

Marshall University Marshall Digital Scholar

Weisberg Division of Computer Science Faculty
Research

Weisberg Division of Computer Science

7-2014

Forensiccloud: An Architecture for Digital Forensic Analysis in the Cloud


Cody Miller

Dae Glendowne

David Dampier
Marshall University, dampierd@marshall.edu

Kendall Blaylock

Follow this and additional works at: https://mds.marshall.edu/wdcs_faculty

 Part of the [Computer and Systems Architecture Commons](#), and the [Forensic Science and Technology Commons](#)

Recommended Citation

Miller, C., D. Glendowne, D. Dampier, and K. Blaylock, "Forensiccloud: An Architecture for Digital Forensics Analysis in the Cloud," *Journal of Cyber Security and Mobility*, Vol. 3, Num. 3, July 2014, pp. 231-262.

This Article is brought to you for free and open access by the Weisberg Division of Computer Science at Marshall Digital Scholar. It has been accepted for inclusion in Weisberg Division of Computer Science Faculty Research by an authorized administrator of Marshall Digital Scholar. For more information, please contact zhangj@marshall.edu, beachgr@marshall.edu.

Forensiccloud: An Architecture for Digital Forensic Analysis in the Cloud

Cody Miller, Dae Glendowne, David Dampier and Kendall Blaylock

*Distributed Analytics and Security Institute, Mississippi State University,
Mississippi State, MS, USA
{miller; dae; dampier}@dasi.msstate.edu, kblaylock@cse.msstate.edu*

Received 30 June 2014; Accepted 20 August 2014
Publication 7 October 2014

Abstract

The amount of data that must be processed in current digital forensic examinations continues to rise. Both the volume and diversity of data are obstacles to the timely completion of forensic investigations. Additionally, some law enforcement agencies do not have the resources to handle cases of even moderate size. To address these issues we have developed an architecture for a cloud-based distributed processing platform we have named Forensiccloud. This architecture is designed to reduce the time taken to process digital evidence by leveraging the power of a high performance computing platform and by adapting existing tools to operate within this environment. Forensiccloud's Software and Infrastructure as a Service service models allow investigators to use remote virtual environments for investigating digital evidence. These environments allow investigators the ability to use licensed and unlicensed tools that they may not have had access to before and allows some of these tools to be run on computing clusters.

Keywords: digital forensics, parallelization, cloud computing, cloud forensics, virtualization, virtual desktop infrastructure, HPC, cluster, infrastructure as a service, software as a service.

1 Introduction

Cyber-crime is a growing trend in the U.S. and around the world. More cyber-crimes are being committed every day, from e-bay fraud to cyber-extortion [27]. Additionally, the amount of data that must be processed in a digital forensic examination continues to rise at a very high rate. Part of this surge is due to the increased storage capacity of hard disk drives. A typical personal computer will often contain a 1 TiB drive with options for expanding to 2, 3, or even 4 TiB in some cases. Computers employed in commercial or government organizations can have even more drives. Add the fact that an investigation may encompass multiple systems, include the possibility of network, live response, and memory data; the processing time for a digital forensics examination rapidly becomes overwhelming. Considering that the rate of computer adoption is not slowing down, this trend appears to be continuing well into the future.

Over the years there has been a significant amount of research calling for increased processing power being applied to digital forensics [29] [31] [32] as well as an improvement of the current tools and techniques we are applying [19] [20] [26]. Research has shown that processing time can be decreased via both methods, increased processing power and more intelligent techniques, and they should be employed together for maximum effectiveness. This research indicates that without a way to take advantage of more processing power and improved tools, it is unlikely that the digital forensics community will be able to keep up with the demand for its services. Another aspect of digital forensics is that centralized processing laboratories at state and federal levels are not suited to taking on an ever-increasing number of local criminal cases containing digital evidence. The backlog at these centralized labs is already long, and increasing the workload will not help to decrease the backlog. Low-priority cases, such as those not involving child victims or significant losses, are probably never going to make it to the top of the queue, as new higher priority cases will get processed first.

This paper presents an architecture for a cloud-based digital forensics processing platform named Forensiccloud. It explores the issues, both technical and judicial, related to performing digital forensics in a remote environment. Finally, it presents a test plan for evaluating various components of a Forensiccloud implementation. The focus of this paper is using the cloud to perform digital forensics not performing digital forensics of the cloud.

1.1 Contributions

In this paper, we describe the architecture for a cloud-based digital forensics analysis platform. Additionally, we identify challenges, both technical and judicial, regarding the implementation of a cloud-based digital forensics analysis platform. We identify a set of tests that can be performed to evaluate various components of a Forensiccloud. The platform takes advantage of a distributed-computing environment to provide faster processing capability for performing digital forensics investigations remotely. We believe this will benefit small law enforcement organizations that could not otherwise afford to purchase their own comparable computing resources to perform in house investigations.

1.2 Motivations

This effort is motivated by over five years of experience building department sized digital forensics laboratories in rural Mississippi. During that time, the National Forensics Training Center (NFTC) at Mississippi State University was engaged in building small laboratories to provide digital forensics capabilities in strategic areas of Mississippi, where it was difficult for departments to either afford a lab or to engage the digital forensics lab in the Cyber-crime Fusion Center in Jackson, due to priorities of the central lab. In many cases, there were trained officers in these small departments, but without a lab of their own, they could not work their own cases. The efforts of the NFTC were to provide regional labs at larger departments where the smaller departments could time share on the equipment and use the larger department's lab to work their own cases. What was discovered is that when the larger department received the lab equipment, they used it a majority of the time, thereby making it difficult to provide the smaller departments with the time they needed in the lab.

Forensiccloud was envisioned to allow any small department to get an account on the cloud server, and with only a small client at the local station, to work on their own cases using a greater level of computing resources. We acknowledge that a significant challenge is the upload of the media containing the potential evidence to the central server, where processing can be accomplished centrally. With normal Internet access, the time required to upload an average sized hard drive image is daunting. We have yet to come up with a long term solution to this problem, but a reasonable stop-gap measure is to have the department deliver the media to the location of the server for upload, or to a facility in the state with a high speed connection to the Internet for upload.

A Forensiccloud environment will provide the law enforcement community substantial processing resources which could not be achieved by a stand-alone workstation environment. Even with the distribution of workstations throughout the state of Mississippi by the National Forensics Training Center, there is still, and will likely always be, an imbalance of utilization between different law enforcement agencies. The on-demand nature of a cloud environment is more robust to the varied processing needs of different law enforcement agencies. It will be able to accommodate and aid a range of departments from those that handle minimal digital forensics cases to those that see sudden surges in crimes requiring digital forensic processing. This versatility is a cornerstone of the motivation for developing Forensiccloud.

The processing flexibility of a Forensiccloud environment will range from being able to process a large number of cases at one time, therefore reducing backlog, to processing a low number of cases with greater processing power, therefore reducing the time of evidence processing. This distribution of resources can also be utilized for high profile cases, such as a child abduction, that must be processed quickly to provide investigators with time sensitive information that may be vital to the outcome of the situation. A Forensiccloud environment will also provide a collaborative capability between departments and examiners that would otherwise be difficult in a stand-alone digital forensic environment.

