Dakota State University

Beadle Scholar

Masters Theses & Doctoral Dissertations

Spring 3-2021

IPCFA: A Methodology for Acquiring Forensically-Sound Digital Evidence in the Realm of IAAS Public Cloud Deployments

Hosam Badreldin
Dakota State University

Recommended Citation

Badreldin, Hosam, "IPCFA: A Methodology for Acquiring Forensically-Sound Digital Evidence in the Realm of IAAS Public Cloud Deployments" (2021). *Masters Theses & Doctoral Dissertations*. 367.
https://scholar.dsu.edu/theses/367

DAKOTA STATE
UNIVERSITY

IPCFA: A METHODOLOGY FOR ACQUIRING FORENSICALLY-SOUND DIGITAL EVIDENCE IN THE REALM OF IAAS PUBLIC CLOUD DEPLOYMENTS

A doctoral dissertation submitted to Dakota State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Cyber Operations

March 2021

By

Hosam Badreldin

Dissertation Committee:

Dr. Ashley Podhradsky
Dr. Omar El-Gayar
Dr. Bramwell Brizendine
Dr. Arica Kulm

DAKOTA STATE
UNIVERSITY

DISSERTATION APPROVAL FORM

This dissertation is approved as a credible and independent investigation by a candidate for the Doctor of Philosophy in Cyber Defense degree and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this dissertation does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department or university.

Student Name: Hosam Badreldin

Dissertation Title: IPCFA: A Methodology for Acquiring Forensically-Sound Digital Evidence in the Realm of IaaS Public Cloud Deployments

Dissertation Chair/Co-Chair: Ashley Podhradsky Date: April 30, 2021
Name: Ashley Podhradsky

Dissertation Chair/Co-Chair: Date:
Name:

Committee member: Omar El-Gayar Date: April 30, 2021
Name: Omar El-Gayar

Committee member: B. Date: April 30, 2021
Name: Bramwell Brizendine

Committee member: Arica Kulm Date: April 30, 2021
Name: Arica Kulm

ABSTRACT

Cybercrimes and digital security breaches are on the rise: savvy businesses and organizations of all sizes must ready themselves for the worst. Cloud computing has become the new normal, opening even more doors for cybercriminals to commit crimes that are not easily traceable. The fast pace of technology adoption exceeds the speed by which the cybersecurity community and law enforcement agencies (LEAs) can invent countermeasures to investigate and prosecute such criminals. While presenting defensible digital evidence in courts of law is already complex, it gets more complicated if the crime is tied to public cloud computing, where storage, network, and computing resources are shared and dispersed over multiple geographical areas. Investigating such crimes involves collecting evidence data from the public cloud that is court-sound. Digital evidence court admissibility in the U.S. is governed predominantly by the Federal Rules of Evidence and Federal Rules of Civil Procedures. Evidence authenticity can be challenged by the Daubert test, which evaluates the forensic process that took place to generate the presented evidence.

Existing digital forensics models, methodologies, and processes have not adequately addressed crimes that take place in the public cloud. It was only in late 2020 that the Scientific Working Group on Digital Evidence (SWGDE) published a document that shed light on best practices for collecting evidence from cloud providers. Yet SWGDE's publication does not address the gap between the technology and the legal system when it comes to evidence admissibility. The document is high level with more focus on law enforcement processes such as issuing a subpoena and preservation orders to the cloud provider.

This research proposes IaaS Public Cloud Forensic Acquisition (IPCFA), a methodology to acquire forensic-sound evidence from public cloud IaaS deployments. IPCFA focuses on bridging the gap between the legal and technical sides of evidence authenticity to help produce admissible evidence that can withstand scrutiny in U.S. courts. Grounded in design research science (DSR), the research is rigorously evaluated using two hypothetical scenarios for crimes that take place in the public cloud. The first scenario takes place in AWS and is hypothetically walked-thru. The second scenario is a demonstration of IPCFA's applicability and effectiveness on Azure Cloud. Both cases are evaluated using a rubric built from the federal and civil digital evidence requirements and the international best practices for

digital evidence to show the effectiveness of IPCFA in generating cloud evidence sound enough to be considered admissible in court.

DECLARATION

I hereby certify that this project constitutes my own product, that where the language of others is set forth, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions, or writings of another.

I declare that the project describes original work that has not previously been presented for the award of any other degree of any institution.

Signed,

____

Hosam Badreldin

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Cybercrime in the era of cloud computing has introduced challenges and complications for existing digital forensics science and its existing models, methodologies, and practices (Choo et al., 2017; Cohen, 2013; Grispos, Storer, & Glisson, 2012; NIST Cloud Computing Forensic Science Working Group, 2014; Simou, Kalloniatis, Gritzalis, & Mouratidis, 2016). While theories such as Bayesian and Dempster-Shafer could be adjusted to help with analyzing digital crime artifacts retrieved from cloud environments (Dykstra, 2013a), there is little documented information about formally tested and legally accepted methods or techniques that can be used to retrieve digital forensics data from Infrastructure-as-a-Service (IaaS) public cloud deployments. According to the American Bar Association (ABA), the foundation for digital evidence to be considered admissible in court is *authenticity* and *integrity* (Dickson, 2011). *Authenticity* has traditionally been tied to expert witness testimony. Furthermore, the testimony can be challenged in courtrooms with the Daubert Standard (Daubert test), which requires that expert witnesses show that they have used a published, known, scientific methodology to perform their collection of the evidence. The second requirement is to maintain the *integrity* of the collected data and the generated evidence, and this has been accomplished through maintaining a clear, undoubted, and well-documented chain of custody.

The National Institute of Justice (NIJ, 2007) provides guidelines for first responders on how to acquire and preserve digital evidence and generate court-admissible evidence; however, public cloud data acquisition was not covered in that publication. Other researchers, such as (Alqahtany, Clarke, Furnell, & Reich, 2015), provided a new methodology for forensic data acquisition in IaaS deployments, but it requires architectural changes and support by the underlying Cloud Service Provider (CSP) to be effective. (Dykstra, 2013a) proposed a novel system for cloud forensics acquisition that was tested successfully on private cloud IaaS deployments such as OpenStack. These systems can be very useful in the case of public cloud investigations, but they require acceptance and adoption by the Cloud Services Providers (CSPs). The question remains, however, whether it is possible to acquire indubitable digital evidence from Infrastructure-as-a-Service (IaaS) deployments in public clouds, present it in court, and gain the acceptance and support of the judges and jurors, all without involving CSPs.

This study introduces a practical methodology for collecting forensic data from IaaS public cloud deployments. The methodology is built on the known boundaries of the existing major IaaS providers, such as Amazon, Google, and Microsoft, and uses the Federal Rules of Evidence (FRE) and the Daubert Standard as guidelines. The methodology aims to ensure admissible and authentic evidence-generation that can be defended in court and validated via expert witness testimony. Data integrity and chain of custody are the bases of the proposed method. Hypothetical cases that simulate successful attacks are used to generate data for a sample case studies. The same hypnotical scenarios are investigated following the various available best practice sources for forensic data acquisition such as RFC 3227 (Killalea & Brezinski, 2002) and the Scientific Working Group on Digital Evidence publications (SWGDE, 2018b, 2020). That feedback is then used to adjust the methodology as needed to produce forensic data that successfully satisfies all the evidence rules and withstand scrutiny in court.

Because there is no actual court case used for the validation of this methodology, the forensic data's court-soundness is assessed against a specially designed rubric based on the existing best practices and recommendations for digital forensic acquisition as well as the various electronically stored information (ESI) legal rules and regulations. The scope of the testing and validation of the methodology is limited to the IaaS service model. This research is grounded in design science (DSR) and adheres to all design science guidelines as outlined by Hevner, March, Park, & Ram (2004). The rest of the dissertation is organized as follows:

- Chapter One provides an overview of cloud computing, digital forensics, cloud forensics, forensics models, digital evidence acquisition, and digital evidence court admissibility.
- Chapter Two contains a literature review of current and foundational research including court admissibility and the Federal Rules of Evidence and other relevant standards. Challenges with cloud forensics and IaaS service models and available tools are highlighted, and the research gap is identified.
- Chapter Three addresses the research methodology and a definition of the success criterion for the proposed artifact, IPCFA, is presented. This chapter provides an overview of the validation, experimentation, and evaluation methods for the proposed methodology, as well as other elements related to the selected research methodology.

- Chapter Four details the proposed methodology (IPCFA) and the various proposed phases, as well as how each phase can be carried out.
- Chapter Five covers the actual applications and demonstration of the proposed methodology.
- Chapter Six discusses the proposed artifact effectiveness, outcomes of the demonstration and validation, as well as it articulates the assumptions.
- Chapter Seven reflects upon the limitations of this research and expands on the possible future research to overcome such limitations and extend the validity of this research. Finally, it provides the conclusion of this research.

CHAPTER ONE

BACKGROUND OF THE PROBLEM

Public Cloud Computing

According to the National Institute of Standards and Technology (NIST), cloud computing is a "computing model where a pool of computing resources is shared, rapidly provisioned and released, and the usage of each resource is measured" (Mell & Grance, 2011). Thus, public cloud computing can be defined as a cloud that can be used by the general public and is physically hosted on the cloud service provider's premises. One of the most common public cloud services is the Infrastructure-as-a-Service (IaaS) model. According to Gartner (2019), the revenue of the worldwide public cloud market was forecasted to grow by 6.3% during the year 2020 alone, and the second fastest-growing market will be the IaaS model, driven by the "increasing adoption of cloud-first strategies in organizations" (Gartner, 2019). The IaaS model is closest to the traditional, virtualized datacenter services, where the customer is responsible for all layers above the virtualization system, or the hypervisor, including the operating systems of the virtualized servers. In this service model, the service provider abstracts the storage and network infrastructure resources, which make up the major portion of the architecture paradigm that provides multitenancy in the cloud. An organization often adapts the IaaS model when it is trying to achieve one of the well-known strategic goals of cloud migration, such as hosting-cost reduction or operations effectiveness, on-demand server (computing) deployments, or flexibility and scalability of resources (elasticity). Some of the largest IaaS providers referenced throughout this research are Amazon Cloud (AWS), Google Cloud (GCP), and Microsoft Cloud (Azure).

Information Security in the Public Cloud

Recently, CSPs have successfully emphasized security and compliance as additional strategic goals in explaining why organizations should move to the public cloud. Most CSPs today consider themselves secure because they follow security best practices or guidelines and are certified with known security compliance bodies, such as PCI ("Official PCI Security

Standards Council Site," n.d.), HIPPA (Health Information Privacy, 2015), or ISO/IEC (ISO/IEC 27001 Information Security Management, n.d.). From their published websites, AWS (Compliance Programs--Amazon Web Services (AWS), n.d.), Azure (Compliance in the Trusted Cloud | Microsoft Azure, n.d.), and Google (Cloud Compliance - Regulations & Certifications, n.d.) list the compliance certifications each CSP holds. The list is comprehensive and shows certifications from almost all global compliance bodies as well as government agencies in the different regions of the world.

Moving to the public cloud does not guarantee security by default. The security of organizational data in the public cloud is a shared responsibility, and the particularity of the level of accountability is determined by type of service model chosen. In an IaaS model, the CSP might be responsible for the physical network and servers (computing), storage, and hypervisor security, and the customer might be responsible for the security of the operating systems, user access management, overlay network, traffic flow management, applications, and application data (Shared Responsibility Model Explained, n.d.). The breakdown of security responsibilities is not fully standardized amongst CSPs, especially the tenant or customer aspect (Shared Responsibility in the Cloud - Microsoft Azure, n.d.) (Shared Responsibility Model - Amazon Web Services, n.d.). This shared security responsibility model brings with it a myriad of legal and technical challenges in how cybersecurity incidents are handled in public clouds, specifically in how digital forensics can take place.

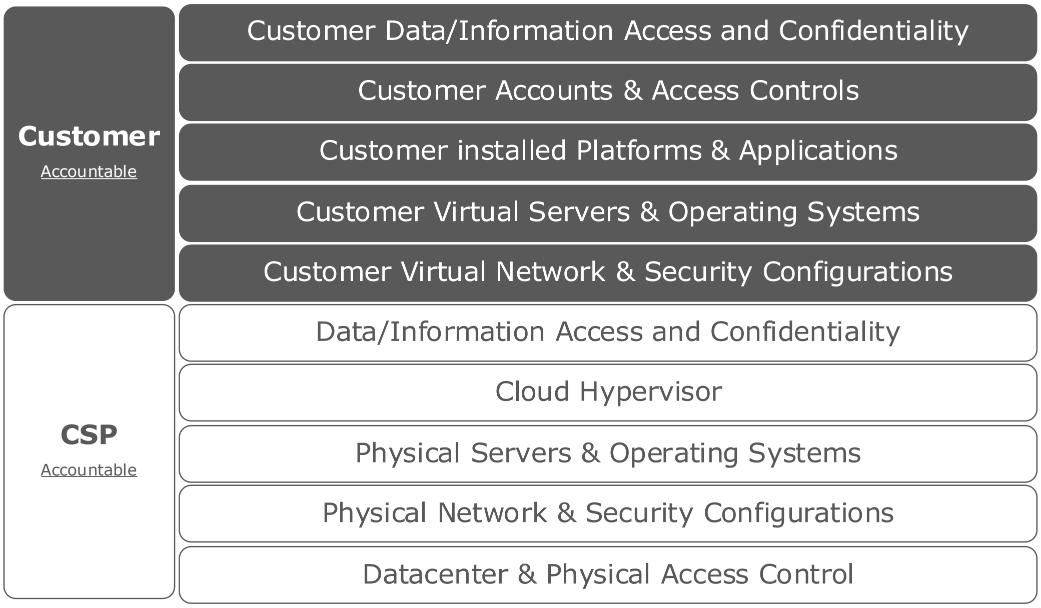

Customer
Accountable
Customer Data/Information Access and Confidentiality
Customer Accounts & Access Controls
Customer installed Platforms & Applications
Customer Virtual Servers & Operating Systems
Customer Virtual Network & Security Configurations
CSP
Accountable
Data/Information Access and Confidentiality
Cloud Hypervisor
Physical Servers & Operating Systems
Physical Network & Security Configurations
Datacenter & Physical Access Control

Figure 1 - The common IaaS shared security accountability model

Digital Forensics

According to the National Institute of Justice, “digital evidence is information stored or transmitted in binary form that may be relied on in court” (NIJ, 2016). Further, NIST defines digital forensics as the practice of collecting and examining digital evidence (NIJ, 2016). NIST SP 800-86 categorizes the digital forensics (DF) model into four high-level phases: Collection (Acquisition), Examination, Analysis, and Reporting (NIST Cloud Computing Forensic Science Working Group, 2014). This model assumes that media is collected and examined, the data analyzed, and the acquired evidence reported. The focus of this dissertation is the collection (acquisition) phase of digital forensics.

An organization wanting to investigate a security breach on their locally hosted servers and possibly generate digital evidence for litigation purposes can follow one of the various published digital forensics processes. Forensic data and evidence acquisition can be performed by the corporate staff or directly by law enforcement agencies with the proper legal authorization. A server can be seized, hard drives can be imaged, and then data can be analyzed. The Chain of Custody (CoC) should be maintained throughout this process. For an organization engaging with this process with data and infrastructure hosted on one of the many public cloud services providers, there are few formally and legally tested guidelines or processes related to obtaining such evidence, aside from trying to convince the CSP to take part in the investigation process. Sometimes, depending on the required evidence, using existing remote forensic acquisition techniques to obtain some of required data is possible.

Digital Forensics - Models

Though NIST in its publication (NIST Cloud Computing Forensic Science Working Group, 2014) has referenced a four-phase digital forensics model, many more models and frameworks also exist dating back to 1999 (McKemmish, 1999). According to (Simou et al., 2016), the digital forensics community proposed more than eighteen models from 1999 to 2016. All of these models share the concept of taking a media or data repository from a criminal case, executing multiple processes on the extracted media, and generating lawful evidence that is scientifically reliable and legally acceptable. Following a preparation or initiation phase for the forensic investigation, in which data related to the case are identified and collected, the collection phase is the starting point. The subsequent phase is the examination, in which the data collected previously are investigated and the most relevant pieces of data extracted. The third stage is the analysis of the extracted data to uncover possible evidence. The last phase is reporting, in which the findings are presented to the relevant entities. Almost all of the proposed models share these four phases, though some use different names or expand these phases into smaller sub-phases. Table 1 compares the phases of eight of the well-known models.

Table 1- The Well-Known Digital Forensic Models and Their Respective Phases

Model	**Proposed Phases**
(McKemmish, 1999)	Identification, preservation, analysis, presentation
(Palmer, 2001)	Identification, preservation, collection, examination, analysis, presentation
(Baryamureeba & Tushabe, 2004)	Readiness, deployment, traceback, dynamite, review
(NIJ, 2008)	Preparation, identification, collection, preservation, packaging, transportation storage, examination, analysis, reporting, documentation
(Agarwal, Gupta, Gupta, & Gupta, 2011)	Preparation, secure scene, survey, document the scene, secure communication channel, collection, preservation, examination, analysis, presentation, results review
(Martini & Choo, 2012)	Evidence source identification and preservation, collection, examination and analysis, reporting and presentation
(Cohen, 2013)	Identification, collection, presentation, transportation, storage, analysis, interpretation, attribution, reconstruction, presentation, destruction
(NIST Cloud Computing Forensic Science Working Group, 2014)	Collection, examination, analysis, presentation
(Zawoad, Hasan, & Skjellum, 2015)	Identification, collection, examination, analysis, presentation, verification, preservation

The abovementioned models and frameworks attempt to address the overall approach of digital forensics and propose various phases to ensure that the extracted evidence is as complete as possible. Few researchers have investigated expanding these phases, proposing processes and methodologies to execute each one properly. This is a significant oversight of our discipline. At the 2006 Digital Forensic Research Conference (DFRWS), Ieong (2006) presented a model called FORZA, which takes a different approach to digital forensics research. He proposed a digital forensics framework that focuses on the various roles involved in an investigation and the creation of a table of roles and responsibilities to help guide the complete life cycle of the digital forensics investigation. By incorporating the business and legal aspects of the investigation, FORZA framework creates processes that can be followed and executed by practitioners that yield trustworthy and defendable results that, in turn, prosecutors can rely upon to draw conclusions.

Adams, Hobbs, and Mann (2013) proposed the advanced data acquisition model (ADAM), which is composed of multiple digital forensic processes that can easily be followed and understood in courtrooms. The authors argue that the "domain of digital forensics is lacking generally accepted processes and procedures to which they and the courts can refer" (Adams et al., 2013, p. 25). While their proposed model has only three phases, they detailed how to execute each phase and provided all required information to successfully carry out the forensic investigation from the beginning to the end. This literature review suggests that no other scholarly papers expand any of the digital forensics models' phases or provide procedural methodologies for the execution of each phase, more specifically for public cloud forensic acquisition.

Digital Forensics - Acquisition

Digital evidence is intrinsically fragile and can be distorted, damaged, forged, or destroyed intentionally or by inappropriate handling, so it must be handled with care. According to the U.S. Department of Justice, digital evidence is always latent and unstable juristically; it can also be time-sensitive (NIJ, 2008). When investigating a cybercriminal activity that involves compute nodes or physical networked servers, there are potentially numerous resources for digital evidence (U.S. Department of Justice; Office of Justice Programs; National Institute of Justice, 2004). Of most concern in this research are the data stored on hard drives (non-volatile) and memory contents (volatile). The process of acquiring evidence from a server can be very complicated and greatly dependent on the specific case. In traditional digital crimes involving a computer or a mobile device, the forensic process can begin by seizing the physical device, removing the hard drive (dead forensic acquisition), connecting it to the examiner-certified computer, and then running the appropriate tools to extract trusted copies of the data.

No single tool can acquire and collect forensic data from all possible venues. The commercial tools most often used today by industry experts and law enforcement agencies (LEAs) are Guidance Software EnCase, Access Data's Forensic Toolkit (FTK), and Magnets AXIOM (Alqahtany et al., 2015). Most of the time when server seizure is not possible for legal or technical reasons, specialized remote acquisition (or remote forensic acquisition) tools can be used to perform data collection remotely. If memory (RAM) imaging is needed (to collect network traffic information, encryption keys, or other volatile data), then a non-invasive process can be followed (live forensic acquisition) to obtain the needed information with as little impact as possible on the investigated system. Whether the forensic examiner decides to perform live or dead acquisition, defining an order of volatility (OOV) is critical to a fruitful and non-erroneous forensic data collection. Figure 2 represents the SWGDE-recommended order of volatility when it comes to acquiring live computer systems. The recommendation is to acquire all data in the RAM; then dump a copy of all running processes, which can partially be collected from the RAM image; get the list of open network connections; capture systems settings and configurations; and finally acquire the non-volatile data stored in the disk media (SWGDE, 2014).

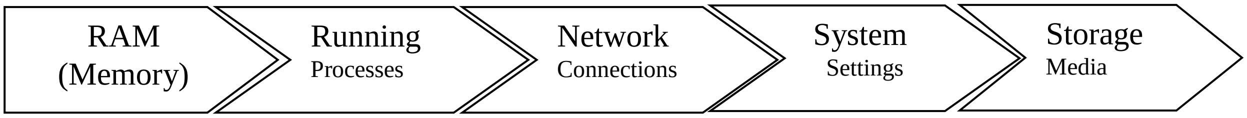


Figure 2 - Sample computer volatility order (SWGDE, 2014)

According to NIST, “identification, collection, and preservation of media can be particularly challenging in a cloud computing environment” (NIST Cloud Computing Forensic Science Working Group, 2014, p. 11). Collecting digital evidence from a physically accessible system is technically feasible: a search warrant can be issued and associated with a specific device and location, and law enforcement officers can be engaged to identify and acquire the needed data. In traditional digital forensics models, the system or device involved can be seized and then the rest of the process applied. The process cannot be followed in the public cloud because of its fundamental architecture where the network is shared so the storage is shared and physically split between multiple physical devices or geographical locations. In fact, physical access might not be permitted at all, so alternative methods must be pursued.

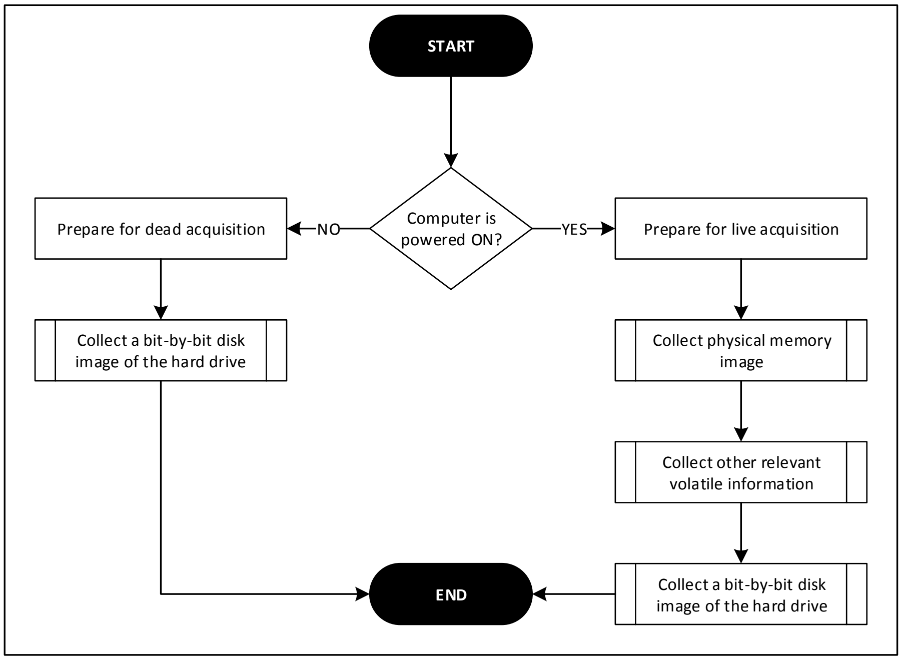

START
Computer is powered ON?
NO
YES
Prepare for dead acquisition
Prepare for live acquisition
Collect a bit-by-bit disk image of the hard drive
Collect physical memory image
Collect other relevant volatile information
Collect a bit-by-bit disk image of the hard drive
END

Figure 3 - Traditional digital forensics acquisition – a high-level overview

As with traditional forensics, no single methodology can be prescribed in every circumstance for the acquisition or collection of forensic data from a cybercrime scene. Each investigation is different and requires a different set of tools and techniques. In the past few years, the Scientific Working Group on Digital Evidence (SWGDE) has published two documents that shed light on some of the best practices for digital evidence forensic acquisition (SWGDE 2014, 2018). Table 2 presents a summary of recommendations, including a subset from SWDGE recommendations, to be followed by forensic examiners during the acquisition or collection phase of a digital forensics' investigation. Another extremely valuable source on best practices related to digital forensics followed by the international DF community is *Digital Forensics Processing and Procedures* by Watson and Jones (2013). The book goes into greater details of the complete life cycle of digital evidence and the importance of chain of custody. This handbook includes technical and non-technical procedures and best practices to comply with international regulations such as ISO 17020, ISO 17025, and ISO 27001. While looking into forensic acquisition from methodical and technical points of view, RFC 3227 (Killalea & Brezinski, 2002) cannot be omitted, as it provides a great amount of detail on the technical procedure and best practices of collection as well as examples of which tools can be used.

In this research, best practices have been extracted from the above-mentioned resources, organized, and merged when possible, then each recommendation is assigned an impact (*trust*) factor, which either enhances or devalues the effectiveness of the collected data. In this case the factor could be the *integrity*, *authenticity*, or *operability* of the collected evidence or data. While integrity and authenticity of the data might have a direct impact on the trustworthiness of the generated evidence, operability contributes to the errorless and seamless extraction of quality information and evidence from the acquired data.

Table 2 - Digital Forensic Acquisition Recommendations and Associated Impact Factor

Best Practice	Impact Factor
1) Prepare order of volatility for each system before starting the acquisition process.	Integrity/Operability
2) Minimize, to the maximum extent possible, changes to the source data.	Integrity
3) Minimize adverse effects as much as possible when choosing the acquisition technique.	Authenticity
4) Document the acquisition process in as much detail as possible.	Authenticity
5) Make all actions taken during the acquisition process auditable, where applicable.	Authenticity
6) Use tools validated for use according to NIST CFTT (trusted binaries).	Integrity
7) Prepare destination media before beginning of the acquisition process.	Authenticity/Integrity
8) Store acquired data on a trusted platform.	Authenticity/Integrity
9) Acquire and save data in raw format, when possible.	Operability
10) Preview the contents of potential data sources prior to acquisition.	Integrity
11) Live acquisition tools should execute trusted binaries from controlled media.	Integrity

12) Execute live acquisition at the lowest possible level of privilege.	Integrity
13) Document "memory smear" if experienced during a live memory acquisition.	Integrity
14) Consider carefully the order in which data are collected.	Integrity/Operability
15) Review acquired data for completeness and certainty.	Integrity/Operability
16) Review tool output and logs for indications of failures.	Integrity/Operability
17) Compute a cryptographic hash value for the acquired data using NIST approved algorithms.	Integrity
18) Document the physical and logical chain of custody properly.	Integrity

Digital Forensics - Admissibility

The general definition of digital evidence admissibility (forensically-sound evidence) has been constantly debated in the forensic science community, most especially when it comes to acceptance in courts of law, nationally or internationally. During the International Hi-Tech Crime and Forensics Conference (IHCFC) of October 1999, the SWGDE and the International Organization on Digital Evidence (IOCE) presented a set of international principles to govern the acquisition and recovery of digital evidence that would be acceptable in courts of law (SWGDE & IOCE, 2000). The proposed principles were well received and adopted by many organizations and governments, including the United States.

Table 3 - SWGDE and IOCE Standards and Principles (SWGDE & IOCE, 2000)

IHCFC Principle
1) Upon seizing digital evidence, actions taken should not change that evidence.
2) When it is necessary for a person to access original digital evidence, that person must be forensically competent.
3) All activity relating to the seizure, access, storage, or transfer of digital evidence must be fully documented, preserved, and available for review.
4) An individual is responsible for all actions taken with respect to digital evidence while the digital evidence is in their possession.

5) Any agency that is responsible for seizing, accessing, storing, or transferring digital evidence is responsible for compliance with these principles.

Furthermore, according to McKemmish (2008), forensic soundness could be defined as the “application of a transparent digital forensic process that preserves the original meaning of the data for production in a court of law” (p. 10). The author in this publication proposed four criteria to evaluate the reliability of an electrically generated evidence:

1. The generated evidence did not get affected by the digital forensic process.
2. All the generated errors can be identified and satisfactorily explained.
3. The entire followed DF process can be independently examined and verified.
4. The forensic analysis been undertaken by an experienced individual

In addition to the admissibility recommendations mentioned above, evidence admissibility in the U.S. court system has always been justified via compliance with a specific set of rules and regulations. For digital evidence to be called upon in U.S. courtrooms—that is, to be “admissible in court,” it must conform to guidelines that govern how the evidence is collected, authenticated, preserved, and presented. These guidelines are called the Federal Rules of Evidence (FRE). According to the American Bar Association, for digital evidence is admissible if it meets five criteria: (1) the evidence must be relevant to what the court case is about (FRE 401, 402, 403); (2) it must be authentic and proven accurate by technical witness expertise (FRE 901); (3) it must not be hearsay or contribute to changing the probability of facts that would otherwise be true (FRE 801); (4) if the evidence is a voice recording or a photograph, the original or accepted duplicate might be presented (FRE 1003); and finally, (5) its probative value must be assessed against its possible unfair prejudice (FRE 104, 105).

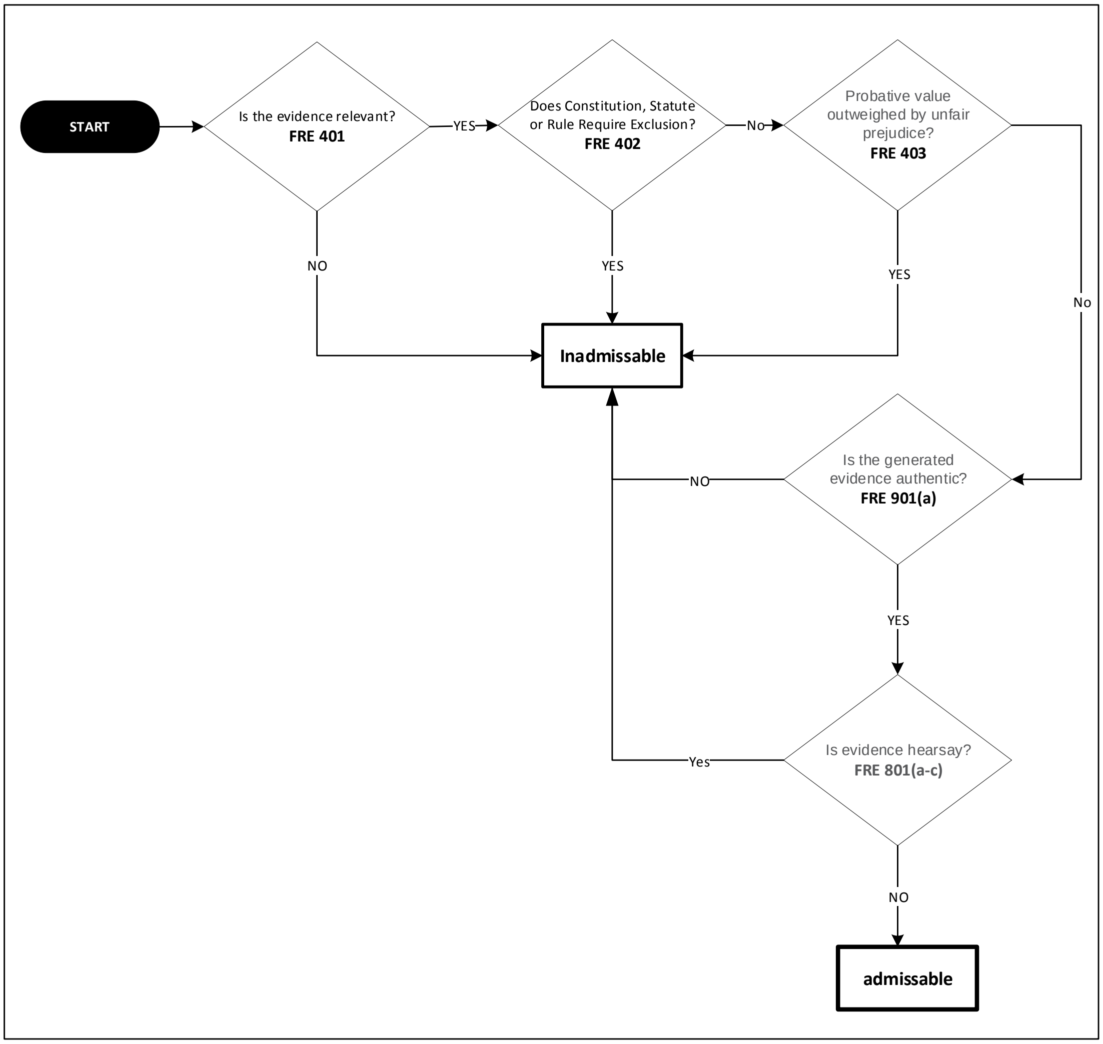

START
Is the evidence relevant?
FRE 401
YES
Does Constitution, Statute or Rule Require Exclusion?
FRE 402
No
Probative value outweighed by unfair prejudice?
FRE 403
NO
YES
YES
No
Inadmissable
Is the generated evidence authentic?
FRE 901(a)
NO
YES
Is evidence hearsay?
FRE 801(a-c)
Yes
NO
admissable

Figure 4 – U.S. court digital evidence admissibility workflow

FRE 902 - Evidence that is self-authenticating is a relatively recent and significant Federal Rule of Evidence that was amended in December 2017. This new rule provides information about how electronic records generated by electronic systems can be self-authenticated. The rule has fourteen items listed that require no extrinsic evidence of authenticity to be considered admissible. The rules that concern digital forensics are 902(13) and 902(14) and their prerequisite rules 902(11) and 902(12). Below are the quoted definitions related to each FRE:

- FRE 902(11): “Certified Domestic Records of a Regularly Conducted Activity.”
- FRE 902(12) “Certified Foreign Records of a Regularly Conducted Activity.”
- FRE 902(13): “Certified Records Generated by an Electronic Process or System.”

