APPROVAL SHEET

Title of Dissertation: Digital Forensics for Infrastructure-as-a-Service Cloud Computing

Name of Candidate: Josiah Alexander Bradford Spoor Dykstra
Doctor of Philosophy, 2013

Dissertation and Abstract Approved: ________________________
Alan T. Sherman
Associate Professor
Department of Computer Science and
Electrical Engineering

Date Approved: ______________________________

CURRICULUM VITAE

Name: Josiah Alexander Bradford Spoor Dykstra.
Degree and date to be conferred: Doctor of Philosophy, May 2013.
Secondary Education: Heelan High School, Sioux City, IA, 1998.
Collegiate institutions attended:

Hope College, B.S., computer science, 2002.
Hope College, B.A., music, 2002.
Iowa State University, M.S., information assurance, 2004.

Major: computer science.
Professional publications:

1. Dykstra, J. and A.T. Sherman, "Design and Implementation of FROST: Digital Forensic Tools for the OpenStack Cloud Computing Platform," *DFRWS Annual Digital Forensics Research Conference*, August, 2013, Monterey, CA (paper accepted).

2. Dykstra, J. "Search Warrant Language for Cloud Computing," In *Proceedings of the Annual Meeting of the American Academy of Forensic Sciences*, Vol. 19, February, 2013, Washington, DC.

3. Dykstra, J, "Seizing Electronic Evidence from Cloud Computing Environments," *Cybercrime and Cloud Forensics: Applications for Investigation Processes.* Ed. Keyun Ruan. Hershey: IGI Global, 2013. 156-85.

4. Dykstra, J. and D. Riehl, "Forensic Collection of Electronic Evidence from Infrastructure-As-A-Service Cloud Computing," In *Richmond Journal of Law and Technology*, Vol. 19, Issue 1, 2012.

5. Dykstra, J. and A.T. Sherman, "Acquiring Forensic Evidence from Infrastructure-as-a-Service Cloud Computing: Exploring and Evaluating Tools, Trust, and Technique," In *Proceedings of the DFRWS Annual Digital Forensics Research Conference*, August, 2012, Washington, DC., S90-S98.

6. Dykstra, J. "Understanding Issues in Cloud Forensics: Two Hypothetical Case Studies," In *Proceedings of the Annual Meeting of the American Academy of Forensic Sciences*, Vol. 18, February, 2012, Atlanta, GA.

7. Dykstra, J. and A.T. Sherman, "Understanding Issues in Cloud Forensics: Two Hypothetical Case Studies," In *Proceedings of the 2011 ADSFL Conference on Digital Forensics, Security, and Law*, 2011, Richmond, VA, 191-206.

8. Dykstra, J. and A.T. Sherman, “Understanding Issues in Cloud Forensics: Two Hypothetical Case Studies,” In *Journal of Network Forensics* , Vol. 3, Issue 1, 2011, 19-31.

9. Dykstra, J., “A framework for network covert channel detection.” Thesis (M.S.)–Iowa State University, 2004.

Professional positions held:

United States Department of Defense. Global Network Exploitation and Vulnerability Analyst. (2004–Present).

Iowa State University. Research Assistant. (2002–2004).

ABSTRACT

Title of Dissertation: Digital Forensics for Infrastructure-as-Service
Cloud Computing

Josiah Alexander Bradford Spoor Dykstra,

Doctor of Philosophy, 2013

Dissertation directed by: Alan T. Sherman, Associate Professor
Department of Computer Science and
Electrical Engineering

We identify important issues in the application of digital forensics to Infrastructure-as-a-Service cloud computing and develop new practical forensic tools and techniques to facilitate forensic exams of the cloud. When investigating suspected cases involving cloud computing, forensic examiners have been poorly equipped to deal with the technical and legal challenges. Because data in the cloud are remote, distributed, and elastic, these challenges include understanding the cloud environment, acquiring and analyzing data remotely, and applying the law to a new domain. Today digital forensics for cloud computing is challenging at best, but can be performed in a manner consistent with federal law using the tools and techniques we developed.

The first problem is understanding how and why criminal and civil actions in and against cloud computing are unique and difficult to prosecute. We analyze a digital forensic investigation of crime in the cloud, and present two hypothetical case studies that illustrate the unique challenges of acquisition, chain of custody, trust, and forensic integrity. Understanding these issues introduces legal challenges which are also important for federal, state, and local law enforcement who will soon be called upon to conduct cloud investigations.

The second problem is the lack of practical technical tools to conduct cloud forensics. We examine the capabilities for forensics today, evaluate the use of existing tools including

EnCase and FTK, and discuss why these tools are incapable of trustworthy cloud acquisition. We design consumer-driven forensic capabilities for OpenStack, including new features for acquiring trustworthy firewall logs, API logs, and disk images.

The third problem is a deficit of legal instruments for seizing cloud-based electronically-stored information. We analyze the application of existing policies and laws to the new domain of cloud computing by analyzing case law and legal opinions about digital evidence discovery, and suggest modifications that would enhance cloud the prosecution of cloud-based crimes. We offer guidance about how to author a search warrant for cloud data, and what pertinent data to request.

This dissertation enhances our understanding of technical, trust, and legal issues needed to investigate cloud-based crimes and offers new tools and techniques to facilitate such investigations.

Digital Forensics for Infrastructure-as-a-Service Cloud Computing

by

Josiah Alexander Bradford Spoor Dykstra

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, Baltimore County in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2013

To Alicia

ACKNOWLEDGMENTS

Alan Sherman has been a helpful advisor, knowledgeable guide, and ardent supporter. He never failed to encourage me to strive for perfection in my writing and research. My work is better because of him.

Thank you to my distinguished committee for guidance and wisdom—Dr. Forno; Dr. Nicholas; Mr. Flynn; Dr. Garfinkel. Thanks also to the UMBC Cyber Defense Lab for listening to my presentations and giving thoughtful feedback. I am particularly indebted to Timothy Leschke, who shared many successes and frustrations with me along the way.

My academic work was made entirely possible through the support and encouragement of the Department of Defense. I appreciate their dedication to continuing education. Thanks especially to supervisors and colleagues who were flexible in allowing me to take the time to work on my research.

Thanks to many friends, colleagues, teachers, and mentors that have reviewed papers and offered helpful advice, especially Matt Georgy, Eoghan Casey, and Mark Rasch. Damien Riehl was a brilliant co-author on Chapter 6 and advisor on legal issues.

More than anyone else, my deepest gratitude and love goes to my wife, Alicia, who encouraged and supported me through this process. She endured many hours of me at the computer and technical conversations over dinner.

Contents

List of Tables

List of Figures

Listings

Chapter 1

Introduction

Cloud computing poses serious implications for judicial policy and practice in the United States and around the world. Crime committed using cloud computing resources and against cloud infrastructures is inevitable. Though real incidents have already taken place against cloud providers such as Google [79], few crimes using the cloud or targeting it directly have been publicized or litigated thus far. Forensic investigators must understand that current tools and techniques are inadequate in the cloud environment where acquisition and analysis will be executed very differently than is done today.

Companies are embracing cloud technology to offload some of the cost and maintenance equipment that they would otherwise have purchased themselves. Cloud infrastructure, with exceptional bandwidth, storage, and computing power, offers an attractive prize for hackers as well. While many have lamented how the users of the cloud and their data lack protection [39, 40, 7], few of these discussions have considered the difficulty of responding to security breaches, including forensics and criminal prosecution.

1.1 Motivation

Cloud computing, while still an emerging technology, will inevitably expand in the coming years. For example, the US Government's "Cloud First" policy mandates that federal agencies must consider cloud solutions [46]. Despite this enthusiastic embrace, few policy makers, law enforcement, and forensic investigators understand the issues or approach to investigating incidents and crimes in the cloud.

Vendors and researchers have explored security relating to data stored in clouds, and some have begun to discuss incident response. No one, however, has fully explained or developed tools to carry out forensics for cloud infrastructure. Understanding these issues is important especially for federal, state, and local law enforcement who will soon be called upon to conduct cloud investigations. This is non-trivial work because the cloud presents not one challenge that can be addressed with technology alone, but many that are interconnected. Further, there are no examples of cloud investigations to illustrate the issues and to educate practitioners. This dissertation addresses those shortcomings by explaining the technical and legal issues, and offering the first concrete solutions.

Cloud computing poses several challenges that are specific and non-trivial to the forensic investigation. First, acquisition of data for analysis is complicated by the fact that the data are remote and controlled by a third-party. Furthermore, physical seizure of evidence may be impractical or impossible given that multiple tenants may reside on a single hard drive, or a single tenant's data may be distributed across many hard drives. Second, cooperation with cloud providers may ultimately determine how successful and thorough the forensic investigation is. A provider that is open about its capabilities and infrastructure, and willing to collect and share forensically-relevant logs, packet captures, or configuration information, will make the investigation much easier than will a provider that is unwilling to assist. Legal assistance in the form of service level agreements and search warrant templates are needed to assist customers, law enforcement, and the judiciary. Third, current forensic tools are unsuited to handle the volume or type of forensic data collected from cloud environments, and today's cloud infrastructure management tools do not offer forensic capabilities.

Popular and wide-spread forensic tools such as Guidance EnCase and AccessData Forensic Toolkit (FTK) are designed to analyze hard drive images. However, cloud providers aiding an investigation are likely to produce raw user data in unknown formats or proprietary virtual machine (VM) images that are unsupported by these tools. State and local law enforcement agencies are already burdened by heavy workloads and small budgets. While they may have dealt with online service providers such as Facebook or Gmail (which can loosely be called Software-as-a-Service cloud offerings), it is almost certain that these agencies would be unprepared to properly investigate a crime that utilized cloud-based infrastructure offerings or cloud-based data storage. Law enforcement today could spend countless hours and dollars trying to analyze hard drives for evidence of cloud usage, to explore data made available by cloud providers, or manually to discern what changed in a VM over time. Currently there are no tools to assist the forensic examiner with these tasks.

Practical forensic tools for cloud computing are possible, but they require a thorough understanding of the issues and careful implementation in real cloud environments. It is possible that existing forensic tools, already in the hands of certified examiners, may fill some requirements. However, they must be carefully evaluated against the added complexity of remote, provider-controlled cloud layers.

The press coverage surrounding cloud security reflects a broad interest in the subject. Our contributions will be immediately useful to the adoption of cloud technology, as a result of reasoned knowledge about forensic investigations and practical technical and legal solutions to address them.

1.2 Thesis Statement

I adopt the following thesis statement:

> *Digital forensics for Infrastructure-as-a-Service (IaaS) cloud computing environments cannot currently be performed in a manner that is consistent with forensic practice federal law today, but can be performed in a manner that is consistent with clarified federal law by the development of practical forensic tools and techniques to facilitate forensic examinations of the cloud.*

This statement reflects two points. First, forensics for cloud computing is different and cannot simply be executed the same way as traditional digital forensics is done today. This shortcoming is problematic since criminal activity in the cloud must be investigated. Second, we will show that forensics for cloud computing is possible given a reexamination of the law and with new tools to empower investigators and law enforcement in their required tasks.

To bound the scope of the work, we consider only one of service models for cloud computing: Infrastructure-as-a-Service. This choice is fully explained in Chapter 2.

1.3 Contributions of this Dissertation

We have conducted research around three key areas: exploration of the issues arising from application of forensics to cloud computing, practical deployment of forensic tools for cloud computing, and analysis of legal issues with a sample search warrant. These three areas are bound together by the overarching goal of conducting digital forensics for cloud computing. Towards that goal, we first developed an approach to reason about the problem, then developed new technical solutions, then prepared the legal community for prosecution of these crimes. Each area expands the body of knowledge about digital forensics and offers unique and timely contributions to the interdisciplinary cloud computing community.

In addition to satisfying the requirements of the forensic examiner, our solutions address the requirements to satisfy the courts and the cloud providers. The courts require integrity and authentication of digital evidence. Cloud providers require scalable solutions that complement the elasticity and scalability of their cloud services. These goals are not always compatible, but we strive to balance the needs of all stakeholders.

Our specific contributions are as follows.

1.3.1 Identification of Cloud-Specific Forensic Issues Using Two Hypothetical Case Studies

The inevitable vulnerabilities and criminal targeting of cloud environments demand an understanding of how digital forensic investigations of the cloud can be accomplished. We present two hypothetical case studies of cloud crimes: child pornography being hosted in the cloud, and a cloud-based website compromised by a hacker. Through the analysis of these scenarios, we highlight shortcomings of current forensic practices and laws. We describe significant challenges with cloud forensics, including forensic acquisition, evidence preservation and chain of custody, and open problems that drive the next phases of the research.

1.3.2 Evaluation of Existing Tools for Cloud Forensics and Analysis of Trust in Cloud Evidence

We expose and explore technical and trust issues that arise in acquiring forensic evidence from IaaS cloud computing and analyze some strategies for addressing these challenges.

First, we create a model to show the layers of trust required in the cloud. Second, we present the overarching context for a cloud forensic exam and analyze choices available to an examiner. Third, we provide for the first time an evaluation of cloud-based acquisition using popular forensic acquisition tools including Guidance EnCase and AccessData Forensic Toolkit, and show that they can successfully return volatile and non-volatile data from the cloud. We explain, however, that with those techniques judge and jury must accept a great deal of trust in the authenticity and integrity of the data from many layers of the cloud model. In addition, we explore four other solutions for acquisition: Trusted Platform Modules, the management plane, forensics-as-a-service, and legal solutions. These alternatives assume less trust but require more cooperation from the cloud service provider. Our work lays a foundation for future development of new acquisition methods for the cloud that will be trustworthy and forensically sound. We suggest that the cloud management plane is a strong candidate for forensic tools because it provides useful forensic data, does not require trust in the guest operating system, can be user-driven, and scales for many cloud customers. Our work also helps forensic examiners, law enforcement, and the court evaluate confidence in evidence from the cloud.

1.3.3 Development of Three Forensic Tools for OpenStack

We describe the design, implementation, and evaluation of FROST—three new forensic tools for the OpenStack cloud platform. Operated through the management plane, FROST provides the first dedicated forensics capabilities for OpenStack, an open-source cloud platform for private and public clouds. Our implementation supports an Infrastructure-as-a-Service cloud and provides trustworthy forensic acquisition of virtual disks, API logs, and guest firewall logs. Unlike traditional acquisition tools, FROST works at the cloud management plane rather than interacting with the operating system inside the guest virtual machines, thereby requiring no trust in the guest machine. We assume trust in the cloud provider but FROST overcomes non-trivial challenges of remote evidence integrity by storing log data in hash trees and returning evidence with cryptographic hashes. Our tools are user-driven, allowing customers, forensic examiners, and law enforcement to conduct investigations without necessitating interaction with the cloud provider. We demonstrate through examples how forensic investigators can independently use our new features to obtain forensically-sound data. Our evaluation demonstrates the effectiveness of our approach to scale in a dynamic cloud environment, making it one of the first "carrier grade" forensic tools. The design supports an extensible set of forensic objectives, including the future addition of other data preservation, discovery, real-time monitoring, metrics, auditing, and acquisition capabilities.

1.3.4 Analysis of Legal Challenges in Cloud Forensics and a Sample Search Warrant

We illuminate legal problems in the United States for electronic discovery and digital forensics arising from cloud computing and argue that cloud computing challenges the process and product of electronic discovery. We investigate how to obtain forensic evidence from cloud computing using the legal process by surveying the existing statues and recent cases applicable to cloud forensics. Using one of our hypothetical case studies, we illustrate the difficulty in acquiring evidence for cloud-related crimes. For the first time, we create a sample search warrant that could be used in this case study, and which provides sample language for agents and prosecutors who wish to obtain a warrant authorizing the search and seizure of data from cloud computing environments. Finally, we present a contrasting view and discusses how defense attorneys might be able to challenge cloud-derived evidence in court.

1.4 Outline

Chapter 2 provides background information that is helpful to understand the remainder of this dissertation. We provide a general introduction to cloud computing, a brief review of related work on digital forensics, and a survey of related works in the law as it relates to cloud computing and forensics.

The remainder of the dissertation is organized into two logical parts: technical issues and proposed solutions, and legal issues and proposed solutions.

Chapters 3, 4, and 5 form Part I. These chapters explore how and why criminal actions in and against cloud computing are unique and difficult to prosecute, and how to conduct forensic examinations of those crimes. We present two hypothetical case studies to reason about the current state of digital forensics for cloud-related crimes. Finally, we describe contributions to the OpenStack cloud platform that enable incident response and forensics. The text in this section is largely from [18], [20], and [21], which are co-authored with Alan T. Sherman.

Chapters 6 and 7 form Part II and cover the legal analysis of cloud forensics and a sample search warrant. We examine legal options for obtaining evidence from the cloud. The text in Chapter 6 is taken from [17], which is co-authored with attorney Damien Riehl. Chapter 7 comes from [15].

We conclude in Chapter 8. Appendix A contains our sample search warrant.

Chapter 2

Background and Related Work

Digital forensics for cloud computing is an intersection of many fields. In this chapter we provide a brief overview and background of the primary technologies and disciplines used throughout the dissertation. We begin with a primer on the types and characteristics of cloud computing in Section 2.1. We provide an overview of digital forensics, and related work in digital forensics, in Section 2.2. We continue in Section 2.3 with an introduction to law and related legal analysis as they relate to digital forensics. We conclude in Section 2.4 with cryptographic concepts used in the dissertation.

2.1 Cloud Computing

This section presents background and related work in cloud computing.

2.1.1 Background

Cloud computing is a broad, generic term with many meanings and definitions. It has infiltrated the vernacular and has been bastardized in marketing and media. It would be unfair to say that cloud computing refers to anything particular other than that it is not the computing device physically in your possession. The National Institute of Standards and Technology (NIST) has the most widely cited definition, which is itself an evolving and non-trivial explanation. This is a living document, and has already gone through more than 15 versions. The first part of the current definition reads:

> "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (*e.g.* networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models" [63].

Cloud computing is an evolution in the history of computers. Starting with single-user standalone computers to multi-user mainframes, the most direct ancestors of cloud computing were utility [9] and grid computing [26].

"Cloud" is a generic term that refers to a network where the physical location and inner workings are abstracted away and unimportant to the usage. "The cloud" was first used to describe telecommunication networks, where the consumer was blissfully unaware of the inner workings of how their telephone conversation was transmitted to the remote end. In early computer networks, it was used to distinguish the boundary between what the provider was responsible for and what the users were responsible for [53]. Most recently it has been used to describe the Internet specifically. Knowing the physical location of `gmail.com`, for example, is unimportant to using that service. Cloud computing also takes advantage of this definition of cloud, as it is also a service connected to a network, often the Internet. However, cloud computing offers specific services whereby customers utilize shared remote computing resources such as processing power or data storage, and provision those resources themselves.

Often overlooked and under-explained is the fact that cloud computing is actually, though not explicitly, broken down into two forms of service. First is data intensive cloud computing, which is the concept of breaking up large computing jobs into smaller subtasks and computing each piece on a different computer. Google's MapReduce [14] and the open-source Hadoop [1] project are examples of this paradigm. Second is utility cloud computing, which describes more generic computing resources (*e.g.*, hard drives, CPUs, etc.) that are exposed to customers as a utility. Microsoft Azure and Amazon Elastic Compute Cloud (EC2) are examples of this paradigm.

Cloud computing as a utility is further broken down into three service models. Each model represents a different separation between how much of the infrastructure the consumer controls and how much the provider controls. *Infrastructure-as-a-Service* is the model for which the consumer has the most control. An Amazon executive described this as the provider having ownership and control of "the concrete to the hypervisor." The consumer has administrator privileges of an entire operating system and everything that runs therein. *Platform-as-a-Service* takes away the consumer's control of the operating system, and *Software-as-a-Service* takes away their control of the application. Unlike Gmail or Facebook, which provide static, concrete services to users, IaaS cloud computing is a canvas that programmers can use to create any service they like.

Four defining characteristics of IaaS cloud computing are of particular importance to my discussion: on-demand self-service, rapid elasticity, location independence, and data replication. First, the customer has complete control over the provisioning and de-provisioning of cloud resources, which they can do quickly and on-demand. Second, because of this ease and elasticity, evidence can appear and disappear at a moment's notice at the customer's bequest. Third, like other resources on the Internet, the physical location of the cloud resources has no bearing on use or provisioning of those resources, which could exist in one or more data centers around the world. Finally, to provide data reliability and fault-tolerance, cloud providers routinely replicate data on several computers in multiple physical locations. Further, cloud environments typically store data in a distributed filesystem and break single files into pieces that could be stored on multiple, independent hard drives.

2.1.2 Related Work

Research on cloud computing to date has focused on service offerings, such as resource allocation strategies, load balancing, large data analysis, and the use of cloud technology in other disciplines including medicine and higher education. Security research for cloud computing is heavily weighted toward data security and privacy, data and service availability, and compliance. This is not unexpected, since those issues most immediately affect standing concerns about adopting cloud computing. However, they are short-sighted and fail to anticipate the investigative tasks required when systems are compromised.

2.2 Digital Forensics

This section presents background and related work in digital forensics.

2.2.1 Background

Digital forensics is a branch of forensic science that uses of scientific and proven methods to analyze and interpret information from digital devices in the reconstruction of criminal events. The job of the forensic examiner is to analyze the digital information and reconstruct a timeline of events that describes, as best as possible, what happened, when it happened, and who did it. While the forensic examiner could be asked to analyze single documents or email messages, the more common traditional task is full hard drive analysis.

Digital forensics holds a unique place in the wider world of forensics distinguished by its meaning, and the dilution of the terminology is detrimental to the field. In forensic odontology or forensic anthropology, for example, forensic investigators are concerned with applying their discipline to evidence of crimes and answering questions of interest to the legal system. In particular, these questions relate to how a crime was committed or how an individual died. "Digital forensics," on the other hand, has come to encompass a wide variety of activities. The term is so encompassing that it often refers to non-legal questions. Some people would say that any file recovery, such as an accidentally deleted term paper, is an example of digital forensics. Others would say that enforcing corporate policy is digital forensics, such as investigating an employee's computer to see if he or she were violating corporate policy against checking sports scores during work hours. This ambiguity in the application of terminology threatens the credibility of the discipline. However, if practitioners maintain consistency in the use of terminology, there should be no loss in credibility when others use the same forensic techniques for other purposes.

In a 2012 online survey of forensic experts—primarily forensic investigators—we found that 61% agreed or strongly agreed with the way the phrase "digital forensics" is used today [16]. The respondents overwhelming felt that digital forensics did not need to involve a civil or criminal offense. However, of five published definitions of "digital forensics," they most agreed with those including the phrases "reconstruction of events found to be criminal" (43.8%) or "in a manner that is legally accepted" (39.3%).

The difficulty lies in the fact that there is no other accepted term to describe forensic-like digital investigations. Both legal and non-legal investigations may use the same software, procedures, and techniques. When the investigation must be legally sound, additional requirements are levied on the process, including chain-of-custody and authenticity. Our survey showed that given five alternative phrases, respondents preferred “digital investigation” and “digital examination” (Figure 2.1).

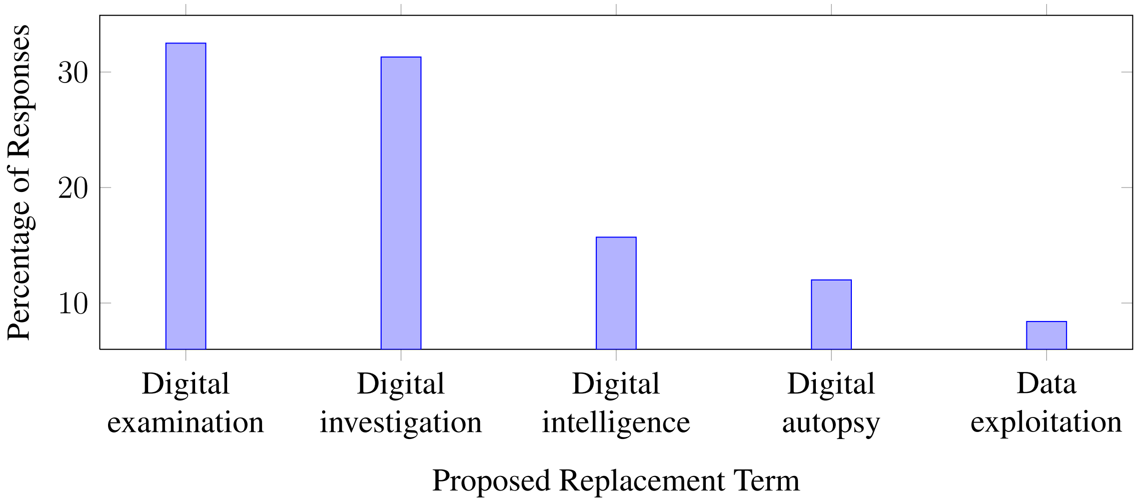


Figure 2.1: Survey results for the question “Which of the following terms would best describe forensic-like activities which are not intended for legal process and are not bound by legal soundness?”

For the purposes of these discussions, we assume that digital forensics is concerned with the acquisition and analysis of digital evidence to inform legal proceedings. Digital forensics is an umbrella term for any digital data that encompass sub-disciplines such as computer forensics, network forensics, database forensics, mobile device forensics, and video forensics. Even modest crimes involving digital devices require blending these disciplines, since nearly every computer is interconnected to another. Cloud computing, by its nature, draws upon computer forensics and network forensics since a networked computer is always involved. Other digital forensic disciplines may also be involved depending on the crime.

In 2013, the NIST Cloud Forensics Working Group began an effort to define “cloud computing forensic science” [64]. The Working Group has not yet agreed upon a definition. For the purposes of this dissertation, we define *cloud forensics* as the application of scientifically-based methodologies for the investigation of events which use or target cloud computing. We also limit the scope in this document to cloud forensics for Infrastructure-as-a-Service cloud computing, even though cloud forensics is necessary for all cloud models.

2.2.2 Related Work

Despite significant research in digital forensics, little has been written about the applicability of forensics to cloud computing environments. Furthermore, no case law exists on which to extrapolate the desire of the courts on the matter. In 2010, Garfinkel [29] suggested that "Cloud computing in particular may make it impossible to perform basic forensic steps of data preservation and isolation on systems of forensic interest." In one of the only published books on cloud forensics, Lillard [51] approaches cloud forensics as a matter of network forensics combined with remote disk forensics. Nevertheless, traditional disk forensic tools are not discussed. While legal complications are introduced, including cloud-based evidence admissibility, no solutions are presented. Wolthusen [88] identified some research challenges, including "discovery of computation structure," "attribution of data," "stability of evidence," and "presentation and visualization of evidence." Lu, et al. [52] and Zho, et al. [90] introduced the idea of data provenance for clouds, applied both to cloud security and data forensics. In 2009, researchers at UC San Diego demonstrated that it was possible to locate a particular VM in Amazon EC2 and mount side-channel attacks by co-locating a new VM with the target [68]. This yielded only crude information. In 2012, this work was extended to show that it was possible to extract cryptographic keys using the side-channel [89].

Public incidents involving cloud computing have skirted the issue of direct forensic investigation of cloud infrastructure. In 2009, Google and 34 other companies were hacked and infected with data-stealing malware. While the attack at Google involved Gmail, a cloud-based email service, the vulnerabilities and exploits were reportedly end-user based and not attacks on the cloud [79]. Researchers recently demonstrated using Amazon's EC2 cloud platform to crack passwords quickly and cheaply, a potentially criminal activity [8]. In 2010, presenters at the DEFCON Conference used EC2 to launch a demonstration denial of service against a small network [48]. In the investigation of individual users, cloud providers have begun to offer services that aid law enforcement. For example, Facebook has the option to download a users entire profile and history on the site [23]. However promising this may be for an investigator, these data cannot be said to be forensically sound. Guidance Software, the maker of EnCase, has produced a training video showing how to recover and analyze Facebook chat artifacts from a local hard drive [33].

The US federal government evaluates some of the most widely used forensic tools to ensure reliability. The National Institute of Standards and Technology's (NIST) Computer Forensic Tool Testing (CFTT) project is charged with testing digital forensic tools, measuring their effectiveness, and certifying them [59]. They evaluated EnCase 6.5 in September 2009, and FTK Imager 2.5.3.14 in June 2008 [62, 61]. They have never tested nor certified the enterprise versions of these products that include remote forensic capabilities. NIST also publishes a Digital Data Acquisition Tool Specification, which "defines requirements for digital media acquisition tools in computer forensic investigations" [60]. The most recent version of the specification was written in 2004, before cloud computing as we know it existed.

Several researchers have pointed out that evidence acquisition is a forefront issue with cloud forensics [19, 18, 69, 80]. Ruan suggested that evidence collection should obey "clearly-defined segregation of duties between client and provider," though it is unclear who should collect volatile and non-volatile cloud data and how. Another lamented about the lack of appropriate tools for data from the cloud, noting that "Many of these tools are standardized for today's computing environment, such as EnCase or the Forensics Tool Kit" [69].

Virtual machine introspection (VMI) is a technique whereby an observer can inspect a virtual machine from the outside through the hypervisor. It was first demonstrated in 2003 as a technique for intrusion detection [30]. In 2009, Symantec presented research on using VMware's VMsafe to inject anti-virus code into a virtual machine from the VMware hypervisor [12]. That same year, researchers proposed applying introspection to live forensic investigations. Terremark is reported to use introspection for monitoring, management and security for their vSphere cloud computing offering [70]. So far no attempt has been made to inject a forensic tool, such as an EnCase servlet, into a virtual machine from the hypervisor.

In 2009, Gartner published a short overview of remote forensic tools and guidance for their use, targeted at enterprise environments [37]. They cited EnCase and FTK as the most widely used products, with the greatest international support. These tools are not without their faults, however: in 2007, a vulnerability was found in the authentication between the remote EnCase agent and the server [31]. In 2011, Guidance Software published a comprehensive examination of legal issues and decisions about electronic discovery [34]. As of that date, the publication had no mention of judicial decisions or statutory law related to the complex legal questions surrounding remote data acquisition.

In addition to disk images, forensic investigators use metadata and system logs to reconstruct an event. Metadata and system logs are typically generated by the operation system's normal operation, rather than as a forensic-specific task. Nevertheless, the logs are useful in an investigation and easily gathered. Consumers of cloud services have few tools available for accessing low-level logs to the cloud infrastructure. Cloud providers and researchers encourage application-level logging [54] and Google, Amazon, and Microsoft allow customers to enable logging for stored object accesses [32, 5, 56]. To our knowledge, no cloud provider makes available customer accessible API call audit logs or VM firewall logs. That is, a customer has no way to know if, when, and from what IP address his or her credentials were used to make API calls.

Data integrity is a critical component of the forensic process. Other authors have developed proposals for ensuring integrity on untrusted machines, such as third-party servers. Clarke [11] proposed a method for validating the integrity of untrusted data using hash trees and a small fixed-sized trusted state. This method differs from our method because it does not check the integrity of subsets of the data. SUNDR (Secure untrusted data repository) [49] is a filesystem for storing data securely on untrusted servers. However, SUNDR requires that each client of the filesystem is able to see each other's file modifications.

Other research has focused on storing content securely on untrusted servers, which could then produce trustworthy forensic data, even from third-party cloud providers. Haber, *et al.* [35] explored in depth the redaction of subdocuments from signed original data, while preserving the cryptographic link of integrity between the two datasets. Haber posited that audit logs can be considered an append-only database, and that an audit report is essentially a database query with certain entries redacted. The proposed redactable signature algorithm is precisely applicable to the cloud logs we will encounter, though it must take into account a constantly changing dataset.

The dissertations of Crosby [13] and Kundu [47] bear striking similarity to our goals despite different motivations. Crosby proposed history tree tamper-evident logs, and suggested that they could "increase the trust in software service and 'cloud computing.'" Kundu was interested in authenticating subsets of signed data objects without leaking structural information about the data structures. Our work was influenced by these designs. We assume that the logger is trusted, and we use our enhanced logging mechanism simply for efficient log storage, retrieval, and integrity validation.

Other recent and emerging research in digital forensics deals with problems shared by non-cloud environments and cloud environments. These issues include the challenges of growing data volumes, encrypted data, solid state hard drives, and triage.

2.3 Law

This section presents background and related work in law.

2.3.1 Background

Forensics is a process governed by legal principles. Digital forensic data falls under the legal umbrella of electronically-stored information (ESI). ESI is information created, manipulated, communicated, stored, and best utilized in digital form, requiring the use of computer hardware and software. Two sets of rules govern criminal and civil procedures, including how ESI can be seized (*e.g.*, search warrants). The Federal Rules for Criminal Procedure (FRCrP) [85] govern criminal prosecutions in United States district courts. The Federal Rules of Civil Procedure (FRCP) [83] are the procedural rules rules that govern civil procedure in US district courts.

The Fourth Amendment to the Constitution protects America citizens and their property:

> “The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.”

This Amendment is invoked when cloud consumers have an expectation of privacy for their content, which we will assume that they do. One Supreme Court case, *Katz v. United States* [2], further extended the application of the Fourth Amendment to protect individuals with a "reasonable expectation of privacy." On the contrary, Fourth Amendment doctrine, known as the "third party doctrine," holds that individuals who entrust data to a third party (*e.g.,* an ISP) relinquish any expectation of privacy. We will examine the implications of the Fourth Amendment more deeply in Chapter 6.

Two laws will feature prominently in future chapters. The Electronic Communications Privacy Act of 1986 (ECPA), codified at 18 U.S.C. §§ 2510-2522, describes the ways the government can collect and use electronic data. ECPA has important implications in the way we construct a search warrant in Chapter 7. The Stored Communications Act (SCA), codified at 18 U.S.C. §§ 2701-2712, describes the voluntary and compelled disclosure of data held by third parties. The SCA contains two definitions that will be important in Chapter 6 because determining which applies to cloud computing determines how data can be seized. The statute defines an *electronic communication service* (ECS) as "any service which provides to users thereof the ability to send or receive wire or electronic communications." It also defines a *remote computing service* (RCS) as "the provision to the public of computer storage or processing services by means of an electronic communications system."