2 Related Works

2.1 Digital Forensics using Cloud Environments

In [21] the authors identify some issues associated with hosting a digital forensics server in a cloud environment. However, most of the paper is about doing digital forensics of the cloud and not using a cloud-based analysis platform to perform digital forensics. They do propose the following issues that need to be addressed when using the cloud to perform digital forensics:

- the evidence should not change when transmitted to and from the cloud and should not change while stored
- local laws should be observed when storing evidence in the cloud
- unauthorized access (either physically or digitally) to the evidence should be prohibited and no one should be able to change the evidence
- only users that have authorization to the evidence should be able to access it

These are all valid points concerning the transmission, authentication, and handling of evidence during an investigation, but the authors do not offer any possible solutions to these issues.

In [33] the authors discuss a model called MapReduce that processes data across many clusters. They explain that the current tools process linearly not in parallel. They implement MPI MapReduce which uses Message Passing Interface (MPI) and Phoenix (a shared memory version of MapReduce) to make the current implementations of MapReduce more efficient. They tested their MPI version of MapReduce and showed that it outperformed Hadoop for CPU, memory, and I/O tasks.

2.2 Increasing Processing Power and Distributed Architectures

In [32], Roussev and Richard note the limits of traditional, single workstation, digital forensics tools and define a set of system requirements for distributed forensics tools. They use the following factors to show that distributed forensics tools are now necessary:

- storage devices are growing in capacity
- the I/O speed of these devices is growing slower than the capacity increases
- digital forensic tools are becoming more and more complex

They propose the following requirements for distributed digital forensics toolkit:

- scalability
- platform-independence
- lightweight
- interactive
- extensible
- robust

The authors made a distributed architecture for digital forensics using these requirements that showed live search improvements that were 18 times faster than a traditional workstation.

In [31] the authors argue that current forensic tools are insufficient because users have not specified performance requirements and that the developers of tools fail to make performance a priority. They suggest that real-time forensics and triage may be a solution. Real-time forensics and triage places a time limit on the computation. They suggest the following goals of acquisition and processing to achieve best performance:

- they should complete at approximately the same time
- their results should be presented immediately upon completion

The authors explain that file system metadata, block forensics, and similarity digests are extremely fast and that file attributes and Windows registry can be processed in under one hour. The authors state that because data carving generates more data to process and has a high false positive rate, it may be time to reconsider why we use carving and if we can achieve its most important results (recovery of files) by other means. With testing they noticed that only a few forensic processes could actually be done in real-time on a traditional computer (8-core). Using a server (48-core) more processes can be completed in the specified time limit. They also note that SSDs may require more resources, as the read speed is much faster than the HDDs used in their testing. They estimate that all processing (at 120 MiB/s read speed) could be completed within the one-hour time limit with two to four 48 or 64 core servers.

In [29] the authors improve a previous tool known as *sdhash* by using parallelization. They also demonstrate that the imaging phase of investigation can also be a processing phase by running their tool as the evidence is mirrored. Using their new tool (*sdhash-dd*), data can be represented at 1.6% of the original size. They showed that they can lookup a small file (16KB) at 1.4TiB/s and they have almost perfect true/false positive rate.

Foremost was turned into a parallelized program by use of a parallel API in [26] to improve the speed at which it processed data. Foremost searches for known file headers and footers on a disk image; it does this sequentially in which it retrieves a chunk of data from the image, searches the chunk, then retrieves another chunk. The authors create an API enabling open source programs to be parallelized. The API uses a communication arbiter that allows the API accesses to the disk image; it has several features such as data safeguards, caching, and data read-ahead. The API also uses a channelized task scheduler that schedules several subtasks to do work. While the parallelized Foremost used more memory than the serial version, it does increase execution speed by 2.5 times on average.

2.3 Better Algorithms for Digital Forensics

Bulk Extractor [20] operates on disks images, files, and directories, and memory dumps and extracts various types of digital evidence including IP addresses, credit card numbers, or user-defined regular expressions. The tool supports parallel execution to improve processing time. It reads the input from start to end and passes the data to scanners that identify the data.

Compressed data is decompressed and sent back through the scanners. Bulk Extractor then creates report files that contain the locations of the identified files on the input. The authors provide a GUI that creates a histogram of the report files allowing quick analysis. Bulk Extractor performed 10 times faster than EnCase when extracting email addresses.

In [19] the author explains that the current age of digital forensics is coming to an end and we need to go in a new direction for digital forensics research. The old ways of performing digital forensics and the tools used need to be updated. The author explains several challenges of current digital forensics research and proposes a new direction for digital forensics research. This new direction requires new data abstractions, modularization and composability of tools, new framework supporting alternate processing models, and support for scaling and validation.

2.4 Scalable Digital Forensics Frameworks

The Sleuth Kit Hadoop [12] is a framework that uses The Sleuth Kit (TSK) on top of Apache Hadoop. It has three phases that it uses to analyze data: ingest, analysis, and reporting. Ingest retrieves information about the file system and the files on the image. The analysis phase uses various modules of TSK to analyze the data. Finally, the reporting phase generates reports on the analysis. TSK Hadoop uses Apache Hadoop [17] to distribute the process of analysis across several nodes. Hadoop has an intergraded distributed file system, a job scheduler, and a Java implementation of MapReduce for parallel processing [17]. Using Hadoop and TSK together benefit from increased processing power from parallelization. However, using TSK Hadoop limits the number of tools to those supported by TSK. One of the goals of Forensiccloud is to enable a variety of tools and techniques to function in the environment.

2.5 Virtualization Architectures

Several virtualization solutions were considered when designing Forensiccloud. Most solutions did not meet all of the requirements needed by Forensiccloud. Xen [16], KVM [7], and OpenVZ [11] have no native support for Virtual Desktop Infrastructure. OpenStack [10] and VMware ESXi [14] are the top two choices we have looked at that meet all the requirements natively. Microsoft Hyper-V [8] and Citrix XenServer [3] were also considered, however they were not as user-friendly, their installations were difficult, or their management software did not have as many features as VMware ESXi and OpenStack.

3 Requirements

There are several challenges that must be addressed for an effective implementation of Forensiccloud. These challenges involve both technical capability and strict digital forensics processes. Each of the challenges listed below can be overcome while upholding the integrity of digital evidence and providing the user with a high level of digital evidence processing capability.

The Scientific Working Group on Digital Evidence Model Quality Assurance Manual for Digital Evidence Laboratories specifies rules that a digital forensics laboratory must follow when computers or automated equipment are used for the acquisition, processing, recording, reporting, storage or retrieval of examination data [25]. The rules are:

1. "Digital forensic tools are documented in sufficient detail and are suitably validated"
2. "The integrity and confidentiality of data entry, data storage, data transmission, and data processing is protected"
3. "Computers and automated equipment are maintained to ensure proper functioning and are provided with environmental and operating conditions necessary to maintain the integrity of examination data"
4. "Unauthorized access is prevented for computer systems used for examining digital evidence"

Each of these is a challenge that all digital forensic laboratories face, and Forensiccloud is no different. Forensiccloud satisfies these rules in the following ways:

1. Forensiccloud will provide all documentation that each tool has available however tool validation will not be done as it is up to the investigator to determine the validity of a tool.
2. As described in the security measures below, the integrity of the data and prevention of unauthorized access will be possible.
3. All hardware and software utilized by a Forensiccloud will be maintained and kept at a high level of operation.
4. Only users whom have access to Forensiccloud will be able to access it and users will not be able to access data that is not apart of their case.

