Digital Forensics in the Cloud

Shams Zawoad, University of Alabama at Birmingham
Ragib Hasan, University of Alabama at Birmingham

Abstract. Today's cloud computing architectures often lack support for computer forensic investigations. Besides this, the existing digital forensics tools cannot cope with the dynamic nature of the cloud. This paper explores the challenges of digital forensics in the cloud, possible attacks on cloud-evidence, and mitigation strategies against those challenges.

Introduction

Cloud computing offers immense opportunities for business and IT organizations by providing highly scalable infrastructure resources, pay-as-you-go service, and low-cost on-demand computing. While clouds attract diverse organizations, the security and trustworthiness of cloud infrastructure has become a rising concern. Clouds can be a target of attacks or can be used as a tool to launch attacks. Malicious individuals can easily exploit the power of cloud computing and can perform attacks from machines inside the cloud. Many of these attacks are novel and unique to clouds.

To illustrate the use of clouds for malicious purpose, we consider the following hypothetical scenario:

Bob is a successful businessman who runs a shopping website in the cloud. The site serves a number of customers every day and his organization generates a significant amount of profit from it. Therefore, if the site is down even for a few minutes, it will seriously hamper not only their profit but also the goodwill. Mallory, a malicious attacker, decided to attack Bob's shopping website. She rented some machines in a cloud and launched a Distributed Denial of Service attack to the shopping website using those rented machines. As a result, the site was down for an hour, which had quite a negative impact on Bob's business. Consequently, Bob asked a forensic investigator to investigate the case. The investigator found that Bob's website records each visiting customer's IP address. Analyzing the visiting customer records, the investigator found that Bob's website was flooded by some IP addresses which are owned by a cloud service provider. Eventually, the investigator issued a subpoena to the corresponding cloud provider to provide him the network logs for those particular IP addresses. On the other hand, Mallory managed to collude with the cloud provider after the attack. Therefore, while providing the logs to the investigator, the cloud provider supplied a tampered log to the investigator, who had no way to verify the correctness of the logs. Under this circumstance, Mallory will remain undetected. Even if the cloud provider was honest, Mallory could terminate her rented machines and leave no trace of the attack. Hence, the cloud provider could not give any useful logs to the investigator.

Fig. 1: Process Flow of Digital Forensics

To identify the actual attacker in the above attack scenario, we need to execute digital forensics procedures in clouds. Currently, extensive research is going on to protect clouds from external or internal attackers. However, in case of an attack, we need to investigate the incident. Besides protecting the cloud, it is important to focus on this issue. Unfortunately, cloud forensics is not yet a popular research topic and there has been little research on adapting digital forensics for use in cloud environments. In this paper, we address the problems of cloud forensics and some mitigation strategies, which have significant real-life implications in investigating cloud-based cyber-crime and terrorism.

Understanding Cloud Forensics

NIST defines digital forensics as an applied science for "the identification, collection, examination, and analysis of data while preserving the integrity of the information and maintaining a strict chain of custody for the data" [1]. Figure 1 illustrates the process flow of digital forensics. Cloud forensics can be defined as applying all the processes of digital forensics in the cloud environment. Ruan et al. defined cloud forensics as a subset of network forensics [2], because cloud computing is based on extensive network access, and network forensics handles forensic investigation in private and public networks. However, cloud forensics also includes investigating file systems, process, cash, and registry history. Different steps of digital forensics shown in Figure 1 vary according to the service and deployment model of cloud computing. For example, the evidence collection procedure of Software-as-a-Service (SaaS) and Infrastructure-as-a-Service (IaaS) will be different. For SaaS, we solely depend on the Cloud Service Provider (CSP) to get the application log. In contrast, in IaaS, we can acquire the virtual machine image from customers and can initiate the examination and analysis phase. In the public deployment model, we rarely can get physical access to the evidence, but this is guaranteed in the private cloud deployment model.

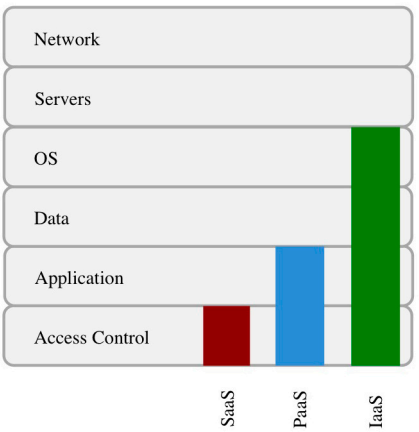

Fig. 2: Customers' control over different layers in different service model

Why Are Clouds Not Forensics Friendly?