The legal system in the United States is greatly influenced by the precedent of case law. *Stare decisis* is a concept whereby courts rule using the precedence of previous court rulings to guide decisions on cases that are similar in nature. Similarly, new rulings have the potential to shape how other courts rule in the future. As Orton, Alva, and Endicott-Popovsky noted, "Since the field of cloud forensics has not established best practices to the level of similar fields such as digital forensics, early rulings could be based on faulty science, complicating the development of adequate case law and setting faulty first precedent" [67].

Specific legal principles that relate to civil and criminal proceedings, including jurisdiction (the area over which a legal body has authority) and venue (the location where a case is heard) will be introduced in Chapter 6.

2.3.2 Related Work

Cloud computing raises unanswered legal questions and issues critical to a forensics exam. No case law exists for scholars to extrapolate the desire of the courts on the explicit applicability of forensics to cloud computing environments. Lillard [51] discusses legal complications of cloud-based evidence admissibility, but presents no solutions. Lawyers and computer scientists have expressed views about remote forensics, a field that closely relates to the cloud. Schwerha and Inch [74] survey legal analysis and case law, as well as a list of remote forensic software, but undertake no application to cloud computing. Law professor Orin Kerr [43] has written extensively on the applicability of the Fourth Amendment to electronic evidence and the Internet. His suggestions on search warrant language for shared resources, such as cloud computing, is relevant to cloud forensic research. Cloud providers are only now beginning to think about compliance practices for subpoenas and search warrants (beyond email), despite the fact that cooperation in the interception of communications for law enforcement purposes today may be difficult. Forensic investigations of cloud-related crimes are likely to fall in the federal domain, given their cross-jurisdictional nature.

We are unaware of any published template for writing a search warrant for cloud data. In 2006, a California attorney published an article titled "Search Warrant Language for Cellular Phones," describing how to obtain data from cell providers [57]. Several search warrants have appeared in the press for services such as Facebook [87] and Gmail [38]. The Department of Justice Search and Seizure Manual [86] includes sample subpoenas, orders, and warrants which we used for guidance, but none of the these were for cloud data.

2.4 Cryptography

A limited number of cryptography concepts are used in this dissertation. They are: cryptographic hashes, hash trees, and public-key cryptography, and digital signatures.

Cryptographic hash functions are one-way algorithms that takes an arbitrary-length message as input and returns a fixed-length output, know as the *hash* or *digest* [73]. A hash is ideally easy to compute. It should be infeasible to generate a message with a given hash. Ideally it is infeasible to modify a message without changing the hash of that message, and infeasible to find two different messages with the same hash value. Digital forensic tools and processes often use the SHA-1 hash function, which produces a 160-bit digest [25].

Hash functions are the basis for hash trees. A hash tree or Merkle tree is a data structure in which every non-leaf node is labelled with the hash of the labels of its children nodes [55]. Hash trees are useful because they allow efficient and secure verification of the contents of larger data structures, such as security logs.

Cloud computing environments commonly employ public-key cryptography, using key pairs to encrypt and sign requests between users and the provider. Each user generates (or is given) a pair of cryptographic keys, a public encryption key and a private decryption key. These keys can be used both for encryption and for digital signatures [73]. Cloud consumers often sign messages to the cloud provider, providing authentication and non-repudiation.

Part I

Issues and Solutions for Digital Forensics of Cloud Computing

Chapter 3

Two Hypothetical Case Studies

In this chapter we consider the investigative response and forensic process of two hypothetical, but plausible, case studies of crimes tied to cloud computing. In Section 3.1 we discuss the applicability of forensic frameworks. Section 3.2 contains our case studies. The first explores a case of child pornography in the cloud, and the trouble with both acquiring and analyzing data. The second case study deals with the cloud as the target of a crime, and the complex issues of chain of custody and trust. We examine issues of attribution, forensic integrity, and chain of custody in Section 3.3, and we conclude in Section 3.4.

3.1 Existing Forensic Frameworks

To frame the approach of forensic investigation of any environment, including the cloud, it is helpful to have a procedure that guides the activity. The cloud environment does not affect the need for a framework, and does not inherently demand a new one. Frameworks for the digital forensic investigation are plentiful: at least 14 have been published since 1995 [77]. Digital forensic labs often choose a combination of approaches, or develop their own process that considers their particular personnel, workload, and budget. The generality of many investigative frameworks makes them applicable under a variety of circumstances and irrespective of technology. While there is hardly a generic computer forensic case that would lend itself to routine and standardized steps, in practice the general forensic process for a particular type of crime tends to look similar each time. For example, the examination of digital artifacts to find evidence of child pornography almost always involves taking a bit-for-bit hard drive image and searching common file system locations and slack space for contraband images.

Consider the “Guide to Integrating Forensic Technique into Incident Response” published by NIST [41]. The NIST process, like others, can be roughly summarized as follows:

- Collection
- Examination
- Analysis
- Reporting.

Collection involves the process of physical acquisition of data. Examination is the process of combing through the data for items of interest. Analysis is the application of the interesting items to the investigative question at hand, and whether it supports or refutes that question. Reporting describes the output of analysis, including the analysis steps taken.

3.2 Case Studies

We have developed two hypothetical case studies to reason about the state of digital forensics for cloud-related crimes. While fictional, they describe computer crimes that are not uncommon today. Case Study 1 uses the cloud as an accessory to a crime. Case Study 2 targets the crime against the cloud. These crimes require a reinterpretation when set in a cloud computing environment. In both scenarios, the following themes emerge that differentiate these investigations from traditional digital forensics:

- Acquisition of forensic data is more difficult.
- Cooperation from cloud providers is imperative for collecting comprehensive forensic data.
- Current forensic tools appear unsuited to process the volume and format of unstructured data stored in clouds.
- Cloud data may lack key forensic metadata.
- Chain of custody is more complex.

We will return to address these issues in more detail in Section 3.3.

3.2.1 Case Study 1

Polly is a criminal who traffics in child pornography. He has set up a service in the cloud to store a large collection of contraband images and video. The website allows users to upload and download this content anonymously. He pays for his cloud services with a pre-paid credit card purchased with cash. Polly encrypts his data in cloud storage, and he reverts his virtual webserver to a clean state daily. Law enforcement is tipped off to the website and wishes both to terminate the service and prosecute the criminal.

This is a case where the computer is incidental to the offense. Let us assume that the cloud model used in this case is IaaS, such as Amazon EC2. In this service model, the provider has responsibility and access to only the physical hardware, storage, servers and network components. In the public interest, law enforcement first contacts the cloud provider with a temporary restraining order to suspend the offending service and account, and a preservation letter to preserve evidence pending a warrant.[1] Tracking down the user is the more difficult task. The onus in this case is on the forensic examiner to piece together a circumstantial case based on the data available.

The examiner has no way to image the virtual machine remotely since the cloud provider does not expose that functionality, and in doing so would alter the state of the machine anyway. Deploying a remote forensic agent, such as EnCase Enterprise, would require the suspect's credentials, and functionality of this remote technique within the cloud is unknown. Today the forensic examiner, with no case law or standard methodology on the matter, may be tempted to attempt standard practices in digital evidence collection. Namely, with proper recording and documentation, the examiner accesses the offending website and takes snapshots or videotaping the collection of the evidence, and saving the web pages locally. Simply viewing the target website is enough to confirm that the content is illegal, but it tells us nothing about who put it there. Additionally, no guarantee can yet be made that the target webserver has not been compromised by an attacker, or that the examiner's request to the web server was not the victim of DNS poisoning, man-in-the-middle, or some other alteration in transit.

Consider other possible sources of digital evidence in this case: credit card payment information, cloud subscriber information, cloud provider access logs, cloud provider NetFlow logs, the web server virtual machine, and cloud storage data. Law enforcement can issue a

[1] 18 U.S.C. §2703(f)(1) ("A provider of wire or electronic communication services or a remote computing service, upon the request of a governmental entity, shall take all necessary steps to preserve records and other evidence in its possession pending the issuance of a court order or other process.")

search warrant to the cloud provider, which is adequate to compel the provider to provide any of this information that they possess. Law enforcement need not execute or witness the search.[2] The warrant specifies that the data returned be an "exact duplicate," the forensic term that has historically meant a bit-for-bit duplication of a drive. Since child pornography is a federal offense, the provider must comply with the order. A technician at the provider executes the search order from his or her workstation, copying data from the provider's infrastructure and verifying data integrity with hashes of the files. Files may have been distributed across more than one physical machine, but they are reassembled automatically as the technician accesses them. Though the prosecution may call the technician to testify, we have no implicit guarantees of trust in the technician to collect the complete data, in the cloud infrastructure to produce the true data, nor in the technician's computer or tools used to collect the information correctly. Nonetheless, the provider completes the request, and delivers the data to law enforcement.

Let us say that Polly had two terabytes of stored data.[3] To transfer that quantity of data, the provider saves it to an external hard drive and delivers it to law enforcement by mail. In addition, the provider is able to produce: account information, 10MB of access logs, 100MB of NetFlow records, and a 20GB virtual machine snapshot. After validating the integrity of the data, the forensic examiner is now charged with analysis.

We would expect the forensic expert to identify the following that would aid in prosecution:

[2] 18 U.S.C. §2703(g)("... the presence of an officer shall not be required for service or execution of a search warrant issued in accordance with this chapter requiring disclosure by a provider of electronic communications service or remote computing service of the contents of communications or records or other information pertaining to a subscriber to or customer of such service.")

[3] Interestingly, 18 U.S.C. §2703(b) allows a cloud provider to disclose the contents of an account used for remote storage without a warrant, and without notifying the customer or subscriber. Kerr suggested that this is unconstitutional [43].

- Understand how the web service works, especially how it encrypts/decrypts data from storage.
- Find keys to decrypt storage data, and use them to decrypt the data.
- Confirm the presence of child pornography.
- Analyze logs to identify possible IP addresses of the criminal.

It is not unreasonable to expect that this activity may take many man hours to analyze. According to performance testing from the manufacturer, AccessData found that their FTK product took 5.5 hours to process a 120GB hard drive fully on a top-of-the-line workstation, and as long as 38.25 hours on a low-end workstation [4]. At that rate, 2TB of data could take 85 hours of processing time. The examiner is likely to dive in first to the data store. The provider may have returned individual files or large files containing “blobs” of binary data. In either case, it will become quickly evident that the data are encrypted. Tools such as EnCase and Forensic Toolkit can analyze VMware data files but not snapshots which include suspended memory. The human analyst will have to fix-up and run the VM snapshot in order to understand the website source and observe how encryption is used. Once the keys are uncovered, and data are decrypted, 2TB of data must be analyzed for evidence. We were already aware of illegal content, but not aware of the data owner. Timestamps or file metadata may prove useful, provided they are available and accurate. Evidence of the owner may be gleaned from NetFlow, timestamp, and potentially in the coding style of the website. We can safely assume that an IP can be found that points to Polly. All of the forensic analysis is documented and presented to counsel.

In the absence of legal precedent, existing case law must be considered in the forensic process used. In 2007, the 100-page opinion by Judge Grimm in *Lorraine v. Markel* issued guidance about the admissibility of original or duplicates of original evidence, as legislated in Rules 1001-1008 of the Federal Rules of Evidence [3]. As mentioned above, service providers are already empowered to conduct searches on behalf of law enforcement. Several important issues regarding the issuance of a warrant were omitted above.

- Search warrants must specify the search of a person or location for evidence of a crime. With cloud computing, a problem emerges because the data may not be location-specific, other than a known public-facing URL or the cloud provider hosting the data. A search warrant must describe the physical place to be searched with particularity.[4] This becomes further complicated if cloud resources are distributed across state or international boundaries. We explore this issue in more depth in Chapter 7.
- The Fourth Amendment presents a preposterous assumption about search preceding seizure, which the courts may be compelled to reinterpret. As Kerr has explored extensively, traditional digital evidence collection is the reverse process of seizure then search [42]. Further, digital evidence, and especially cloud evidence, is never "seized" in the sense that it ceases to exist in one place, but the data are the target of the seizure, which are copied and the original remains. We return to this issue in Chapter 6.

Given the procedure undertaken above, consider the issues which the defense may raise to introduce doubt in the examination:

- Since raw bit-for-bit copies of hard drives were not provided, how do we know that the cloud provider provided a complete and authentic forensic copy of the data? Can

[4] Search warrants for online webmail have traditionally specified only the email address as the "place to be searched." See the search warrant for a Gmail mail account at http://docs.justia.com/cases/federal/district-courts/michigan/miedce/2:2009mc50275/237762/2/

the authenticity and integrity of the data be trusted? Can the cloud technician, his/her workstation and tools be verifiably trusted?

- Were the data located on one drive, or distributed over many? Where were the drives containing the data physically located? Who had access to the data, and how was access control enforced? Were the data co-mingled with other users' data?
- If data came from multiple systems, are the timestamps of these systems internally consistent? Can the date and time stamps be trusted, and compared with confidence?
- Does the virtual machine have a static IP address? How can the prosecution tie the malicious activity on the virtual machine to Polly?
- What jurisdiction governs the data in question? If the cloud provider's jurisdiction, then which of their geographic locations or data centers?

Some of the digital evidence collection from the cloud mirrors traditional collection. In other respects the process is new, such as data dispersed over many storage systems and virtual machine use. Current tools are ill-equipped to process the data in this case easily. The case in almost every respect hinges on how the cloud provider cooperated. Without greater transparency into how the provider operates, it is difficult or impossible to counter the above objections from the defense.

Finally, we note that cloud providers have a legal obligation to purge child pornography from their systems. Many providers keep duplicate copies of stored data, which here requires that they know where all copies are located and how to verifiably delete the contraband. Even if human employees at the cloud provider are unaware of where each bit of data are, the computers that implement the cloud environment must be able to locate and reconstruct cloud data. Microsoft and Amazon declined to comment about their compliance abilities in this situation.

3.2.2 Case Study 2

Mallory is a hacker who intends to exploit victims by placing a malicious webpage in the cloud. She uses a vulnerability to exploit the cloud presence of Buzz Coffee, a legitimate company. From there, she installs a rootkit that injects a malicious payload into web pages displayed, and hides her malicious activity from the operating system. She then redirects victims to the website, which infects them with malware. Users complain to the legitimate company that they are being infected, so the company seeks to fix the problem and investigate the crime.

This example is a different type of computer crime, one where the target is the computer. Let us assume that Buzz Coffee uses a Software-as-a-Service provider, such as RackSpace. In this service model, the provider has responsibility and access to the hardware, the operating system, and the hosting platform. Buzz wishes to make an example of this hacker, and hires a lawyer to prosecute the attacker. The attorney contracts a forensic specialist to conduct the digital investigation. Using experience as a guide, the investigator constructs a plan to access the cloud provider remotely over a secure channel using Buzz Coffee's credentials and retrieve the website source files. However, when the data are returned, nothing malicious is found since Mallory's rootkit hid the files from the host operating system and the provider's APIs. The forensic investigator determines that the following are additional possible sources of data: cloud provider access logs, cloud provider NetFlow logs, and the web server virtual machine.

The prosecutor approaches the cloud provider with a subpoena and requests all of this data, including a forensic copy of the virtual machine.[5] The provider is willing to conduct an internal investigation; however, it is reluctant to produce the raw data citing confidential and proprietary information. In fact, the Service Level Agreement lacks any language requiring compliance with intrusion response or remediation. The attorney is able to convince a judge that there is likely evidence of a crime inside the cloud, and a search warrant is issued to the provider.[6] Even in this case, the provider complies to the extent that its legal counsel feels is appropriate, which in this case includes: NetFlow logs, web access logs, and files from the virtual machine that comprise Buzz Coffee's website. Any further data from the operating system or hosting platform, they claim, would threaten their business and competitive advantage.

A technician at the provider executes the court order from his workstation, copying data from the provider's infrastructure and verifying integrity with MD5 hashes. This information is burned to DVD, and contains 2 MB of NetFlow logs, 100 MB of web access logs and 1 MB of web source code. Using this information, we wish our investigator to uncover the following:

- A chronology that shows when the web pages have been viewed and modified/accessed/created
- Determine the malicious webpage and how the system was compromised
- Analyze the scope of the intrusion, and possible spread to other systems

[5] Unlike warrants, subpoenas do not require probable cause and can be issued by prosecutors without judicial approval, as long as they are not unreasonably burdensome. See William J. Stuntz, Commentary, *O.J. Simpson, Bill Clinton, and the Transsubstantive Fourth Amendment*, 114 HARV. L. REV. 842, 857-58 (2001).

[6] See examples in NIJ's *Investigations Involving the Internet and Computer Networks*, http://www.ncjrs.gov/pdffiles1/nij/210798.pdf

- Identify the origin of the malicious activity.

Comparing the original website files created by Buzz Coffee (assuming they still exist in the cloud) to the data returned from the cloud provider would be a constructive first step. Here the technique employed during collection becomes paramount. If the host operating system was used to retrieve the files, Mallory's rootkit would have hidden the malicious files. If files were acquired by reading the physical disk, bypassing the operating system, the complete collection of files will be accurate. Constructing a timeline is a common practice for forensic examiners, and one important in determining when Mallory's files were created. Unfortunately, the procedure employed by the provider again determines whether the investigator receives useful metadata, such as file creation timestamps.

Web access logs are likely the most definitive evidence of the original intrusion, corroborated by NetFlow records. The suspected attacker IP is identified in the logs, which is presented alongside the complete analysis in the subsequent forensic report. Prudent readers might also approach this problem by analyzing the malware installed after visiting the now-hacked webpage, and trying to determine who wrote it or to where it beacons back, but that is not considered here.

Taken to court, the following are questions that could be raised by the defense to discredit the forensic process used in this case:

- Was the chain of custody preserved throughout the process?

- Can the malicious page be definitively attributed to Mallory? Who else had access to create/modify this page? Were other clients hosted on the same infrastructure that could have had access?

- What process did the cloud provider use to copy and produce the webpages? Can they make any claims about the forensic integrity of this process? Are timestamps across the different evidence (NetFlow, web logs, etc.) synchronized enough to create an accurate timeline?

- What was the physical location of the virtual machine that is run by the hosting website? By what laws/regulations is it governed?
- What detection and protection mechanisms are employed by the provider to keep their infrastructure secure and to identify intrusions?
- Since the provider refused to provide operating system evidence, can the prosecution have enough evidence to prove that a compromise actually occurred?

In this case the closed nature of the provider was the primary hindrance to a routine investigation. The provider has an incentive to keep as much of its infrastructure private as possible, since it may give them a competitive advantage. Unfortunately, this decision hinders the investigative process and may discredit the legal proceedings that follow.

3.3 Analysis

Whether in the cloud or not, forensic investigation can be an intensive process. Exams are almost always limited by time and budget, since clients are unwilling or unable to support them indefinitely. Cloud computing, for better or worse, gives customers an ability to terminate virtual machines or revert them to a saved state almost instantaneously. Providers and investigators may also benefit from easy data duplication, system copying/imaging, and extensive business logging. Investigators must recognize the extreme fragility of the evidence. These attributes are indeed positive and contribute towards well-rounded security preparation for incident response. The hindrances seen in these case studies illustrate areas for continued research and development. Consider how we might address the five issues presented at the beginning of Section 3.2.

First, in our case studies, acquisition was accomplished using legal vehicles of subpoena and search warrant. While somewhat cumbersome given the complex legal system, if a forensic investigation is to support a potential criminal proceeding, this approach is necessary. More efficient mechanisms for the secure transfer of data from providers and law enforcement would be ideal.

Second, cloud consumers will need to negotiate or lobby providers for an appropriate level of cooperation and transparency about how their infrastructure works, the amount of support available during incident response, and forensically-sound practices for assisting law enforcement. One potential approach is a forensic service level agreement (SLA) appended to the existing SLA signed by providers and subscribers. This legal backing would give customers assurance about the support available to them from their provider during an investigation, a quantitative measure by which to compare providers.

Third, it is clear that remote forensic tools applied to cloud computing are prone to scrutiny, and local processing tools of cloud-stored data are not designed to handle the format or scope of the data. In the case of IaaS, analysis will certainly include the investigation of a virtual machine. Forensic analysts need a tool for parsing, searching and extracting information from virtual machine snapshots, including suspended memory state.

Fourth, the lack of forensic metadata may be addressed in several ways. One proposal is to introduce data provenance in order to track the history and access of cloud objects. In 2007, a report from the Department of Justice recommended asking "what is the chronology of the access to or changes in the data?" of persons providing digital evidence [58]. Another proposal is to introduce preemptive forensics in the cloud, the forensically-sound logging of information at all times without evidence of a crime in order to specifically support forensic investigations after a crime takes place. For example, keeping regular virtual machine snapshots would create a forensic record back in time once an event arises. When created as a standard business record, this computer-generated evidence can be protected against hearsay arguments, a viewpoint now recognized by most courts.

Finally, chain of custody remains complex given the number of people that may have access to the evidence, and the third-party collection as discussed above. In traditional digital forensics, a chain of custody exists for both physical evidence (*e.g.*, the computer) and its associated data. In the cloud case, data are the only evidence. As such, pristine copies of the data, and associated integrity information such as MD5 checksums, must be carefully handled. Since chain of custody is the legal equivalent of secure provenance, transfers of custodianship could be documented by a digital provenance system.

Note that we have not addressed the issue of responsibility and fault in either case study. In Case Study 1, we have not established what liability the cloud provider has for hosting the illegal content. In all likelihood, the cloud provider demonstrated no negligence, and is simply a data custodian unaware of the activity. Nonetheless, the law demands they identify and remove all illegal content. In Case Study 2, can users who were infected sue the legitimate company or the cloud provider for negligence? Could Buzz coffee sue the hosting provider if they failed to secure their infrastructure, or to notice the intrusion? These questions may be answerable using an interpretation of current laws. Additionally, we have not explored the investigative complexity of cloud service resellers who themselves offer services that utilize cloud technology. The layering of providers may further complicate the preservation and acquisition of evidence.

Finally, both case studies assume trust in the provider, its employees and infrastructure. Providers have their business reputation and customer base to lose if trust is lost in their ability to provide secure and reliable service. However, if an adversary or corrupt insider gains control over the cloud infrastructure—particularly the hypervisor—no data or computational results in the hosted virtual machines can be trusted.

3.4 Conclusions

Cloud security is a much discussed topic, but planning about incident response and forensics needs to happen in parallel. The move of data and services to the cloud is already underway, and research and development in the forensic research community must keep pace. These two case studies illustrate larger issues that exist beyond the scope of our specific examples. Forensic acquisition is a renewed challenge, one unsuited for today's tools, which will possibly be addressed by a combination of technological and legal approaches. We evaluate the ability of popular forensic tools to obtain evidence from a cloud environment in Chapter 4. Cooperation with providers will empower consumers to understand their risks and give them leverage to prosecute crimes. The preservation and availability of forensically-relevant metadata remains an open problem.

We have highlighted the issues of common crimes that vary from today only in their use of the cloud. This technology alone introduces peculiarities and open problems that demand immediate attention. As we have shown, deficiencies in both law and technology can be addressed with proper advances.

Chapter 4

Acquiring Forensic Evidence from Infrastructure-as-a-Service Cloud Computing

Discovery and acquisition of evidence in remote, elastic, provider-controlled cloud computing platforms differ from that in traditional digital forensics, and examiners lack appropriate tools for these tasks. While there are many important issues in this new field, we focus explicitly on data acquisition. Crimes that target or use cloud computing will undoubtedly emerge in this landscape, and investigators will rely on their existing expertise in tools such as Guidance EnCase or AccessData FTK unless alternative tools and techniques are provided.

As we found in Chapter 3, digital forensics for cloud computing brings new technical and legal challenges. Cloud computing makes forensics different, particularly given the remote nature of the evidence, lack of physical access, and trust required in the integrity and authenticity. While the goals of the forensic examiner are the same as before, the non-conventional difficult problems include forensically sound acquisition of remote data, large data volumes, distributed and elastic data, chain of custody, and data ownership.

Seizure and acquisition of digital artifacts are the initial steps in the forensic process [10]. Two possible scenarios exist: remote investigators could collect forensic evidence themselves from the source, or providers could deliver it. Each scenario requires a different degree of trust in the data returned. Further, each scenario uses different technical implementations to recover the data. Given years of development, acceptance by the judicial system, and expertise in the field, market leaders in the commercial forensic tool space including EnCase and FTK are perhaps the companies best prepositioned for the cloud forensic challenge [76]. One question that remained until now, however, was an evaluation of the ability of such tools to acquire and analyze cloud-based evidence.

In this chapter, we assume that the target system of the forensic investigation still exists in the cloud. The elastic nature of cloud computing makes it possible for a criminal to commit a crime and then immediately destroy the evidence, but that situation is not considered here. While some cases will involve the cloud as the instrument of the crime, others will involve the cloud-hosted service as the target of the crime. The later is the scope of this chapter.

In draft guidance [24, p. 21] on the secure use of cloud computing, the Federal Chief Information Officers Council states that "incident response and computer forensics in a cloud environment require fundamentally different tools, techniques, and training." In this chapter, we evaluate the validity of that statement with respect to data acquisition. Contributions of our work include:

- Results from three experiments that exercise existing tools for persistent and non-persistent data collection in a public cloud, Amazon's *Elastic Compute Cloud* (EC2).

- Analysis of alternatives for forensic acquisition at lower levels of the infrastructure stack, for cases when there is insufficient trust in data acquisition using the guest operating system.
- A demonstration of how to use virtual machine introspection to inject a remote forensic agent for remote acquisition.
- Exploration of four strategies for forensic data acquisition with an untrusted hypervisor.

The rest of the chapter is organized as follows. Section 4.1 reviews previous and related work. Section 4.2 explores the forensic examination of cloud-based crimes, including a model of cloud trust. Section 4.3 presents our experiments in using the native capabilities of EnCase, FTK, Fastdump, and Memoryze for data acquisition in EC2. Section 4.4 discusses considerations of the experiments. Section 4.5 suggests alternative approaches. Section 4.6 concludes the chapter.

4.1 Previous and Related Work

In Chapter 3, our analysis of two hypothetical case studies illustrated the non-trivial issues with collecting evidence from a cloud crime. We also discovered that we needed to evaluate existing remote forensic tools.

EnCase Enterprise and FTK include a client-server feature for remote forensics. In each case, the investigator installs a small executable on the client machine (EnCase calls the executable a "servlet"; FTK calls it an "agent"). Figure 4.1 illustrates how the server, built into or on top of the vendors' forensic analysis software, communicates with the client over a secure connection, and can command the client to return forensic data including a hard

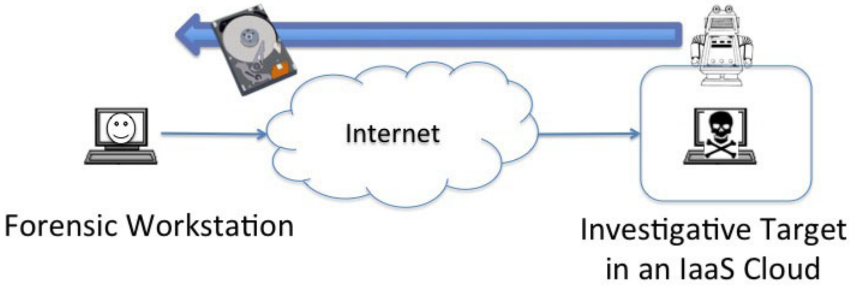


Figure 4.1: General technique for remotely acquiring forensic evidence over the Internet, where a trusted forensic workstation controls a remote agent on the cloud-based target to acquire a disk image.

drive image. The forensic examiner may conduct some forensics remotely on the client, or return to the server for local analysis. Large enterprises employ remote forensics where machines may be geographically disperse, but the incident response team centralized.

4.2 The Cloud Forensic Examination

In this section we explore the forensic examination of a cloud-based crime. As a foundation, we first present a model to reason about the trustworthiness of evidence from the cloud, since the level of trust influences the choices for how an exam should be conducted. Second, we pose choices that determine how to approach a forensic investigation.

4.2.1 Layers of Trust

Before evaluating tools for acquisition, it is important to understand trust in the cloud environment. When brought to court, the judge or jury must ultimately decide if they believe and trust the evidence. This choice embodies a specific confidence about whether the result is accurate and reliable. In traditional forensics, where the target machine is physically present, some of the same trust issues exist, as we shall explain.

Consider an example where a criminal used a single desktop computer to plan a murder. If law enforcement removes the hard drive for imaging, they must trust their hard drive hardware to read the disk correctly. If they run forensic tools on the live computer, they may have to trust the integrity of the host operating system in addition to the hardware. If the suspect computer was hosted in the cloud, new layers of possible uncertainty are inherently introduced. We do not consider the trust in the forensic acquisition tools themselves nor in the human agents executing those tools, since these components, while important, are outside the cloud environment.

Cloud Layer	Acquisition Method	Trust Required
Guest application/data	Depends on data	Guest operating system (OS), hypervisor, host OS, hardware, network
Guest OS	Remote forensic software	Guest OS, hypervisor, host OS, hardware, network
Virtualization	Introspection	Hypervisor, host OS, hardware, network
Host OS	Access virtual disk	Host OS, hardware, network
Hardware	Access physical disk	Hardware
Network	Packet capture	Network

Table 4.1: Six layers of the IaaS cloud environment and potential forensic acquisition techniques for each, including the cumulative trust required by the investigator and courts at each layer.

Table 4.1 models trust in IaaS cloud computing in six layers. The other cloud models, Platform-as-a-Service and Software-as-a-Service, would have additional layers on top to account for the platform or service provided. In IaaS, the consumer retains administrative control over the Guest OS Layer and Guest Application Layer, despite no physical access. Furthermore, the forensic acquisition activity would be different at each layer. Each layer requires a different amount of confidence that the layer is secure and trustworthy; the further down the stack, the less cumulative trust is required. In public clouds, all layers require some trust in the provider, especially trust against malicious insiders. Ultimately, it is the judge or jury that must have confidence in the data to render a legal decision. This model also assumes no hardware root of trust which could mitigate some trust issues. We explore this possibility in Section 4.5.1.

Imagine a situation where a forensic investigator has remote access to the guest virtual machine operating system. The investigator could collect evidence contained inside the VM, install a forensic tool and obtain live evidence remotely, or suspend/terminate the VM and analyze it offline. Unfortunately, acquisition at this layer requires trust that the guest operating system, hypervisor, host operating system, underlying hardware, and network produce complete and accurate evidence data, and are free from intentional and accidental tampering, compromise, or error.

As a risk mitigation strategy, the forensic examiner should examine evidence at multiple layers. This technique allows an investigator to check for consistency between the layers. Arranging individual contexts into groups is a basic concept from archaeology, known as straigraphic interpretation [36]. We recommend a new area of research to identify suspicious activity at different layers of the cloud, and in corroborating forensic hypotheses across layers.

Investigators may be tempted to conduct their investigation remotely on a running machine particularly given the size of the remote data, the time and cost to retrieve a full drive image, and the propensity to conduct live forensics. These are valid goals, and ones we will return to in Section 4.3.

Currently, law enforcement asks the provider for data. Law enforcement issues a search warrant or subpoena to the provider, and the provider executes the search, collects the data, and returns it to law enforcement. Though this process frees law enforcement from needing remote acquisition technology and from the burden of understanding details of the cloud environment, it does not free them from significant trust in the result nor from needing to process the data. Instead, the examiner and jury must now trust the integrity of the technician at the provider to execute the search in a trustworthy manner, the technician's hardware and software used to collect the data, and the cloud infrastructure (at least network and hardware) to retrieve, reassemble, and report the data.

4.2.2 Choices in Cloud Forensics

We now consider how to conduct a forensic exam of IaaS cloud computing by considering the following issues. The layers explored in Section 4.2.1 are also choices of where to conduct a forensic investigation. In particular, the investigator can choose at what layer of the cloud the forensic process will be executed. Considerations for this decision revolve first around the technical capability to conduct forensics at that level, and second the trust in the data returned. The layer also influences what type of forensic data are available for collection, such as packet captures at the Network Layer, physical files at the Hardware Layer, or virtual files at the Host OS Layer. For each data type the data must adhere to strict chain of custody and must include a mechanism for integrity checking.

One must choose who will conduct the exam and where will it be conducted. Possible choices for who will execute the exam include law enforcement, an employee of the cloud provider, or an independent examiner. Choices for where the exam will take place include at the provider's corporate headquarters, at one of the provider's remote data centers, at a remote law enforcement facility, or at an independent third party facility. These choices are as much about practicality and logistics as about law and the examiner's qualifications. Requiring a non-employee of the provider to conduct an exam on provider premises would impose an unacceptable logistic burden to the provider.

Cost is another choice affecting how an investigator conducts an exam. When forensic data are requested, the cost in dollars and labor to preserve and produce records might be passed on to the requester, or sold as a service by the cloud vendor.

Technical choices of how to conduct a forensic exam of cloud computing are numerous but closely mimic the choices in a traditional exam. First, the specific crime dictates whether the forensic process will be conducted on a live or dead machine. Second, regardless of whether the forensic data come from a workstation or the cloud, the forensic goal of determining what happened is the same, except that the volume and format of data may differ. The examiner's choice of analysis tools may be influenced by the format of data collected (*e.g.*, traditional files vs. cloud "blobs"), volume of data, and data type (*e.g.*, netflow logs, billing records, drive images).

Cloud computing introduces one powerful new option: virtual machine snapshots. With many cloud implementations that utilize virtualization it is possible to take a snapshot of a running machine and later restore and run the snapshot offline as if it were live. This offers the ability to create a historical record, as well as do "live" forensics after the fact.