3.1 Client Security

One of the primary focuses of Forensiccloud is the widespread availability of significant computing resources. This capability will allow officers to utilize hardware and software that would otherwise be difficult or too expensive for

them to acquire and maintain. To provide this functionality it is imperative that officers be able to access Forensiccloud to not only set up their desired evidence processing, but to also analyze the output from that processing. Access Forensiccloud will be done over the Internet and must therefore be secured. To help ensure the security of the clients that are accessing Forensiccloud will use traditional user access, such as confirmed users who will have unique login credentials. The client security will be taken a step further by requiring a client machine be approved and validated to access Forensiccloud. These client machines will be limited to stand-alone machines that are in secure areas of a law enforcement or government office. These client machines must also be under the control of a trusted user that has been approved to use Forensiccloud system and perform digital forensic investigations. Evidence and case data will only be accessible by the machine(s) and user(s) that are authorized to examine that specific case.

3.2 Data Security

As with any digital evidence investigation, the security of the evidence throughout the entire digital forensic process is paramount. The idea of Forensiccloud naturally includes the transmission and remote storage of sensitive information. Maintaining the confidentiality and integrity of the data while ensuring it is available to forensics examiners is crucial. In order to protect the evidence being transmitted across the Internet data encryption will be used. The data will not be encrypted is when it is being processed, when the investigator is viewing it, when it is stored at Forensiccloud and the local image on the investigator's computer. Since the processing engine is not open to the Internet and the evidence is isolated from other users, the data will be secure.

3.3 Network Latency

Unlike traditional digital forensic practices where the working copy of the evidence can be loaded on a stand-alone workstation in the examiners lab, Forensiccloud will need to have the capability to receive digital evidence from remote locations. Currently, the average upload speed in the U.S. is 7.7 Mbps [6]. At this rate, it would take approximately 25 days to transmit 2 TiB of data across the Internet. With the advent of Internet2 speeds are significantly increased, which will allow for much faster upload capability. With the speed of Internet2 it will be possible for examiners to upload the working copy of their evidence to Forensiccloud in a practical amount of time. Currently, there are 7 Internet2 participants distributed throughout the state of

Table 1 Theoretical transfer of files

Size of evidence	100 Mbit/s	10 Gbit/s
100 GiB	02:57:29	00:01:44
500 GiB	14:47:28	00:08:40
1 TiB	30:17:31	00:17:44
5 TiB	151:27:39	01:28:44

Mississippi [9]. Each of these locations has a 10 Gbps connection. Forensiccloud will utilize these locations by having upload stations at each that have a 10 Gbps connection to the Forensiccloud storage. Table 1 shows the theoretical time it would take to transfer different sizes of evidence with a 30% overhead on the network. These times do not represent reading from a disk and transmitting it across the Internet, they assume that there is no disk bottleneck. The actual transmission speed will be limited by the disk speed when using Internet2. Uploading the evidence is a one-time cost because the evidence will only be transmitted once and will be stored during the entire investigation of the case. The investigator will only download the reports generated by the tools used and will not download all the evidence.

The drawback of Internet2 is that fact that it is not available in every location and it may be unreasonable for some examiners to travel to an Internet2 location to upload their evidence. For a scenario where an examiner is unable to travel to an Internet2 or Forensiccloud upload site location they will need to ship the working copy of their evidence to either an Internet2 location or directly to Forensiccloud. This solution will work as long as the shipping provider will uphold chain of custody for the evidence.

3.4 Data Authentication

An important part of any digital forensics examination is the authenticity of the digital evidence and case files. This is no different in the Forensiccloud environment. Cryptographic hashes, such as MD5, SHA1 and SHA256 will be used to authenticate the data. This will be done when the evidence files are sent to Forensiccloud to ensure the data that is submitted is the same as the data that was originally seized by the local investigators. These hashes will also be used to ensure the data is not changed during the processing and storage of the evidence.

3.5 Data Storage

The purpose of the cloud is to leverage the greater resources of a high performance computing system while providing a plethora of tools to examiners. Based on the practices outlined in the Scientific Working Group on Digital

Evidence Model Quality Assurance Manual for Digital Evidence Laboratories Forensiccloud will be a short-term data storage area and considered to be a working area for the digital investigation. Because of this, the evidence and case data that is stored on the cloud will only be stored while the case is considered to be an active investigation. Evidence and case files will not be retained longer than 90, unless special permission is given or further analysis is needed [25]. This permission will be given for special circumstances only and the approved user of the case must submit a request.

3.6 User Interface

The method in which the examiner interacts with Forensiccloud is critical. It is imperative that the interface for the Forensiccloud be both versatile and user-friendly. There are essentially two points of interaction between the investigator and a Forensiccloud system. There will be an upload-and-request interface and a processing-and-review interface.

The upload-and-request interface will be a client that will run on the examiner's workstation and the upload facility workstation. This client will give the examiner the ability to create a new case, upload evidence files, and request specific processing items that he or she would like to have done to each evidence item.

The processing-and-review interface will give the examiner the ability to review the evidence as well as any results from processing. They will also have the capability to perform any additional processing that may be beneficial for their case. This interface will utilize a virtual machine to give the examiner the look and feel of a traditional digital forensic workstation. As a part of this look and feel, AccessData's FTK [5] and Guidance Software's Encase [4] will be available to the examiner. The availability of these two tools is important because they are two of the most widely used commercial tools for digital forensic examination [23]. Including both FTK and Encase is an advantage because it brings two of the most well known digital forensic tools to users that may not have access to those tools, either for economic reasons or technical capability. In addition to the tools that will be built into Forensiccloud for processing the processing-and-review interface will allow the examiner to install their own tools into the virtual machine. This provides the examiner with the versatility that they would have with a stand-alone digital forensics workstation while still leveraging the greater computing power of a Forensiccloud system.

Due to the fact that the majority of examiners use a graphical user interface (GUI) instead of command line tools [23], it is important that the upload-and-request interface and processing-and-review interface primarily utilize a GUI but have the flexibility and control for the examiner to use command line tools when needed. Both interfaces will be GUIs that provide the user with the options they will need for the tasks that are built into Forensiccloud service. The virtual machine environment will allow the user to use tools that they are familiar with and interface, command line or GUI, that they are comfortable with.

Another benefit of Forensiccloud is the ability for streamlined collaboration. The cases are stored in a centralized location, other examiners or investigators will have the ability to view or process the same case, as long as each are approved by the case manager. This allows for new examiners to get assistance from an experienced examiner without the need for one to travel to the other. This will also help with collaboration between departments when evidence items may be involved in multiple crimes.

3.7 Chain-of-Custody

As with any investigation, tracking digital evidence is crucial. The procedures for the use of Forensiccloud will require that any evidence files that are uploaded must have chain of custody documentation. The same is true if an examiner or investigator makes a request to obtain the evidence files from the Forensiccloud storage. In addition to chain-of-custody documentation, the system will maintain logs of all evidence and user events, such as logins, tool processing, and evidence upload and downloaded. These logs will automatically be generated and will contain a timestamp, a user id, and a report of the action.

