Edith Cowan University
Research Online

Australian Digital Forensics Conference | Conferences, Symposia and Campus Events

2017

A centralised platform for digital forensic investigations in cloud-based environments

Shaunak Mody
Security & Forensics Research Group, Auckland University of Technology, Shaunakmody14892@gmail.com

Alastair Nisbet
Security & Forensics Research Group, Auckland University of Technology

DOI: DOI 10.4225/75/5a8392cd1d280
Published as: Mody, S., & Nisbet, A. (2017). A centralised platform for digital forensic investigations in cloud-based environments. *Paper presented in Valli, C. (Ed.). The Proceedings of 15th Australian Digital Forensics Conference 5-6 December 2017, Edith Cowan University, Perth, Australia.*
This Conference Proceeding is posted at Research Online.
http://ro.ecu.edu.au/adf/174

A CENTRALISED PLATFORM FOR DIGITAL FORENSIC INVESTIGATIONS IN CLOUD-BASED ENVIRONMENTS

Shaunak Mody, Alastair Nisbet,
Security & Forensics Research Group, Information Technology & Software Engineering Department
Auckland University of Technology, Auckland, New Zealand
Shaunakmody14892@gmail.com, anisbet@aut.ac.nz

Abstract

Forensic investigations of digital media traditionally involve seizing a device and performing a forensic investigation. Often legal and physical obstructions must be overcome so that the investigator has access to the device and the right to secure it for investigation purposes. Taking a forensic image of a hard disk may need to be done in the field but analysis can usually be performed at a later time. With the rapid increase in hard disk size, the acquiring of a forensic image can take hours or days. This poses significant issues for forensic investigators when potential evidence resides in the cloud. What is highly desirable is the ability to perform the acquisition of the image and the data recovery whilst the data remains in the cloud. The comparatively small amount of recovered data can then be downloaded from the cloud. This may solve legal, time and physical obstacles with one relatively simple method. This research describes the development of cloud-based software to perform a digital forensic investigation in the cloud and describes the efficiency of the process under several different configurations utilising Amazon Web Services cloud solutions.

Keywords: cloud, forensics, network security, investigation

INTRODUCTION

As the increase in Internet speeds has seen rapid growth, the uptake of basing data in the cloud for many businesses and individuals has followed that growth. Many businesses are utilising the cloud for some, or all of their data needs, including platforms, software and infrastructure such as their corporate databases. The advantages are many with cheaper access to software that may be prohibitively expensive to purchase, the latest iterations of operating systems and other software for no extra fee and the ability for employees to access the same data from anywhere in the world. One drawback of cloud-based data for organisations is that they store all their corporate data in one place controlled by a third party. This single point for a company's data may also be attractive for the nefarious hacker who targets the data for malicious destruction, modification or theft.

The cloud computing spending in the overall IT industry is expected to increase greatly by the end of 2017. In a recent report published by Gartner in the year 2016 and cited by Cooke, it is predicted that the worldwide spending on the cloud is expected to grow by 18% by the end of 2017. The expected investment in the cloud computing industry is approximated to reach a total of US$247 billion dollars, up from US$209 billion dollars in 2016. A major share of this is expected to come from IaaS and PaaS. A growth of 37% is projected in 2017 in IaaS and PaaS and SaaS which is expected to grow by 20% in 2017 (Viveca Woods, 2016) (Cooke, 2016).

However, whilst the corporations are enjoying the many benefits of cloud-based services, the forensic investigators are finding these same services provide significant challenges to forensic investigations. Firstly there may be legal hurdles to overcome to gain access to the cloud-based data. Whilst the Budapest Convention on Cybercrime has seen many nations agree to cooperate in seizing and sharing of data, many countries have resisted signing up to the agreement amongst concerns over bypassing the usual checks and balances of privacy and security provided by the law, including New Zealand. The legal process in some countries may be extremely challenging and time-consuming and when time-constraints exist, both for time-critical investigations but also to prevent deletion of evidence, this alone can foil an investigation. If the legal hurdles can be overcome, then the next challenge is to have access to the data. Usually this will reside on one or many hard drives on a cloud server residing in an often undisclosed location in the world. The investigator may have to trawl through terabytes or more of information looking for the evidence required. This 'logical' view of the data is not what is normally of interest as often the best evidence has been recently deleted. Therefore, a 'physical' view of the evidence, which is a direct bit by bit copy of the hard drive(s) is much preferred. This forensic image often allows the investigator to locate and recover deleted files. However, there is currently little ability to do this rather than attempt to copy

