Using Open Source Based Distributed Agents to Perform Digital Investigation in Virtual Environments

Daniel Spiekermann[1], Jörg Keller[2], Tobias Eggendorfer[3]

Abstract: To solve the challenges of digital investigation in virtual environments, we propose the use of distributed agents to ensure valid and continuous network traffic observations in these environments. As one of the most relevant new evolution in information technology, cloud computing demands a flexible and highly dynamic infrastructure, provided by the virtualization of systems, networks and storage. However investigating computer related crime in these environments a necessity. Current techniques like computer or network forensic investigation are not suitable for these environments. The migration of virtual machines changes the environment permanently, thus every running investigation is aggravated. Our approach of open source based, distributed agents is able to implement viable investigations in these virtual environment.

Keywords: network forensic, virtual environment, cloud computing, digital investigation, cyber crime

1 Introduction

The relevance of virtual environments increases with the evolution of cloud computing. Hence a modern datacenter uses dynamic, flexible and highly automated infrastructures to provide the services. Static implementations are not flexible enough, therefore virtual machines (VM) replacing physical servers and virtual networks overlaying the physical network.

The necessity of digital investigation is still existing, even in virtual environments. Current methods of digital investigation in traditional environments are proved and well-known. The analysis of personal computers, servers or network data is done with tools and frameworks, which support the investigator and automate necessary steps. But mostly all tools and techniques of nowadays digital investigations are mainly limited to static systems and fail in virtual environments with their frequent changes. These environments raise different issues which impede the digital investigation of a system of interest (SOI) in law enforcement [SEK15].

[1] FernUniversität in Hagen, daniel.spiekermann@fernuni-hagen.de
[2] FernUniversität in Hagen, joerg.keller@fernuni-hagen.de
[3] HS Weingarten, tobias.eggendorfer@hs-weingarten.de

doi:10.18420/in2017_224

In this paper we present an approach of using open source based, distributed agents to implement digital investigation in these environments. The agents communicate with a central server and collect predefined data inside the environment autonomously. If an agent finds relevant data, the server is noticed and depending on the policy various actions are available. A first implementation of such an autonomous system named ForCon was presented at [SKE17]. ForCon ensures valid and ongoing network forensic investigation in OpenFlow networks in combination with Open vSwitch instances. The expansion of ForCon to gather different data improves computer forensic investigation in virtual networks. The distributed agents monitor the environment and in case of a starting migration of the SOI, the server is informed and as a consequence subsequent steps are initiated. Thus a combination of network and computer forensic investigations supported by distributed agents increases the success of a given examination.

The rest of this paper is structured as follows. In Section 2 we list related work regarding the fields of digital investigation in law enforcement and forensics examination in virtual environments. Background information regarding virtual environments is summarized in Section 3, the field of digital investigation is described in Section 4. The use of distributed agents to improve digital investigation in cloud environments is discussed in Section 5. Section 6 concludes this paper and gives an outlook on our future research.

2 Related Work

Different research has been done in the area of digital investigation and virtual environments. Most of the work is limited to aspects like the analysis of attacks or network security. The special demands of law enforcement are discussed in [AFCF13], [Na01] and [Na04].

Research with a focus on digital investigations in virtual environments are presented in [Eu16] with a focus on lawful interception and retained data or in [DS13] with a first implementation named FROST to perform digital investigation in OpenStack clouds. [FF14] describe the detection of anomalies and identification of the source in software defined networks (SDN). [Ag14] implements tracing in SDN to fulfill digital investigation.

The use of open source in digital investigation is discussed in [Ca03] and in [Ga10] with a view of the possibilities and limitations of open source tools and the possible benefit of closed source tools.

3 Virtual Environment

The architecture of a modern datacenter differs from traditional environments. In the past, single servers were installed to host only a few specific services like webservers, databases or applications. Today virtual machines connected to a virtual networks are deployed to the customer. Hundreds or thousand of VMs are hosted on very powerful servers called compute nodes simultaneously. The management is performed by a virtual machine monitor [SN05].

The use of VMs improves the overall resource usage, increases the flexibility and simplifies the administration, but the maintenance of these VMs is still cumbersome. Aspects like security, isolation and interconnection raise additional tasks to the provider. Common tools and protocols like MPLS or VLAN are able to separate the different VMs, but they are limited in the number of possible networks and fail to provide all requirements of virtual environments. Hence the next step led to the use of virtual networks, which use new network protocols like VXLAN [Ma14], GENEVE [Gr17] and NVGRE [GW15] or new paradigms like SDN or network function virtualization (NFV) [Eu14].