4 Forensiccloud Architecture

Cloud computing, as defined by NIST, is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models [24]. Essential Characteristics:

- On-demand self-service
- Broad network access

- Resource pooling
- Rapid elasticity
- Measured service

Service Models:

- Software as a Service (SaaS)
- Platform as a Service (PaaS)
- Infrastructure as a Service (IaaS)

Deployment Models:

- Private cloud
- Community cloud
- Public cloud
- Hybrid cloud

The architecture of Forensiccloud will fulfill the five essential characteristics. Forensiccloud will

- be accessible by investigators to analyze evidence as needed (on-demand self-service)
- be available to an entire state (broad network access)
- use virtualization to pool several servers resources (resource pooling)
- be able to expand or shrink based on the needs of its users (rapid elasticity)
- schedule processing based on priority and consumption of the investigator (measured service)

By providing software and an environment to use the provided software, Forensiccloud uses both SaaS and IaaS models. The deployment model used will be a community cloud. In this instance, the community is the collection of law enforcement agencies within a given state. Digital forensics evidence by nature is sensitive information. As such, public clouds like Amazon's EC2 should not be used for either the investigation or processing components.

4.1 Architecture Overview

Forensiccloud will enable an investigator to upload, process, and analyze evidence. Investigators will have access to a client that will allow them to connect to Forensiccloud and upload digital evidence. The client will first authenticate the investigator then it will create a secure connection to Forensiccloud. Using this connection, an investigator will set up a job by selecting the digital forensic tools that will be run. The investigator will then upload a disk image to Forensiccloud using the client. While the image is being

uploaded some of the selected forensic tools will run; these tools do not require the full disk image. Tools that require the entire image will be executed when the image has finish uploading. The tools will run on a cluster and the results will be saved to the job directory.

During the processing and analysis of the evidence, the investigator will authenticate and connect to a virtual machine remotely. The connection to the virtual machine will be secured. The investigator will only be able to connect to the virtual machine that has the evidence that they uploaded. As tools finish on the cluster their results become available in the virtual machine. In this virtual machine they have access to tools that cannot or do not benefit from being run on a cluster. The investigator can use the output of these tools as well as other tools on the virtual machine to analyze and investigate their case. Virtual machines are isolated from each other and the virtual machine is not able to access evidence that is owned by another investigator.

The analysis virtual machine and the cluster have access to a data store that has the evidence uploaded by the client. The evidence is encrypted from the client, decrypted on the cluster, and the results of the processing will be encrypted back to the data store. The analysis virtual machine will then decrypt the data for the investigator.

There are three main parts to Forensiccloud: the client, the investigation component, and the processing component. The client will do initial job setup and will upload the evidence to Forensiccloud. The investigation component provides the user with a virtual machine to use to analyze the evidence. The processing component will run forensics tools using a cluster.

4.2 Forensiccloud Client

The client used to upload and setup jobs will be installed on the investigators computer. Once installed it will connect to Forensiccloud and validate the system and itself. Validation requires the law enforcement agency to register their system(s) with the providers of Forensiccloud. The client will also validate itself to ensure it is updated and has not been altered. Only departments that have registered will be capable of accessing Forensiccloud. The investigator must also authenticate with Forensiccloud. If the client is valid and the investigator is authenticated a secure connection will be made and the upload and job setup can continue. The client enables the investigator to upload and download files from Forensiccloud. The investigator can upload tools that will be available to them and download analysis reports. The client will mirror the evidence not only to Forensiccloud, but also to an image on the investigators

computer. This eliminates the need of a second mirroring step, which speeds up the overall investigation time.

4.3 Investigation Component

Virtual machines provide investigators an environment for investigation of digital evidence. Investigators will use the same credentials used with the Forensiccloud client to authenticate with this virtual machine. When an investigator creates a job, a fresh virtual machine is cloned from a base virtual machine. This virtual machine will have tools installed with licenses for the investigator to use. If the investigator needs a tool that is not already installed they can simply install it themselves. When an investigator is done analyzing evidence or when the job expires (jobs expire in 90 days by default), the virtual machine is deleted and all data corresponding to the job is deleted.

A virtualization manager is needed to control the creation and deletion of these virtual machines. Commercial software such as VMware ESXi and VMware vCenter [14] or open source software such as OpenStack [10] will fulfill all the needs of the investigation component. It will allow creation, cloning, modification, and deletion of virtual machines. It also handles virtual machine isolation [2]. In order for investigators to access virtual machines on-demand and remotely virtual desktop infrastructure (VDI) is needed. VDI provides on-demand access to a virtual machine that has either been provisioned upfront or as the investigator connects. It provides both input (keyboard, mouse, etc.) and output (monitor, sound, etc.). The particular VDI solution used is VMware Horizon View. VMware View provides the input and output as well as USB redirection, which allows a local USB device, such as a flash drive, to be connected to the remote virtual machine [1]. VMware Horizon View provides VDI by use of a standalone client that can authenticate with Forensiccloud over a secure connection.

4.4 Processing Component

Forensic tools will be run using computing clusters. There are two options we are currently considering. The first is to use a virtualization cluster that has a similar architecture to that used in the investigation component; the second is to run the tools on an HPC cluster. Either solution has the ability to scale up or down based on the needs of the system and both will have measured service.

Using the virtualization cluster several virtual machines will handle the workload for the tools. The virtual machines will have an agent running that will accept payloads and process them with forensic tools. A standard blade

server will contain one or more nodes (virtual machines). This option brings both advantages and disadvantages. Advantages are:

- virtual machines can run any operating system required by the particular tool being used
- they have dedicated resources
- the particular tool should not have to be modified to run on them
- a virtualization cluster can also be as big or as small as it needs to be since nodes of the cluster can be added and removed with little difficulty

There are also several disadvantages to using a virtualization cluster:

- virtualization overhead
- the need for a custom scheduler to schedule incoming jobs
- no default message passing interface (MPI)
- the need for a custom agent on each node to accept incoming jobs

Using an HPC cluster, a job is submitted to a scheduler node on the cluster and the other nodes of the cluster handle the work load. The cluster will have a scheduler for the jobs built in. Some of the benefits of using a cluster are:

- little overhead
- no custom scheduler
- MPI support

There are also a few disadvantages to using a HPC cluster:

- uses Linux for all the nodes; this means that all the tools running on the cluster must support Linux
- forensic tools may need to be modified to support the particular MPI used by the cluster
- nodes are more expensive than virtualization cluster nodes. However, each node of an HPC cluster will generally have superior hardware than the nodes of a virtualization cluster
- nodes are more specialized to handle computing jobs rather than general purpose jobs

Either solution can be used depending on the resources available. Figure 1 shows how each part of Forensiccloud is connected to each other.

4.5 Investigation on Forensiccloud

The process of investigation on Forensiccloud follows three phases.