4.3 Cloud Forensics Using Today's Tools

In this section, we measure and evaluate the ability of EnCase Enterprise and AccessData FTK to remotely acquire forensic evidence from cloud computing and measure their effectiveness. Both products are widely deployed today, benefit from tool expertise in the field, are trusted by the courts, and have a remote acquisition feature that has been targeted at geographically dispersed corporate LANs. Our goal is to evaluate the ability of these features to acquire forensic data from cloud computing environments over the Internet. We also test live forensic acquisition tools using Fastdump from HBGary, Memoryze from Mandiant, and FTK Imager from AccessData. These experiments evaluate the success at gathering evidence, the time to do so, and the trust required.

4.3.1 Motivation

Experimentation and testing of today's most popular forensic tools have not previously been applied to cloud computing. We propose three experiments using the IaaS cloud model, since that gives the examiner the most access and control of all cloud models. In particular, we use a public cloud, EC2 from Amazon Web Services (AWS), as a live test bed. *Experiment 1* collects forensic data from the Guest OS Layer. *Experiment 2* collects data from the Virtualization Layer. *Experiment 3* collects data from the Host OS Layer.

The goal of these experiments is to evaluate the ability of five tools to acquire forensic data from cloud computing environments over the Internet. Consider how an investigator might approach his or her first case involving cloud computing. The investigator would likely pick the most popular volatile and non-volatile forensic software acquisition tools and seek to use them in the cloud environment. The first tools we chose were Guidance

EnCase and AccessData FTK. We also chose three memory acquisition tools—Fastdump, Memoryze, and FTK Imager—to determine their success in Amazon's cloud environment.

4.3.2 Extracting Data From Amazon EC2

Extracting data from Amazon EC2 requires more steps than imaging a physical computer or acquiring data from a remote corporate desktop. Here we explain what we learned and ultimately used to acquire forensic data.

One choice for acquiring remote, persistent storage is to download a copy of the volume, or a snapshot thereof. Amazon stores virtual hard drives, called Elastic Block Storage (EBS) volumes, in its Simple Storage Service (S3), but they are not exposed to the end user for downloading.

Two options exist to obtain the data for an entire volume. The first is to create a snapshot of the drive being investigated, create a volume from that snapshot, attach the new volume read-only to a trusted Linux instance in EC2, and then create a raw disk image of the volume that could be downloaded. The second is to detach the target volume from the host under investigation, attach it to a trusted Linux instance in EC2, and use a low-level copying utility (*e.g.*, the Unix data duplication tool *dd*) to create a block copy which can be stored in S3 and downloaded.

Amazon provides a service to export data from S3 onto a physical device and ship it to the requester, but the customer must provide the storage device and is billed $80 per storage device handled plus $2.49 per data-loading-hour [6].

In neither of these cases is it is possible to verify the integrity of the forensic disk image. Amazon does not provide checksums of EBS volumes from either the management website or through the API, so one cannot positively assert that the image retrieved is identical to the original. Further, no hardware write blocker can be used to protect the integrity of the exhibit. However, it is possible to guarantee that the data have not been modified in transit (*e.g.*, hashing the image before export and again after it has arrived from shipping). We will also demonstrate in Section 4.3.4 a case where Amazon provides a checksum.

4.3.3 Methods

For each experiment, we used a non-cloud based standalone machine as a control to evaluate the success of the test. The control was a Dell workstation with 32-bit Windows 2008 R2, a single 30GB disk drive and 2GB RAM. We connected the machine to the Internet and installed the Apache web server. We created several web pages with identifying names and content. Some files were deleted. We artificially compromised the machine using a web-based vulnerability, and assumed that a criminal and forensic investigation had commenced. We imaged the drive with EnCase and FTK.

Experiment 1 tested the advertised ability of popular tools to collect forensic data remotely in the cloud at the Guest OS Layer. Success or failure would be measured by (a) if the tool was able to collect evidence remotely, and (b) how accurately the data compared to those from a standalone control machine. We prepared a single, Internet-connected (proxied), forensic examiner workstation with 64-bit Windows 7 Enterprise. We installed EnCase Enterprise 6.11, including the SAFE (Secure Authentication For EnCase), according to the manufacturer's instructions. We also installed FTK 3.2. In Amazon EC2, we provisioned a new virtual machine to simulate the target of an investigation. This machine was an Amazon-provided Windows 2008 R2 32-bit image with a single 30GB disk drive and 1.7GB RAM. We configured the Amazon firewall to allow only Remote Desktop Protocol (RDP) (tcp/3389).

We connected to the target machine using RDP and proceeded to exercise normal behavior of a user configuring a webserver. We downloaded and installed Apache and created several web pages with identifying names and content. Some files were deleted. We again artificially compromised the machine using a web-based vulnerability and assumed that a criminal and forensic investigation had commenced.

EnCase Servlets and FTK Agents are the remote client programs that communicate with their host server controllers. Each can be deployed in a variety of ways. In a corporate environment, the investigator typically deploy agents to Windows machines over the network using Windows file shares. The products also allow manual file delivery (*e.g.*, USB). In our experiment, we transferred the agent to the target virtual machine over RDP and executed it with Administrator privileges. We modified our firewall to allow communication with the agent: the EnCase servlet used tcp/4445 and the FTK agent used our user-defined port of tcp/3399.

We also tested FTK Imager Lite version 2.9.0. We copied the product over the Remote Desktop connection from our examiner's workstation and ran it interactively. FTK Imager Lite does not require installation, and runs self-sufficiently once uncompressed. For this experiment we attached a second 100 GB storage volume onto which we saved an image of the primary volume captured by FTK Imager.

Finally, we ran Fastdump, Memoryze and FTK Imager to acquire images of system memory, resulting in three 1.7GB images.

Experiment 2 tested popular forensic tools at the virtualization layer by injecting an agent into the virtual machine (Virtualization Layer). We again measured success or failure by (a) the ability of the tool to collect evidence, and (b) how accurately the data matched those from a standalone control machine. We prepared an installation of the Eucalyptus cloud platform [22] from the Ubuntu distribution on a Dell workstation. Eucalyptus supports the Xen hypervisor for managing virtual machines, and LibVMI [50] is a library for monitoring guest operating systems in Xen. We used the LibVMI library to write into memory of the guest virtual machine. With this capability, we demonstrated injecting an EnCase Servlet and FTK Agent directly into a running guest. As with *Experiment 1*, we communicated with the agent over the network.

Experiment 3 tested forensic acquisition at the host operating system level by exercising Amazon's Export feature (Host OS Layer). This experiment most closely resembles the process probably used to satisfy subpoenas and search warrants, since the data are exported from Amazon's internal network at a data center. Additionally, AWS maintains a chain of custody for the storage device while it is in its custody. We measured success or failure by (a) the ability of the technique to collect evidence, and (b) the accuracy of the data as compared to those from the standalone control machine. AWS Export involves a service request to Amazon and shipping them a storage device. Unfortunately, it is currently possible only to export data from an S3 bucket and not from an EBS volume. To meet that requirement, we attached the EBS volume from the compromised machine to a Linux VM, and used *dd* to store an image of the volume in an S3 bucket. We requested from AWS an export of this bucket, and shipped a Seagate FreeAgent eSATA external hard drive. Amazon returned the storage device with a copy of the data.

4.3.4 Results

The manual installation of the EnCase Servlet and FTK Agent in *Experiment 1* was successful and we were able to acquire a hard drive and memory image remotely. Analyzing these images in EnCase Forensic and FTK Investigator respectively correctly revealed a timeline of activity, including the installation of Apache and the webpages we created and deleted. The analysis revealed no unusual artifacts of the virtual environment, nor any apparent

Experiment	Tool	Evidence Successfully Collected	Time (hrs)	Trust Required
1	EnCase	✓	12	OS, HV, Host, Hardware, Network
1	FTK	✓	12	OS, HV, Host, Hardware, Network
1	FTK Imager (disk)	✓	12	OS, HV, Host, Hardware, Network
1	Fastdump	✓	2	OS, HV, Host, Hardware, Network
1	Memoryze	✓	2	OS, HV, Host, Hardware, Network
1	FTK Imager (memory)	✓	2	OS, HV, Host, Hardware, Network
1	Volume Block Copy	✓	14	OS (imaging machine), HV, Host, Hardware, Network
2	Agent Injection	✓	1	HV, Host, Hardware, Network
3	AWS Export	✓	120	AWS Technician, Technician's Host, Hardware and Software, AWS Hardware, AWS Software

Table 4.2: Results of three experiments acquiring cloud-based forensic evidence using popular tools, including the time to retrieve the data and trust required in the data.

anomalies to raise doubt about the integrity of the data. The speed of the acquisition process was limited by our learning how to use the remote agents and the network bandwidth to transfer the data. The later took approximately 12 hours each for EnCase and FTK to transfer the 30GB disk image and 2GB memory image using our university's OC-12 connection. We do not suspect a network limitation at Amazon, but the latency is indicative of a 10Mb/s network segment somewhere along the communications path, perhaps inside our university.

Experiment 2 successfully resulted in a complete image of the drive and a correct timeline. VM introspection is a powerful tool for forensics and allows live investigation of a host without revealing the presence of the investigator. However, introspection is a special feature which must be implemented by the cloud service provider. This was the only experiment where we were able to verify cryptographically the integrity of the image, since we had access to the physical disk and could compare hash values of the EnCase image and the original disk.

The AWS Export process in *Experiment 3* also successfully returned a complete image of the drive. We were able to load this drive into EnCase and FTK with no difficulties, and verified the contents of the drive. An added benefit of this method is that AWS generates a log report with metadata for each file exported. This report contained the following for each file: date and time of the transfer, location on the storage device, MD5 checksum, and number of bytes. Amazon saved these data in an S3 bucket that we specified in the export request. Using expediated shipping, it took five days to receive our data, at a cost of $125. We imagine that this process would closely mimic the steps taken by AWS when complying with a search warrant or subpoena.

EnCase and FTK were easiest to use. Despite setup and learning time required to use the remote capabilities, the features of the tools were familiar and easy to execute. The 12-hour time required to retrieve our disk image was significantly shorter than the 120 hours required for the AWS Export process for this data volume. Downloading data achieved an average of 2.5 GB per hour. AWS Export spent 4 hours loading our data, while the remaining 116 hours were spent in transit. At these rates, the most time effective choice is the export process when more than 240 GB of data will be retrieved.

Table 4.2 summarizes the results of data acquisition in EC2. Each tool and technique successfully resulted in evidence production, but each requires substantial trust in the cloud infrastructure at all levels.

4.4 Discussion

The nature of online remote forensics introduces security considerations. For example, a forensic examiner's workstation must have access to the Internet to acquire the evidence. While precautions such as firewalls and proxies may help shield the workstation from attack and compromise, the possibility of infection becomes more likely than if the workstation were standalone or on an isolated network. This risk must be accepted, or remediated with appropriate technology (*e.g.*, monitoring, patching).

One attractive feature of allowing examiners to use existing tools, as in *Experiment 1*, is that no changes to the cloud infrastructure are necessary, and no assistance from the provider is required. Introspection, as in *Experiment 2*, requires considerable change to the environment made by a provider, even though an examiner could exercise that feature without the provider's intervention. Data export, as in *Experiment 3*, requires no change to the infrastructure, but the provider must execute the process.

Our experiments assume that the cloud consumer is the victim of the crime and the plantiff in the investigation. However, an equally likely scenario is one in which a criminal creates a system in the cloud, uses it to commit a crime, and removes the cloud system entirely. This situation demands proactive logging of data by the provider which may be of forensic relevance in the future. Shields, *et al.* [78] created a proof-of-concept continuous forensic evidence collection system that could be used to record the creation and deletion of cloud provisions. Finally, if the cloud provider is the criminal, the forensics service is also suspect and another alternative must be considered to investigate the crime.

A forensic shortcoming, and potential legal problem, is the lack of validation for the disk images. Forensic examiners are accustomed to using cryptographic hashes to validate that the copy of a hard drive that they have taken is identical to the original. With no hash available for the original data source, examiners and jurors could question the integrity of the result and reject the evidence. In our experiments, we were unable to verify cryptographically that our cloud images were identical to the standalone control because of differences such as different hardware (thus drivers) and network configurations. These differences did not affect the ability to reconstruct the crime.

The EnCase Servlet and FTK Agent used for our experiments had some limitations. These programs typically have System privileges, giving them unfettered access to memory and disks. However, as with all software, they are vulnerable to malicious code that may have already compromised the target machine. The agent could be installed at any time in the lifecycle of the virtual machine; installing at the time the VM is provisioned prevents the disruptive installation after an incident has taken place. Cloud providers such as Amazon employ user-configurable firewalls that must also be opened to allow the agents to communicate with the command and control node. Though not inherently a vulnerability, open ports do increase the attack surface. To mitigate this potential vulnerability, we opened the firewall only for the time necessary for imaging. We also configured the firewall to only allow traffic from the IP address of our imaging machine. Fortunately, EnCase and FTK also employ network encryption between the client and server to provide confidentiality and authentication.

Investigators must consider the cost associated with a remote forensic analysis. Imaging and retrieving a virtual hard drive and its associated memory will incur potentially significant bandwidth costs. Our experiment used an instance with a 30 GB virtual disk and 1.7 GB memory. Amazon currently bills outbound data transfer at $0.150 per GB, for the first 10 TB / month. Therefore, the retrieval of the disk and memory images totaled only $3.60. One TB of data would cost $150. Data transfer costs could be eliminated if analysis were done with another EC2 instance.

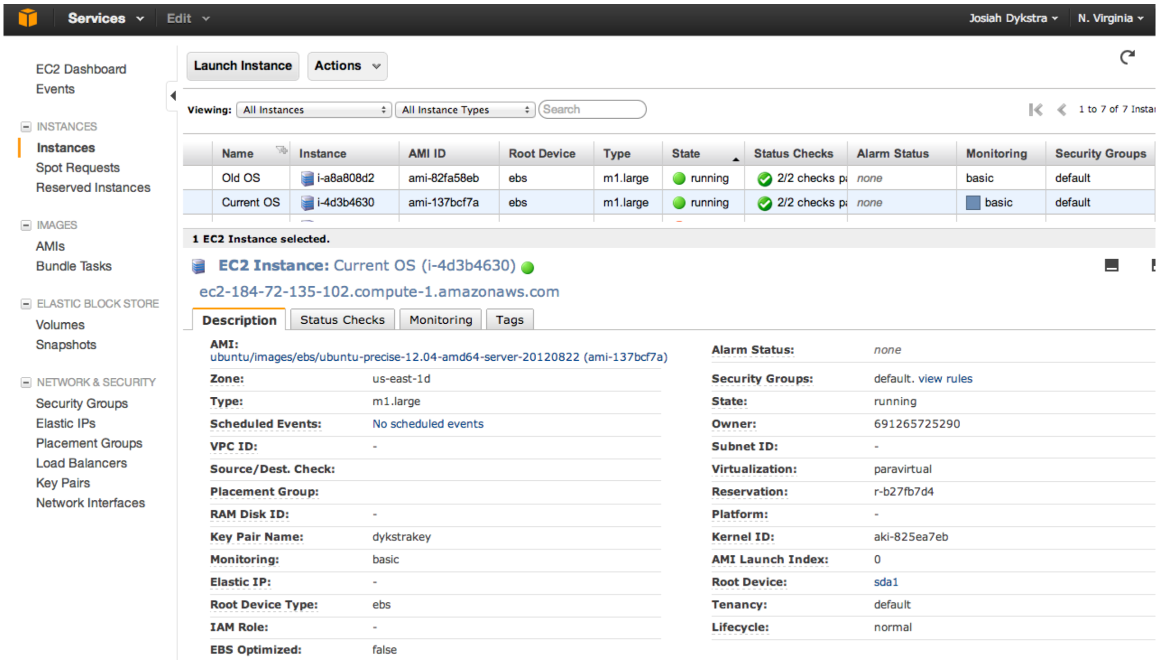
4.5 Alternatives for Forensic Acquisition

In this section we briefly propose four alternate solutions to acquiring cloud-based data: Trusted Platform Modules (TPMs), the cloud management plane, forensics-as-a-service, and contract support. The adoption of one or more of these alternatives would make remote acquisition more trustworthy than acquisition using EnCase or FTK since trust is rooted at lower cloud layers.

4.5.1 Rooting Trust with TPMs

The deployment of TPMs would root trust in cloud computing hardware. Several researchers have previously suggested this approach [44, 45, 71, 72]. A TPM can provide one or more capabilities: machine authentication, hardware encryption, signing, secure key storage, and attestation. Previous solutions for TPMs in cloud computing focus on provisioning trusted guest VMs rather than on attestation of the host platform. If TPMs were installed in each cloud server, the hardware and associated software could validate what software is installed on each machine and verify the health and status of each machine. Despite this benefit and low cost, TPMs have limitations of their own and are not perfectly secure.

While appropriate for future consideration, we believe the primary hindrance to this approach today is that cloud vendors have large amounts of heterogeneous, commercial hardware which is replaced as needed rather than all at once, much of which does not have a TPM. While future hardware may include a TPM, the provider cannot guarantee that each server in its cloud has one today. Future CPUs may even include integrated TPMs. Nevertheless, customer demand today or in the future may drive providers to introduce trusted hardware for some or all customers. Providers could also chose to provide TPMs at an additional fee.



Name	Instance	AMI ID	Root Device	Type	State	Status Checks	Alarm Status	Monitoring	Security Groups
Old OS	i-a8a808d2	ami-82fa58eb	ebs	m1.large	● running	✓ 2/2 checks p	none	basic	default
Current OS	i-4d3b4630	ami-137bcf7a	ebs	m1.large	● running	✓ 2/2 checks p	none	basic	default

1 EC2 Instance selected.

EC2 Instance: Current OS (i-4d3b4630) ●

ec2-184-72-135-102.compute-1.amazonaws.com

AMI:
ubuntu/images/ebs/ubuntu-precise-12.04-amd64-server-20120822 (ami-137bcf7a)

Zone:	us-east-1d	Alarm Status:	none
Type:	m1.large	Security Groups:	default. view rules
Scheduled Events:	No scheduled events	State:	running
VPC ID:	-	Owner:	691265725290
Source/Dest. Check:		Subnet ID:	-
Placement Group:		Virtualization:	paravirtual
RAM Disk ID:	-	Reservation:	r-b27fb7d4
Key Pair Name:	dykstrakey	Platform:	-
Monitoring:	basic	Kernel ID:	aki-825ea7eb
Elastic IP:	-	AMI Launch Index:	0
Root Device Type:	ebs	Root Device:	sda1
IAM Role:	-	Tenancy:	default
EBS Optimized:	false	Lifecycle:	normal

Figure 4.2: A screenshot of the AWS Management Console for EC2.

4.5.2 Collection from the Management Plane

Cloud computing has a unique attribute that could be used to support trustworthy forensics: consumers manage and control virtual assets via a management plane, an out-of-band channel that interfaces with the cloud infrastructure. In Amazon Web Services, this system is called the AWS Management Console. This web-facing system interfaces with the provider's underlying filesystem and hypervisor, and is used to provision, start and stop virtual machines, and manipulate the firewall (Figure 4.2).

The management plane is particularly attractive because it is user driven. The provider, end users, and law enforcement could download log files, disk images, and packet captures from the management plane on demand. Further, with forensic acquisition occurring under the hypervisor, retrieving VM images and other data would require trust only in the Virtualization Layer and below.

While attractive, this solution does require trust in the management plane, a potential vulnerability that differs from trust assumptions with non-virtualized, physical computers. As a web-facing interface, the management plane opens a new attack surface which must be protected by the provider. Access to the management plane should be logged and strictly enforced with identity and access management. Communication between the user and the management plane endpoint should be done securely (*e.g.*, using SSL).

4.5.3 Forensic Support as a Service

Provider support for forensic acquisition is a natural choice. While this type of support is not self-service, the provider is already pre-positioned to preserve and collect the data since they control the infrastructure, not only from a virtual machine, but also from infrastructure logging mechanisms, packet captures and billing records. Technology for remote acquisition would be moot if the provider and its infrastructure were trusted and the provider was willing and able to provide evidence to the investigator directly. At their choosing, providers could offer these services to their clients with little effort and cost. Voluntarily doing so would demonstrate their care for security, and put reluctant security-minded clients at ease knowing that investigation was possible. At least one provider, Terremark, offers forensic-as-a-service [81]. Potential drawbacks to a forensic support service include response time (potentially mitigated by the Service Level Agreement) and the provider's lack of knowledge about how customers are using the cloud to meet their goals.

Consider the following protocol for trust-preserving, provider-assisted evidence production. Law enforcement serves a cloud provider with a search warrant for data related to a particular IP address, including the client records for the user of that IP and the virtual machine serving content. A technician at the provider, certified as a forensic examiner by an independent third party, sits down at a forensic workstation connected to the back-end cloud infrastructure. The provider executes the warrant and gathers the data requested, validating the data with cryptographic checksums. Among the data requested are historical snapshots of virtual machines, access logs from the Management Console, data provenance logs, netflow records for the requested IP, and firewall logs. The data are copied to media for law enforcement. This protocol works at the Virtualization Layer, which requires trust in the host operating system, hardware, network, and the technician in this case. Though the protocol still requires trust in the hardware (which could be mitigated by using a TPM), it provides basic assurances that the operating system, network, and technician are trustworthy.

4.5.4 Legal Solutions

Laws could require investigative support from a cloud provider. Contrary to forensics-as-a-service, this support would be legally mandated and might take the form of entitlements to law enforcement for monitoring and surveillance of suspected criminal activity.

No provider has publicly advertised the options for forensic collection available to law enforcement. It is unknown whether the Communications Assistance for Law Enforcement Act (CALEA) [82], a federal law that codifies how telecommunication carriers must support law enforcement in wiretaps, or others like it might apply to cloud computing. CALEA demands certain technical interfaces on the part of the provider to facilitate this collection. Such capabilities are necessary if the courts decide that CALEA, or similar legislation, applies to cloud providers. Even if wiretaps are a sufficient legal instrument for collecting data, the technical implementation must make such collection easy.

In Chapter 6 we analyze the unique legal problems raised by the application of current law to cloud computing, particularly for search and seizure of data from cloud providers. These issues are intertwined with the technical ability to acquire data, and range from whose law governs cloud data to who can legally execute the warrant. An exemplar search warrant for cloud evidence, described in Chapter 7, gives law enforcement a starting point to request the relevant data from a provider.

4.6 Conclusion

We have demonstrated that today's most widely used forensic tools are technically capable of remote acquisition of data in Amazon EC2. We have also shown that given the many layers of trust required, technology alone is insufficient to produce trustworthy data and solve the cloud forensic acquisition problem. The four alternatives we presented offer options that bridge technology and provider support.

Our recommendation for forensic acquisition of IaaS cloud computing is the management plane. This option potentially offers an attractive balance of speed and control with trust. We encourage cloud providers to make forensic data available to users in this way as we show with our own implementation in Chapter 5. While EnCase and FTK successfully returned evidence, we do not recommend using them for remote forensics in the cloud because too much trust is required.

Several areas remain for future work. First, our experiments are specific to IaaS using EC2. These results do not carry to other cloud models and environments, such as Microsoft Azure or Google AppEngine, where forensic software cannot be installed and run as they can in EC2. Future work will be required to find suitable parallels on those platforms. Second, as we show in Chapter 5, cloud users would benefit from consumer-driven forensic capabilities exposed to them by the provider. We intend to work with providers to allow clients to retrieve forensic logs and metadata (*e.g.*, cryptographic checksums of disk volumes) directly from the online management console. Third, investigators need solution to preserve evidence and prevent the loss of forensic evidence when cloud resources are released. Finally, in

Chapter 6 we will explore legal questions of acquisition, particularly those arising from Fourth Amendment concerns about search and seizure, jurisdiction, and ownership in future work.

Cloud computing is gaining momentum and where the people, the data, and the money go, so does crime. Our work lays a foundation and path to enable forensic examiners to take the initial steps in the forensic investigation of cloud-based crimes.

Chapter 5

Design and Implementation of FROST: Digital Forensic Tools for the OpenStack Cloud Computing Platform

Today, cloud computing environments lack trustworthy capabilities for the cloud customer or forensic investigator to perform incident response and forensic investigation. Consequently, customers of public cloud services are at the mercy of their cloud provider to assist in an investigation. Law enforcement relies on the cumbersome and time-consuming search warrant process to obtain cloud data, and requires the cloud provider to execute each search on behalf of the requester. In Chapter 4 we concluded that the management plane is an attractive solution for user-driven forensic capabilities since it provides access to forensic data without needing to trust the guest VM or the hypervisor, and without needing assistance from the cloud provider. Storing and acquiring trustworthy evidence from a third party provider is non-trivial. This chapter describes and evaluates our implementation of this solution in a laboratory instantiation of the OpenStack cloud platform, which we call Forensic OpenStack Tools (FROST).

FROST provides forensic capabilities built directly into OpenStack. We adopt the NIST definition of cloud computing as a model for on-demand access to a pool of resources “that can be rapidly provisioned and released with minimal management effort or service provider interaction” [63]. Our forensic extensions allow for efficient, trustworthy, and user-driven incident response and forensic acquisition in a cloud environment.

This work implements practical tools on the theoretical foundations that we established in Chapter 3.2. FROST collects data at the cloud provider, at the host operating system level underneath the guest virtual machines, and makes that data available within the *management plane*. The management plane, exposed through a website and application programming interface (API), is how users of OpenStack control the cloud, and where they start and stop virtual machines. Because the user collecting forensic data does not communicate with a virtual machine, forensic data are preserved against a compromised or untrustworthy virtual machine. Consider an arbitrary cloud customer Alice who wants to investigate suspiciously high bandwidth usage from her cloud-hosted webserver. Aside from the logging of web requests that she does inside of her own VM, Alice would have a more complete picture of activity if she could also get a record of management activity and metadata about her VMs. Our solution collects and provides trustworthy API logs, guest firewall logs, and virtual disks. These data can be used to help construct a timeline of activity and understand an incident.

OpenStack [66] is an open-source cloud computing platform, conceived as a joint project between the National Aeronautics and Space Administration (NASA) and Rackspace. OpenStack users include many large organizations such as Intel, Argonne National Laboratory, AT&T, Rackspace, and Deutsche Telekom. The cloud platform comprises six primary modular components: Nova, the compute platform and cloud controller; Swift, the object storage system; Glance, the service for managing disk images; Keystone, the identity service; Horizon, the web-based dashboard for managing OpenStack services; Quantum, network services for virtual devices. OpenStack is a non-trivial software package, with over 600,000 lines of code and 415 active developers [65]. It is a widely used platform for private cloud instances, but it is also compatible with commercial cloud offerings. OpenStack has APIs compatible with Amazon EC2 and S3.

Without loss of generality, our approach makes the following assumptions. First, the user-driven forensic capabilities are applicable in situations where a cooperative cloud customer is involved in the investigation. That is, if a malicious customer uses the cloud to commit a crime, the cloud provider will still be required to assist law enforcement in the investigation. Second, the proposed solution assumes a trusted cloud provider and cloud infrastructure. Evidence from our forensic tools could be manipulated unless the underlying layers of the cloud infrastructure, such as the host operating system and hardware, have integrity. We assume that the hardware, host operating system, hypervisor, and cloud employees are trusted, but we do not assume trust in the guest machine. Third, we do not consider legal issues associated with the process or product of cloud-based forensic data acquisition; Chapter 6 explores those issues.

Our contributions are as follows:

- Description of the architecture, design goals, and implementation of user-driven forensic acquisition of virtual disks, API logs, and firewall logs from the management plane of OpenStack.
- An algorithm for storing and retrieving log data with integrity in a hash tree that logically segregates the data of each cloud user in his or her own subtree.

- Evaluation results showing that the proposed solution satisfies technological and legal requirements for acceptance in court and scales appropriately for a cloud environment.

The rest of the chapter is organized as follows. Section 5.1 describes the requirements, specifications, and capabilities of FROST. Section 5.2 explains the architecture of our solution, Section 5.3 discusses the design, and Section 5.4 explains our API and management console implementations based on the architecture. Section 5.5 presents a concept of operations, Section 5.6 evaluates the solution, and Section 4.4 discusses advantages, limitations, and trust assumptions. Section 5.7 concludes the chapter.

5.1 Requirements, Specifications, and Capabilities

We describe the requirements, specification, and capabilities for FROST. We identify the stakeholders and use cases that will help determine the tool requirements. We also discuss the accepted legal and forensic community requirements, and how we will meet them.

Cloud-based crimes take two general forms that determine the stakeholders who would use FROST. One form is a crime committed against an innocent cloud-based victim who is cooperative in an investigation. The other is a crime committed by an uncooperative party using the cloud as an instrument of a crime. In the first case, the legitimate cloud customer and/or law enforcement will use FROST. In the second case, law enforcement or the provider will use FROST. In both cases the requirement is to minimize interaction with personnel at the cloud provider. The cloud provider deploys FROST, but has no other responsibilities (subject to the assumptions above).

5.1.1 Scientific, Technical, and Legal Requirements

There is no single, authoritative source for requirements development of new forensic tools. Our solution, however, is informed by accepted practices and written guidance. The Scientific Working Group on Digital Evidence (SWGDE) [75] asserts that “Digital Evidence submitted for examination should be maintained in such a way that the integrity of the data is preserved. The commonly accepted method to achieve this is to use a hashing function.” On the requirements for acquisition the National Institute for Standards and Technology (NIST) [59] says “The two critical measurable attributes of the acquisition process are completeness and accuracy. Completeness measures if the all the data was acquired, and accuracy measures if the data was correctly acquired.” Integrity and completeness of the data will be of foremost importance.

The cloud environment dictates the technical requirements. Any digital forensic tools for cloud computing should be compatible with cloud characteristics of on-demand self-service, rapid elasticity, and scalability. The following technical requirements are consistent with these characteristics:

1. **Be compatible with existing forensic formats.** Instead of creating new data formats, the new capabilities output data in existing formats to be easily ingested by other forensic tools. Our logs and disk images are provided in standard formats, and all are accompanied by a Digital Forensic XML (DFXML) file [28]. DFXML is used to express the cryptographic hashes and provenance information.

2. **Be easy to generate.** It must be easy to modify existing cloud deployments to add forensic capabilities. It must also be intuitive and simple for a user to request forensic data. Our changes to a stock installation of OpenStack can be made by running an

installation script. Users can request forensic data with a single command or web click.

3. **Be open and extensible.** The implementation must be available for any OpenStack administrator. Developers should be able to extend and contribute new forensic capabilities. The platform we developed allows other developers to integrate other forensic tools quickly and easily. The software will be submitted to the OpenStack project.

4. **Be scalable.** The forensic tools must be usable for single cloud instances, while also supporting millions of cloud customers and virtual machines. FROST can support any number of instances and is limited only by the processing time it takes the host operating system to retrieve the forensic data.

5. **Follow existing practices and standards.** Where possible, cloud forensic tools should follow standard forensic practices. The forensic data we provide adheres to accepted practices and can be ingested by standard forensic tools such as Guidance EnCase.

For acceptance in court, the Federal Rules of Evidence 901(b)(0) explain that "To demonstrate authenticity for computer-generated records, or any records generated by a process, the proponent should introduce '[e]vidence describing a process or a system used to produce a result and showing that the process or system produces an accurate result'" [84]. In most cases, the reliability of a computer program can be established by showing that users of the program actually do rely on it on a regular basis, such as in the ordinary course of business. Our solutions use ordinary data, such as firewall logs, even when we have enhanced the storage of data to add increased data security.

5.1.2 Specifications and Capabilities

FROST has three primary components. First, a cloud user can retrieve an image of the virtual disks associated with any of the user's virtual machines, and validate the integrity of those images with cryptographic checksums. Second, a cloud user can retrieve logs of all API requests made to the cloud provider made using his or her credentials, and validate the integrity of those logs. The API is used for administering virtual machines, such as creating and starting VM instances. Third, the cloud user can retrieve the OpenStack firewall logs for any of the user's virtual machines, and validate the integrity of those logs. The OpenStack firewall operates at the host operating system, and the API is used to administer it, such as allowing or blocking network ports. These three components are useful and offer forensic data that are not available directly to cloud users today. In our informal discussions with cloud users and administrators of two large private clouds and forensic experts, they all requested capabilities that were consistent with these features.

Cloud users interact with their provider and manage cloud resources through the management plane using a web interface and API. FROST is accessible from each of those management plane interfaces. The implementation is modular to allow additional forensic capabilities to be added later.

5.2 Architecture

We describe the architecture of our solution. We show how we will integrate with OpenStack, the type and format of the data we will collect, and the methods for returning data to the requestor.

5.2.1 Integration with OpenStack

OpenStack has many components, but we focus on the two where we have integrated FROST: Nova and Horizon. Nova provides the compute service through virtual servers similar to those in Amazon EC2 and implements the compute API. Horizon provides the web-based user interface for OpenStack, and communicates with Nova through the compute API. Figure 5.1 highlights where we modified Nova and Horizon to integrate FROST.