Several characteristics of cloud computing complicate the process of cloud forensics. As the storage system is no longer local, law enforcement agents cannot confiscate the suspect's computer and get access to the digital evidence even with a subpoena. In a cloud, each server contains files from many users. Hence, it is not feasible to seize servers from a data center without violating the privacy of many other benign users. Moreover, even if the data belonging to a particular suspect is identified, separating it from other users' data is difficult. The trustworthiness of the evidence is also questionable, because other than the cloud provider's word, there is no usual way to link a given evidence to a particular suspect. The following issues make cloud forensics challenging.

• In traditional computer forensics, investigators have full control over the evidence (e.g., router logs, process logs, and hard disks). Unfortunately, in a cloud, the control over data varies in different service models. Figure 2 shows the control of customers in different layers for the three different service models – IaaS, PaaS, and SaaS. Cloud users have highest control in IaaS and least control in SaaS. This physical inaccessibility of the evidence and lack of control over the system make evidence acquisition a challenging task in the cloud. For example, in SaaS, customers do not get a log of their system, unless the CSP provides the logs. In PaaS, it is only possible to get the application log from the customers. To get the network log, database log, or operating system log we need to depend on the CSP. In IaaS, customers can only get the operating system logs, they do not have access to network or process logs. For example, Amazon does not provide load balancer logs to the customers [3], and it is not possible to get MySql log data from Amazon's Relational Database Service [4].

• Cloud computing is a multi-tenant system, while traditional computing is a single owner system. To give an analogy, the cloud can be compared to a motel, while the other can be compared to a personal house. In a cloud, multiple Virtual Machines (VM) can share the same physical infrastructure, i.e., data for multiple customers can be co-located. An alleged suspect may claim that the evidence contains information of other users, not her. In this case, the investigator needs to prove to the court that the provided evidence actually belongs to the suspect. Conversely, in traditional computing systems, a suspect is solely responsible for all the digital evidence located in her computing system. Moreover, in the cloud, we need to preserve the privacy of other tenants. The multi-tenancy characteristic also brings novel side-channel attacks [5] that are difficult to investigate.

• Volatile data cannot sustain without power. Data residing in a VM are volatile, as after terminating a VM, all the data will be lost. In order to provide the on demand computational and storage service, CSPs do not provide persistent storage to VM instances. There is, though, a way to preserve VM data by storing an image of the VM instance. An attacker can exploit this vulnerability in the following way: after doing some malicious activity (e.g., launch DoS attack, send spam mail), an adversary can terminate her VM that will lead to a complete loss of the evidence and make the forensic investigation almost impossible. A malicious user can also fraudulently claim that her instance was compromised by someone else who had launched a malicious activity. In the absence of any evidence, it will be difficult to prove her claim as false via a forensic investigation [6].

• Chain of custody is one of the most vital issues in traditional digital forensic investigation. Chain of custody should clearly depict how the evidence was collected, analyzed, and preserved in order to be presented as admissible evidence in court [7]. In traditional forensic procedure, it is trivial to maintain an access history of time, location, and person to access the computer, hard disk, etc. of a suspect. On the other hand, in a cloud, we do not even know where a VM is physically located. Also, investigators can acquire a VM image from any workstation connected with the internet. The Investigator's location and a VM's physical location can be in different time zones. Hence, maintaining a proper chain of custody is challenging in clouds.

• Currently, investigators are completely dependent on CSPs for acquiring cloud evidence. However, the employee of a cloud provider, who collects data on behalf of investigators, is most likely not a licensed forensics investigator and it is not possible to guarantee his integrity in a court of law. A dishonest employee of a CSP can collude with a malicious user to hide important evidence or to inject invalid evidence to prove the malicious user is innocent. On the other hand, a dishonest investigator can also collude with an attacker. Even if CSPs provide valid evidence to investigators, a dishonest investigator can remove some crucial evidence before presenting it to the court or can provide some fake evidence to the court to frame an honest cloud user. In traditional storage systems, only the suspect and the investigator can collude. The three-way collusion in the cloud certainly increases the attack surface and makes cloud forensics more challenging.