- Phase 1 – The investigator will specify the tools that will be ran on the evidence as well as other information about the case used for reporting

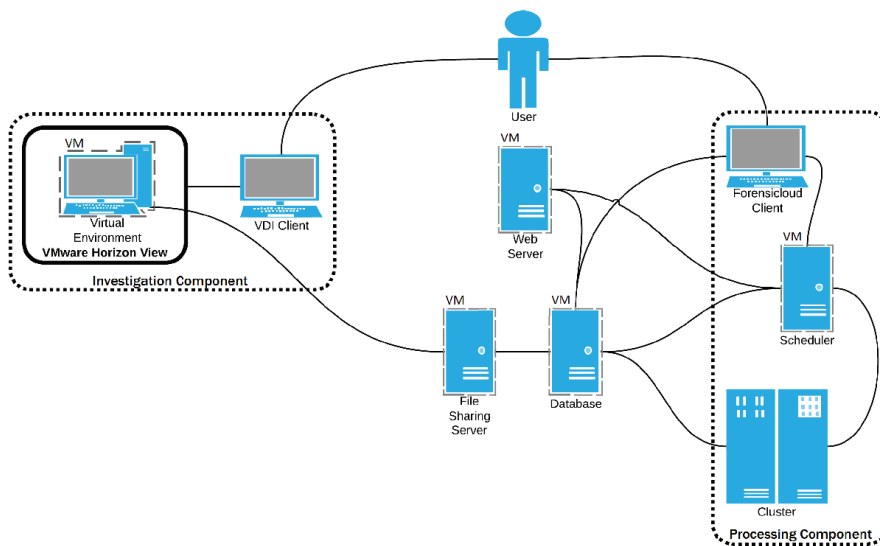


Figure 1 The components of Forensiccloud

(case type, investigators, etc.). The investigator will upload any extratools they need to be available on the virtual machine used in phase 3. The investigator will then upload their evidence to Forensiccloud using the Forensiccloud client. Figures 2 and 3 illustrate the job setup and image uploading.

- Phase 2 – While the evidence is uploading specific tools that the investigator selected are run against the uploaded data. Only bitstream or single file capable tools will be run as the evidence is being uploaded. When the evidence has finished uploading the rest of the processing can be done on the evidence.
- Phase 3 – Using the investigation component, the investigator will be provided a virtual machine to examine the results returned from the processing component. They may also further analyze the evidence using

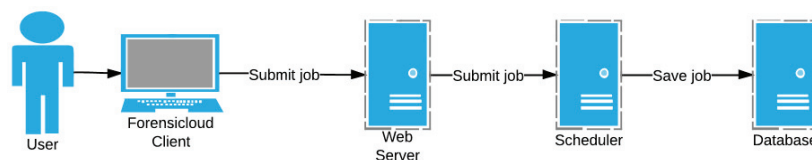


Figure 2 Phase 1 - Job setup

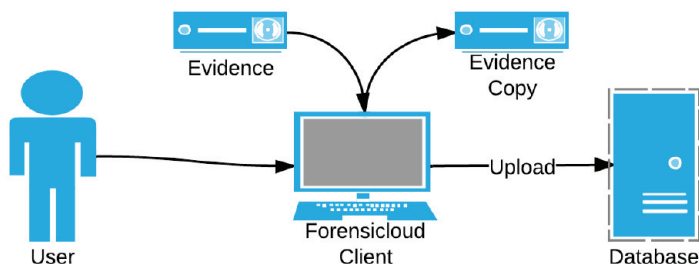


Figure 3 Phase 3 - Analysis retrieval

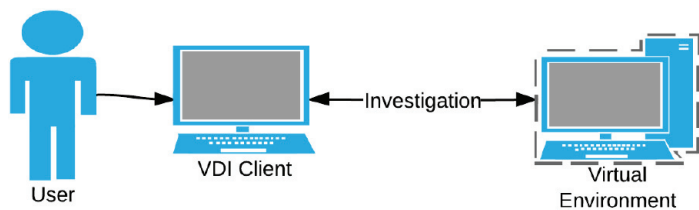


Figure 4 Phase 3 - Investigation

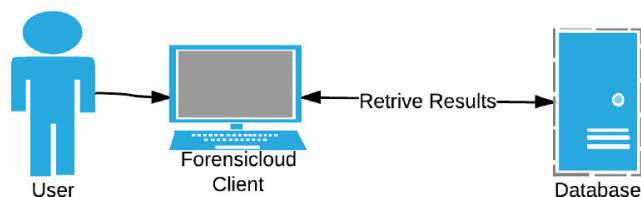


Figure 5 Phase 3 - Analysis retrieval

tools provided in the virtual machine as well as any tools they have uploaded. Figure 4 shows the investigation step. As the investigator performs analysis on the evidence they will be able to retrieve it for use in court. The retrieval of this data will be done using the Forensicloud client. Figure 5 displays the analysis retrieval part of phase 3.

5 Experiments

This section presents an experimental plan for testing various components of the Forensicloud architecture. For each phase of the process there are tests that should be conducted to determine the feasibility of the implementation. Some

tests are also designed to establish benchmarks for the underlying components. Many of the files and images used in the following tests are from the digital corpora hosted by NPS [18].

Table 2 lists the systems that we are planning to use to test our implementation of Forensiccloud. The HPC Cluster is named Talon and is a 34.4 TeraFLOPS IBM iDataPlex cluster with 256 nodes and is hosted by the High Performance Computing Collaboratory of Mississippi State University. The Virtualization Cluster nodes are running virtualization software so one node can act as several nodes. In the following tests each physical node has two virtual machines that are acting as nodes. This allows the virtualization cluster to have four nodes total with 12 threads and 24 GiB of RAM each. Table 2 lists the hardware specifications for one physical node.

5.1 Phase 1 Testing

Phase 1 encompasses setting up the initial processing tasks that should be performed on the evidence, the possible imaging of the evidence, and the encryption of the evidence for transmission across the Internet. All tests for this phase should be performed on a workstation as defined in Table 2 as all these tests relate to the Forensiccloud client.

5.1.1 Simultaneous imaging and upload

Current research [31] [29] indicates that processing of data should start at the same time that the imaging process begins. Imaging and uploading, at least on Internet2, are I/O bound tasks when the target drive is a SATA HDD. Roussev et. al [31] report that a commodity SATA HDD can be read sequentially at 120 MB/s. This would take 2.5 hours to read a 1 TB drive. To facilitate faster processing we believe the client should support transmitting the data and creating the image in parallel.

Test – Determine the effect on performance of simultaneous imaging and network transmission of the evidence. The client should be run twice: once just to upload the test data to the server; once to upload the test data and

Table 2 Systems used in testing

Name	RAM	CPU(s)	Cores	Threads	Disk Speed
HPC Cluster	24 GiB	2 Intel Xeon X5660 @ 2.8 Ghz	12	24	100Mib/s
Virtualization Cluster	128 GiB	2 E5-2670 @ 2.6 GHz	16	32	1Gib/s
Workstation	8 GiB	Intel Core i7-860 @ 2.8 GHz	4	8	100Mib/s

make an image of it in parallel. This will be used to determine if the client can support both operations simultaneously while maintaining the optimal 120 MB/s upload rate.

5.1.2 Encryption

To ensure the confidentiality and integrity of the data during transmission, encryption must be applied to the evidence being transmitted between the client and server. The Advanced Encryption Standard (AES) is supported by NIST and is a well-known, reliable, symmetric key encryption protocol [13]. We consider this sufficient for the purpose of securing data during network transmission.

Test – Determine if AES encryption has any effect on the overall upload rate of the evidence being transmitted across the Internet. The client should be run twice: once just to upload test data to the server; once to encrypt the test data prior to transmission. If the upload rate is significantly affected by the encryption of the data, other reliable encryption standards may be explored.

