Using Network Data to Improve Digital Investigation in Cloud Computing Environments

Daniel Spiekermann
FernUniversität Hagen
58084 Hagen, Germany
daniel.spiekermann@fernuni-hagen.de

Tobias Eggendorfer
Hochschule Ravensburg-Weingarten
88250 Weingarten, Germany
tobias.eggendorfer@hs-weingarten.de

Jörg Keller
FernUniversität Hagen
58084 Hagen, Germany
joerg.keller@fernuni-hagen.de

***Abstract*—With the rise of cloud computing environments and the increasingly ubiquitous utilization of its opportunities, the amount of data analysed in a traditional digital forensic examination is increasing significantly, thus increasing the risk to miss evidence. Without adopting new methodology or different approaches investigators are unable to guarantee a valid digital forensic investigation. Due to the large amount of cloud platforms it is hardly feasible to identify them when investigating a computer. Knowing all different services of cloud computing platforms is impossible for a human. The paper therefore proposes to investigate raw network data in order to improve the complete digital investigation process by correlating network and computer forensic parts. We present a new method to analyse network traffic to find information about the usage of cloud specific data. With the possibility to automate this extraction and the comparison with a cloud service knowledge base, the error rate of a forensic investigation is reduced. It also reduces the risk of human errors.**

I. Introduction

A. Cloud Computing

Cloud Computing is a new, rapidly evolving paradigm of information technology. It combines virtualization technologies with modern web architecture like web services or service oriented architectures (SOA) [1] and could be offered at a cost or free of charge and offers different services to users, some are free of charge, some are paid on demand. Nevertheless there are various definitions of cloud computing. We consider the definition by [2] as the most useful.

> *"Cloud computing 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."*

With cloud computing, users profit from storing and working with data and applications without the need to manage these computers. The cloud service provider (CSP) takes care of it. As a consequence, user data is not stored on local harddisks any more, but on the CSP's hardware [3].

Users have access to their files via network connections, no matter whether they are connected with a local-area-network (LAN) or the mobile communications networks like UMTS or LTE. They may use different applications or the web interface, depending on what the CSP offers. So one of the most important features is the connection to the internet which realises the ubiquitous access to the cloud services.

Cloud computing is often defined as Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS) [2].

Another classification of clouds is by the owner of the infrastructure: Public, private, hybrid or community cloud. Private clouds offer their services to a defined group such as all students of a university or all staff. Public clouds offer their resources likes servers, storage or applications to the general public. A hybrid cloud is a mixture of these two, in which a private cloud is used for most services and a public cloud supports load-balancing or high-availability. In a community cloud a group of users with common interest are working together to improve their infrastructure or to increase their return of investment.

B. Forensics

Digital forensics helps investigating crimes involving computers, often described as cyber crime. The amount of data to be forensically examined is increasing and thus impeding an efficient analysis. The traditional cyber forensic approach of analysing locally stored data is not longer appropriate in cloud computing environments.

Forensic analysis can comprise an online phase (e. g. short term monitoring a suspect's network connection to identify cloud usage) and an offline phase (e. g. analysing a suspect's harddisk for artefacts of cloud usage), where the latter phase can be better targeted with information from the former phase. The online phase is only used in network forensic investigation to capture the relevant traffic. Traditional digital forensic investigation mostly consists of just the offline phase by analysing the given hardware.

The ability to perform valid forensic investigation in cloud computing environments is impeded by a lot of problems facing different parts of organizational, technical or jurisdictional fields [4].

- Access to data
  The user data is stored in datacenters all over the world.

The storage environment is designed to react automatically to problems like full hard disks, exceeded thin provisioning environment or hardware failure. In case of this, the storage controller moves the relevant data from one server to another independently. Due to this, even the CSP might not know where the data is actually stored.