- FRE 902(14): "Certified Data Copied from an Electronic Device, Storage Medium, or File**."** (Federal Rules of Evidence, 2017)

The 902 FRE rule also addresses the point of digital identification, where hash values can be used to authenticate that the electronic record is what it purports to be. According to the ABA, FRE 902(13) specifically provides more information on the self-authentication of "A record generated by an electronic process or system that produces an accurate result, as shown by a certification of a qualified person that complies with the certification requirements of Rule 902(11) or (12). The proponent must also meet the notice requirements of Rule 902(11)" (Toft, 2018). This rule has a significant impact on the global digital forensics' community in terms of practices, costs, timelines, and unexpected evidentiary arguments.

Courts regularly require that expert witnesses validate evidence claims based on the Daubert (Ryan, 2009a) or Frye standards (Ryan, 2009b). Daubert is the most commonly used test today, deriving from the 1993 Supreme Court case, Daubert v. Merrell Dow Pharmaceuticals Inc., 509 U.S. 579. Daubert uses five factors to determine if the methodology followed is acceptable: "(1) whether the theory or technique in question can be and has been tested; (2) whether it has been subjected to peer review and publication; (3) its known or potential error rate; (4) the existence and maintenance of standards controlling its operation; and (5) whether it has attracted widespread acceptance within a relevant scientific community" (Ryan, 2009a). Frye, the older and more rarely used standard, is derived from Frye v. United States, 293 F. 1013 in 1923. While Daubert has multiple factors to determine whether the methodology followed by the expert witness is valid, the Frye standard calls only for general acceptance within the scientific community (Ryan, 2009b).

Another set of rules to follow, more specifically designed for ESI and evidence acquired through a subpoena for a civil case, is the Federal Rules of Civil Procedure (FRCP). The two rules that may be applicable when addressing a CSP are FRCP 34 (Rule 34. Producing Documents, Electronically Stored Information, and Tangible Things, or Entering onto Land, for Inspection and Other Purposes, n.d.), which allows a request to preserve, collect, and possibly inspect data, and FRCP 45 (Rule 45. Subpoena, n.d.), which allows for a subpoena. In criminal cases, subpoenas can be obtained by applying the Federal Rules of Criminal Procedure and invoking Rule 41 (search warrant). FRCP Rule 41 is used by prosecutors after the case has been filed. Search warrants are used prior to the case being filed with the prosecutor. In civil matters, however, subpoenas are used quite frequently. All of these rules were crafted for traditional forensics investigations and evidence collection, and then adopted for digital forensics.

Statement of the Problem

The acquisition phase of a digital forensics' investigation is one of the most critical and fundamental phases of the entire process. If the investigation has not been performed appropriately, it can be challenged in court with regards to chain of custody, completeness of documentation, integrity of the generated evidence, and methodology used to acquire the evidence (Montasari, 2017). Though telecommunications companies have to comply with federal laws—such as the Communications Assistance for Law Enforcement Act (CALEA), which permits law enforcement agencies to use wiretaps—no such law exists for dealing with CSPs (Dykstra, 2013a). Access to direct authentic evidence that is physically collected by law enforcement officers might not a viable option when dealing with incidents in the public cloud. Following the traditional forensic acquisition methodologies and techniques, by design, is not feasible to conduct court-grade investigations in the realm of the public cloud.

The problem is there is a lack of digital forensic methodologies that can be followed to collect court-grade digital evidence from IaaS public cloud environments (Barrett, 2018; Dykstra, 2013a). The aim of this research is to complement and extend the existing traditional digital forensic acquisition methodologies and processes to encompass public cloud deployments, while continuing to comply with federal and civil laws. The aim is to preserve the *authenticity* and *integrity* of the collected evidence, thus supporting the credibility of expert witnesses, and ease the decision for jurors and judges. Therefore, we are presented with the following research question: What digital forensic tools and acquisition techniques are applicable for court-grade evidence acquisition from IaaS deployments in the public cloud?

CHAPTER TWO

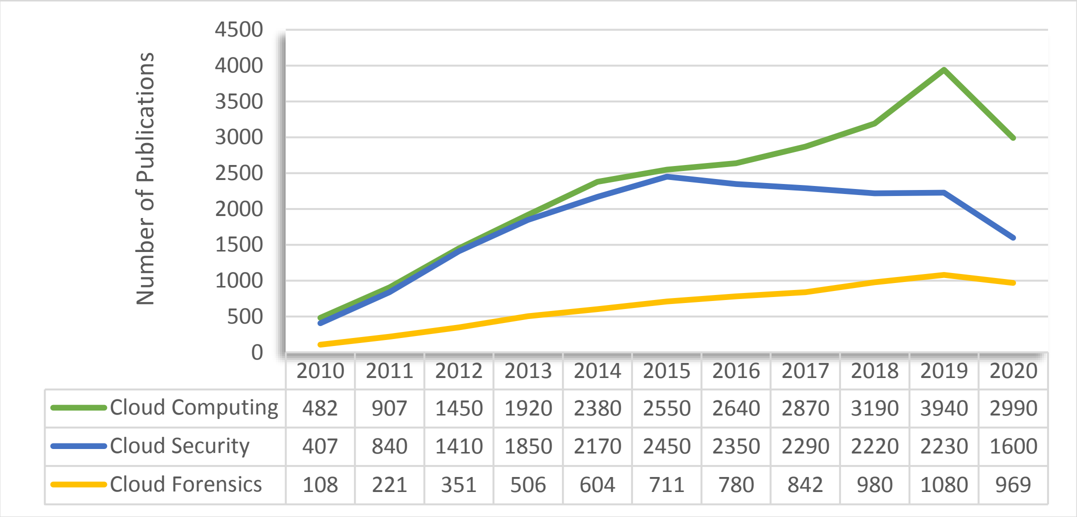
LITERATURE REVIEW

Research trends and developments

In the past two decades, researchers have published many scholarly articles, papers, and books that shed light on the possibilities and limitations of cloud computing in delivering services. According to Dykstra (2013a), until 2012, all cloud-focused research was about 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” (p. 12). Though the main research area for this proposal is the IaaS deployments of the public cloud computing model, it is also interesting to note the trends in literature and research related to cloud computing overall as compared to research that focuses on cloud security and on cloud forensics. It is not feasible to accurately locate and catalogue every publication related to cloud computing but using statistical approximation can help us understand generally what has been published so far. Finding a single association or online library that had a copy of all possible published work was also very challenging, so Google Scholar was the best tool that could search hundreds of research associations and organizations to get more comprehensive results.

This high-level systematic search encompassed all scholarly work from around the world and was not filtered to publications from the major associations only. The publication timeframe was limited to 2010 through 2020 (inclusive), and a-per-year count was recorded. This search filtered for compound phrases in the metadata as well as the various sections of the published work. The outcome of this search shows the serious lack of research when it comes to addressing forensic science applications in the realm of cloud computing. The limitation to this search is work that has not been published electronically or made available through online libraries. Sample compound searches used include the following:

- *"cloud computing" "cloud * computing" OR "Public Cloud" OR "private cloud" -"cloud security" -security -forensics*
- *"cloud security" "cloud * security" OR "Public Cloud" OR "private cloud" -forensics*
- *"cloud forensics" OR "cloud computing" OR "Public Cloud" OR "private cloud" "digital forensics"*



	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
🟩 Cloud Computing	482	907	1450	1920	2380	2550	2640	2870	3190	3940	2990
🟦 Cloud Security	407	840	1410	1850	2170	2450	2350	2290	2220	2230	1600
🟨 Cloud Forensics	108	221	351	506	604	711	780	842	980	1080	969

Figure 5 - Cloud Forensics research trend in the past ten years

Landscape of known challenges

In the past six years, the security of the data and information hosted in the public cloud has become a significant and controversial topic with comparisons made between securing on-premises assets and securing the same in the multitenant public cloud space. As a result, the research community started resolving the ambiguities and attempted to settle the debates about cloud security. They have uncovered many challenges and proposed some successful solutions. One of the challenges is Incident Response (IR) and Digital Forensics (DF) in the public cloud, which is being researched presently (Alqahtany et al., 2015). Researchers have been actively publishing about the challenges of exercising digital forensic science related to Anything-as-a-Service (XaaS) incidents (Freet, Agrawal, John, & Walker, 2015). Almost all researchers have categorized the challenges into three categories: technical, architectural, and legal. Few researchers have decided to categorize challenges according to the various frameworks or cloud services models (Freet et al., 2015; Grispos et al., 2012). NIST, on the other hand, has categorized challenges into nine categories in its draft report (NIST Cloud Computing Forensic

Science Working Group, 2014). This literature review revealed that most of the challenges usually span more than a single category, concentrating on the legal or judicial domain.

During the past decade, researchers have identified the following high-level digital forensics challenges, and although there are many more, the list below highlights the most commonly referenced ones. Those highlighted challenges can exist in any or all of the commonly adapted cloud service models such as Software-as-a-Service (SaaS), Platform-as-s-Service (PaaS), and IaaS. By design, the more control users have over a service layer, the fewer challenges they face. For example, IaaS should have fewer forensic challenges than SaaS. Solutions have also been proposed for some of these challenges, but that doesn't mean they have been widely adopted or accepted by CSPs, regulatory bodies, or forensic tools development organizations. Table 4 shows some of the commonly referenced challenges and their respective categories, as well as how sometime a challenge can span multiple categories.

Table 4 - Common Challenges and Corresponding Categories

	Challenge	Technical	Architectural	Legal
1)	No access to the hypervisor logs or volatile data (Freet et al., 2015) (Simou et al., 2016) (Grispos et al., 2012) (Alqahtany et al., 2015) (Cohen, 2013) (Zawoad et al., 2015)	✓	✓	
2)	High dependency on the CSP (Simou et al., 2016) (Freet et al., 2015) (Alqahtany et al., 2015) (Dykstra, 2013a) (Zawoad et al., 2015)		✓	✓
3)	No physical access & difficulty of imaging (Grispos et al., 2012) (Simou et al., 2016) (Alqahtany et al., 2015) (Dykstra, 2013a) (Barrett, 2018) (Zawoad et al., 2015)	✓	✓	✓
4)	Trusting the collected data integrity (data provenance) (Simou et al., 2016)			✓

	(Freet et al., 2015) (Alqahtany et al., 2015) (Dykstra, 2013a) (Cohen, 2013) (Zawoad et al., 2015)			
5)	Time zones synchronization (Grispos et al., 2012) (Simou et al., 2016) (Alqahtany et al., 2015) (Cohen, 2013)		✓	
6)	Maintaining chain of custody (Grispos et al., 2012) (Simou et al., 2016) (Alqahtany et al., 2015) (Dykstra, 2013a) (Cohen, 2013) (Zawoad et al., 2015)			✓
7)	Legal authority and geographical boundaries (location) (Grispos et al., 2012) (Simou et al., 2016) (Miranda Lopez, Moon, & Park, 2016) (Alqahtany et al., 2015) (Dykstra, 2013a) (Barrett, 2018)		✓	✓
8)	Lack of standardized cloud forensics acquisition tools (Grispos et al., 2012) (Simou et al., 2016) (Freet et al., 2015) (Alqahtany et al., 2015) (Dykstra, 2013a) (Barrett, 2018) (Zawoad et al., 2015)	✓		✓
9)	Lack of standardized cloud forensics examination tools (Grispos et al., 2012) (Simou et al., 2016) (Freet et al., 2015) (Alqahtany et al., 2015) (Dykstra, 2013a) (Barrett, 2018)	✓		✓
10)	Lack of literacy of court jurors (Grispos et al., 2012) (Alqahtany et al., 2015)			✓

(Cohen, 2013) (Zawoad et al., 2015)		
11) Lack of literacy for public cloud customers (Simou et al., 2016) (Alqahtany et al., 2015) (Dykstra, 2013a)	✓	
12) Lack of literacy for cloud investigators (Alqahtany et al., 2015) (Cohen, 2013)		✓

The literature overview done by (Simou et al., 2016) resulted in a comprehensive summary table referencing all previously identified challenges and the proposed solutions. Though their research findings are perhaps now antiquated, their highlighting of the challenges expressed, and the solutions proposed between 2011 and 2016 remains useful. Their literature review shows that most of the challenges can only be addressed by the CSPs. Although some of the challenges happen to capture the attention of the research community, who have proposed multiple viable solutions, no evidence suggests that those solutions were adopted by CSPs. These proposed solutions vary in implementation requirements, but to address the most critical challenges and help build legally acceptable forensics practices, CSP involvement is inevitable (Alqahtany et al., 2015). Clearly, most of the challenges fall into the collection/acquisition/preservation phase of the various digital forensics models. Someone wanting a quick reference and overview of existing challenges (table of problem/solution pairs) can look into (Simou et al., 2016), (Miranda Lopez et al., 2016), (Freet et al., 2015), and (Alqahtany et al., 2015), all of whom have compiled a holistic summary of the challenges and the respective proposed solutions in the past decade.

Cloud Forensics - Models

Cloud forensics has been defined as “the application of scientific principles, technological practices and derived and proven methods to reconstruct past cloud computing events through identification, collection, preservation, examination, interpretation and reporting of digital evidence” (NIST Cloud Computing Forensic Science Working Group, 2014, p. 2). NIST’s definition implies the application of the NIST digital forensic model to guide forensic investigation in a cloud computing environment, which contradicts the findings of academic researchers in the field, such as Barrett and Kipper (2010) and Martini and Choo (2012).

The research performed by Martini & Choo (2012) is considered one of the few notable studies that attempted to tackle the bigger picture of cloud forensics and proposed an integrated digital forensics model that could actually be adapted and followed to conduct investigations in various cloud deployments. In their publication, the authors started by going over the various cloud forensic challenges highlighted by the research community, then reviewed some of the commonly used models, namely Kent, Chevalier, Grance, & Dang (2006) and McKemmish (1999), analyzed the gap, and established new requirements to be able to carry-on forensic investigation in cloud environments. Their proposed model is a hybrid of both Kent et al., (2006) and McKemmish (1999), consisting of four phases: Evidence Source Identification and Preservation; Collection; Examination and Analysis; Reporting and Presentation. The major distinction of the proposed model occurs during the examination and analysis phase. The authors reason that during this phase in cloud forensic investigation, some iterations might be required as "cloud computing usage would most likely be discovered based upon the examination and analysis of physical devices and this would lead to a second (or more) iteration(s) of the process" (Martini & Choo, 2012, p. 75).

The same researchers, one year later, attempted to add more rigor and validation to their proposed model and applied it successfully to cloud storage open source services OwnCloud (Martini & Choo, 2013). Furthermore, in the same year, Quick & Choo (2013c) attempted to generalize the previously proposed methodology and demonstrated its effectiveness to successfully acquire forensic data from three of the most popular cloud file storage providers: Dropbox (Quick & Choo, 2013a), Google Drive (Quick & Choo, 2014), and Microsoft SkyDrive (Quick & Choo, 2013b). While this model was successful in producing forensically-sound data from SaaS deployment, the literature shows no demonstrations of IaaS-related deployments or investigations.

Cloud Forensics - Tools

One of the biggest challenges facing cloud forensics is the lack of digital forensics tools that are designed with cloud computing in mind. There is a tremendous need for tools to perform acquisition, complete analysis, and forensic examination of the cloud data remotely (Almulla, Iraqi, & Jones, 2014). The Computer Forensic Tool Testing (CFTT) project at NIST attempted to create a list of trusted tools to perform traditional digital forensics as well as cloud forensics. The effort generated tool specifications, test procedures, test criteria, test sets, and test hardware (Allen, 2017). As of March 2021, the published “cloud forensics” tools on the CFTT website are all purposed for SaaS investigations such as OneDrive, Flickr, Dropbox, and such. There is no tool related to IaaS investigations listed in the database. While there are many popular, documented, traditionally tested and widely used digital forensics tools such as the Linux-based “dd” or the commercial grade “FTK,” the literature did not show any scholarly work related to systematically testing such tools on public cloud acquisitions.

According to Dykstra (2013a), traditional forensic tools are not well suited to the vast quantity of data nor the type of data in the cloud, and often cloud infrastructure management tools are lacking in forensic options. Moreover, few available examples of cloud investigations can be used by practitioners for educational purposes. Prior to Dykstra developing FROST (Dykstra, 2013a), no known useable tools had been designed specifically for cloud forensics. FROST is the closest tool available today to perform real credible forensics on the management plane of private clouds (OpenStack). The idea was to convince CSPs to take a similar approach to making their offered services forensically capable. This led Dykstra to evaluate and analyze different leading commercial digital forensics tools to see if they could be used for cloud forensics. Though FROST was developed in response to cloud forensic challenges, some existing commercial tools, such as FTK Remote Agent, EnCase Remote Agent, and X-Ways, have also been used to perform live and remote forensics to some extent.

Dykstra argues that another challenge is the format of the data. Cloud providers who aid in an investigation might provide the investigators with data that is in propriety VMs or in unsupported formats. These could be difficult or impossible to analyze with popular tools such as EnCase or Access Data's Forensic Toolkit (FTK). Thus, state or local law enforcement agencies may be unable to fully investigate cybercrimes that involve cloud-based infrastructure or cloud storage. According to Zawoad et al. (2015), another important challenge is the capturing of volatile data. When a virtual machine is powered off, all such volatile data is lost unless an image of the instance or a stateful snapshot is made. Some data, such as registry files, may therefore be lost, and someone could intentionally exploit this to ensure the loss of volatile data.

Few researchers have argued that CSPs can restructure and introduce valuable forensic capabilities at the hypervisor level (Poisel, Malzer, & Tjoa, 2013). Without changing the system state, the hypervisor is able to provide access to computing resources at a low level. The hypervisor can allocate computing resources, such as disk space, CPU, and memory, as well as networking. Hypervisors are fully capable of secretly monitoring, introspecting, and interacting with VM guests with complete transparency, which can be immensely useful for cloud forensics. However, most hypervisors fail to expose APIs at a level sufficient for fine-grained introspection that can be customized and remain transparent (Poisel, Malzer, & Tjoa, 2013). Thus, CSPs have to start implementing forensic services at the management plane or hypervisor kernel level to help resolve the majority of the challenges being highlighted.

Cloud Forensics - Acquisition

While reading this section of the research and moving forward, one must keep in mind the unblemished distinction between acquiring data for incident response purposes and performing digital forensic acquisition in order to pursue a litigation. The focus of this research is the latter one. NIST affirms that the "identification, collection, and preservation of media can be particularly challenging in a Cloud Computing environment" (NIST Cloud Computing Forensic Science Working Group, 2014, p. 8). Most existing methodologies and frameworks assume a tool capable of collecting and preserving the needed data, while maintaining chain of custody until delivered to court. With traditional digital forensics, one of the challenges has been seizure and imaging of physical discs; this is not possible with cloud forensics. As (Dykstra, 2013a) argues, physical seizure may be difficult for many reasons, such as that multiple tenants may have data on one physical drive, or that one tenant's content could be held by many hard drives. Also, that the data would be stored in disk-arrays and not individual disks is a very high possibility. Imaging of media could be impractical in a cloud and having only a partial image could be called into question legally as not being whole and complete. In addition, the disk drives could be spread out over many geographical areas, such as on different countries or continents (Dykstra, 2013a; Miranda Lopez et al., 2016, Simou et al., 2016).

Dykstra writes that two possibilities exist for data acquisition with cloud forensics. First, the forensic examiner could collect it remotely, and second, the CSP could furnish it to the examiner. According to Dykstra, if forensic examiners had access remotely to the compromised server operating system (OS), then they could potentially collect evidence in a couple of ways. 1) They could install a forensic tool to remotely acquire evidence; or, 2) the VM could be suspended or stopped and then subjected to offline analysis. All this effort, however, assumes a great deal of confidence and trust in the hypervisor, the host OS, the hardware, and the guest OS; in this case, it would provide evidence that is not only whole but with full integrity. It assumes no omissions, no mistakes, and no tampering (Dykstra, 2013a). Alternatives for forensic acquisition include Trusted Platform Modules (TPMs), the cloud management plane forensics, and Forensics-as-a-Service.

While it might be possible to collect data remotely from IaaS public cloud deployments using some of the existing tools today, there is no methodology or standardized process to perform such a sensitive task. If attempted, the process would have to be based on the forensic practitioner's personal understanding of the investigated cloud platform; there is no document to be followed to repeat such acquisition if requested in courts of law. In 2014, the Scientific Working Group on Digital Evidence published a document to help guide forensic examiners acquire data from live systems (SWGDE, 2014). The document focused on defining order of operations and types of data to be acquired from a computer, and finally addressed the possible limitations and the negative implications of systems' state changing due to live acquisition. While this publication did not explicitly refer to public cloud deployment or IaaS environments, it is still applicable when it comes to imaging memory and collecting volatile data from virtual machines deployed in the public cloud.

The same organization published a document entitled "SWGDE Best Practices for Computer Forensic Acquisitions," which gives practitioners high-level guidelines on how to carry out a successful and court-grade forensic acquisition and evidence preservation (SWGDE, 2018a). The document is focused on forensic acquisition from computers/servers and attached storage media, and it gives recommendations on each one of the possible acquisition types: physical, logical, and targeted acquisition (SWGDE, 2018a). While this document does not mention the keyword "cloud" anywhere and does not constitute a defined methodology, it can be applied partially in the case of IaaS acquisitions because the document includes high-level guidelines and recommendations to follow when capturing any computer system (Storage & Memory).

In September 2020, as this dissertation was being written, SWGDE published a document that directly addresses forensic acquisition when it comes to the cloud services providers. The SWGDE (2020) best practice document is considered a very exceptional document, and it contributes greatly to the knowledge base of cloud forensics and the DF community. At the beginning of this research, there was no such document in the literature that explicitly guided the forensic examiner on acquiring data from public clouds. The authors of the best practice document did not focus on a specific cloud service model, but rather generalized the scope to cover all of the possible service and deployment models. The document approaches forensic efforts from a legal perspective and shares various possible methods of acquiring forensic data, such as asking the CSP to furnish the data, exporting the data using native cloud tools, using exposed APIs and commercial DF remote agents, and finally physical seizing the cloud hardware (SWGDE, 2020). The document splits the acquisition process into three main phases and provides high-level guidelines on what should be done in each phase:

- Prior to acquisition phase
    - Identify the CSP, data sources, data types, timeline, and utilized services.
    - Ask the CSP legally to preserve the data sought.
    - Identify and choose which acquisition methodology will be used.
- During acquisition phase
    - Document all evidentiary data.
    - Acquire all attached media, local and in the cloud.
    - Confirm the acquisition methodology can produced the required data.
    - Acquire the data with the selected methodology.
    - Take photographs/screenshots of the relevant data if unable to acquire data.
- After acquisition phase
    - Calculate integrity hash values for acquired data.
    - Verify that the executed acquisition was able to acquire all required data.
    - Document the whole process using the organization defined policies
    - Document any received media from the CSP.
    - Store all acquired data following the organizational policies (SWGDE, 2020).

These three best practice publications from SWGDE are excellent starting points to performing forensic acquisition in the public cloud. The documents also provide high-level guiding principles to govern the DF processes, such as how to perform live acquisition, define OOV, acquire data from computers and storage, and addresses public cloud investigations. While the information in SWGDE, 2020 is the most relevant to the research question, it still leaves a gap about how to carry out each one of the proposed phases. SWGDE, 2020 argues that the process can be taken directly with the CSP, or without the CSP, and attempt to acquire the data. The described procedure is very high level by design, and it keeps a lot of room for the forensic practitioner to make their own decisions. The procedure does not have default processes to fit small businesses or organizations that do not have policies and standards to follow when it comes to digital forensics. The document omits chain of custody and physical personnel who performed acquisition or interacted with the digital evidence or forensic data. The process also lacks reference to the cloud metadata, and it doesn't discuss the possible types of data that can be acquired from a CSP. Each type of data can require a different acquisition mechanism, and there is not a single selected mechanism of the described methodology that can capture all sought data. These excellent recommendations from SWGDE thus act as a basis for the present research and are enhanced and motivated to address IaaS deployments in the public cloud realm.

While there are a plethora of models, methodologies, and procedures related to the overall lifecycle of digital forensic evidence, few resources go into detail about how the process should be followed to provide optimized results. There are only guidelines and instructions provided by the various DF software vendors on how they think the process should be carried out. While cybercrime investigations usually vary by which methodology or tool can be used, and each case can be very different, having a common ground or methodology to follow is efficient and advantageous. It is crucial for the forensic analyst or examiner to put a plan together to determine how they will approach each case before attempting to perform any task.

That said, regardless of which process is being followed, the anticipated outcome is expected to be forensically sound, and, if litigation is expected, it must be able to stand in courts of law. Appendix A presents the various acquisition methodologies and procedures encountered during literature review. Only the live acquisition and logical collection phases of the methodologies has been captured in Table 14. The literature shows that almost all of the encountered methodologies do not go into the details of how the forensic practitioner should approach the cybercrime scene, and they leave it up to the individual organization incident response policies. This might be a possibility, but assuming that all small and large organizations have a defined cloud forensics or digital forensics policies, and procedures is not realistic.

Another significant problem highlighted in the literature is the post-acquisition examination. Some disk images acquired from cloud platforms cannot be forensically validated or rendered not traditionally operable. This in itself could pose a substantial legal challenge, which could ultimately frame the evidence as intangible. With traditional digital forensics, a hash of the intact physical hard drive can confirm that a copied disk is the same as the original. This is not always possible with cloud forensics, and in court the lack of data integrity could cause results to be rejected (Dykstra, 2013a). Within the acquisition phase of traditional digital forensics, data are to be acquired in such a fashion that both authenticity and integrity are maintained. Though impractical for cloud forensics, often this constitutes the use of a write blocker, thereby ensuring that nothing else is written to the disk. Imaging software is to be used, providing either a bit-for-bit image or a logical image made from active files and directories. In order to provide assurance with regard to the integrity of the collected data, it is essential that hashing be employed. Different solutions attempt to address data integrity at a stage prior to collecting the data, such as FROST; some SaaS forensics tools, such as Kumodd and Kumofs, as developed by Roussev, et al.; Cloud FaaS; and Almulla's snapshot-based forensics framework.

Because cloud data storage is decentralized, located in multiple jurisdictions, some legal issues arise that are very difficult to address and can create challenges when commonly issued search warrants are pursued. With the warrant, it is necessary to identify a particular place where the evidence is located. That may not be possible with cloud forensics. Even if a warrant is granted, there could be an issue in maintaining chain of custody because of the multiplicity of jurisdictions, differing procedures, and the use of proprietary hardware and software (SWGDE, 2014). Another legal challenge is the availability of cloud forensic acquisition tools that have been proven to acquire data in the cloud that meet the legally required rules for a specific jurisdiction. While it can be argued that some existing tools that are used today collect data from traditional computers or servers can be reused in the cloud, few tools are natively designed with the cloud in mind. Thus, the court could challenge a lack of tools certified for the cloud. Accordingly, legal requirements for forensically sound evidence also remain nebulous.

Cloud Forensics - Laws and Regulations

Throughout the work investigated here, researchers agree that the reason for the existence of almost all of these challenges is ultimately legal (laws, regulations, and court compliance and acceptance). Miranda Lopez et al. (2016) and Dykstra (2013a) reached a similar conclusion after evaluating two hypothetical use cases of criminal activities in cloud computing settings. In their draft publication regarding cloud forensic challenges, NIST has highlighted 65 challenges, 33 of which fall under legal or architecture categories. The drive behind performing digital forensics in the first place has always been uncovering criminal behavior and presenting them to their respective authorities (NIST Cloud Computing Forensic Science Working Group, 2014). The collected evidence must follow certain rules to be admitted at a trial level if litigation takes place. If there is no litigation, then the collected information/data/evidence does not need to be authenticated against any federal rules or regulatory body standards. Thus, most challenges are not categorized under cloud forensics, but rather are dispersed among cloud incident response frameworks and corporate security policies.

Researchers have found that the largest US-based IaaS-offering CSPs—AWS, GCP, and Azure—have published basic guidelines on how to technically acquire a copy of virtual assets or media, such as acquiring a raw image of a virtual machine storage or a copy of the volatile memory. Neither the CSPs nor their competitors have published information related to acquiring evidence that can be admissible in court. The abovementioned CSPs have made available some blog posts and whitepapers about how to proactively automate incident response (threat mitigation) in their respective service offerings. These write-ups have focused mainly on processes related to anomaly detection and log correlation and targeting organization’s incident response and security policy teams (Building a Cloud-Specific Incident Response Plan, 2017; Data Incident Response Process | Documentation, 2018; rkarlin, 2018). One recent publication showed how acquisition could be accomplished in public clouds using the CSP-exposed APIs; the researchers focused on Amazon Cloud (AWS) and successfully acquired full disk and memory images using the provided API calls, as well as some third-party tools (Orr & White, 2018).

The same CSPs have provided information on their websites about how law enforcement agencies, attorneys, or private investigators could approach them for search warrants or subpoenas (Amazon, 2015; Law Enforcement Requests Report – Microsoft Corporate Social Responsibility, 2018; Orton, Alva, & Endicott-Popovsky, 2013). Though this information is helpful, it is absolutely up to the CSP to decide what data they can share, what they can share partially, and what they can reject sharing (Amazon, 2015; Evans, 2017). The CSP's decision about what type of data to provide is dependent on many factors, such as their understanding and compliance with governing laws, regulations, and common service provider's privacy practices.

Research Gap

After reviewing the literature over the past decade related to cloud forensics—including the challenges that have been voiced and the solutions proposed— the researcher believes that for these proposed solutions to be effective, the CSPs must perform architectural changes to their underlying hypervisors or cloud engines to allow for native forensic capabilities and expose the appropriate API calls. The other option is for the legal community and legislators to start creating a different set of acceptance criteria for cloud-generated digital evidence, taking into consideration the limitations that exist today. Arguments are also made for new laws and regulations to govern the public cloud space to allow for more control when it comes to law enforcement agencies (LEAs) interactions. Alqahtany et al. (2015) explains the challenging nature of demonstrating the integrity of cloud-based evidence in court in such a manner that it can be admissible without being contested. This can be due to the number of different stakeholders involved in the data preservation stage. While some researchers have highlighted the legal challenges to cloud forensics, they have been addressing all challenges together at a high level, with no focus on specific challenges (Dykstra, 2013a).

The few practical solutions proposed for cloud acquisition have always argued for the involvement of the CSP in order to adjust the cloud architecture or allow external interfaces direct access to the cloud infrastructure. The reviewed literature does not mention the use of the existing digital forensic acquisition tools to perform acquisition in the public cloud.

In the present research, the gap between cloud forensic acquisition and the U.S. court system has been closely examined, and a forensically-sound methodology to prove the collected data’s authenticity is proposed. The methodology inherits its main structure from the various international work groups (SWGDE, IOCE, and IETF), standards bodies (NIST, ISO2700), and U.S. federal rules and regulations (FRE, FRCP, Daubert). The proposed methodology extends the existing knowledge base of digital forensics to encompass public cloud IaaS deployments. The focus of this research is to deliver indubitable and highly admissible forensic evidence to the U.S. courts. Further, the goal is to allow the same methodology to be extensible, and thus easily applicable to any other court system or corporate incident response policies.