the entire logical image of the hard dive or drives to the investigator's computer using the Internet. This is a relatively slow process and with the very large hard drives that are common in the cloud may not be practical. What is urgently required is the ability to perform the investigation on the data whilst it remains in the cloud, and this includes the forensic acquisition of the image as well as the locating and recovery of the evidence. This research attempts to show, through development of new software in the form of a bash script that can be uploaded to the cloud, that acquisition and analysis of data in the cloud is possible and realistic.

The following section discusses the cloud environment and the options available to users.

LITERATURE REVIEW

The cloud offers a 'shared responsibility model' which describes the responsibilities of the parties involved. This shared responsibility model generally has three main parties involved, the cloud provider, the service provider and the service consumer. A cloud provider represents an entity which is responsible for providing and building the infrastructure for running cloud-based services. The cloud providers are generally the owners of the infrastructure where the services are set up. A service provider is an entity which is responsible for running services, applications and provides an interface for the end-users to use. A service consumer is the entity which utilises the services provided by the service providers (Mohamed Al Morsy, 2016).

Forensic investigations follow accepted practices to ensure that the evidence acquired retains its integrity and can be presented in a court of law if necessary. Whilst guidelines exist on the finer points of acquiring and analysing evidence the basic investigation process includes seizing the device, examining the data recovered for evidence, analysing the evidence for probative value and finally reporting and presenting the evidence in a sound manner (NIST 2006).

The digital forensic guidelines provided by NIST, ACPO and other law enforcement agencies cannot at this stage be extended to the cloud environments as they only describe how to perform examinations on the infrastructure where they have physical access to the evidence. These guidelines do not mention the steps that need to be taken by examiners while conducting investigations in a distributed environment such as that of the cloud. This lack of a cloud-specific framework on how to perform investigations in the cloud has made it difficult for the examiners to conduct examinations in the cloud (Rani & Geethakumari, 2015). Several researchers have suggested using a combination of various existing frameworks for conducting investigations on the cloud. A widely used combination is that of NIST and the McKemmish framework where the acquisition and recovery is performed in the same phase followed by examination, analysis and reporting. This combination does aid an investigation but still does not guide an investigator on how to perform investigations in the cloud (Rani & Geethakumari, 2015).

The current most common commercial forensic tools such as EnCase and FTK do not provide the functionality to perform investigations on the cloud and require administrator permissions to access the data (Alqahtany, Clarke, Furnell, & Reich, 2016). These extra permissions that are pose a threat as these agents need to be trusted for transferring confidential data from the cloud to the local server (Rani & Geethakumari, 2015). Tools such as Oxygen Forensic Extractor have been found to be successful in examining the evidence acquired from a distributed computing environments but it has also been observed that it cannot be used for performing acquisitions on the cloud environment (Rani & Geethakumari, 2015). The trust factor required by tools that can acquire an image but then necessitate downloading the image to a local computer means they have failed to penetrate the cloud forensics industry (Dykstra & Sherman, 2012).

A framework called 'The integrated conceptual digital forensic framework for cloud computing' is a combination of the framework proposed by NIST and McKemmish. The NIST framework proposes a four step forensic approach including data collection, examination, analysis and reporting (Mell & Grance, 2014). The Mckemmish framework consists of the identification of the digital evidence, the acquisition of the digital evidence, the analysis of digital evidence and the preservation of the digital evidence (McKemmish, 1999).While the NIST framework focuses on the examination and reporting of the data, the McKemmish framework focuses of identification and preservation of the evidence and is an iterative framework which focuses on the identification of the source of the data for the examination (Martini & Choo, 2012).

The California Fire Assistance Agreement (CFAA) model proposed a central command and control server responsible for performing all the forensic activities on the cloud (Alqahtany et al., 2016). This proposed framework aims at solving the restrictions of cross-border data privacy and security laws that exist in a cloud environment by providing them with a platform deployable on the cloud itself (Sibiya, Venter, & Fogwill, 2015). For a trusted third party to be able to perform forensic investigations on the cloud, they have to be given access to the cloud environments and also have to be verified by both the cloud service providers and the customer

whose cloud infrastructure they will be accessing for performing examinations which is a serious drawback. (Meera, Alluri, Powar, & Geethakumari, 2015). The majority of these guides fail to address the concern of the efficiency of the tools and frameworks (Casey, Katz, & Lewthwaite, 2013).