- VM-Life-cycle
  Virtual machines (VM) have different states. A *stopped* machine is transferred to the state *starting*, when the user launches the instance. After finishing the boot process, the VM reaches the state *running*. Depending on the user interaction two states are possible: "Stopped" or "Terminated". The latter means the VM is not used anymore, the cloud controller deletes the VM and releases the storage. Because of this highly dynamic environments free storage is quickly overwritten. The loss of relevant information occurs faster than in traditional data center architectures.
- Multitenancy
  In a virtual environment the paradigm of *one server - one application* is not usable any more. Nowadays one physical server hosts tens or hundreds of VM. Because of this, the connected hard disks store data of more than one specific user. With techniques like migration or thin provisioning the problem is further increased. Hence it is impossible to retrieve a specific user's data.
- Chain-of-Custody
  In digital investigation the chain-of-custody (CoC) refers to the correct chronological documentation and describes where the evidence was at a given time and who had access to it. That is done to ensure, evidence is not tampered with during the investigation process.
  In cloud computing evidence is hardly accessible for the investigator. This moves the beginning of the CoC to the CSP, who creates the forensic copy of the storage. The investigator is no longer able to control and monitor the forensic image creation and has to trust the CSP.

This problems impede the process of digital investigation in cloud computing environments, but the lack of software support to detect cloud usage is more serious. A valid detection of cloud computing usage is still not implemented in most of the traditional forensic frameworks, which leads to a obscurity of usage.

- Obscurity of usage
  The number of cloud applications is enormous [5], with new services being offered daily and other services disappearing as quickly. In our study we found users hardly notice whether they use cloud services or not. Therefore asking the suspect on whether cloud technologies were used is not feasible - besides the right to deny self incriminating information, neither is it possible to identify all cloud services available.
- Lack of software support
  On the other hand there is a lack of software support aiming at the forensic process. A lot of network forensic software exists, but none of these products is able to support the investigator to find relevant data. Even if such a forensic framework would exist, techniques like encryption would complicate the process of evaluating cloud data.
  Different solutions (programs or hardware appliances like firewall) offer a recognition of cloud usage, mostly restricted to specific services without the possibility to extend this service list. But all of these perform only real time examination of the network traffic, which is useful for network protection or rejecting unwanted traffic according to company policies. None of these solutions supports a network forensic examiner to investigate captured network data at a later time.

But the inherent use of network connections can be used to improve these facts. We present a new approach to use network data *and* existing network forensic techniques to improve the cloud forensic analysis in cases where cloud usage might have been present.

To validate our results we developed a proof-of-concept (PoC) as a prototype framework to automate the forensic process. This prototype tool uses a knowledge base specifying how use of a certain cloud provider or service can be detected in the network data. Such a knowledge base can be maintained by experts, relieving investigators from this tedious task.

Our studies show how to improve the digital investigation.

The remainder of this article is structured as follows. Section II gives an overview on related research.

Section III explains in detail, which information and protocol fields of the existing network protocols are relevant and usable for forensic investigations. Different services and their communication behaviour are analysed to identify relevant network data of each service in section IV.

Without automation the digital forensic process is time consuming, exhausting, complex and error-prone. The use of different software tools support a more streamlined, faster and more reliable forensic process. Section V demonstrates the PoC to automate the cloud identification based on raw network data.

Section VI concludes and gives an outlook on intended future research.

II. Related Work

A lot of research is done in the wide field of network forensics or cloud computing.

The multitude of different implementations, services or models lead to a lot of problems which are discussed in [6] or [7]. So the need for new approaches and processes to realise forensic investigation in cloud computing environments is inevitable. [8] clarifies the use of new approaches and a revolution in the digital investigation process.

Different research is done to analyse specific cloud environments like *Dropbox* [9] or *Eucalyptus* [10]. An analysis of different cloud storage provider like *dropbox*, *Google Drive* or *Microsoft Skydrive* is discussed in [11].

