<http://www.diva-portal.org>

This is the published version of a paper presented at *ACMSE 2017 The Annual ACM Southeast Conference*.

Citation for the original published paper:

Alobaidli, H., Nasir, Q., Iqbal, A., Guimaraes, M. (2017)

Challenges of Cloud Log Forensics.

In: *Proceedings of the SouthEast Conference* (pp. 227-230). ACM
ACM SE '17

<https://doi.org/10.1145/3077286.3077302>

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-207032>

Challenges of Cloud Log Forensics

Hanan Alobaidli, Qassim Nasir
University of Sharjah, Sharjah
UAE
U00023157@sharjah.ac.ae

Asif Iqbal
KTH Royal Institute of Technology,
Stockholm
Sweden
asif.iqbal@ee.kth.se

Mario Guimaraes
Saint Martin's University, Lacey,
Washington
USA
mguimaraes@stmartin.edu

ABSTRACT

The forensics¹ investigation of cloud computing is faced by many obstacles originating from the complex integration of technologies used to build the cloud and its sheer size. In this research we aim to provide an insight into cloud computing log forensics, as logs are an important source of forensic evidence in the cloud. This is followed with conclusions regarding the issues faced by researchers in log forensics in cloud computing that will aid the research process.

KEYWORDS

Log, Security, Cloud computing, Digital Forensics

1 INTRODUCTION

Cloud computing is considered one of the significant areas in the future of computing, giving the end user computing resources similar to other utilities such as water and electricity. This move to Cloud computing needs to take into consideration the security and legal requirements of users and organizations using the cloud. In a case of a crime or a misunderstanding, how would an organization retrieve evidence from the cloud? Their dilemma would be that, in most cases, they don't have physical access to the computing resources that are required in order to undertake a regular forensics investigation. Another issue would be, even if they have physical access, how would they deal with the cloud environment. For example, in a public cloud the resources are used by multiple tenants and data may well be stored in decentralized locations. This makes the forensic investigation process more complex. Using techniques such as imaging a drive for retrieving deleted files or finding evidence in slack space becomes a daunting task. From that perspective, we considered that logs may provide a good set of evidential artifacts that might aid in identifying what needs to be found in the cloud environment and logs are more accessible compared to other forensics artifacts. That being said, logging in the cloud also has problems such as the decentralization of logs, volatility of logs,

multiple tiers and layers, different archiving and retention policies, accessibility of logs, nonexistence of logs, the absence of critical information in logs and incompatible / random log formats [1]. Hence we need to understand the research undertaken in this realm and identify the possible improvements that can be made to logging mechanisms in clouds.

The first section of the paper provides an introduction to the topic, followed by discussion of log forensic challenges in cloud computing according to NIST. While the third section provides a literature review of the research that has been carried out in the realm of cloud log forensics and relates the solutions to the challenges identified by NIST. This is followed by a discussion and conclusions of the research direction for cloud log forensics.

2 FORENSICS LOGGING CHALLENGES IN CLOUD ENVIRONMENT

In 2014 the NIST [2] published a report that discussed the digital forensic challenges in the cloud computing ecosystem. The aim of this research was to gain a deep understanding of these challenges and identify technologies and standards to mitigate them. Out of the 65 cloud forensics challenges identified, 13 of the challenges were related to logs. These challenges are as follow: Timestamp synchronization, Log format unification, Log capture, Limited knowledge of log records, Identification of the storage media containing the log files and its physical location, Volatility of log files stored in VMs, Log file chain of dependencies, Evidence identification, The integrity of the logs used in investigation, Decentralization of logs between different layers, Accessibility of logs, The use of logs in hypervisors is not well understood, Potential segregation of evidence in logs for a multi-tenant environment to protect tenants privacy.

Some of the log forensics challenges are inherited from different technologies but are exaggerated in the cloud environment as a result of its design. For example, timestamp synchronization is a challenge in network forensics, but made more complicated as it must be synchronized across multiple physical machines that may be spread across multiple geographical regions. Similarly, log format unification made more difficult in the cloud environment as a result of the massive range of technologies used, which may produce different log formats and may use proprietary or unusual formats. The integrity of the logs used in an investigation is also made more complicated by the cloud computing architecture. The multi-tenant environment of clouds, where the storage of data is shared among multiple computers, and locations and accessible by multiple parties, makes ensuring the integrity of log records a difficult task.