5.2 Phase 2 testing

Phase 2 of Forensicloud is the processing of the digital evidence by the backend cluster. The goal of this phase is to efficiently process incoming data in order to provide the investigator with analyzable results as soon as possible. Testing in this phase relates to the number of nodes, or cores, necessary to carry various tasks within the constraints of the system (network transmission time, I/O speed, etc.).

5.2.1 Real-time stream processing

Streamlining the analysis process requires processing the data as it becomes available to the server. In [20] Garfinkel mentions two approaches to processing digital evidence: *bulk data analysis* and *file-based approaches*. *Bulk data analysis* processes blocks of data without regard to the underlying structure. That is, it does not require filesystem structural information in order to carry out its task. These types of techniques are ideal for processing the stream of data being transmitted from the Forensicloud client. Given that we expect the process to be I/O bound from the client and that SATA HDDs are still more common than solid state drives (SDDs), the transfer rate will likely be limited to 120 MB/s. This requires 120 – 200 cores to maintain processing at this speed [31]. We are currently aware of two tools designed for processing a stream of data: bulk extractor [34] and sddhash [30].

Test – Determine the required number of cores for a Forensiccloud implementation to run bulk extractor and sdhash when data is being streamed from the client. Received data needs to be chunked and passed to each tool with nodes being allocated as needed.

File-based approaches process at the file level and many tools use this method. By extracting files from a stream in realtime, it is possible for the Forensiccloud cluster to perform file level analysis. The latency-optimized target acquisition prototype mentioned in [31] shows it is possible to reassemble files from a stream in realtime and make them available to clients. In this case, the clients are the nodes. Once the files are available, they can be passed to appropriate tools for file-centric processing. While many tools operate on files, not all of them make sense to run during the initial processing. Password cracking tools for instance will only need to be run against a small subset of files on a given system and then with very specific configurations. Exif data however is something many file types possess that can be extracted in an automated manner.

Test – Determine required number of cores for ExifTool [22] to process files extracted from an uploaded data stream in realtime.

5.3 Phase 3 Testing

During phase 3 the investigator will connect to a remote virtual machine. They will use this virtual machine to perform analysis on the evidence and the reports generated by the various tools run on the cluster. As reports are made the investigator can use the Forensiccloud client to retrieve them. The investigator can run certain tools on the cluster and others will be run on the virtual machine itself. Using the client they can also upload additional tools to use within the virtual machine.

5.3.1 Remote desktop connection

The investigator connects to a remote virtual machine by use of a remote desktop protocol that is built into VDI. If VMware Horizon View is used for VDI there are two protocol options: PCoIP and Microsoft's RDP. PCoIP has more features, such as USB redirection and multiple-monitor support.

Test – This test will determine the usability of the remote desktop protocol used. To determine this connect to Forensiccloud using VDI from an average internet connection. Determine if the virtual machine is still usable from these locations by performing typical investigation activities, such as viewing

reports and using forensic tools. The virtual machine should be responsive and not have any noticeable video delay.

5.4 Performance Tests

To determine the optimal processing option for processing digital evidence in ForensicCloud, it is important to first get a baseline of how fast a forensics workstation can process digital evidence. For each tool described below, the tool will use data from the digital corpora hosted by NPS [18] unless otherwise specified. For each tool, the amount of time that the tool takes to complete the processing will be used as a benchmark to determine the overall performance improvement of the distributed processing options.

After the benchmark for each tool has been set, each tool will be executed using the virtualization cluster and HPC cluster to determine the speeds that can be achieved when each tool is utilizing parallelization. Based on the speeds of each tool in the different environments, it can be determined which distributed processing solution is most useful.

Test – Bulk Extractor – the goal of this test is to compare HPC cluster versus virtualization cluster performance of Bulk Extractor [34]. Using a disk image of 1 TiB or greater, break the image into overlapping fragments. The number of fragments will be equal to the number of nodes and they will be of equal size. Write the output data back to the file server. Record and analyze processing time for each environment.

Test – sdhash – the goal of this test is to compare HPC cluster versus virtualization cluster performance of sdhash [30]. Using a disk image of 1 TiB or greater, break the image into overlapping fragments. The number of fragments will be equal to the number of nodes and they will be of equal size. Write the output data back to the file server. Record and analyze processing time for each environment.

Test – Password Cracking with John the Ripper – the goal of this test is to compare HPC cluster versus virtualization cluster performance of John the Ripper [28]. Construct and store Windows NTLM hashes corresponding to password lengths of 4, 8, 12, and 16. Run John the Ripper will use MPI on the HPC cluster, using its ‘node’ option on the virtualization server, and using only threading on the forensics workstation. Record and analyze processing time for each environment.

Test – Extract metadata from files with ExifTool – the goal of this test is to compare HPC cluster versus virtualization cluster performance of ExifTool [22]. Store the Govdocs1 corpus files on the file server. For each test run

multiple ExifTool instances at the same time. The number of instances equals the number of threads per node multiplied by the number of nodes used. Record and analyze processing time for each environment.

Test – Extract artifacts from Windows memory dump using Volatility – the goal of this test is to compare HPC cluster versus virtualization cluster performance of Volatility [15]. Create memory images of size 4 GiB, 8 GiB, and 16 GiB. Run a single Volatility command on each node. Record and analyze processing time for each environment.

5.5 Other Testing

This section contains tests that should be performed that do not directly test Forensiccloud. However, these tests should be performed to ensure Forensiccloud will not be limited by external complications.

5.5.1 Nodenize

Nodenize is a tool made for Forensiccloud that allows forensic tools that only work with single input files work in parallel. Nodenize takes as input a directory containing files to be processed. It also takes a tool that will be used to process files in the directory as input. Nodenize will be run on each node used for the task; each of these nodes will know what set of the input files it will process by using a node identifier. If the identifier is 1 out of 4 the node will process the first 25% of the data, if it is 2 out of 4 will process the second 25% of the data, etc. It then gets a list of all the files in the directory and determines which section of them it will process. Finally, it processes each file using the tool selected.

Test – The goal of this test is to determine if nodenize has any negative influence on the performance of forensic tools ran with it. To test this run a tool with nodenize and run the same tool without nodenize. Running the tool without nodenize will require batch execution of the tool on the different input. Determine if nodenize had any noticeable influence on performance.

5.5.2 Workload

There are many areas of in a Forensiccloud workflow that will need to be tested to establish the most efficient and effective use of that specific Forensiccloud implementation. These tests layout procedures that should be used to determine the capability and processing load that a specific Forensiccloud implementation can handle.

Test – Single Upload and Download Capability – the goal of this test is to determine upload capability to a Forensiccloud environment. Using varying data set sizes, upload and download each data set individually from different Forensiccloud facilities. Record and document how much time it took to upload each file. Determine the amount of time that would be prohibitive for a single user.

Test – Multiple Upload and Download Capability – the goal of this test is to determine multiple upload capability to a Forensiccloud environment. Using a set data set size of 500GB or greater, upload and download a data set from 2 Forensiccloud facilities at the same time. Record how much time it took to upload each file. Increment the upload locations by 1 and upload the files again. Record how much time it took to upload each file. Continue adding one facility until the time for upload would be prohibitive to any user.