Network forensics is an established part of digital investigations. Like computer or mobile phone forensics it is necessary for correct and valid results. Characterization of network traffic facilitates network forensic investigators to accurately identify the data traversing the network. [12] discusses the state of the art approaches of network traffic identification. This characterization of network traffic is implemented in different kinds of security appliances like firewalls or intrusion detection systems, which might improve the security of the protected networks. Modules or signatures for *dropbox* [13] or *iCloud* [14] are existing since a long time. But the benefit of these modules is mostly given for real time analysis and is limited to permitting or rejecting the communication, a later investigation of stored network data is impossible. Sometimes a deeper inspection of these traffic is not possible [15].

Digital forensics frameworks improve the process and create reliability and reproduce-ability in the whole forensic investigation. [16] lists different models for digital forensics, but none of them are suitable for cloud forensics. [4] creates a new model called *Cloud-Forensic-Maturity-Model* with special parts dedicated to cloud forensics.

This paper discusses how to extract relevant data without accessing the cloud and how this data can be combined with traditional forensic processes. Some research is done to implement a forensic investigation inside the cloud [17].

III. Methodology

This sections explains our new approach. The main idea is to search for cloud based information in raw network data to conclude the cloud usage. If relevant network data is found, we provide this information to search for certain stored data.

Our approach separates the digital forensic process in two supportive phases, which might offer new indication for the traditional digital forensic investigation process. The digital forensic investigation is not limited to reactive work, especially the network forensic investigation requires an active intervention to collect all network data before the final investigation starts. After capturing all network traffic the collected data is analysed.

A lot of tools exist for this analysis, but none of them offers a presentation of the found services. It is possible to filter relevant network traffic, e. g. with display filter in *wireshark*, but not all of our investigated services offer a simple detection. Our approach uses the combination of different information to conclude the detection of a service.

The result of this analysis might be used to perform a specific search on the suspects hardware.

A. Use of network data

When a cloud application is used, all user data is transmitted via internet. So in principle all data might be intercepted. However since most applications use encrypted protocols, plain text data is hardly available in most cases.

Obviously the unencrypted transferred data is most useful for forensic purpose however the knowledge of the communication might help to improve the digital investigation. Information on who communicates with whom is referred to as *meta-information*.

Based on this meta-information cloud servers among others might be identified and seized. In order to correctly identify used services the beginning of a communication is most useful. With the identification of signatures and communication fingerprints such as accessed servers and ports distinguishing connected services becomes feasible.

B. Evaluating Network Data

Obviously the quantity of transferred and intercepted data is unlimited. In case of intensive usage of internet services it might be very high. Without filtering based on predefined rules network data investigation is time-consuming and the forensic process is still not improved.

The partitioning into abstract layers of the OSI-model [18] was used to determine relevant connection data. The first layer providing relevant information according to connections in the internet is the network layer (layer 3). Additional information might be found on the transport layer (layer 4) and particularly on the application layer (layer 7) [19]. But each information used without the combination with other data is not meaningful enough.

Because of this the relevant protocols are IP on layer 3, TCP and UDP on layer 4 and the application protocol like http or https on layer 7. Most of the analysed services transfer the data via https, so additional examination of the ssl protocol might be necessary.

The straight forward approach to only rely on socket data (ip-address and port number) however this is very likely to fail due to the dynamic nature of the Internet. Thus examination has to rely on more information.

Based on our evaluation of raw network data information in table I is sufficient to identify cloud services:

TABLE I
Relevant network data

Layer	**Protocol**	**Field**
Network	IP	version
Network	IP	protocol
Network	IP	source address
Network	IP	destination address
Transport	TCP / UDP	source port
Transport	TCP / UDP	destination port
Application	HTTP	header informationen
Application	DNS	response-type
Application	DNS	FQDN
Session	SSL	Common Name
without affiliation	-	No. of packets
without affiliation	-	as number

The IP version is important for a correct processing later during automated analysis. The IP protocol field defines the encapsulated layer 4 protocol, which is normally the transmission control protocol (TCP) or the user datagram protocol (UDP) [20].

The header information contains potentially relevant meta-information. If available, this information is used to identify communication partners.