What is required is an efficient solution that will address the legal and technical issues involved in cloud forensics. This research attempts to do this by describing the development of custom-designed software that can be uploaded to an organisation's cloud-based data centre as a file that will be accepted onto the cloud. The file is a custom-written script that performs 2 distinct operations. Firstly, it acquires an image of the partition, hard drive or possibly multiple hard drives, and secondly it processes that image by analysing it for the desired files, including deleted files. These files may be documents, graphics or any other file that the examiner chooses to look for.

The efficiency is present because the acquisition and recovery of the evidence is done at the same stage in the cloud. The acquired image is then sent into a data stream which accepts the evidence as input and then runs a recovery and generates a directory containing extracted data. The software runs an automated phase for examination and analysis of the extracted data to improve the efficiency by reducing the time required for performing an examination, as this is a desirable feature of forensic software (Casey et al., 2013). It was expected that performing the acquisition and analysis in the cloud would be very efficient as there is no data transfer time required until the recovered files are downloaded after the forensic processes have completed.

The following section describes the experimental setup and testing of the new software.

RESEARCH DESIGN

This research focuses on the Linux-based environments as according to Amazon, more than half a million servers running on AWS are Linux (Amazon Web Services, 2017a). The software was designed and developed on the Ubuntu operating system version 14.04.01. The scripting language used for developing this software is bash scripting with 1400 lines of code and the automation platform uyilises cssh. Amazon Web Services (AWS) was selected as reports published by Gartner and Forbes state that currently AWS is the leading cloud service provider in the market as discussed by Cooke (Cooke, 2016). The Cloud offers its users the option to choose an operating system from a wide range of distributions such as Microsoft Windows Servers, Linux Servers including Ubuntu Server, Red Hat Enterprise Linux, Suse Linux and CentOS. This type of flexibility makes it easy for the customer to deploy a hybrid infrastructure much more easily on the cloud than with local on-premise infrastructure.

Furthermore, once the user has selected the operating system, the cloud service provider offers an option of selecting the configuration of the server based on the purpose of the deployment. Each server has a 10GB hard disk, so that multiples of 10GB must be acquired and analysed. For example, the single server required a 10GB acquisition, 2 servers requires 20GB and 3 servers 30GB. The timings of these acquisitions and analysis can then be used as a guide for forensic investigators who can multiple the time taken to then calculate the time required for larger disks. For example, the timing of 10GB on a single server is multiplied by 100 if the hard disk in the cloud is a single server with 1TB, and so forth. The RAM and CPU configurations range from a minimum of 512 MB of RAM and 1 vCPU to a terabyte of RAM and 128 vCPU's.

The software was tested and developed on the T2 series of server models offered by Amazon Web Services. The software was developed and tested on the three of the biggest models offered by the T2 series, T2.large, T2.xlarge and T2.2xlarge. These three server models were tested against all the server models offered by T2 family of Amazon web services. The client nodes that used were as follows t2.micro, t2.small, t2.medium, t2.large, t2.xlarge, t2.2xlarge. Each model represents servers with different hardware configurations. Each server was tested against all the client nodes for benchmarking the performance of the server to identify the most optimal server.

Each experiment involved uploading the script to the cloud where the target data resides. Then the script was executed which acquired a forensic image of the selected partition. Once the image was acquired, a forensic analysis was made. This involved extensive keyword searches for both logical files and deleted files. The results were compared against each other using the time required to perform the actions as a comparison. Each experiment was run 18 times, as this gave an acceptable but manageable number of iterations to manage and average that was then used for the final result.

RESULTS

The first experiment involved testbed-1. The tested server is t2.large and is equipped with 8GB of RAM and 4 vCPUs. Six different client nodes were tested ranging from t2.micro to t2.2xlarge. The next experiment, testbed-2, utilised t2.xlarge with 16GB of RAM and 8 vCPUs. Similar ranges of client nodes were used. Finally, the third testbed, testbed-3, was T2.2xlarge with 32GB of RAM and 8vCPUs with similar ranges of client nodes. It was felt that this would give a good comparison in performance for the client nodes with the increasing ability of the server to process the information. The number of client servers present during a given iteration varied from a single server to a maximum of three servers during the experiments. Figure 1 shows the time taken to acquire the image using T2 Large.