The new protocols use encapsulation to transfer the given network packets and eradicate the limitations of MPLS and VLAN by increasing the number of usable logical networks. The use of SDN provides a higher flexibility by using a central controller and the programmability of the network. NFV converts common hardware devices in software functions, which simplifies the deployment of needed network functions.

4 Digital investigation

Law enforcement agencies perform digital forensic investigation to examine crimes in information technology, often described as cyber crime, by recovering and analyzing information found in digital devices. A common investigation uses three well-defined steps:

- **Saving**
  The relevant data is copied and stored for the subsequent tasks.
- **Analyzing**
  The stored data is analyzed to extract relevant information, and find important data.
- **Presenting**
  The results are processed to a report which lists the performed tasks and provides a comprehensible documentation.

Depending on the target system and the environment, different specialized branches of digital investigation are used. This leads to various possible implementations like computer, network or mobile phone forensics and application-specific areas like database or multimedia forensics.

5 Distributed agents

As described in Section 4, different areas of digital investigation exist, each of them with adapted methods, frameworks and tools. To perform digital investigation in a datacenter, the most relevant fields are computer and network forensic investigation. Especially the network forensic investigation is used by law enforcement agencies to capture the network traffic of a SOI. In combination with computer forensics, the analysis of the collected data is simplified because of the additional information.

But these current methods fail in modern virtual environments. Problems like data access, VM migration, flexibility, legal issues and multitenancy impede all phases of digital investigation [SE16]. One of the most critical issues is the possibility to migrate a VM from one physical host to another. The movement of this SOI raises different challenges to the investigator, which might affect computer as well as network forensics. Capture processes in network forensic investigations are interrupted if the SOI moves to another compute node and subsequently uses another network interface to transmit the network data. Carving techniques to retrieve deleted data will find different file fragments, but the assignment to the SOI gets more complicated in dynamic environments with lots of movements.

ForCon is an implementation for digital investigation in OpenFlow networks, which uses distributed agents to identify a SOI and to prepare the capture process. On the one hand, the so-called *sdn-agents* extract various data out of the connected devices and monitor the involved compute nodes and gather relevant information of changes in the network. If these changes are crucial for the running capture process, ForCon is able to reconfigure the capture process to ensure the ongoing packet capture. On the other hand, special agents named *mirror-agents* establish tunnel connections between the involved sdn-agent and itself to transfer the copied network traffic. The agents communicate with the ForCon and interact autonomously inside the virtual environment.

We propose the use of distributed agents to also improve the analysis of involved compute nodes. These hosting systems might be involved only for a short period of time, but in this interval relevant information might be created and stored on the virtual hard disk. If the VM is moved to another server, these information get lost because this system releases the storage space after the migration is finished. If the storage space is not allocated anymore, the hypervisor assigns it to other VMs. So all information of the SOI stored in this area are lost after the migration. This might occur again and again, and the movement leads to perpetually changing compute nodes and different storage locations. Without knowing the actual storage area the investigator only has the notice of the last compute node, which limits the ability to acquire relevant storage. The information of prior storage pools is stored in different files and devices like the hosting system, a logging device or the cloud controller. The extraction is faced with various issues like different file format, diverging designation of systems and the overflowing number of log entries in a highly dynamic infrastructure, which requires a cumbersome analysis, which can hardly be automated.

The use of ForCon and the implemented monitoring in combination with a subsequent extraction of these information eradicates these problems and improves computer forensic investigations. If an agent informs ForCon about changes inside the network, various information are exchanged. ForCon extracts the relevant information and informs the involved agents with predefined messages. By announcing a special message (*announcement*-message) containing information about the compute node, the investigator is notified and might start additional processes like imaging and carving of the virtual filesystem or live forensic investigation. A timely notification facilitates the extraction of volatile memory, whereby the subsequent analysis might retrieve important information like encryption keys, open network connections and recently used commands [Li14]. Without this monitoring convenient live forensic investigations are hardly feasible.

6 Conclusion

The evolution of modern DC from static implementations to flexible and dynamic virtual environments is still ongoing. Traditional methods of digital investigation in these infrastructures will fail because of the dynamic and flexibility. We presented a new approach of distributed agents acquiring the data of involved compute nodes running a SOI. By expanding the agent-based framework ForCon, which implements network forensic investigation in OpenFlow networks, additional information gathered by the agents eradicate issues of computer forensic investigation.