¹Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

ACM SE '17, April 13-15, 2017, Kennesaw, GA, USA

© 2017 Association for Computing Machinery.

ACM ISBN 978-1-4503-5024-2/17/04...\$15.00

<http://dx.doi.org/10.1145/3077286.3077302>

Another challenge is evidence identification, which is essential in any computer forensics investigation. The effect of the cloud design on this challenge is shown by the fact that most of the artifacts normally used to identify certain actions in an investigation are either not available or are not generated or stored in the same way as they would be in a traditional non-cloud environments. On the other hand some challenges are unique to the cloud environment, such as log capture, as there is no standard method used to collect specific logs across different cloud providers. The Identification of the storage media containing the log files and its physical location is also challenging in the cloud. That is because the physical machines can be spread across different geographic locations, combined with the facts that cloud computing utilizes technologies such as virtualization and have features such as elasticity that makes the identification of the physical location a difficult task. Another issue related to the use of virtual machine technology is that the removal of a VM would result in the loss of its log files, which makes them, in a way, volatile if not saved in a centralized location. Other issues unique to the cloud ecosystem are the accessibility and decentralization of logs between the different layers. The accessibility of the logs differs between different cloud service models, where the IaaS user has more access compared to SaaS user. Taking that in consideration, the accessibility is also limited for IaaS users where the user would require the support of the cloud service provider (CSP) to get logs for layers such as hardware and networking. The multi-tenant environment also provides a challenge for forensic investigation, as the investigator needs to access the data of one user without breaching the confidentiality and privacy of the other tenants. This task is difficult, according to NIST, as the technology used for the segregation of tenants is not effective enough. Identifying the log file chain of dependences is also challenging in cloud as the CSP may depend on another CSP to provide them with computing services. Hence, in this case, a forensic investigator needs to investigate each dependency link in the dependency chain of CSPs.

Considering that cloud computing is in its infancy, it also provides a set of challenges related to the lack of knowledge of its inner workings and how that affects forensic investigations. An example of the challenges presented in the NIST report is the lack of training materials to educate the forensics investigator on cloud computing technology and cloud forensics operating policies and procedures. This is also combined with the limited knowledge of logs and records available in the cloud environment. The NIST challenges for log forensics in cloud computing can be divided in four categories which are, challenges related to log content, challenges resulting from the cloud architecture, challenges related to data collection in a cloud environment, and finally challenges related to the lack of training and research in cloud computing forensics (see Fig. 1).

3 LITERATURE REVIEW OF CLOUD LOG FORENSICS

A current research direction in cloud forensics is the utilization of logging to preserve evidence in the cloud and reduce dependency on CSP. An example of this research perspective is the work done by Sang [3]. In his work he studied the utilization of logging on both SaaS and PaaS service models as a source of forensic evidence. They stated that modifying the logging mechanism for PaaS and SaaS is more beneficial in terms of potential evidence compared to IaaS. When it comes to IaaS logging outside the Virtual machine will not provide much useful evidence, but we do not agree with this statement. In his opinion most of the evidence for this service model can be acquired by getting an image of the VM and its logs, but the work didn't consider the effect of deleting the VM. Other work carried out by Zawoad et. al. [4] [5] presenting Secure-Logging-as-a-Service (SecLaaS), which stores a virtual machine's logs and provides access to forensic investigators, while ensuring the confidentiality of other cloud users. They claim that their logging mechanism preserves proof of past logs (PPL) and thus protects the integrity of the logs from dishonest investigators or CSP.

With regard to IaaS Dykstra et. al. [6] developed and designed the Forensic Open-Stack Tools (FROST). The tool focuses on the trustworthy forensics acquisition of virtual disks, API logs, and guest firewall logs. The collection is done from the CSP at the host OS level. The collected data are then made available within the management plane to the user. On other hand, Thorpe et. al. [7] discussed the desired data and logs that can be gathered from the kernel hypervisor. They have categorized the desired information into 4 categories which are: 1) Information that is available to the VM host OS and recorded on non-volatile media. 2) Information that is available to the VM host system, but is not recorded. 3) Information that is not currently available to the VM host OS, but could be made available. 4) Information that it is impossible to obtain from a VM host OS.