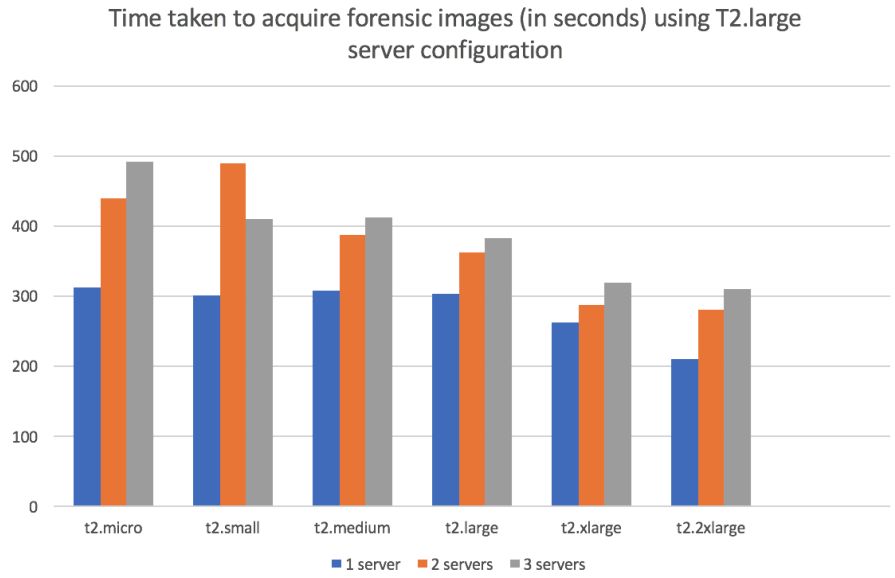

Figure 1: T2.large acquisition results

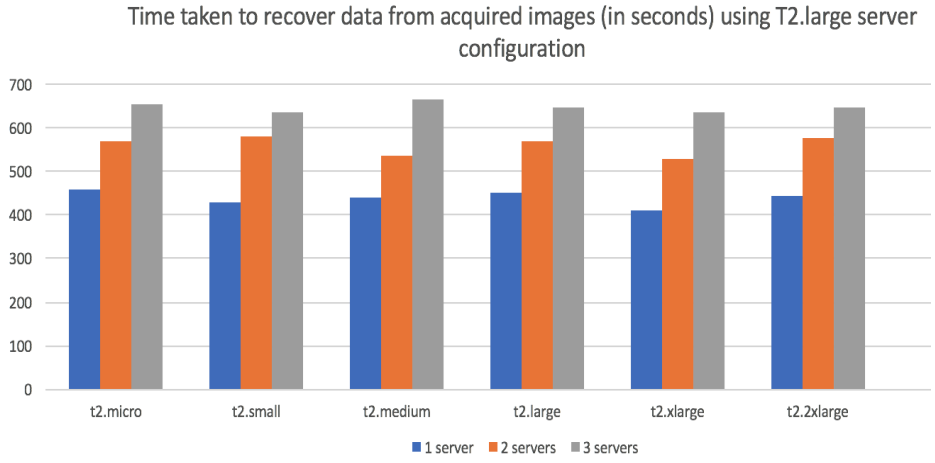

Figure 2: T2.large data recovery results

Figure 2 shows the results for a data recovery performed on this image whilst it resides in the cloud. The results from sever T2.large tested against the 6 client nodes with 1 to 3 servers indicate that the more servers available the slower the time to acquire an image. Generally, a linear decrease in time is seen across all clients as they move from micro through to t2.2xlarge. Image time remains approximately the same for all client nodes. Table 1 expands on these acquisition results by showing the T2.large server tested against the T2.large client in its 3 differing configurations.