The fully-qualified-domain-name (FQDN) presents the unique hostname with additional information about the network it is connected to [21]. Depending on the response type of the dns packet further information like CNAME[1] can be extracted.

AS numbers define an autonomous system in the internet, which is a collection of connected networks under the control of a single entity. The use of as numbers helps to identify the controlling entity of these servers. So even equivocal services might be assigned to the correct and known provider. This numbers are requested from a local database, which stores a current relation between ip-address and as-number.

All this data is usually unencrypted and therefore easily available through network capture.

The most common methods to capture network traffic are:

- **port-mirror** (formally known as SPAN-Port[2]) which are offered by most of the professional network switches.
- **TAP** denotes a special hardware device which is inserted between two communication partner. All traffic between these passes unimpeded, but all traffic is copied to the so-called monitor-port, enabling a third party to capture the data.
- **Bridge-mode** using an intercepting computer equipped with two network cards acting as a bridge, the traffic passes unimpeded, but might be captured, too.

We decided to use a PC running Linux configured as a bridge to capture traffic. Slightly different ways to capture network traffic, e. g. within a VM, exist, but this scenario eliminates different sources of error. Since network data cannot evade a bridge, this approach guarantees capture of all passing packets.

IV. Identification

The evaluated network data enables a detection of many services, because all of them offer different kind of meta information which might be used to identify them. Our experiments only consider to the mostly used services of the three service models, but an extension with other services is possible. Our approach implements a knowledge base, which is customizable by adding or changing the entries.

A. Services

Dropbox is a SaaS-Application and realize a multi platform synchronisation. After installing the software client the user authenticates this system at the dropbox servers. Till now the service synchronises all files in the defined directory, first to the dropbox storage systems and after that to all connected and authorized clients[3].

Apple iCloud is a SaaS-application too. All Mac OS X systems contain an installed, but still not configured version. After configuring the service, defined applications transfer their data to Apple-server, e. g. calender information or taken photos of the software *photostream*. The iCloud is accessible via the website *www.icloud.com* and allows authorized users to access these information from all over the world[4].

Google App Engine is a PaaS-environment, where users and developers are able to host their tools. Google provides this development environment to facilitate developers a scalable and flexible runtime environment for their applications[5].

Amazon Simple Storage Service (S3) is a storage service. It is a IaaS-environment and offers a filehosting environment to everyone. The files are stored on Amazon servers in a datacenter. Amazon replicates the stored data to other datacenters in the defined regions by itself and guarantees an availability over 99%[6].

Amazon Elastic Compute Cloud (EC2) is a web service that offers cloud hosting environment. These environments allows users to rent virtual computer on which they are able to administrate their own services[7].

B. Analysis

Working in a separated testing environment, we performed defined processes of examination, e. g. upload or download data to the provider, using the software client or the website. Within the testing environment, we used PCs with *Microsoft Windows 7*, *Ubuntu Linux 12.04 LTS* or *Apple MacBook* with *Mac OS X 10.9 Mavericks* installed on it. This variety addresses the most important operating systems, enabling us to examine a deviating behaviour by accessing the different services.

These PCs were connected with a *Cisco 2960* layer2-switch, the access to the internet is offered by a standard vdsl-router. All network data was captured with a interposed software bridge based on a *Debian Linux* machine with two network cards. The captured data was examined with *wireshark*[8] and *tshark* as opensource multiplatform protocol analyser [22].

We use predefined scenarios in a our separated test environment, executing only one service at the same time. This approach guarantees a capture file containing only relevant data.

All of the services indicate some options for identification as shown in table II.

However only one of these fields is not sufficient for a proper identification. Only a combination of these fields obtain a valid identification.

Not all services provide the same amount of usable information. Because of this a valid detection of these services is not guaranteed. The amount of information depends on the kind of service, e. g. whether it is IaaS, PaaS or SaaS.