While there is less focus on the existing commercial forensic tools, some existing open-source and well-known digital forensic acquisition tools have been tested against public cloud IaaS resources. Some newly made available means of collecting forensic data from public clouds have been tested, such as applications programming interfaces (APIs) exposed by the various CSPs, as well as some of the most recently developed cloud forensic gadgets. In the present research, two hypothetical scenarios are carefully crafted to mimic realistic forensic investigation cases in IaaS deployments. The existing forensic acquisition methodologies and best practices are tested against the forensic cases, then the newly developed methodology, IPCFA, is attempted. The focus of this research is to maintain the authenticity of the collected forensic evidence, meet U.S. courts requirements, and help pass the Daubert test using a uniquely developed rubric to execute and validate the proposed methodology.

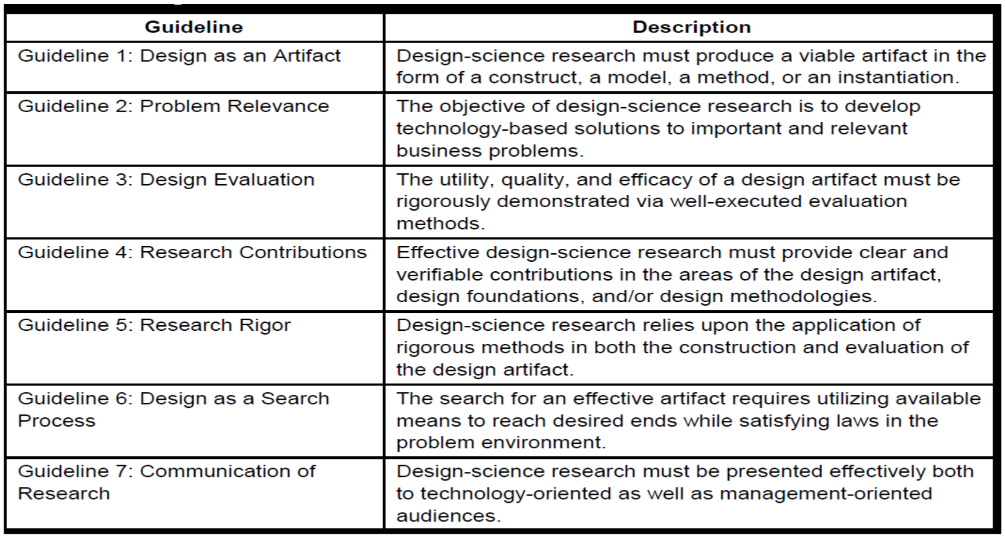
CHAPTER THREE

RESEARCH METHODOLOGY

Overview

This research is grounded in Design-Science Research (DSR) and adheres overall to all design science guidelines as outlined by Hevner et al. (2004). Hevner et al. (2004) assert that design science is concerned with the process of discovery through artifacts that have been created in order to address certain problems. Design science can be very productive, allowing for new and useful knowledge to emerge from the creation of an artifact. That knowledge can then make a significant contribution to the collective knowledge of the cybersecurity discipline. The authors proposed a framework that consists of seven guidelines to "assist researchers, reviewers, editors, and readers to understand the requirements for effective design-science research" (p. 82). Hevner et al. (2004) does not provide any explicit information about the order in which the following guidelines should be followed, so in this research the chronological order was followed as it produced the intended results.

Table 5 - DSR Guidelines from Hevner et al. (2004)



Guideline	Description
Guideline 1: Design as an Artifact	Design-science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.
Guideline 2: Problem Relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.
Guideline 3: Design Evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.
Guideline 4: Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.
Guideline 5: Research Rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.
Guideline 6: Design as a Search Process	The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
Guideline 7: Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

Note. Reprinted from "Design Science in Information Systems Research" by Alan Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram, retrieved from https://www.researchgate.net/profile/Alan_Hevner/publication/201168946_Design_Sc ience_in_Information_Systems_Research/links/5405d4670cf23d9765a75fc2.pdf. Copyright 2004.

Design as an Artifact

According to Barrett (2018), digital forensics as a discipline is "lacking the tools, published processes, and guidance for proper acquisition of digital evidence in cloud computing environments" (p. 1362). Thus, this research proposes a digital forensics acquisition methodology to help investigate cybercrimes related to the IaaS Public Cloud service model to complement most of the published digital forensic models. The methodology sheds light on the use of the existing digital forensic acquisition tools as well as the newly made available tools from the various CSPs. It helps fill the gap temporarily until CSPs provide support for reliable cybercrime investigations and upgrade their cloud-means to a forensic-aware cloud architecture.

Researchers have provided adequate starting points for cloud forensics as relates to acquisition. Zawoad et al. (2016), in their proposed Open Cloud Forensic model (OCF), provide directions for CSPs on how to perform reliable, continuous forensics acquisition. Alqahtany et al. (2016) proposed a novel cloud forensic methodology and covered the acquisition process in great detail. In both scenarios, the caveat is the dependency on CSPs to support the proposed solutions. In this research, the proposed methodology, IPCFA, does not involve the provider and is composed of multiple phases, aiming to ensure the *integrity*, *authenticity*, and the *operability* of the collected data and the produced evidence. Legally, this means fulfilling FRE 901 (Rule 901, 2011) and FRE 902 (Rule 902, 2011) and successfully navigating the Daubert standard (Ryan, 2009a).

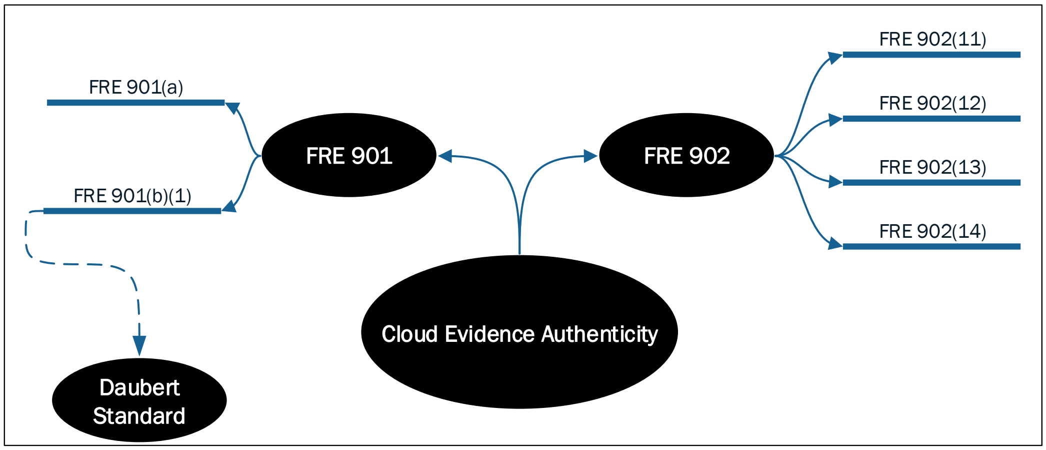


Figure 6 - Digital evidence and U.S. courts – Digital evidence admission map

Maintaining the *integrity* and *authenticity* of collected forensic data are the first two imperative requirements for court-admissible evidence generation. A third factor is the *operability* of the collected data, which refers to the ease of analysis and examination using commonly available and trusted forensics examination tools. The most critical aspects of forensically-sound evidence are reliability and completeness; if they are questionable, the evidentiary value is significantly weakened (McKemmish, 2008). According to Orton et al. (2013), “decisions in court cases rely on the *authenticity* and *reliability* of the evidence presented” (p. 186).

In the proposed methodology, the first requirement is to address the *integrity* problem by applying continuous integrity checks throughout the execution of process, starting with forensic server creation, and moving throughout various phases to the transfer of the collected data back to the forensics server. The second requirement, *authenticity*, which is related to maintaining a trusted and unbroken chain of custody, is mostly addressed by making the platform owner executing the process save the collected data directly to the final analysis destination to avoid creating untrusted copies. The same two requirements should be met by following FRE 901/902 to provide authentic evidence. *Operability*, the third and final requirement, is achieved by making sure the extracted data is always relevant, complete as relates to the request, and tool-agnostic.

Table 6 - Artifact Requirements Summary

Methodology Requirement	**Sub-requirements**

Maintain the integrity of the acquired data	- Maintain simple and trustful chain of custody
Maintain the authenticity of the acquired data	- Adhere to FRE 901 and FRE 902 - Withstand or moot the Daubert test
Generate reliable and operable forensic data	- Generate complete and uncorrupted data - Collected data must be certainly examinable

Problem Relevance

According to Hevner et al. (2004), a problem is the variance between a goal state and the current state of a system". The present research attempts to benefit the digital forensics community by helping practitioners acquire forensic data from public cloud instances, while having the confidence that it will be admitted and trusted in the U.S. courts of law. The security of organizational data in the public cloud is a shared responsibility, and distinction in the level of accountability is determined by the type of service model chosen. In an IaaS model, the responsibilities of the CSP and the consumer are weighted as equally important, which has significant impact on security. This shared security responsibility model brings a lot of challenges in how cybersecurity incidents are handled in the public cloud, specifically how digital forensics can successfully take place.

Though telecommunications companies have to comply with federal laws—such as the Communications Assistance for Law Enforcement Act (CALEA), which permits law enforcement agencies to use wiretaps—no such law exists for dealing with CSPs (Dykstra, 2013a). Access to direct authentic evidence collected by law enforcement officers is not an option when dealing with incidents in the public cloud. Traditional DF acquisition methodologies and techniques are by design not probable in cloud environments. The challenge is to innovate new and/or extend existing acquisition methodologies or processes that continue to comply with federal and civil laws and preserve the authenticity and integrity of the collected evidence, thus gaining the credulity of expert witnesses, jurors, and judges. Based on the existing literature review, and the published methodologies, techniques, and procedures, there are no formally written and adopted methodologies for acquiring forensic evidence from public clouds, more specifically from IaaS instances. There are numerous research papers focusing on SaaS deployments and how to investigate such crimes, while IaaS has always been overlooked and thought of as if the existing digital forensics tools and methodologies could be applied.

The only publication that directly addresses public cloud digital forensics is the SWGDE *Best Practices for Digital Evidence Acquisition from Cloud Service* (2020), which provides a very good starting point for cloud practitioners. The SWGDE best practice document does not have default processes to fit small businesses or organizations that do not have policies and standards to follow when it comes to digital forensics or incidents response. The document omits chain of custody and physical personnel who performed acquisition or interacted with the digital evidence or forensic data. The process also lacks reference to the cloud metadata and does not discuss the possible types of data that can be acquired from a CSP. Each type of data can require different acquisition mechanisms, and no single mechanism of the described methodology can capture all sought data. Yet this document serves as a base for the technical aspects of this research that is then enhanced and focused to address IaaS deployments in the public cloud realm.

Design Evaluation

The designed artifact of the research is a methodology, IPCFA, which includes numerous procedures to be followed and executed to achieve the desired results. Two hypothetical scenarios (cases) with a validation rubric are proposed. One scenario is hypothetically walked-thru while the second scenario will utilize an observational case study to demonstrate the effectiveness of the proposed artifact. According to Peffers et al. (2007), a case study can be used to demonstrate the use of the artifact to solve one or more instances of the problem. According to Ramirez, Mukherjee, Vezzoli, & Kramer (2015), using a scenario-based scholarly form of evaluation “helps to challenge existing assumptions and to identify novel lines of inquiry” (p. 70). The authors also argue that scenario-based evaluations can help produce research that is interesting, rigorous, and actionable (Ramirez et al., 2015). Choosing to validate the IPCFA in multiple public cloud service providers is essential to exhibit the independence of the proposed methodology that it is applicable to any CSP.

Following DSR guidelines by Hevner et al. (2004), the designed artifact evaluation in this research falls under Observational Design Evaluation Methods. The proposed methodology is evaluated in terms of *functionality*, *reliability*, and *usability* as relates to core requirements, which are *integrity*, *authenticity*, and *operability*, of the collected data. These evaluation criteria meet Hevner et al. (2004)’s definition of a well-developed artifact that is “complete and effective when it satisfies the requirements and constraints of the problem it was meant to solve" (p. 85). In this research the main constraints are the lack of involvement of the CSP during data acquisition and that the data collected should be able to satisfy the Federal Rules of Evidence as well as supports the plaintiff withstand the Daubert test. Two hypnotical scenarios (use cases) are assumed in two different and well-known CSPs—AWS and Azure. The IaaS deployments selected for the scenario are very common and have a high probability of a large number of implementations among large enterprises as well as small businesses today. The following two scenarios are used to evaluate the existing digital acquisition methodologies versus the proposed methodology:

Scenario #1 – Amazon Web Services (AWS use case):

Dakwa LLC is an online agricultural wholesale venue that allows importers, exporters, and direct buyers to buy, bid, and pay online for crops and raw oils. They are hosted in the public cloud and utilize various IaaS components such as virtual machines, databases, and network services. They have been set up to accommodate large spikes of traffic by utilizing native features of public clouds such as auto-scaling groups to meet the demand during large auctions or right before harvest sessions. Their cloud platform is set up as a 3-tier architecture to provide modularity, scalability, and high availability.

J0anneB, a hacker who was hired by a buyer transacting on Dakwa LLC platform, to dictate the competition and derive more revenue. J0anneB is very well versed in exploiting public cloud architectures, tools, and environments, and tends to exploit areas that are emerging such as productions pipelines. She also uses automation techniques of deploying and maintaining web applications and systems. She successfully launches her indirect attack and gains access to the transactions database, for which she is able to adjust many factors related to auctions and place bids for a long period of time, thus the buyer is able to deceptively make more financial gains. Some buyers and sellers have noticed the behavior and determined it is not normal, raising concerns to the company. The company now seeks to fix and investigate a fraud case, bring case to justice, and demand financial paybacks.

Assumptions:

- IaaS is hosted on Amazon Cloud (AWS) and is consuming IaaS native resources.

- The environment consists of six EC2 instances (two web servers, two application servers, two RDS servers). All servers are part of auto-scaling groups and load balancing pools.
- Infrastructure and code deployments are mostly through pipelines and various DevOp processes. There is no use of serverless or function-based applications.
- Dakwa LLC has an IT security department but they lack digital forensics expertise, thus they decided to hire an attorney to bring forensic experts along to investigate.

Scenario #2 – Microsoft Cloud (Azure use case):

Oba7 is a hacker who sells distributed computing CPU cycles to underground communities for cryptocurrencies miners, DDoS attackers, and other malicious-intended purposes. He is an expert in utilizing exploit kits to infect legitimate websites with potential high traffic to distribute malware to visitor's workstations. His malicious code renders to regular webpages, gets downloaded behind the scenes onto clients' workstations, and then executed to connect back to Oba7's command and control server. He was able to acquire the Azure Public Cloud account information and credentials of Zool Corporation, which owns and operates a small-town casino. The hacker accessed the associated cloud account and used the readily available computing and storage resources to allow anonymous buyers to borrow resources in exchange for cryptocurrency. The buyers pre-transfer the cryptocurrency amount that translates to the desired hours/resources to be borrowed, then come to the website, place a single text file with the transaction hash and few other encrypted details, and leave. Oba7 utilizes the power of automation and ephemerality of resources available in the public cloud to automatically obliterate all relevant text files right after being accessed and downloaded to the local workstation. Thus, the virtual machine and the storage are always in a clean state or even do not exist.

Zool Corporation's accounting department noticed a major increase in their recent cloud utilization bills. They reached out to the CSP and complained that they have been over overcharged for several months. The CSP advised the corporation to review their accounts and their utilization trends in order to reduce their resources consumption. The corporation is determined to investigate their cloud account misuse, find the root cause behind their overages, and attempt to legally ask for settlement.

Assumptions:

- IaaS is hosted on Microsoft Cloud (Azure) and is consuming IaaS native resources.
- The environment consists of six virtual machines (three web servers, three application servers, two Azure SQL servers). All servers are part of availability sets and are behind load-balancers.
- Infrastructure and code deployments are mostly through pipelines, scripts, and various DevOp processes. There is no use of serverless or function-based applications.
- Zool Corporation has a cloud security analyst's team, but they lack digital forensics expertise, thus they decided to hire an attorney to bring forensic experts along to investigate.

An observational case study will be conducted on the Azure scenario for demonstration purposes, while the AWS scenario will only be hypothetically walked through and compared to various existing recommendations and DF processes from the following publications:

1. Internet Engineering Task Force (RFC#3227) (Killalea & Brezinski, 2002)
2. Capture of Live Systems (SWGDE, 2014)
3. Best Practices for Computer Forensic Acquisitions (SWGDE, 2018a)
4. Best Practices for Computer Forensic Acquisitions from CSPs (SWGDE, 2020)
5. FRE 901 - Authenticating or Identifying Evidence (Rule 901, 2011)
6. FRE 902 - Evidence that is Self-Authenticating (Rule 902, 2011)
7. Daubert Standard (Ryan, 2009a)

Given the above resources, a small focus group will be carefully assembled in order to walk-thru the AWS hypothetical scenario and provide all the steps required in order to collect forensically-sound data to support the Dakwa LLC case. Assembling a small focus group is important to understand the capabilities of the current practitioners in the market and determine any application gaps. They will be instructed to follow the SWGDE (2020) acquisition process in addition to their expertise and personal knowledge of U.S. legal systems. In order to recruit five highly skilled digital forensic practitioners with legal and public cloud practical knowledge, the following selection criteria are required:

- A U.S.-based and currently practicing digital forensics senior examiner with more than five years of practical experience.

- A holder of valid and relevant U.S. technical and legal digital forensics credentials (education, certified training, experience).
- Practical understanding of the federal and state rules of evidence and their applicability to digital forensics.
- Working knowledge of digital forensics expert witness legislations and U.S. court systems.
- Working knowledge and valid credentials of working with the existing public cloud ecosystem (hands-on with architecture, deployment models, and service models).

The main task of the forensic investigation is to adhere to the above-mentioned resources and to provide the projected steps taken to acquire cloud forensic information that can be used to generate court-sound evidence. The acquired cloud forensic data and process must prove *integrity*, *authenticity*, and *operability*. After both hypothetical cases have been attempted with the existing methodologies and procedures from the common body of knowledge, the newly proposed methodology will be executed, and the results compared to the outcome of previous attempts. The following investigative outcomes are expected from the collected data:

- Determining the chronology of the attack.
- Identifying the source and scope of the impact of malicious activity.
- Uncovering the origin of the attack and tie back to possible entity.
- Enabling the client to prosecute the attacker(s) and suspected buyer in courts of law.

According to the University of Texas, a rubric is a “scoring guide used to evaluate performance, a product, or a project” (Build a Rubric, 2017, p. 1). A rubric is a criterion-referenced assessment tool (O’Reilly & Cyr, 2006). In this research, a customized Admissibility Likelihoods Rubric (ALR) is developed from the federal requirements for digital evidence and the digital forensic acquisition best practices from the various institutions, technological standards, and regulatory bodies. The added value of this rubric is that it bridges the gap between the legal requirements and technological requirements, and thus simplifies the decision making from technological perspective.

The rubric presented in Table 7 is used to evaluate the results of executing the methodology and validating the generated data authenticity, thus determining the probability of being accepted as admissible evidence in the U.S. judicial system. The levels of performance used in the rubric are *Trustworthy*, *Doubtful*, and *Untrustworthy*. Each performance level has been assigned a score. These performance levels are not meant to determine actual admissibility in U.S. courts, but rather acts as trusted probability vector indicating whether the court is more or less likely to accept the evidence. They also provide a clear indicator if the forensic methodology followed can withstand defense scrutiny. ALR evaluation is based on the overall scope of admissibility probability out of 90 possible points. This evaluation is subjective; thus, no given score value can determine 100% admissibility in a trial. The highest the number of points on ALR predicts a high level of trust in the forensic process and the generated evidence, thus a higher chance of admissibility in federal and state courts. The opposite is also true: if the ALR score is low, the lesser the chance that evidence will be rendered admissible.

Table 7 - Digital Evidence Admissibility Likelihood Rubric (ALR)

Criteria	**Trustworthy (6-10)**	**Doubtful (2-5)**	**Untrustworthy (1-0)**
General/Legal			
1) Adopted a structured and published forensic acquisition method or process	A structured acquisition process has been followed, and it is a tested and published process.	Acquisition process has been followed without a reference to published processes.	No acquisition process was followed or referenced.
2) Forensic acquisition practitioner certified abilities	Acquisition performed by documented, certified, and knowledgeable examiner.	Acquisition performed by undocumented but knowledgeable examiner.	Acquisition performed by inexperienced and undocumented personnel.
3) Trustworthy capturing of the whole acquisition process	The acquisition process was captured and can be clearly referenced to examine and verify or repeat each step of the process, if required.	The acquisition process was captured but cannot be clearly referenced to examine and verify some steps of the process.	The acquisition process was not captured or was unclearly captured and cannot be reproduced.
4) Provided case-supporting artifacts	All evidence-supporting logs and events from all platforms were fully captured and preserved properly.	Some evidence-supporting logs and events from various platform were captured or fully captured but	No evidence-supporting logs and events were captured nor preserved properly.

		not preserved properly.	
5) Very well documented and validated chain of custody	Chain of custody has been fully documented throughout the acquisition, preservation, and access processes.	Chain of custody has been partially documented but not covering all phases of the DF process.	Chain of custody has been poorly documented or completely undocumented.
Data acquisition			
6) Used trusted acquisition tools	The tools used are tested and trusted (compiled from trusted source), CFTT report is available, or a well-known tool was used whose integrity is validated from vendor or maintainer.	The tools used are trusted, but the trust has not been validated and documented prior the execution.	The tools used are not well-known and their integrity were neither validated nor documented.
7) Data is captured in operable format	Data is captured and saved in raw or known format and can be examined using common and trusted tools.	Data is captured and saved in a proprietary format that is widely used. The data can be examined using few proprietary tools.	Data is captured and saved in a format that is not commonly used. The data cannot be examined using trusted tools.
8) Utilized secure and immutable evidence storage	Storage is defined, sterilized, secure, isolated, and immutable with limited network and access.	Storage is defined, secure, and isolated with limited network and account access.	Storage is not defined or is partially defined and is not isolated or secured nor immutable.
9) Validated the captured forensic data	NIST-approved hash algorithm value calculated for all the captured evidence data files. All hashes comparisons were accurate.	NIST-approved hash algorithm value calculated for some of the captured evidence data files. Some hash comparison values were accurate.	NIST-approved hash algorithm value not calculated for the collected evidence files, or incorrect values were returned during hash comparison.

As rubrics by design are objectively deriving the evaluation and grading of the objective are usually matched with the pre-defined objective of each criterion. The ALR consists of nine criteria. Each one has been derived from the core requirements that enhances the admissibility of digital evidence in the U.S. Court system.

1) Adopted a structured and published forensic acquisition method or process

Article VII of the Federal Rules of Evidence governs expert witness testimony requirements. FRE 702c and FRE 702d (Rule 702. Testimony by Expert Witnesses, n.d.) both require the expert witness to apply a defined methodology to the artifacts presented to provide facts. Thus, using a pre-defined methodology to perform the digital forensic acquisition is an important factor in providing a reliable testimony in courts of law. While attempting a forensic acquisition based solely on personal knowledge of the matter might be sufficient in some situations, this is ultimately not a reliable approach if a trial is expected.

- This criterion receives ten points if the forensic examiner has followed a published methodology that addresses the acquisition nature of the specific case. The steps taken to achieve the desired goal must be documented prior to executing the observed methodology.
- Five points are given if a methodology that has been followed is not published to the knowledge base or has been adopted from another form of acquisition that is not directly applicable to the investigated case environment. If the examiner created and followed their own acquisition process, it is up to the court to determine if the acquisition process followed can be used to generate sound evidence. The court might invoke a Daubert pre-trial testimony of expert witnesses that thus can be challenged in court.
- The lowest number of points—one point—is expected if the forensic practitioner did not follow any process neither published nor created for the purpose of the referenced acquisition. This weakens the credibility of the derived evidence, diminishing its value in court significantly. Furthermore, it renders the evidence not compliant with Federal Rules of Evidence 702c and 702d.

2) Forensic acquisition practitioner-certified abilities

This is another essential criterion that revolves around solidifying the credibility of the collected evidence as well as enduring the probable Daubert expert witness testimony. If the acquisition is carried out by a court-appointed expert or by law enforcement agency, this requirement is automatically satisfied prior to hearing by following FRE 602 and FRE 706, but in this case the forensic process is carried out by private investigators. When entering into a trial, judges usually determine the credibility of expert witnesses by executing the Daubert test (Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 1993). Prior to a Daubert pre-trial, all experts presenting evidence in the court must submit a summary report that includes their analysis and findings. This report is shared with both parties, which allows for cross-examination of witnesses. The submitted report must follow the FRCP Rule 26 (2015), which explicitly requires the report to include subject matter expertise of the witness, experience, education, special trainings, industry certifications, publications, and similar cases in which the expert testified in the past four years; this is not an exhaustive list of qualifications.

- The full ten points are possible if the acquisition was performed by a digital forensic expert who has previous relevant and in-depth experience in the subject matter. Having served as an expert witness in similar cases adds greater trust to the expert testimony. Experience must be documented and verifiable such as authored publications, court-case appointment letters, training certifications, employment duties and assignment letters, and formal relevant education diplomas.
- Five points are awarded if the expert has knowledge and experience but is not properly documented and verifiable; thus, the expertise can be disputed. This also applies to experts who are new to the U.S. court system and have limited or very recent digital forensic knowledge.
- If the acquisition was performed by an inexperienced and/or completely undocumented examiner, then the lowest score is given.

3) Trustworthy capturing of the whole acquisition process

Making the case stronger, admissible, and indubitable is the main driver behind this criterion. While the DF acquisition process can take long hours or days depending on the case, being able to show a great amount of transparency in the forensic process can be very rewarding when evidence is presented in courtrooms. While it might not be possible to capture the whole process of a very large file being copied between drives, showing the initial and finish state of the transfer with timestamps is possible, and this makes a difference. Judges, juries, and defense attorneys having a clear understanding of how evidence was collected can remove a lot of questions and help the trial move forward.

The referenced publication does not qualify the type of notes if it is written or recorded or in any other format, but it emphasized on the detail level of the notes and explicitly requires notes to be detailed to a point where the whole process can be duplicated just by following the provided notes. According to the National Institute of Justice special report (U.S. Department of Justice; Office of Justice Programs; National Institute of Justice, 2004), taking very detailed notes of the acquisition process from start to finish is crucial to having complete case information that is presentable in court. When looking at digital forensic evidence requirements from an international perspective, the task force formed between SWGDE and IOCE (2000) approved the recommendation of having detailed documentation of all actions taken during the acquisition process.

- Ten points are awarded if the acquisition is documented with detail that allows for inspection of each step of the process as well makes it possible to reproduce the whole process and the expected results. This can be done with written notes, screenshots, screen recording, or a combination of methods as long as the results provide a full capture of the process.
- Five points are awarded if the documentation process does not capture the complete forensic acquisition. This can be due to some acquisition steps being omitted or notes not having enough detail to carry out the process based on the provided notes only.
- If the acquisition process was not documented at all, or was documented at only a very high level, then the instance receives one point.

4) Provided supporting case artifacts

This criterion encompasses two types of supporting artifacts: the first includes artifacts that support the acquisition process itself; the second supports and ties evidence back to a claim. While all provided data and evidence can be argued upon in a court setting, having supporting information about the presented data can be the differentiator between relevant and admissible evidence and inadmissible and rejected evidence. In IaaS acquisition cases, the main data sources are the network, the disk drives, and the ephemeral memory storages such as RAM. Collecting data duplicates from these sources can be done in many ways and can be documented.

Providing information that is originally generated by software and processes used during the acquisition process can have a very high value in courts. Important artifacts that are always overlooked are metadata about all forensic data and evidence. The most recent change in Federal Rules of Evidence and the introduction of FRE 902 (Rule 902, 2011) provides guidance for self-authenticated information that can be admitted in courts without further investigation. FRE 902(11) states that electronic records generated by systems or electronic processes can be considered self-authenticated. The same applies to the second case where logs, events, or metadata generated during the investigated activity are collected and provided to support the advocated claim. Being able to provide supporting logs to validate IP activity on the network, for example, is one way of providing self-authenticating evidence.

- A full ten points are awarded if supporting artifacts such as logs, events, and any other auto-generated information are captured, preserved, and provided alongside the documented acquisition process. Also, it is required that any extracted evidence has supporting artifacts to validate the claimed evidence.
- The second highest number of points (up to five points) is given if supporting artifacts are generated for some processes or evidence, but not for all important forensic or acquisition tasks. Also, the same number of points can result if all artifacts are captured by not preserved properly.
- One point is awarded if no supporting artifacts are captured or presented in conjunction with the forensic process documentation and the collected data or evidence.

5) Very well documented chain of custody (CoC)

Chain of custody is the heart of the whole forensic process, and without it any collected data loses its evidentiary value. Maintaining a complete, clear, and undoubted CoC by ensuring that any and every action taken during the forensic process is documented contemporaneously is crucial. According to NIST SP 800-101, the most important information to include is “each person who handled the evidence, the date/time it was collected or transferred, and the purpose for any transfers” (Jansen & Ayers, 2004, p. 67).

While the information required is considered for all types of forensic evidence, digital evidence requires more information to be documented in order to be complete. Additional information includes the collection method, storage location and type, transport medium, integrity checking hashes for each item, and access tracking and purposes.

- This criterion receives a full score if the CoC is provided, and it has all the information required in order to assure the evidence integrity from collection to presentation at the trial without skipping any part of the forensic process. Information about the creation and preservation of the CoC itself is included as well.
- Five points is given if the CoC is provided, but it has minimal information about personnel, evidence collection, transport, and access, or the CoC is not comprehensive and has a few gaps in one or more of the forensic process phases.
- 1 point is given if the CoC is not documented or provided or was provided but not documented properly, such as with missing dates, personal information, or process information.

6) Used trusted acquisition tools

The DF process usually involves executing a plethora of tools starting from case initiation and management going through acquisition and all the way to presentation software. While all involved tools share the same importance in the process, the acquisition tools are thought of as the most critical tools to consider as they are used to collect and generate the actual evidence for the investigated case. According to NIST Computer Forensics Tool Testing Program – CFTT (Allen, 2017), law enforcement agencies are severely lacking computer forensic tools that are trusted to generate reliable data. The program was established in order to provide a testing methodology and a database to categorize tools, their functions, and the quality of the generated results. Tools that are tested and validated under CFTT are considered credible and can generate the intended results if used properly. Each tool has a corresponding report that shares all the details of applying the CFTT testing methodology. While the NIST database does not have tools that are designated for DF in the IaaS public cloud, some listed tools can be re-used in the context of IaaS investigations. While trust in the well-known tools is presumed, this trust can still be challenged in courts of law, thus it is very important to come ready with defensible arguments about any and all tools used throughout the forensic acquisition.

- The highest score (ten points divided equally amongst all tools) is given if the tool used during the acquisition is either 1) compiled from known good source code and hash values that are tracked before and after execution; 2) the tool is certified with CFTT and has a valid and relevant report, and the hash values of executables are validated; 3) the tool is well-known to the forensic community with a verifiable vendor or maintainer and executable hash values. Some other tools not designed for forensic analysis such as web browsers, remote access scripts, graphical-user interfaces, command-line interfaces, and direct API callers that require authentication must be validated by proving authenticity of the connection and the access granted.
- Five points (divided equally amongst all tools) can be achieved if a tool is trusted per the previous criterion, but the trust is not validated before executing the tool, or the authenticity of non-forensic tools was not captured and documented.
- One point (divided equally amongst all tools) is given if the forensic examiner utilized tools that are not well-known and not trusted,

7) Data is captured in operable format

Digital evidence might be the deciding factor in a criminal case, but if it has not been stored and preserved in appropriate format it becomes unusable (Garfinkel, Malan, Dubec, Stevens, & Pham, 2006). In the context of IaaS forensics, one might think data formats are more of a concern when collecting disk and memory images, but appropriate storage and preservation matters for all relevant collected data. In public cloud forensics, the acquisition is logical, thus a lot of metadata is acquired alongside disk and memory images. Metadata is one of the most important aspects of acquiring data from the public cloud, as it contains most of the fingerprinting information that can be referenced at any stage during the investigation. All collected data must be operable, and disk images need to be complete and verifiable; the same applies for memory images, individual files, and screen captures. Without DF data in the right format, extracting quality and court-sound digital evidence is unmanageable.

- The highest score is given if all collected files are stored in their respective operable formats, and they can be examined to the fullest. The stored files must be verifiable, complete, and electronically fingerprinted.

- The middle score is given to data that has been collected in uncommon formats that are still operable using uncommon or questionable tools and techniques. The data still needs to be verifiable in order to maintain integrity.
- The least score is given if the data collected is not operable by any trusted means as well as not fingerprinted, and no hashes are calculated when the data was generated.