The first source of information is the most accessible and utilized in an investigation. While the other sources of evidence require researchers to study them. Thorpe [8] stated that the hypervisor system logs can be used to track VM incidences which may later be used to compile potential evidence for a cloud investigation. He also discussed several methods that can be used to acquire logs from that cloud which are the Trusted Platform Modules (TPMs), the cloud management plane (management console),

Figure 1 : Mind Map of NIST Logging Forensics Challenges in the Cloud Categorization

forensics as-a-service, contractual support and virtual machine introspection (VMI). Log contents were also discussed in the paper where the authors stated that the desired information in a log based forensics investigation needs to answer few questions. These questions are: Who, what, when, how and where, relating to the VM user, machine and actions done by users or on VMs. They suggested the use GPS based coordinate, global unique identifiers in logs in order to assist in the identification of the physical location of data and log records.

Böck et. al [9] researched the methods that can be utilized to provide secure log data which can be used as trustable evidence in a forensics investigation. Their proposed model used a TPM and AMD secure virtual machine (SVM) concept to establish a root of trust for client SYSLOG daemons. This would provide a hardware based trust in the log production. On the other hand, Zafarullah et. al. [10] analyzed Eucalyptus, which is an open source cloud framework, with the aim to identify logs useful for forensic investigation of cloud applications or a CSP. In their work they concentrated on identifying SYSLOG, Snort and other log entries that could help detecting attacks on the cloud and aid in forensic investigation. With regards to SaaS logging, Marty [1] discussed a set of logging frameworks and guidelines that can provide a proactive logging approach to ensure that the data needed for forensics investigation has been generated and collected. The logging guidelines discussed in this work included when to log, what to log and how to log. With regard to what to log he stated that the minimum information that needs to be present in every log which are: Timestamp, Application, User, Session ID, Severity, Reason, Categorization. This work would provide more information for investigating a SaaS service model, but one of its limitations is that it didn't discuss the other layers of the cloud such as the virtual machine logs, hypervisor logs and other networking and hardware logs. This limitation is a result of the access provided by the SaaS service model implementation, but nevertheless this information is needed for a forensics investigation.

Sibiya [11] proposed a Live Digital Forensic Framework for a Cloud (LDF2C) environment. This framework attempted to address the live forensics techniques for sound evidence collection in the cloud, and the association of the collected data to a specific user. This was an improvement to [1] by attempting to collect the volatile evidence while the system is running alongside the log records created in the system. The developed framework utilized data mining techniques in order to extract data from log files from locations of interest for an investigation, such as the hosted remote desktop and the accessing client device. On the other hand, Pătraşcu [12] proposed a framework for cloud computing logging. The proposed framework was designed as a set of five. A local logging module is integrated into the virtualization layer, specifically in the running KVM hypervisor in order to reliably gather the log records. Using this model, the investigator can gather raw log records from a specific VM on the hypervisor or can monitor the whole system during a specified time frame, along with specifying the data quantity. That being said the authors didn't discuss the exact

data that can be gathered using the logging module nor the method used to extract the required data.

Thorpe et. al. [13] also discussed technical issues facing cloud log forensics with the focus on the hypervisor log examination, its challenges and how it would affect the investigation. While in [14] the author discussed integrating Virtual Machine Log Auditor VMLA in an open-stack cloud configuration in order to enable remote forensic collaborating agencies to work seamlessly via suitable service level agreements and privacy preservation constraint mechanisms. On the other hand, Nakahara and Ishimoto [15] discussed concerns about the accountability of actions taken in the cloud and how to log them to insure this accountability. They specified that accountability issues in the cloud are caused by the following: Non locality based issues, linkage between applications, outsource infrastructure. These issues make the log collection process more complicated. The solutions they specified to overcome these issues are the implementation of higher layer state monitoring and state reporting log management system, introducing a Log-Linking Mechanism between Multiple Services and Multiple Servers, introducing a Time Synchronization Mechanism into Sequential Service Logs. They also specified Requirements to Incorporate Accountability into Log Data as well as accountable information structure.