[1] canonical name
[2] SwitchPortANalyzer
[3] Details can be found at http://www.dropbox.com
[4] Details can be found at http://www.icloud.com
[5] Details can be found at http://appengnine.google.com
[6] Details can be found at http://aws.amazon.com/s3/
[7] Details can be found at http://aws.amazon.com/ec2/
[8] http://www.wireshark.org

TABLE II
IDENTIFICATION OF CLOUD SERVICES

Parameter	Dropbox	iCloud	Google App Engine	EC2	S3
HTTP-Header	x	-	-	-	-
Port number	x	x	x	-	x
FQDN	x	x	x	x	x
CNAME	x	x	x	-	x
x.509-Information	x	x	-	-	-
AS	x	x	x	x	x
Website	x	x	x	x	x

SaaS like *Dropbox* provides the largest amount of relevant information. The fewest information is provided by IaaS like *S3* or *EC2*.

The customization of the service makes this obvious. In SaaS the whole communication with the connected servers and thus the meta-information is mostly the same or differs just a little in case of load balancer or the horizontally scaling of web servers.

In IaaS environments the user is able to configure the offered services and the transferred data like http information or the existence of SSL/TLS. Only meta-information like as-numbers obtain a identification, but the lack of usable data leads to fuzzy results.

Some problems are still existing even with the use of the knowledge base.

Most traffic between the client and the server is transmitted encrypted, mostly by using the SSL/TLS protocol. Depending on the kind of service sometimes unencrypted traffic is submitted, but most of the traffic is transmitted without plain text information.

Additionally in order to avoid any Internet traces anonymizer services might be used. An example being TOR[9] (The Onion Router) hiding the client's IP using encryption and a network of intermediate routers.

Virtual Private Networks (VPN) intended to provide a secure connection over public networks to allow private data transfer by connecting to a dedicated VPN gateway from where the destination host is then contacted.

Due to the nature of these technologies, network analysis based on data transmitted within a network is impossible. However most perpetrators still do not consider these options according to the authors' experience.

V. SYSTEM OVERVIEW

A. Network forensic software

Automation is significant in computer forensic examination. Without automation, the process of digital investigation is time consuming and inefficient. Forensic software is used in digital investigation to extract data, carve for deleted files or analyse a huge amount of data.

[9]Details can be found under http://www.tor.org

Network forensic tools interpret the captured data, dissect each frame according to the transferred protocols and present the decoded information to the investigator. The analysis of cloud specific network data requires additional features. We analysed three of the most common and proven network forensic tools to determine the usability for cloud traffic detection. In detail we checked the following three tools:

- Wireshark
- Wildpackets Omnipeek Enterprise
- Network-Miner

Table III lists the additional features and the results for each of the examined tools.

TABLE III
SUMMARY OF NETWORK FORENSIC TOOLS

	Wireshark	Omnipeek	Network-Miner
Parsing of network streams	x	x	x
Analysis of dns-traffic	x	x	x
Extract HTTP-header information	x	x	x
Extract SSL-Information	x	x	-
List communication partner	x	x	-
Store cloud-based communication	-	-	-
Present cloud-based traffic	-	-	-

As table III summarizes, none of the examined tools supports the entire process to extract cloud based traffic which helps the investigator to identify the use of cloud services. Because of this, we developed a PoC which realizes the wanted features.

B. Design

Our PoC reads captured network data stored in pcap-files. So far there is no option to read pcap-ng or raw-files, but a file conversion is easily done with *editcap*, delivered with the wireshark installation.

It uses the *dpkt framework*[10] for basic identification and stream extraction. *pygeoip*[11] is used for determining the as-numbers. Our prototype consists of about 700 lines of code.

All relevant and dynamic information according to the services are stored in a xml-file. We donate this file as a *knowledge base* of cloud service detection. All listed services can be customized in a flexible manner. If needed, additional services can be added by experts analysing a given service. The investigators may use the new results of these researches without knowing the details of the services. The following listing details exemplary a short piece of the relevant detection

[10]http://code.google.com/p/dpkt/

[11]https://github.com/appliedsec/pygeoip

information of *dropbox* and *Amazon S3*. The information of the other analysed services are deposited in the same way.