8) Utilized a secure and immutable destination for captured data

This criterion is a major differentiator between authentic and trusted forensic acquisition, and a doubted acquisition. Forensic data is fragile and unstable by design, and thus brings a lot of debate in federal or state courtrooms (Dykstra, 2013a). Being able to demonstrate that the acquired data has not been changed and could have not been changed since it was saved to the storage media is a vital winning argument during a trial. Evidence doubts from the defendant can range from questions about the sterilization of the storage media all the way to the appropriate use of write blockers to attain credible forensic data. The native architecture of cloud computing opens opportunities for digital forensics examiners to utilize the elasticity, modularity, and security functions of the cloud to attain a high level of evidence preservation until prosecution date.

- The highest score is given if the forensic data is stored in sterile, secure, isolated, and immutable storage. While all these aspects are very important, immutability (single write, multiple reads) is the one that is emphasized as it can stand both the authenticity and integrity arguments in courtrooms. This can be thought of as having the originals stored securely, while copies can be made for investigation and analysis.
- The second highest score can be awarded if the evidence is stored in a secure storage that is not isolated nor immutable, but the security of the storage can be verified, and the integrity of the files can be maintained. This can be argued in court and can be accepted as evidence, but it can also lead to evidence dismissal.
- Storing forensic evidence and data in shared storage media that cannot validate the integrity of all files receives the lowest score possible.

9) Validated the integrity of the captured forensic data

This criterion has been touched upon within multiple criteria within the ALR. Having authentic evidence means the evidence was collected in a trustworthy way, persevered until presentation with no alterations (NIST Cloud Computing Forensic Science Working Group, 2014). The Advisory Committee behind FRE 902(14) describes the process of the "digital identification process" as the utilization of special software to compare the hash values of two documents. This process can render the document authentic in court. The committee also states that the new rule is flexible enough that it allows "certifications through process[es] other than comparison of hash value, including by other reliable means of identification provided by future technology" (Federal Rules of Evidence, 2017, p. 14). While this statement opens the door for other means of validation such as witness testimony, generating and authenticating and evidence with a hash value is the ultimate means of identification today.

- Up to ten points is given if forensic data collected can be identified and validated using a hash algorithm that is approved by NIST, such as SHA-1, SHA-2, or SHA-3.
- The second highest score can be awarded if most of the data collected can be identified using a hash algorithm that is approved by NIST such as SHA-1, SHA-2, or SHA-3. Moreover, most evidence files must have correct hash values when compared during trial.
- The lowest score is given if hash values are not calculated, or a non-approved hash functions is used, or the hashes are calculated but do not match evidence original hash values during trial.

While there seems to be a lot of overlap between the various criteria proposed in the ALR, each criterion is centered on specific qualities of the captured forensic data that supports the admissibility claim of the presented evidence. Similarly, the proposed evaluation and validation scenarios, while they are fictional, describe deployments and cybercrimes that are ubiquitous today. The focus of executing the proposed methodology is two-fold: first is to perform the forensics acquisition using various existing and new tools and techniques following pre-defined steps defined, and second is to apply the ALR to evaluate the overall admissibility probability if evidence is to be presented in a U.S. court. Digital evidence can always be challenged in court, and so the disposition is to provide evidence that is systematically sound enough to be assumed highly authentic and comprehensively provocative to be disputed.

Research Contribution

When compared to other prevalent research methodologies such as natural or social science, design science research tends to be more specific on the expected outcome and contribution of the research. According to Hevner et al. (2004), there are three possible outcomes expected from DSR grounded research, which are the “novelty, generality, and significance” of the designed artifact. Any given artifact of design science research must meet one or more of these three outcomes. Furthermore, the artifact should be developed to address a specific problem in a business environment not previously resolved; this can be done by expanding on an existing knowledge to resolve a problem or use existing knowledge in a new and innovative way to resolve a problem (Hevner, 2004).

In this research, the artifact is a practical methodology to acquire court-admissible digital evidence from public cloud IaaS deployments. The three major requirements that makes this research significant and novel to the cybersecurity community are the following: 1) It helps enterprises and small business forensic teams in acquiring court-admissible forensics data from the major public cloud IaaS deployments; 2) No involvement of the cloud services providers throughout the DF process is required; and 3) The acquired data and evidence meets almost all U.S. governing rules and regulations related to digital evidence. The methodology builds on the existing knowledge base of digital forensics (models, methodologies, mechanisms, and tools) and extends the knowledge to include DF acquisition in public clouds IaaS environments, expands best practices for acquiring U.S. court-admissible digital evidence, and, as a byproduct of this methodology application, sheds light on the capabilities and limitations of the existing DF tools as well as the not-so-popular in the digital forensic community, the cloud-native Application Programmable Interfaces (API).

Research Rigor

Research rigor in design science “provides past knowledge to the research project to ensure its innovation” (Hevner, 2007, p. 90). Rigorous research is highly dependent on the researcher performing a deep and thorough literature review of the past knowledge base related to the researched topic, which helps to guarantee the research outcome is a design science research and not a routine design (Hevner, 2004). Furthermore, Hevner et al. (2004) argue that the application of rigorous methods should be applied in both the construction of the artifact as well as the evaluation. In this research, rigor is addressed in both phases.

Prior to the design of the artifact, the preceding research and knowledge base related to digital forensics methodologies, models, and technical procedures are thoroughly investigated, studied, and categorized. Then, the same is performed for cloud forensics in order to address challenges and proposed solutions presented by researchers as well as the cybersecurity community. During the construction of the artifact, the knowledge acquired from the past knowledge base is considered; in this case, this includes the various methodologies of how digital forensics is being conducted today and of how the digital data and digital evidence are being acquired.

The proposed methodology inherits the existing digital forensics methodologies, best practices, and available tools of acquiring evidence from non-cloud-based environments. It is an improvement to the acquisition phases proposed by the following researchers, standards organizations, and governance organizations, to enable public cloud DF acquisition:

- Scientific Working Group on Digital Evidence (SWGDE, 2018b)
- Scientific Working Group on Digital Evidence (SWGDE, 2020)
- Internet Engineering Task Force (RFC#3227) (Killalea & Brezinski, 2002)
- NIST special publications (Kent et al., 2006)
- U.S. Department of Justice special reports (NIJ, 2008)
- An integrated conceptual digital forensic framework for cloud computing (Martini & Choo, 2012)
- Digital Forensics Processing and Procedures (Watson & Jones, 2013)

While there are a lot of publications deliberating and trying to address digital forensics from different perspectives, authors of the abovementioned publications propose recommendations and procedures to acquire forensically-sound data from electronic media while keeping the possible legal challenges in mind. According to the literature review, besides the recent SWGDE, 2020, there is no other work published that guides forensic examiners or incident responders on how to address acquisition in public cloud related cases. Adams (2013) too includes a small section discussing acquisition from cloud environments but does not go into the details of how the acquisition could take place taking into consideration the elastic nature of cloud computing resources. The authors of RFC#3227 (Killalea & Brezinski, 2002) were able to create the most relevant process to perform a trusted forensic acquisition that the new methodology proposed in this research builds upon. The artifact in this research, IPCFA, expands on these methodologies to accommodate cloud computing cases, more specifically public cloud IaaS-related investigations. The new methodology proposed in this research takes into consideration possible limitations of performing forensic investigations in the realm of the public cloud today.

The proposed methodology correspondingly stems from the knowledge base of digital forensics models, methodologies, and procedures. The evaluation consists of a scenario to be executed; the methodology is applied and the outcome captured, then validated against the metrics composed in specially-designed rubric that encompasses all best practices of DF acquisition as well as the U.S. Federal Rules of Evidence (Federal Rules of Evidence, 2017). Included are the extra metrics, guidelines, and recommendations produced and validated while experimenting with public cloud acquisition throughout this research process.

Design as a Search

Design science is naturally iterative, and iterations of search are inevitable to solving problems in information systems. Solving a problem following DSR requires utilizing all available resources (*means*) to construct the desirable solution following the presented constraints (*ends*), while navigating the uncertainty in the environment (*laws*). As means, ends, and laws are refined and made more realistic, the design artifact becomes more relevant and valuable (Hevner, 2004).

The proposed methodology, IPCFA, is motivated by performing digital forensic acquisition in the public cloud, and more specifically in two of the major public clouds available today, Amazon's AWS, and Microsoft's Azure. Yet, it can represent neither all possible CSPs nor all possible incidents that could occur in the realm of public cloud. The intent of this research is to address incidents related to their IaaS compute engines where forensic data can fundamentally reside on the host, memory, attached disk images, CSP logs, and other security logs. The methodology is generic by design to apply to as many existing CSPs as possible and to allow for future adoption and alterations to fit various other use cases. The same limitation applies to the deployment models of the public cloud, as this research only attempts to solve digital forensics challenges related to Infrastructure-as-a-service deployment. The proposed solution is modular and consists of numerous phases to allow for future research extensibility, modification for adoption, and unit testing.

The research question has been formalized after performing significant literature review and technical experiments related to cloud forensics. The design starts by performing a search on the relevant knowledge base to determine all accessible resources (*means*) that can be utilized to facilitate solving the presented research problem. The global footprint of public clouds is looked at, more specifically their presence and adoption in the US, and their published work is analyzed to understand what they offer in terms of digital forensics capabilities, subpoena request abilities, and their openness and cooperation with law enforcements. All the encountered DF federal and international rules, models, methodologies, theoretical publications, commercial and open-source tools, and techniques are considered part of the resources that the research builds upon to develop the methodology. This phase of the search helps determine the initial direction for which the search efforts is focused to solve the problem.

After all the means have been determined and categorized, the next step is to search constraints, limitations, and effectiveness of the possible solutions proposed by the digital forensics research community in the literature (*ends*) that might help solve the same problem presented in this research. The final goal is to solve the problem, navigate the existing limitations, and come up with a working solution that satisfies the constraints. The constraints examined are related to the Cloud Services Providers' terms of service related to subpoena, the data exposed via API calls, and various data sources and granularity levels provided by the CSP for customers. The other major set of constraints that limit this research design is the U.S. Federal Rules of Evidence, as well as privacy rules that govern personal identification information exposure. On the technical side, the limitations are to perform the acquisition without involving the CSP, to avoid changing the forensic data by the invoked forensic acquisition tool or technique, and to ensure the data collected gets transmitted securely to the destination, the forensic examination workstation.

While the U.S. federal and civil rules governing the admission of digital evidence in courts appear straightforward, in actuality they are not; and all aspects of the digital forensics case can be challenged during a litigation (*laws*). All produced evidence can be argued not trustworthy, and the mechanism for acquiring evidence can be tested with assessments such as the Daubert standard. The natural workflow of trials and their uncertainties bring challenges when attempting to solve problems related to digital forensics in the cloud. In this research, the final solution must stand up against the Daubert test, and the produced evidence is designed to become self-authenticated to an extent in order to attain trustworthiness of U.S. court systems.

Communication of Research

As with other information systems related artifacts, the technical solution is just a part of the big picture, and understanding the surrounding issues related to processes, standards, governance, and other relevant laws and regulations is key to the successful implementation and practicing of the matured artifact. The technical specifics of the proposed solution are detailed and broken down into various phases to ease its implementation and facilitate easier adoption. While this research is heavily focused on constructing the practical methodology itself, the big picture of the problem and the existing challenges related to performing successful digital forensics in the public cloud are explained with reference to the new methodology.

The evaluated and generated DF processes, skillset, recommendations, and other required knowledge are highlighted and focused on in order for organizational leadership teams to determine how and why they can adopt the newly developed artifact, and how it differs from the existing solutions. The research and associated artifact will be published and made available via the Dakota State University Beadle Scholar system. This initial publication will make the research available to all 550+ Institutions that use the Digital Commons Network. The intent for this research is to make it available on as much as possible of the online digital libraries to be available in the global knowledge base of digital forensics.

CHAPTER FOUR

IPCFA - IAAS PUBLIC CLOUD FORENSIC ACQUISITION

The technical details and internal deployments of the various public clouds might differ, but they all still share the same high-level characteristics, and almost all offer IaaS. In IaaS, the owner can customize and start up virtual machines (computing instances) and containers (computing pods), install the desired applications, harden as needed, permit the necessary communication flows for inbound and outbound access, and manage the lifecycle of the created instances. The proposed acquisition methodology assumes there is a need to acquire as much forensic data as possible to extract court-sound evidence that supports the litigation claim. The investigation involves one or multiple compromised virtual server(s) that is part of an IaaS environment. The investigation could be led by the owner or a delegate of the affected platform incident response team, while keeping in mind that litigation is always an option. The core concepts that make up the guidelines for the proposed methodology (IPCFA) are depicted in Table 8.

Table 8 - High-Level Guidelines for IaaS Cloud Evidence Acquisition

Principle	Description
1) Digital seizure	When possible, preserve evidence source, and isolate the server to a secure location where the only permitted access is from the forensic examiner workstation; in order to minimize additional changes and lessen the impact on the investigated server, the server must be isolated to a secure location.
2) Iterative collection	Expect cycles of collection because case-relevant data may be located in many places in the cloud environment. While the collected data may be analyzed on the forensic server, additional data might be required to construct a timeline or uncover a relevant event.

3) Secure data transfer	If data transfer is required, transfer the data using the appropriate transfer mode for the file type (binary/ASCII). This makes sure the data is transferred bit by bit without losing information. The file transfer should be encrypted or local between the forensic workstation and the examined server/destination storage.
4) Audited Acquisition	Capture all commands (log/output) throughout the process of the acquisition on the forensic workstation. These logs are provided alongside the collected evidence as proof of the acquisition process followed for documentation purposes. This step is crucial to assure the integrity of the process.
5) Trusted forensic tools	Use tools compiled from the source prior to being used, capturing the calculated executable hashes via the continuous auditing mechanism, or tools that are well known and trusted within the community (preferably certified under NIST CFTT) or natively provided by the provider authenticated interface. The authenticity of the tools used must be validated and captured via a continuous integrity mechanism.
6) Data and evidence integrity	Calculate the generated output files hashes after the execution of every one of the required tools on the compromised server. The hash value must be captured and referenced via a continuous auditing mechanism.
7) Data and evidence validity	Check all collected evidence with analysis tools to validate the hash value, the completeness of the acquired data, and the possibility of gathering the needed information from the collected data. This is important because collected data can be corrupted, inoperable, or irrelevant to the ongoing investigation (for the purpose for which it was generated).

8) After-the-fact integrity checks	Once files have been transferred to the forensic server, check the hashes for all collected files against the original values. If confirmed, then forensic examiner can carry on the forensic analysis.
9) Immutable storage	Save all artifacts generated from the acquisition process directly on a write-once read-many storage. This requirement helps ensure the integrity of all artifacts and evidentiary data throughout the forensic lifecycle.
10) Near-event forensic examination and analysis	Store evidence near where the data was collected, and perform as much as possible of the forensic examination in an isolated network on the investigated public cloud. Since the incident occurred in the public cloud and data acquisition took place in the public cloud, it makes sense to isolate the evidence in order to avoid mishandling of the evidence. This measure also helps greatly to reduce the amount of data that must traverse the network between the customer network and the cloud services provider network. This requisition helps reduce the chain of custody.

The proposed methodology, IPCFA, has three major phases: initiation, execution, and closure. Each phase is composed of multiple processes to be executed. Each phase supports one or more of the guidelines in Table 8, indicating how this phase can be executed to help generate forensically sound evidence. Some of the proposed phases might be labeled as *optional* or *iterative*, and they should be determined as required or discretionary during the preparation phase of the forensic process. *Iterative* means this step can be re-executed in the future to collect more relevant data, as the investigation progresses. The order of execution of these phases is highly dependent on the type of incident, the purpose of the investigation, and the cloud provider-enabled capabilities. The specifics of each phase are listed in detail as follows:

1. **Initiation phase**

1.1. Stand up a forensic server

Why? A trusted forensic server contributes to the authenticity claim of the generated data by making sure the forensic workstation is built from a trusted and pre-built image with all required tools pre-installed. This step also involves setting up security policies, forensic accounts, isolated networks, and relevant storage. This removes any distrust related to where the collected forensic data are being sent and saved for analysis.

How? Collecting data directly and securely saving the data to the forensic workstation with immutable storage can add further credibility to the acquired data's integrity and confidentiality. The forensic workstation needs to be instantiated from a pre-configured and hardened cloud image in an isolated network with separate security policies and rules. Appropriate access needs to be given to allow only the forensic examiner's remote machine to connect to the specific required ports to operate on the workstation. The whole process of setting up the forensic workstation and enabling the required restricted communications can be scripted and automated to be executed seamlessly when needed. In some instances, the cloud platform might offer cross-account storage and volume mounting; in such cases, it is recommended to initiate the forensic workstation in a dedicated account to avoid inquiries about the overall security of the account running the compromised instance.

1.2. Enable continuous integrity monitors

Why? Enabling continuous integrity monitors supports the authenticity claim of the evidence by ensuring the complete transparency of the acquisition process as a whole, including the collected data and the forensic examiner actions and interactions with the various tools. It also plays a vital role in ensuring unquestioned chain of custody.

How? This phase sets the stage for the whole process, so it must be executed properly. As one of the major sub-phases in this methodology, it has direct impact on the authenticity, thus the admissibility, of the acquired data. When the preparation phase of the chosen digital forensics model is complete (that is, the forensic examiner is identified and furnished with an appropriate and secure workstation, the forensic data sources have been identified, required documentation has been approved and prepared, the preferred tools to perform this acquisition have been chosen, and the forensic examiner is ready to collect and acquire data), then this phase of the methodology begins.

It is not a default behavior for all public cloud components to generate and keep logs, thus giving the consumer various options to which logs to collect and how and where to store them. As logging in the public cloud is shared responsibility, it is important to make sure to enable logging if not already enabled on the Cloud Console or management plane and cloud audit/action logs before engaging in the acquisition process. While there are many measures that can be taken to maintain the integrity of evidence, enabling audit logs on the forensic workstation or server is one of the easiest to implement and present in courtrooms.

No matter which tools the examiner decides to use—browser, command-line, or proprietary tool—all supplied commands and interactions must be captured and recorded properly. Audit logs capture all actions taken by the examiner during the examination timespan (commands entered, clicks, processes executed, exit codes, and more). In an audit log, each log line is timestamped to the second using the local time zone of the examiner's computer. This helps jurors and judges reference specific log entry during the trial. By having audit logs enabled on the cloud and in the forensic workstation with timestamps, fewer doubts can arise about any generated or auto-generated piece of information or data during the process.

A secondary measure that can significantly support the evidence to withstand scrutiny in courtrooms is to properly take clear screenshots of critical outputs and artifacts. The screenshot must show the generated artifacts, the timestamp, as well as any calculated hashes. These screenshots later can be referenced later in courtrooms and tied together with audit logs to enhance the integrity and validity of the acquisition process.

1.3. Authenticate the forensic tools

Why? Authenticating the forensic tools contributes to the authenticity claim of the generated data by ensuring the integrity of the forensic acquisition tool used throughout the process. Using authentic cloud-native tools to generate the forensic evidence helps validate the authenticity of the evidence without the need for expert witness in courtrooms.

How? Many commercial and open-source digital forensic tools can be used for the collection and acquisition phases of the investigation. Trust in these tools is based on common-use or community feedback, and the collected data have always been thought of as trustworthy. This leaves an open area for the defendant to question the authenticity and accuracy of the tool used. This setup in the methodology establishes trust in the forensic tool and removes possible ambiguities related to the integrity of the tool. To achieve full trust in a tool, there are two possible procedures to be followed. If scripts are to be used, the first option is applicable: the source code is analyzed to determine the security of the source code, and then it can be compiled, and hashes calculated for the generated executables. If a community-approved or a commercial tool is used, where the source code is not available to download, the second procedure applies: the tool should be downloaded from a trusted provider repository and paired with a calculated hash provided by the vendor/developer/maintainer of the tool. Once the tool has been installed or unarchived, the hash of the binaries should be compared to ensure that it is a match with the one provided by the vendor, proving the authenticity and integrity of the tool.

If one of the above-mentioned procedures has been followed, the tool can be formally used in the forensic investigation with less doubt about its integrity and the integrity of the generated data. For some non-traditional and newer tools, such as the public cloud management consoles, these procedures might not be applicable. Most of the CSPs furnish their customers with at least four possible means to access and operate their cloud environment: a browser-based console, a command-line console, a list of exposed APIs, and finally a software development kit console (SDK). In the case of a browser-based console, the authenticated account used as well as the digital certificate of the website should be validated and captured before executing the forensic acquisition. In the case of a command-line console, the access keys used to authenticate to the CSP should be validated and compared against the ones recorded on the browser-based console. Access keys should also be validated when interacting with the CSP via direct API calls or through an SDK. Following such validation schemes can help remove concerns about the integrity of tool and its generated data, suppressing possible further scrutiny in court.

1.4. Capture the architecture and metadata

Why? Capturing the architecture and metadata contributes to the authenticity claim of the generated evidence by ensuring the validity and consistency of data sources and that the investigated digital asset matches the identified data sources during the preparation phase of the forensic investigation.

How? The amount of consumer independence and the level of automation differentiate the cloud computing environment from traditional computing farms. Most of today's cloud deployment follow a process called DevOps, where developers and operators come together to streamline product development and delivery. DevOps involves setting up a cloud environment with multiple automated tasks that can start from creating the virtual machine all the way to decommissioning it. This process and others can occur without any human intervention. There are also processes in the middle that perform many tasks on the virtual machines or virtual networks to adapt to changes or workloads. Crucial to the success of the investigation is that the customer cloud architecture is captured and analyzed. This sometimes requires reading documentation and talking to DevOps team members and cloud architects to understand the full scope and possible impact of the investigation; it helps greatly to setup appropriate OOV.

Metadata about a VM hosted in a cloud environment provides data about the virtual machine or resource that uniquely identifies it. Documenting the investigated instance VM metadata, such as region, hosted zone, hostname, IP address, MAC address, instance ID, launch date, signature, associated security policies, auto-scale groups, access groups, routing groups, and any other connected attributes is vital before starting forensic engagement. Another important factor to be documented is the time difference between the systems and the forensic server (time-drift). These characteristics help determine how to execute the next phases of the methodology as well as confirm the identity of the target. Without capturing metadata, determining appropriate order of volatility to prepare for a successful acquisition is not possible. Metadata need to be captured via a screen capture tool (or the continuous integrity process) and saved as an image or acquired via the cloud API, in a location where the image can be viewed at its final destination. The saved metadata file should have its hash calculated and saved. The same information needs to be documented in the investigation track book or the case document, including the calculated hash.

1.5. Define order of volatility (OOV)

Why? Referencing the metadata and all possible data sources determined during the preparation phase, order of volatility should be defined to optimize the collection and acquisition of relevant evidence. This helps ensure the completeness and correctness of the acquired evidence.

How? Determining the appropriate sequence for interacting with volatile data is crucial to supporting a digital forensic case. Since the focus in this research and the target for this methodology is IaaS environments, which are mainly constituted of compute, storage, and network resources, following OOV as defined in the Internet Engineering Task Force RFC 3227 would suffice the initial volatility imperative for a Windows- or Linux-based server. In the public cloud context, a secondary volatility order that needs to be taken into consideration is the logging retention defined by SLA or in agreement with the cloud provider architecture. Logs and possible data sources with shorter retention periods or log rotations (such as network traffic) need to be captured before logs with longer retention and rotation timespans (such as console and account audit logs). This part of the volatility order differs per the organization's contract with the CSP, and the default data retention values vary between CSPs. Thus, the order of operations for the next phase of the methodology is affected by the outcome of this sub-phase of the methodology.

2. Execution phase

2.1. Volatile data acquisition (Optional, iterative)

Why? Volatile data acquisition contributes to the operability and completeness of the acquired data and supports the generation of the foreseen evidence to defend the claimed contentions. Though acquiring ephemeral data might not be required for the investigated case, this is completely dependent on the preparations phase and the volatility order defined in the previous phases. Thus, if offline investigation is required or preferred, all volatile data are lost and there are no other means to recover them as related to this instance.

How? The prior defined OOV should be followed**.** By design, all computer programs are loaded to the RAM or physical memory-like structure before being passed on to the CPU (or cache registers beforehand) to be processed. A complete image of the physical memory should be taken, which includes information not limited to the current running processes but also including any allocations of memory spaces, network and communications statistics, passwords, cryptographic keys, unencrypted data, and also some hidden data. While it can be argued that performing this memory acquisition might have negligible impact on the investigated server, it is strongly recommended that this acquisition occur remotely without logging onto the investigated virtual machine to avoid state changes and possible objections during trials. Further, it is recommended that the acquisition is performed from the cloud-native API call or command-line tools; if that is not possible, then it can occur by utilizing pre-installed forensic remote acquisition agents or tools. If neither option is available, then it can be accomplished from within the investigated machine while documenting all system changes during the course of the acquisition process. The collected image must be transferred directly to the forensic workstation immutable storage, with hashes calculated and captured via the continuous integrity mechanism.

2.2. Digital Seizure (Optional, iterative)

Why? As in a traditional forensic investigation, seizing the actual evidence is a very important step of the process. While actual seizure of the digital device or data sources is not possible in the public cloud, it is possible to isolate the investigated instance into a dedicated, separate, and secure network while following the pre-defined OOV. This limits the potential harm from propagating into the network, helps remove any maintained or undetected unauthorized access, and assists in maintaining the machine state during the various acquisition phases. In some instances, cloud providers furnish customers with some ephemeral storage options to be attached to their compute instances instead of the regular block and persistent storage; in such cases if the VM gets shutdown or terminated, all associated data is lost, not only the commonly known volatile information. This makes digital seizure and live acquisition the only options for constructing an appropriate case. Console access should be granted specifically to the forensic workstation. This sub-phase contributes directly to the preservation phase of the digital forensic model, as well as to ensuring the integrity of the data source and the trustworthiness of the acquired data. This sub-phase plays a vital role in the incident response effort of an organization.

How? One of the major selling points of cloud computing—and especially public clouds—has been always the elasticity (hitless scalability and dynamic high availability operations) of resources. In IaaS deployment, virtual machines or any other type of compute resources hosted in the public cloud are part of scaling groups that provide greater accessibility and workload adaptability. They can also be set up behind traffic load balancers to allow for high availability and load distribution. In such scenarios, if an instance is compromised or being investigated, then the auto-scaling function can change the instance metadata and sometimes the internal configuration parameters. If the investigated VM/server falls under this scenario, it should be removed from scaling groups before commencing the actual acquisition. If the examined instance is behind load balancers, it should also be removed from the load balancing group to deny the active network traffic from reaching the server. Instance metadata should be checked before and after the removal from these groups to reconfirm the investigated instance identity.

2.3. Non-volatile data acquisition (Iterative)

Why? Non-volatile data acquisition contributes to the operability and completeness of the acquired data and supports the generation of the foreseen evidence to defend the claimed contentions. The disk image is an exact (bit-by-bit image) of the server or instance storage that contains all persistent data saved on the disk. Though live forensic investigations can take place and some relevant or volatile information can be extracted, the disk image might still be needed as a reference at a later stage in the process or might need to be examined further to acquire more relevant data or evidence supporting data.

How? Most of the investigations involving a workstation and servers require a full disk image to be investigated by a forensic examiner or to be sent to be processed by specialized software for analysis. As referenced in the literature, the common methodology of acquiring disk images is by seizing the disk, attaching write-blockers to the disk drives, and attempting to perform a bit-by-bit copy of the disk. In a public cloud environment, a partially similar approach might be possible, by taking a snapshot of the running instance-associated disk volumes and sending them directly to the forensic workstation. This process can be executed and automated via the cloud-native API calls and CLI or GUI interfaces. While the generated disk duplicates might not be in raw format, they can be mounted with read-only permission in an isolated environment, and then, if required, raw format images can be produced using the traditional imaging techniques and tools. The complete acquisition process must be captured via continuous integrity processes. Metadata about the snapshot as well as the produced raw image files (if any) must be captured and saved immediately following the completion of collection process, and hashes are calculated for the metadata captures.

2.4. Collect supporting artifacts (Iterative)

Why? Collecting supporting artifacts contributes to the operability and completeness of the acquired data and supports the claimed arguments and the produced evidence. Cloud-generated audit logs or events related to network traffic flows, security tools, or any underlying infrastructure logs are invaluable to the success of digital forensic investigations. Especially with audit logs, the key is to answer special questions such as who did what, where, and when. Other important data to collect is DevOp tools and gadgets that service the suspected cloud instance.

How? Though sometimes collecting RAM and disk images can be enough to identify the thumbprints and fill the gaps in a cybercriminal case involving IaaS deployments, collecting supporting information to complete the picture of how the attack happened is also worthwhile. In IaaS, other components in addition to the compromised server itself can often generate valuable logs. Audit logs and event logs from load balancers, security managers, antivirus software, web application firewalls, web proxies, network firewalls, cloud account management consoles, and other related logging facilities should be collected and downloaded to the forensic workstation. The logs to be collected can be defined during the preparation phase of the forensic investigation, when data sources are identified. It is important to plan on collecting relevant logs after checking the SLA document.

The period of retaining logs and events can be from one week to two years, depending on the contract signed by the CSP and the customer. A copy of the logs should be downloaded directly in the forensic instance, dating back as far as deemed necessary for the investigated case. Saving the logs in raw text format, if possible, is recommended; sometimes the CSP allows the logs to be downloaded directly from the cloud instance management interface or the command-line interface.

Some of the main reasons' businesses are moving to the cloud is to utilize DevOps mechanisms to automate workloads and ease deployments of their software products. Thus, during forensic investigations it is very important for examiners to take a quick look at the mechanisms, code, and gadgets being utilized to make sure there is nothing hidden there. Almost all of the actions that can be done via the cloud console can also be done in a more efficient faction from the cloud native DevOps tools as well as in various integrations with the open-source DevOp tools. Data related to DevOps should be captured by downloading involved code or taking screenshots of the cloud console of the automation dashboards. This data acquisition process must be captured via the continuous integrity mechanism; hashes for all of the downloaded or created files must be calculated, and the byte size and last modified timestamp of each file must be noted.

3. **Closure phase**

3.1. Complete the continuous integrity process

Why? Completing the process for continuous integrity contributes to the overall integrity of the collected data as well as the forensic methodology followed. This step helps ensure that a very short and limited chain of custody is produced, as all information and supporting documents reside on the same immutable storage.

How? At this point in the investigation, the actual acquisition and collection of data is completed and documented. Hash values should be computed for all of the generated and produced log files and screenshots. These files and hash values should be stored directly on the forensic workstation.

3.2. Validate the collected data

Why? Validating the collected data contributes to the operability and completeness of the collected data. The data have been collected and saved on the forensic workstation. The last step in the methodology is validating that the data have been captured properly and are readable by the various digital forensic tools. The acquisition is now complete and is passed along to the forensic examiners to continue the analysis.

How? The necessary data have been collected according to a prepared list of data sources; this includes, but is not limited to, physical memory image, disk image, and cloud account and underlay logs. All logs and screenshots should be checked for readability and accuracy. All log files should be opened in a text editor to validate the readability of the files' contents. The collected artifacts such as disk, memory, cloud logs, systems logs, and account access logs should be loaded into the selected forensic examination tool to validate the content and operability of the data files.

As it is crucial to understand the various environments that are used to execute the methodology, Figure 7 is a swimlane flowchart depicting the various environments that can act as a reference for forensic acquisition processes. The first environment is the forensic examiner workstation, where all initial tasks take place, and it is used to connect to the cloud portal. The second environment that should be ready is the security environment. The security environment is in a separate cloud network or account, with limited and controlled access for the forensic and incident response team only. This account is used to host the forensic server and the connected immutable storage. The third environment is labeled as a production, and this one refers to the network segment or account where the incident took place, or where the suspected virtual machine resides. Depending on how the methodology is being adopted, and the required evidence, a fourth environment might be required, which is the quarantine or the logical seizer environment. This environment can be used to isolate the suspected instance if necessary.