Table 1: Server T2.large with client nodes T2

Client node configuration	No. of servers	Time taken to acquire forensic image
T2.large	1	5 minutes 3 seconds
	2	6 minutes 2 seconds
	3	6 minutes 22 seconds
T2.xlarge	1	4 minutes 23 seconds
	2	4 minutes 47 seconds
	3	5 minutes 20 seconds
T2.2xlarge	1	3 minutes 30 seconds
	2	4 minutes 40 seconds
	3	5 minutes 9 seconds

What is clear is that predicting the time taken from the number of servers cannot be done precisely but a reasonable estimate of the time required is possible. This may be useful when acquiring the image as it is desirable that the data is not be modified by users when the image is acquired. Figure 3 shows the results from acquiring an image when utilising T2.xlarge as the server.

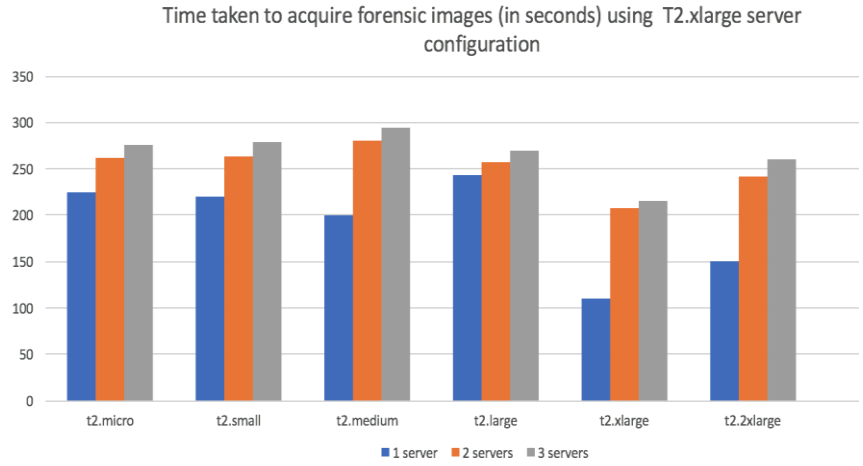

Figure 3: T2x.large acquisition results

The next step is to recover data from the forensic image and the time taken to complete this task with 1-3 servers is shown in figure 4.

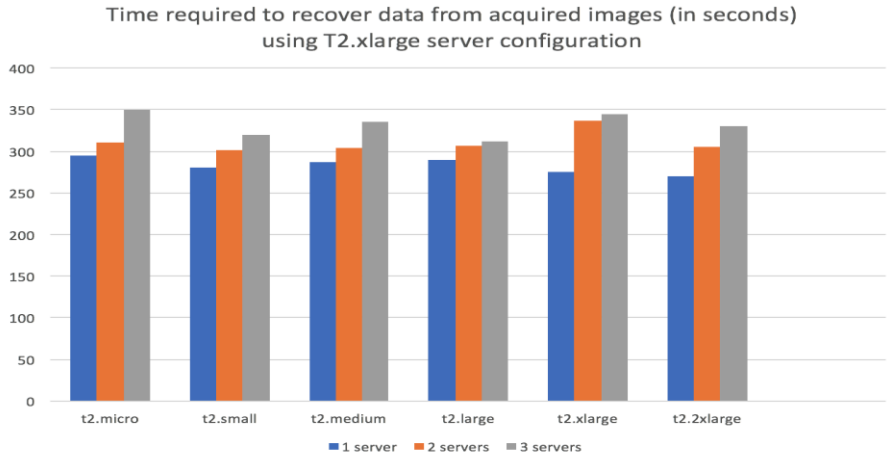

Figure 4: T2.xlarge acquisition results

Figures 3 and 4 indicate that as the servers utilised are more powerful in terms of increased RAM and virtual CPUs, the time required to acquire an image and recover data from the image decreases significantly.