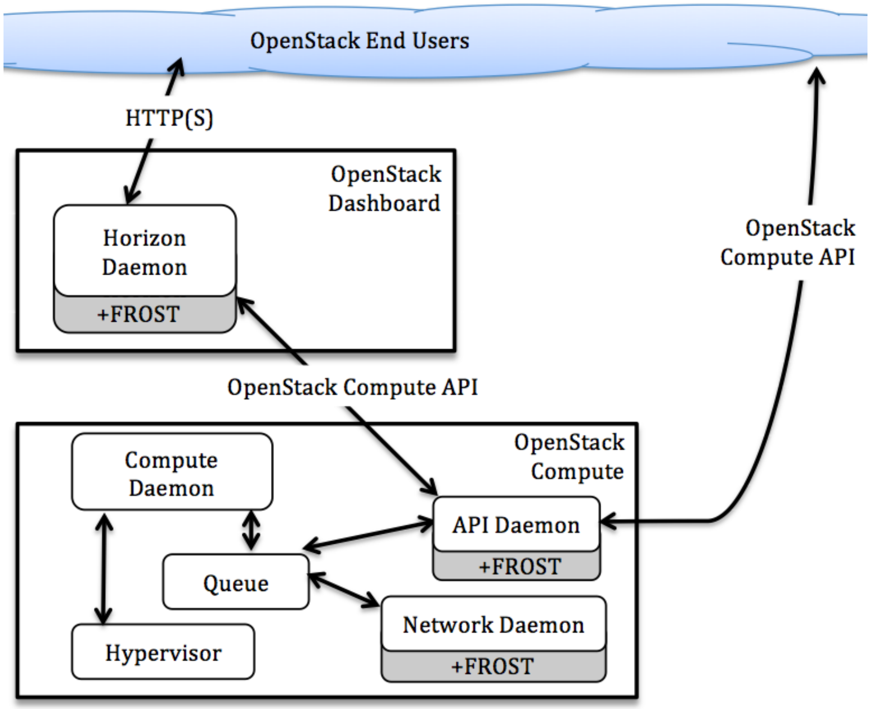


Figure 5.1: Pictorial snippet of the OpenStack architecture showing where OpenStack Compute (Nova) and OpenStack Dashboard (Horizon) have been modified to add FROST. Horizon provides a web interface to the management plane and Nova provides an API interface to the management plane. The majority of changes for FROST were to the API Daemon.

We will add new Nova API calls that correspond to our forensic features. Cloud users who interact with OpenStack using the compute API will be able to exercise our capabilities from command-line tools and in their own programs.

Horizon is built using Django and Python, and implements dashboards for OpenStack. We modify the specification for the dashboard that displays instance information and creates a new tab. This tab will have links to our forensic capabilities. These links will return data from their corresponding API calls.

OpenStack has a variety of credentials for different purposes. Our tools assume that OpenStack has authenticated the user making the request. The Horizon web interface requires only a username and password. The command-line API requires either an access key and secret access key (which can be retrieved using the API), or an X.509 certificate and private key. API requests are digitally signed using the private key, and this signature is transmitted to OpenStack along with the certificate. Nova also has a root certificate that can sign documents. We will use this root certificate to add integrity to the storage of log data, which we call the *Authenticated Logging Service* (ALS).

5.2.2 Data Retrieval

Each of the three FROST capabilities accesses unique data that are already stored by OpenStack or which we can easily enable for storage. Retrieval of data for the user depends on how and where the data are stored.

Retrieval of virtual disks is the most straightforward task. For each virtual machine, OpenStack creates a directory on the host operating system that contains the virtual disk, ramdisk, and other host-specific files. The file format of the virtual disk varies according to the hypervisor used. Since we use KVM as our hypervisor, the format of our virtual disks is QEMU QCOW2 images. The ability to retrieve the original virtual disks must support snapshots of disks from machines that are running, as well as downloads of disk images from stopped machines. QEMU provides utilities to convert QCOW2 images to raw format, and libewf can convert raw images to the EWF-E01 format.

Cloud users may run a firewall inside their VM, but OpenStack provides firewall services beneath the VM. By default OpenStack uses the Linux iptables firewall on the host machine to implement network security for the guest machines. A new chain, or group of rules, is created for each instance. Several default rules are automatically created, such as allowing the host to communicate with the guest. Cloud users are then able to create custom rules manually, such as allowing inbound SSH or HTTP traffic. OpenStack has no inherent configuration options to log network connections that match the firewall rules or connections that are denied by the firewall. However, iptables natively has this ability. We will enable logging on all denied network connections and enable the user to retrieve logs for their OpenStack instances.

OpenStack has the ability to log request successes and failures when a user issues a request to Nova. For example, when a user uses the API to request a new VM, this request can be recorded. These logs are stored on the host operating system, and therefore are typically not available to cloud users. FROST should store these same data, but in a method that allows the data to be segregated for each user and that includes integrity checking information.

5.3 Design

The goals of enhanced API and firewall logging are to enable a cloud user to retrieve and validate the integrity of forensically-relevant log data. The ALS will supplement Nova's default logging capability. This service will store the same data as the traditional log, but a new hash tree will segregate users' data and integrity checking information with minimal overhead for record storage or retrieval. Each OpenStack user account will have his or her own subtree under the root.

When a user provisions a new virtual machine in OpenStack, a universally unique identifier (UUID) is assigned to the machine. These UUIDs become children of the owner's root, and logs for that machine are appended as follows. The subtree of any virtual machine has a depth of four for the year, then month, then day of the log entry, with the log messages as leaves of the tree. Because the tree is constantly changing as new log entries are added, hash values for the intermediate hash tree nodes are re-calculated daily. This structure enables a user to request any date range for any or all virtual machines, while minimizing the additional overhead required.

ALS guarantees integrity of the log data using cryptographic hashes. Integrity checking allows the user to validate if data has been inserted, removed, or modified. For example, if Alice requests her logs for December, she can calculate the hash values that she expects in the tree and compare them to what the provider claimed they should be. If an attacker modified the log data in transit, the integrity check would fail and alert Alice to errors or manipulation.

5.4 Implementation

We provide details about the implementation of FROST and show how users interact with the tools.

We implemented the forensic extensions using DevStack, an OpenStack development environment, on Ubuntu 12.04. We used OpenStack Folsom, which was released September 27, 2012. We used the Xen hypervisor and Ubuntu guests, but our implementation can support any hypervisor and guest operating systems.

5.4.1 Authenticated Logging Service

The Authenticated Logging Service uses Merkle trees [55] as the data structure for storing API and firewall log data. Unlike previous work, we are not concerned with hiding the structural information associated with the tree, nor about prohibiting redaction in exported subtrees.

Hash trees offer three advantages. First, storing summary information about a larger dataset enables efficient validation and minimal data transmission. For any subset of data in the tree, the algorithm hashes chunks of the data, and uses those hashes to compute the hash of the whole tree. It is unnecessary to reveal or transmit the entire tree. Second, given the way we organized the tree, a user can easily query for data over any date range. Third, the hash tree natively enables a user to validate the integrity of a subset of log data.

Our algorithm for storing API and firewall logs is as follows. These two sets of data are stored separately. Since the design is the same for each, we describe only the storage of API logs. As shown in Figure 5.2, the cloud provider maintains a single, append-only hash tree for all users. When a new user joins the cloud service, a subtree is created for the user under the root. The user's tree root is signed using the user's public key. All API logs associated with that user are stored in his or her subtree. Data under the user's root are organized in five layers, corresponding to the machine instance, year, month, and day of the respective log entry. Raw records are found at the leaves, stored as children of the day. The value at each branch node is calculated by concatenating the values of its children and computing the hash of that aggregate. Every minute, the provider computes a hash of the children at each node and updates the value of the node with the new hash. The provider also signs the root of the tree, and the root of each cloud customer, using the Nova root certificate.

When a user wishes to retrieve the logs associated with a particular instance, the cloud provider returns the raw log messages and any hash values necessary to validate the integrity

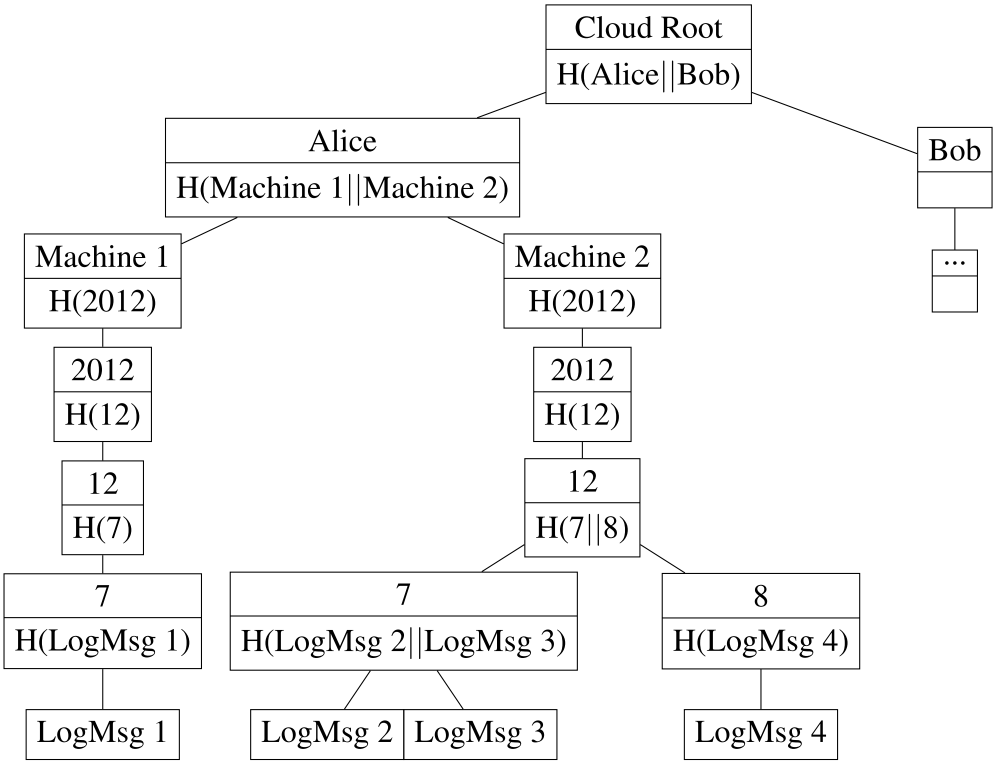


Figure 5.2: Tree structure used to store API logs by user, machine, year, month, and day, showing log entries for Alice's two virtual machines on December 7-8, 2012. The value at each branch node is the hash of the concatenation of its children. Hash values allow integrity validation for a subtree of the whole.

of the result up to the user's root. For example, in the most trivial case shown in Figure 5.2, the provider would only return a single log message and the hash value at node "Alice." Using the Nova root certificate, the provider also hashes and signs all data being returned and records these values in a DFXML log file which is returned to the user. Alice could then compute the hashes and validate that the value she calculated for "Alice" matches what the provider claimed.

5.4.2 API Implementation

Many users interact with cloud platforms with command line tools that call API functions. The Nova API daemon is the endpoint for API queries. Our API extension file contains the code to implement our features. We register these extensions with Nova, and add the ability to call them from the dashboard and the command-line `novaclient`. New API calls are added to OpenStack by placing their functionality in a contribution directory, and modifying `novaclient` to allow the user to call the API. Each of our forensic capabilities was implemented in this manner. We then hook the Nova logging handler to send log messages to our replacement logging service, described below. We also hook the iptables manager to label firewall messages with the instance ID associated with them. The Nova Network daemon then carries out the work of correctly modifying the iptables rules as the system and the user creates them.

To use FROST a user must have already authenticated to OpenStack with his or her private key or credentials. The authenticated user can access only the logs for machines that he or she owns, as enforced by Keystone, the OpenStack identity service. The API validates that the requestor has permission to access the instance for which he or she is requesting forensic data.

Nova logs are stored in /var/log/nova/ on the host operating system. When a user requests his or her Nova logs, FROST searches this file for lines that contain that user's personal identifier.

Listing 5.1 shows the output of using FROST from the command line to retrieve the Nova logs for a single virtual machine. FROST returns the Nova entries that match that UUID, and also creates a DFXML file named report.xml. The DFXML file contains provenance information about the execution of FROST and a hash of the log data for integrity validation.

Listing 5.1: Execution of the FROST API to retrieve the Nova logs for virtual machine 0afcfbcd-b836-4593-a02c-25d8d3a94b00 showing user "admin" provisioning a new virtual machine. These data are available only to users with FROST or with provider assistance.

```
$ nova get-nova-logs 0afcfbcd-b836-4593-a02c-25d8d3a94b00 verify.xml
[truncated]
2012-12-01 13:30:49 INFO nova.api.openstack.wsgi [req-0afcfbcd-b836
   -4593-a02c-25d8d3a94b00 admin demo] POST http://10.34.50.142:8774/
   v2/5ee3040fa890428387f56111576cf819/servers
2012-12-01 13:30:49 DEBUG nova.quota [req-0afcfbcd-b836-4593-a02c-25
   d8d3a94b00 admin demo] Created reservations [’915e9c89-b3bc-4091-8
   b75-3b555961ec3e’, ’72c39d24-0a96-42ca-96f1-593da3aa9f81’,
   ’57843316-872b-4b40-a853-2aa7c730262e’] from (pid=16036) reserve /
   opt/stack/nova/nova/quota.py:697
2012-12-01 13:30:50 DEBUG nova.compute.api [req-0afcfbcd-b836-4593-
   a02c-25d8d3a94b00 admin demo] Going to run 1 instances... from (
   pid=16036) _create_instance /opt/stack/nova/nova/compute/api.py
   :492
[truncated]
```

Firewall logging must be enabled, since it is not enabled by default in OpenStack. Since OpenStack creates default rules for each running virtual machine, we append another rule that logs all dropped packets to /var/log/syslog. For each instance, we prepend a special prefix to the log messages that labels the UUID of the machine. Doing so enables us to parse the log file and identify those lines that correspond to the particular virtual machine that the user requests.

Listing 5.2 shows the output of using FROST from the command line to retrieve the firewall logs for a single virtual machine. FROST returns the firewall logs that match that UUID, and also creates a DFXML file named report.xml.

Listing 5.2: Execution of the FROST API to retrieve the firewall logs of virtual machine 0a18799f-c198-4dbb-b369-b49184e3dfbc showing traffic to ports 443 and 53 being dropped. This level of logging is exposed only to users with FROST or with provider assistance.

```
$ nova get-firewall-logs 0a18799f-c198-4dbb-b369-b49184e3dfbc verify.
    xml
0a18799f-c198-4dbb-b369-b49184e3dfbc: Nov 28 11:13:38 domU
    -12-31-39-17-29-5D kernel: [  310.765760] IPTables-Dropped: IN=
    eth0 OUT= MAC=12:31:39:17:29:5d:fe:ff:ff:ff:ff:ff:08:00 SRC
    =130.85.36.72 DST=10.97.42.171 LEN=52 TOS=0x00 PREC=0x00 TTL=48 ID
    =29222 DF PROTO=TCP SPT=55739 DPT=443 WINDOW=1002 RES=0x00 ACK
    URGP=0
0a18799f-c198-4dbb-b369-b49184e3dfbc: Nov 28 11:13:36 domU
    -12-31-39-17-29-5D kernel: [  309.623023] IPTables-Dropped: IN=
    eth0 OUT= MAC=12:31:39:17:29:5d:fe:ff:ff:ff:ff:ff:08:00 SRC
    =172.16.0.23 DST=10.97.42.171 LEN=103 TOS=0x00 PREC=0x00 TTL=64 ID
    =42188 PROTO=UDP SPT=33905 DPT=53 LEN=83
[truncated]
```

Disk images are stored in the filesystem of the host operating system. The file path includes the name of the instance, which is used to identify the correct image to return to the user.

Our implementation supports the retrieval of disk images from virtual machines that are powered off. New versions of QEMU and Libvirt include functionality to execute shapshots of running instances, but these features have not yet been added to OpenStack.

Listing 5.3 shows the output of using FROST from the command line to retrieve a disk image for a single virtual disk with volume name myvol-e9a5612d. FROST returns the disk image for myvol-e9a5612d, and also creates a DFXML file named report.xml in the same way as above. The requestor can validate the integrity of the image by comparing the hash value in the DFXML, as computed by the cloud provider, with the hash value computed by the requestor.

Listing 5.3: Execution of the FROST API to retrieve a disk image of volume myvol-e9a5612d. Integrity validation is easily performed.

```
$ nova get-disk myvol-e9a5612d report.xml
MD5:  b17ee04095b2a3d81eed98628072eab6
SHA1: 399f5ffaccd09fe43d642d5f0d996875ca650c9f

$ sha1sum myvol-e9a5612d
399f5ffaccd09fe43d642d5f0d996875ca650c9f myvol-e9a5612d
```

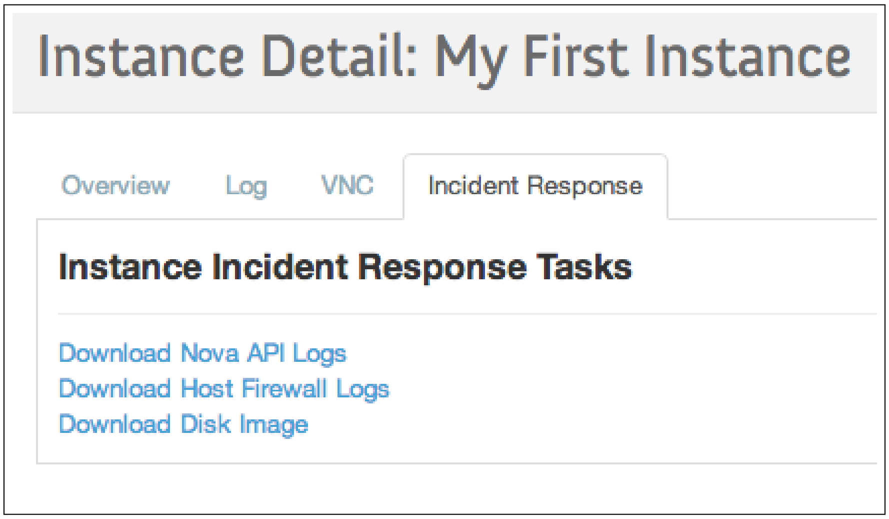
5.4.3 Management Console Web Implementation

The Management Console for OpenStack Compute contains an Instance Detail page for each virtual machine guest created by the user. We added a new tab for "Incident Response" to the Instance Detail section. This tab contains our forensic tools, and provides a space for future forensics and incident response related features.

Figure 5.3 shows the Incident Response page for a virtual machine. On this page a user can click to retrieve Nova logs, firewall logs, and a disk image. These links return a zip file that contains the data requested and a DFXML file.

5.5 Concept of Operations

Here we explain how Alice, who we introduced in Section 5.3, might use FROST to investigate an incident involving one of her virtual machines. Let us assume that one of Alice's machines has a webserver and that an attacker has compromised it and gained access



Instance Detail: My First Instance

Instance Incident Response Tasks

Download Nova API Logs
Download Host Firewall Logs
Download Disk Image

Figure 5.3: Screenshot of the OpenStack web interface showing our new incident response tab and links to FROST functions to download Nova logs, firewall logs, and disk images for one virtual machine. These links provide easy access to forensic functions for cloud users.

to the machine. Alice becomes aware of this incident, and engages law enforcement who open an investigation.

Alice's provider uses OpenStack with FROST. When she first created her account, the ALS initialized a subtree for her. When she created her virtual machine, subsequent requests were logged to the tree. Furthermore, FROST has exposed the ability to download virtual hard drive images from both the web management plane and via the Nova API.

After the incident, Alice uses her private key to retrieve forensically-sound firewall logs, Nova logs, and virtual machine images of the compromised machine, and provides them to the authorities. The firewall logs may show the attacker scanning Alice's machine before hacking it, and the disk image contains evidence of what the attacker did once he got access. This evidence is available only with assistance from Alice's provider, or with FROST, and it gives strong forensic evidence about the crime that can be used in court.

5.6 Evaluation

We conducted two evaluations of FROST. The first is an objectives-based assessment to validate that FROST can scale and produce correct results. The second is a consumer-oriented demonstration and independent appraisal to gather feedback from potential users.

We tested FROST by creating 100 fictitious users and used the API to launch five virtual machines for each user simultaneously. For each virtual machine, we associated firewall rules that allowed only SSH. With 500 virtual machines running, we used a network scanner to scan ports 1–1024 on each machine. This was done to trigger the firewall to block network traffic on the prohibited ports. We then chose a random user's key from the list of 100 users, and a random instance from the list of 500, and used the API to try and stop the virtual machine. There was only a 1% chance that the chosen user owned the chosen virtual machine, and this generated Nova logs for both successful and unsuccessful attempts.

We then chose 20 users at random and for each user requested the API logs, firewall logs, and disk image for each of the user's instances. We validated the integrity of each log and disk image returned by computing the hash of the data and comparing it to the hash value in the DFXML file. No anomalies were observed.

To scale to more users the logging mechanism needs only more storage space. Each API and firewall log entry is limited to 1KB, the syslog's maximum message size. Using SHA-1 as the cryptographic hash algorithm requires 160 bits for each tree node (user, VM, year, month, day). In the worst case this is 1664 bytes per entry. Therefore, the logging mechanism can store more than 645,000 log entries in 1GB of storage. We believe that modern servers can easily handle this load. Cloud providers could choose to share this cost with customers who wish to enable the logging service.

Cloud providers can expect minimal performance impact after deploying FROST. The overhead of calculating checksums and providing them to users is negligible. The time and bandwidth required for a user to download his or her logs or disk images is dependent upon the size of the data. We also expect users to request large data volumes, such as disk images, infrequently.

We demonstrated FROST to 12 users and administrators of a large private government cloud; their reactions were positive. One administrator said "[FROST] is exactly what OpenStack has been missing" and "I appreciate shifting the load [of investigation] away from me and onto our users." The audience was confident that FROST would be useful in incident response and forensics due to its ease of use. Users exercised FROST's web and API interfaces and described them as "intuitive and consistent with OpenStack's design." Most users anticipated automating their use of FROST, such as for collecting logs on a daily basis. They were also interested in using FROST for non-forensic purposes, such as troubleshooting and compliance. The administrators plan to deploy FROST to this cloud in mid-2013.

This evaluation shows that the integrity, completeness, and accuracy of the forensic data are intact, as identified by SWGDE and NIST in Section 5.1.1. The legal requirements are similarly met. Our solutions use computer data which are already collected and used in standard practice, or like firewall logs, are standard practice in computer networks and are easily enabled in OpenStack.

5.7 Conclusion

We have introduced the FROST suite for OpenStack, the first collection of forensic tools integrated into a cloud architecture. These tools enable cloud consumers, law enforcement, and forensic investigators to acquire trustworthy forensic data independently. In addition to incident response and forensics, FROST can also be used for real-time monitoring, metrics, or auditing.

FROST offers concrete user-accessible forensic capabilities to cloud consumers. While many businesses are still hesitant to adopt cloud solutions because of security concerns, FROST arms them with powerful and immediate response capabilities. Similar tools should be a part of all commercial cloud services, and we look forward to the creation and adoption of more such tools to enhance forensic readiness for cloud computing.

Part II

Legal Aspects of Digital Forensics for Cloud Computing

Chapter 6

Forensic Collection of Electronic Evidence from Infrastructure-as-a-Service Cloud Computing

6.1 Introduction[1]

As cloud computing becomes ubiquitous, the criminal targeting and criminal use of cloud computing is inevitable and imminent. Similarly, the need for civil forensic analyses of cloud computing has become more prevalent. Forensic investigation of cloud computing matters first requires an understanding of the technology and issues associated with the collection of electronically stored information ("ESI") in the cloud. The misuse of the broad term "cloud computing" has caused some confusion and misinformation among legal and

[1] This chapter was co-written with attorney Damien Riehl. As a piece of scholarly legal writing, the chapter uses the Bluebook style guide.

technology scholars, leading to a muddied and incomplete analysis of cloud-based discovery issues. Cases and academic analyses have dealt primarily with popular online services such as Gmail and Facebook, but they omit discussions of commercial cloud computing providers' fundamental infrastructure offerings.[2] Even worse, legal analysis about electronic discovery is largely devoid of jurisprudence concerning cloud-computing services.[3] As cloud computing becomes a large and necessary part of our computing existence, policymakers and jurists should carefully analyze how the law should best approach forensic acquisition and analysis of digital artifacts hosted by remote cloud service providers.

In early 2011, Sony was the victim of an online data breach that took the PlayStation Network offline.[4] To commit that crime, the intruder used Amazon's public cloud.[5] The FBI investigated the crime, but very little information was made public. For example, neither Amazon nor the FBI would comment on whether the former was served with a search warrant or subpoena.[6] This is a single publicly known case of a cloud-related crime, though many more are bound to emerge. Civil cases more frequently address online discovery—

[2] *See infra* Section 6.2.1.

[3] *See generally* H. Marshall Jarrett et al., U.S. Dep't of Justice, Searching and Seizing Computers and Obtaining Electronic Evidence in Criminal Investigations 115-51(2009) [hereinafter "DOJ Manual"], available at http://www.justice.gov/criminal/cybercrime/docs/ssmanual2009.pdf. The vendors of the two most popular forensic tools, Guidance EnCase and AccessData FTK, also publish documents describing the electronic discovery process and cases where their products were used; neither mentions cloud forensic acquisition, analysis, or legal precedent. *See generally* Guidance Software, EnCase Legal Journal (2011), http://www.guidancesoftware.com/DocumentRegistration.aspx?did=1000017380&id=2525; AccessData Corp., The Rules of Digital Evidence and AccessData Technology, http://accessdata.com/downloads/media/Rules_of_Digital_Evidence_and_AccessData_Technology.pdf.

[4] News: Consumer Alerts, Playstation Network, http://us.playstation.com/news/consumeralerts/ (last visited Aug. 22, 2012).

[5] *See* Joseph Galante, Olga Kharif & Pavel Alpeyev, Sony Network Breach Shows Amazon Cloud's Appeal for Hackers, Bloomberg (May 16, 2011, 4:45 PM), http://www.bloomberg.com/news/2011-05-15/sony-attack-shows-amazon-s-cloud-service-lures-hackers-at-pennies-an-hour.html.

[6] *Id.*

most often in the context of services such as Gmail or Facebook—but fewer cases have addressed cloud-computing infrastructures such as Amazon's Elastic Compute Cloud (EC2), Microsoft Azure, or Rackspace.[7] Given cloud computing's intricacies, the courts will likely continue to struggle with addressing the technology's inherent complexities.

This chapter discusses some challenges involved with electronic discovery and digital forensics arising from cloud computing infrastructure as a service, arguing that the nature of cloud computing challenges the process and product of electronic discovery. We conclude that although existing rules and doctrines—the Federal Rules of Civil Procedure, Federal Rules of Criminal Procedure, and the Fourth Amendment—are appropriately applied to the forensic acquisition and analysis of cloud-based ESI, this technology requires adapting these rules with novel interpretations. We make the following claims: (1) online users have an expectation of the geographic location of their data and thus, the laws protecting that data; (2) cloud providers should not be permitted to execute subpoenas and search warrants on behalf of law enforcement without rigorous guidelines, including challenges to the searches' scope and procedure; and (3) remote forensics of the remote service provider's forum should be governed by the laws of the remote service provider.

Section 6.2 defines the technologies and clarifies terms. Section 6.3 surveys cases involving cloud forensics, discussing how the Federal Rules of Civil Procedure, Federal Rules of Criminal Procedure, and the Fourth Amendment apply to cloud forensics. Section 6.4 takes a contrasting view, analyzing how parties might undermine cloud-derived evidence.

[7] Compare Equal Emp't Opportunity Comm'n v. Simply Storage Mgmt., LLC, 270 F.R.D. 430, 432 (S.D. Ind. 2010) (discussing discovery regarding social media sites), with Global Sessions LP v. Travelocity.com LP, No. 6:10cv671, 2012 WL 1903903, at *10 (E.D. Tex. May 25, 2012) (discussing discovery regarding EC2), and RealPage, Inc. v. Yardi Sys., Inc., No. CV 11-00690-ODW, 2012 WL 443730, at *6 (C.D. Cal. Feb. 13, 2012) (discussing discovery of generic cloud computing services such as Rackspace).

6.2 Overview of Cloud Technology for Legal Professionals

Cloud computing is still an emerging technology, but its use is expanding at a blistering pace.[8] In 2011, the United States Government implemented a "Cloud First" policy, requiring that before federal agencies make any new investments, they must evaluate cloud-computing solutions—citing the "considerable benefits to efficiency, agility, or innovation."[9] As such, several government agencies have already implemented cloud solutions,[10] and many more are anticipated to do so in the coming years.[11] Despite this mandate and rush to cloud computing, some policy makers, law enforcement, and forensic investigators do not appear to understand the nuances to investigating incidents and crimes in the cloud, nor do they fully appreciate the implications in civil discovery. Private companies are similarly rushing to cloud computing at a blistering pace.[12] Surveys indicate that most companies use cloud

[8] *See* Saul Berman et al., The Power of Cloud, IBM, 2-3 (Feb. 2012), http://www.ibm.com/cloud-computing/us/en/assets/power-of-cloud-for-bus-model-innovation.pdf.

[9] *See* Vivek Kundra, Federal Cloud Computing Strategy 19 (Feb. 8, 2011), available at http://www.cio.gov/documents/federal-cloud-computing-strategy.pdf.

[10] *See*, *e.g.*, Steve Hoffman, GSA Becomes First Federal Agency to Move Email to the Cloud Agencywide (Dec. 1, 2010), http://www.gsa.gov/portal/content/208417; see also Government Cloud Computing, Cloudbook, http://www.cloudbook.net/directories/gov-clouds/government-cloud-computing.php (last visited May 31, 2012) (compiling list of government agencies that have adopted cloud computing, including Department of Energy, NASA, National Science Foundation, National Institute of Standards & Technology, and others).

[11] Google's SaaS cloud service has obtained ISO 27001 certification for security techniques. Thomas Claburn, Google Apps Clears Key Security Hurdle, InformationWeek (May 29, 2012 3:05 pm), http://www.informationweek.com/news/cloud-computing/software/240001126. Microsoft announced a separate cloud product for government: Office 365 for Government. Kirk Koenigsbauer, Announcing Office 365 for Government: A US Government Community Cloud, Office 365 (May 30, 2012), http://blogs.office.com/b/microsoft_office_365_blog/archive/2012/05/30/announcing-office-365-for-government-a-us-government-community-cloud.aspx. Both of these developments are sure to rapidly increase government adoption of cloud services.

[12] *See*, *e.g.*, Berman et al., *supra* note 8.

computing,[13] and many use multiple cloud services.[14] Given its rapid adoption, cloud computing has serious legal implications in the United States and around the world. But before analyzing and developing the law, however, one must first understand the technology.

Legal scholars and practitioners have long made analogies between computer hard drives and filing cabinets.[15] Paul Ohm observed: "Warehouses—and even less so filing cabinets—are insignificant containers of information compared to today's hard drives, and the analogy will only become more mismatched over time."[16] To extend Ohm's analogy (warehouses and filing cabinets) to cloud-based data requires the following modification: cut up each document, store each piece in a different locked filing cabinet, and distribute all those cabinets to different warehouses around the world. As Ohm concluded, "Today's technology poses a constitutional puzzle that is different in kind, not just in degree, from the one solved only a few decades ago."[17]

[13] *See* Smriti Sharma, 74 Percent Companies Using Cloud Services, Global Services (Apr. 20, 2012), http://www.globalservicesmedia.com/IT-Outsourcing/Infrastructure-Management/74-Percent-Companies-Using-Cloud-Services/22/6/12123/GS1204209710723.

[14] *See* Meghan Kelly, 86 Percent of Companies Use Multiple Cloud Services, Says Study, Venture Beat (May 10, 2012), http://venturebeat.com/2012/05/10/cloud-services-data/ (surveying one company's 3,200 customers in 80 different countries).

[15] *See*, *e.g.*, Gruenspecht, "Reasonable" Grand Jury Subpoenas: Asking for Information in the Age of Big Data, 24 Harv. J. L. & Tech. 543, 552 (2011) (citing In re Grand Jury Subpoena Duces Tecum Dated Nov. 15, 1993, 846 F. Supp. 11, 12-13 (S.D.N.Y. 1994)); Orin S. Kerr, Searches and Seizures in a Digital World, 119 Harv. L. Rev. 531, 550 (2005) (acknowledging the usefulness of treating a computer like a container).

[16] Paul Ohm, Massive Hard Drives, General Warrants and the Power of Magistrate Judges, 97 Va. L. Rev. In Brief 1, 8 (2011).

[17] *Id.*

6.2.1 Cloud Computing

As we presented in Chapter 2, cloud computing is a broad, generic term with many proffered meanings and definitions.[18] It has infiltrated the vernacular and has been debased in marketing and media. It would be an oversimplification to say that cloud computing refers to anything "in general" other than it is not the computing device in your physical possession.[19] Ultimately, cloud computing is a waypoint in conglomeration of decades of technology evolution. Starting with single-user standalone computers and multi-user mainframes, cloud computing's most direct ancestors were utility and grid computing.[20]

First, it is important to distinguish between cloud *services* and cloud *computing*. Facebook and Gmail are remote cloud services, but they are not cloud computing.[21] Examples of cloud computing are Amazon Elastic Compute Cloud (EC2), Microsoft Azure, and Rackspace web hosting.[22] "Cloud" is a generic term that refers to a network where the

[18] *See* Peter Mell & Tim Grance, The NIST Definition of Cloud Computing, Nat'l Inst. of Standards & Tech., 2 (Sept. 2011), http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf. *See generally* Nicole Galli & Edward Gecovich, Cloud Computing and the Doctrine of Joint Infringement: "Current Impact" and Future Possibilities, 11 J. Marshall Rev. Intell. Prop. Law 673, 676 (2012).

[19] Michael Armbrust et al., Above the Clouds: A Berkley View of Cloud Computing, UC Berkeley Reliable Adaptive Distributed Systems Lab, 4 (Feb. 10, 2009), http://inst.cs.berkeley.edu/˜cs10/fa10/lec/20/2010-11-10-CS10-L20-AF-Cloud-Computing.pdf.

[20] *See* Sourya Biswas, Cloud Computing vs Utility Computing vs Grid Computing: Sorting the Differences, CloudTweaks (Feb. 1, 2011, 7:44 AM), http://www.cloudtweaks.com/2011/02/cloud-computing-vs-utility-computing-vs-grid-computing-sorting-the-differences/.