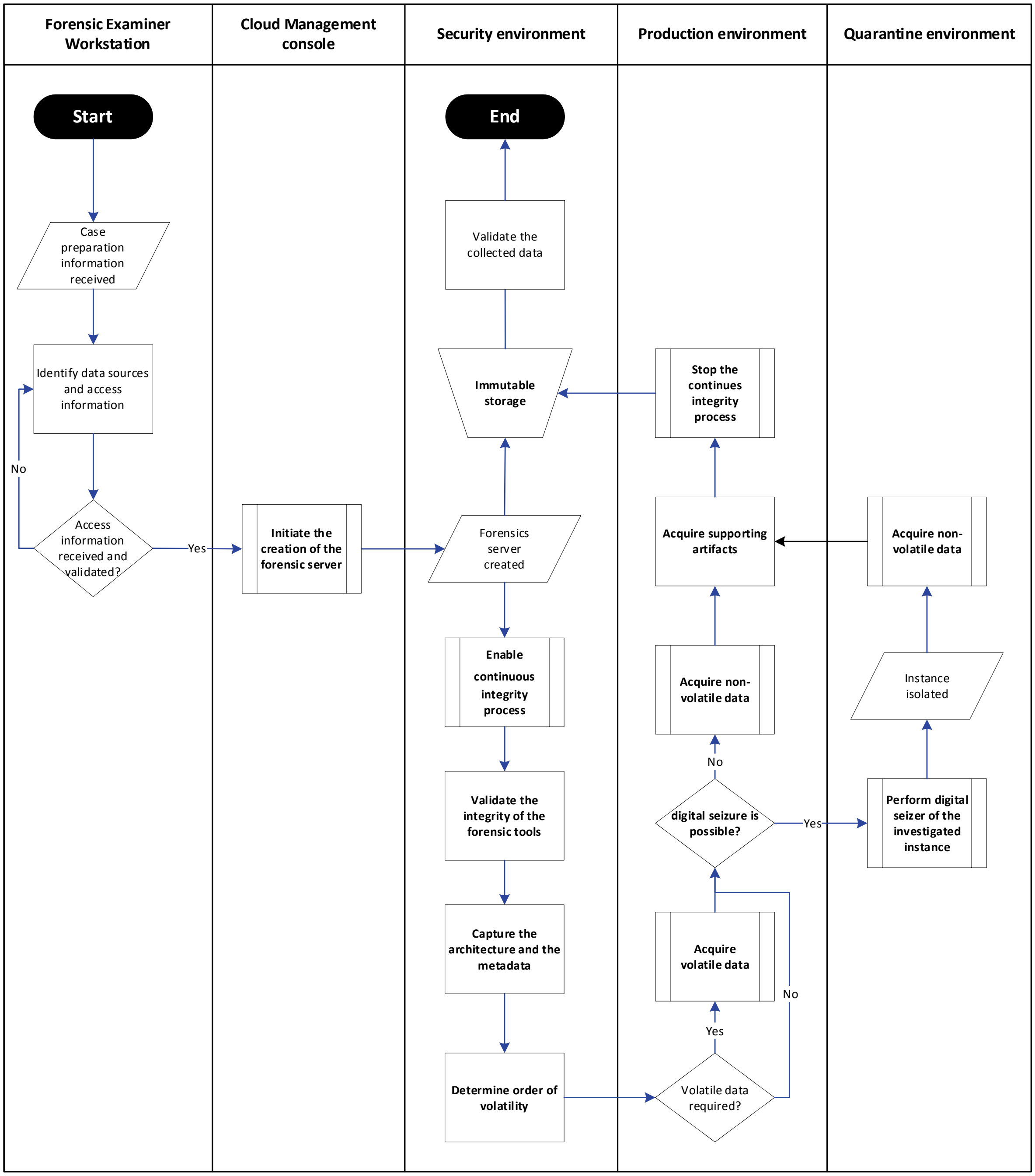

Forensic Examiner Workstation
Cloud Management console
Security environment
Production environment
Quarantine environment
Start
End
Case preparation information received
Validate the collected data
Identify data sources and access information
Immutable storage
Stop the continues integrity process
No
Access information received and validated?
Yes
Initiate the creation of the forensic server
Forensics server created
Acquire supporting artifacts
Acquire non-volatile data
Enable continuous integrity process
Acquire non-volatile data
Instance isolated
No
Validate the integrity of the forensic tools
digital seizure is possible?
Yes
Perform digital seizer of the investigated instance
Capture the architecture and the metadata
Acquire volatile data
No
Yes
Determine order of volatility
Volatile data required?

Figure 7 - IPCFA Proposed Flowchart

Assumptions

IPCFA was designed with today’s rules and regulations in mind and applied to the existing leading public cloud providers to demonstrate its efficacy. This makes it subject to changes related to both public cloud computing ecosystems and relevant U.S. federal rules and regulations that govern digital evidence. Examination of the history of federal rules of evidence, federal rules of civil procedures, and the Daubert standard shows that continuous amendments and improvements occur. In the past three decades, these changes have transpired for numerous reasons such as gaps found in rules, inconsistency observed between rules, and overall enhancements and amendments due to new use cases or relevant legislations. While there are many amendments for every rule of federal or civil procedure, which is normal, not all changes are equally impactful. For example, FRE 902 was amended eight times between 1987 and 2017, but only two amendments are considered impactful improvements (Rule 902, 2011).

Yet over a similar timeframe, major changes to the computing world have taken place. In 1990 the WWW was invented, seven years later cloud computing was defined, and in 1999 Salesforce offered the first form cloud services (SaaS). In 2006 AWS was launched as the first public cloud provider to revolutionize the traditional datacenter and server hosting paradigm and offer IaaS, initially. So, this clearly shows the advancement in technology that can take place within a short time. Furthermore, the pattern of improvement and amendment to technology is likely to continue. Thus, it is probable that legal changes will be necessary in the future. Such changes can affect the applicability of IPCFA as described in this research. Table nine captures regulatory and technical assumptions that are critical to the successful application of IPCFA.

Table 9 - IPCFA Future Application Assumptions

Assumption	Category
Federal Rules of Evidence are the relevant regulations for evidence admissibility in U.S. courts	Regulatory
FRE 902(13) and 902(14) are active, and the definition of self-authentic evidence is unchanged	Regulatory
Daubert standard is the relevant expert witness challenge, and its content is unchanged	Regulatory
Public CSPs offer APIs to operate (create/update/delete) the hosted IaaS recourses such as compute, storage, network, and other core services elements	Technological
Public CSPs offer APIs to execute remote commands on the hosted IaaS components	Technological

Public CSPs offer APIs to collect logs and metrics about the hosted IaaS components	Technological
Public CSPs offer multiple storage access granularity levels (read, write, write-once-read-many, write-append)	Technological
Public CSPs offer different types of storage to accommodate the various use cases (binary data, OS data, file shares, archive data)	Technological

Recommendations

Each IaaS deployment might have a different set of corporate or compliance body policies governing all aspects of security and incident response. They may differ on such items as how the instantiated systems must be hardened, what security tools must be installed, and what logging and auditing processes and configurations need to be enabled and applied. Establishing a strong hardening baseline that helps enable forensic capabilities for each instance before being approved to go into production is imperative. Furthermore, and before engaging with the CSP, it is vital that the contract states the appropriate Service Level Agreement (SLA) terms that permit the customer to investigate cybersecurity incidents on their deployed platforms and promote cloud forensic readiness. Below is a list of recommendations and preconditions that, if met, can conceivably contribute to yielding a high level of trustworthiness in executing the proposed methodology. The list has been constructed from many resources based on experienced researchers and skilled practitioners, as well as the researcher's understanding and experience performing forensics in the public cloud throughout the tenure of this dissertation (*Building a Cloud-Specific Incident Response Plan*, 2017; *SWGDE Best Practices for Digital Evidence Collection*, 2018; Gonzales et al., 2007a, 2007b; Grobler & von Solmes, 2009; Kent et al., 2006; Killalea & Brezinski, 2002; Orr & White, 2018; Simou et al., 2016; Stone, 2015; SWGDE, 2014, 2018; Watson & Jones, 2013).

Table 10 - Recommendations for Cloud Forensic-Enablement

Cloud Forensics Recommendation	**Technology**	**Policy**	**SLA**
Have a well-documented digital forensic process. This can be part of the incident response policy and procedures, but it needs to cover the complete the		✓	

forensic investigation/evidence lifecycle. The required training and expertise of the forensic examiners should also be covered under the same policy.			
Establish logging requirements for cloud platforms (compute/server, network, or storage) and set the severity level, timestamp format, and retention and rotation periods. Define where else logs should be sent (event correlation tools—SIEM).	✓	✓	✓
Pre-define set of corporate-approved tools to be installed as part of any instance deployment in the cloud platform, which includes forensic-enablement tools.	✓	✓	
Build a forensic playbook that describes procedures to be followed as part of the corporate incident response procedures and policies.		✓	
Enable logging and auditing in the cloud environment (all pipelines, compute, network, and storage instances) by corporate policy.	✓	✓	
Enable consistent time synchronization among all systems within the private cloud zone or region.	✓	✓	
Establish training requirements for staff members handing security incident response and digital forensics. Make sure staff has active and valid certificates pertaining to their specialty.		✓	
Ready forensic server image (forensic acquisition and analysis tools, cloud API command-line tools) and version it for all deployed CSPs. This includes processes on how the IR team obtains the required access to be able to carry all of their required investigations' tasks.	✓	✓	

Acquire and document from the chosen CSP a clear process guiding how law enforcement agencies can approach the CSP for information requests.		✓
Ensure that the SLA clearly defines the timeframes and mechanisms for the CSP to notify consumers about data breaches or security incidents to the consumer platform or the cloud underlay infrastructure.		✓
Define with the CSP which logs from the underlay infrastructure can be exposed during forensics examinations, if requested, what the request process will be, and how long the log retention periods.		✓
Describe the types of logs to be maintained regarding the deployed IaaS cloud and the retention periods for each type of log. Specify recoverability after deletion options for each type of log.	✓	✓
Define ownership of the data residing in the corporate cloud instances and any derivatives of it.	✓	✓
Define data storing location(s) by agreeing with the CSP on the corporate authority to decide where all the data is stored for the deployed environment (country, state, and jurisdiction).	✓	✓
Clarify the process of retrieving encryption keys from the CSP (if required or lost) and capture any limitations to data decryption.		✓
Define data-volatility recovery options (elastic auto-scaling groups and resources). What type of data can be recovered for each type of the ephemeral resources, and what are the associated time constraints?		✓

CHAPTER FIVE

DEMONSTRATION

To demonstrate and validate the proposed methodology, two use cases were crafted. The use cases cover common aspects of highly probable cybercriminal activity that could involve the public cloud. The first scenario (Case #1) assumes a public cloud deployment that was deliberately compromised (that is, a criminal act takes place in the public cloud), and in the second scenario (Case #2) the public cloud deployment is used as a means to compromise other non-cloud-based components (in other words, the criminal act takes place outside the public cloud). Both cases assume investigation conducted with probable litigation, thus; U.S. courts evidence admissibility requirements are kept in mind. In these scenarios it was assumed the victims would want to perform their own investigation and forensic exanimation, utilizing their in-house resources or hiring external forensic investigators to validate their claims before reaching out to law enforcement. The collected data are expected to meet the following investigative goals:

- Determine the chronology of the attack.
- Identify the source and scope of the impact of malicious activity.
- Uncover the origin of the attack and tying it back to possible intruder, if any.
- Enable the client to prosecute the attacker(s) in courts of law.

The forensic examiner (the researcher) has thirteen years of experience in the field of information technology and cybersecurity (M.Sc, CISSP) and network and operating systems (CCNP, CCDP). Experienced with cloud computing, the examiner has three years of technical hands-on performing digital forensic acquisition in the various major public clouds. All qualifications are documented and provided, including diplomas, employment letters of duties, and professional certifications. The examiner is familiar about the international, federal, and state rules and regulations of digital evidence. The same examiner (the researcher) performs acquisition in all presented scenarios.

Hypothetical application of SWGDE 2020 – AWS use case

The forensic examiner prepares for acquisition by obtaining all information related to the incident, possible data sources, deployed IaaS architecture, cloud account information, and possible timeline of the attack. The input to the forensic examiner is the story from the case described in Chapter Three, the architecture of the platform and how it works, the key personnel who manage the environment, and any other relevant information such as which security or best practices are implemented. The examiner documents all the information in their workstation. They create a folder named "Case001 - Dakwa LLC." The examiner creates an Excel file (logbook) or relies on another case management tool with all possible evidence resources that might be vital to be acquired and presented for analysis. As this is a sub-phase of the whole forensic processes; this logbook can be a part of a larger logbook documenting the whole process. At this stage all steps are to be followed on SWGDE (2020) while relying on the focus group examiner's knowledge and expertise. The following steps resample the combined steps proposed by the focus group examiners walk-thru of the Dakwa LCC acquisition:

1. **Steps to Take Prior to acquisition**

   This phase of recommended process corresponds to the preparation phase of the many available DF models. Based on the case information, the examiner needs to authenticate to the AWS environment where the incident is suspected. The examiner has to confirm the architecture and metadata provided by the team and check scaling groups, load balancers, networks, databases type and setup, policies, and make sure permissions are in place. Based on this review, the examiner can then determine the best path for acquisition and the type of data to be collected, as well as how it can be used to extract evidence to support the claim. The examiner can determine whether it will be required to have some volatile data and disk images of the suspected VMs (snapshots or raw disks), access logs, database audit logs, database transactions' logs, and events logs from the account console. AWS logging facilities to be collected by the examiner include CloudTrail, CloudWatch Logs, RDS (Relational Database Services) Logs, GuardDuty, and Security Hub logs. It is important to document all collected screenshots, accesses, and data sources touched in a logbook.

Based on the above information, the examiner can decide on the best acquisition mechanism, and in this case, it is a combination of cloud native data export tools (cloning or snapshots) and other well-known DF tools (open source or commercial). The examiner must be very familiar with the AWS architecture and available tools and integrations. The destination media must be prepared according to best practices. The examiner could decide to have a local media or a remote media on the cloud itself (AWS S3 bucket) with read-only capabilities and good size to be able to fit all possible data to be collected. Some of the tools the examiner can prepare in this phase include but are not limited to EnCase Enterprise, AWS_IR (Incident Response), AWS_CLI, AWS Browser Console, and Margarita Shotgun.

The acquisition location in this case is virtual, and acquisition can be performed from a trusted and secure environment such as the examiner's workstation with appropriate credentials. There are no physical servers to seize or preserve; all servers are logical. The examiner is furnished with all diagrams and information about the setup of the virtual private cloud where Dakwa LLC hosts their platform. As the examiner is working directly with the company; any encryption keys can be provided.

2. Steps to Take During acquisition

The second phase of SWGDE is to execute the selected methodology and use the designated tools to collect the needed forensic data. The examiner begins by preparing an electronic notebook to document all the processes with actions, timestamp, data source, and data type. Photographs and screenshots are also taken to support the written notes. The examiner cannot avoid the need for volatile data that might reside on the suspected VMs; thus, live acquisition is required. The examiner determines the appropriate OOV, such as RAM, running processes, network connections, system settings, and storage media. The examiner can launch AWS CLI from their workstation and acquire a lot of information from RAM such as hidden processes, network stats, encryption keys, and other relevant data by running a tool like AWS_IR or Margarita Shotgun to collect memory dumps of the suspected VMs.

The examiner has the option to take snapshots or clone the VMs, and while these are considered a full disk image, they do not include RAM or volatile data. The examiner cannot turn the VMs off, as some data might be required from the volatile memory. When it comes to the database, since AWS RDS is being used, the examiner can execute a SQL dump tool such as Mysqldump to obtain a complete snapshot of the database. Amazon RDS runs on EC2 as well, so the examiner can decide to just take snapshots of the RDS VM instances instead. Some data need to be collected manually from the actual environment setup, such any existing proxy setup and the use of a Content Delivery Network (CDN) and network configuration. All collected snapshots can be downloaded to the forensic examiner workstation and saved.

If the organization has a policy regarding how chain of custody needs to be documented, then the examiner follows it. In this case, the company does not have a chain of custody specific policy. The examiner documents chain of custody following SWGDE recommendations and notes at least the following: the full name and signature of the person keeping the forensic data, the full name and signature of the person receiving the data, the date of data transfer, the purpose of the data transfer, and the method of the data transfer (SWGDE, 2018a). The examiner might decide to deliver the case directly to the forensic workstation used for analysis and examination for evidence extraction; this makes fewer entries into the chain of custody logbook, which can be beneficial for the court case.

3. Steps to Take After acquisition

The examiner collects disk images of VMs, database dumps, and volatile information in different files, saves them into the designated media, and calculates hashes for each file right after acquisition. This acquisition can take place by executing the disk imaging tool directly from AWS CLI, or via indirect SSH access using AWS Systems Manager. The same can also be done by taking snapshots of the attached media of the EC2 instances, basically snapshotting the EBS (Elastic Block Storage) volumes and cloning the VM. The hash algorithm used to maintain the integrity of the acquired files is SHA-128 which is one of the approved hash functions by NIST. The hashes are written to the notebook.

The examiner reviews all acquired data to make sure they cover all possible sources of relevant evidence. Dakwa LLC does not have their own policy about documentation of forensic cases; thus, the examiner follows best practices. The examiner documents in a logbook all outputs from the tools used and makes sure to document any errors and explain how they were overcome. All calculated hash values are documented referencing the relevant output files. The generated document contains the following information: a unique assets identifier, source of evidence, case number, acquisition type, hash values for each of the acquired data, photos of evidence, examiner full name, acquisition date, errors encountered, and any other additional information that are relevant to the acquisition (SWGDE, 2018a).

Table 11 - ALR for SWGDE 2020 Application – AWS Walk-Thru

Criteria	**Trustworthy (6-10)**	**Doubtful (2-5)**	**Untrustworthy (1-0)**
General/Legal			
1) Adopted a structured and published forensic acquisition method or process	-	(5) Process followed SWGDE, which is recent and not tested, but recommended.	-
2) Forensic acquisition practitioner certified abilities	(9) Examiner abilities are certified, but the examiner has no previous court involvement.	-	-
3) Trustworthy capturing of the whole acquisition process	(9) The process is captured on notes and screenshots.	-	-
4) Provided case-supporting artifacts	-	(5) Few supporting artifacts are collected, while not required by SWGDE. Preservation is not discussed.	-
5) Very well documented and validated chain of custody	(9) CoC is documented, not mentioning DF process or stage of documentation.	-	-
Data acquisition			
6) Used trusted acquisition tools	-	(5) While the tools used are known, validation and authentication are not discussed.	-

7) Data is captured in operable format	(9) Data has been captured in operable and known formats. Data is not validated.	-	-
8) Utilized secure and immutable evidence storage	(6) Storage is defined and secure. Isolation and immutability are not addressed.	-	-
9) Validated the captured forensic data	(10) SHA-128 is calculated for all the captured evidence data files.	-	-

Hypothetical application of IPCFA – AWS use case

Along the same lines, we can apply the methodology proposed in this research to acquire forensic data from the Dakwa LLC AWS environment. The story from the hypothetical was taken as an input and how the platforms work, and the deployment architecture were captured. All relevant management and access data was captured. The examiner documents all the information in their workstation. They create a folder named “Case001 - Dakwa LLC.” The examiner creates an Excel file (logbook) or relies on another case management tool with all possible evidence resources that might be vital to be acquired and presented for analysis. As this is a sub-phase of the whole forensic processes, this logbook can be a part of a larger logbook documenting the whole process. The following tasks can take place to generate court-sound evidence.

1. **Initiation - Stand up a forensic server**

   The forensic examiner requests a new environment to be created that is dedicated for this forensic case. The account needs to be isolated: it must not reside on any of the common networks and must have a security policy to allow it to access the suspected environment components; nothing is allowed to access any resources in the environment. Then in the created environment the examiner attaches a S3 (Simple Storage Services) bucket to be used to upload the forensic workstation image and another immutable storage, which could be S3 or Glacier, depending on the configuration. This immutable storage is used as a destination for all forensic data as well as the various log sources. The image of the forensic servers is converted into an AMI (Amazon Machine Image) and installed on the AWS dedicated environment as an EC2 instance. The immutable storage is attached to the forensic server. During this demonstration, any time the examiner is saving forensic data to the forensics storage, it can be assumed that the save is to immutable storage, unless otherwise stated.

The various forensic tools are uploaded to the S3 bucket attached to the forensic server instance, then downloaded into the server. Tools include a web browser, EnCase Enterprise with agent, putty terminal emulator, AWS_IR, AWS CLI, AWS Console, LiMe, or MargaritaShotgun, and all prerequisites' packages and libraries. Tools should not be installed at this stage, but binaries should be validated and made available to be installed after the continuous integrity mechanism has been deployed and activated. Usually with expert forensic examiners, they will have their tools integrated into a forensic server image that is ready to be deployed when the need arises. At this stage the forensic server is set up securely and ready to conduct the forensic investigation; for this hypothetical scenario, it is assumed to be a Windows 2016 server. The last step in the process is for the examiner to document the metadata information for the forensic server/EC2, attached S3 buckets, group policies, and permissions applied in the investigation logbook.

2. **Initiation - Enable continuous integrity and logging**

The examiner enables AWS CloudWatch and Cloud Trail logs on the forensic server and for the environment with logs to be sent to immutable storage. The forensic EC2 instance is set up with time-synchronizing protocol to be in the same time zone as the suspected environment. The examiner installs a screen capture tool with imaging and video-recording abilities. The tool is set up to capture artifacts with timestamps and location tracking and configured to send the captures directly to a designated folder in the immutable storage hierarchy. The last tool to be installed and/or enabled is an event- or activity-tracking tool on the forensic server that captures all clicks and commands and saves the output directly on the immutable storage. If the tools capturing a command-line session, the output can be extracted at the end of the session and sent directly to the immutable storage.

In this case the forensic server is Windows based, thus we rely heavily on these tools to capture mouse clicks and pop-up actions. The commands and other information are captured by the screen capture tool. Throughout the next steps in this process, it is assumed the screen capture tools are enabled and available, and the generated logs/events are being sent to the immutable storage. The examiner validates this ability before moving on to the next phase. Any screen capture file should have its hash calculated and be saved right after the screen has been captured. The same information needs to be documented in the investigation logbook or the case management document, including the calculated hash.

3. **Initiation - Authenticate the forensic tools**

Coming into this phase, the forensic server is being monitored and is auditable, so it is ready to commence operations. The available forensic tools are installed and configured under monitoring provided by the continuous integrity and logging processes initiated in the previous phase. Before installation, the integrity of tools downloaded from vendors repositories must be validated by comparing the hashes provided by the vendors to the actual hashes calculated on the forensic server. If the hash matches, it must be logged in screen capture and saved. The screen capture should show the hash on vendor's website as well as the calculated one. For example, this can be applied to the browser and the commercial forensic tools, for example. The same process can be applied to open-source tools, with an additional step of reviewing the available source code for possible unintended features or security flaws. When possible, open-source tools are compiled from sources based on the maintainer's recommended settings. This applies especially to AWS native and related tools such as AWS_CLI, AWS_IR, and AWS SDK. All tools must be set up with logging enabled and with a default output to a folder on the immutable storage.

In this scenario, three of the main forensic tools that must be indisputable are the command-line or terminal emulators, the internet browsers, and the AWS_CLI and its components. The command-line emulator is used for most cases to execute commands directly on the cloud relying on some other tools, such as AWS_CLI, thus it is important to use a reliable, authentic, auditable, and community-recommended terminal emulator. As most cloud providers provide a web console to administer cloud resources, such as AWS Console, browsers play a vital role in cloud forensic investigations as the majority of the work is carried out by the examiner using the browsers to collect data and validate the cloud setup. Finally, the cloud-native tools provided, in this case AWS_CLI and other dependent tools such as AWS_IR tools, rely heavily on configuration files that contain cryptographic certificates, key pairs, or credentials; thus, it is important these tools are set up properly and functioning as intended. All these tools, configuration files, certificates, and keys must be captured during this phase to show the appropriate ones are used to connect to the suspected instances with appropriate policies and authorization.

4. **Initiation - Capture architecture and metadata**

While the Dakwa LLC team has already provided the architecture of their platform, it is recommended to capture the actual setup directly from the AWS cloud console right before starting the forensic acquisition. In this case, the examiner authenticates the AWS console and attempts to capture the status of infrastructure deployed, that includes the automation and DevOps processes, the interactions between components, the virtual network IP schema, the network flows and policies, access management, and any other enabled capabilities. Then the focus is on the investigated instances, which includes networks, virtual machines, and databases. The VPC network information and information such as all egress and ingress point of the VPC is captured; this includes public IP ranges, private IP ranges, VPN gateways, transit gateways, and VPC peering.

The examiner takes a screen capture of the metadata of each EC2 instance that is part of the suspected platform. Information in the capture includes but is not limited to the AWS region, hosted zone, hostname, and attached storage, types of storage attached, size of storage, IP address, MAC address, instance ID, launch date, signature, associated security policies, auto-scale groups, access groups, and routing groups. Partially similar metadata is captured for the attached RDS MySQL Database instances. The examiner calculates the integrity hashes for each screen capture taken or documents generated and saved on the immutable storage. These hashes are also saved on the logbook. At this point the Dakwa LLC cloud architecture is known to the examiner and documented. The information about how the environment components are working together is clearly understood.

5. **Initiation - Define order of volatility (OOV)**

Referencing the metadata and all possible data sources suggested during the previous phases, the examiner attempts to put together the best possible course of action. Taking into consideration the OOV proposed by SWGDE (2018b) as well as the one proposed by (Killalea & Brezinski, 2002), the examiner puts together a recommended OOV. The three major sources of volatile data are the memory images of all involved EC2 instances that host the applications and the RDS databases, the data located in any attached instance stores, and the swap files. The following actions take place, from top-down, to guide the rest of the acquisition process:

1. Acquire full memory dumps from all *EC2* instances.

2. Acquire all data located in the *instance stores* (Ephemeral storage) attached to the *EC2* instances.

3. Acquire all data located in swap files of the *EC2* instances.

4. Acquire full disk images for all *EBS/EFS* attached to the *EC2* instances.

5. Acquire full copies for all attached *S3 buckets* in the environment.

6. Capture any existing automation, deployment, or delivery code that can be located on AWS *CodeCommit*, *CodePipeline*, *CodeDeploy*, or other DevOp tools.

7. Capture the existing network configurations, security policies, groups, and rules.

8. Export account and services logs & events that are setup to send logs to CloudWatch, CloudTrail, GuardDuty, and AWS Security Hub.

9. Export other logs that are setup to log locally, this can be any or all the VPC Flow logs, ELB logs, S3 bucket logs, CloudFront access logs, and RDS logs.

6. Execution - Volatile data acquisition

Following the defined OOV, the examiner has three volatile data sources: RAM, instance store, and swap files. The best scenario for acquiring memory images is remotely without running or attempting to execute the tool from within the compromised instance. The examiner executes a memory dump tool such as MargaritaShotgun which uses SSH keys to connect to an EC2 instance and dump the raw memory into a file. This tool can be executed straight from the forensic server command-line, and the generated file can be named properly and saved directly to the immutable storage. The same process can be executed using AWS Systems Manager, which allows one to invoke remote commands over SSH via AWS APIs at the remote host without having to authenticate to the host. Once the snapshots/disk images have been collected, integrity hashes are to be computed and documented via the continuous integrity tools as well as the logbook.

The second type of volatile information that the examiner needs to capture is any data that is stored on EC2 attached Instance Store volumes. This can be done in multiple ways. The examiner can use the AWS CLI to provision a clean EBS volume with disk size that is equivalent or larger than the Instance Store that is targeted; then they can copy the data into the newly provisioned EBS volume. Another option that is easier if the instance store happens to have few files is to use AWS CLI to copy all files and directories directly to the forensic server immutable S3 bucket. We assume the examiner has chosen to use the second option and has copied all files into a dedicated and pre-defined folder in the investigation S3 bucket.

The last type of volatile data is the swap files, which might contain very important information since it is an extension to the memory. Swap files are captured as part of the disk image acquisition. Once the copies have been made, integrity hashes are to be computed and documented via the continuous integrity tools as well as the logbook. At this stage it is unknown to the examiner which one of the EC2 instances has been compromised, thus the examiner collects volatile information from all involved servers.

7. **Execution - Digital Seizure (if possible)**

Usually performing digital seizure and isolation is a task for the corporate incident and response team when possible, but in this case the corporation does not have an IR team. It is also not easy to isolate all involved VMs from the network, as they are production, and it is not yet determined which VMs are involved or all of them have been compromised. Thus, this step is not executed yet.

8. **Execution - Non-volatile data acquisition**

Following the defined OOV, the examiner collects disk images from all involved EC2 instances and their attached storage volumes. The examiner collects volume information about each EC2-attached EBS volume, then uses the AWS CLI to attempt to perform API calls to create snapshots of the EBS volumes and send them directly to the forensics S3 bucket. This operation captures disk images, including swap files that are usually stored under the root home or other directories on the associated disk. Sometimes swap files are also created and saved on Instance stores; in that case they are captured as part of the non-volatile acquisition phase. If swap files are stored on another attached EBS volume, then a snapshot of that storage volume is captured. The same mechanism can be used to duplicate data stored in any attached S3 buckets, saving the data and sending it directly to the forensic storage.

The same operations can be done remotely utilizing AWS APIs calls or AWS Systems Manager to execute Linux native disk images tool such as dd. With dd, the output is 100% raw disk image, which can be read by many digital forensics analysis tools, including community, open-source, or commercial tools. In this investigation, the examiner only takes snapshots, as they can later be attached to EC2 instances for examination. If there is a need to have raw images, they can be recreated from the snapshots as well. The examiner calculates the integrity hashes for each screen capture taken or documents generated and saved on the immutable storage. These hashes are also saved in the logbook.

9. Execution - Collect supporting artifacts

In this phase of the methodology, the remaining items from the predefined OOV are collected. The examiner attempts to log in to the suspected environment using a web browser and authenticates to the AWS Console. The examiner makes sure they are visiting the appropriate console by checking the SSL certificates of the visited URL before authenticating. Then authentication takes place, and the examiner captures the SSL certificate information, console URL, account number, and username information via the continuous logging mechanism, which automatically saves this output to the forensic storage. Henceforth, all work is done via the web console and can be commenced via the established, authenticated, and encrypted session.

The examiner navigates to the AWS CodeCommit page (source control), and downloads/clones all existing repositories that should contain CloudFormation YAML or JSON code that might be used to automate the deployment and delivery for products and infrastructure components. The same code can also be used to execute any AWS SDK API functions to perform any changes to the environment.

The next section that the examiner attempts to capture is any configured or setup Codebuild projects. AWS CodeBuild allows end users to compile and build infrastructure components from source code that is stored in an S3 bucket or other repositories. The examiner screen-captures existing projects and the content of each project. Then delivery pipelines listed under AWS CodePipeline are captured in a similar manner. The last component of the automation to be captured is the AWS CodeDeploy configurations, which might contain instructions to automatically deploy an application or a bundle of applications. This completes the acquisition of all DevOp and automations code that might be relevant to the incident.

The next data that need to be captured are the network and security configurations. While most of this data is likely to have been captured during the metadata and architecture captures, in this phase the examiner validates the data provided by the customer via console access. The examiner uses the AWS Console to navigate to the following network configuration areas and validate the existing setup against the information provided during the metadata and architecture capturing phase, capturing any differences: route tables, prefix list, internet gateways, carrier gateways, NAT rules, DHCP options, VPC peering, VPC endpoints, Elastic IP addresses, DirectConnect interfaces, transit gateways, and VPNs. The same approach is taken to validate security configurations and settings from the following areas: identify and access management (IAM), logging and monitoring, Security groups, Network ACLs, VPC Flow Logs, and Route53/DNS. At this stage the examiner has validated all network and security configurations and settings, capturing any differences.

The examiner next attempts to capture all possible logs in the environment that might contribute to solving the investigation code questions. This phase is highly dependent on which log facilities are enabled and being used. From the previous step, assuming the environment is set up properly and the main log facilities are enabled, the examiner exports logs directly to the forensic S3 bucket for the past 30 days from the following log facilities: AWS CloudWatch, CloudTrail, GuardDuty, VPC Flow Logs, and Security Hub. These are the main locations where logs are getting aggregated from EC2, RDS, IAM, Automation tools, CloudFront, load balancers, and other enabled cloud services or systems. Some other logs are captured as part of the EBS snapshots if they are saved on local facilities on EC2 instances. It is very important for the examiner to notice the retention of logs during the previous step in order to determine how far back to go in each one of the log facilities.

10. Closure - Complete the continuous integrity process

Reaching this phase in the process means the examiners are confident that they have collected all needed data to generate the sought-after evidence. The examiner navigates to the immutable storage to make sure every collected file has its hash value calculated. If a file is found without a corresponding hash value, a hash value is calculated and saved. All hashes should be calculated using SHA-128 or greater. The examiner then stops all running screen capture tools as well as any scripts that were executed to capture keyboard input during the forensic process. The final logs from the screen capture tools are moved to the immutable storage and their hash values calculated and saved.

11. Closure - Validate the collected data

The examiner reviews data sources and makes sure all sought-after data has been captured. The examiner opens each collected file using the respective analysis tool to validate content, ensuring it is appropriate and not empty or corrupt. All collected logs and screenshots are opened for readability and accuracy. A similar approach is taken by the examiner to review the collected disk and memory, where the images are loaded into examination tools to validate content, timeframes, and data availability.