Dynamic fields are listed as regular expressions to match different kinds of spellings.

```
<provider name="Dropbox">
<sld>dropbox.com</sld>
<server>
<webserver>
<web>www.dropbox.com</web>
<port>80</port>
<port>443</port>
</webserver>
<webstore>dl-web.dropbox.com</webstore>
<sync>client(\d*|-lb)\.dropbox\.com</sync>
<store>dl-client\d*\.dropbox\.com</store>
</server>
<as>AS19679</as>
```

The Amazon S3 service did not offer a native sync client or storage server, so the appropriate entry is empty.

```
<provider name="Amazon S3">
<sld>amazonaws.com</sld>
<port>80</port>
<port>443</port>
<server>
<web>console.aws.amazon.com/s3</web>
<sync></sync>
<store></store>
</server>
<cname>s3-\d-w\.amazonaws\.com</cname>
<as>AS16509</as>
<uri>authorization AWS</uri>
</provider>
```

The python script is divided in three main parts, which are used in sequence:

- Parsing the pcap-file
- Analysis according to services
- Presentation of the results

C. Parsing

The parsing engine examines every frame captured in the file once. Each frame is dissected and the extracted data is stored for a later analysis. The parser stores the data, any further examination is done by different analysing sub-routines of the PoC to improve speed and flexibility of the examination process.

After parsing the pcap-file the next steps are depending on the found data.

D. Service rating

The parser stores all relevant data according to their purpose in different data structures like lists or dictionaries. By choosing different data structures the time of sorting and searching data might be minimized. The analyser uses these data structures to compare the information with the user defined knowledge base.

The analysing part is most time-consuming because of the expensive string comparisons. The bigger the pcap-file the longer the examination process will run. But this problem is already existing in traditional forensic investigation. The time needed for a valid investigation (which includes mostly a full one-to-one copy of a hard disk) is getting longer depending on the size of the volume. Each service listed in the xml-file is analysed successively.

To avoid false-positives we implement a blacklist with fqdn-entries. This blacklist is checked to ignore hostnames like *www.google.com*. The PoC checks all valid hostnames with regular expressions and compares the as-number to the given number in the xml-file. Some hostnames (especially in case of IaaS-detection) match the regular expression and the as-number, too. The blacklist helps to eliminate those entries and keeps the final result more meaningful and precise.

E. Presentation

No examination process is usable without a presentation of the results. This presentation should be understandable to persons without technical know-how or special training, e. g. judges or lawyers. Because of this our PoC offers a short form of result presentation which is presented below for the detection of *dropbox* usage.

```
-----------------Dropbox----------------
DNS-data of dropbox.com found
Providerbased traffic found
Datatransfer found
No connection to website detected
Syncinformation found
Store of data detected
*** Usage of Dropbox detected ***
```

The main statement is written in the last line. With this information the forensic investigator is able to examine the other evidences to find specific dropbox-files or to use special tools for investigating dropbox-data like Dropbox Decryptor offered by Magnet Forensics[12].

For well trained person the verbose mode of our script enables a deeper inspection of the relevant parts which leads to a more chatty output as listed below:

```
-----------------Dropbox----------------
CNAME Information found
AS True
DNS-data of dropbox.com found
Providerbased traffic found
Datatransfer to FQDN found
No connection to website detected
Data in HTTP-Header found
Syncinformation found
-------
```

[12]http://www.magnetforensics.com/dropbox-decryptor-a-free-digital-forensics-tool/

```
Name: dl-client615.dropbox.com /
 IP-address : 50.16.224.172
receiving-rate 66.30% /
 transmitting-rate 33.70%  on port 443
receiving data: 4793 Byte /
transmitting data: 2436 Byte
probably download
```

The investigator gets additional information about the extracted data, e. g. http or fqdn. At the end the amount of transferred data is listed, in combination with the fqdn, ip-address and port a suggestion is made for up- or downloading files.