Test – Nodes per Task – the goal of this test is to determine the number of nodes that is optimal for each task. Using a data set of 1TiB or greater run each tool available to the user. With every run of the tool add 1 for the tool to use. Continue this until there is either no more nodes remaining or there is no longer an increase in level of performance by adding another node. Document the optimal number of nodes for each task.

Test – Nodes per case – the goal of this test is to determine the number of nodes for each user of Forensiccloud. Starting with a single case, allocate all nodes to all cases equally. Increment the number of cases by one until the processing performance is no longer optimal. Record the number of nodes a single case needs to be effective.

6 Preliminary Test Results

We were unable to perform a full evaluation of the system due to various technical difficulties. However, one test was completed that tests various aspects of the particular cluster being used. Bulk Extractor was run on a forensic workstation, virtualization cluster, and HPC cluster. Table 3 displays the evidence used in the tests. The first test used only one node of each cluster; the second test used bulk extractor's parallelization option '-Y' to separate the evidence into four chunks. Each chunk was processed by one node of each cluster. Bulk Extractor was run with all default scanners active. Table 4 displays the results of this test with only one node per cluster. The time and processing rate reported for node of each cluster is shown in Tables 5–7. These tables also display whether Bulk Extractor thought the process was CPU or I/O bound.

Table 3 Details of the evidence used in testing

	ubnist1.gen3.E01	nps-2009-domexusers.E01
Compressed	854 MiB	4.07 GiB
Uncompressed	1.96 GiB	40 GiB

Table 4 Results of one node per cluster

	ubnist1.gen3.E01			nps-2009-domexusers.E01		
	HPC	Virtualization	Workstation	HPC	Virtualization	Workstation
Seconds	112.3	44.8	101.0	892.1	566.4	894.2
MiB/s	18.75	47.0	20.9	48.2	75.8	48.0
MiB	2106	2106	2106	42949	42949	42949
Bound	CPU	CPU	CPU	CPU	None	CPU
Threads	12	12	8	12	12	8
RAM	24	24	8	24	24	8

Table 5 The workstation cluster results. Each workstation had 8 GiB RAM and 8 threads

	ubnist1.gen3.E01				nps-2009-domexusers.E01			
	Node 1	Node 2	Node 3	Node 4	Node 1	Node 2	Node 3	Node 4
Seconds	35.9	44.9	26.0	5.1	462.3	203.5	196.0	47.9
MiB/s	14.5	11.6	20.0	107.4	23.2	52.7	54.8	224.1
MiB	520	520	520	546	10737	10737	10737	10737
Bound	CPU	CPU	CPU	None	CPU	CPU	CPU	I/O

Table 6 The HPC cluster results. Each node had 24 GiB RAM and 12 threads

	ubnist1.gen3.E01				nps-2009-domexusers.E01			
	Node 1	Node 2	Node 3	Node 4	Node 1	Node 2	Node 3	Node 4
Seconds	45.6	64.1	31.3	10.9	480.6	260.2	165.1	53.4
MiB/s	11.4	8.1	16.6	50.1	22.3	41.3	65.0	201.0
MiB	520	520	520	546	10737	10737	10737	10737
Bound	CPU	CPU	CPU	None	CPU	CPU	CPU	I/O

Table 7 The virtualization cluster results. Each node had 24 GiB RAM and 12 threads

	ubnist1.gen3.E01				nps-2009-domexusers.E01			
	Node 1	Node 2	Node 3	Node 4	Node 1	Node 2	Node 3	Node 4
Seconds	20.2	26.1	14.3	6.1	240.3	171.3	163.7	47.5
MiB/s	25.8	19.9	36.3	90.3	44.7	62.7	65.6	225.9
MiB	520	520	520	546	10737	10737	10737	10737
Bound	None	None	None	None	CPU	None	CPU	I/O

The HPC cluster performed similarly to the workstation. We believe that the primary issue with our HPC cluster is that it currently lacks a fast storage device. Due to technical difficulties we are unable to provide it with the same or

similar storage device that is used by the virtualization cluster. If faster storage was used it would improve the processing speed. The virtualization cluster outperformed both the workstation and the HPC cluster by up to 225%. Further testing needs to be conducted with a shared storage device. Even though the HPC cluster and workstation cluster performed similarly, we believe the HPC cluster is still more economical and feasible than a workstation cluster. It would take 384 workstations to reach the 3072 threads in the HPC cluster we used. Thread for thread, it would take 96 virtualization nodes to equal our HPC cluster's threads.

7 Future Work

There is still significant work to be done to fully implement a working prototype of Forensiccloud. One goal of this paper is to enumerate the challenges Forensiccloud faces in order to start a conversation within the community. This will provide us useful feedback that we can incorporate into the prototype we are developing.

Only a small subset of tools were discussed in this paper, we will complete a comprehensive review of digital forensics tools to determine which tools can be parallelized appropriately for use in Forensiccloud. This includes identifying those tools that operate on discrete elements such as files and those that are designed to be executed in parallel.

An aspect of Forensiccloud we have not mentioned in this paper is its use as Platform as a Service (PaaS). If tool developers have access to a Forensiccloud they can use it to build and test parallelized tools for digital forensics. We will implement an API that will allow investigators to easily interface with Forensiccloud. With this API a tool can be made that utilizes parallel processing without the need to write parallel code.

8 Conclusion

The quantity and diversity of digital evidence continues to increase. What was once a matter of investigating thousands of files has turned into investigating millions of files of various types. An investigator can no longer manually search evidence; smarter tools are required that can automate as much of the process as possible. These tools work on traditional forensics workstations; however, they can take hours or even days to finish on larger evidence. Forensiccloud decreases the time needed to process data by leveraging the

power of a high performance computing platform and by adapting existing tools to operate within this environment. Forensiccloud further improves the investigation by providing an environment for remote investigations that gives investigators access to licensed tools, such as Access Data's FTK and Guidance Software's Encase that they may not have had before.

In this paper we have presented an architecture for a cloud-based digital forensics analysis platform. A Forensiccloud system provides several benefits:

- it reduces the overall processing time of large quantities of data by leveraging the power of a high performance computing platform and adapting existing tools to operate within this environment
- it provides smaller departments that may not have access to certain commercial software the ability to use this software remotely
- it enables collaboration. With the evidence stored in the cloud, it is possible to allow someone to access the case, provided sufficient authorization from the case owner, to provide additional analysis. An attorney could also have access to the case on the cloud to view the investigators analysis.

We have detailed several challenges we believe an implementation of Forensiccloud must overcome in order to gain acceptance from both the judicial and technical communities. We have presented guidelines that address these challenges based on existing standards, where applicable.

Finally, we have presented a test plan for evaluating various components of a Forensiccloud implementation. Using these tests one can determine the feasibility of the architecture for the particular implementation of Forensiccloud.

References

- [1] Vdi: A new desktop strategy. Technical report, VMware Inc., Palo Alto, CA, 2006.
- [2] vsphere security esxi 5.1. Technical report, VMware Inc., Palo Alto, CA, 2012.
- [3] Citrix xenserver. <http://www.citrix.com/products/xenserver/overview.html>, 2014. Accessed: 2014-07-20.
- [4] Encase forensic. <https://www.guidancesoftware.com/products/Pages/encase-forensic/overview.aspx>, 2014. Accessed: 2014-07-20.