Table 12 - ALR for IPCFA Application – AWS Walk-Thru

Criteria	**Trustworthy (6-10)**	**Doubtful (2-5)**	**Untrustworthy (1-0)**
General/Legal			
1) Adopted a structured and published forensic acquisition method or process	-	(2) Process followed IPCFA, which is not published or tested yet, but is structured.	-
2) Forensic acquisition practitioner certified abilities	(7) Examiner abilities are certified, but examiner has no previous court involvement.	-	-
3) Trustworthy capturing of the whole acquisition process	(10) The process is captured on notes, screenshots, and screen video for multi-process operations.	-	-
4) Provided case-supporting artifacts	(10) All possible events and logs for the cloud platform are captured.	-	-

5) Very well documented and validated chain of custody	(9) CoC is limited to logbook initialization, then all original forensic data resides in the immutable storage.	-	-
Data acquisition			
6) Used trusted acquisition tools	(9) Most tools are cloud-native; thus authenticity and authentication has been validated before execution.	-	-
7) Data is captured in operable format	(9) Data is captured in operable and known formats. Data is validated.	-	-
8) Utilized secure and immutable evidence storage	(10) Secure, immutable storage is used to store all forensic data.	-	-
9) Validated the captured forensic data	(10) SHA-128 is calculated for all captured evidence data files.	-	-

Demonstration of IPCFA –Azure use case

The initial input to the forensic examiner is the story from the Azure case described in Chapter Three. The examiner documents all the preparation information in their workstation and creates a folder named "Case001 - Dakwa LLC." The examiner creates an Excel file (logbook) or relies on another case management tool, listing all possible evidence resources that might need to be acquired and presented for analysis. Since acquisition is a sub-phase of the forensic procedure, this logbook can form part of a larger logbook documenting the entire process. The examiner is then is furnished with a copy of IPCFA to be followed to collect data that can aid in the investigation of the intrusion. The examiner's task is to acquire authentic data from the fictional Zool Corporation cloud environment and make it available for analysis in a forensic server. During the preparation phase of the investigation, the examiner is briefed about Zool Corporation, its cloud architecture, and management philosophy. The department head and senior engineer provide an architecture diagram of the production environment where the incident took place. They provide the following information to the examiner:

- The IT team consists of eight engineers: two DevOps (sysadmin, OS, storage, automation), one DB admin (SQL servers), two NetOps (network, automation), and two SecOps (security, audit, incident response), and a team manager.
- The platform is designed following Microsoft N-tier architecture (N-Tier Architecture Style - Azure Application Architecture Guide, 2018), which allows for scalability, modularity and high availability. The platform is designed as a 3-tier application (Figure 8).
- Zool has two environments, production (prod) and development (dev). Dev is a replica of prod, with minor changes related to security policies. All infrastructure components of a single environment belong to a single Azure resource group to allow for ease of deployment, maintenance, and monitoring. The incident took place in the production environment.
- The production environment consists of four different subnets using private addresses: management (automation and bastion virtual machine), web (web servers), app (application servers), and db (database servers).
- The IaaS platform consists of six Linux CentOS 7.2 servers (web/app virtual machines), and two Windows 2019R2 servers (Azure SQL virtual machines), all in availability-sets and different availability zones. This ensures each virtual machine is in a separate fault and update domain. The first three Linux servers are dedicated for the web interface and located in the DMZ network, which is accessible from the Internet. The remaining three Linux servers run the Casino management system (CASM), which also allows bidders to play, book tables, and bid on live events online. The two Windows servers run on Azure SQL Database systems that hold all the application and financial data.
- All servers are split among three zones hosted in Azure East U.S. Datacenters. This provides additional fault tolerance and meets the corporate 99.5% SLA goal.
- The Azure SQL Servers are hosted in two different zones and active replication is established between them. The databases are set up with Always-On architecture, so both can transact simultaneously.

- All incoming traffic is load balanced between the three front-end webservers via prod-frontend-lb; then all traffic from the webservers is load balanced between the three application servers via prod-backend-lb. All traffic from applications servers targets the active/standby SQL servers directly.
- All infrastructure is deployed using infrastructure-as-a-code (IaC), utilizing Azure PowerShell. All applications, software updates, and OS updates are managed via DevOp and automation processes that are hosted in a bastion virtual machine hosted in the production environment (prod-mgmt subnet). Access to all IaaS components is only permitted from the bastion machine, thus no direct access is permitted.
- Network Security Groups (NSG) are deployed to allow only the required application flows between the various IaaS components; other traffic is implicitly denied.
- Role-Based Access Control (RBAC) is adopted to manage authentication, auditing, and accountability of team members; all actions are logged; and access is only permitted to specific areas of the cloud platform. This is enforced by the SecOps team.
- Azure Security Center agents are installed in all virtual machines, and all logs are aggregated and sent to the security center. Network Security Group (NSG) flow logs are enabled for network traffic actions. Azure Monitor is enabled to all resources in the prod resource group.
- While the department has a set of information technology-related policies, there are no policies related to incident response, e-discovery, or data retention.
- Since the last occurrence of the incident, all engineers have changed their Azure Portal logon credentials and created new private keys for the API authentication.
- Zool suspects the jump server to be compromised; thus, they have decommissioned the old one and created a new one from scratch in order to mitigate the unknown access to the platform.
- Zool still notices from time to time that the production environment is using more resources than expected, for which they incur extra unplanned charges. They suspect there is some unauthorized activity taking place in their environment.

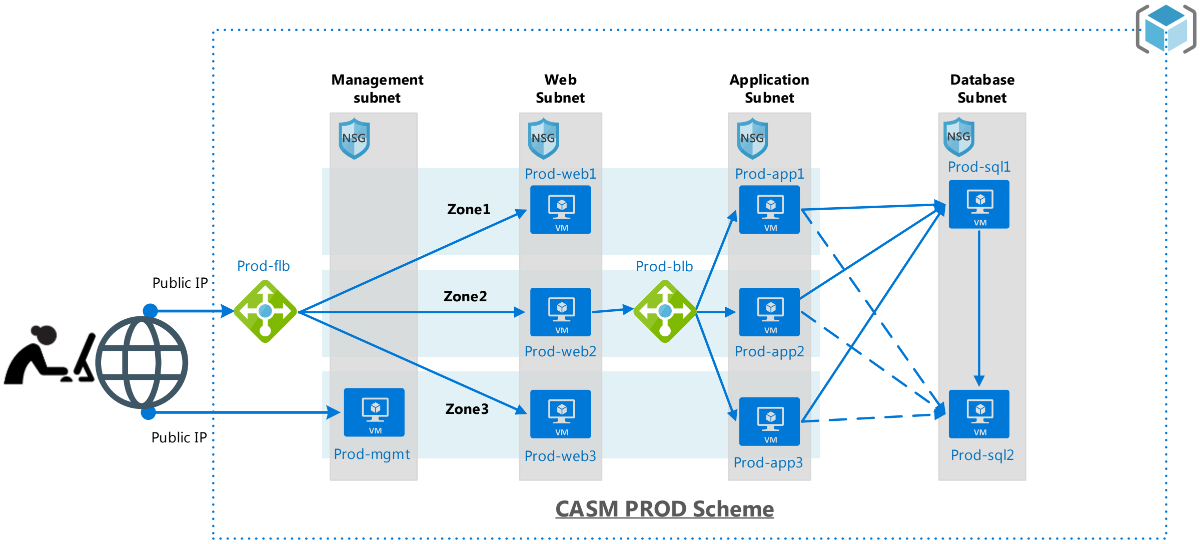


Figure 8 - Zool Corp. Production Environment

Prior to engagement, the examiner should be ready to perform acquisition in the various public cloud providers. Being prepared includes but is not limited to having a cloud-agnostic or cloud-specific forensic server image with all needed tools to perform acquisition and analysis. The forensics image must be in the examiner's Azure account, AWS account, and GCP account for minimum coverage. The examiner should have knowledge of at least the various cloud components of the major public clouds that make up the IaaS environment. Automation and DevOps scripting knowledge and experience are also very useful for the examiner to have. With all this knowledge and these tools available, and while anticipating direct interaction with Zool Corp staff, we attempted to follow the IPCFA methodology to collect the sought-after data from the emulated IaaS deployment:

1. Initiation - Stand up a forensic server

The examiner requested from the corporation IT staff to create a dedicated virtual network to accommodate the investigation. The environment would have no access to any other virtual networks and would have no roles associated with it. The corporate engineers furnished us with a new virtual network "prod-secure". The examiner already has an Azure account with a forensic image ready for deployment with all tools and gadgets needed to perform a digital forensic acquisition and analysis. All tools are located in a folder called "DFIR" and are to be authenticated, then installed when continue monitoring is enabled. The image is in the examiner's Azure account image gallery. The examiner shared the images with the Azure subscription of Zool Corporation by granting them read-only permissions to the image disk.

The examiner provided the Zool Corp. team with an Azure API-based script that automatically deploys the forensic image (Windows 10 Pro - generalized) in the new, dedicated environment. The script also creates the immutable storage utilizing Azure blob storage with a legal hold and timer-based policies, allowing for blob append operations. Furthermore, it creates an Azure file share to be mounted directly by the forensic server in order to perform other operations in copies of the collected data. The script then connects the immutable blob storage to the VM using NFSv3.0 protocol. This allows for all files being written to the immutable storage to be unchangeable once written and closed, while allowing for new files to be added/copied into the storage. This is basically a cloud write-blocker.

The examiner asked the team to add an NSG to allow access out from this new forensic server to the zoolprod environment on service ports for DNS, SSH, TLS, and RDP. This is accomplished via VNet (virtual network) Peering to allow inter-vnet traffic between the prod-secure and the Zoolprod VNETs. Opening these ports enables the forensic servers to have direct access to the IaaS environment to perform the various acquisition operations. The same NSG rules are applied on the opposite side to allow for connection establishment. The examiner's workstation public static IP address is the only access permitted on the newly deployed server. At this stage the forensic server has been deployed, and the examiner can access it remotely. This completes this sub-phase of the process, and the forensic server has been initiated successfully.

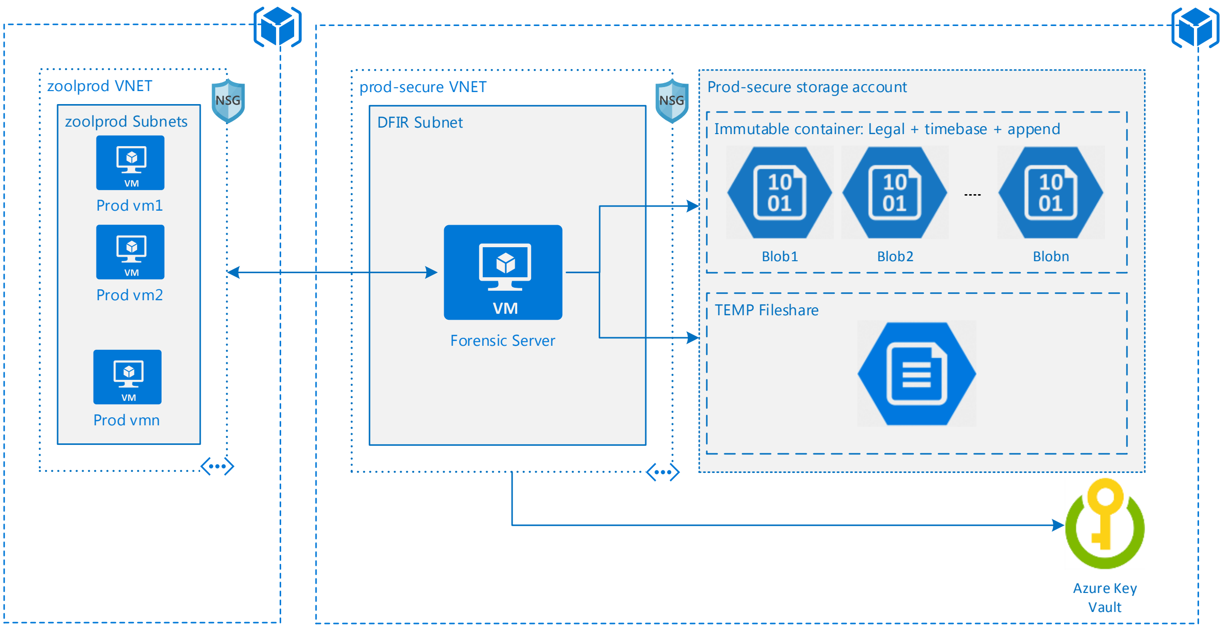


Figure 9 - The forensic server deployment architecture

2. Initiation - Enable continuous integrity monitoring and logging

Since the whole acquisition process is executed from the forensic VNet and resources group initiated in the previous sub-phase, all actions should be logged and referenceable. The following tasks have been executed on the forensics resource group components to enable logging, auditability, and action cross-referencing:

- Created an Azure Log Analytics workspace in the forensics resource group and enabled Azure Monitor.
- Enabled logs for Azure Key Vault and saved the logs container on the same forensic storage. Enabled Azure Monitor for Key Vault.
- Installed Azure log agent on the forensic server and enrolled the forensic server to be monitored by Azure monitor and log analytics. The agent captures all logs from the forensic server and sends them to the Azure Log Analytics.
- Enabled Azure Monitor for the Azure blob storage and Azure file share in order to capture read, write, and delete actions as well as all authenticated and anonymous access. All logs are sent to the forensics Azure Log Analytics workspace and Azure Monitor.

- Enabled NSG flow logs to have true visibility on all network activities that traverse the VNet. These logs operate at Layer 4 of the OSI model, and they record all traffic in and out of an NSG. All logs are sent to the forensics Azure Log Analytics workspace.

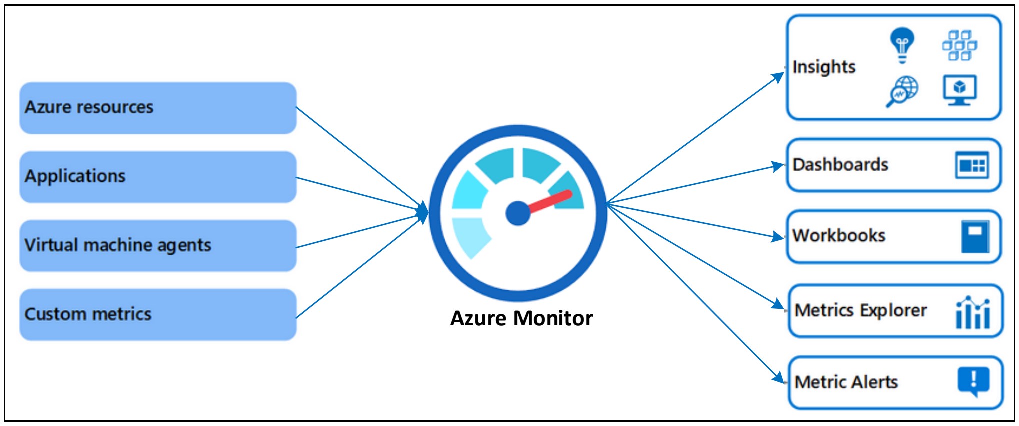
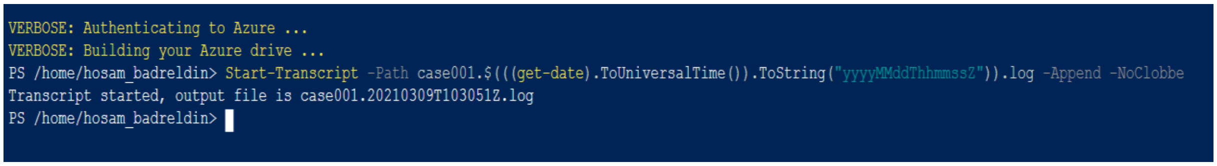


Figure 10 - Azure Monitor logs and metrics

After enabling comprehensive logging in the cloud forensic environment, additional continuous integrity monitoring measures and software packages are included as part of the forensic server image. Azure Monitor Agent is installed on the forensic server and reports all logs to Azure Monitor. The installers for Greenshot and OBS Studio are also included in the forensic server image. Both are open source, lightweight, and feature-rich screen capture software tools.



```
VERBOSE: Authenticating to Azure ...
VERBOSE: Building your Azure drive ...
PS /home/hosam_badreldin> Start-Transcript -Path case001.$(((get-date).ToUniversalTime()).ToString("yyyyMMddThhmmssZ")).log -Append -NoClobbe
Transcript started, output file is case001.20210309T103051Z.log
PS /home/hosam_badreldin> █
```

Figure 11 - Capturing all actions on Azure Cloud Shell

3. Initiation – Authenticate the forensic tools

The forensic server contains a folder with a plethora of digital forensic tools for acquisition, analysis, and presentation. The examiner puts together a list of tools that might be needed for this acquisition and validates and installs them. While the examiner's list might include all needed tools, at any point during the investigation the examiner can go back to the tool repository and check out other tools following the same process. The examiner initially pre-selected the following tools to be used for the acquisition:

- Provided scripts for disk, memory, and logs acquisition (open source)
- Greenshot: Screen image capture software (open source)
- OBS Studio: Screen video capture software (open source)
- Mozilla Firefox web browser: Azure secure web portal access (open source)
- azCopy: Tool that copies data between forensics VM and the immutable storage
- Windows Powershell: Powershell interpreter (native)
- Python27: Tool that executes Python-based scripts to facilitate the acquisition process
- Volatility: Memory analysis tool (open source)
- MobaXterm: Bash shell emulator to execute Linux commands and scripts
- Notepad++: Code editing, log reviewing, and diff tool

The examiner connected remotely from their workstation to the newly initiated forensic server via Remote Desktop. Before installing any of the selected tools, the installer file hash is captured, any certificates or public keys fingerprints attached to executables are recorded, and the version of the software is logged. The content for all scripts is reviewed, and hash is calculated and saved. All hashes and certificates are saved to the Azure Key Vault dedicated as part of the forensic environment. The examiner then installed each selected tool and captures software versions after each install. Timestamped screenshots are taken and saved to the forensic storage showing the authentication of each tool. At this point all acquisition tools are installed and ready to be launched.

4. Initiation - Capture architecture and metadata

Under a video capture, the examiner begins by launching the Mozilla Firefox browser and browsing to the Microsoft Azure portal. The examiner opens the website identity, website certificate, and website encryption protocol section before proceeding with the logon. Once these actions are validated and captured, the examiner attempts to login to the portal using his personal credentials. After successful authentication, the examiner changes the directory to assume the role assigned to them on the Zool Corp. subscription. Now the examiner can see and interact with all resources within Zool Corp. environments. The examiner opens Azure Cloud Shell, switches contexts to Powershell, and enables session logging to a file that saves directly to the forensic storage. The examiner first executes a command and a screen image capture of the status and identity information of immutable storage that is crucial during this investigation. Then, the examiner uses Azure Powershell to execute a set of commands to retrieve the following: Azure subscription information, virtual network, resources groups, all resource groups, network topology (capture from GUI), storage topology, and finally Azure DevOps repositories' structure and profiles.

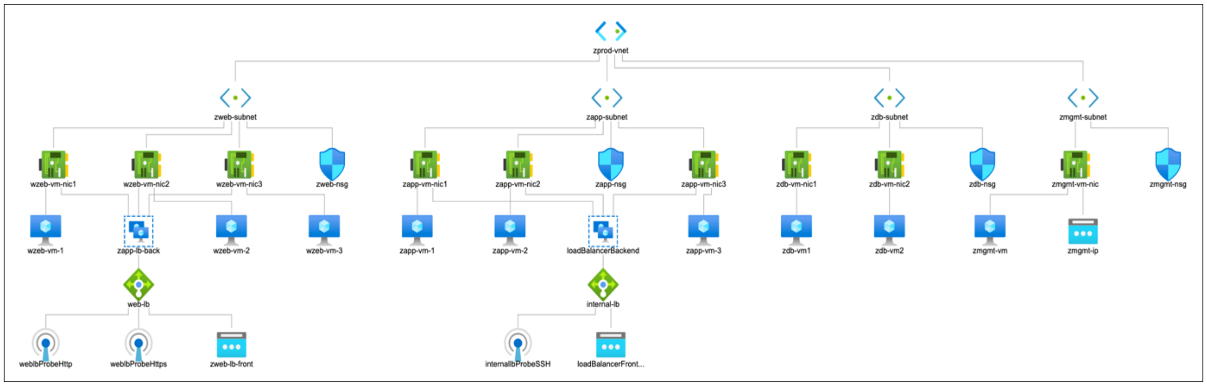


Figure 12 - Zoolprod topology generated from Azure portal

5. Initiation - Define order of volatility (OOV)

Once the examiner scrutinizes the data collected in the previous sub-phase, the architecture of the zoolprod environment becomes clear; data location and the sizes and types of storage containers, log retention periods, and virtual machines are identified. The examiner established and presented the proposed OOV to the client to review, explaining the actions that will be taken and their impacts, if any, to the production environment. The examiner proposed the following OOV to collect data to support the investigation:

1. Acquire full memory dumps from all virtual machines.
2. Acquire volatile information from Windows VMs (processes, connections ... etc).
3. Acquire full disk images for all storage attached to the virtual machines.
4. Acquire full copies for all attached blob storage in the environment.
5. Capture all existing automation, deployment, and delivery code in Azure DevOps.
6. Capture the existing network configurations, security policies, groups, and rules.
7. Export 90 days (when possible, Azure monitor default is 30) of cloud logs from Azure Activity logs, Azure Monitor, Azure Analytics workspaces, and Azure Network Watcher logs.

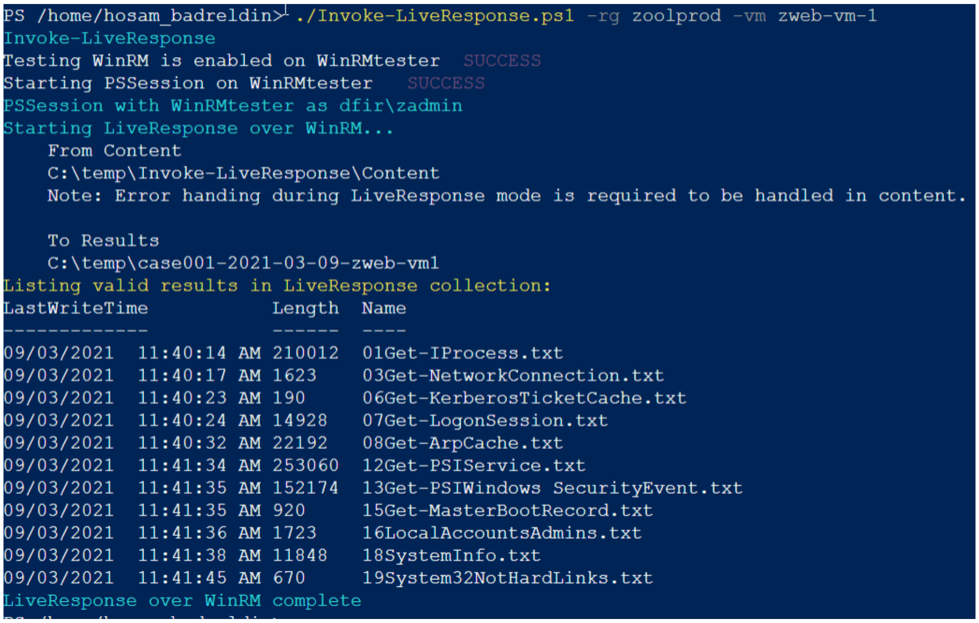
8. Export logs from virtual machines—from Windows using Azure Diagnostics storage and from Linux via Azure Monitor agents.

6. Execution - Volatile data acquisition

Following the defined OOV, the examiner attempted to collect a complete memory dump from all VMs within zoolprod environment, this includes all Linux and Windows VMs. IPCFA recommends performing the acquisition with the least interaction with the VM as possible, while interaction with the VM sometime is not avoidable, the examiner tried to utilize the cloud console to remotely execute lightweight tools on the VMs to perform the memory dump, then transferred the dump to the immutable storage. For Linux VMs, the examiner created a simple Azure Powershell script that connects to the VM via the cloud backplane console (Using Azure Cloud Shell), load *LiMe*, compiles it for the specific kernel, and execute it. Once the dump is ready, the script automatically calculates the SHA-256 of the collected image and transfer the generated raw image from the VM to the immutable storage. The collected hash values are stored directly onto Azure Vault dedicated on the forensic environment.

The same line of thought is followed for the Windows VMs utilizing *Winpmem.exe*. The applications used to dump memory have been chosen due to their light interaction with the VM processes, and their direct functionality of only dumpy the full memory of the VM. Due to the nature of memory sharing of the public cloud, acquiring memory images is one of the areas that public cloud providers usually restrict access to it, and it requires customers to open support cases in order to retrieve memory dumps from their own VMs.

The examiner moves to the next step of collecting additional volatile data from the Windows VMs, as this data will be automatically captured on the Linux memory dump, but not the Windows. The examiner tweaks the script *Invoke-LiveResponse* Powershell from (Green, 2018) and injects it directly on the VM via Cloud Shell to collect information from both Windows VMs such as: running processes, network connection, routing table, Kerberos tokens, logon session and more. The collected file then transferred to the immutable storage after its hash value gets calculated and stored. This process is executed under the continuous integrity monitoring via executing OBS video capturing to show the whole process then the video files are saved to the immutable storage.



```
PS /home/hosam_badreldin> ./Invoke-LiveResponse.ps1 -rg zoolprod -vm zweb-vm-1
Invoke-LiveResponse
Testing WinRM is enabled on WinRMtester  SUCCESS
Starting PSSession on WinRMtester   SUCCESS
PSSession with WinRMtester as dfir\zadmin
Starting LiveResponse over WinRM...
    From Content
    C:\temp\Invoke-LiveResponse\Content
    Note: Error handing during LiveResponse mode is required to be handled in content.

    To Results
    C:\temp\case001-2021-03-09-zweb-vm1
Listing valid results in LiveResponse collection:
LastWriteTime           Length  Name
-------------           ------  ----
09/03/2021  11:40:14 AM 210012  01Get-IProcess.txt
09/03/2021  11:40:17 AM 1623    03Get-NetworkConnection.txt
09/03/2021  11:40:23 AM 190     06Get-KerberosTicketCache.txt
09/03/2021  11:40:24 AM 14928   07Get-LogonSession.txt
09/03/2021  11:40:32 AM 22192   08Get-ArpCache.txt
09/03/2021  11:41:34 AM 253060  12Get-PSIService.txt
09/03/2021  11:41:35 AM 152174  13Get-PSIWindows SecurityEvent.txt
09/03/2021  11:41:35 AM 920     15Get-MasterBootRecord.txt
09/03/2021  11:41:36 AM 1723    16LocalAccountsAdmins.txt
09/03/2021  11:41:38 AM 11848   18SystemInfo.txt
09/03/2021  11:41:45 AM 670     19System32NotHardLinks.txt
LiveResponse over WinRM complete
```

Figure 13 - Collecting data remotely from Windows VMs on Azure

7. Execution - Non-volatile data acquisition

While collecting disk images from IaaS deployment is possible using any of the traditional forensic tools available today, it requires a lot of attention to the integrity and completeness of the collected images. IPCFA recommend generating complete disk image while performing the investigation close to the source, so in this case the forensic examination will take place on Azure; thus, it made sense to generate images that can be examined in Azure environment. Collecting images that are native to the cloud means the image can be used to create a new VM, a replica of the suspected VM, for investigation. It also means we can generate raw images if required, as it is basically going to be a matter of covering the mage from a custom image to raw image, which can be done using a lot of the community tools.

The examiner prepared a simple script that uses the Azure Powershell API and Cloud shell to execute and create snapshot of the disk image of each one of the VMs. When the disk image is created, the script will calculate its hash value, save the hash value to a dedicated and encrypted Azure Vault, and then save the collected image to the immutable storage. If the VMs were encrypted while in production, encrypted keys can be passed to the script to be used for decryption of the VMs to be able to collect snapshots, and the same keys can be used later to encrypt the snapshots at rest.

The next data to be collected based on the defined OOV is to download blob storage content that are attached to the VMs, and in this scenario there was no blob storage used, only managed disks, which were captured as part as the snapshots. Then capturing automations tools, logs, pipelines, and code archive. This data was captured partially already when the management VM snapshot was taken, as it is their automation server. What is missing is to capture via Greenshot some images of the Azure DevOps origination settings, projects, pipelines, and repos. While it is possible to download repos locally, some organizations will not allow external resources to download their private code, if any. As far as capturing network topology, NSG, NAT, Route tables, and other Azure configuration items, these items were covered initially during the architecture and metadata capturing phase of IPCFA, so it is not needed to be captured here again, unless some data was missed. All these processes are executed under the continuous integrity monitoring via executing OBS video capturing to show the whole process then the video files are saved to the immutable storage.

8. Execution - Collect supporting artifacts

Based on the input from Zool Corp team, and the architecture that has been captured, all logs are being sent to Azure Monitor and Azure Log Analytics. This includes Azure Activity, operating systems logs, and all other application logs configured on the VMs, such as Syslog-ng driven logs and Windows custom-events. In all large public cloud providers, these logs usually can be accessed via the Web Console, Command-line console, and API calls. This is the case in Azure Cloud. The examiner invoked simple API calls to *Get-azlog* and *Get-AzureRmOperationalInsightsDataSource* to check the volume and type of logs available, then based on that a small script is written and invoked to send all output to multiple files in the immutable storage blob. Once the files have been written, the hash values are calculated and saved to the Azure Vault. This process is executed under the continue integrity monitoring via executing OBS video capturing to show the whole process. The generated video files are saved to the immutable storage.

9. Closure - Complete the continuous integrity

Once this phase is reached, that means the actual acquisition and collection of data is completed and documented. Thus, any running continuous monitoring mechanism must be turned off. The examiner navigates to the immutable storage to validate that all files have been stored and they do have data and not empty files. The examiner matches the number of files on the immutable storage with the number of hash values calculated and saved in the key Vault. Once all is validated, the examiner and under OBS video monitoring, turned off the Azure Shell running command-line capturing tool, calculated its hash, and sent the last file to the immutable storage. At this stage the actual collection process has been concluded. If any of the phases needs to be repeated for any reason, the continuous monitoring will need to be enabled first; then the desired phase can be re-executed.

10. Closure - Validate the collected data

This is the phase where the collected data is validated to either be passed to the next examiner who will perform the analysis, or if the analysis will be carried out by the same examiner who completed the acquisition. The examiner starts by opening each screenshot and video captured to validate it is readable and operable. Then they open the collected log files using Notepad++ to check if the log files contain logs messages with timestamps that are readable. The examiner executed *Volatility* against the collected memory dumps and attempted basic operations to make sure the memory dump is functional and not corrupted. The last check was for the examiner to spin up test VMs from each collected disk image and be able to browse the VM content. Once all these steps have been taken, the forensic data is ready for analysis and to continue the rest of the digital forensic process as would with non-cloud acquisition.

It is important to note that this demonstration focused on following IPCFA to acquire IaaS related forensic data from Azure but did not focus on specific set of tools or technologies. The demonstration also did not go into data analysis to extract evidence, as that is not the goal behind IPCFA; rather the goal is to collect the intended data from CSP involvement and to maintain the integrity and authenticity of the collected data until it is presented in courts of law.

CHAPTER SIX

DISCUSSION

In this design science-based research, we examined the applicability of the existing digital forensic tools and acquisition techniques in acquiring U.S. court-admissible evidence from IaaS deployments in the public cloud. The literature showed a lack of cloud-specific forensic tools and scholarly work related to acquisition processes and methodologies. The published document from SWGDE (2020) that sheds light on acquisition from public cloud providers was the only valid reference that directly addresses cloud forensics. This section of the research focuses on presenting the findings and results stemming from the demonstration and validation of the proposed artifact that attempts to solve the research question.

As discussed in Chapter Three, the core requirements to determine the artifact's effectiveness are the following: 1) maintaining the integrity of the acquired data and providing a trusted, simple, and short chain of custody; 2) maintaining the authenticity of the data by adhering to FRE 901/902 and withstanding the Daubert test in courts; and finally, 3) generating operable, non-corrupted forensic data that is easily examinable using commonly available tools. The artifact should also encompass the functionalities required to acquire forensic data and should generate reliable results while being practically useable and cloud-agnostic. These requirements and some of the most stringent qualities for digital evidence for U.S. court admissibility have been merged to produce the ALR. Because ALR describes the most important factors that contribute to digital evidence court admissibility, it has been used to quantify the effectiveness of the executed digital forensic methodology or process.

Core Requirements

Ensuring the integrity of the evidence and the whole forensic process is crucial to a successful forensic acquisition operation. Failing to document a step can render the whole lengthy process unreliable. This makes the digital chain of custody the center of the forensic process. While the SWGDE 2020 publication does not include any direct reference to chain of custody, it references another publication (SWGDE, 2018b) that sheds light on how to document chain of custody. The 2018 publication only states that chain of custody should be documented to include personnel information when evidence is transferred from one person to another (SWGDE, 2018b), then references another publication for more details. This cross reference, while useful, does not make the SWGDE cloud acquisition process clear or complete since it relies on a document not created with the complexity of the cloud environment in mind. IPCFA, on the other hand, takes another approach in order to guarantee a trusted chain of custody and data provenance: IPCFA recommends chain of custody to be as short as possible by acquiring the forensic data and sending it directly to secure storage that is read-only. Then as many copies can be made as possible for investigators, judges, or any third party who might be interested in validating the integrity or authenticity of the collected evidence. This new methodology eliminates the need for a long and complicated chain of custody, limiting it to the forensic examiner, who maintains access to the secure storage and documents who received access for what reason. All this access can be maintained electronically via private keys and tokens, thus there is no need for actual documents, as access logs can be extracted from the Role-based Access Control system.