[21] Some authors have mistakenly tied online services to cloud computing. See, *e.g.*, Marc Aaron Melzer, Copyright Enforcement in the Cloud, 21 Fordham Intell. Prop. Media & Ent. L.J. 403, 405 (2011) ("To illustrate this point . . . three sites that can readily be considered examples of SaaS cloud computing: Facebook, the social networking site and number one website by traffic; Yahoo! Mail, the number one webmail provider by accounts; and YouTube, a video sharing site . . . ."). Facebook, Yahoo! Mail, and YouTube do not meet the NIST definition. See Mell & Grance, *supra* note 18. Further, they are supported by advertising, not billed to the customer based on their usage.

[22] *See* Amazon Elastic Compute Cloud (Amazon EC2), Amazon Web Services, http://aws.amazon.com/ec2/ (last visited Aug. 7, 2012); Cloud Services on Windows Azure, Windows Azure, http://www.windowsazure.com/en-us/home/scenarios/cloud-services/ (last visited Aug. 7, 2012);

physical location and inner workings are abstracted away and unimportant to the usage.[23] The Internet is one type of cloud.[24] For example, to use Gmail, one need not know the physical location of Gmail's servers. Cloud computing also takes advantage of this definition of a cloud, as it is also a service connected to a network, often the Internet.[25] But cloud computing offers customers additional functionality in the form of raw remote computing resources, such as processing power or data storage, and the ability to provision[26] those resources themselves.[27]

Unlike Gmail or Facebook, which provide users with specific services, cloud computing is a canvas that programmers can use to create any service they choose.[28] This chapter limits discussion to public clouds rather than private clouds on a company's premises. Few have

Open Public, Private, and Hybrid Clouds, Rackspace, http://www.rackspace.com/cloud/ (last visited Aug. 7, 2012).

[23] Mell & Grance, *supra* note 18.

[24] Cloud was first used to describe telecommunication networks, where the customer was blissfully unaware of the inner workings of how their telephone conversation was transmitted to the remote end. The term was later used to describe computer networks, and ultimately to describe the Internet specifically. See Antonio Regalado, Who Coined "Cloud Computing"?, Tech. Rev. (Oct. 31, 2011), http://www.technologyreview.com/business/38987/.

[25] *See* Mell & Grance, *supra* note 18, at 3. Cloud computing by definition exposes resources over a network—Public clouds offer these services over the Internet; Private clouds offer services on a private network; such as a private, internal company network; Hybrid clouds link the Internet's public resources with an organization's private resources. *See id.*

[26] "Provisioning" of cloud resources refers to the act of requesting, purchasing, and acquiring the resource so that it is ready for use. This process is often done by filling out a simple form on a management webpage. After the request is received, the storage or computation services can be available to users in as little as a few seconds. See, *e.g.*, Amazon Elastic Block Store (EBS), Amazon Web Services, http://aws.amazon.com/ebs/ (last visited Aug. 7, 2012).

[27] *Id.*

[28] *See generally* Geva Perry, How Cloud & Utility Computing are Different, GigaOM (Feb. 28, 2008, 4:42 PM), http://gigaom.com/2008/02/28/how-cloud-utility-computing-are-different/.

analyzed the thorny legal issues that arise in electronic discovery of utility cloud computing, a topic explored in Section 6.3.

Four of cloud computing's defining characteristics are particularly important to legal analysis: (1) on-demand self-service; (2) rapid elasticity; (3) location independence; and (4) data replication.[29] First, within the limits defined by the cloud provider, the customer has complete control over the provisioning and deprovisioning of cloud resources, which they can do quickly and on-demand.[30] Second, because of this ease and elasticity, customers can cause evidence to appear and disappear at a moment's notice.[31] Third, like other resources on the Internet, the cloud resource's physical location has no bearing on the use or provisioning of those resources, which could exist in one or more data centers around the world.[32] Finally, to provide data reliability and fault-tolerance, cloud providers routinely replicate data on several computers in multiple physical locations.[33] Further, cloud environments typically

[29] *See* Mell & Grance, *supra* note 18, at 2.

[30] *See* Yung Chou, Cloud Computing for IT Pros, Part I: What is Service, TechNet Blogs (Dec. 15, 2010, 4:06 PM), http://blogs.technet.com/b/yungchou/archive/2010/12/15/cloud-computing-concepts-for-it-pros-1-3.aspx.

[31] *See generally* Alberto G. Araiza, Electronic Discovery in the Cloud, 2011 Duke L. & Tech. Rev. 8, 35 (2011) (discussing the increased legal risks of deleting ESI under the cloud). But see Simson Garfinkel et al., Practical Unix and Internet Security 675 (3d ed. 2003) (suggesting that the deletion of data is not permanent).

[32] *See* Perry, *supra* note 28 ("Although it is difficult to come up with a precise and comprehensive definition of cloud computing, at the heart of it is the idea that applications run somewhere on the "cloud" (whether an internal corporate network or the public Internet) – we don't know or care where.").

[33] *See*, *e.g.*, Amazon Web Services, Amazon Web Services: Overview of Security Processes 7 (May 2011), http://d36cz9buwru1tt.cloudfront.net/pdf/AWS_Security_Whitepaper.pdf ("Data stored in Amazon S3, Amazon SimpleDB, or Amazon Elastic Block Store (EBS) is redundantly stored in multiple physical locations as a part of normal operation of those services and at no additional charge."); Jeffrey Richter, Understanding Cloud Storage, Windows Azure, http://www.windowsazure.com/en-us/develop/net/fundamentals/cloud-storage/ (last visited Aug. 3, 2012) ("In order to achieve highly available and scalable applications, Windows Azure offers multitenant storage machines within the various Windows Azure data centers. These machines replicate your data ensuring that if one replica fails, others are still viable.").

store data in a distributed file system, breaking single files into pieces that can be stored on multiple independent storage devices, such as hard drives.[34]

Under the Stored Communications Act, which governs service providers' voluntary and compelled disclosure of electronic communications and records, cloud computing likely fits the definition of a "remote computing service" ("RCS").[35] When Congress enacted this legislation in 1986, it likely never contemplated anything akin to modern cloud computing.[36] At that time, many businesses could not afford largescale computation or storage, so data were stored by providers and accessed remotely.[37] Congressional discussion of remote computing services essentially described them as timesharing services.[38] Those systems are distant relatives of today's cloud-computing offerings. By nature of their In-

[34] *See generally* Sean Gallagher, The Great Disk Drive in the Sky: How Web Giants Store Big—and we mean big—Data, Ars Technica (Jan. 26, 2012, 9:00 PM EST), http://arstechnica.com/business/2012/01/the-big-disk-drive-in-the-sky-how-the-giants-of-the-web-store-big-data/ (explaining how Google, Microsoft, and Amazon, have adopted distributed file systems and the architecture behind such storage systems).

[35] *See* 18 U.S.C. § 2711(2) (2006) ("the term 'remote computing service' means the provision to the public of computer storage or processing services by means of an electronic communications system"); see also William Jeremy Robison, Note, Free at What Cost?: Cloud Computing Privacy under the Stored Communications Act, 98 Geo. L.J. 1195, 1212-14 (2010) (examining cloud computing as a remote computing service under the Stored Communications Act).

[36] *See* Crispin v. Christian Audigier, Inc., 717 F. Supp. 2d 965, 971 n.15 (C.D. Cal. 2010); Derek Constantine, Note, Cloud Computing: The Next Great Technological Innovation, the Death of Online Privacy, or Both?, 28 Ga. St. U. L. Rev. 499, 502 (2012).

[37] ECPA Reform and the Revolution in Cloud Computing: Hearing Before the Subcomm. on the Constitution , Civil Rights, and Civil Liberties of the H. Comm. on the Judiciary, 111th Cong. 19-20 (written testimony of Richard Salgado, Senior Counsel, Law Enforcement and Information Security, Google Inc.).

[38] *See* H.R. Rep. No. 99-647, at 23 (1986) (citing Electronic Communications Privacy Act: Hearing on H.R. 3378 Before the Subcomm. on Courts, Civil Liberties, and the Admin. of Justice of the H. Comm. on the Judiciary, 99th Cong. 78 (1986) (statement of P. Nugent, Chairperson, Committee on Computer Systems and Communications Privacy)). Nugent's examples of remote computing services were current for the day, including "the service customer's sales people use terminals to electronically transmit sales order information from geographically dispersed locations to the service vendor's computer center." Id; see also S. Rep. No. 99-541, at 10 (1986) (describing RCS as "essentially a timesharing arrangement"). Contrasted from 25 years ago, today's cloud computing environment is fundamentally different, offering more general computing services that customers can quickly and easily provision on demand. The district court in Viacom International v. YouTube ruled that YouTube was a remote computing service. 253 F.R.D. 256, 264 (S.D.N.Y. 2008); see also Flagg v. City of Detroit, 252 F.R.D. 346, 363 (E.D. Mich. 2008) (holding that "the archive maintained by

ternet connectivity and client-server model, cloud providers also provide an "electronic communication service" ("ECS").[39] When selling raw infrastructure resources that include network bandwidth, providers broadly deliver the ability to send or receive any kind of Internet communication.[40]

Cloud-hosted computers can play the same roles in a case as can any other types of computers.[41] In criminal cases, a cloud-hosted computer could involve or constitute contraband, evidence, fruits, or instrumentalities.[42] Similarly, cloud-hosted computers may contain rich troves of evidence in civil matters.[43] But despite conventional wisdom, seizing a cloud provider's hardware in a criminal matter is often unfruitful.[44] And in a civil matter,

[the service provider] constitutes "computer storage," and that the company's maintenance of this archive on behalf of the City is a "remote computing service" as defined under the SCA").

[39] 18 U.S.C. § 2510(15) (2006) (defining "electronic communication service" as "any service which provides to users thereof the ability to send or receive wire or electronic communications"). A cloud provider is also likely an "electronic communications system," defined as "any wire, radio, electromagnetic, photooptical or photoelectric facilities for the transmission of wire or electronic communications, and any computer facilities or related electronic equipment for the electronic storage of such communications." 18 U.S.C. § 2510(14) (2006).

[40] Customers that set up services in the cloud may or may not also be an ECS, depending on whether or not they provide the ability to send communications to third parties. See Becker v. Toca, No. 07–7202, 2008 WL 4443050, at *4 (E.D. La. Sept. 26, 2008).

[41] *See* State v. Bellar, 217 P.3d 1094, 1110 (Or. Ct. App. 2009) (finding no distinction between data stored on a personal computer and data copied and stored on another medium in the context of privacy rights).

[42] *See* David A. Couillard, Note, Defogging the Cloud: Applying Fourth Amendment Principles to Evolving Privacy Expectations in Cloud Computing, 93 Minn. L. Rev. 2205, 2218-20 (2009).

[43] *See*, *e.g.*, Fed. Trade Comm'n v. First Universal Lending, LLC, 773 F. Supp. 2d 1332, 1342 (S.D. Fla. 2011). *See generally* Steven C. Bennett, E-Discovery Meets the Cloud, N.Y. St. B.J., May 2011, at 45-46 (discussing discovery and a litigator's duties in the context of cloud computing).

[44] Unfortunately, the DOJ Search and Seizure Manual still recommends it. DOJ Manual, *supra* note 3, at 70-71; see also Liquid Motors, Inc. v. Lynd, No. 3:09–cv-0611–N (N.D. Tex. Apr. 3, 2009) (order ruling that the FBI had reasonable cause to seize computer servers of a cloud-like provider, even though data from other innocent customers were co-mingled with the search warrant's target).

such a seizure it is often unduly burdensome or logistically impossible.[45] Cases should evolve to contemplate and address the nuances of cloud computing, whose normal operations involve breaking up files and storing them across many servers in many locations.[46] In fact, as part of normal operations, cloud-based data can move easily and transparently to different servers or storage locations.[47] This is not a sufficient argument for a party to request every server on each of the cloud provider's premises. A forensic examiner analyzing conventional computer hardware has the benefit of being able to search for and sometimes recover lost or deleted data still resident on the disk.[48] Although this may be possible with a copy of a virtual machine, it requires additional evidence for a storage service such as Amazon's Simple Storage Service (S3).[49] For example, if the provider keeps logs of what files are deleted, who deleted them, and when were they deleted, that could be useful metadata, even if content proves unrecoverable.

[45] *See* Fed. Judicial Ctr., Managing Discovery of Electronic Information: A Pocket Guide for Judges 17 (Barbara Rothstein et al. eds. 2007).

[46] Given cloud computing's distributed nature, courts in such cases should move beyond the concept of a server as a singular document repository. See, *e.g.*, United States v. Hay, 231 F.3d 630, 637 (9th Cir. 2000) (upholding seizure of entire computer as contraband in child pornography case); Davis v. Gracey, 111 F.3d 1472, 1480 (10th Cir. 1997) ("[T]he computer equipment was more than merely a "container" for the files; it was an instrumentality of the crime."). Unlike Davis, where the court observed that "the obvious difficulties attendant in separating the contents of electronic storage from the computer hardware during the course of a search," cloud computing makes this separation natural and convenient. 111 F.3d at 1480.

[47] Jeff Boles, The Benefits of Cloud-based Storage, Part 2, InfoStor (Nov. 10, 2008), http://www.infostor.com/index/articles/InfoStor-Article-Tool-Template/_saveArticle/articles/infostor/backup-and_recovery/cloud-storage/the-benefits_of_cloud-based.html.

[48] *See*, *e.g.*, United States v. Gourde, 440 F.3d 1065, 1071 (9th Cir. 2006); Hay, 231 F.3d at 635-36; United States v. Crist, 627 F. Supp. 2d 575, 578 (M.D. Pa. 2008); see also DOJ Manual, *supra* note 3, at 69.

[49] In normal operations, the cloud fabric does this reassembly automatically. If the cloud provider is the criminal defendant or a civil party (not a third party)—or if there is doubt in the trustworthiness of the fabric—then the data's veracity may be suspect.

In federal criminal cases, the decision of whether to seize hardware also weighs into the choice between a Rule 41 search warrant under the Federal Rules of Criminal Procedure and an Electronic Communications Privacy Act ("ECPA") warrant.[50] While a Rule 41 warrant might be justified for seizing hardware and imaging hard drives on-site, courts have traditionally issued such warrants only for objects physically in the judicial district where the court is located.[51] For ECPA warrants, the statute permits issuance from any court of "competent jurisdiction."[52] The Justice Department's Search and Seizure Manual contains a sample ECPA warrant for email hosted by an ISP[53] as well as a sample Rule 41 warrant for removing computers from the premises.[54]

For civil cases, issuing a subpoena under Rule 45 of the Federal Rules of Civil Criminal Procedure is a process similar to that under Rule 41 of the Federal Rules of Criminal Civil Procedure. Rule 45 permits subpoena service in three instances: (1) within the issuing court's district; (2) "within 100 miles of the place specified for the . . . trial, production, or inspection;" or (3) within the state of the trial, production, or inspection.[55] State courts have a limited geographic jurisdiction within their state's borders, so a party cannot enforce

[50] *See* United States v. Daccarett, 6 F.3d 37, 46, 53 (2d Cir. 1993); In re United States, 665 F. Supp. 2d 1210, 1214 (D. Or. 2009); DOJ Manual, *supra* note 3, at 112, 133-34.

[51] *See* DOJ Manual, *supra* note 3, at 84.

[52] 18 U.S.C. § 2703(a) (2006).

[53] DOJ Manual, *supra* note 3, at 255-62.

[54] DOJ Manual, *supra* note 3, at 241-50.

[55] Fed. R. Civ. P. 45(b)(2).

extraterritorial subpoenas.[56] As such, a civil party seeking an out-of-state subpoena may choose to initiate an action in a court in the jurisdiction where the hardware is located.

But because cloud-computing data may be distributed throughout the country or around the world, determining the physical location of such a "production" or "inspection" raises several questions. Is cloud-computing data "produced" at the locations of dozens of servers around the world? If a civil party seeks to "inspect" documents, rather than have them "produced," that party historically would have traveled to a document repository, requiring subpoena service near that repository.[57] With cloud computing, however, does it matter that a requesting party could conceivably conduct such an "inspection" using a computer physically located anywhere in the world, including the venue jurisdiction? Given these quandaries, subpoena service location may be unclear, but the most obvious service location is a cloud service provider's headquarters or principle place of business.

6.2.2 Digital Forensics for Cloud Computing

Today, ESI is ubiquitous and plays a role in constitutes a part of nearly every legal case.[58] Digital forensics uses scientific and proven methods to analyze and interpret ESI to reconstruct events.[59] The forensic examiner is tasked with analyzing ESI to reconstruct a timeline

[56] *See* 98 C.J.S. Witnesses § 28 (2002) ("Service of a subpoena of a state court outside of the state where it issued is a nullity.").

[57] *See* Fed. R. Civ. P. 34; Fed. R. Civ. P. 45(b)(2).

[58] *See* Joseph A. Martin & Christine S. Baxter, A Practical Guide to Admitting ESI at Trial, 19 American Bar Association Business Torts Litigation Newsletter, no. 4, Summer 2012, at 2, available at http://www.archerlaw.com/files/articles/A%20Practical%20Guide%20to%20Admitting%20ESI%20at%20Trial.pdf.

[59] *See* Brian Carrier, Defining Digital Forensic Examination and Analysis Tools, Digital Forensics Research Workshop, Aug. 2002, at 2, available at http://www.dfrws.org/2002/papers/Papers/Brian_carrier.pdf. Many people use the term forensics in non-criminal contexts because no other term describes digital investigations in non-criminal situations, such as civil cases, intelligence gathering, and internal corporate investigations.

that describes, as best as possible, what happened and when.[60] Although the forensic examiner could be asked to analyze single documents or email messages,[61] traditional forensics focuses on analyzing entire hard drives.[62] Cloud computing injects new and non-trivial challenges to this task, including remotely located data, lack of control, layers of complexity, and authenticity. Recall Case Study 2 from Chapter 3:

> Mallory is a hacker who intends to exploit victims by placing a malicious webpage in the cloud. She uses a vulnerability to exploit the cloud-hosted website of a legitimate company, Buzz Coffee. After hacking into the server, she installs software that infects victims who browse the website. Users complain to Buzz Coffee that they are being infected, so the company seeks to fix the problem and investigate the issue.

This realistic scenario illustrates some of the forensic task's legal issues. If Buzz Coffee owned, operated, and housed the server, then the technical and legal process of acquiring evidence would be routine. Even if Buzz Coffee leased the server from a third party that who housed it remotely, it would add very little complexity. However, because this scenario involves cloud computing, Buzz Coffee owns no hardware and it might have no idea where any of its data are stored. As discussed in Section 6.3, many conventional questions—such as those of jurisdiction, subpoenas, search warrant issuance and execution, and trustworthy evidence—take on unconventional complexity.

[60] *See* Ovie L. Carroll et al., Computer Forensics: Digital Forensic Analysis Methodology, 56 United States Attorneys' Bulletin, no. 1, Jan. 2008, at 4; Christopher Pogue, Sniper Forensics: GFIRST Edition, Government Forum of Incident Response and Security Teams, 34 (2011), http://www.us-cert.gov/GFIRST/presentations/2011/Sniper_Forensics.pdf.

[61] *See* Carroll et al., *supra* note 60, at 3.

[62] *See* Tyler Newby & Ovie L. Carroll, Rethinking the Storage of Computer Evidence, 56 United States Attorneys' Bulletin, no. 1, Jan. 2008, at 60.

Amazon is unusually open and candid about its internal processes and support for e-discovery in their cloud offerings known as Amazon Web Services ("AWS").[63] In one risk-management white paper, Amazon describes how it meets customers' needs for electronic discovery, stating that "[c]ustomers are responsible for responding appropriately to legal procedures involving the identification, collection, processing, analysis, and production of electronic documents they store or process using AWS," and "[u]pon request, AWS may work with customers who require AWS' assistance in legal proceedings."[64] Unlike some cloud providers, Amazon does not explicitly offer services such as forensics or incident response assistance.[65] Rather, Amazon and other public cloud providers largely work with parties and law enforcement to the extent required by law.[66]

6.3 Obtaining Forensic Evidence from the Cloud

Numerous constitutional and statutory provisions govern searching and acquiring forensic evidence from cloud providers. On the federal level, the analysis focuses on the Federal Rules of Civil Procedure, the Federal Rules of Criminal Procedure, and the Fourth Amendment. In this section, we discuss how each applies to acquiring cloud-based ESI.

[63] *See generally* Amazon Web Services: Risk and Compliance, Amazon Web Services, 11 (January 2013), http://media.amazonwebservices.com/AWS_Risk_and_Compliance_Whitepaper.pdf (addressing whether AWS's cloud services meet e-discovery procedures and requirements).

[64] *Id.*

[65] E.g., Terremark Worldwide, Investigative Response & Forensics, http://www.terremark.com/services/security-services/investigative-response.aspx (last visited Aug. 15, 2012) (advertising managed forensics and incident response, whereby the customer pays and the provider performs the work).

[66] *See* AWS Service Terms, Amazon Web Services, https://aws.amazon.com/serviceterms/ (last visited Aug. 15, 2012) (stating that Amazon removes content "pursuant to the Digital Millennium Copyright Act or as required to comply with law or any judicial, regulatory or other governmental order or request").

6.3.1 Federal Rules of Civil Procedure and Federal Rules of Criminal Procedure

Criminal and civil cases use similar analyses to determine which party is the proper discovery target. Both Federal Rule of Civil Procedure 34 and Federal Rule of Criminal Procedure 16 permit a party to request data "in the responding party's possession, custody, or control."[67] For cloud computing, the "responding party" is usually the cloud provider (as with third-party subpoenas), or a cloud provider's customer.[68] However, the question of who has "possession, custody, or control" is more complex.

For example, Dropbox is an online storage service that uses AWS for data storage.[69] Customers negotiate directly with Dropbox, not Amazon.[70] If a Dropbox customer is sued or placed under criminal investigation, the opposing party could potentially request data from Dropbox, Amazon, or both. As demonstrated below, the seeking party's choice of target depends on what data are sought.

When a customer uploads data to the cloud, that customer also arguably transfers the data's custody and possession to the cloud service provider—yet the customer may still retain "control."[71] Depending on the services provided and the parties' contractual relationship,

[67] Fed. R. Civ. P. 34(a)(1); see also Fed. R. Crim. P. 16(a)(1)(B)(i).

[68] *See generally* Fed. R. Civ. P. 34(b)(2); Fed. R. Crim. P. 16(d) (describing the steps required from a responding party).

[69] *See* Where does Dropbox store everyone's data?, Dropbox http://www.dropbox.com/help/7 (last visited Aug. 15, 2012) (stating that "all files stored online by Dropbox are encrypted and kept securely on Amazon's Simple Storage Service (S3) in multiple data centers located around the United States").

[70] *See generally* id. (noting that Amazon owns Dropbox's physical servers); DropBox is just a frontend to Amazon S3 with a killer sync feature, Cloudiquity (Mar. 25, 2012, 12:58 PM), http://www.cloudiquity.com/2012/03/dropbox-is-just-a-frontend-to-amazon-s3-with-a-killer-sync-feature/ (noting that Dropbox employs a frontend sync feature that syncs files stored on Amazon's S3 servers).

[71] AWS Customer Agreement, Amazon Web Services, at § 8.1, https://aws.amazon.com/agreement/ (last updated Mar. 15, 2012) (stating that "As between [AWS] and [content owner], [owner] or your licensors own

the cloud provider may well act as the customer's agent. Generally, to establish an agency relationship, the agent must be authorized to act for the principal, thereby binding the principal by the agent's words or actions.[72] The issue of whether an agency relationship exists is largely fact-dependent.[73] More so than for other types of cloud services (*e.g.*, PaaS or SaaS), the agency relationship for parties to an IaaS contract appears clearer because the customer has additional control.[74] For example, AWS customers can instruct the provider to execute automatic actions based on particular events.[75] For instance, the customer can instruct AWS as follows: if a customer's website becomes overwhelmed with too many requests, then AWS should automatically start another virtual machine to assist with the load.[76] This arrangement could be interpreted as one of express actual authority:

all right, title, and interest in and to Your Content . . . including any related intellectual property rights"); see also Security, Dropbox, https://www.dropbox.com/teams/security (last visited Aug. 15, 2012) (specifying that users gain "added control" over their data because Dropbox provides extra security and password protection).

[72] *See* Black's Law Dictionary 70 (9th ed. 2009) (defining "agency" as "[a] fiduciary relationship created by express or implied contract or by law, in which one party (the agent) may act on behalf of another party (the principal) and bind that other party by words or actions."); Harold Gill Reuschlein & William A. Gregory, The Law of Agency and Partnership § 1, at 3 (2d ed. 1990); see also Asa-Brandt, Inc. v. ADM Investor Servs., Inc., 344 F.3d 738, 743 (8th Cir. 2003) ("[I]n determining whether an agency relationship exists, the question hinged on the principal's right to exercise control over the activities of the agent." (citing Gunderson v. ADM Investor Servs., Inc., No. 99-4032, 2000 WL 1154423, at *2 (8th Cir. Aug. 16, 2000))); United States v. Bonds, 608 F.3d 495, 505 (9th Cir. 2010) (analyzing the Second Restatement's ten factors, noting that the "essential ingredient . . . is the extent of control exercised by the employer." (quoting NLRB v. Friendly Cab Co., 512 F.3d 1090, 1096 (9th Cir. 2008) (alteration in original)).

[73] *See* Section 6.2.1 C.J.S. Agency Generally § 7 (1972) (noting that determining if an agency relationship exists is a question of fact).

[74] *See*, *e.g.*, Sample Technology Statements of Work (SOWs), U.S. Gen. Services Admin., http://www.gsa.gov/portal/content/133795 (last updated Aug. 22, 2012) (providing samples of IaaS contracts for many different aspects of cloud storage and data protection).

[75] *See* AWS Management Console, Amazon Web Services, http://aws.amazon.com/console/#eb (last accessed Aug. 24, 2012) (describing the different cause and effect mechanisms available from AWS).

[76] *See id.* (describing the "Elastic Beanstalk" feature which "handles the details of capacity provisioning, load balancing, auto-scaling, and application health monitoring").

the customer acts as principal and AWS acts as the customer's agent.[77] If AWS acts as an agent, the customer's fiduciary, then the customer would also arguably have "control" over its cloud-computing data.[78] As such, the customer could be required to produce the cloud-computing data that it controls.

Where discovery requests and subpoenas are issued to cloud providers directly and without reference to whether the provider "controls" those data, those situations require a different analysis. To discuss what data are in the cloud provider's possession, custody, or control, one should first understand what data might be available.[79] IaaS can be viewed as a multi-layered cake, with each layer independently comprising part of the cloud service. The cake's top layer contains the customer's data and applications, which Internet users may utilize as a webpage or database.[80] These data are the first that may be available and by definition of IaaS, the data are owned and controlled by the customer.[81] The next layer is the guest virtual machine, which in IaaS is also owned and controlled by the customer.[82]

[77] *See* FTC v. First Universal Lending, Llc, 773 F. Supp. 2d 1332, 1347-49 (S.D. Fla. 2011) (discussing a party's data on servers owned by cloud service Salesforce, which constituted the "backbone" of a party's business).

[78] Compare AWS Customer Agreement, *supra* note 71 (describing customer control in AWS services), with C.J.S. Agency, *supra* note 73 (describing the agency relationship).

[79] The complete set of forensic data available to a requestor is categorically unknown. The public cloud providers have thus far withheld their capabilities, possibly because they are protecting the proprietary implementation that gives them competitive advantage. We speculate about data that are likely available, but cannot speculate about the provider's practical ability to collect these data.

[80] *See* Bill Loeffler et al., What is Infrastructure as a Service, TECHNET (Sept. 13, 2011, 7:07 AM), http://social.technet.microsoft.com/wiki/contents/articles/4633.what-is-infrastructure-asa-service.aspx (illustrating and comparing the different types of cloud service models).

[81] *See* Infrastructure as a Service (IaaS), U.S. Gen. Services Admin., www.gsa.gov/iaas (last updated July 5, 2012).

[82] *See* Infrastructure as a Service, CDW, 2 (2011), available at http://www.edtechmagazine.com/higher/sites/edtechmagazine.com.higher/files/108289-wp-inf_service_df.pdf; Information Supplement: PCI DSS Virtualization Guidelines, PCI Security Standards Council, 23 (June 2011), https://www.pcisecuritystandards.org/documents/Virtualization_InfoSupp_v2.pdf.

The cake's third layer is the hypervisor, which is special software that runs on a provider's computer (called the host), thus allowing many virtual machines to run independently on a single physical machine.[83] Below the physical machine is the distributed array of storage disks.[84] The cake's base is the computer networking that interconnects the components, providing high bandwidth to the Internet.[85]

To date, the major cloud providers have not yet released their policies regarding their responses to civil or criminal requests, nor have they described the types of records and data that they will make available.[86] Nevertheless, cloud providers do have data that could prove useful in criminal and civil matters. For example, cloud providers maintain data related to subscriber information and billing records.[87] Because customers are billed based on their usage, records relating to service usage should also be available.[88] Beyond these obvious requests, providers often keep other data for some time. A provider uses connection information (sometimes called NetFlow records), to records an Internet communication's

[83] Francoise Gilbert, Cloud Service Contracts May Be Fluffy: Selected Legal Issues to Consider Before Taking Off, 14 No. 6 J. Internet L. 1, Dec. 2010, at 17, 19; PCI Security Standards Council, *supra* note 82, at 7.

[84] *See* Loeffler et al., *supra* note 80; PCI Security Standards Council, *supra* note 82, at 6.

[85] *See* Loeffler et al., *supra* note 80; PCI Security Standards Council, *supra* note 82, at 6.

[86] *See* Ashish S. Prasad, Cloud Computing and Social Media: Electronic Discovery Considerations and Best Practices, The Metropolitan Corporate Counsel, Feb. 2012, at 26, 27, available at: http://www.metrocorpcounsel.com/articles/17454/cloud-computing-and-social-media-electronic-discovery-considerations-and-best-practic; cf., John Soma et al., Chasing the Clouds without Getting Drenched: A Call for Fair Practices in Cloud Computing Services, 16 J. TECH. L. & POL'Y 193, 220-21 (2011).

[87] *See* Joshua S. Parker, Note, Lost in the Cloud: Protecting End-User Privacy in Federal Cloud Computing Contracts, 41 Pub. Cont. L.J. 385, 396-97 (2012).

[88] *See id*; see also Architecture for Managing Clouds, Distributed Management Task Force, 39 (June 18, 2010), http://www.dmtf.org/sites/default/files/standards/documents/DSP-IS0102_1.0.0.pdf.

two endpoints; this non-content data can be useful as a historical record.[89] When a customer wishes to procure or manage cloud services, that customer typically visits a special website to manage those actions.[90] That website and its underlying components may also be an attractive source of forensic evidence. The provider might be able to produce logs showing successful and unsuccessful logins, the logins' IP addresses and geographic origins, and their time and date. If services can be provisioned programmatically, then similar logs may be available.

Although the cloud system's operation may not require *humans* to know where data are located (*e.g.*, such as in a server or data center), the underlying *infrastructure* must know that information.[91] The provider may record system logs that describe where the data are, who created them, and when they were created, modified, or deleted. In sum, no universal template currently exists for parties and law enforcement seeking cloud data; often, they do not know what they can or should ask for. Moreover, the data that cloud service providers house can be as diverse as the cloud service providers themselves. We present a sample search warrant in Chapter 7.

Regarding IaaS, data inside a customer's virtual machine (*e.g.*, webpages) are hidden even from the provider unless the customer makes that data available.[92] The cloud provider, whose ownership and responsibility extend to the hypervisor and below, could access the computer files that make up the virtual machine and when responding to discovery, they

[89] *See* Jamie Epstein, Get in the Know, NetFlow is the Way to Go, TMCnet.com (July 30, 2012), http://netflow.tmcnet.com/articles/300888-get-the-know-netflow-the-way-go.htm.

[90] *See* Loeffler et al., *supra* note 80.

[91] *See* Distributed Management Task Force, *supra* note 88, at 26.

[92] *See* Wely Lau, Comparing IAAS and PAAS: A Developer's Perspective, ACloudyPlace (Jan. 13, 2012), http://acloudyplace.com/2012/01/comparing-iaas-and-paas-a-developers-perspective/.

could provide a copy of that virtual machine.[93] Providers are also capable of collecting content and non-content forensic evidence in their possession, custody, and control. For example, they could collect network packet captures of all ingress and egress network traffic from their cloud, they could collect logs showing the data's physical storage locations, and they have billing data about the provisioning and usage of cloud resources.[94]

Providers' contractual language with their customers will determine the extent to which those customers may access these data.[95] To complicate matters, providers possess some data over which their customers may not have access (*e.g.* such as infrastructure logs), as well as other data over which the providers may not have control (*e.g.*, such as customer's data).[96]

[93] *See* Wayne Jansen & Timothy Grance, Guidelines on Security and Privacy in Public Cloud Computing, Nat'l Inst. of Standards & Tech., 12, 18 (Dec. 2011), http://csrc.nist.gov/publications/nistpubs/800-144/SP800-144.pdf.

[94] *See id.* at 12, 20-21 (explaining that cloud service providers have access to a lot of information that the user does not have access to). The Communications Assistance for Law Enforcement Act of 1994 (CALEA) requires telecommunications carriers to assist law enforcement in performing electronic surveillance pursuant to court orders. 47 U.S.C. §§ 1001-1010 (2006). However, the term "telecommunications carrier" does not include "persons or entities insofar as they are engaged in providing information services." 47 U.S.C. § 1001(8)(C)(i). The law does not require cloud providers to provide real-time interception capabilities. In a statement before the House Judiciary Committee, the FBI and others identified this as a shortcoming. See, *e.g.*, Going Dark: Lawful Electronic Surveillance in the Face of New Technologies: Hearing Before the Subcomm. on Crime, Terrorism, and Homeland Security of the H. Comm. on the Judiciary, 112th Cong. 23-24 (2011) (statement of Susan Landau, Fellow at the Radcliffe Institute for Advanced Study at Harvard University); FBI - Going Dark: Lawful Electronic Surveillance in the Face of New Technologies, Federal Bureau of Investigation (Feb. 17, 2011), http://www.fbi.gov/news/testimony/going-dark-lawful-electronic-surveillance-in-the-face-of-new-technologies (posting the testimony of Valerie Caproni, General Counsel to the Federal Bureau of Investigation, before the Subcommittee on Crime Terrorism, and Homeland Security).