Requirements For Forensics-Enabled Cloud

To mitigate the challenges that we discussed above, we identified the following characteristics that a forensics-enabled cloud should have:

• As CSPs do not provide persistent storage to VMs, turning off or rebooting a VM will eventually lose all the data residing in that VM. Data that are volatile in nature must be stored in persistent databases so that even if a malicious user terminates her virtual machine, we can still gather the evidence. One possible solution to this problem is that CSPs will provide a continuous synchronization API to customers. Using this API, customers can preserve the synchronized data to any cloud storage e.g., Amazon S3, or to their local storage. However, if the adversary is the owner of a VM, this mechanism will not work. Trivially, she will not be interested in synchronizing her malicious VM. To overcome this issue, CSPs by themselves can integrate the synchronization mechanism with every VM and preserve the data within their infrastructure. CSPs can constantly monitor all the running VMs and store the volatile data in a persistent storage. The volatile data can be network logs, operating system logs, and registry logs. When a VM is in active state, CSPs can track which data belongs to which VM. Hence, while preserving the data, CSPs can take care of segregating the data according to VM owner. In this way, multiple VM owners' data will not be co-mingled.

• After preserving all the evidence, CSPs need to ensure the integrity of the evidence in order to prevent collusion between CSPs, investigators, and cloud users. Without integrity preservation, the validity of the evidence will be questionable and the defense and the jury can object about it. Generating a digital signature on the collected evidence and then checking the signature later is one way to validate the integrity. Another way is preserving the proofs of past data possession [8]. Preserving the proofs of files can significantly decrease the continuous synchronization cost and at the same time ensure the integrity and confidentiality of cloud evidence. Trusted Platform Module (TPM) can also protect the integrity of cloud evidence. By using a TPM, we can get machine authentication, hardware encryption, signing, secure key storage, and attestation. It can provide the integrity of the running virtual instance, trusted logs, and trusted deletion of data to customers.

• Besides preserving the integrity of evidence, CSPs also need to provide proper chain of custody information. As provenance provides the history of an object, by implementing cloud provenance, CSPs can provide the chronological access history of evidence, how it was analyzed, and preserved, which can ensure the chain of custody for cloud forensics. However, as all the evidence and the access histories are under the control of CSPs, they can always tamper with the provenance record. Moreover, from the provenance data in the cloud, an attacker can learn confidential information about the data stored in the cloud. To protect provenance information from these types of attack, we need a secure provenance scheme [9].

• Considering CSPs are preserving all the evidence, investigators will be still dependent on CSPs to collect evidence, as all the cloud evidence resides in the providers' data center. CSPs can play a vital role in this step by providing a web-based management console or providing secure API to law enforcement agencies. Using web console or API, customers as well as investigators can collect network, process, database logs, and other digital evidence as well as the provenance records of those evidence.

Moving Towards Regulatory Compliant Cloud

As cloud computing does not provide the facility of proper forensics investigations, it cannot be used to store healthcare, business, or national security-related data, which require audit and regulatory compliance. Auditability is a vital issue to make the cloud compliant with the regulatory acts, e.g., The Sarbanes Oxley (SOX) Act [10] or The Health Insurance Portability and Accountability Act (HIPAA) [11]. According to SOX, financial information must reside in auditable storage that the CSPs cannot provide currently. Business organizations cannot move their financial information to a cloud, as it does not comply with the SOX act. As cloud infrastructures do not comply with HIPAA's forensic investigation requirement, hospitals also cannot move their patients' confidential medical records to cloud storage. A forensics-enabled cloud architecture that satisfies all the requirements stated in the previous section will definitely increase the auditability of a cloud environment. By deploying such an architecture, we will be able to store and provide the types of evidence from which we can get all the activities of cloud users.

Business and healthcare organizations are the two most data consuming sectors. Hence, cloud computing cannot reach its goal without including these two sectors. These sectors are spending extensively to make their own regulatory-compliant infrastructure. A regulatory-compliant cloud can save this huge investment. We need to solve the audit compliance issue to bring more customers into the cloud world. Implementing an architecture that allows cloud forensics investigations will make clouds more compliant with such regulations, leading to widespread adoption of clouds by major businesses and healthcare organizations.

Conclusion

In this paper, we discussed the technical challenges of executing digital forensic investigations in a cloud environment and presented the requirements to make clouds forensics-friendly. Collecting trustworthy evidence from a cloud is challenging as we have very little control over clouds compared to traditional computing systems. For now, investigators need to depend on the CSP to collect evidence from a cloud. To make the situation even worse, there is no way to verify whether the CSP is providing correct evidence to the investigators, or the investigators are presenting valid evidence to the court. Thus, we need to build a trust model to preserve the trustworthiness of evidence. For forensics data acquisition, CSPs can shift their responsibility by providing a robust API or management console to acquire evidence. However, the CSPs need to come forward to resolve most of these issues. Creating a secure model for cloud forensics is very important as it will lead to more trustworthy clouds, allowing their adoption in sensitive application domains such as defense, business, and healthcare.