IPCFA adopts the industry standard for validating the integrity of any electronic file, which is the hash fingerprint. The hash fingerprint is almost an integral part of any digital forensic investigation and is always looked for in connection with chain of custody. Throughout the various phases of IPCFA, the forensic examiner is asked to calculate hash values for every evidence file they collect as well as for any supporting files. IPCFA recommends calculating hash values using approved algorithms and automating the calculation when possible, saving all hash values directly to sterilized and immutable storage while capturing the actual fingerprinting process via continuous integrity monitoring. While the SWGDE publication does direct the practitioner to calculate hash values of the acquired data, it does not refer to an approved hash algorithm from NIST, for instance. Computing hash values comes in the final step of their proposed three-step acquisition process, which is not appropriate for such a paramount requirement of the forensic process. SWGDE does a great job referencing organizational policy use, in the cases where that exists, as well as emphasizing reaching out to CSPs to see if they are willing to investigate the allegation.

The next major requirement for the artifact is to generate authentic evidence and in case the evidence’s authenticity or the expert’s testimony is challenged, to help pass a Dauber hearing. Confirming the authenticity of electronic evidence usually entails validating the claims associated with the evidence, and this can be done in multiple ways. IPCFA includes two possible authentication mechanisms. The first one is authenticating a piece of evidence by validating its integrity and authenticating its source, and the second is calling an expert witness (FRE 702) to independently authenticate the evidence.

The heart of the proposed methodology is the utilization of the cloud management plane to produce the needed forensic data, thus making sure all data that has been produced or captured is either generated by an electronic system (FRE 902(13)) or has been copied from an electronic device or storage medium (FRE 902(14)). In both cases the generated evidence is considered self-authenticating based on the most recent amendment of Federal Rules of Evidence 902 and has a great chance of preventing the defense from pursuing a Daubert hearing (Federal Rules of Evidence, 2017). Nevertheless, the defense can still ask for a Daubert hearing to challenge the methodology used to acquire to the claimed evidence. This is usually the case if there are multiple methodologies that can be followed to generate the evidentiary data. In this case the authenticity claim of the evidence produced by the expert witness is challenged via Daubert standard.

The Daubert test is one of the toughest legal challenges the defendant can use to cast doubt on the validity of the expert's testimony, possibly disqualifying it altogether. So, the best way to defeat the Daubert test is by avoiding it completely, which, admittedly, is not always possible. According to FRE 702, a witness may qualify as an expert on the basis of educational and experiential qualifications in a general field closely related to the subject in question (Rule 702. Testimony by Expert Witnesses, 2011). For those cases when expert witness testimony is challenged by Daubert, the Supreme Court has identified five non-exclusive important factors to assess the testimony. Though these factors have been delineated, in Kumho Tire Co v. Carmichael, the court said all five factors "do not all necessarily apply in every instance in which the reliability of scientific testimony is challenged" (Michaels, 2008). According to Michaels (2008), federal courts have also taken the position that the Daubert test should not be followed a checklist and was "relying instead on the ability of federal judges to properly determine admissibility" (p. 16).

All these factors have been taken into consideration during the IPCFA design; the methodology requires an experienced, certified, and documented forensic examiner to execute it. The methodology also gives the examiner control over the process by predefining OOV, and the steps required to generate the evidence and document each step rigorously. This helps make the process regeneratable by the court or any other expert witness. The aim behind the methodology, by design, is that once it gets published, it gains acceptance by the forensic community, thus meeting Daubert's second requirement, which mandates a peer reviewed and published methodology. In contrast, none of the expert witness-related scenarios were incorporated into the only existing cloud forensic recommendation document from SWGDE.

The last core requirement for the artifact is that it must be able to generate operable forensic data that is not corrupted and is easily examinable using commonly available tools. While having operable forensic data is a moot point if evidence generation is required, IPCFA makes sure there is a written reference to be followed. IPCFA requires all generated data to be checked for operability before moving to the next phase of the digital forensic process. This includes reviewing data for operability, correctness, content types, and readability. In SWGDE 2020's recommended process, data validation is required only when data is furnished by the CSP, and this is during the acquisition phase. After acquisition, there is no reference to data validation or correctness checks.

Current Knowledge

Before attempting the walk-thru of the hypothetical use case, a small, focused group was assembled of highly skilled digital forensic practitioners. Initially it was planned to have a focus group of 10 certified examiners, but due to the high level of expertise needed and the broad range of other requirements, the group contained only five highly accomplished experts. All were provided with the reference documents from Chapter Three (SWGDE, FRE, Daubert) and the AWS hypothetical scenario, and were asked to provide their walk-thru step by step in three to seven pages in a finalized read-only PDF document. Table 13 shows the credentials and qualifications of the focus group members.

Table 13 - Credentials of the Formed "Existing Knowledge" Focus Group

Member	Experience	Qualifications

1	20 Years	• B.Sc. – Computer Systems and Networking • M.Sc. – Information Systems and Assurance • CISSP - Certified Information Systems Security Professional • GCED - GIAC Certified Enterprise Defender • GCFE - GIAC Certified Forensic Examiner • AWS CSA - Certified Solution Architect • AWS CSS - Certified Security Specialist • Microsoft Certified Azure Engineer
2	10 Years	• Ph.D – Cybersecurity, incident response frameworks • SAFA - SANS Advanced Forensics Analyst • CEH - Certified Ethical Hacker • CDPSE - Certified Data Privacy Solutions Engineer • CompTIA Security+
3	18 Years	• BSc - Criminal Justice • M.Sc. - Computer Forensics • U.S. court expert witness experience • California Criminal Investigations Instructor • Certified High-Tech Crime Specialist
4	16 Years	• B.Sc. – Computer Science • Certified Cyber Threat Management • CompTIA Security+ • Microsoft Certified Azure Engineer
5	9 Years	• B.Sc. – Management Information Systems • CISSP - Certified Information Systems Security Professional • U.S. court expert witness experience

An analysis of the feedback-captured output from the focus group participants shows that while many common steps were taken and common tools used to acquire the required data, discrepancies were also found. Almost all practitioners attempted to use cloud-native tools and AP calls to collect logs and disk images. Three proposed the use of commercial tools to acquire memory images, and two proposed the use of open-source tools designed for AWS memory acquisition. While tools and data sources were common among the examiners, each participant followed a different starting and ending point in the collections process. Furthermore, only two of them mentioned chain of custody without stating how it would be securely and accurately maintained. Very basic attention was given to capturing additional logging information to support the evidence claims.

None of the practitioners addressed important admissibility arguments related to issues such as write blockers when copying snapshots, specifications for evidence and forensics data storage, or data transfer and access to examiners for examinations. While all practitioners were furnished with the SWGDE process to follow, the SWGDE provides only high-level guidance; major differences were clearly observed in the routes taken by each examiner to satisfy the need for court-soundness. This behavior contributed to inconsistent data being collected with missing FRE requirements, resulting in the collection of forensic data that might not yield forensically-sound evidence if presented in courts of law. This very short focus group study supports the contention that the field lacks a common practical methodology for cloud forensic acquisition, an important gap in the digital forensic knowledgebase shown in the literature.

ALR Comparison

This section of the discussion focuses on showing the strengths and weaknesses of both SWGDE's recommendation and IPCFA's methodology in acquiring forensic data that can lead to generating court-sound evidence based on the ALR criteria. This comparison is based on hypothetically walking-thru the AWS investigation scenario presented in Chapter Three first following SWGDE best practices, and then following IPCFA methodology.

1) Adopted a structured and published forensic acquisition method or process

SWGDE's *Best Practices for Digital Evidence Acquisition from Cloud Service Providers* (2020) is a published document that includes a structured process to be followed to collect forensic evidence from cloud providers. Thus, in this important criterion it wins with no doubts. Because IPCFA is yet to be published, it could not receive the full points as SWGDE did, but rather received partial credit for providing details and a structured approach for the acquisition. Keeping in mind that this criterion stems from Daubert's standard five factors, it is very important for the process to be published or adopted by the forensic community, which is not yet the case for IPCFA.

2) Forensic acquisition practitioner certified abilities

The full points for this criterion are awarded to both SWGDE and IPCFA as both acquisitions were performed by researchers with their credentials, qualifications, and experience documented before the beginning of the forensic process. Both SWGDE and IPCFA explicitly mention the importance of capturing the digital examiner’s credentials and experience. The full points were not given due to the examiner’s lack of real-life court experience and not serving as an expert witness before.

3) Trustworthy capturing of the whole acquisition process

SWGDE recommends for notes to be taken during the acquisition process, and when notes are not possible to be taken, screenshots and photographs are to be taken in lieu. This is indeed a very important point mentioned in the recommendation, but the recommendation lacks direction on how to maintain these photographs and guarantee their integrity in order to support the integrity of the collected evidence. IPCFA calls for a similar approach to be taken, but it replaces the notes with images and video captures that are saved directly into immutable storage with their hash values calculated. Capturing these hashes is very important as these videos and images can also be challenged during court sessions. IPCFA gets the higher points due to maintaining the integrity of the captured images.

4) Provide case-supporting artifacts

Section 6 of the SWGDE publication provides an overview of the possible acquisition methodologies that can be used during the “acquisition” phase of the process. While the acquisition process methodology chosen might be case-dependent, the publication does not mention the need to acquire any supporting data, leaving it completely up to the practitioner to decide what data might be needed. In the case of traditional forensic investigations, the supporting data might be very minimal or nonexistent, but in cloud environments it is paramount to authenticate any claim of evidentiary data as well as its method of access, collection, or transferal. The application of SWGDE on the AWS hypothetical case walk-thru shows some logs being collected and stored to support the evidence’s existence, but no proper preservation and integrity checks were enforced as they are not recommended by the published process. The opposite applies to IPCFA where the methodology mandates specifically the types of supporting logs and data that must be captured as part of any IaaS acquisition in the public cloud. IPCFA also shows the possible sources of supporting data when applied to the AWS case. IPCFA scores the highest score on this criterion while SWGDE gets the second score.

5) Very well documented and validated chain of custody

IPCFA addresses the chain of custody by shortening the length of it, only initializing it at the beginning of the process and leaving the rest to be automated via the role-based access control imposed on the immutable storage that logs all access, timestamps, and types of operations. While this is not the known methodology of documenting chain of custody, it is an updated method that can be accepted in courts if expert witnesses are able to testify to the trustworthiness of electronically produced CoC. On the other hand, SWGDE does not mention CoC in its 2020 publication, but refers to its 2018 forensic acquisition best practice document, which recommends that the examiner maintain chain of custody for all collected forensic artifacts. For this criterion, SWGDE gets the highest score, but not the full score as this recommendation is not directly included in its cloud forensics document. IPCFA gets a very close score to SWGDE as it describes the importance of CoC in cloud acquisitions and provides an unconventional and shorter CoC.

6) Used trusted acquisition tools

This is another area where SWGDE lacks guidance on how the examiner attempts the acquisition using the various type of tools (open source, commercial, well-known /tested, and cloud-native). There is no motion of forensic tools authentication at all in SWGDE, while the publication does mention the use of cloud API as a possible data acquisition venue. In IPCFA, tools are categorized, the authentication process of each type of tool is presented, and recommendations are proposed—and followed during the AWS case walk-thru. IPCFA also recommends using tools that are approved by NIST for the specific acquisition media, when possible. IPCFA scores the higher score in this section, while SWGDE gets the middle score due to missing tool authentication, which is a possible court hearing showstopper.

7) Data is captured in operable format

Both SWGDE and IPCFA recommend capturing the data in operable format, and both attempt the same when tested against the AWS use case. While SWGDE does recommend obtaining the data in raw format as well, it does not recommend or propose a data validation step. Thus, data can be generated in a well-known format, but when examination is attempted, it might be corrupted or inoperable. Both IPCFA and SWGDE get equal points for this criterion, as IPCFA generates disk images that are operable in AWS, but not raw formats (raw can also be generated from the AWS images when needed via any forensic tools such as dd), and

SWGDE is missing data validation steps while collecting data in both raw format and AWS images.

8) Utilized secure and immutable evidence storage

SWGDE recommends evidence preservation and storage to be in sterile, secure, and isolated storage, which is the common requirement for digital evidence storage and a best practice. While this might be enough in the case of traditional forensics, as well as in some cloud forensic cases, access and evidence access beyond the documented chain of custody is very difficult to attest to. Having a storage system that is read-only is invaluable to show that the data saved in this storage is not editable after storage. Immutable storage makes easy the presentation of evidence as authentic as the evidence has the appropriate integrity checks and thus is very difficult to be contested. IPCFA clearly states the storage of the digital evidence should be immutable and auditable, thus removing any questions relevant to preservation of the collected forensic artifacts. IPCFA gets the full points in this area, while SWGDE comes second.

9) Validated the captured forensic data

This criterion focuses on validating the integrity of all collected forensic data before closing the acquisition phase of the forensic process. While this might seem normal to any forensic practitioner, SWGDE does not mention validation after complete acquisition takes place—the examiner might be calculating hashes on the fly for all sought-after evidence files but might miss a file or a few files due to the use of a combination of various acquisition techniques. This might be what a defense attorney is looking for—a single evidence file that is missing its hash value during a developing court session that can be called inauthentic and inadmissible. Furthermore, SWGDE does not mention or recommend hash algorithms that can be used, as there are many that are outdated and no longer considered secure. On the other hand, IPCFA recommends the use of hash algorithms accepted and tested by NIST. Both IPCFA and SWGDE score high scores, but IPCFA gets few points more due to the level and specificities of validation it requires.

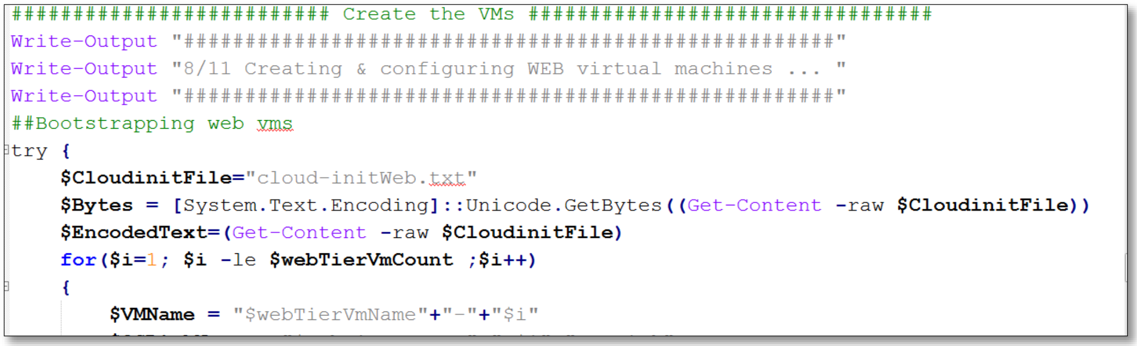
SWGDE is indeed a recent and very valuable document that sheds light on cloud forensics and provides a process to be followed by practitioners to conduct investigations in the realms of cloud service providers. It is broad by design to allow examiners to define their own processes and follow organizations' documented processes and procedures, which works well for larger organizations and corporations. In the case of smaller organizations that do not have applicable policies and procedures, the examiner bears the responsibility to define a process and follow it, hoping to generate authentic and court-admissible evidence. IPCFA fills in the gap, and proffers recommendations, ideas, and strategies for practitioners to keep legal evidence-generating in mind from the beginning of the process. In this very short, unconventional, and hypothetical comparison, IPCFA shows several advantages over the SWGDE publication, and IPCFA scores an overall higher ALR value when taking into consideration the challenging nature of cloud evidence.

Table 14 - ALR Comparison: SWGDE 2020 vs. IPCFA

Criteria	SWGDE	IPCFA
1) Adopted a structured and published forensic acquisition method or process	5	2
2) Forensic acquisition practitioner certified abilities	7	7
3) Trustworthy capturing of the whole acquisition process	9	10
4) Provided case-supporting artifacts	5	10
5) Very well documented and validated chain of custody	9	9
6) Used trusted acquisition tools	5	9
7) Data is captured in operable format	9	9
8) Utilized secure and immutable evidence storage	6	10
9) Validated the captured forensic data	9	10

IPCFA Demonstration

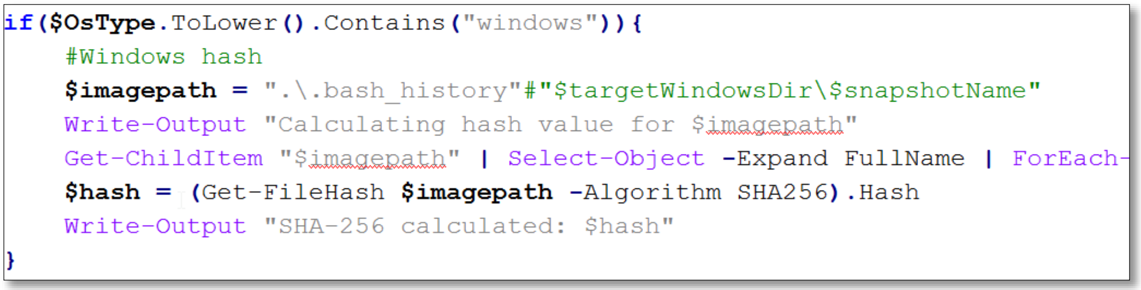
During the demonstration, we did not focus on the forensic evidence extraction and analysis process, but rather focused on the acquisition of cloud forensic data to demonstrate the use of IPCFA in an observational case study. Usually, once the data has been acquired, the forensic analysis process can progress in much the same way as it would for non-cloud investigations. The demonstration focused on exploring the various cloud tools that can produce forensic data that can be considered acceptable in courts of law. The use of these tools in conjunction with IPCFA allowed us to generate forensic data that can be thought of as admissible due to its integrity, authenticity, and ease of delivery to examiners and judges, if needed. The demonstrations involved building a sandbox to simulate the investigated environment, as well as creating a forensic server image and hosting it in a different account that is assumed to be owned and operated by the external forensic firm. A set of scripts were created to simplify the build and tear down of the sandbox environment to accommodate multiple tests and to fix issues encountered during the acquisition process.



```
########################## Create the VMs #################################
Write-Output "######################################################"
Write-Output "8/11 Creating & configuring WEB virtual machines ... "
Write-Output "######################################################"
##Bootstrapping web vms
try {
    $CloudinitFile="cloud-initWeb.txt"
    $Bytes = [System.Text.Encoding]::Unicode.GetBytes((Get-Content -raw $CloudinitFile))
    $EncodedText=(Get-Content -raw $CloudinitFile)
    for($i=1; $i -le $webTierVmCount ;$i++)
    {
        $VMName = "$webTierVmName"+"-"+"$i"
```

Figure 14 - Code snippet to initiate the POC environment

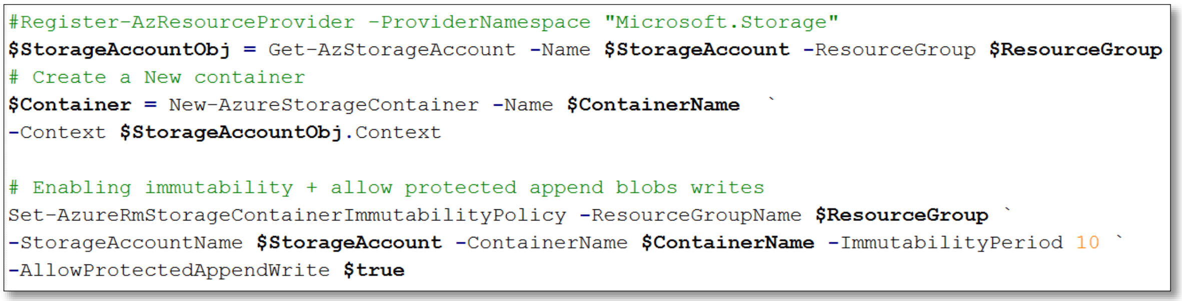
The demonstration attempted to shed light on the various public cloud tools and mechanisms available for the forensic practitioner to use to simplify the acquisition process. While most API calls and the made-available interfaces are cloud-specific—in this case Azure-specific—all public cloud providers have similar interfaces created for various reasons. Some cloud providers might offer more or less powerful and useful API calls. For instance, the Snapshot API call that allowed us to clone a VM disk has been created for disaster and recovery purposes, and that has nothing to do with digital forensics or investigations. The same applies to the rest of tools used in this demonstration. These unconventional tools are very effective in capturing digital evidence from public cloud deployments as long as they have been executed while keeping the integrity and authenticity of the collected data in mind.



```
if($OsType.ToLower().Contains("windows")){
    #Windows hash
    $imagepath = ".\.bash_history"#"$targetWindowsDir\$snapshotName"
    Write-Output "Calculating hash value for $imagepath"
    Get-ChildItem "$imagepath" | Select-Object -Expand FullName | ForEach-
    $hash = (Get-FileHash $imagepath -Algorithm SHA256).Hash
    Write-Output "SHA-256 calculated: $hash"
}
```

Figure 15 - Code snippet to calculate hash values

We have found that the possibility of creating storage that is immutable until certain conditions are met is a feature of almost all of the major public cloud providers, and it is indeed a beneficial feature when it comes to cloud forensics. This feature allows for a cloud storage blob to become the most secure storage for digital evidence when it is bundled with role-based access control and enabled audit logs. Having storage that can be accessible securely over the internet with specific access control lists is invaluable. This allows for access based on what Microsoft called SAS tokens (Shared Access Signature), which enables an investigator to access, read, or copy the evidence remotely and securely under audit. The access can be granted under many conditions including number of access times, allowed operations (read/copy/download), and access expiration token.



```
#Register-AzResourceProvider -ProviderNamespace "Microsoft.Storage"
$StorageAccountObj = Get-AzStorageAccount -Name $StorageAccount -ResourceGroup $ResourceGroup
# Create a New container
$Container = New-AzureStorageContainer -Name $ContainerName  `
-Context $StorageAccountObj.Context

# Enabling immutability + allow protected append blobs writes
Set-AzureRmStorageContainerImmutabilityPolicy -ResourceGroupName $ResourceGroup `
-StorageAccountName $StorageAccount -ContainerName $ContainerName -ImmutabilityPeriod 10 `
-AllowProtectedAppendWrite $true
```

Figure 16 - Code snippet to create immutable storage

Actual collection was straightforward, from collecting disk images using the Azure-provided API calls to obtaining various auditing and server logs. Azure does not provide an API to interface with hosted virtual machines' memory, and explicitly asks forensic examiners who want to capture memory dumps to submit an Azure support case and ask for memory dump for a specific VM. Other cloud providers such as AWS and GCP are the same; we were able to overcome this limitation by injecting a memory extraction tool like LiMe (Linux Memory extractor) into the VM command-line and execute the extraction tool with the immutable storage designated as a remote destination. This method prevents the direct logon and interaction with the virtual machine to run the memory dump tools manually. This was possible by utilizing the Azure run-command API calls, which allows for the VM admin to execute remote commands on the virtual machines if locked out or unable to login to the machines for any reason.

While the application of IPCFA in this context was on a hypothetical case, it showed that a great deal can be accomplished in the realm of IaaS public cloud deployments when it comes to digital forensics. The application of IPCFA sheds light on the various capabilities of the public clouds that enable forensics investigators to capture authentic forensic data more easily than expected. The proof-of-concept deployment (zoolprod), while experimental, is indeed representative of a major deployment highlighted by public cloud providers today—that is the, "*3-tier architecture*" IPCFA was carried out successfully in the Azure context and was able to generate forensically-sound data that meets the U.S. federal digital evidence rules as well as the best practices recommended for an authentic digital evidence. All this was done by utilizing cloud subsystems and components to generate the data copies; thus, it can comply with FRE 902(13). All generated data including logs and disk images were directly logged and captured from system commands during the generation process. The use of scripted API calls would make the application of IPCFA in any cloud provider a replicable process that could be re-executed anytime to validate evidence claims and attest to the effectiveness of the methodology.

Implications of Practice

This research has touched upon numerous areas related to the technological and regulatory aspects of cloud forensics. These areas affect all stakeholders involved in the forensic process, from large and small businesses and cloud services providers to forensic examiners and practitioners, court staff, and policymakers.

In the past decade, large organizations and enterprises have been proactively preparing for the worst, putting comprehensive incident response frameworks and policies in place. An integral part of incident response activities is the handling of digital forensics evidence. Standards and regulatory bodies such as NIST SP800-86 (Kent et al., 2006), RFC 3227 (Killalea & Brezinski, 2002), and ISO 27043:2015 (ISO, 2015) have published guidelines and standards to provide organizations with formalized processes and governance activities to reinforce their cyber resilience. While these standards have substantially helped organizations to enable forensic practitioners to better respond to security incidents requiring digital evidence, the standards do not address cloud forensics.

The present research provides recommendations for organizations of any size on how to enable cloud forensics from the technical, legal, and organizational points of view. The delivered artifact educates the technical team on the specifications and processes of cloud forensics. It also provides the technical and legal knowledge needed for more efficient discussions between the legal and technical sides of the organization. If integrated into one of the above standards, the artifact can help fill in the gap of cloud forensic readiness in organizations and enterprises.

While cybersecurity enhancements have always seemed to revolve around large businesses, organizations, and enterprises, hackers have begun to realize that they can destabilize large businesses by compromising the weak cybersecurity of small businesses. A small business can act as a doorway for attackers to be able to gain a foothold in larger organizations, as small businesses are always behind when it comes to cybersecurity (Rebner, 2019). According to the Cyber Readiness Institute (Josue, 2020) only 40% of the small businesses that participated in a survey performed in late March 2020 have implemented any kind of cybersecurity-related policy. Yet, even without established cybersecurity policies, which are the foundation for any cybersecurity efforts, small businesses are still leading the industry in the adoption of public cloud services (Lava, 2020). The present research addresses the cybersecurity needs of small businesses by providing them with useful guidelines for addressing cloud forensics. These guidelines recommend having an incident response policy that encompasses digital forensic efforts, explain the legal side of digital forensics in terms easily understood by typical IT admins, and propose IPCFA as a practical methodology that can be adopted as part of the business's incident response processes and procedures. IPCFA promotes a cost-effective forensic acquisition process by abstracting the legal terms so the digital evidence can be acquired by the local IT staff with appropriate background.

The lack of formal standards among public cloud providers makes the work of forensics practitioners very unpredictable. To investigate an incident in the public cloud today, the forensics examiner must have detailed knowledge about that specific cloud provider's platforms, architecture, and offered services; otherwise, the examiner will not be able to successfully extract sufficient forensic data to provide court-ready evidence. While it might seem that providers offer very similar services—for instance, IaaS components are similar—there are no standards or documents that guide the examiner on how to approach any cloud with confidence. IPCFA proposes those generalized procedures, offering a starting point for examiners to undertake cloud forensics in any public cloud provider with confidence. While IPCFA is high level, it still provides low-level details to help the forensics examiner make correct decisions and allocate appropriate data sources. In addition, this work also sheds light on the importance of utilizing cloud-native tools and APIs to acquire data for a forensics case; this is one way that public cloud services providers can change their platform to be forensically-enabled without changing their architecture. Exposing more APIs to the end users and allowing them to collect data from the various data sources (volatile, non-volatile, and supporting data) helps remove a lot of the burden and challenges related to cloud forensics today.

In cloud forensics, one challenge is the lack of relevant literature for the court staff (Grispos et al., 2012; Simou et al., 2016). During a trial, an expert witness can present evidence that has been collected from a cloud system with the possible challenge of having to explain to jurors what cloud computing is and how the witness generated the evidence without confiscating a physical device. In addition, jurors and judges might not be able to make an informed decision without understanding the concept of cloud forensics and how it differs from traditional digital forensics. In Australia, Adams et al. (2013) proposed a digital forensic model that can be easily explained in court. It uses UML diagrams and textual representations to address court staff, making the forensic process clear so judges and juries can make informed decisions.

The present artifact and its research, including IPCFA, can benefit court systems by preparing trial staff for litigations involving cloud computing or cloud evidence, more specifically the rapidly adopted public cloud computing model. This work provides a simple breakdown of the differences between traditional forensics and cloud forensics as well as an explanation of those differences. This research also provides IPCFA with a flowchart that represents the various stages of the acquisition process and where each process takes place. This can be coupled with one of the digital forensics models that puts some emphasis on cloud forensics such as Martini and Choo (2012), creating a useful resource to the average juror who is usually an individual from the general public who might not have heard of or used cloud computing in a business context. Furthermore, the research provides attorneys with the means to prepare for trials and cover their bases with respect to their claimed evidence. If ALR is executed prior to the court date, attorneys can get a reasonable probability of how court-ready their evidence is.

While this research focused on forensic acquisition carried out by the plaintiff or delegated third party, law enforcement agencies (LEAs) and policymakers can also benefit from this study. LEAs can adopt and use IPCFA just as a civilian forensic practitioner would. Moreover, having a documented process to follow to perform the acquisition can make a big difference in the effort required to train and prepare forensic practitioners. The lack of standards, procedures, guidelines, and certified tools related to cloud forensics should also encourage LEAs and policymakers to dedicate more resources to fill the gap in this area. This research also informs policymakers about the current gap between digital forensics and cloud forensics in terms of legislation and digital evidence rules. While federal rules of evidence do apply to cloud evidence, these rules should be reviewed and refined to address the cloud as a possible source of forensic data as well as cloud chain of custody in order to better articulate how to legally handle and authenticate cloud evidence.

CHAPTER SEVEN

CONCLUSION

This research aimed to study and identify which digital forensic tools and acquisition techniques are applicable for court-grade evidence acquisition from IaaS deployments in the public cloud. By analyzing the existing digital forensics models, processes, best practices, legal requirements for digital evidence, available digital forensic tools libraries, and published guidelines from the largest public cloud providers in the US, we conclude that there is a dearth of structured methodologies and processes to guide practitioners in digital forensic investigations in the realm of public clouds, including IaaS. We have also concluded that there is no lack of non-volatile data forensic acquisition tools to for IaaS, rather the opposite, as cloud-native tools makes the acquisition simple, effective, and efficient. This research proposed an IaaS Public Cloud Forensic Acquisition (IPCFA) practical methodology to acquire sound forensic evidence from public cloud IaaS deployments. IPCFA provides direction and guidance to digital forensic investigators who have cases involving IaaS public cloud environments. IPCFA translates the legal needs into technical strategies and tools in order for them to collect data considered admissible in courts of law.

The proposed methodology enables small organizations and private investigators to pursue a cloud forensics case while not jeopardizing the authenticity of the collected evidence. It enables the extraction of court-grade evidence without going through the cloud service provider. Based on the native tools available from the public cloud providers and the existing digital forensic tools, we suggest it is possible to execute a forensic acquisition in an IaaS public cloud environment, and it can be comparable to a physical acquisition of servers from a datacenter. We recommend following a structured methodology, such as the proposed IPCFA, that addresses all possible evidence admissibility issues related to evidence authenticity and integrity.

Research Limitations

Although this design-based research developed a useful methodology for obtaining court-admissible evidence, the research performed does have limitations. The original proposal and scope of this research was to develop a practical methodology for IaaS acquisition in public clouds and to validate the methodology by sharing it with experts in the field. The validation was planned to start from the design phase by sharing the expertise, capturing feedback, and adjusting the artifact accordingly until reaching acceptance from the experts. The last bit of validation was planned to simulate the methodology application in a real court case to show its effectiveness. The validation strategies had to change in response to developments during the research process as well as the limited access to resources. The validation shifted to an observational case study instead, which is not a strong evaluation methodology.

Recruiting experts with specific criteria to perform a Delphi study or form a focus group was not successful. The candidate pool was not strong enough; there were not enough digital forensics, cloud, and practical legal knowledge experts to form strong opinions and receive sound feedback to validate our work. Testing the methodology on past cases was not achievable during the research timeframe due to our limited access to previously archived forensic cases that involved public cloud or IaaS. While there have been some potential cases, access to the complete data behind these cases was not available to us so a genuine and effective comparison could not be created.

More changes to the scope of artifact validation and verification were introduced due to the publication of SWGDE's *Best Practices for Digital Evidence Acquisition from Cloud Service Providers* (2020). SWGDE (2020) came with acquisition recommendations that were very close to the ones proposed by IPCFA phases and the suggested processes; thus, ALR was created, and the Azure experimental case study was introduced. While the Azure case study focused on the technical application of IPCFA and observed the outcome, ALR integrated the various legal requirements for admissible evidence in U.S. courts of law and connected them to the technical processes followed and executed by forensic examiners during cybercrime forensic investigations.