The flexibility of IaaS complicates a valid investigation, because the transferred meta information is not completely predictable. But our approach uses all of the transferred meta information to present a result according to the given pcap-file.

```
-----------------Amazon S3-----------------
CNAME Information found
AS True
DNS-data of amazonaws.com found
Providerbased traffic found
No datatransfer to FQDN found
No connection to website detected
No data in HTTP-Header found
.................
Found 'raw' data belonging to the \
   autonomous system AS16509
IP-address : 72.21.195.33
receiving-rate 64.71% /
  transmitting-rate 35.29% on port 80
receiving data: 805 /
  transmitting data: 439
probably download
```

VI. Conclusion

In this paper we presented a new approach for digital investigation in cloud computing. We analysed network data by extracting different information to identify the use of cloud services. This chapter describes the results of our new approach and present our future research.

A. Summary

The digital investigation process in cloud forensic environment is complex and error-prone. Lots of problems base on global, highly dynamic and frequently changing provision of cloud services. Traditional digital forensic investigations are not able to evade this problems.

We presented a new approach to use network data to identify cloud based traffic. In combination with traditional digital forensic processes an investigator is able to use this identification to focus on the identified services. Our approach proved, that an analysis of network data offers enough information, even if a lot of traffic is encrypted. We demonstrated that the meta-information contain enough relevant information to clarify a usage of cloud computing services. The identification depends on the kind of service (IaaS, PaaS, SaaS), a higher customization makes the identification more inaccurate.

Our prototype framework uses a knowledge-base which characterizes the communication of each service. This knowledge-base might be customized by experts to improve the number of identifiable services.

In 2012 [23] pointed out that new methodologies, techniques and tools are needed for the challenges of digital investigation in cloud computing:

> *"Incident response and computer forensics in a cloud environment require fundamentally different tools, techniques, and training to assess a situation and capture appropriate evidence ..."*

Cloud computing is an increasingly growing field, which entail new challenges for forensic experts. As described above, the problems concern different aspects of the forensic process like jurisdiction or organisational parts.

Our new approach improves the cloud forensic investigation. A big problem nowadays is the discovery of data belonging to cloud services. Because of the great amount of CSP's and offers, a forensic investigator is overwhelmed with the detection of cloud data which leads to a low detection rate.

Combined with network forensic methods these low detection rate might be drastically increased. First attempts to use our script to improve the detection of cloud service data were made in practice. In different cases the existence of cloud data was demonstrated, which cause the investigator to search for this special service and its stored data. Without our approach, the detection of these services would have failed. Without the captured network data, the detection would have failed, too. So in practice all network data needs to be recorded, which depending on the legislation might require a search warrant..

The knowledge about the use of cloud services enables a deeper inspection of the cloud service itself, again in accordance with the respective legal system.

B. Future Work

Cloud forensic still requires a lot of research. Our new approach presents the scope of network forensic techniques in combination with traditional digital forensic processes. Our approach focusses on the connection between client and cloud server providing a good starting point for further research to improve digital forensic investigation in cloud environments.

Network virtualization makes the network forensic process inside the cloud environment more complex, because none of the traditional techniques seem to be successful. Virtual network cards impede the data recording, new network protocols like VXLAN complicate the analysis.

Our focus in future work is to analyse this difficulties and improve the whole forensic investigation process, including all network data arising in the cloud environment.

References

[1] C. Baun, M. Kunze, J. Nimis, and S. Tai, *Cloud Computing - Web-Based Dynamic IT Services.* Springer, 2011.

[2] P. Mell and T. Grance, "The nist definition of cloud computing," *National Institute for Standards and Technology*, vol. Special Publication 800-145, p. 7, 2011.

[3] M. Taylor, J. Haggerty, D. Gresty, and D. Lamb, "Forensic investigation of cloud computing systems," *Network Security*, vol. 2011, no. 3, pp. 4–10, 2011.

