td>Obtain and review the steps of the EC2 first response plan</td>
<td>This step may be taking the plan from this paper, editing it to fit the environment, reviewing the plan with the other IT members and management, and then taking the action in the plan.</td>
</tr>
<tr>
<td>EC2 Response plan</td>
<td>This begins the first responder’s plan of action</td>
</tr>
<tr>
<td>Activate AWS CloudTrail through the account console</td>
<td>CloudTrail is an Amazon web service that records AWS API calls for the account and delivers a complete log file for all activity (AWS CloudTrail, n.d.). This is VERY important to journal ALL of the steps that a first responder takes on the EC2 console and with system objects.</td>
</tr>
<tr>
<td>Activate AWS Config through the account console</td>
<td>AWS Config is a managed service that provides an inventory, configuration history, and configuration change notifications (AWS Config, n.d.).</td>
</tr>
<tr>
<td>A. Use AWS Config in</td>
<td>This will discover resources that exist in the EC2</td>
</tr>
<tr>
<td>Step</td>
<td>Description</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>discovery mode</td>
<td>account and record their configuration. This will include information on resources that have been deleted. This comprehensive snapshot of all resources and their configuration attributes provides a complete inventory of resources.</td>
</tr>
<tr>
<td>B. Use AWS Config in continuous assessment</td>
<td>This will generate a set of reports related to the governance and compliance configurations for AWS resources under your control. This is a very important step that gives both the first responder and forensic investigator information about the environment itself.</td>
</tr>
<tr>
<td>Prepare a network diagram</td>
<td>This diagram is extremely important. This is NOT a logical topology, but a diagram of the network given the network interfaces on the instances. The reports from AWS config will help with this step. It is very important to document the full “as-is” environment. Make certain the network diagram includes ALL connected systems to the EC2 environment.</td>
</tr>
<tr>
<td>Scan with an external scanning agent</td>
<td>The PCI Security Standards Council has a program called Approved Scanning Vendor (“ASV”) (Payment Card Industry (PCI) Data Security Standard Approved Scanning Vendors: Program Guide Version 2.0, 2013). An ASV will allow you to scan your external IP addresses using the PCI scanning criteria. This is a good baseline standard for external scanning.</td>
</tr>
<tr>
<td>Scan with an application testing tool</td>
<td>There are several of these tools on the market that will scan a Web site, focusing on the OWASP (OWASP Top Ten Project, n.d.) and SANS CWE (SANS CWE, n.d.) criteria. These tools are very good at finding application vulnerabilities that can be exploited. Why are we scanning? Neither of these tools will find issues resulting in the brute force attack, but they may discover the injected code and other vulnerabilities that could be exploited against the company and their customers.</td>
</tr>
<tr>
<td>Capture PCAP files</td>
<td>Assuming that there is a Web server accepting all traffic and an IP gateway that the systems are assigned to send out-bound packets, it is very important to capture full tcpdump pcap files for each of these inbound and outbound interfaces. In the example case, these pcap files will show bad activity. They may also show if there is more than one bad-actor in the company’s system at the same time. Unfortunately, the downside is that analysis of large pcap files is time consuming for first responders and may need to be sent to the forensic investigator for detailed review.</td>
</tr>
<tr>
<td>Quick analysis (may take up to a day)</td>
<td>Analyze the scans and pcap data. This is a quick analysis and may not show bad activity without a</td>
</tr>
</tbody>
</table>
### Incident Response in Amazon EC2

**Step** | **Description**
--- | ---
| | deeper dive. For the pcap files, focus on filtering the egress IP’s from the systems. Filter the source IP as the system where the credit card data are collected from end-users. See if all of the destination IP addresses can be explained and identified.
| From the AWS Config reports, identify instances | Identify and document the InstanceID’s for all instances that are operating during the time of the potential security incident.
| Take snapshots of each instance | Take a snapshot of each instanceID. Be sure to document the snapshotID of each.
| Create volumes for each snapshot | For each of the snapshotID’s create a volume for each. This will be the disk volume that will be acquired for the deep forensic analysis. Make a clear note of the volumeID. At the end of this step, you should have a matrix of all the systems running in EC2 that list the following:
InstanceID ➔ SnapshotID ➔ VolumeID

**Most Important Note:** If your EC2 site runs in multiple AWS Availability Zones (AWS Availability Zone, n.d.), the top three steps MUST be performed in each zone. Remember, the bad-actor who controls a system in one Availability Zone is not restricted to that zone, they can move laterally into all of your zones (assuming that security controls allow such activity. Simple use of PSEXEC will make the movement between instances).

At this point, the basic steps for the first responder are done. The objective now is to analyze this data and make a decision on what short-term containment action to perform next. There are several containment choices that include, but are not limited to: (1) suspending the system instances; (2) building new system instances to take over processing; (3) wait for forensic response to take next steps; or, (4) follow management’s order to contain the systems.