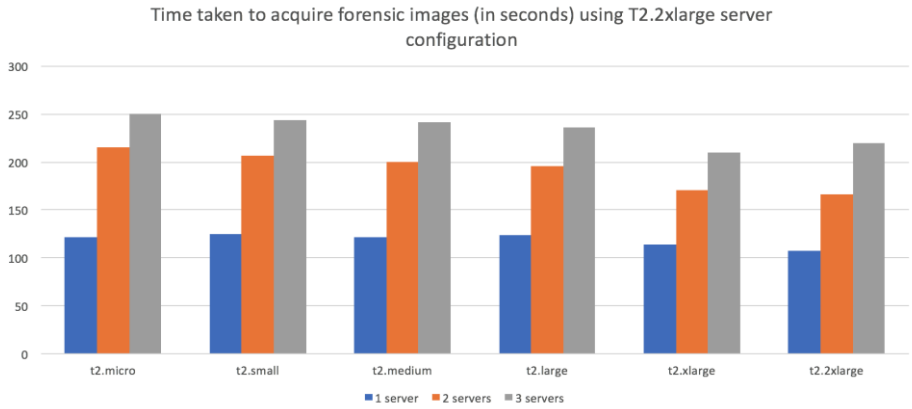

Figure 5: T2.2xlarge acquisition results

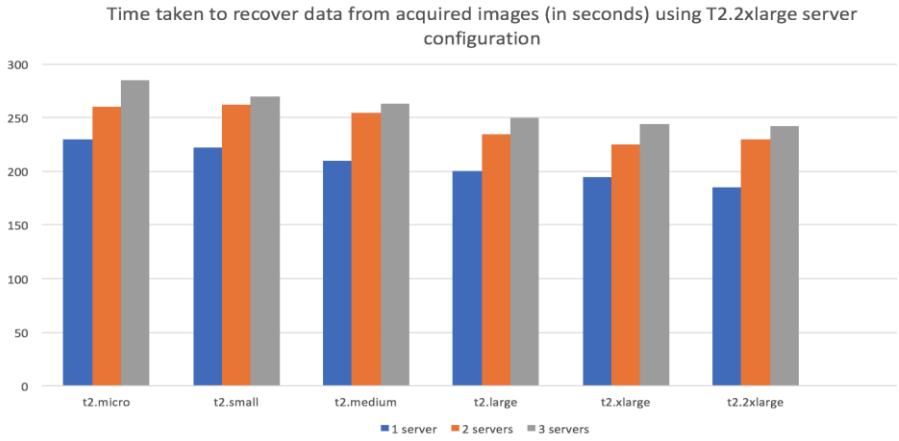

Figure 6: T2.xlarge acquisition results

Figures 5 and 6 show that an increase in RAM and vCPUs lead to increasing efficiency in acquisition and data recovery. The time decreases from using the less powerful T2.large but not significantly. The results obtained from running the experiment with central forensic server as t2.large model showed that t2.micro took the most time for acquiring images and recovering data. The quickest client node was t2.2xlarge. The average time required to acquire images from the client nodes was between 4 minutes and 6 minutes. The results also showed that in order to acquire the evidence more efficiently the server performs best with client node running t2.2xlarge. Further, while recovering data from the acquired images it was found that t2.2xlarge performed the best when the server was configured as t2.large. While identifying the most efficient server model for this software deployment, it was found that the time difference between the most efficient node and the other two nodes is minimal. While the results of the test cases show that t2.micro was the slowest performing client node, t2.2xlarge was the most efficient client node for performing recovery and a combination of t2.2xlarge, t2.large and t2.small was identified as the most efficient for acquisition. T2.2xlarge was the identified as the most efficient model having 8 vCPU's and 32GB of RAM, the highest configuration server amongst the three.

The average time taken to acquire the image and perform recovery is shown in table 2. This shows that t2.2xlarge is the best performing server of the 3 tested. This increase in efficiency is measured by time and given as a percentage compared to the baseline.

Table 2: Percentage difference from benchmark T2.2xlarge

Client node configuration	No. of servers	Time difference for acquisition (in percentage)
T2.micro	1	149%
	2	157%
	3	159%
T2.small	1	143%
	2	146%
	3	157%
T2.medium	1	146%
	2	138%
	3	133%
T2.large	1	144%
	2	129%
	3	123%
T2.xlarge	1	125%
	2	102%
	3	103%
T2.2xlarge	1	100%
	2	100%
	3	100%

DISCUSSION

It can be seen from the performance of the script in the cloud that acquisition and recovery times are a few minutes for 10GB hard drives, and even expanding this to 30GB with 3 servers requires only a relatively small increase in time of up to 50%, something that is still very acceptable. As more servers are added, the time increases by a small amount and this follows the configuration changes for the client nodes, where more powerful nodes consequently reduce the time for acquisition and recovery but only by a relatively small amount.