[4] K. Ruan and J. Carthy, "Cloud forensic maturity model," in *International Conference on Digital Forensics & Cyber Crime*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 114. Springer, 2012, pp. 22–41.

[5] A. Anjomshoaa and A. M. Tjoa, "How the cloud computing paradigm could shape the future of enterprise information processing," in *Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services*. ACM, 2011, pp. 7–10.

[6] S. Zawoad and R. Hasan, "Cloud forensics: A meta-study of challenges, approaches, and open problems," *Computing Research Repository*, vol. abs/1302.6312, 2013.

[7] B. Grobauer and T. Schreck, "Towards incident handling in the cloud: challenges and approaches," in *Cloud Computing Security Workshop*, A. Perrig and R. Sion, Eds. ACM, 2010, pp. 77–86.

[8] G. Grispos, T. Storer, and W. B. Glisson, "Calm before the storm: The challenges of cloud computing in digital forensics." *International Journal of Digital Crime and Forensics*, vol. 4, no. 2, pp. 28–48, 2012.

[9] I. Drago, M. Mellia, R. Sadre, A. Sperotto, A. Pras, and M. Munafo, "Inside dropbox: Understanding personal cloud storage services," *Internet Measurement Conference*, 2012.

[10] Zafarullah, F. Anwar, and Z. Anwar, "Digital forensics for eucalyptus," in *Frontiers of Information Technology*. IEEE Computer Society, 2011, pp. 110–116.

[11] D. Quick, "Cloud storage forensic analysis," Masterthesis, University of South Australia, October 2012.

[12] A. Tongaonkar, R. Keralapura, and A. Nucci, "Challenges in network application identification," in *Proceedings of the 5th USENIX Conference on Large-Scale Exploits and Emergent Threats, ser. LEET*, vol. 12, 2012, pp. 1–1.

[13] Cisco Security, "IPS-signatures: Dropbox file sharing client," October 2011, last access: 12.05.2015. [Online]. Available: http://tools.cisco.com/security/center/viewIpsSignature.x?signatureId =38686&signatureSubId=0&softwareVersion=6.0&releaseVersion=S604

[14] Cisco Security, "IPS-signatures: Apple icloud traffic," November 2012, last access: 12.05.2015. [Online]. Available: http://tools.cisco.com/security/center/viewIpsSignature.x?signatureId =1570&signatureSubId=0&softwareVersion=6.0&releaseVersion=S682

[15] Palo Alto Networks, "List of applications excluded from ssl decryption," February 2015, last access:12.05.2015. [Online]. Available: https://live.paloaltonetworks.com/docs/DOC-1423

[16] M. Pollitt, "An ad hoc review of digital forensic models," in *Systematic Approaches to Digital Forensic Engineering*. IEEE Computer Society, 2007, pp. 43–54.

[17] J. Dykstra and A. T. Sherman, "Design and implementation of frost: Digital forensic tools for the openstack cloud computing platform," *Digital Investigation*, vol. 10, pp. S87–S95, 2013.

[18] H. Zimmermann, "Osi - reference model-the iso model of architecture for open systems interconnection," *IEEE Transactions on Communications*, 1980.

[19] R. W. Stevens, *TCP/IP Illustrated*. Addison Wesley, 1994.

[20] C. Hunt, *TCP/IP Network Administration*. O'Reilly, 1994.

[21] C. Liu, P. Albitz, and M. Loukides, *DNS and BIND*, 3rd ed. Sebastopol, California: O'Reilly, September 1998.

[22] G. Combs, "About wireshark," Wireshark.org, 02 2015, last access: 17.02.2015. [Online]. Available: http://www.wireshark.org/about.html

[23] M. Iorga, "Challenging security requirements for us government cloud computing adoption," November 2012, last access: 18.02.2015. [Online]. Available: http://www.nist.gov/customcf/get_pdf.cfm?pub_id=912695