The agent running on the involved compute node monitors the system and manipulates flows to capture the network traffic of the SOI. By expanding this functionality, the involved agent running on a specific compute node might start imaging or carving processes.

The combination of network and computer forensics improves the entire digital investigation process in law enforcement, requiring forensic valid and sound evidence. Whereas ForCon as a network forensic framework is successfully implemented and tested in different situations, the use of distributed agents to extract additional information still needs to be evaluated and tested in detail, which is a topic for future research. ForCon is using an open source approach which facilitates the development of adapted agents for different vendors and implementations like VMWare, KVM and Hyper-V.

References

[AFCF13] Al Fahdi, M.; Clarke, N.L.; Furnell, S.M.: Challenges to digital forensics: A survey of researchers practitioners attitudes and opinions. In: 2013 Information Security for South Africa. pp. 1–8, Aug 2013.

[Ag14] Agarwal, Kanak; Rozner, Eric; Dixon, Colin; Carter, John: SDN traceroute: tracing SDN forwarding without changing network behavior. In: Proceedings of the Third Workshop on Hot Topics in Software Defined Networking. HotSDN 14. ACM, pp. 145–150, 2014.

[Ca03] Carrier, Brian: Open Source Digital Forensic Tools: The Legal Argument. Technical report, @stake Research Report, September 2003.

[DS13] Dykstra, Josiah; Sherman, Alan T.: Design and implementation of FROST: Digital forensic tools for the OpenStack cloud computing platform. Digital Investigation, 10:87–95, 2013.

[Eu14] European Telecommunications Standards Institute: Network Functions Virtualisation (NFV) - White Paper 3. Technical report, European Telecommunications Standards Institute, 2014. last access: 10.04.2017.

[Eu16] European Telecommunications Standards Institute: Lawful Interception (LI); Cloud/Virtual Services for Lawful Interception (LI) and Retained Data (RD). Technical Report ETSI TR 101 567, European Telecommunications Standards Institute, January 2016.

[FF14] Francois, Jerome; Festor, Olivier: Anomaly traceback using software defined networking. In: 2014 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE, pp. 203–208, 2014.

[Ga10] Garfinkel, Simson L: Digital forensics research: The next 10 years. Digital Investigation, 7:64–73, 2010.

[Gr17] Gross, J.; Sridhar, T.; Garg, P.; Wright, C.; Ganga, I.: Geneve: Generic network virtualization encapsulation. Proposed standard, IETF, 2017.

[GW15] Garg, P.; Wang, Y.: NVGRE: Network Virtualization Using Generic Routing Encapsulation. RFC 7637, IETF, September 2015.

[Li14] Ligh, Michael Hale; Case, Andrew; Levy, Jamie; Walters, Aaron: The Art of Memory Forensics: Detecting Malware and Threats in Windows, Linux, and Mac Memory. Wiley Publishing, 1st edition, 2014.

[Ma14] Mahalingam, M.; Dutt, D.; Duda, K.; Agarwal, P.; Kreeger, L.; Sridhar, T.; Bursell, M.; Wright, C.: Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks. RFC 7348, IETF, August 2014.

[Na01] National Crime Justice: A Guide for First Responders. Electronic Crime Scene Investigation, 187736, 2001.

[Na04] National Institute of Justice: Forensic examination of digital evidence: a guide for law enforcement. NIJ special report. U.S. Deptartment of Justice, 2004.

[SE16] Spiekermann, Daniel; Eggendorfer, Tobias: Challenges of Network Forensic Investigation in Virtual Networks. Journal of Cyber Security and Mobility, 5(2), 2016.

[SEK15] Spiekermann, Daniel; Eggendorfer, Tobias; Keller, Jörg: Using network data to improve digital investigation in cloud computing environments. In: International Conference on High Performance Computing & Simulation (HPCS),. IEEE, pp. 98–105, 2015.

[SKE17] Spiekermann, Daniel; Keller, Jörg; Eggendorfer, Tobias: Network forensic investigation in OpenFlow networks with ForCon. Digital Investigation, 20:66–74, 2017.

[SN05] Smith, James E.; Nair, Ravi: Virtual machines: Versatile platforms for systems and processes. Morgan Kaufmann Publishers, Amsterdam and Boston, 2005.