4 DISCUSSION

From the research discussed above, we can outline several issues that require the attention of researchers in the field of cloud computing log forensics. These issues are as follow: First, what log data should be collected in order to assist a forensics investigation. Secondly, the need to address the concern regarding data explosion and log content size. Thirdly, how to handle the integrity of the collected log data. Also, how to correlate between log sources in the cloud. As well as, what are the benchmarks that researchers can use to identify the success of their research with regard to log forensics in the cloud. Finally, how to visualize the log content collected from the cloud to assist a forensics investigation As the field of cloud computing is just at its starting point, forensics research has the advantage of being able to manipulate the structure of the cloud in order to resolve the issues presented above. A possible research direction that we can suggest is improving the cloud logging framework to provide correlated forensics evidence from different layers of the cloud. This direction requires answers being found for several of the issues mentioned above, in order to implement effective solutions.

Table 1 shows that most of the research done focus on the process of data collection, which is essential to the cloud log digital forensics area, but there is a lack in the evidence identification and log format unification issues. This lack will also affect the data collection process as without resolving these issues the collection as well as the analysis problems can become time consuming and difficult. We believe that identifying the optimal cloud log forensics content can play a vital role in the cloud forensics investigation process. Some of the research attempted to identify this optimal content such as [1] [7] [15] discussed it briefly. No metrics or standards were discussed in details identifying the strength of the proposed content.

Table 1 Log forensics research mapped to NIST log forensics challenges

Paper Ref.	NIST Log Forensics Challenge discussed	Challenge Category
Sang, [3]	Log accessibility for SaaS & PaaS	Data Collection
Zawoad, [4] [5]	Log integrity	Data Collection
Dykstra, [6]	Log collection and accessibility of logs	Data Collection
Thorpe, [7]	Knowledge of VM Kernel logs in a cloud environment, What information should be included in cloud logs.	Training and research, Log Content
Thrope, [8]	Knowledge of logs in a cloud environment and its effect on a forensics investigation	Training and research
Boeck, [9]	Log integrity and confidentiality.	Data Collection
Zaferullah [10]	Knowledge of logs in Eucalyptus	Training and research
Marty, [1]	Collection of logs from different cloud components, discussed the minimum data that should be available in a log record	Data collection, Log Content
Sibiya, [11]	Using Data mining to collect the needed log records for an investigation	Data Collection
Pătrașcu [12]	Collection of specific logs	Data Collection
Thorpe, REF [14]	Brief discussion of a method for collecting hypervisor log data	Data collection
Nakahara, [15]	Evidence identification and log format unification	Log Content

Table 2 shows the identified optimal content for cloud forensics investigation.

TABLE 2 Identified Optimal Cloud log forensics content

content	Marty [1]	Thorpe [7]	Nakahara [15]
Timestamp	✓	✓	✓
Application User /who	✓	✓	
Session ID	✓		
Severity	✓		
Reason	✓	✓	
Categorization	✓		
Role Identify			✓
Place/where		✓	✓
Method			✓
Data			✓
Result		✓	✓

The most robust attempt to identify the optimal cloud log content can be seen in the Cloud Auditing Data Federation (CADF) standard. This standard [16] is a product of the DMTF Cloud Auditing Data Federation working group [17]. It defines an event model specifying the essential data needed to certify, self-manage and self-audit application security in cloud environments. Never the less case studies and tests are required to prove if that data identified in the standard are sufficient for a cloud forensics investigation or if we need to identify more log content. That being said is more data is what required, how do we identify what would provide the needed evidence. These are questions that need to be tackled. This needs to be considered because too much data can harm an investigation as well. The investigator might be lost in mountains of unnecessary

information for his/her investigation. Huge spaces are also required to store these logs and more log content will result in larger log files. Hence research will also need to be done on how to optimally store these files.

5 CONCLUSIONS

The field of cloud log forensics is wide open for research as shown in the discussion section. Identifying the optimal cloud log content can assist in answering most of the research questions presented. Never the less these identified log content need to be scrutinized in order provide the most help to a forensics investigation. Along with log content other research directions are data collection and preservation, data integrity and identification of log records in different cloud platforms.