Therefore, the results indicate that performing a data acquisition followed by file recovery is possible in the cloud. Further, the increased efficiency by performing these actions whilst the data remains in the cloud means that this type of forensic process could realistically be performed on very large hard drives. Whilst the time to acquire and recover files could take many days or more, this may be of little concern if the primary concern is to recover the forensic evidence. The alternative and more usual practice of gaining physical access to the hard drive(s) is likely not possible in a cloud environment. Legal issues, especially present in many countries where the data may be stored, along with Internet speeds that are simply too slow to transfer terabytes or more of data, mean that traditional recovery is often not realistic. Additionally, downloading a large amount of data in an acquisition and then recovering a small amount of data that is of interest means that most of the data is of no value to the investigator. Overall, results indicate that this process has significant benefits for forensic investigators, and whilst more development is required to expand its use to more platforms, services and configurations, the basic concept shows significant promise.

CONCLUSION

The cloud continues to see extensive growth and acceptance by organisations and individuals, especially as faster Internet speeds are offered globally. However, the challenges posed to forensic investigators by this technology are not keeping pace with the developments of cloud technologies. Currently, the legal, technical and other practical challenges make forensic investigations of cloud-based data extremely difficult and therefore often unsuccessful. These results show that these challenges can largely be overcome by deploying this type of software in the cloud alongside the forensic data sought by the investigator. The results of these experiments show that this is an effective and efficient method for collecting forensic data as no data transfer is required during the process, other than downloading the recovered files of interest once the processes have completed. Whilst this software is sufficient to perform the forensic processes in the tested environment, many different

services running different hardware and software configurations exist in the cloud. Therefore, continued development of this type of software is required so that all services on all platforms can be investigated in this manner, leading to timely and effective forensic data gathering by investigators around the world.

REFERENCES

Alqahtany, S., Clarke, N., Furnell, S., & Reich, C. (2016). A forensic acquisition and analysis system for Iaas. Cluster Computing, 19 (1), 439-453.

Amazon Web Services. (2016). AWS cloud formation. Blog.

Amazon Web Services. (2017a). Amazon ec2 instance types. Retrieved from https://aws.amazon.com/ec2/instance-types/

Casey, E., Katz, G., & Lewthwaite, J. (2013). Honing digital forensic processes. Digital Investigation, 10 (2), 138-47.

Cooke, L. (2016, August). What's changed: Gartner's 2016 cloud infrastructure-as-a-service magic quadrant. Retrieved from https://solutionsreview.com/cloud-platforms/whats-changed-gartners-2016-cloud-infrastructure-as-a-service-magic-quadrant/

Dykstra, J. & Sherman, A. T. (2012). Acquiring forensic evidence from infrastructure-as-a-service cloud computing: exploring and evaluating tools, trust, and techniques. Digital Investigation, 9, S90-S98.

Meera, G., Alluri, B. K. R., Powar, D., & Geethakumari, G. (2015). A strategy for enabling forensic investigation in cloud iaas. International conference on computer and communication technologies (icecct), 2015 ieee (pp. 1-5).

Mell, P. & Grance, T. (2014). Nist cloud computing forensic science challenges. Draft Nistir, 8006.

Mohamed Al Morsy, I. M., John Grundy. (2016). An analysis of the cloud computing security problem. Computer Science & Software Engineering, Faculty of Information & Communication Technologies Swinburne University of Technology, Hawthorn, Victoria, Australia.

Martini, B. & Choo, K.-K. R. (2012). An integrated conceptual digital forensic framework for cloud computing. Digital Investigation, 9 (2), 71-80.

McKemmish, R. (1999). What is forensic computing? Australian Institute of Criminology, Trends and Issues in crime and criminal justice.

NIST (2006). Guide to Integrating Forensic Techniques into Incident ResponseSpecial Publication 800-86.

Rani, D. R. & Geethakumari, G. (2015). *A meta-analysis of cloud forensic frameworks and tools*. IEEE Conference on power, control, communication and computational technologies for sustainable growth. 11-12 December 2015. Kurnool, India. doi: 10.1109/PCCCTSG.2015.7503922

Sibiya, G., Venter, H. S., & Fogwill, T. (2015, May). Digital forensics in the cloud: the state of the art. In 2015 IST-Africa Conference. 6-8 May 2015. Lilongwe, Malawi. doi:10.1109/ISTAFRICA.2015.7190540

VivecaWoods, R. (2016, January). Gartner says worldwide public cloud services market is forecast to reach US$204 billion in 2016. Retrieved from http://www.gartner.com/newsroom/id/3188817