As the researchers do not have the practical legal experience nor professional exposure to the U.S. court systems, we have relied heavily on the limited literature to highlight the connection between the technical and legal requirements of digital evidence admissibility, and thus generated the ALR to quantify technical decisions toward court admissibility. While the ALR might not be the sturdiest strategy to perform such an exercise, the literature shows no other tool or mechanism available today that attempts to provide a weighted probability of digital evidence admissibility.

Future Research

Continuation of this research would be beneficial to the digital forensic community; thus, addressing the limitations provide a starting point. Applying more rigorous evaluation methodologies to IPCFA such as a well-assembled focus group study or a Delphi study would help address any gaps that might have missed with the observational study, solidifying the generated artifact. Another opportunity to extend this research is to reframe the proposed artifact to address any drastic changes underlying technical and legal assumptions.

Looking forward, expanding on IPCFA to encompass the other cloud services model would be beneficial to the digital forensics' community. While the literatures show some research exists in the areas of cloud storage and SaaS, the focus has been on the tooling and technical collection only. Adding the legal aspects to such research will help address the gap we have today between investigations in the public cloud realm and our court system, which demonstrates the real value of digital forensics versus corporate incident response practices. IPCFA expansion can also be on the ALR front, as it can be extended and hardened to be able to produce a very strong and accurate probability about digital evidence admissibility into the U.S. Court system.

While it was a coincidence that IPCFA and SWGDE (2020) happened to have very similar structures with three major phases and similar definitions of each phase, this opens opportunities for collaboration. Recommendations and proposed processes from IPCFA and SWGDE (2020) can be merged to form a solid cloud forensic acquisition framework that could encompass cases of both CSP-involved and CSP-free investigations. This framework could be developed as extensible to allow for other public cloud service models such as SaaS or CaaS to be integrated. The framework would act as a baseline to allow for faster developments of such extensions.

From technical perspective, there are still a lot of areas in the public cloud that has been untouched to date, such as serverless and PaaS forensics, where the CSP takes care of your infrastructure elements, and you only handle your application and code.   This brings with it all possible challenges to cloud forensics as your reliance on the cloud providers gets stronger and deeper. Another arguable area of research would be investigations that embrace Microservices architecture and Containerized platforms (CaaS). While they do run on an IaaS or PaaS layers, each container provides a service that might fall within an investigation scope. These are just a sample of the open areas of research when it comes to cloud forensics.

REFERENCES

Adams, R. (2013). The emergence of cloud storage and the need for a new digital forensic process model. In K. Ruan (Ed.), *Cybercrime and cloud forensics: Applications for investigation processes*. Hershey, PA: IGI Global. Retrieved from https://researchrepository.murdoch.edu.au/id/eprint/19431/1/emergence_of_cloud_storage.pdf

Adams, R., Hobbs, V., & Mann, G. (2013). The Advanced Data Acquisition Model (Adam): A process model for digital forensic practice. *Journal of Digital Forensics, Security and Law, 8*. doi:10.15394/jdfsl.2013.1154

Agarwal, A., Gupta, M., Gupta, S., & Gupta, S. C. (2011). Systematic digital forensic investigation model. *International Journal of Computer Science and Security, 5*(1), 118-131.

Allen, T. A. (2017, May 8). Computer Forensics Tool Testing Program (CFTT). *NIST*. Retrieved from https://www.nist.gov/itl/ssd/software-quality-group/computer-forensics-tool-testing-program-cftt

Alqahtany, S., Clarke, N., Furnell, S., & Reich, C. (2015). Cloud forensics: A review of challenges, solutions, and open problems. *2015 International Conference on Cloud Computing (ICCC)*, 1–9. doi:10.1109/CLOUDCOMP.2015.7149635

Alqahtany, S., Clarke, N., Furnell, S., & Reich, C. (2016). A forensic acquisition and analysis system for IaaS. *Cluster Computing, 19*(1), 439–453. doi:10.1007/s10586-015-0509-x

Amazon. (2015). Amazon.com help: Amazon information request reports. Retrieved from https://www.amazon.com/gp/help/customer/display.html?nodeId=GYSDRGWQ2C2CRYEF

Ayers, R., Brothers, S., & Jansen, W. (2014). *Guidelines on mobile device forensics* (NIST SP 800-101r1; p. NIST SP 800-101r1). National Institute of Standards and Technology. doi:10.6028/NIST.SP.800-101r1

Barrett, D. (2018). Forensic investigations in cloud computing. In D.B.A.M. Khosrow-Pour (Ed.), *Encyclopedia of information science and technology* (4th ed.) (pp. 1356-1385). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-2255-3.ch116

Barrett, D. & Kipper, G. (2010). *Virtualization and forensics: A digital forensic investigator's guide to virtual environments.* Burlington, MA: Syngress.

Baryamureeba, V., & Tushabe, F. (2004). The enhanced digital investigation process. https://www.dfrws.org/sites/default/files/session-files/paper-the_enhanced_digital_investigation_process_model.pdf

Beebe, N. L., & Clark, J. G. (2005). A hierarchical, objectives-based framework for the digital investigations process. *Digital Investigation*, *2*(2), 147–167. doi:10.1016/j.diin.2005.04.002

Build a Rubric. (2017). *The University of Texas at Austin.* Retrieved April 15, 2021 from https://facultyinnovate.utexas.edu/sites/default/files/build-rubric.pdf

Building a Cloud-Specific Incident Response Plan. (2017). *Amazon Web Services.* https://aws.amazon.com/blogs/publicsector/building-a-cloud-specific-incident-response-plan/

Center, E. P. I. (n.d.). EPIC - The CLOUD Act. Retrieved June 14, 2019, from https://epic.org/privacy/cloud-act/

Cloud Compliance—Regulations & Certifications. (n.d.). *Google Cloud.* Retrieved September 23, 2019, from https://cloud.google.com/security/compliance/

Cohen, F. (2013). Challenges to digital forensic evidence in the cloud. In K. Ruan (Ed.), *Cybercrime and cloud forensics: Applications for investigative processes* (pp. 59-78). Hershey, PA, IGI Global. doi:10.4018/978-1-4666-2662-1.Ch003.

Compliance in the trusted cloud | Microsoft Azure. (n.d.). Retrieved September 23, 2019, from https://azure.microsoft.com/en-us/overview/trusted-cloud/compliance/

Compliance Programs—Amazon Web Services (AWS). (n.d.). *Amazon Web Services, Inc.* Retrieved September 23, 2019, from https://aws.amazon.com/compliance/programs/

Data incident response process | Documentation. (2018). *Google Cloud*. https://cloud.google.com/security/incident-response/

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993). https://supreme.justia.com/cases/federal/us/509/579/

Decusatis, C., Carranza, A., Ngaide, A., Zafar, S., & Landaez, N. (2015). Methodology for an open digital forensics model based on CAINE. *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, 935–940. doi:10.1109/CIT/IUCC/DASC/PICOM.2015.61

Dickson, K. B. (2011). *Admissibility and evidentiary issues with electronic evidence.* National CLE Conference. Retrieved from https://silo.tips/download/american-bar-association-section-of-labor-and-employment-law-national-cle-confer

Dykstra, J. (2013a). *Digital forensics for infrastructure-as-a-service cloud computing* (Doctoral dissertation). Retrieved from https://www.csee.umbc.edu/~dykstra/dissertation

Dykstra, J. (2013b). Seizing electronic evidence from cloud computing environments. In K. Ruan (Ed.), *Cybercrime and cloud forensics: Applications for investigation processes* (pp. 156-185). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-2662-1.Ch007.

Dykstra, J., & Sherman, A. T. (2011). Understanding issues in cloud forensics: Two hypothetical case studies. *Annual ADFSL Conference on Digital Forensics, Security, and Law*. https://mdsoar.org/bitstream/handle/11603/12821/viewcontent.pdf?sequence=1

Earford.pdf. (n.d.). Retrieved January 31, 2019, from http://www.cs.tufts.edu/comp/116/archive/fall2016/earford.pdf

Federal Rules of Civil Procedure. (2019, December 1). United States Courts. Retrieved September 23, 2019, from https://www.uscourts.gov/sites/default/files/federal_rules_of_civil_procedure_dec_1_2019_0.pdf

Federal Rules of Evidence. (2017, December 1). United States Courts. Retrieved April 15, 2021, from https://www.uscourts.gov/sites/default/files/evidence-rules-procedure-dec2017_0.pdf

Forensic examination of digital devices in civil litigation: The legal, ethical and technical traps. (2016). *The Professional Lawyer*, *24*(1). https://www.americanbar.org/groups/professional_responsibility/publications/professional_lawyer/2016/volume-24-number-1/forensic_examination_digital_devices_civil_litigation_legal_ethical_and_technical_traps/

Freet, D., Agrawal, R., John, S., & Walker, J. J. (2015). Cloud forensics challenges from a service model standpoint: IaaS, PaaS and SaaS. *Proceedings of the 7th International Conference on Management of Computational and Collective Intelligence in Digital EcoSystems - MEDES '15*, 148–155. doi:10.1145/2857218.2857253

Garfinkel, S., Malan, D., Dubec, K.-A., Stevens, C., & Pham, C. (2006). Advanced forensic format: An open extensible format for disk imaging. In M. S. Olivier & S. Shenoi (Eds.), *Advances in digital forensics II* (Vol. 222, pp. 13–27). New York: Springer. doi:10.1007/0-387-36891-4_2

Gartner. (2019, April 2). Gartner forecasts worldwide public cloud revenue to grow 17.5 percent in 2019. Retrieved from https://www.gartner.com/en/newsroom/press-releases/2019-04-02-gartner-forecasts-worldwide-public-cloud-revenue-to-g

Gonzales, A. R., Schofield, R. B., & Hagy, D. W. (2007a). *Digital evidence in the courtroom: A guide for law enforcement and prosecutors*. U.S. Department of Justice. Retrieved from https://www.ojp.gov/pdffiles1/nij/211314.pdf

Gonzales, A. R., Schofield, R. B., & Hagy, D. W. (2007b). *Investigations involving the Internet and computer networks*. U.S. Department of Justice. Retrieved from https://www.ojp.gov/pdffiles1/nij/210798.pdf

Goodison, S. E., Davis, R. C., & Jackson, B. A. (2005). *Digital evidence and the U.S. criminal justice system: Identifying technology and other needs to more effectively acquire and utilize digital evidence*. Priority Criminal Justice Needs Initiative. Retrieved from https://www.ojp.gov/pdffiles1/nij/grants/248770.pdf

Green, M. (2021). *Mgreen27/Invoke-LiveResponse* [PowerShell]. https://github.com/mgreen27/Invoke-LiveResponse (Original work published 2018)

Grimm, P. W., & Brady, K. F. (2007). *Checklist of Potential Authentication Methods*.

Grispos, G., Storer, T., & Glisson, W. B. (2012). Calm before the storm: The challenges of cloud computing in digital forensics. *International Journal of Digital Crime and Forensics, 4*(2), 28-48. doi:10.4018/jdcf.2012040103.

Grobler, M. & von Solmes, B. (2009). *A best practice approach to live forensic acquisition.* Paper presented at Information Security South Africa Conference 2009. Retrieved from https://www.researchgate.net/profile/Marthie-Grobler-2/publication/220803169_A_Best_Practice_Approach_to_Live_Forensic_Acquisition/links/56bc790208aebaa770e8b9c2/A-Best-Practice-Approach-to-Live-Forensic-Acquisition.pdf

Guven, N. (2017). *Your responsibility in cloud security*.

Hartman, K. (2018). *Digital forensic analysis of Amazon Linux EC2 instances*. Retrieved from https://www.sans.org/reading-room/whitepapers/cloud/digital-forensic-analysis-amazon-linux-ec2-instances-38235.

Health Information Privacy. (2015, August 26). U.S. Department Health and Human Services. https://www.hhs.gov/hipaa/index.html

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian Journal of Information Systems, 19*(2):87-92. Retrieved from https://www.researchgate.net/profile/Alan-Hevner/publication/254804390_A_Three_Cycle_View_of_Design_Science_Research/links/0c96053b4a9f2d7adc000000/A-Three-Cycle-View-of-Design-Science-Research.pdf

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in Information Systems research. *MIS Quarterly*, *28*(1), 75–105. doi:10.2307/25148625

Ieong, R. S. C. (2006). FORZA – Digital forensics investigation framework that incorporate legal issues. *Digital Investigation*, *3*, 29–36. doi:10.1016/j.diin.2006.06.004

ISO. (2015). *ISO/IEC 27043:2015(en), Information technology—Security techniques—Incident investigation principles and processes*. https://www.iso.org/obp/ui/#iso:std:iso-iec:27043:ed-1:v1:en

ISO/IEC 27001 Information security management. (n.d.). ISO. Retrieved September 13, 2019, from http://www.iso.org/cms/render/live/en/sites/isoorg/home/standards/popular-standards/isoiec-27001-information-securit.html

Jansen, W., & Ayers, R. P. (2004). *Guidelines on PDA forensics* (NIST SP 800-72; 0 ed., p. NIST SP 800-72). National Institute of Standards and Technology. doi:10.6028/NIST.SP.800-72

Josue. (2020, April 5). The new realities of a remote workforce increase cybersecurity concerns for half of all small business owners, but policies, training still lag, Cyber Readiness Institute Survey finds. *Cyber Readiness Institute.* Retrieved from https://cyberreadinessinstitute.org/news-and-events/the-new-realities-of-a-remote-workforce-increase-cybersecurity-concerns-for-half-of-all-small-business-owners-but-policies-training-still-lag-cyber-readiness-institute-survey-finds/

Kent, K., Chevalier, S., Grance, T., & Dang, H. (2006). *Guide to integrating forensic techniques into incident response* (NIST SP 800-86; 0 ed., p. NIST SP 800-86). National Institute of Standards and Technology. doi:10.6028/NIST.SP.800-86

Killalea, T., & Brezinski, D. (2002). *Guidelines for evidence collection and archiving*. https://tools.ietf.org/html/rfc3227

Lava, Shari. (2020). *U.S. small and medium-sized business: The state of cloud adoption*. IDC: The Premier Global Market Intelligence Company. https://www.idc.com/getdoc.jsp?containerId=US43649719

Law Enforcement Requests Report – Microsoft Corporate Social Responsibility. (2018). Microsoft. Retrieved from https://www.microsoft.com/en-us/corporate-responsibility/lerr

Legal Issues with Cloud Forensics—ProQuest. (2015). Retrieved from https://www.ezproxy.dsu.edu:2085/docview/1681299975?rfr_id=info%3Axri%2Fsid%3Aprimo

Legal process for user data requests FAQs—Transparency Report Help Center. (n.d.). Retrieved February 16, 2019, from https://support.google.com/transparencyreport/answer/7381738?hl=en

Mapping the forensic standard ISO/IEC 27037. (2013). *Cloud Security Alliance.* https://cloudsecurityalliance.org/artifacts/mapping-the-forensic-standard-isoiec-27037-to-cloud-computing/

Martini, B., & Choo, K.-K. R. (2012). An integrated conceptual digital forensic framework for cloud computing. *Digital Investigation*, *9*(2), 71–80. doi:10.1016/j.diin.2012.07.001

Martini, B., & Choo, K.-K. R. (2013). Cloud storage forensics: OwnCloud as a case study. *Digital Investigation*, *10*(4), 287–299. doi:10.1016/j.diin.2013.08.005

Marturana, F., Tacconi, S., & Italiano, G. F. (2015). A forensic-as-a-service delivery platform for law enforcement agencies. In Information Resources Management Association (Ed.), *Cloud technology: Concepts, methodologies, tools, and applications* (pp. 2288-2306). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-6539-2.ch109

McKemmish, R. (1999). What is forensic computing*? Trends & Issues in Crime and Criminal Justice, 118*. Retrieved by https://www.aic.gov.au/sites/default/files/2020-05/tandi118.pdf

McKemmish, R. (2008). When is digital evidence forensically sound? In I. Ray & S. Shenoi (Eds.), *Advances in digital forensics IV* (pp. 3–15). US: Springer. https://doi.org/10.1007/978-0-387-84927-0_1

Mell, P., & Grance, T. (2011). The NIST definition of cloud computing. Retrieved from http://faculty.winthrop.edu/domanm/csci411/Handouts/NIST.pdf

Michaels, D. (2008). *Doubt is their product: How industry's assault on science threatens your health*. Oxford: Oxford University Press.

Miranda Lopez, E., Moon, S. Y., & Park, J. H. (2016). Scenario-Based digital forensics challenges in cloud computing. *Symmetry*, *8*(10), 107. doi:10.3390/sym8100107

Montasari, R. (2017). A standardised data acquisition process model for digital forensic investigations. *Journal of Information and Computer Security 9*(3), 229-249.

Nelson, S. D., & Simek, J. W. (2009). *Computer Forensics Best Practices*. 24.

Neware, R., & Khan, A. (2018). Cloud computing digital forensic challenges. *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 1090–1092. doi:10.1109/ICECA.2018.8474838

NIJ. (2008). Electronic crime scene investigation: A guide for first responders, Second Edition: (511502010-005) [Data set]. American Psychological Association. doi:10.1037/e511502010-005

NIJ. (2016). Digital evidence and forensics. National Institute of Justice. https://www.nij.gov:443/topics/forensics/evidence/digital/pages/welcome.aspx

NIST Cloud Computing Forensic Science Working Group. (2014). NIST cloud computing forensic science challenges (NIST Internal or Interagency Report (NISTIR) 8006 (Draft)). *National Institute of Standards and Technology*. https://csrc.nist.gov/publications/detail/nistir/8006/draft

N-tier architecture style—Azure Application Architecture guide. (2018). https://docs.microsoft.com/en-us/azure/architecture/guide/architecture-styles/n-tier

Official PCI Security Standards Council site. (n.d.). Retrieved September 13, 2019, from https://www.pcisecuritystandards.org/

O'Reilly, L., & Cyr, T. (2006). Creating a rubric. *University of Colorado Denver*. http://www.ucdenver.edu/faculty_staff/faculty/center-for-faculty-development/Documents/Tutorials/Rubrics/1_what_is/index.htm

Orr, D. A., & White, P. (2018). Current state of forensic acquisition for IaaS cloud services. *Journal of Forensic Sciences*, *10*(1). 555778. doi:10.19080/JFSCI.2018.10.555778.

Orton, I., Alva, A., & Endicott-Popovsky, B. (2013). Legal process and requirements for cloud forensic investigations. In Keyun Ruan (Ed.), *Cybercrime and cloud forensics: Applications for investigation processes*. Hershey, PA, IGI Global. doi: 10.4018/978-1-4666-2662-1.ch008

Palmer, G. (2001). A road map for digital forensic research. *Dfrws*. https://www.dfrws.org/conferences/dfrws-usa-2001/sessions/road-map-digital-forensic-research

Quick, D., & Choo, K.-K. R. (2013a). Dropbox analysis: Data remnants on user machines. *Digital Investigation*, *10*(1), 3–18. doi:10.1016/j.diin.2013.02.003

Quick, D., & Choo, K.-K. R. (2013b). Digital droplets: Microsoft SkyDrive forensic data remnants. *Future Generation Computer Systems*, *29*(6), 1378–1394. doi: 10.1016/j.future.2013.02.001

Quick, D., & Choo, K.-K. R. (2013c). Forensic collection of cloud storage data: Does the act of collection result in changes to the data or its metadata? *Digital Investigation*, *10*(3), 266–277. doi:10.1016/j.diin.2013.07.001

Quick, D., & Choo, K.-K. R. (2014). Google Drive: Forensic analysis of data remnants. *Journal of Network and Computer Applications*, *40*, 179–193. doi:10.1016/j.jnca.2013.09.016

Ramirez, R., Mukherjee, M., Vezzoli, S., & Kramer, A. M. (2015). Scenarios as a scholarly methodology to produce “interesting research.” *Futures*, *71*, 70–87. doi:10.1016/j.futures.2015.06.006

Rebner, S. (2019). *Council post: The state of cybersecurity pertaining to small business*. *Forbes*. Retrieved from https://www.forbes.com/sites/theyec/2019/09/18/the-state-of-cybersecurity-pertaining-to-small-business/

rkarlin. (2018). *Respond to security incidents with Azure Security Center*. Retrieved from https://docs.microsoft.com/en-us/azure/security-center/security-center-incident-response

Rule 26. Duty to Disclose; General Provisions Governing Discovery. F.R.C.P. (2015). Retrieved February 2, 2021, from https://www.law.cornell.edu/rules/frcp/rule_26

Rule 34. Producing Documents, Electronically Stored Information, and Tangible Things, or Entering onto Land, for Inspection and Other Purposes. F.R.C.P. (2015). Retrieved from https://www.law.cornell.edu/rules/frcp/rule_34

Rule 45. Subpoena. F.R.C.P. (2013). Retrieved from https://www.law.cornell.edu/rules/frcp/rule_45

Rule 105. Limiting Evidence That Is Not Admissible Against Other Parties or for Other Purposes. F.R.E. (2011). Retrieved from https://www.law.cornell.edu/rules/fre/rule_105

Rule 401. Test for Relevant Evidence. F.R.E. (2011). Retrieved from https://www.law.cornell.edu/rules/fre/rule_401

Rule 402. General Admissibility of Relevant Evidence. F.R.E. (2011) Retrieved from https://www.law.cornell.edu/rules/fre/rule_402

Rule 403. Excluding Relevant Evidence for Prejudice, Confusion, Waste of Time, or Other Reasons. F.R.E. (2011). Retrieved from https://www.law.cornell.edu/rules/fre/rule_403

Rule 702. Testimony by Expert Witnesses. F.R.C.P. (2011). Retrieved from https://www.law.cornell.edu/rules/fre/rule_702

Rule 801. Definitions That Apply to This Article; Exclusions from Hearsay. F.R.E. (2014). Retrieved from https://www.law.cornell.edu/rules/fre/rule_801

Rule 901. Authenticating or Identifying Evidence. F.R.E. (2011) Retrieved from https://www.law.cornell.edu/rules/fre/rule_901

Rule 902. Evidence That Is Self-Authenticating. F.R.E. (2011) Retrieved from. https://www.law.cornell.edu/rules/fre/rule_902

Rule 903. Subscribing Witness. F.R.E. (2011). Retrieved from https://www.law.cornell.edu/rules/fre/rule_903

Rule 1003. Admissibility of Duplicates. F.R.E. (2011). Retrieved from https://www.law.cornell.edu/rules/fre/rule_1003

Ryan, M. M. (2009a). *Daubert Standard*. LII / Legal Information Institute. https://www.law.cornell.edu/wex/daubert_standard

Ryan, M. M. (2009b). *Frye Standard*. LII / Legal Information Institute. https://www.law.cornell.edu/wex/frye_standard

Sampana, S. S. (2019). FoRCE (Forensic Recovery of Cloud Evidence): A digital cloud forensics framework. *2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3)*, 212–212. doi:10.1109/ICGS3.2019.8688215

Scanlon, M. (2017). Enabling the remote acquisition of digital forensic evidence through secure data transmission and verification. *ArXiv:1712.02529 [Cs]*. Retrieved from http://arxiv.org/abs/1712.02529

Schatz, B. (n.d.). AWS EC2 cloud storage acquisition with Evimetry. *Schatz Forensic*. Retrieved January 31, 2019, from https://schatzforensic.com/insideout/2017/09/aws-ec2-cloud-storage-acquisition-with-evimetry/

Scientific Working Group on Digital Evidence (SWGDE). (2014). *SWGDE capture of live systems*. Retrieved from https://www.irisinvestigations.com/wp-content/uploads/2019/05/SWGDE-Capture-of-Live-Systems-090514.pdf

SWGDE. (2018a). *SWGDE best practices for cComputer orensic acquisitions*. Retrieved from https://drive.google.com/file/d/1KeEI1DUkSE2DSPZyPFEFIGfzbZS3-zZC/view

SWGDE. (2018b) *SWGDE best practices for digital evidence collection*. Retrieved from https://drive.google.com/file/d/1zP4OgpRrj-t9sVGNcqndqIgsemq7u5XQ/view

SWGDE. (2020). *SWGDE best practices for digital evidence acquisition from cloud service providers*. Retrieved from https://drive.google.com/file/d/1j_0HoVGdRigyqy-DKna4AoasIQUDw0Va/view

SWGDE & International Organization on Digital Evidence (IOCE). (2000). *Digital evidence: Standards and principles*. Retrieved from https://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/april2000/swgde.htm

Shah, J. J., & Malik, L. G. (2014). An approach towards digital forensic framework for cloud. *2014 IEEE International Advance Computing Conference (IACC)*, 798–801. doi:10.1109/IAdCC.2014.6779425

Shared responsibility in the cloud—Microsoft Azure. (n.d.). Retrieved January 12, 2021, from https://docs.microsoft.com/en-us/azure/security/fundamentals/shared-responsibility

Shared Responsibility Model Explained. (n.d.). *Cloud Security Alliance.* Retrieved January 12, 2021, from https://cloudsecurityalliance.org/blog/2020/08/26/shared-responsibility-model-explained/

Shared Responsibility Model—Amazon Web Services. (n.d.). *Amazon Web Services, Inc.* Retrieved January 12, 2021, from https://aws.amazon.com/compliance/shared-responsibility-model/

Simou, S., Kalloniatis, C., Gritzalis, S., & Mouratidis, H. (2016). A survey on cloud forensics challenges and solutions. *Security and Communication Networks*, *9*(18), 6285–6314. doi:10.1002/sec.1688

sptramer. (n.d.). *Get started with Azure CLI*. Retrieved February 12, 2019, from https://docs.microsoft.com/en-us/cli/azure/get-started-with-azure-cli

Stone, A. (2015). Chain of custody: How to ensure digital evidence stands up in court. *GovTechWorks*. Retrieved from https://www.govtechworks.com/chain-of-custody-how-to-ensure-digital-evidence-stands-up-in-court/

Toft, A. (2018, June 22). New rules for self-authenticating electronic evidence. *American Bar Association.* Retrieved from https://www.americanbar.org/groups/litigation/committees/trial-evidence/practice/2018/new-rules-electronic-evidence/

U.S. Department of Justice; Office of Justice Programs; National Institute of Justice. (2004). *Forensic Examination of Digital Evidence: A Guide for Law Enforcement: (378092004-001)* [Data set]. American Psychological Association. doi:10.1037/e378092004-001

Watson, D. & Jones, A. (2013). *Digital forensics processing and procedures: Meeting the requirements of ISO 17020, ISO 17025, ISO 27001 and best practice requirements*. Waltham, MA, Syngress.

Zawoad, S., & Hasan, R. (2016). Trustworthy digital forensics in the cloud. *Computer, 49*(3), 78–81. doi:10.1109/MC.2016.89

Zawoad, S., Hasan, R., & Skjellum, A. (2015). OCF: An Open Cloud Forensics model for reliable digital forensics. *2015 IEEE 8th International Conference on Cloud Computing*, 437–444. doi:10.1109/CLOUD.2015.65

APPENDICES

APPENDIX A: EXISTING DF ACQUISITION METHODOLOGIES AND PROCEDURES

Table 15 - Existing DF Acquisition Methodologies and Procedures

Methodology/Procedure	Phases and processes
"Guidelines for Evidence Collection and Archiving" (Killalea & Brezinski, 2002)	- List systems from which the evidence will be collected. - Establish what is likely to be relevant and admissible (collect less, more efficient). - Plan order of volatility for each system. - Remove external avenues for change. - Follow order of volatility; collect the evidence with relevant tools. - Record the extent of the system's clock drift. - Question what else may be evidence to be collected. - Document each step. - Document people present at the crime scene. - Generate checksums and cryptographically sign the collected evidence.
"Guide to Integrating Forensic Techniques into Incident Response" (Kent et al., 2006)	- Develop a plan for acquisition. - Likely value - Volatility - Amount of effort required - Acquire the data. - Verify the integrity of the data.

"Electronic Crime Scene Investigation: A Guide for First Responders, Second Edition" (NIJ, 2008)	- If the computer is powered on, and there is a digital evidence seizure trained personnel, do not turn it off. - Locate and secure all evidence within the scope of authority for the specific circumstances. - Document, log, and photograph all computers, devices, connections, cables, and power supplies. - Log and secure all evidence according to agency policies pending forensic examination.
"Digital Forensics Processing and Procedures: Meeting the Requirements of ISO 17020, ISO 17025, ISO 27001 and Best Practice Requirements" (Watson & Jones, 2013)	- Inform relevant parties of consequences of live capture. - Establish remote connectivity. - Start acquisition in order of volatility. - Take network traffic dumps. - Reserve/protect the evidence. - Intercept evidence at scene. - Create detailed records of scene.
"SWGDE Best Practices for Digital Evidence Collection" (SWGDE, 2018b)	- Prepare possible acquisition tools, destination media, and to collect memory and ancillary data. - Consider acquisition location, environment, encryption, and boot loaders restrictions. - Preview the contents of potential data sources prior to acquisition to reduce the amount of data acquired (triage). - Execute the appropriate acquisition method (physical, logical, targeted, clone):

	- If live acquisition is used, tools should execute trusted binaries from controlled media. - If live acquisition is used, software should execute at the lowest level of privilege needed to ensure all possible data is available for acquisition. - Validate the integrity of the data as acquired. - Document digital evidence acquisition per organizational policy. - Document chain of custody as required by organizational policy.
"SWGDE Best Practices for Digital Evidence Acquisition from Cloud Service Providers" (SWGDE, 2020)	- **Prior to acquisition phase:** - Identify the CSP, data sources, data types, timeline, and utilized services. This might be billing information for the cloud account, DNS records, or privacy policies. - Ask the CSP legally to preserve the data sought. This can be carried out via the involvement of law enforcements. - Identify and choose which acquisition methodology to use. If the CSP involvement is required, contact the CSP via their legal means. - **During acquisition phase:** - Document all evidentiary data via handwritten notes, screen captures, or photographs. - Acquire and do not omit all attached media, local and in the cloud. For media acquisition,

	follow the acquisition process outlined in (SWGDE, 2018a). - Confirm if the type of sought data can be acquired using the acquisition methodology selected before. - Acquire the data with the selected methodology. If the provider is asked to provide the data, keep in mind it might need to be unencrypted or transformed into another format to be processes. - If it was not possible to execute the selected methodology, and the acquisition fails, then take photographs and screenshots of the relevant data. - **After acquisition phase:** - Calculate integrity hash values for acquired data. If the CSP provided the data, re-check the digital signatures. - Verify that the executed acquisition methodology was able to acquire all required data. If data was furnished by the CSP, data still needs to be verified. - Document the whole process using the organization’s defined policies and procedures. - Document any data received from the CSP in digital media or storage. - Store all acquired data following the organizational policies and procedures that address digital evidence storage.

APPENDIX B: AZURE ACQUISITION SANDBOX

Table 16 - Azure Sandbox – Zool Corp. PROD Environment

Software/Tool name	Purpose
Linux CentOS 7.9	Web & App Linux servers
Nginx 2.5.6	Web servers – reverse proxy
WordPress 13.2	App servers – vulnerable CMS (WordPress)
Microsoft Windows 2016	DB servers
Microsoft SQL Server 2016	MySQL RDMS
Azure Load balancers	Traffic Load balancer between tiers
Azure Storage accounts	Storage account for all tiers

Table 17 - Azure Sandbox – Forensic Environment

Component name	Purpose
Microsoft Windows 10 Pro (with DF tools)	Forensic Server Image
Azure Blob storage	Immutable storage for evidence
Azure Fileshare	A file share to connect to forensic server
Azure Key Vault	Secret vault to save hash values

Table 18 - Azure Sandbox – Scripts to Deploy the Sandbox and Perform the Acquisition

Script name	Purpose
azDF_pocDeploy.ps1	Deploys the Zoolprod environment
azDF_pocDeploy.init	Cloud-init script
azDF_DFenvPrep.ps1	Deploys the forensic server environment
azDF_dsiskSnap.ps1	Acquires disk images
azDF_lnxMemSnap.ps	Acquires Linux memory images with LiMe
azDF_lnxMemSnap.sh	LiMe deployment bash script
azDF_winMemSnap.ps1	Acquires Windows memory images with WinpMem
azDF_winMemSnapLocal.ps1	Local execution of WinpMem
azDF_cloudLogs.ps1	Collect various logs from Azure subscription

All testing scripts can be found in the following GitHub repository: https://github.com/azdfir/ipcfa-poc

Sandbox Notes:

- All scripts have been created for testing purposes only, and they are not part of the IPCFA methodology.
- Requires two Azure pay-as-you-go subscriptions:
    - One hosts the DF image in a Shared Gallery.

  - One is the sandbox to pull the Shared Gallery image.
- Manually mount Azure File Share onto the forensic server.
- Manually mount immutable storage using NFS3.0 to the forensic server.
- Install zCopy in order to send screenshots and video captures to the protected blob.
- Use Azure Cloud Shell with either Azure PS or CLI to add hash values to the created Vault.