REFERENCES

- [1] R. Marty. 2011. Cloud application logging for forensics. In Proceedings of the 2011 ACM Symposium on Applied Computing (SAC '11). ACM, New York, NY, USA, 178-184. DOI=<http://dx.doi.org/10.1145/1982185.1982226>
- [2] M. Herman, M. Iorga. NIST Cloud Computing Forensic Science Challenges (Draft NISTIR 8006) National Institute of Standards and Technology, U.S. Department of Commerce, 2014 Retrieved on 1th of July 2015 from: http://csrc.nist.gov/publications/drafts/nistir-8006/draft_nistir_8006.pdf
- [3] T. Sang, 2013. A Log Based Approach to Make Digital Forensics Easier on Cloud Computing. In proceedings of the Third International Conference on Intelligent System Design and Engineering Applications (ISDEA), 2013 Hong Kong, pp. 91-94, 2013
- [4] S. Zawoad, A. Kumar Dutta, and R. Hasan. 2013. SecLaaS: secure logging-as-a-service for cloud forensics. In Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security (ASIA CCS '13), ACM, New York, NY, USA, pp. 219-230, 2013
- [5] S. Zawoad, A. K. Dutta, and R. Hasan. 2016. Towards building forensics enabled cloud through secure logging-as-a-service, IEEE Transactions on Dependable and Secure Computing, vol. 13, no. 2, pp. 148-162, Mar. 2016.
- [6] J. Dykstra and A. T. Sherman. 2013. Design and implementation of FROST: Digital forensics tools for the OpenStack cloud computing platform. Digital Investigation, vol. 10, pp. S87-S95, Aug. 2013.
- [7] S. Thorpe, I. Ray, T. Grandison and A. Barbir. 2012. Cloud Log Forensics Metadata Analysis. Computer Software and Applications Conference Workshops (COMPSACW), 2012 IEEE 36th Annual, Izmir, pp. 194-199, 2012.
- [8] S. Thorpe. 2012. An experimental survey towards engaging Trustable Hypervisor log evidence within a cloud forensic environment. International Journal of Computer Science and Information Technology, vol. 4, no. 6, pp. 125-141, Dec. 2012.
- [9] B. Böck, D. Huemer and A. M. Tjoa. 2010. Towards More Trustable Log Files for Digital Forensics by Means of "Trusted Computing", 24th IEEE International Conference on Advanced Information Networking and Applications, Perth, WA, pp. 1020-1027, 2010.
- [10] Z. Zafarullah, F. Anwar and Z. Anwar. 2011. Digital Forensics for Eucalyptus. Frontiers of Information Technology (FIT), Islamabad, pp. 110-116, 2011.
- [11] G. Sibiya, H. Venter, T. Fogwill. 2012. Digital forensic framework for a cloud environment. In Proceedings of the 2012 IST Africa Conference, ISBN: 978-1-905824-34-2, 2012.
- [12] A. Patrascu and V.-V. Patriciu. 2015. Logging for cloud computing forensic systems. International Journal of Computers Communications & Control, vol. 10, no. 2, p. 222, Feb. 2015.
- [13] S. Thorpe, T. Grandison and M. B. Blake. 2014. Cloud computing log forensics-the new frontier. IEEE SOUTHEASTCON 2014, Lexington, KY, pp. 1-4, 2014.
- [14] S. Thorpe, I. Ray, T. Grandison, A. Barbir. 2011. The Virtual Machine Log Auditor. In Proceedings of Information Assurance and Letters (IASL), vol. 2, no. 1, pp 37-43, 2011.
- [15] S. Nakahara and H. Ishimoto. 2010. A study on the requirements of accountable cloud services and log management. Information and Telecommunication Technologies (APSITT), 2010 8th Asia-Pacific Symposium on, Kuching, pp. 1-6, 2010.
- [16] Distributed Management Task Force, Cloud Auditing Data Federation (CADF) - Data Format and Interface Definitions Specification version 1.0.0. Retrieved 29 of May 2016, From: www.dmtf.org/sites/default/files/standards/documents/DSP0262_1.0.0.pdf
- [17] M. Waschke. 2014. Cloud-Specific Standards, Cloud Standards: Agreements That Hold Together Clouds. Apress, Berkeley, CA., 2014, pp. 289-331 . ISBN: 978-1-4302-4111-9