- [5] Forensic toolkit. <http://www.accessdata.com/solutions/digitalforensics/ftk>, 2014. Accessed: 2014-07-20.
- [6] Household upload index - united states. <http://www.netindex.com/upload2,1/United-States>, 2014. Accessed: 2014-07-21.
- [7] Kvm. http://www.linux-kvm.org/page/Main_Page, 2014. Accessed: 2014-07-20.
- [8] Microsoft hyper-v. <http://www.microsoft.com/en-us/server-cloud/solutions/virtualization.aspx>, 2014. Accessed: 2014-07-20.
- [9] Mississippi optical network. <http://mission.mississippi.edu>, 2014. Accessed: 2014-07-25.
- [10] Openstack. <http://www.openstack.org>, 2014. Accessed: 2014-07-20.
- [11] Openvz. http://openvz.org/Main_Page, 2014. Accessed: 2014-07-20.
- [12] Sleuth kit hadoop. http://www.sleuthkit.org/tsk_hadoop/, 2014. Accessed: 2014-07-20.
- [13] Standards and guidelines tested under the cavp. <http://csrc.nist.gov/groups/STM/cavp/standards.html>, 2014. Accessed: 2014-07-27.
- [14] Vmware esxi. <http://www.vmware.com/products/vsphere/hypervisor>, 2014. Accessed: 2014-07-20.
- [15] The volatility framework 2.3.1. <https://code.google.com/p/volatility/>, 2014. Accessed: 2014-07-27.
- [16] Xen project. <http://www.xenproject.org>, 2014. Accessed: 2014-07-20.
- [17] Welcome to apache hadoop. <http://www.hadoop.apache.org>, (Accessed July 20 2014).
- [18] Simson Garfinkel, Paul Farrell, Vassil Roussev, and George Dinolt. Bringing science to digital forensics with standardized forensic corpora. *digital investigation*, 6:S2–S11, 2009.
- [19] Simson L Garfinkel. Digital forensics research: The next 10 years. *Digital Investigation*, 7:S64–S73, 2010.
- [20] Simson L Garfinkel. Digital media triage with bulk data analysis and bulk_extractor. *Computers & Security*, 32:56–72, 2013.
- [21] George Grispos, Tim Storer, and W Glisson. Calm before the storm: The challenges of cloud computing in digital forensics. *International Journal of Digital Crime and Forensics*, 4(2):28–48, 2012.
- [22] Phil Harvey. Exiftool 9.69. <http://www.sno.phy.queensu.ca/~phil/exiftool>, 2014. Accessed: 2014-07-27.
- [23] Hanan Hibshi, Timothy Vidas, and Lorrie Faith Cranor. Usability of forensics tools: a user study. In *IT Security Incident Management and IT Forensics (IMF), 2011 Sixth International Conference on*, pages 81–91. IEEE, 2011.

- [24] Peter Mell and Tim Grance. The nist definition of cloud computing. 2011.
- [25] Scientific Working Group on Digital Evidence. Swgde model quality assurance manual for digital evidence laboratories, 2012.
- [26] Marc Parisi, David A Dampier, Rayford Vaughn, and Yoginder Dandass. Improving foremost execution speed by data and task level parallelization. 2009.
- [27] Nicole Perlroth. Tally of cyber extortion attacks on tech companies grows. <http://bits.blogs.nytimes.com/2014/06/19/tally-of-cyber-extortion-attacks-on-tech-companies-grows/?php=true&type=blogs&r=0>, Accessed: 2014-07-20.
- [28] Openwall Project. John the ripper 1.7.9 jumbo 7. <http://www.openwall.com/john>, 2014. Accessed: 2014-07-27.
- [29] Vassil Roussev. Scalable data correlation. In *Eighth annual IFIP WG*, volume 11, 2012.
- [30] Vassil Roussev. sddhash 3.4. <http://roussev.net/sddhash/sddhash.html>, 2014. Accessed: 2014-07-27.
- [31] Vassil Roussev, Candice Quates, and Robert Martell. Real-time digital forensics and triage. *Digital Investigation*, 10(2):158–167, 2013.
- [32] Vassil Roussev and Golden G Richard III. Breaking the performance wall: The case for distributed digital forensics. In *Proceedings of the 2004 Digital Forensics Research Workshop*, volume 94, 2004.
- [33] Vassil Roussev, Liqiang Wang, Golden Richard, and Lodovico Marziale. A cloud computing platform for large-scale forensic computing. In *Advances in Digital Forensics V*, pages 201–214. Springer, 2009.
- [34] Naval Postgraduate School. bulk extractor 1.5 alpha 6. https://github.com/smsong/bulk_extractor, 2014. Accessed: 2014-07-27.

Biographies



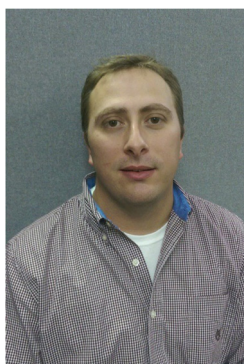
Cody Miller is a Research Associate for the Distributed Analytics and Security Institute at Mississippi State University. Cody's research interests are in Cloud Computing, Computer Security, and Digital Forensics. He has a B.S. and M.S. Degree in Computer Science & Engineering from Mississippi State University. In his graduate studies he worked for the National Forensics Training Center at Mississippi State University where he taught law enforcement officers digital forensics.



Dae Glendowne is an Assistant Research Professor at the Distributed Analytics Security Institute at Mississippi State University. He is currently pursuing his Ph.D. in Computer Science at Mississippi State University. He has a B.S. Degree in Computer Science from the University of Tennessee at Martin and an M.S. Degree in Computer Science from Mississippi State University. His research interests include malware analysis, memory forensics, and applying machine learning to computer security problems.



Dr. Dave Dampier is a Professor of Computer Science & Engineering at Mississippi State University specializing in Digital Forensics and Information Security. He currently serves as Director of the Distributed Analytics and Security Institute, the university level research center charged with Cyber Security Research. In his current capacity, Dr. Dampier is the university lead for education and research in cyber security. Prior to joining MSU, Dr. Dampier spent 20 years active duty as an Army Automation Officer. He has a B.S. Degree in Mathematics from the University of Texas at El Paso, and M.S. and Ph.D. degrees in Computer Science from the Naval Postgraduate School. His research interests are in Cyber Security, Digital Forensics and Software Engineering.



Kendall Blaylock received his M.S. and B.S. degrees from Mississippi State University. During that time he worked as a research assistant in the area of computer forensics. After graduating from MSU he then went on to work for the National Forensic Training Center at MSU. At the NFTC Kendall is currently serving as a Research Associate III. The research associate position

at the NFTC requires Kendall to be an instructor as well as a researcher in the area of digital forensics. As an instructor for the NFTC, Kendall provides training for law enforcement officers and Military Veterans. In addition to being an instructor for the NFTC, he also oversees and conducts research projects at the NFTC. These projects are intended to benefit the digital forensics community and allow law enforcement to conduct investigations in a more effective and efficient manner. Kendall's background in the College of Business at MSU enables him to research where digital forensics is involved with business operations, such as the area of e-discovery and internal corporate investigation.