Acknowledgement:

This research was supported by a Google Faculty Research Award, the Office of Naval Research Grant #N000141210217, the Department of Homeland Security Grant #FA8750-12-2- 0254, and by the National Science Foundation under Grant #0937060 to the Computing Research Association for the CI Fellows Project.

REFERENCES

1. K. Kent, S. Chevalier, T. Grance, and H. Dang, "Guide to integrating forensic techniques into incident response," NIST Special Publication, pp. 800–86, 2006.
2. K. Ruan, J. Carthy, T. Kechadi, and M. Crosbie, "Cloud forensics: An overview," in proceedings of the 7th IFIP International Conference on Digital Forensics, 2011.
3. AWS, "Amazon web services," <http://aws.amazon.com>, [Accessed July 5th, 2012].
4. R. Marty, "Cloud application logging for forensics," in In proceedings of the 2011 ACM Symposium on Applied Computing. ACM, 2011, pp. 178–184.
5. T. Ristenpart, E. Tromer, H. Shacham, and S. Savage, "Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds," in Proceedings of the 16th ACM conference on Computer and communications security. ACM, 2009, pp. 199–212.
6. D. Birk and C. Wegener, "Technical issues of forensic investigations in cloud computing environments," Systematic Approaches to Digital Forensic Engineering, 2011.
7. J. Vacca, Computer forensics: computer crime scene investigation. Delmar Thomson Learning, 2005, vol. 1.
8. S. Zawoad and R. Hasan, "Towards building proofs of past data possession in cloud forensics," ASE Science Journal, 2012.
9. R. Hasan, R. Sion, and M. Winslett, "Preventing history forgery with secure provenance," ACM Transactions on Storage (TOS), vol. 5, no. 4, p. 12, 2009.
10. Congress of the United States, "Sarbanes-Oxley Act," <http://thomas.loc.gov>, 2002, [Accessed July 5th, 2012].
11. Centers for Medicare and Medicaid Services, "The health insurance portability and accountability act of 1996 (hipaa)," <http://www.cms.hhs. gov/hipaa/>, 1996, [Accessed July 5th, 2012].

ABOUT THE AUTHORS

Shams Zawoad is working as a graduate research assistant in SECuRE and Trustworthy Computing Lab (SECRETLab) and a Ph.D. student at the University of Alabama at Birmingham (UAB). His research interest is in cloud security especially in cloud forensics, and in location provenance. He received his B.Sc. in Computer Science and Engineering from Bangladesh University of Engineering and Technology in 2008. Before joining UAB, Zawoad had been working in software industry and developed authentication and authorization framework for several critical business applications, including a online payment system of Bangladesh Post Office.

Phone: 205-915-4262
E-mail: zawoad@cis.uab.edu

Ragib Hasan, Ph.D. is a tenure-track Assistant Professor at the Department of Computer and Information Sciences at the University of Alabama at Birmingham. With a key focus on practical computer security problems, Hasan explores research on cloud security, mobile malware security, secure provenance, and database security. Hasan is the founder of the SECuRE and Trustworthy Computing Lab (SECRETLab) at UAB (http://secret.cis.uab.edu). He is also a member of the UAB Center for Information Assurance and Joint Forensics Research. Prior to joining UAB in Fall 2011, Hasan was an NSF/CRA Computing Innovation Fellow and Assistant Research Scientist at the Department of Computer Science, Johns Hopkins University. He received his Ph.D. and M.S. in Computer Science from the University of Illinois at Urbana Champaign in October, 2009, and December, 2005, respectively. Before that, he received a B.Sc. in Computer Science and Engineering and graduated summa cum laude from Bangladesh University of Engineering and Technology in 2003. He is a recipient of a 2013 Google RISE Award, a 2012 Google Faculty Research Award, the 2009 NSF Computing Innovation Fellowship and the 2003 Chancellor Award and Gold Medal from Bangladesh University of Engineering and Technology. Dr. Hasan's research is funded by the Department of Homeland Security, the Office of Naval Research, and Google. Hasan is also the founder of Shikkhok.com – a grassroots movement and platform for open content e-learning in South Asia.

Phone: 205-934-8643
E-mail: ragib@cis.uab.edu