[95] *See*, *e.g.*, Flagg v. City of Detroit, 252 F.R.D. 346, 354 (E.D. Mich. 2008) (the court ruled that text messages held by a provider were subject to the city's control, given that the city had some contractual right of access to the data).

[96] *See* Jansen & Grance, *supra* note 93, at 20-21.

Preservation is an essential tool in electronic discovery, particularly with data that are highly volatile or elastic.[97] For criminal matters, compelling a provider to preserve a snapshot of potential evidence requires a very low bar.[98] For civil matters, the bar for obtaining forensic data is higher and more time-intensive, so civil parties who require such ephemeral data are wise to start the process of acquisition quickly.[99]

If they do not have one already, cloud providers should have some mechanism for preservation. On one hand, providers have an advantage in preserving large data volumes since they advertise broad storage resources.[100] Additionally, IaaS resources such as virtual machines inherently permit providers to take snapshots of the running machines at any time.[101] On the other hand, providers may not be able to prevent their customers from deprovisioning resources or deleting data. Consider the following example, where current cloud practices could inhibit preservation:

[97] Erik Harris, Note, Discovery of Portable Electronic Devices, 61 Ala. L. Rev. 193, 197, n.24 (2009); cf. Cameron G. Shilling, Electronic Discovery: Litigation Crashes into the Digital Age, 22 Lab. L., 207, 227 (2007).

[98] ECS and RCS providers "upon the request of a governmental entity, shall take all necessary steps to preserve records and other evidence in its possession pending the issuance of a court order or other process" for 90 days, which can be renewed for an additional 90 days. 18 U.S.C. § 2703(f) (2006). Section 2704 describes how a governmental entity, in a subpoena or court order, may order the provider to create a backup copy of the contents of the communications. 18 U.S.C. § 2704(a) (2006).

[99] *See* Shilling, *supra* note 97, at 214 (explaining that counsel should discuss E-discovery issues early on in the litigation process); Justin P. Murphy, E-Discovery in Criminal Matters—Emerging Trends & The Influence of Civil Litigation Principles Post-Indictment E-Discovery Jurisprudence, 11 Sedona Conf. J. 257, 259, 262-64 (2010) (pointing out that electronic discovery for criminal matters do not have to follow the much more strict discovery requirements under the Federal Rules of Civil Procedure).

[100] Cf. Viacom Int'l Inc. v. YouTube Inc., 253 F.R.D. 256, 262 (S.D.N.Y. 2008) (ruling that the need for 12 terabytes of data outweighed the expense and burden of production).

[101] *See* Shathabheesha, Virtualization Security in Cloud Computing, INFOSEC Institute (June 21, 2012), http://resources.infosecinstitute.com/virtualization-security-cloud-computing/ (explaining that anyone with access to the host disk files on a virtual machine can create a snapshot).

> Cloud resources, such as virtual machines, are launched using a user's private key. A hacker steals a key from a legitimate user, launches hundreds of machines that flood a popular website, and takes it offline. The opposing party may request data from the legitimate user, seeking activity logs to show who launched the machines, as well as copies of the machines themselves. Nevertheless, the legitimate user may have no logs to produce and the attacker may have tried to cover her tracks by deleting the hundreds of malicious machines.

In traditional digital forensics, investigators would create a mirror image of a hard drive that the examiner can then search for deleted files.[102] Tragically, although cloud providers likely know when files in their storage array are deleted and although they may have logs to prove it, they probably lack the ability to recover deleted files or to produce complete hard disk images.[103]

Because cloud computing is elastic, its corresponding data are often ephemeral.[104] While some courts have noted that ephemeral data "are not discoverable in most cases,"[105] some courts have held that in certain cases, ephemeral data, such as random access memory

[102] *See* Franz J. Vancura, Using Computer Forensics to Enhance the Discovery of Electronically Stored Information, 7 U. St. Thomas L.J. 727, 728-29 (2010).

[103] Microsoft Azure's service level agreement reads "You're responsible for backing up the data that you store on the service . . . . Data that is deleted may be irretrievable." Microsoft Services Agreement, Microsoft, http://windows.microsoft.com/en-US/windows-live/microsoft-service-agreement (last visited Aug. 27, 2012).

[104] *See* Amazon Elastic Compute Cloud Getting Started Guide, Amazon Web Services, 5 (Mar. 11, 2008), http://ec2dream.webs.com/AWS-Management-Console.pdf.

[105] H. James F. Holderman et al., Seventh Circuit Electronic Discovery Pilot Program 14 (2009), available at http://www.ilcd.uscourts.gov/Statement%20-%20Phase%20One.pdf (listing categories of data "not discoverable in most cases," including hard drives" "deleted" or "unallocated" data, RAM, "ephemeral data," temporary files, cache frequently updated metadata, duplicative backup data, and other ESI requiring "extraordinary affirmative measures"); see also Phillips v. Netblue, Inc., No. C-05-4401 SC, 2007 WL 174459, at *2-3 (N.D. Cal. Jan. 22, 2007) (holding a party's argument that hyperlinks should have been preserved was absurd).

(RAM) data, are discoverable.[106] At least one court has affirmed the discoverability of IP addresses.[107] In the cloud, both RAM and IP addresses are potentially fleeting and quickly inaccessible.[108] Although a civil party must preserve evidence when it reasonably anticipates litigation,[109] the Federal Rules of Civil Procedure also relieve parties of the duty to preserve if the data are "lost as a result of the routine, good-faith operation of an electronic information system."[110] At a minimum, cloud providers are more likely to retain data about when resources are provisioned and deprovisioned since those activities directly determine a customer's bill.[111]

Contracts between the cloud provider and customers often detail such issues of ownership.[112] Clear contractual provisions can help to avoid later litigation and expense. Where

[106] E.g., Columbia Pictures, Inc. v. Bunnell, 245 F.R.D. 443, 453 (C.D. Cal. 2007); Victor Stanley, Inc. v. Creative Pipe, Inc., 269 F.R.D. 497, 524 (D. Md. 2010) ("[t]he general duty to preserve may also include deleted data, data in slack spaces, backup tapes, legacy systems, and metadata"); Tener v. Cremer, 931 N.Y.S.2d 552, 555-57 (N.Y. App. Div. 2011) (remanding for determination of several questions, including the data's current availability, custodians, and cost for retrieval).

[107] *See* Columbia Pictures, 245 F.R.D. at 451.

[108] *See generally* Conrad J. Jacoby, E-Discovery Update - Discovery of Ephemeral Digital Information, Law and Technology Resources for Legal Professionals (Jul. 27, 2007), http://www.llrx.com/columns/fios19.htm (explaining how RAM is constantly rewritten and therefore a fleeting storage space).

[109] *See* Zubulake v. UBS Warburg LLC, 220 F.R.D. 212, 216 (S.D.N.Y. 2003).

[110] Fed. R. Civ. P. 37(e).

[111] *See generally* Matthew Wachs et al., Exertion-based billing for cloud storage access, Proceedings for the 3rd USENIX Workshop on Hot Topics in Cloud Computing (Hot Cloud '11), (June 14-15, 2011), available at http://www.pdl.cmu.edu/PDL-FTP/CloudComputing/hotcloud11-final62.pdf (discussing the different rates for charging cloud customers depending on their amount or usage).

[112] Amazon Web Services has such an agreement. See AWS Customer Agreement, *supra* note 71. This contract defines "content" as "software (including machine images), data, text, audio, video, images or other content." *See id.* at § 14. In Section 8.1, Amazon clams "no rights under this Agreement from you or your licensors to Your Content, including any related intellectual property rights." *Id.* at § 8.1. The document defines "Service Offerings" as "the Services (including associated APIs), the AWS Content, the AWS Marks, the AWS Site, and any other product or service provided by us under this Agreement." *Id.* at § 14. In Section 8.4, Amazon clams that "we or our affiliates or licensors and reserve all right, title, and interest in and to the Service Offerings." *Id.* at § 8.4. In other words, the customer explicitly owns their virtual machines, and

contracts do not sufficiently discuss ownership, however, parties must look to case law. For example, the court in *Flagg v. City of Detroit* found that the city had a contractual right to text messages held by a third party provider.[113] The *Flagg* court did not address the ownership of other data, such as the provider's logs.[114] For discovery requests, subpoenas, or search warrants, any requesting party would be wise to determine what data are in what party's possession or custody, whether the provider, the customer, or both.

Determining jurisdiction in cloud-computing environments is unlike any prior jurisdiction analysis. Even more than websites, cloud computing is neither jurisdictional nor multi-jurisdictional. It is non-jurisdictional in that physical geography frequently does not matter. Even for existing cases discussing online data, those cases almost exclusively revolve around websites.[115] Although online services such as Facebook and Gmail frequently comply with discovery requests, those cases rarely, if ever, discuss the nature of the services' back-end geographic location and the locations of the resultant data.[116] In the cloud, the issue compounds since data are likely stored in several jurisdictions and possibly even across international borders among countries with conflicting laws. For example, in one criminal case, the defendant was tried in California because she was accused of violating a social networking site's terms of service and the site's owner was located in California.[117]

does not own the IP address, hardware, or cloud-hosting infrastructure. Microsoft contracts contain similar language. See Microsoft Services Agreement, *supra* note 103 ("Except for material that we license to you, we don't claim ownership of the content you provide on the service. Your content remains your content."). But unlike Amazon's agreement, Microsoft's Service Agreement does not define "content." *See id.*

[113] 252 F.R.D. 346, 354 (E.D. Mich. 2008).

[114] *See generally* id. (discussing the discoverability of text messages).

[115] *See*, *e.g.*, Facebook v. Connectu LLC, No. C 07-01389 RS, 2007 U.S. Dist. LEXIS 61962 *10-22 (N.D. Cal. Aug. 13, 2007) (discussing jurisdiction as relates to the Plaintiff's website).

[116] *See*, *e.g.*, id. at *14-15.

[117] United States v. Drew, 259 F.R.D. 449, 458 (C.D. Cal. 2009).

Courts frequently apply the "effects test" for personal jurisdiction, which is based on "(1) intentional actions (2) expressly aimed at the forum state (3) causing harm, the brunt of which is suffered—and which the defendant knows is likely to be suffered—in the forum state."[118] Under this framework, one would expect most cloud-based litigation to occur in the cloud customer's forum state.[119] The effects test assumes that most often, the crimes, infringements, or torts are committed against the data owners in their forum state, without any intent to cause harm in the forum state of the data.[120]

For crimes involving cloud computing, following Rule 18—that "the government must prosecute an offense in the district where the offense was committed"[121]—is not straightforward. Where the object of the crime is the cloud, a criminal case could potentially be tried in one of four venues: that of the perpetrator, the cloud provider, the cloud customer, or the online data location. Cloud service providers may dictate the venue in their contract, but that may not be criminally binding.[122] Barring a contractually chosen venue, 18 U.S.C. § 3237 allows for criminal offenses committed in one district to "be inquired of and prosecuted in any district in which such offense was begun, continued, or completed."[123] Courts have described the determination of a proper venue "as a substantial contacts rule that takes into

[118] CoreVent Corp. v. Nobel Indus. AB, 11 F.3d 1482, 1486 (9th Cir. 1993) (citing Calder v. Jones, 465 U.S. 783 (1984)).

[119] *See generally* Facebook, 2007 U.S. Dist. LEXIS at *14-15 (explaining how jurisdiction has typically been evaluated by the courts regarding the effects test).

[120] *See*, *e.g.*, Brayton Purcell LLP v. Recordon & Recordon, 606 F.3d 1124, 1128 (9th Cir. 2010).

[121] Fed. R. Crim. P. 18.

[122] *See*, *e.g.*, Cisco Connect Cloud Terms of Service, CiscoConnectCloud, http://ciscoconnectcloud.com/ui/ustatic/termsofservice/1.0.0/termsofservice-en-US.html (last visited August 19, 2012) (providing an example of a contract in which the provider dictates choice of venue).

[123] 18 U.S.C. § 3237 (2006).

account a number of factors—the site of the defendant's acts, the elements and nature of the crime, the locus of the effect of the criminal conduct, and the suitability of each district for accurate fact finding."[124] In cloud-based crimes, none of these factors creates an obvious choice. Any of those four locations could arguably be a proper venue.

Cloud computing and most other web services exist without deference to geographical location.[125] Customers generally have a reasonable expectation of location for their data; they generally believe that if they are using a service provided by a U.S. company, then their data reside in the United States.[126] Looking at their providers' top-level domain names, users may assume that data stored by "www.state.md.us" is located in the United States, while data stored by associated with a web address including "mail.ru" is located stored in Russia.[127] Most service-level agreements for online services do not specify the location where data will be stored.[128] Absent any reason to believe otherwise,[129] customers and end-users will make assumptions about the data's location, of their data as well as and the laws governing it.

[124] United States v. Beddow, 957 F.2d 1330, 1335 (6th Cir. 1992) (quoting United States v. Williams, 788 F.2d 1213, 1215 (6th Cir. 1986)).

[125] *See generally* Mell & Grance, *supra* note 18, at 2 (describing the location independence of resources as an essential characteristic of cloud computing).

[126] *See* Joseph A. Schoorl, Note, Clicking the "Export" Button: Cloud Data Storage and U.S. Dual-Use Export Controls, 80 Geo. Wash. L. Rev. 632, 648 (2012) (stating that cloud users are generally unaware that their data are transferred across national borders).

[127] The Internet Assigned Numbers Authority (IANA) assigns top-level domain names based on the International Organization for Standardization (ISO) 3166–1 alpha-2 country codes. The United States is assigned .us and Russia is assigned .ru. See ICP-1: Internet Domain Name System Structure and Delegation (ccTLD Administration and Delegation), ICANN, http://www.icann.org/en/icp/icp-1.htm (last visited Sept. 20, 2012).

[128] *See*, *e.g.*, CiscoCloudConnect, *supra* note 122.

[129] Amazon Web Services, for example, allows customers to specify the geographic region where data are stored. See AWS Customer Agreement, *supra* note 71.

In criminal cases, several vehicles can compel data from a provider. As with any other data, 18 U.S.C. § 2703 offers prosecutors five mechanisms to obtain certain information from a provider: (1) Subpoena; (2) Subpoena with prior notice to the subscriber or customer; (3) § 2703(d) court order; (4) § 2703(d) court order with prior notice to the subscriber or customer; and (5) Search warrant. [130]

The Department of Justice prefers using "a subpoena or other less intrusive means to obtain evidence from disinterested third parties, unless use of those less intrusive means would substantially jeopardize the availability or usefulness of the materials sought."[131] Losing the availability of data is of paramount concern given the cloud's elasticity. Regardless of the vehicle used, some data may be in the provider's possession, custody, or control, whereas other data may be in the cloud customer's possession, custody, or control. To further complicate matters, the Stored Communications Act (18 U.S.C. § 2701 et seq.) has been interpreted to prohibit a provider from disclosing user content in response to a civil subpoena.[132] This decision on communication and the SCA provides drastically different protections for data storage in an ECS versus a provider of RCS, where 18 U.S.C. § 2703(b) allows a cloud provider, acting as a provider of RCS, to disclose the contents of an account used for remote storage without a warrant and without notifying the customer or subscriber.[133] One scholar, Orin S. Kerr, has suggested that this disparate treatment is unconstitutional.[134]

[130] 18 U.S.C. § 2703(a)-(d) (2006).

[131] *See* DOJ Manual, *supra* note 3, at 111.

[132] *See* Flagg v. City of Detroit, 252 F.R.D. 346, 350 (E.D. Mich. 2008) ("[The Stored Communications Act] lacks any language that explicitly authorizes a service provider to divulge the contents of a communication pursuant to subpoena or court order.").

[133] *See* 18 U.S.C § 2703(b) (2006).

[134] *See* Orin S. Kerr, Applying the Fourth Amendment to the Internet: A General Approach, 62 Stan. L. Rev. 1005, 1029 (2010).

Another issue to consider is time. Rule 45 of the Federal Rules of Civil Procedure does not specify a minimum time period of time within which a responding party must comply with a subpoena.[135] Typically, the issuing party will permit the responding party to comply in ten to thirty days, except where the issuing court's local rules dictate another minimum period for compliance.[136] Given the ease with which cloud data can be either be overwritten or destroyed, as well as providers' lack of evidence preservation mechanisms, the threat of spoliation dramatically increases.[137] One solution is to require faster subpoena compliance.[138] However, the difficulties with this approach are that it would require human intervention at the cloud provider and it does not scale.[139] Another solution is empowering data owners and investigators to gather forensic evidence themselves.[140] This option would

[135] *See generally* Fed. R. Civ. P. 45 (discussing the form in which documents must be produced). Courts, however, "may specify conditions for the discovery." Fed. R. Civ. P. 45(d)(1)(D).

[136] *See* David J. Lender et al., Federal Practice: Responding to a Subpoena, Practical Law Company, 7 (2010), available at http://www.weil.com/files/Publication/925ba5e1-3ebb-4758-8e83-a1424fdff940/Presentation/PublicationAttachment/e8247337-b86d-4df9-b01b-a953f20b0545/10.18.10-Federal%20Practice%20Responding%20To%20A%20Subpoena%20(1-503-1741)%20(2)%20(2).pdf.

[137] *See* The Sedona Conference, The Sedona Conference Commentary on Cloud Computing (Draft) 28-31 (The Sedona Conf. Working Paper Grp. 1, 2011).

[138] But see Erin E. Rhinehart, Civil Subpoenas in Federal Court: Complying with Third-Party Subpoenas, American Bar Association (2012), http://apps.americanbar.org/litigation/committees/pretrial/articles/0923_civil-subpoenas-2.html (discussing what is considered a "reasonable time to comply" under the current version of Rule 45). The Rule 45 "reasonable time" requirement may prove antagonist to the goal of faster subpoena compliance.

[139] *See* Meera Unnithan Sossamon, Subpoenas And Social Networks: Fixing The Stored Communications Act In A Civil Litigation Context, 57 Loy. L. Rev. 619, 642-43 (highlighting the unreasonably high costs and expenses associated with cloud computing providers' subpoena compliance); see also Steven S. Gensler, The Intersection of Facebook and the Law: Symposium Article: Special Rules for Social Media Discovery?, 65 Ark. L. Rev. 7, 35 (2012) (addressing a discussion held by the Discovery Subcommittee over whether a "detailed rule" regarding "when the duty to preserve is triggered and what must be preserved" is necessary, or whether that rule would be too limiting given rapid developments in technology).

[140] This area is being actively explored by one of this article's authors.

shift the burden from the provider (who lacks monetary or legal incentive to quickly comply) to the parties themselves (who have every incentive to collect evidence quickly and inexpensively). FROST, presented in Chapter 5, is one example of this approach.

6.3.2 Fourth Amendment

Search and seizure of evidence regarding crimes committed in or against the cloud should be valid under the Fourth Amendment.[141] This topic has become a focal point of discussion over recent years; scholars have carefully looked at the interplay between privacy and cloud computing.[142]

For simplicity, we will assume that cloud-computing customers have a reasonable expectation of privacy for their data.[143] We also proceed under the current case law applying the Fourth Amendment to online data.[144] Therefore, under *Katz v. United States* and its progeny, obtaining cloud data constitutes a search and violating the reasonable expectation

[141] *See* U.S. Const. amend. IV ("The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized."); see also David J. Goldstone & Daniel B. Reagan, Practice Tips: Social Networking, Mobile Devices, and the Cloud: The Newest Frontiers of Privacy Law, 55 B.B.J. 17, 20-21 (2011) (noting that the "court held that [the defendant] had a reasonable expectation of privacy in his emails stored by the ISP, finding that emails are subject to the same Fourth Amendment protections as letters and phone calls" (citing United States v. Warshak, 631 F.3d 266, 288 (6th Cir. 2010))).

[142] *See*, *e.g.*, Konstantinos K. Stylianou, An Evolutionary Study of Cloud Computing Services Privacy Terms, 27 J. Marshall J. Computer & Info. L. 593, 594-98 (2010); David S. Barnhill, Cloud Computing and Stored Communications: Another Look at Quon v. Arch Wireless, 25 Berkeley Tech. L.J. 621, 638, 642-47 (2010).

[143] *See* Couillard, *supra* note 42, at 2205-06 ("[U]sers expect their information to be treated the same on this virtual cloud as it would be if it were stored on their computer, phone, or iPod.").

[144] *See* R. Bruce Wells, The Fog of Cloud Computing: Fourth Amendment Issues Raised by the Blurring of Online and Offline Content, 12 U. Pa. J. Const. L. 223, 225-29 (2009) (featuring a proposal to protect online data under an entirely new doctrine); see also Couillard, *supra* note 42, at 2205.

to privacy implicates the Fourth Amendment.[145] More difficult are the issues surrounding warrant execution for cloud data.[146]

Warrants for web-based email can specify particular senders, recipients, and timeframes, thereby preventing the unnecessary production of the entire email corpus.[147] In IaaS, the warrant may similarly narrow the search for data by filename, creation time, or author.[148] Recently Kerr criticized the *ex ante* regulation of computer search and seizure.[149] Despite the potential for an unprecedented and overwhelming volume of ESI from cloud crimes, search warrants in these cases have a unique opportunity to address the particularity issue often associated with digital searches. Unfortunately, because cloud providers are often opaque about their infrastructure, it would be impossible or unwise for the warrant to specify the search strategy or approach of execution.[150] With a basic understanding of cloud-computing

[145] Katz v. United States, 389 U.S. 347 (1967); see also Wells, *supra* note 144, at 226-27.

[146] *See generally* Barnhill, *supra* note 142 (discussing reasonable expectation of privacy in the workplace and in data migrating to the cloud).

[147] But see Constantine, *supra* note 36, at 518-520 (discussing the dilemma of determining whether a part of the e-mail should be considered “content” or “non-content”, and the implications for a search warrant based on this distinction).

[148] *See*, *e.g.*, Marlo Arredondo Aff. for Search Warrant 2, Aug. 7, 2008. But see Kerr, *supra* note 15, at 543-48 (discussing the ability for users to alter these characteristics, making certain data nearly impossible to find). Given the nature of digital evidence, this does not overcome the need to scan the container for the evidence. Just as one would leaf through a filing cabinet looking for a particular document, so too must the investigator scour the computer looking for the particular file. Unfortunately, distributed cloud data may require the leafing through many filing cabinets in many warehouses in many locations, where data are co-mingled with other users’ data. *Id.* at 576-77 (including a discussion on the plain view doctrine).

[149] Orin S. Kerr, Ex Ante Regulation of Computer Search and Seizure, 96 Va. L. Rev. 1241, 1246 (2010) (“[Arguing] that ex ante regulation of computer warrants is both constitutionally unauthorized and unwise.”).

[150] *See* Constantine, *supra* note 36, at 501 (“Considering the expansive nature of the terms of Google’s general service agreement and assuming consumers actually read the agreement rather than blindly clicking “agree,” users may wonder what level of privacy their files will have if uploaded or sent through one of Google’s services.”); see also Ari Schwartz et al., Storing Our Lives Online: Expanded Email Storage Raises Complex Policy Issues, 1 J.L. & Pol’y Info. Soc’y 597, 597 (2005) (“[I]t is sometimes hard to determine what a specific provider’s policy is, especially with respect to deletion of mail from inactive accounts or deletion of older mail from active accounts.”). But see Kerr, *supra* note 15, at 565 (“The Framers of the Fourth Amendment included

technology, courts should decline to impose limits as conditions on issuing the issuance of cloud-targeted warrants.

Today, most search warrants for online data are served upon providers, who subsequently execute them.[151] The provider's legal authority for the provider to execute a warrant comes from both statutory and case law.[152] The practical reason is also germane: law enforcement officers have neither the resources nor expertise to execute warrants surrounding cloud computing.[153] This structure is consistent with traditional search warrants. When officers go to an office building looking for evidence, they do not ask the occupants to locate that evidence. They know what they are looking for, so it is more efficient for them to do the search, rather than relying to rely on the occupant who lacks incentive to be thorough. For cloud computing, however, when the cloud provider executes the warrant at the bequest of law enforcement, it may become the government's agent.[154] Cloud providers may also

a particularity requirement to disallow general searches: all warrants must describe ex ante the particular place to be searched and the particular person or thing to be seized.").

[151] *See* United States v. Warshak, 631 F.3d 266, 288 (6th Cir. 2010); Winston Maxwell & Christopher Wolf, A Global Reality: Government Access to Data in the Cloud, Hogan Lovells, 4 (May 23, 2012), available at http://www.hoganlovells.com/files/News/c6edc1e2-d57b-402e-9cab-a7be4e004c59/Presentation/NewsAttachment/a17af284-7d04-4008-b557-5888433b292d/Revised%20Government%20Access%20to%20Cloud%20Data%20Paper%20(18%20July%2012).pdf (last updated July 18, 2012).

[152] *See*, *e.g.*, 18 U.S.C. § 3105 (2006) ("A search warrant may in all cases be served by any of the officers mentioned in its direction or by an officer authorized by law to serve such warrant, but by no other person, except in aid of the officer on his requiring it, he being present and acting in its execution."); 18 U.S.C. § 2703(g) (2006) ("Notwithstanding section 3105 of this title, the presence of an officer shall not be required for service or execution of a search warrant issued in accordance with this chapter requiring disclosure by a provider of electronic communications service or remote computing service of the contents of communications or records or other information pertaining to a subscriber to or customer of such service."); United States v. Bach, 310 F.3d 1063, 1066-67 (8th Cir. 2002) ("The Fourth Amendment does not explicitly require official presence during a warrant's execution, therefore it is not an automatic violation if no officer is present during a search.").

[153] *See* Couillard, *supra* note 42, at 2217.

[154] *See* Lugar v. Edmondson Oil Co., 457 U.S. 922, 953 (1982); People v. McKinnon, 500 P.2d 1097, 1106 (Cal. 1972); People v. Scott, 117 Cal. Rptr. 925, 926 (Cal. Ct. App. 1974). In *United States v. Richardson*,

look for ways to empower customers and law enforcement to acquire forensic data through self-help. This capability is admirable and would free the provider from the burden of doing all the work. It would also be an attractive feature to potential security-minded clients. Regardless of who does the search, whether the provider or law enforcement, this approach raises two new questions, which apply equally to civil litigation: first, where can the search be done, and second, what law applies?

6.3.3 Jurisdictional Difficulties and Implementation Costs

Consider an example that illustrates this problem. Imagine that a cloud provider incorporated in California has a data center in Virginia. A Washington, D.C. court issues a warrant for data residing in the Virginia data center. A New York resident owns the data. If the provider executes the search, it does so from a computer terminal in California. The provider also provides the FBI with access to search remotely from their offices in D.C. We propose that where the search is done (inside the United States) is immaterial and that California law should control. The interconnected, networked nature of a national or global company makes where the search is conducted irrelevant. Even if the provider physically executes the search in California, it still accesses the data remotely, flowing across many interstate networks to the Virginia data center. It follows, however, that the location of the provider (in this example, incorporated and governed by California law) should be the operative jurisdiction, regardless of where the search occurs.

the Fourth Circuit has held that AOL was not acting as an agent for the government when it uncovered and reported child pornography in a customer's email. 607 F.3d 357, 366-67 (4th Cir. 2010). This activity was not done at the government's request, but reported pursuant to an unrelated statute that requires mandatory reporting of suspected violations of child pornography regulations. *Id.* Professor Steven R. Morrison has suggested that ISPs be treated as state actors for any search of user's email. See Steven R. Morrison, What the Cops Can't Do, Internet Service Providers Can: Preserving Privacy in Email Contents, 16 Va. J. L. & Tech. 253, 257 (2011).

Upon execution of a warrant, the cost of cloud-based ESI collection and production could be expensive.[155] The situation is not entirely analogous to the civil case of *Zubulake v. UBS Warburg* ("*Zubulake IV*").[156] In *Zubulake IV*, the majority of the $273,649 production costs stemmed from restoring five offline magnetic tapes and attorney fees.[157] Data stored in the cloud is clearly online and available for access. But the physical act of locating and copying the data may still take considerable time. For example, Amazon offers an export service, which copies and mails customers' data in a storage device.[158] This service costs $80 eighty dollars per storage device handled plus $2.49 per data-loading hour.[159] These costs are unlikely to approach the costs of magnetic tape restoration, but the costs to analyze large data volumes will likely dwarf the data production costs.[160] Importantly, an IaaS cloud provider may be unable to search the corpus of data and produce specific evidence (*e.g.*, a particular file), but rather would have to hand over the whole data set.[161]

[155] *See* David Degnan, Accounting for the Costs of Electronic Discovery, 12 Minn. J. L. Sci. & Tech. 151, 151 (2011).

[156] 216 F.R.D. 280 (S.D.N.Y. 2003).

[157] *Id.* at 289-90.

[158] AWS Import/Export, Amazon Web Services, http://aws.amazon.com/importexport/ (last visited Sept. 4, 2012).

[159] *Id.*

[160] If a cloud customer arbitrarily had two terabytes of data in the cloud, it would take nearly 10 hours to copy to a USB hard drive, totaling $104.90. *Id.* One article estimates forensic analysis averaging $1000 per gigabyte, bringing two terabytes to $2 million. See Degnan, *supra* note 155, at 162.

[161] *See* David Colarusso, Note, Heads in the Cloud, A Coming Storm the Interplay of Cloud Computing, Encryption, and the Fifth Amendment's Protection Against Self-Incrimination, 17 B.U. J. Sci. & Tech. L. 69, 91 (2011).

6.4 Responsive Strategies

In Section 6.2, we discussed the logistics and pitfalls of obtaining cloud data. This section describes defenses and responses that could discredit that evidence. Some issues parallel the scrutiny of any evidence, including the *Daubert* or *Frye* tests.[162] Other issues arise explicitly from the use of cloud technology, such as environment complexity and jury comprehension.

It is worth noting that the law deals in imperfect analogies.[163] This makes explaining the relation between the cloud and the law difficult for all involved. Despite cursory similarities between searching a cloud-based file system and a physical filing cabinet, the injustice served by that analogy should raise doubt about its applicability.

By their nature, cloud-computing environments are more complex than a single computer or a server.[164] Cloud environments have many layers of implementation that must be trusted to produce authentic data.[165] In 2009, for example, researchers demonstrated a working exploit to break out of a virtual machine and attack the host.[166] In a real-word situation, this could have destroyed confidence in the forensic evidence. Courts have repeatedly ruled that merely showing that an action is possible does not prove that it is so.[167] Nevertheless,

[162] *See generally* Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579 (1993); Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

[163] *See* Serena Mayeri, Reconstructing the Race-Sex Analogy, 49 Wm. & Mary L. Rev. 1789, 1837-38 (2008) (discussing the Race-Sex analogy in Regents of University of California v. Bakke, 438 U.S. 265 (1978)).

[164] *See* William R. Denny, Survey of Recent Developments in the Cloud Computing and Software as a Service Agreement, 66 Bus. Law. 237, 237 (2010).

[165] *See supra* Section 6.2.1.

[166] *See* Video Demonstrating Cloudburst Module, Immunity Inc., http://www.immunityinc.com/documentation/cloudburst-vista.html (last visited Aug. 19, 2012).

[167] *See*, *e.g.*, Noblesville Casting Div. of TRW, Inc. v. Prince, 438 N.E.2d 722, 731 (Ind. 1982) (mere possibilities will not suffice to place a fact in issue; "[o]f course, an expert's opinion that something is "possible" or "could have been" may be sufficient to sustain a verdict or award when it has been rendered in conjunction

computer malfunction and malfeasance must be investigated and can cause fact-finders to question the evidence. The hypervisor is especially vulnerable to scrutiny given its powerful position to see and manipulate all virtual machines that it controls, including concomitant data.[168] Many cloud service providers use custom proprietary hypervisors that the global security community has neither seen nor independently audited.[169]

This evidentiary complexity can challenge judges and juries who lack knowledge about cloud computing. Such complex evidentiary analysis might leave the lay juror "spinning with information too strange to digest and often too intimidating to ponder."[170] Much has been written, particularly over the last twenty years, about how juries comprehend complex evidence, including highly scientific evidence such as DNA.[171] Jurors have almost certainly

with other evidence concerning the material factual question to be proved"). The "what if" scenarios for data tampering in the cloud are numerous, a non-comprehensive list of which includes: (1) data could be tampered with in transit over the network; (2) redundant copies of the data could have gotten out of sync; (3) the data owner's credentials could have been compromised, resulting in false data creation or data tampering; (4) there are opportunities for many insider threats at the provider; (5) the hypervisor may be insecure allowing a malicious user to manipulate other virtual machines; (6) the host operating system could be insecure; or (7) there could be weak or no encryption on the provider's internal infrastructure for data in transit or data at rest.

[168] *See* Jansen & Grance, *supra* note 93, at 2 (noting that a hypervisor "is an additional layer of software between an operating system and hardware platform that is used to operate multi-tenant virtual machines and is common to IaaS clouds" and "supports other application programming interfaces to conduct administrative operations, such as launching, migrating, and terminating virtual machine instances," which is vulnerable to compromise because it "causes an increase in the attack surface" via the "additional methods (*e.g.*, application programming interfaces), channels (*e.g.*, sockets), and data items (*e.g.*, input strings) an attacker can use to cause damage to the system").

[169] *See*, *e.g.*, Clive Longbottom, Will Hypervisors need a Supravisor?, VNUnet, 1-2 (2008), available at http://www.quocirca.com/media/articles/042008/220/Will%20Hypervisors%20need%20a%20Supravisor.pdf.