If the choice is either containment or to shut the systems down, performing the next steps may yield some more interesting information and assist with the containment actions. The next steps involve capture of volatile data from the systems. This should be done prior to any suspension or deactivation of the servers. Volatile data is stored in the memory of the system and can expose several types of malicious software or root kits. Each of these tools may have droppers that will reinstall the evil malware upon reboot.

**Step** | **Description**
--- | ---
| Capture volatile data from each server | Volatile data are configurations and information stored in volatile memory. If the instance (server) is
<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>system</td>
<td>suspended or terminated, this data will immediately be lost. It is extremely important to capture this data.</td>
</tr>
<tr>
<td>A. For Windows</td>
<td>This means that a tool to capture the memory of the system will need to be run. If the system has 16gb of memory, the tool will be dumping a big file. So the problem here is to create a new Volume (large enough to hold the memory dump) in EC2 and attach that volume to the target instance (server). Once attached, this will be the drive where the dumped memory image will be written. There are several tools that can be used to achieve this. Work with the forensic investigator to find out what tool is their preference.</td>
</tr>
<tr>
<td>B. For Linux</td>
<td>Dumping Linux memory is not always the easiest thing to achieve and many variants of Linux don’t like to having their physical or logical memory read. As such, there are tools that will capture important information about processes, open files, network state and other areas. Create a new Volume in EC2 and attach that volume to the target instance (server). Once attached, this will be the drive that volatile data will be dumped. Should a memory dumping tool be the choice of action, setup a test environment to determine what, if anything, will happen to your system when the tool runs. Work with the forensic investigator to find out what tool is their preference.</td>
</tr>
<tr>
<td>Prepare for arrival of forensic investigator</td>
<td>This will be the virtual “arrival” of the forensic investigator. This is an important step where a EC2 Console account using AWS Identity and Access Management (AWS Identity and Access Management, n.d.) is used to create an account for the forensic investigation. This account should have full privileges to the EC2 console. All actions by this account will be tracked using AWS CloudTrail, which was turned on during an earlier step. It is very important that this account be enabled to perform all administrative actions, including but not limited to, creating instances, creating volumes, attaching volumes to instances, and creation and access to snapshots. This account will also need to have privileges to all Availability Zones. Even though the investigator will create an “investigative” instance in a specific Availability Zone, this account must be able to access volumes and objects within other Zones.</td>
</tr>
<tr>
<td>Document results</td>
<td>It is very important to document the Availability Zone and volumeID of each of the volumes used to capture the volatile data from the servers. Include notes on how this data is organized and identified, so the forensic investigator clearly understands which machines this data came from.</td>
</tr>
</tbody>
</table>
### 4.5. Investigative Evidence

Remember, the first responder is always a member of the company’s IT local staff. They are the person who is immediately available to do something about the reported incident. Their primary job is to begin short-term containment while preserving important evidence for the forensic investigator and analyst.

After the first responder has completed their work, they can produce the following items for the forensics response team:

1. A complete record and journal of ALL activity performed within the EC2 console and environment. The AWS CloudTrail generates this log and record, which was one of the first steps taken by the first responder.
2. Documentation on the environment, including: network diagram, details from the AWS Config system that has an inventory of all assets.
3. PCAP files from the tcpdump captures
4. Scan reports from the ASV and application testing
5. Listing of InstanceID ➔ SnapshotID ➔ VolumeID for every instance listed in the inventory
6. Volatile data for the servers and systems
7. Status report on the state of each instanceID (suspended, terminated, whatever)

From this data, the forensic investigator will ask to create an instance of a forensic workstation within the Availability Zone for the EC2 environment where the company’s snapshots and volumes exist. The investigator will use this environment to capture forensic images for full and deep analysis. The cool part of this step is that this activity can be performed remotely. Conceptually, this is the logical arrival of the investigator.

### 5. Lessons Learned and Preparation

In summary, having a clear and well-documented incident response plan that talks directly to the first responders is extremely important. Remember that cloud environments, like Amazon EC2, can present special challenges and opportunities for the responder. A response to one of these environments is quite
different than a response to one of the traditional data center or hosting environments.

In closing, it would be important to highlight:

- Know who the primary and secondary first responders are in your organization
- Have a clear plan for the response by the first responder
- Establish a relationship with a forensic investigator that understands EC2 and how to correctly respond.
- Have the investigator create a test instance and run a real incident response test
- Learn from all responses and update your plan after everything
- Don’t leave your head in the sand. Security incidents occur and cannot be fully prevented
References


Amazon Elastic Compute Cloud (n.d.) User Guide for Linux Instances

AWS Availability Zone (n.d.)


AWS Identity and Access Management (IAM). (n.d.)


*Computer Security Incident Handling Guide (SP 800-61r2).* (August 2012). NIST.


OWASP Top Ten Project (n.d.)

  https://www.pcisecuritystandards.org


SANS CWE (n.d.) https://www.sans.org/top25-software-errors/?cat=top25