[170] Keith E. Broyles, Taking the Courtroom into the Classroom: A Proposal for Educating the Lay Juror in Complex Litigation Cases, 64 Geo. Wash. L. Rev. 714, 714 (1996); see also Donald E. Shelton, Forensic Science in Court: Challenges in the Twenty First Century 117 (Gregg Barak ed. 2010).

[171] *See*, *e.g.*, Joe S. Cecil et al., Citizen Comprehension of Difficult Issues: Lessons from Civil Jury Trials, 40 Am. U. L. Rev. 727, 728-29 (1991) (citing J. Joseph F. Weis, Jr. et al., Fed. Courts Study Comm., Report of the Federal Courts Study Committee 97 (1990), available at http://www.fjc.gov/public/pdf.nsf/lookup/repfcsc.pdf/$file/repfcsc.pdf (recommending comprehensive examination of how courts handle scientific and technological complexity in litigation)).

used the Internet,[172] but this says nothing about their comprehension of how it or their computer works. Cloud computing is one of today's most complex computing environments and it is likely to challenge even the most technically inclined juror. As such, evidence and expert witness testimony must be presented artfully.

Cloud providers currently execute search warrants and subpoenas for law enforcement and litigants.[173] In this regard, cloud providers act no differently than any other Internet-based entity. But doing so may raise a conflict of interest.[174] Cloud providers are interested in protecting their reputations, so they are not likely disinterested.[175] Furthermore, the provider may have neither the discernment nor the authority to determine what other evidence is relevant, responsive, or in plain view.[176] Lastly, in civil matters, providers lack

[172] *See* Internet Adoption, Pew Internet & Am. Life Project (2012), http://www.pewinternet.org/Static-Pages/Trend-Data-(Adults)/Internet-Adoption.aspx (noting that, as of April 2012, 82% of American adults use the Internet).

[173] *See supra* Section 6.3.2.

[174] *See*, *e.g.*, David D. Cross & Emily Kuwahara, E-Discovery and Cloud Computing: Control of ESI in the Cloud, 1 EDDE J., no. 2, Spring 2010 at 3, available at http://www.crowell.com/documents/e-discovery-and-cloud-computing-control-of-esi-in-the-cloud.pdf (noting that "[w]ith a third-party in possession of data that parties to litigation may view as their own (or a court may view as belonging to them), issues surrounding the duties to preserve and produce become more pronounced."); see also Gruenspecht, *supra* note 15, at 545, 551 (noting that cloud service providers have a "lack of interest in disputing governmental requests," but that, for document creators, "[t]he privacy problem presented is clear: searching [electronic storage] in a comprehensive way can expose both crimes and embarrassing private information that can be admissible in court under the plain view exception") (internal quotation marks omitted).

[175] *See* Achieving Data Privacy in the Cloud, Ponemon Institute LLC 3 (June 2012), available at http://download.microsoft.com/download/F/7/6/F76BCFD7-2E42-4BFB-BD20-A6A1F889435C/Microsoft_Ponemon_Cloud_Privacy_Study_US.pdf. But see Gruenspecht, *supra* note 15, at 550-51 ("A third-party subpoena recipient rarely disputes the request, or even the delay of notice. The problems with subpoenas to cloud computing data service providers go beyond the service providers' lack of interest in disputing governmental requests.").

[176] *See* Gruenspecht, *supra* note 15, at 551 ("[C]loud computing data holders, unlike traditional business records holders, may not be in a position to address the questions of relevance and particularity, since they do not know what information they possess. Even a data holder willing to dispute a subpoena may not have sufficient knowledge to argue against its unreasonableness."). Courts disagree about what constitutes "plain view" in digital evidence. Compare United States v. Williams, 592 F.3d 511, 521-22 (4th Cir. 2010) (holding

the incentive to do thorough and accurate searches, particularly because such searches can be expensive.[177] Because the requesting parties often lack technical knowledge of the cloud providers' systems—and are often physically remote from the providers who execute the searches—those parties' oversight over those searches is often limited or nonexistent.[178] Rigorous guidelines, such as how to challenge the scope and procedure of the search, are currently lacking or absent. Barring these changes, it would be preferable for an independent third party to execute the warrant or subpoena upon a cloud provider.[179] Until the process of how a provider executes a search is well understood, however, the requesting party would be wise to call the technicians to testify about their methodology.[180] As already noted, a party "need not call each of the technicians who did the search so long as it presents a witness who 'can explain and be cross-examined concerning the manner in which the records are made and kept.'"[181]

that evidence viewable on a computer or electronic media may be seized under the plain view exception), with United States v. Mann, 592 F.3d 779, 782, 785-86 (7th Cir. 2010) (holding that evidence uncovered while searching a computer pursuant to a warrant falls within the plain view exception).

[177] *See* Fed. R. Civ. P. 26(b)(2)(B) ("A party need not provide discovery of electronically stored information from sources that the party identifies as not reasonably accessible because of undue burden or cost.").

[178] *See* Cross & Kuwahara, *supra* note 174, at 1 (noting that "cloud computing may dramatically expand the number of places that ESI may reside—and may significantly increase the complexity and difficulty of locating and obtaining that data"). But cf. Fed. R. Civ. P. 26(f); Shira A. Scheindlin & Jonathan M. Redgrave, Special Masters and E-Discovery: The Intersection of Two Recent Revisions to the Federal Rules of Civil Procedure, 30 Cardozo L. Rev. 347, 356 (2008) (noting a Rule 26(f) conference requires that "parties must be prepared to disclose information about their computer systems, including where and for how long information is maintained").

[179] *See* Jerry Archer et al., Security Guidance for Critical Areas of Focus in Cloud Computing V3.0, Cloud Security Alliance, 42-43 (2011), https://cloudsecurityalliance.org/guidance/csaguide.v3.0.pdf (noting that a cloud service provider "might be tempted to reply" to a request for client data by providing a broad range of data to the requestor without questioning the validity of the request).

[180] *See* Fed. R. Evid. 702.

[181] United States v. Cameron, 733 F. Supp. 2d 182, 188 (D. Me. 2010) (citing Wallace Motor Sales, Inc. v. Am. Motors Sales Corp., 780 F.2d 1049, 1061 (1st Cir. 1985)).

The cloud's nebulous nature makes evidentiary admission difficult. The *Daubert*[182] and *Frye*[183] standards are used to measure the scientific validity and relevance of forensic evidence. The *Daubert* factors include determining whether a theory or technique has been tested, whether it has been subject to peer review and publication where there is a known error rate, and whether the theory or technique is generally accepted within the relevant scientific community.[184] Similarly, the *Frye* standard requires that the method "be sufficiently established to have gained general acceptance in the particular field."[185] The Supreme Court in *Daubert* held that Federal Rule of Evidence 702 superseded *Frye* as the applicable standard for admitting expert scientific evidence in federal courts,;[186] but some state courts still follow the "general acceptance" standard articulated in *Frye*.[187]

Because cloud forensics is a new discipline, establishing all of these factors is difficult.[188] Courts have held that popular forensic tools such as EnCase have passed the *Daubert* test in part because of their commercial availability, testing by the government,[189] long-term use,

[182] Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 589 (1993) (holding that although scientific evidence does not have to be generally accepted, any evidence admitted must be both relevant and reliable).

[183] Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923) (finding that experts should be permitted to testify only about scientific principles that are generally accepted in their fields).

[184] Daubert, 509 U.S. at 593-94.

[185] Frye, 293 F. at 1014.

[186] Daubert, 509 U.S. at 589 n.6 ("[W]e hold that Frye has been superseded.").

[187] *See*, *e.g.*, State v. Sercey, 825 So.2d 959, 978 (Fla. Dist. Ct. App. 2002) ("Notwithstanding the U.S. Supreme Court's ruling in *Daubert* that the Federal Evidence Code had superceded [sic] the *Frye* test in federal court proceedings, Florida has continued to adhere to *Frye.")*.

[188] *See* Archer et al., *supra* note 179, at 42 (noting that questions regarding authentication, admissibility, and credibility are not easily resolved by establishing that the information was stored in the cloud).

[189] The National Institute of Standards and Technology's (NIST) Computer Forensic Tool Testing (CFTT) project is charged with testing, measuring the effectiveness of, and certifying digital forensic tools. NIST evaluated EnCase 6.5 in September 2009, but has never evaluated EnCase Enterprise, which includes the

and extensive scientific acceptance.[190] But in the forensic community, techniques for remote forensics, let alone cloud forensics, rarely enjoy any consensus.[191] As we saw in Chapter 4, forensic practitioners who are unfamiliar with cloud environments are often tempted to use their existing tools such as EnCase.[192] Even the advertised features of commercial tools such as EnCase, that can be used for remote forensics, have not been tested for accuracy or error rate, nor have they been tested in court.[193] This software is not unassailable. In 2007, experts analyzed vulnerability was allegedly found in the authentication between the remote EnCase client and the server, allegedly finding vulnerability that could purportedly allow an attacker to corrupt or falsify data.[194]

As one district court noted, "[i]t is the rare case that a litigant does not allege some deficiency in the production of electronically stored information."[195] Producing cloud-based evidence is no different, particularly since that kind of evidence will likely remain novel for years to come.

remote forensic features. See CFTT Project Overview, Nat'l Inst. of Standards and Tech. (Aug. 20, 2003), http://www.cftt.nist.gov/project_overview.htm.

[190] *See* John Patzakis et al., EnCase Legal Journal, Guidance Software, 55-66 (April 2004), http://home.engineering.iastate.edu/~guan/course/CprE-536/paperreadinglist606/LegalJournal.pdfGuidance Software, *supra* note 3, at 59-92 (summarizing trial and appellate court decisions addressing the admissibility of EnCase software).

[191] *Id.* at 1.

[192] *See* Chapter 4.

[193] *See* Archer et al., *supra* note 179, at 97 (explaining that until accepted best practice guidelines are developed, it is unclear whether the analysis results for cloud will stand up in court).

[194] *See* U.S. Computer Emergency Readiness Team, Vulnerability Note VU912593: Guidance EnCase Enterprise uses weak authentication to identify target machines, U.S. Dep't of Homeland Security (Nov. 9, 2007), http://www.kb.cert.org/vuls/id/912593 (last updated Nov. 20, 2007).

[195] Covad Commc'ns. Co. v. Revonet, Inc., 258 F.R.D. 5, 13 (D.D.C. 2009).

Many issues can be raised about the deficiency of production of cloud-based ESI. Such questions may include:

1. Who from the provider executed the search warrant, what were their credentials, and how was the search conducted?
2. Can the technician who executed the search attest to the data's reliability and authenticity, including:
   a.) the security of the workstation used to execute the search,
   b.) the security of the network to prevent data tampering over the network, and
   c.) a record of who had access to the data?
3. Does the provider maintain aggressively enforced records management policies that can provide authenticity and authentication of the data, perhaps in the form of data provenance?
4. Can the provider attest to the reputation and integrity of the cloud infrastructure, including the hypervisor and host operating system?
5. Is it possible that important evidentiary data once existed and has been deleted, and if so, is there any record of it?

As these questions illustrate, the most vulnerable aspects of cloud discovery are expert-witness testimony and the forensic methodology used.[196]

Finally, cases addressing cloud-based evidence are unlikely to produce much definitive judicial guidance because the technology is new and unfamiliar.[197] Cloud computing

[196] *See* Cross & Kuwahara, *supra* note 174, at 5.

[197] *See* Christine Soares, Applying E-Discovery Best Practices to Cloud Computing, Law.com, (Feb. 10, 2012), http://www.law.com/jsp/lawtechnologynews/PubArticleLTN.jsp?id=1202541881944.

technology has evolved over time and continues to change regularly.[198] Adjudicating too narrowly on cloud-specific issues would be premature even though courts can, and do, broadly apply certain established principles (*e.g.*, civil and criminal rules of evidence, Fourth Amendment search and seizure).[199] In fact, Justice Sotomayor's recent concurring opinion discusses potentially changing attitudes about the expectation of privacy in the digital age: "people reveal a great deal of information about themselves to third parties in the course of carrying out mundane tasks."[200]

6.5 Conclusion

Cloud computing is a tremendous advancement in the history of computation, due in large part to technological convergence.[201] The economics of the paradigm will drive companies and individuals to increase growth and adoption rates. Where the people, the data, and the money go, so follows crime and litigation.[202] While investigators and litigators struggle

[198] Amazon Web Services has announced new features or service changes at least one time per month during 2011 and 2012. Amazon Web Services Releases, Amazon, http://phx.corporate-ir.net/phoenix.zhtml?c=176060&p=irol-newsWebservices (last visited Sept. 5, 2012). Other providers have a similar pace of change.

[199] *See* Matthew A. Verga, Cloudburst: What Does Cloud Computing Mean to Lawyers?, 5 J. Legal Tech. Risk Mgmt. 41, 48-49 (2010) (discussing the application of the Federal Rules of Civil Procedure to cases involving cloud computing).

[200] United States v. Jones, 132 S. Ct. 945, 957 (2012) (Sotomayor, J., concurring) ("More fundamentally, it may be necessary to reconsider the premise that an individual has no reasonable expectation of privacy in information voluntarily disclosed to third parties. This approach is ill suited to the digital age, in which people reveal a great deal of information about themselves to third parties in the course of carrying out mundane tasks.") (internal citations omitted).

[201] *See* Araiza, *supra* note 31, at 7-8; Couillard, *supra* note 42, at 2216.

[202] *See* Archer et al., *supra* note 179, at 35; J. Mark Ramseyer, Litigation and Social Capital: Divorces and Traffic Accidents in Japan, in The Harvard John M. Olin Center Faculty Discussion Paper Series, No. 727, at 6 (2012), available at http://www.law.harvard.edu/programs/olin_center/papers/pdf/Ramseyer_727.pdf.

with the emerging problems of acquiring and analyzing cloud data, the law must prepare for evidentiary challenges associated with acquiring and presenting cloud data. The first public cases involving cloud-based ESI are emerging and those involved in those cases have a rare opportunity to develop electronic discovery.[203]

The issues presented here are not wholly unique to cloud computing, and we stress that these issues can also be raised regarding other Internet-derived data, such as social networks and web-based email. Some major important choices must be made, which stand to improve the approach to online data. We have proposed three new ideas. First, online users should have a reasonable expectation of their online data's geographic location.[204] Second, cloud providers should not be permitted to execute search warrants or subpoenas without the introduction of more rigorous operating guidelines.[205] Third, remote forensics should be permitted from anywhere, guided by the laws of the provider's forum.[206] If implemented, these changes will likely provide a stronger foundation to gathering and analyzing cloud-computing evidence in ways that are more robust and defensible.

[203] *See* Verga, *supra* note 199.

[204] *See supra* Section 6.3.1.

[205] *See supra* Section 6.3.2.

[206] *See supra* Section 6.4.

Chapter 7

Search Warrant Language for Cloud Computing

As we saw in Chapter 6, one way that the government seizes electronic data is with a search warrant. This legal instrument requires probable cause and a description of what the government wishes to seize. While the Department of Justice provides sample warrants in the Search and Seizure Manual [86], they do not provide a warrant dealing specifically with cloud environments. The Cloud Security Alliance Legal Information Center sponsored a whitepaper describing other laws and vehicles that the government can use to access cloud data [27].

This chapter explores how cloud-specific details map into a traditional search warrant. While Chapter 6 detailed jurisdictional difficulties, implementation costs, and practical concerns of warrant execution, we now assume that the courts have decided that a warrant is appropriate for cloud data. In Section 7.1 we highlight the basic requirements for a warrant. Section 7.2 explains when a warrant is not required. Section 7.3 presents language for a cloud-based warrant and enumerates examples of data to seize from the cloud provider. In

Section 7.4 build an example warrant using Case Study 1 from Chapter 3. The complete sample warrant can be found in Appendix A and online at http://www.cisa.umbc.edu/warrant. We conclude in Section 7.5.

7.1 Requirements for a Warrant

Before a lawful warrant can be issued, a number of basic requirements must be met. First, the warrant must be approved by a court of law. Second, the Fourth Amendment requires probable cause, in a sworn statement, that the law enforcement officer requesting the warrant believes that the search will reveal criminal activity. Third, the Amendment requires that a warrant "particularly" describe the person, place, or thing to be searched. This third requirement presents an issue for cloud-based crimes.

In previous chapters we examined the location-independent nature of cloud computing. This is clearly counter to the requirements of the Fourth Amendment. In addition to physical location, pinpointing the data to be searched is also problematic. In recent years, warrants for web-based email could specify a particular sender, receiver, and timeframe, preventing the unnecessary production of the entire corpus of email. In IaaS, the warrant may equally narrow the search for data by filename, creation time, or author. Lawyers like to use an analogy between hard drives and filing cabinets. Given the nature of digital evidence, electronic searches do not overcome the need to scan the container for the evidence. Just as one would leaf through a filing cabinet looking for a particular document, so too must the investigator interrogate the computer looking for the particular file [42]. Unfortunately, distributed cloud data may require the leafing through many filing cabinets in many warehouses in many locations, where data are co-mingled with other users' data. Despite the potential for an unprecedented and overwhelming volume of ESI from cloud crimes, search warrants in these cases have a unique opportunity to address the particularity issue often associated with digital searches. Unfortunately, since cloud providers are opaque about their infrastructure, it would be impossible for the warrant to specify the search strategy or approach of execution ahead of time. With a basic understanding of cloud computing technology, magistrates should decline to impose certain conditions of issuing cloud-targeted warrants.

7.2 Exceptions to the Warrant Requirement

The Fourth Amendment does not apply, and warrants are not required, in a number of circumstances. Two of these situations—consent and plain view—are considered here. In cloud crimes where a crime has been committed against an innocent data owner, that party can give consent to a search, and a warrant is not required. Someone whose website, hosted in the cloud, has been hacked or whose data has been stolen, for example, is likely to cooperate with law enforcement in the criminal investigation.

Another type of cloud-based crime is that where the party controlling the cloud resources is committing the crime. For example, if some party is distributing child pornography on a cloud-hosted website, it would be immediately apparent to an officer that incriminating evidence is hosted on the site. This situation is known as plain view, and no warrant is required to seize that contraband. Bear in mind, however, that after an officer identifies the contraband in plain view, a warrant based on the plain view evidence discovered is required for subsequent searches.

7.3 Cloud-Specific Language for a Warrant

The first part of a search warrant must describe what is to be seized. The law requires "reasonable particularity" in the description of the evidence, contraband, fruits, or instrumentality of crime that the agents hope to obtain by conducting the search. In cloud computing environments, the "property to be seized" should contain a description of information (such as computer files) rather than physical hardware, regardless of the role of the computer in the offense. By definition, the physical hardware of a cloud provider is not owned by the suspect (unless the provider is the subject). Seizure of physical hardware yields no benefit that data alone cannot provide, and in fact may be disruptive to other cloud clients sharing that hardware. The "property to be seized" described in the warrant should fall into one or more of the categories listed in Federal Rules of Criminal Procedure (FRCrP) Rule 41(b):

1. *"property that constitutes evidence of the commission of a criminal offense"*

   This is a very broad authorization, covering any item that an investigator reasonably believes would reveal information that would aid in the investigation. "Property" has come to include tangible and intangible property. Case law has established that electronic data are also "property" that may be searched and seized.

2. *"contraband, the fruits of crime, or things otherwise criminally possessed"*

   In cloud environments, contraband could take one of the following forms. Contraband, including child pornography, pirated software, and other copyrighted materials, may be kept in cloud storage or inside of cloud virtual machines. When a hacker breaks into a machine hosted in the cloud, that machine could be the fruits of the crime that property acquired as the result of the crime of unauthorized access.

3. *"property designed or intended for use or which is or had been used as a means of committing a criminal offense"*

   Cloud environments could be used as the instrument of a crime in several ways. Cloud storage could be used to transmit child pornography, and cloud-based virtual machines could be used to produce it. A virtual machine could be used for hacking, or used to host websites with illegal content. In each case, the cloud contains property used to commit an offense.

The second step in drafting a warrant is to describe the property's location. The law, rooted in the physical world, is interested in where the property is. The location, which must be noted with reasonable particularity, has historically been a safeguard to citizens that limit the scope of the warrant. Search warrants for online webmail have traditionally specified only the email address as the "place to be searched." "Location" requires special consideration when dealing with online data, especially with cloud computing. Only rarely will data be stored on a single server at the address of the data custodian. In many cases the servers will be dispersed across state or international boundaries. Further, cloud data are often replicated to multiple data centers. This fact seemingly presents a problem when describing the "location to be searched," since the agent or prosecutor may not know where the data containers are.

The search warrant for cloud-based data should not specify a physical address to be searched, lest the search exclude data stored at other physical locations. Instead, the warrant should specify the desired data and the warrant served to the data custodian.

Here is an example of how to describe the location of cloud-based data in some data center owned and controlled by Amazon:

Data, metadata, and account information created, stored, or controlled by Amazon Web Services LLC, 410 Terry Avenue North, Seattle, WA 98109-5210, related to IP address 1.2.3.4 for the time period beginning 12:01 a.m. CST (January 1, 2012) through 12:01 a.m. CST (July 1, 2012).

The terms "data" and "metadata" include all of the foregoing items of evidence in whatever form (such as virtual machines, user-created content, log data, packet captures, intrusion detection alerts, billing records) and by whatever means they may have been created or stored, including any electrical, electronic, or magnetic form (such as volatile and non-volatile information on an electronic or magnetic storage device, including hard disks, backup storage, live memory, as well as printouts and readouts from any storage device), in any physical location controlled by the provider where the data may reside.

The third step in drafting a warrant is to set the parameters for executing the warrant. Federal warrants allow the specification for the time of day during which to execute the warrant, and the date by which to execute the warrant. These are further safeguards to ensure a limited lifetime of the warrant and minimal disruption (*e.g.*, "in the daytime between 6:00 a.m. to 10 p.m.") to the subject of the warrant.

The elasticity and near-instant provisioning and de-provision of data poses a legal challenge in cloud computing. Unless physical machines are seized or virtual machines are turned off, execution of the warrant is unlikely to impact or disrupt the data owner, but in fact risks spoliation if announced. The search warrant can be executed at any time in the day or night, but should be executed as soon as possible to preserve evidence. The traditional response time of 10 days should be shortened as much as possible, within reason of the logistic constraints of the cloud provider.

An affidavit to justify the search and seizure of cloud-based computer data should include, at a minimum, the following sections: (1) definitions of any technical terms used in the affidavit or warrant; (2) a summary of the offense, and, if known, the role that a targeted computer plays in the offense; and (3) an explanation of the agents search strategy.

While agents and prosecutors should resist the urge to pad affidavits with long, boilerplate descriptions of well-known technical phrases, cloud computing is a new discipline and currently requires special attention to defining new terms. As a rule, affidavits should only include the definitions of terms that are likely to be unknown by a generalist judge and are used in the remainder of the affidavit. Figure 7.1 shows a sample definition for "cloud computing" which could be used in the affidavit. This, and several others, are included in the sample search warrant later in the chapter.

Virtual Machine ("VM")
Virtualization is a technique whereby special software, called the hypervisor, can run many virtual (rather than physical) machines. The hardware on the single machine is emulated so that each virtual instance of a computer, called a virtual machine ("VM"), does not require dedicated physical hardware, but each VM believes it has its own hardware. The hypervisor has special access to control all of the virtual guests, but it should also be able to isolate the guests from each other.

Figure 7.1: Definition of "Virtual Machine" for use in a search warrant.

These concepts are embodied in the sample search warrant that follows. The key point to remember is that the seizure should focus on data rather than hardware, and that the data may be distributed across physical locations.

7.4 Case Study Search Warrant

To illustrate the application of the concepts presented so far, we use Case Study 1 from Chapter 3 and construct a sample search warrant for it.

Recall the hypothetical situation:

> *Polly is a criminal who traffics in child pornography. He has set up a service in the cloud to store a large collection of contraband images and video. The website allows users to upload and download this content anonymously. He pays for his cloud services with a pre-paid credit card purchased with cash. Polly encrypts his data in cloud storage, and he reverts his virtual webserver to a clean state daily. Law enforcement is tipped off to the website and wishes both to terminate the service and prosecute the criminal.*

We assume that the scenario took place in Amazon EC2. Since law enforcement would not have cooperation from the yet unidentified criminal, the cloud provider must provide assistance. In this case, we will use a search warrant to seize the data related to the crime.

In Chapter 3 we identified the following list of potentially relevant forensic data:

- Credit card payment information
- Cloud subscriber information
- Cloud provider management plane access logs
- Cloud provider NetFlow logs
- Cloud consumer virtual machine
- Cloud consumer data in cloud storage.

Appendix A contains a full example of an affidavit for a search warrant in the hypothetical case study. The request focuses on data rather than on hardware. For this reason, it is written as an Electronic Communications Privacy Act (ECPA) § 2703(d) warrant. An FRCrP Rule 41 warrant would have been used to seize hardware or imaging disk drives on-site. The affidavit is an academic example and is not legal advice. The warrant should not be used in practice without seeking legal counsel.

Other academics have suggested that search warrants, and in particular ECPA warrants, may be inappropriate for commanding production of cloud-based data. Orton, Alva, and Endicott-Popovsky [67] cited definitions in the Stored Communications Act as grounds for dismissing the application of the SCA to cloud data. These views present an alternative interpretation to ours here, but neither has been settled by the courts.

The warrant is structured as follows. Paragraph 1 establishes the request for cloud data in investigation of the crime. Paragraph 4 details the cloud crime and presents probable cause that the provider has relevant evidence. The technical background in paragraphs 5-12 is specific to cloud computing, using Amazon as the example. They describe how the service works and what data may be available. Paragraphs 13-23 are similar to language found in any request for electronic evidence.

7.5 Conclusion

The legal community is at the threshold of a wave of cloud-based crimes. Our exploration of seizing electronic evidence from cloud computing provides a foundation to forensic investigators and legal professionals as they investigate and prosecute cloud-based crimes.

Because cloud crimes are not yet widespread and public, it is difficult to predict how the legal system will handle them. Public cases could reasonably be predicted in the next one or two years. These proceedings will test the viability of search and seizure of ESI in cloud environments. Successful legal prosecution will rely on continued education of the players involved, legal interpretation by the courts, and technical capabilities of forensic investigators.

Examining the legal process against a concrete case study highlighted the practical implication of the complex considerations for acquiring evidence. However, the case study introduced a context against which to build a search warrant. As a first public example, this language arms law enforcement agents with topics to consider when they draft their first warrant for cloud data. The warrant serves both to educate legal professionals about how to author such a document and to inform technologists about the cloud concepts important to the courts.

Chapter 8

Conclusions

The technical and legal ability to investigate incidents in and against IaaS cloud environments is key to the execution of justice in the modern age. Equipped with the proper tools and techniques, digital forensics for cloud computing can be performed in a manner that is consistent with federal law. We have shown that forensics for cloud computing exacerbates existing forensic challenges and introduces challenges unique to the cloud. We have also produced tools to enable trustworthy forensics of a cloud environment, and guidance for applying the law and seizing cloud data.

8.1 Summary

The economics of the cloud paradigm are driving growth and adoption rates from companies and individuals. Where the people, the data, and the money go, so does crime. While investigators struggle with the new problems of acquiring and analyzing cloud data, the law must prepare for the legal challenges associated with acquiring and presenting cloud data in court. We have explored how to conduct digital forensics examinations in IaaS cloud computing, identifying important technical, trust, and legal issues, and developing new practical forensic tools and techniques.

Examining two concrete case studies highlighted cloud-specific forensic issues and the practical implication of acquiring evidence. We revealed shortcomings of current forensic practices and laws. Our hypothesis that acquisition was the primary technical and legal challenge was confirmed after analyzing the case studies. These analyses demonstrated the need for an evaluation of existing forensic tools at gathering cloud-based data.

We began our evaluation of existing tools by considering the cloud forensic examination itself. The foundation rested in layers of trust for cloud-based evidence. We found that remote forensic acquisition tools, such as Encase and FTK, are able to acquire data at the Guest OS Layer over the Internet from Amazon's cloud; however, requiring trust at that level is unsatisfactory. Four alternative solutions were identified: TPMs, the management plane, forensics-as-a-service, and legal solutions.

Having concluded that the management plane may be desirable, we proceeded to design and implement forensic acquisition at the Host OS Layer. Using OpenStack, we gave users the ability to acquire copies of virtual disks, API access logs, and host firewall logs with FROST. The integrity of these data could be validated independently. These capabilities were user-driven and did not require assistance from the cloud provider. Testing showed that the solution could scale well, and feedback from potential users was positive.

Early in this research, we viewed the challenges of cloud forensics to be primarily technical. Mid-way through the research, a course in cybersecurity law and policy offered me the opportunity to develop the legal issues more deeply. This introduction to cyber-related statutes, recent case law, and outstanding legal questions in technical situations fostered the insights and careful study of how the legal system would apply jurisprudence to the new domain of cloud computing.

Many public cases involving cloud-based ESI are likely to appear soon, and the people involved in those cases have a unique opportunity to set a new legal precedent. When these cases emerge, each players' actions will be shaped by an interpretation of how traditional discovery rules govern the cloud crime. As we saw, applying these rules can be murky and unclear. Preservation, ownership, jurisdiction, and search warrant execution are just some areas with non-trivial challenges.

Examining a concrete case study helped highlight the practical implication of the complex considerations for acquiring evidence. It introduced a context against which to build a search warrant. As a first public example, this language arms law enforcement agents with specific details to include when they draft their first warrant for cloud data. Arming the prosecution also led us to outline some of the areas that defense teams could incorporate into their own strategies.

Now is an exciting time for cloud computing as innovative new product offerings emerge. The legal community is also at the threshold of a wave of cloud-based crimes. Our exploration of seizing electronic evidence from cloud computing provides a foundation to forensic investigators and legal professionals as they investigate and prosecute cloud-based crimes.

8.2 Open Problems and Future Work

Because cloud crimes are not yet widespread and public, it is difficult to predict how the legal system will handle them. Public cases could reasonably be predicted in the next one or two years. These proceedings will test the viability of search and seizure of ESI in cloud environments. Successful legal prosecution will rely on continued education of the players involved, legal interpretation by the courts, and technical capabilities of forensic investigators.

FROST introduced the capability to collect forensic evidence at the host operating system level. As a platform, FROST opens the opportunity for additional forensic extensions and expanded use for real-time monitoring, metrics, and auditing. However, the hypervisor itself presents an unexplored area for research. Hypervisor forensics may be necessary to record transactions taken by the hypervisor, save state information about the hypervisor, collect and produce logs from the hypervisor, and assertions about the integrity of the hypervisor. It also emphasizes the need for transparency from providers so that third parties can validate the integrity of proprietary cloud components including non-public custom hypervisors.

Some cloud environments are, by nature, rapidly changing in the course of normal operation. For example, data or virtual machines may move without human intervention in order to provide increased reliability or availability. New advances in software defined networking (SDN) are likely to make cloud environments more dynamic and malleable even in their networks. Moreover, raw data may become meaningless if data or metadata are changed in the very act of accessing them. This situation means that forensic data may be a snapshot in time that perhaps cannot be replicated easily.

Cloud computing may challenge traditional notions of data authenticity and integrity. As with live memory forensics, every snapshot of the environment could be different, as normal operations continually alter the environment. It is nearly impossible to hash the acquired snapshot and also the original, live environment to compare them before the live environment changes. Data authenticity, therefore, relies on the correctness of the forensic tools. When the forensic tools are executed on the system being analyzed, this raises the possibility that a compromised system could manipulate the forensic data. Alternate roots of trust will be necessary.

The global, distributed nature of cloud computing will require scholars in international jurisdictions to consider how laws in their countries may apply to cloud crimes. Further, ample work remains for establishing how law enforcement will cooperate in cross-boundary cloud investigations.

Visualization is a powerful tool for forensic investigation. In his outlook for the next ten years, Garfinkel explicitly called out data visualization and visual analysis as topics demanding research attention and new approaches [29]. In their cover article "Cyber Forensics in the Cloud," Zimmerman and Glavach [91] predict, "In incidents where acquisition is a challenge, next generation forensic tools must visualize the physical and logical data locations. The visualization must indicate obtainable and unobtainable artifacts, easing the collection burden and preservation estimates." Assuming that the authors are correct that it is possible to use data visualization techniques to show the physical and logical data locations, and to indicate which artifacts are obtainable, research is needed to develop tools and validate them to support the authors' claims.

8.3 Final Thoughts

We have shown that it is possible to conduct digital forensic examinations of IaaS cloud computing environments that are consistent with federal laws through the development of practical forensic and legal tools. The culmination of an analysis of technical issues was FROST, and the culmination of legal analysis was a sample search warrant.

While end users of FROST appreciated the ability to acquire forensic evidence from a remote cloud environment, the key consideration that will make the solution workable and well accepted is that cloud providers can deploy the system on a large scale and offload many of the manpower-intensive investigative tasks of today. This unintentional by-product of our design choices makes it compelling to cloud providers. At the individual cloud user level, users reacted positively to the intuitive interface that they could control.

This dissertation enhances our understanding of technical, trust, and legal issues needed to investigate cloud-based crimes and offers new tools and techniques to facilitate such investigations. We have distilled the broad and complex topic of cloud forensics into manageable and understandable components. It is our hope that this knowledge, and the adoption of our solutions, will help law enforcement and forensic experts solve investigations.

Appendix A

Search Warrant for Cloud Computing

UNITED STATES DISTRICT COURT

FOR THE WESTERN DISTRICT OF WASHINGTON

IN THE MATTER OF THE SEARCH OF INFORMATION ASSOCIATED WITH THE WEBSITE POLLYONLINE.NET THAT IS STORED AT PREMISES CONTROLLED BY AMAZON WEB SERVICES, LLC	Case No.

APPLICATION OF THE UNITED STATES
FOR AN ORDER PURSUANT TO 18 U.S.C.§ 2703(d)

I, JOHN DOE, being first duly sworn, hearby depose and state as follows:

INTRODUCTION AND AGENT BACKGROUND

1. I make this affidavit in support of an application for a search warrant for information associated with certain Amazon Web Services (AWS) accounts and Internet Protocol ("IP") address that are stored at premises owned, maintained, controlled, or operated by Amazon Web Services (the "Company"), LLC, a web services company headquartered at 410 Terry Avenue North, Seattle, Washington, 98109 (the "Premises"), which functions as an electronic communications service provider and remote computing service. The information to be searched is described in the following paragraphs and in Attachment A. This affidavit is made in support of an application for a search warrant under 18 U.S.C. §§ 2703(a), 2703(b)(1)(A) and 2703(c)(1)(A) to require AWS to disclose to the government records and other information in its possession, pertaining to the subscriber or customer operating the web site.

2. I am a Special Agent (SA) with the Federal Bureau of Investigation (FBI), and have been since January 2003. I am currently assigned to the Baltimore Field Office, Cyber Squad. Since joining the FBI, I have been involved in investigations of computer intrusions, intellectual property right violations and Internet fraud. I have also been assigned to investigate Sexual Exploitation of Children (SEOC) violations of federal

law. I have gained experience conducting such investigations through training and everyday work related to these investigations.

3. The facts in this affidavit come from my personal observations, my training and experience, and information obtained from other agents and witnesses. This affidavit is intended to show only that there is sufficient probable cause for the requested warrant and does not set forth all of my knowledge about this matter

PROBABLE CAUSE

4. On April 1, 2012, an anonymous tip was submitted to the FBI's Baltimore Field Office that the website www.pollyonline.net contained and was distributing child pornography, in violation of 18 U.S.C. §§ 2252 and 2252(a). I determined that the IP addresses for the website hosting the material resolved to one assigned to Amazon Web Services. On April 3, 2012, a preservation request was sent to AWS related to this website and its IP address. Accordingly, this application sets forth specific and articulable facts showing that there are reasonable grounds to believe that the materials sought are relevant and material to an ongoing criminal investigation.

TECHNICAL BACKGROUND

5. Based on my training and experience, I use the following technical terms in this Affidavit and Attachments A and B to this Affidavit:

   a.) "Cloud" is a generic term that refers to a network where the physical location and inner workings are abstracted away and unimportant to the usage. "The cloud" was first used to describe telecommunication networks, where the consumer was blissfully unaware of the inner workings of how their telephone conversation was transmitted to the remote end. The term was later used to describe computer networks, and ultimately to describe the Internet specifically. Knowing the physical location of a website is unimportant to using that service. Cloud computing also takes advantage of this definition of cloud, as it is also a service connected to a network, often the Internet. However, cloud computing offers specific services whereby customers rent remote computing resources such as processing power or data storage, and provision those resources themselves.

b.) "Cloud computing" is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (*e.g.*, networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models.

   i. "Infrastructure-as-a-Service" (IaaS) allows a consumer to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (*e.g.*, host firewalls).

   ii. "Platform-as-a-Service" (PaaS) allows a consumer to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.

iii. "Software-as-a-Service" (SaaS) allows a consumer to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser (*e.g.*, web-based email), or a program interface. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.

c.) "Cloud Service Provider" (CSP) is the entity that offers cloud computing services. CSPs offer their customers the ability to use infrastructure, platform, or software as a service. These services may include offerings such as remote storage, virtual machines, or web hosting. Service is billed as a utility based on usage.

CSPs maintain records pertaining to the individuals or companies that have subscriber accounts with it. Those records could include identifying and billing information, account access information in the form of log files, account application information, and other information both in computer data format and in written record format. CSPs reserve and/or maintain computer disk storage space on their computer system for the use of the cloud service subscriber for both temporary and long-term storage of electronic data with other parties and other types of electronic data and files. Such temporary, incidental storage is defined by statute as "electronic storage," and the provider of such a service is an "electronic communications service" provider. A cloud service provider that is available to the public and provides long-term storage services to the public for electronic data and files, is providing a "remote computing service."

CSPs may be able to provide some of the following, depending on the type of services they provide:

> NetFlow Full Packet Captures Firewall and Router Logs Intrusion Detection Logs Virtual Machines Customer Account Registration Customer Billing Information

d.) "Virtual Machine" (VM) is a system where the hardware is virtual rather than physical. Vitalization is a technique whereby special software, called the hypervisor, can run many virtual (rather than physical) machines. The hardware on the single machine is emulated so that each virtual instance of a computer, called a VM, does not require dedicated physical hardware, but each VM believes it has its own hardware. The hypervisor has special access to control all of the virtual guests, but it should also be able to isolate the guests from each other.

e.) "NetFlow Records" are collections of network statistics collected by a service provider about traffic flows. A traffic flow is a sequence of data packets from a source to a destination. NetFlow is collected when it is impractical to collect all of the data packets for a flow. Providers may use these logs for quality control, security, or billing. For any particular network flow, NetFlow can include the source and destination IP addresses, network ports, timestamps, and amount of traffic transferred. A provider may only collect a sample of all possible sessions, and may only store the NetFlow for a short time.

6. Amazon Web Services (AWS) is an IaaS Cloud Service Provider, a subsidiary of Amazon.com, Inc., that does business online at http://aws.amazon.com. AWS allows

its users to establish accounts with the company, and users can use their accounts to purchase the use of a variety of cloud computing resources offered by AWS.

7. AWS requires users to provide basic contact information during the registration process. This information includes the user's full name, contact e-mail address, physical address (including city, state, and zip code), telephone number, credit card information, and billing address. Users must read and agree to the AWS Customer Agreement. The final step in the registration process is identity verification where an automated system at AWS calls the phone number provided with a verification code that must be entered online.

8. AWS users have the ability to store and retrieve data in the Amazon Simple Storage Service (S3). S3 can store an unlimited number of data objects, which may be documents, photos, videos, or other data. Each object is retrieved using a unique, user-specific key. AWS users are billed based on the amount of data stored, and the transfer into and out of the cloud.

9. AWS provides its users the ability to purchase computing resources on the Amazon Elastic Compute Cloud (EC2). EC2 is a virtual computing environment that allows users to create, use, and manage an unlimited number of virtual machines. Each virtual machine is associated with the user that created it. The user has complete freedom to configure and use the VM as they wish, including installing software and services such as a webserver. AWS users are billed based on the type of VM they choose, and the number of hours that the VM is running.

10. AWS stores user-generated data in more than one physical location. They state "Objects are redundantly stored on multiple devices across multiple facilities in an

Amazon S3 Region." (http://d36cz9buwru1tt.cloudfront.net/AWSRiskand ComplianceWhitepaperJanuary2012.pdf). User-generated data are unlikely to be stored at the Premises. However, system administrators, using the software that controls the cloud infrastructure, have the ability to identify the physical and geographic storage location of the disk drives containing the data.

11. Cloud Service Providers such as AWS typically retain information about their users' accounts, such as the types of service utilized, the date and time of when the services were started and stopped, and connection information (such as the Internet Protocol "IP") address from where the request initiated).

12. Therefore, the computers of AWS are likely to contain all the material just described, including user-created content, stored electronic communications, and information concerning subscribers and their use of AWS, such as account access information, transaction information, and account application.

INFORMATION TO BE SEARCHED AND THINGS TO BE SEIZED

13. I anticipate executing this warrant under the Electronic Communications Privacy Act, in particular 18 U.S.C. §§ 2703(a), 2703(b)(1)(A) and 2703(c)(1)(A), by using the warrant to require Amazon Web Services to disclose to the government copies of the records and other information (including the content of communications) particularly described in Section I of Attachment B. Upon receipt of the information described in Section I of Attachment B, government-authorized persons will review that information to locate the items described in Section II of Attachment B.

14. As described above and in Attachment A, this application seeks permission to search and seize records that might be found on the Premises or data centers controlled by

AWS, in whatever form they are found. I submit that for some computers or electronic medium found on the Premises or in data centers controlled by AWS, there is probable cause to believe those records will be stored in that computer or electronic medium, for at least the following reasons:

f.) Based on my knowledge and experience, I know that Cloud Service Providers bill customers based on the usage of services, and that current and historical billing records are likely to be kept for resources currently being used.

g.) I know that Cloud Service Providers have a tremendous amount of storage capacity, and that this storage is distributed across physical storage media (i.e., hard drives) in multiple data centers in multiple geographic locations. I also know that software keeps track of how data is stored in this environment, and that it has the ability to identify the physical location of any piece of data and reconstruct the pieces into their original format.

h.) I know from training and experience that child pornographers generally prefer to store images of child pornography in electronic form as computer files. The computer's ability to store images in digital form makes a computer an ideal repository for pornography. Even a small portable disk or computer hard drive can contain many child pornography images. The images can be easily sent to or received from other computer users over the Internet. Further, both individual files of child pornography and the disks that contain the files can be mislabeled or hidden to evade detection. In my training and experience, individuals who view child pornography typically maintain their collections for many years and keep and collect items containing child pornography over long periods of time; in fact, they rarely, if ever, dispose of their sexually explicit materials.

15. In this case, the warrant application requests permission to search and seize images of child pornography, including those that may be stored on a virtual machine. These things constitute both evidence of crime and contraband.

16. I know that when an individual uses a website to distribute child pornography over the Internet, the web server will generally serve both as an instrumentality for committing the crime, and also as a storage device for evidence of the crime. The computer is an instrumentality of the crime because it is used as a means of committing the criminal offense. The computer is also likely to be a storage device for evidence of crime. From my training and experience, I believe that a computer used to commit a crime of this type may contain: data that is evidence of how the computer was used; data that was sent or received; notes as to how the criminal conduct was achieved; records of Internet discussions about the crime; and other records that indicate the nature of the offense.

17. Because several people share the Premises as customers of the cloud service, it is possible that the Premises will contain data that are predominantly used, and perhaps owned, by persons who are not suspected of a crime. If agents conducting the search nonetheless determine that it is possible that the things described in this warrant could be found with those intermingled data, this application seeks permission to seize that data as well.

18. Based upon my knowledge, training and experience, I know that searching for information stored in cloud providers may result in a large amount of electronic storage to be searched later by a qualified computer expert in a laboratory or other controlled environment. This is often necessary to ensure the accuracy and completeness of

such data, and to prevent the loss of the data either from accidental or intentional destruction. This is true because of the following:

i.) The volume of evidence. Computer storage devices (such as hard disks) can store the equivalent of millions of pages of information. Cloud computing offers a vast amount of storage for very little cost. Additionally, a suspect may try to conceal criminal evidence; he or she might encrypt the data or store it in random order with deceptive file names. This may require searching authorities to peruse all the stored data to determine which particular files are evidence or instrumentalities of crime. This sorting process can take weeks or months, depending on the volume of data stored.

j.) Technical requirements. Searching computer systems for criminal evidence sometimes requires highly technical processes requiring expert skill and properly controlled environment. The vast array of computer hardware and software, and non-traditional data formats used to support a cloud environment requires even computer experts to specialize in some systems and applications, so it is difficult to know before a search which expert is qualified to analyze the system and its data. In any event, however, data search processes are exacting scientific procedures designed to protect the integrity of the evidence and to recover even ”hidden,” erased, compressed, password-protected, or encrypted files. Because computer evidence is vulnerable to inadvertent or intentional modification or destruction (both from external sources, destructive code embedded in the system, or malicious insiders, a controlled environment may be necessary to complete an accurate analysis.

19. The information requested should be readily accessible to Amazon Web Services by computer search, and its production should not prove to be burdensome.

20. I anticipate executing this warrant under the Electronic Communications Privacy Act, in particular 18 U.S.C. §§ 2703(a), 2703(b)(1)(A) and 2703(c)(1)(A), by using the warrant to require AWS to disclose to the government copies of the records and other information (including the content of communications) particularly described in Section I of Attachment B. Upon receipt of the information described in Section I of Attachment B, government-authorized persons will review that information to locate the items described in Section II of Attachment B.

CONCLUSION

21. Based on my training and experience, and the facts as set forth in this affidavit there is probable cause to believe that on the computer systems in the control of Amazon Web Services there exists evidence of a crime, contraband, and fruits of a crime. Accordingly, a search warrant is requested.

22. This Court has jurisdiction to issue the requested warrant because it is "a court of competent jurisdiction" as defined by 18 U.S.C. § 2711, 18 U.S.C. §§ 2703(a), (b)(1)(A) and (c)(1)(A). Specifically, the Court is "a district court of the United States… that – has jurisdiction over the offense being investigated." 18 U.S.C. § 2711(3)(A)(i).

23. Pursuant to l8 U.S.C. § 2703(g), the presence of a law enforcement officer is not required for the service or execution of this warrant.

REQUEST FOR SEALING

It is respectfully requested that this Court issue an order sealing, until further order of the Court, all papers submitted in support of this application, including the application and search warrant. I believe that sealing this document is necessary because the items and information to be seized are relevant to an ongoing investigation into the criminal organizations, as not all of the targets of this investigation will be searched at this time. Based upon my training and experience, I have learned that online criminals actively search for criminal affidavits and search warrants via the Internet and disseminate them to other online criminals as they deem appropriate, i.e., post them publicly online through the carding forums. Premature disclosure of the contents of this affidavit and related documents may have a significant and negative impact on the continuing investigation and may severely jeopardize its effectiveness.

Respectfully submitted,

JOHN DOE

Special Agent

Federal Bureau of Investigation

Subscribed and sworn to before me on __________:

____

UNITED STATES MAGISTRATE JUDGE

ATTACHMENT A

Property to Be Searched

This warrant applies to information associated with the website www.pollyonline.net resolving to IP address 23.20.70.250 that is hosted at premises owned, maintained, controlled, or operated by Amazon Web Services, a company headquartered at 410 Terry Avenue North, Seattle, Washington, 98109.

ATTACHMENT B

Property to Be Searched

I. Information to be disclosed by Amazon Web Services

To the extent that the information described in Attachment A is within the possession, custody, or control of AWS, AWS is required to disclose the following information to the government for the IP address listed in Attachment A:

(a) All contact information, including full name, user identification number, birth date, contact e-mail addresses, physical address (including city, state, and zip code), telephone numbers, screen names, websites, and other personal identifiers of the user or users of services associated with the IP address;

(b) IP logs, including all records of the IP addresses that logged into the accounts associated with the IP address;

(c) Firewall, router, and intrusion detection logs associated with the IP address;

(d) The length of service (including start date), the types of service utilized by the user or users associated with the IP address, and the means and source of any payments associated with the service (including any credit card or bank account number).

II. Information to be seized by the government

All information described above in Section I that constitutes fruits, evidence and instrumentalities of violations of 18 U.S.C. §§ 2252 and 2252(a) involving www.pollyonline.net from April 1, 2012 to April 30, 2012, including information pertaining to the following matters:

(a) The virtual machine assigned to the IP address in question on April 1, 2012;

(b) A list of other IP addresses assigned to the virtual machine in question, and the dates and times they were assigned;

(c) Packet captures of traffic to and from the virtual machine in question;

(d) Data stored in any other cloud service, including S3 and DynamoDB, associated with the account running the virtual machine;

(e) Records relating to who created, used, or communicated with the website.

Bibliography

[1] Apache Hadoop. Available at http://hadoop.apache.org/. Last accessed Novmber 11, 2011.

[2] Katz v. United States. 389 U.S. 347 (1967).

[3] Lorraine v. Markel American Insurance Company. 241 F.R.D 534 (D.Md. May 4, 2007).

[4] ACCESSDATA. FTK Performance Testing. Available at http://www.accessdata.com/downloads/media/FTKPerformanceTesting.pdf, 2010. Last accessed December 10, 2010.

[5] AMAZON WEB SERVICES. Amazon Web Services: Overview of Security Processes. Available at http://awsmedia.s3.amazonaws.com/pdf/AWS_Security_Whitepaper.pdf, 2011. Last accessed October 28, 2012.

[6] AMAZON WEB SERVICES. AWS Import/Export. Available at http://aws.amazon.com/importexport/, 2011. Last accessed December 28, 2011.

[7] ANTHES, G. Security in the cloud. *Communications of the ACM 53* (November 2010), 16–18.

[8] BAGH, C. Amazon EC2 helps researcher to crack Wi-Fi password in 20 minutes. Available at http://www.ibtimes.com/articles/100314/20110112/amazon-ec2-password-wi-fi-hacking-cracking-brute-force-attack-wpa-ht, 2011. Last accessed January 12, 2011.

[9] BUYYA, R., AND BUBENDORFER, K. *Market Oriented Grid and Utility Computing*. Wiley Press, New York, NY, USA, 2008.

[10] CASEY, E. *Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet. 2nd ed.* Amsterdam: Elsevier Academic Press, 2004.

[11] CLARKE, D. E. *Towards Constant Bandwidth Overhead Integrity Checking of Untrusted Data*. PhD thesis, MIT, 2005.

[12] CONOVER, M., AND CHIUEH, T. Code Injection From the Hypervisor: Removing the need for in-guest agents. In *Proceedings of Blackhat USA* (2008). Last accessed November 1, 2011.

[13] CROSBY, S. A. *Efficient Tamper-Evident Data Structures for Untrusted Servers*. PhD thesis, Rice University, 2009.

[14] DEAN, J., AND GHEMAWAT, S. Mapreduce: simplified data processing on large clusters. *Communications of the ACM 51* (January 2008), 107–113.

[15] DYKSTRA, J. Seizing electronic evidence from cloud computing. In *Cybercrime and Cloud Forensics*, K. Ruan, Ed. IGI Global, Hershey, PA, 2012.

[16] DYKSTRA, J. [Survey of digital forensics and the law]. Unpublished raw data, 2012.

[17] DYKSTRA, J., AND RIEHL, D. Forensic Collection of Electronic Evidence from Infrastructure-As-A-Service Cloud Computing. *Richmond Journal of Law and Technology 19* (2012). Available at http://jolt.richmond.edu/?p=463.

[18] DYKSTRA, J., AND SHERMAN, A. T. Understanding Issues in Cloud Forensics: Two Hypothetical Case Studies. In *Proceedings of the 2011 ADFSL Conference on Digital Forensics Security and Law* (2011), ASDFL, pp. 191–206.

[19] DYKSTRA, J., AND SHERMAN, A. T. Understanding Issues in Cloud Forensics: Two Hypothetical Case Studies. *Journal of Network Forensics 3*, 1 (2011), 19–31.

[20] DYKSTRA, J., AND SHERMAN, A. T. Acquiring forensic evidence from infrastructure-as-a-service cloud computing: Exploring and evaluating tools, trust, and techniques. In *The Proceedings of the Twelvth Annual DFRWS Conference* (August 2012), vol. 9, pp. S90–S98.

[21] DYKSTRA, J., AND SHERMAN, A. T. Design and Implementation of FROST: Digital Forensic Tools for the OpenStack Cloud Computing Platform. Manuscript submitted for publication, 2013.

[22] EUCALYPTUS. Eucalyptus: The Open Source Cloud Platform. Available at http://open.eucalyptus.com/, 2011. Last accessed November 1, 2011.

[23] FACEBOOK. Help Center: How can I download my information from Facebook? Available at http://www.facebook.com/help/?page=18830, 2011. Last accessed October 10, 2011.

[24] FEDERAL CIO COUNCIL. Guidelines for the Secure Use of Cloud Computing by Federal Departments and Agencies (Draft Version 0.41), 2011.

[25] FIPS PUB 180-1. Secure Hash Standard, SHA-1. Available at http://www.itl.nist.gov/fipspubs/fip180-1.htm, 1995. Last accessed March 24, 2013.

[26] FOSTER, I., ZHAO, Y., RAICU, I., AND LU, S. Cloud computing and grid computing 360-degree compared. In *Grid Computing Environments Workshop, 2008. GCE '08* (Nov. 2008), pp. 1–10.

[27] FRANCOISE GILERT. What Rules Regulate Government Access to Data Held by US Cloud Service Providers. Available at https://downloads.cloudsecurityalliance.org/initiatives/clic/CLIC-Govt-access-to-data-20130221.pdf, 2013. Last accessed March 24, 2013.

[28] GARFINKEL, S. Digital forensics xml and the dfxml toolset. *Digital Investigation 8*, 3–4 (2012), 161–174.

[29] GARFINKEL, S. L. Digital forensics research: The next 10 years. In *The Proceedings of the Tenth Annual DFRWS Conference* (August 2010), vol. 7, pp. S64–73.

[30] GARFINKEL, T., AND ROSENBLUM, M. A virtual machine introspection based architecture for intrusion detection. In *Proceedings of the 10th Annual Symposium on Network and Distributed System Security (NDSS 2003)* (2003), pp. 191–206.

[31] GIOBBI, R., AND MCCORMICK, J. Vulnerability Note VU912593: Guidance EnCase Enterprise uses weak authentication to identify target machines. Available at http://www.kb.cert.org/vuls/id/912593, 2007. Last accessed September 21, 2011.

[32] GOOGLE. Access Logs & Storage Data (experimental) - Google Cloud Storage. Available at https://developers.google.com/storage/docs/accesslogs, 2012. Last accessed October 28, 2012.

[33] GUIDANCE SOFTWARE. Facebook Chat Examinations? Available at http://www.encaseondemand.com/EnCast/EnCastVideos/tabid/1383/ProductID/99/CategoryID/129/List/1/Level/1/Default.aspx, 2009. Last accessed October 10, 2011.

[34] GUIDANCE SOFTWARE. EnCase Legal Journal. Available at http://www.guidancesoftware.com/DocumentRegistration.aspx?did=1000017380, 2011. Last accessed September 21, 2011.

[35] HABER, S., HATANO, Y., HONDA, Y., HORNE, W., MIYAZAKI, K., SANDER, T., TEZOKUY, S., AND YAO, D. Efficient signature schemes supporting redaction, pseudonymization, and data deidentification. In *Proceedings of the ACM Symposium on Information, Computer & Communication Security (ASIACCS'08)* (March 2008), pp. 353–362.

[36] HARRIS, E. C. *Principles of Archaeological Stratigraphy, 2nd Edition*. Academic Press, 1989.

[37] HEISER, J. Remote forensics software. Gartner RAS Core Research Note G00171898, 2009.

[38] HORN, C. V. Chris Coleman documents and search warrants. Available at http://www.examiner.com/article/chris-coleman-documents-and-search-warrants, 2009. Last accessed July 4, 2012.

[39] KAUFMAN, L. Data security in the world of cloud computing. *Security Privacy, IEEE 7*, 4 (July-Aug. 2009), 61–64.

[40] KAUFMAN, L. Can public-cloud security meet its unique challenges? *Security Privacy, IEEE 8*, 4 (July-Aug. 2010), 55–57.

[41] KENT, K., CHEVALIER, S., GRANCE, T., AND DANG, H. Guide to Integrating Forensic Techniques into Incident Response. Available at `http://csrc.nist.gov/publications/nistpubs/800-86/SP800-86.pdf`, 2006. Last accessed September 8, 2010.

[42] KERR, O. S. Search Warrants in an Era of Digital Evidence. *Mississippi Law Journal 75* (2005), 85–135.

[43] KERR, O. S. Applying the Fourth Amendment to the Internet: A General Approach. *Stanford Law Review 62* (2010), 1005–1050.

[44] KRAUTHEIM, F. J. *Building Trust into Utility Cloud Computing*. PhD thesis, Department of Electrical Engineering and Computer Science, University of Maryland, Baltimore County, Baltimore, Maryland, 2010.

[45] KRAUTHEIM, F. J., PHATAK, D. S., AND SHERMAN, A. T. Trusted Virtual Environment Module: Managing Trust in Cloud Computing. In *3rd International Conference on Trust and Trustworthy Computing* (2010), pp. 211–227.

[46] KUNDRA, V. Federal Cloud Computing Strategy. Available at `http://www.cio.gov/documents/federal-cloud-computing-strategy.pdf`, 2011. Last accessed October 11, 2011.

[47] KUNDU, A. *Data in the Cloud: Authentication without Leaking*. PhD thesis, Purdue University, 2010.

[48] LEMOS, R. Cloud-Based Denial Of Service Attacks Looming, Researchers Say. Available at http://www.darkreading.com/smb-security/167901073/security/perimeter-security/226500300/index.html, 2010. Last accessed August 4, 2010.

[49] LIAND, J., KROHN, M., MAZIÈRES, D., AND SHASHA, D. Secure Untrusted Data Repository (SUNDR). In *Proceedings of the 6th conference on Symposium on Opearting Systems Design & Implementation - Volume 6* (Berkeley, CA, USA, 2004), OSDI'04, pp. 121–136.

[50] LIBVMI. Virtual Machine Introspection (VMI) Tools. Available at http://vmitools.sandia.gov/, 2011. Last accessed November 1, 2011.

[51] LILLARD, T. V. *Digital Forensics for Network, Internet and Cloud Computing: A Forensic Evidence Guide for Moving Targets and Data*. Syngress, 2010.

[52] LU, R., LIN, X., LIANG, X., AND SHEN, X. S. Secure provenance: the essential of bread and butter of data forensics in cloud computing. In *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security (ASIACCS '10)* (New York, NY, USA, 2010), ACM, pp. 282–292.

[53] MARK LAUBACH. Minutes of the IP Over Asynchronous Transfer Mode Working Group. Available at ftp://ftp.isi.edu/ietf/ipatm/atm-minutes-93jul.txt, 1993. Last accessed March 24, 2013.

[54] MARTY, R. Cloud application logging for forensics. In *Proceedings of the 2011 ACM Symposium on Applied Computing* (New York, NY, USA, 2011), SAC '11, ACM, pp. 178–184.

[55] MERKLE, R. C. A digital signature based on a conventional encryption function. In *A Conference on the Theory and Applications of Cryptographic Techniques on Advances in Cryptology* (London, UK, 1988), CRYPTO '87, Springer-Verlag, pp. 369–378.

[56] MICROSOFT. About Storage Analytics Logging. Available at http://msdn.microsoft.com/en-us/library/windowsazure/hh343262.aspx, 2012. Last accessed November 12, 2012.

[57] MORGESTER, R. Search Warrant Language for Cellular Phones. Available at http://www.olemiss.edu/depts/ncjrl/pdf/May-June%202006%20Final%20Copy.pdf, 2006. Last accessed July 4, 2012.

[58] NATIONAL INSTITUTE OF JUSTICE. Digital Evidence in the Courtroom: A Guide for Law Enforcement & Prosecutors. Available at http://ncjrs.gov/pd_les1/nij/211314.pdf, 2007. Last accessed September 7, 2010.

[59] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. Computer forensic tool testing (CFTT) project overview. Available at http://www.cftt.nist.gov/project_overview.htm, 2003. Last accessed September 21, 2011.

[60] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. Digital Data Acquisition Tool Specification. Available at http://www.cftt.nist.gov/Pub-Draft-1-DDA-Require.pdf, 2004. Last accessed September 21, 2011.

[61] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. Test Results for Digital Data Acquisition Tool: FTK Imager 2.5.3.14. Available at http://www.ncjrs.gov/pdffiles1/nij/222982.pdf, 2008. Last accessed September 21, 2011.

[62] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. Test Results for Digital Data Acquisition Tool: EnCase 6.5. Available at http://www.ncjrs.gov/pdffiles1/nij/228226.pdf, 2009. Last accessed September 21, 2011.

[63] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. The NIST Definition of Cloud Computing. Available at http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf, 2011. Last accessed January 8, 2012.

[64] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. NIST Cloud Computing Forensic Science Working Group. Available at http://collaborate.nist.gov/twiki-cloud-computing/bin/view/CloudComputing/CloudForensics, 2013. Last accessed March 6, 2013.

[65] OPENSTACK. Keynote Recap, Day 2: Why We Do What We Do. Available at http://www.openstack.org/blog/2012/10/keynote-recap-day-2-why-we-do-what-we-do/, 2012. Last accessed October 26, 2012.

[66] OPENSTACK. OpenStack Open Source Cloud Computing Software. Available at http://www.openstack.org/, 2012. Last accessed December 13, 2012.

[67] ORTON, I., ALVA, A., AND ENDICOTT-POPOVSKY, B. Legal process and requirements for cloud forensic investigations. In *Cybercrime and Cloud Forensics*, K. Ruan, Ed. IGI Global, Hershey, PA, 2012.

[68] RISTENPART, T., TROMER, E., SHACHAM, H., AND SAVAGE, S. Hey, you, get off my cloud: Exploring information leakage in third-party compute clouds. In *Proceedings of the 16th ACM Conference on Computer and Communications Security (CCS '09)* (New York, NY, USA, 2009), ACM, pp. 199–212.

[69] RUAN, K., CARTHY, J., KECHADI, T., AND CROSBIE, M. Cloud forensics: An overview. In *Advances in Digital Forensics VII* (2011).

[70] SANTANA, M. Cloud Security: Beyond the Buzz. Available at http://www.linuxworldexpo.com/storage/10/documents/CI7%20Mario%20Santana.pdf, 2009. Last accessed September 21, 2011.

[71] SANTOS, N., GUMMADI, K., AND RODRIGUES, R. Towards trusted cloud computing. In *Proceedings of USENIX HotCloud* (2009), pp. 3–3.

[72] SATO, H., KANAI, A., AND TANIMOTO, S. A Cloud Trust Model in a Security Aware Cloud. In *Proceedings of the 2010 10th IEEE/IPSJ International Symposium on Applications and the Internet* (2010), SAINT '10, pp. 121–124.

[73] SCHNEIER, B. *Applied Cryptography (2nd ed.)*. John Wiley & Sons, Inc., New York, NY, USA, 1995.

[74] SCHWERHA, J. J., AND INCH, S. Remote Forensics May Bring the Next Sea Change in E-discovery: Are All Networked Computers Now Readily Accessible Under the Revised Federal Rules of Civil Procedure? *Journal of Digital Forensics, Security and Law 3* (2008), 5–28.

[75] SCIENTIFIC WORKING GROUP ON DIGITAL EVIDENCE (SWGDE). Data Integrity Within Computer Forensics. Available at https://www.swgde.org/documents/Current%20Documents/2006-04-12%20SWGDE%20Data%20Integrity%20Within%20Computer%20Forensics%20v1.0, 2006. Last accessed September 16, 2012.

[76] SCMAGAZINE. Best computer forensic tool. *SCMagazine* (2011). Last accessed November 1, 2011.

[77] SELAMAT, S. R., YUSOF, R., AND SAHIB, S. Mapping process of digital forensic investigation framework. *IJCSNS International Journal of Computer Science and Network Security 8*, 10 (2008).

[78] SHIELDS, C., FRIEDER, O., AND MALOO, M. A system for the proactive, continuous, and efcient collection of digital forensic evidence. In *The Proceedings of the Eleventh Annual DFRWS Conference* (August 2011), vol. 8, pp. S3–13.

[79] SYMANTEC. The Trojan.Hydraq Incident: Analysis of the Aurora 0-Day Exploit. Available at http://www.symantec.com/connect/blogs/trojanhydraq-incident-analysis-aurora-0-day-exploit, 2011. Last accessed January 21, 2011.

[80] TAYLOR, M., HAGGERTY, J., GRESTY, D., AND LAMB, D. Forensic investigation of cloud computing systems. *Network Security 2011*, 3 (2011), 4–10.

[81] TERREMARK. Secure Information Services. Available at http://www.terremark.com/uploadedFiles/Services/Security_Services/TMRK_SIS_Gatefold2_4pagelayout_Screen.pdf, 2009. Last accessed November 1, 2011.

[82] UNITED STATES CODE. Communications Assistance for Law Enforcement Act (CALEA). 47 USC 1001-1010, 1994.

[83] UNITED STATES GOVERNMENT. Federal Rules of Civil Procedure. Available at http://www.law.cornell.edu/rules/frcp, 2010. Last accessed March 24, 2013.

[84] UNITED STATES GOVERNMENT. Federal Rules of Evidence 901. Available at http://www.uscourts.gov/uscourts/rules/rules-evidence.pdf, 2011. Last accessed March 24, 2013.

[85] UNITED STATES GOVERNMENT. Federal Rules of Criminal Procedure. Available at http://www.law.cornell.edu/rules/frcrmp, 2012. Last accessed March 24, 2013.

[86] US DEPARTMENT OF JUSTICE. Searching and Seizing Computers and Obtaining Electronic Evidence in Criminal Investigations. Available at http://www.justice.gov/criminal/cybercrime/ docs/ssmanual2009.pdf, 2009.

[87] WILLAMETTE WEEK. You Look Like Obama: FBI Seeks Facebook Records for Person of Interest in Mosque Arson. Available at http://www.wweek.com/portland/blog-26890-you_look_like_obama_fbi_seeks_facebook_records_for.html, 2011. Last accessed July 4, 2012.

[88] WOLTHUSEN, S. D. Overcast: Forensic Discovery in Cloud Environments. In *Proceedings of the 2009 Fifth International Conference on IT Security Incident Management and IT Forensics (IMF '09)* (Washington, DC, USA, 2009), IEEE Computer Society, pp. 3–9.

[89] ZHANG, Y., JUELS, A., REITER, M. K., AND RISTENPART, T. Cross-vm side channels and their use to extract private keys. In *Proceedings of the 2012 ACM conference on Computer and communications security* (New York, NY, USA, 2012), CCS '12, ACM, pp. 305–316.

[90] ZHOU, W., SHERR, M., MARCZAK, W. R., ZHANG, Z., TAO, T., LOO, B. T., AND LEE, I. Towards a data-centric view of cloud security. In *Proceedings of the Second International Workshop on Cloud Data Management (CloudDB '10)* (New York, NY, USA, 2010), ACM, pp. 25–32.

[91] ZIMMERMAN, S., AND GLAVACH, D. Cyber Forensics in the Cloud. *IAnewsletter 14* (2011), 4–7.