Digital Forensics Challenges to Big Data in the Cloud

Xiaohua Feng  
Department of Computer Science and Technology  
University of Bedfordshire  
Luton, UK  
Xiaohua.feng@beds.ac.uk

Yuping Zhao  
Microwave Wireless Telecommunications Research Lab  
School of Electronics Engineering and Computer Science  
Peking University, Beijing, China  
yuping.zhao@pku.edu.cn

Abstract — As a new research area, Digital Forensics is a subject in a rapid development society. Cyber security for Big Data in the Cloud is getting attention more than ever. Computing breach requires digital forensics to seize the digital evidence to locate who done it and what has been done maliciously and possible risk/damage assessing what loss could leads to. In particular, for Big Data attack cases, Digital Forensics has been facing even more challenge than original digital breach investigations.

Nowadays, Big Data due to its characteristics of three “V”s (Volume, Velocity, and Variety), they are either synchronized with Cloud (Such as smart phone) or stored on the Cloud, in order to sort out the storage capacity etc. problems, which made Digital Forensics investigation even more difficult. The Big Data-Digital Forensics issue for Cloud is difficult due to some issues. One of them is physically identify specific wanted device. Data are distributed in the cloud, customer or the digital forensics practitioner cannot have a fully access control like the traditional investigation does.

The Smart City technique is making use of ICT (information communications technology) to collecting, detecting, analysing and integrating the key information data of core systems in running the cities. Meantime, the control is making intelligent responses to different requirements that include daily livelihood, PII (Personally identifiable information) security, environmental protection, public safety, industrial and commercial activities and city services. The Smart City data are Big Data, collected and gathered by the IoT (Internet of Things).

This paper has summerised our review on the trends of Digital Forensics served for Big Data. The evidence acquisition challenge is discussed. A case study of a Smart City project with the IoT collected services Big data which are stored at the cloud computing environment is represented. The techniques can be generalised to other Big Data in the Cloud environment.

Keywords – Big data, Cloud Computing, Smart City system, IoT, Cyber security and Personally identifiable information (PII) data protection, Personally Information against cyberstalking

I. INTRODUCTION

Digital Forensics is important in cyber security. Evidence data is crucial for Digital Forensics. Big data evidence collecting is challenging in the cloud.

While Big Data is defined as a large collection of data sets which is too complex to be processed by people’s hand. Instead, it needs database management tools to be handled. Big Data is a science/technology which analyses huge amount of data in order to find out the rules, collect valuable opinions and predict complex problems. It maybe involves capture, storage, search, sharing, transfer, analysis and visualization.

Big Data includes Volume, Velocity, and Variety (Veracity); especially in the Cloud environment these three “V”s determine their characteristics.

Volume is the amount of data, usually when data capacity reaches up to Terrabyte (1024 GB) and Pettabyte (1024 TB, 1,048,576 GB).

Velocity is the speed of collecting, processing and using data. Big Data Security often needs speedy processing or even real time processing data.

Velocity becomes a vital measure of big data usage, especially when real time data process is required. Variety is the type of big data involved. In current era, Big Data usages include text, audio, video and multimedia types of data format.

Currently, there is large volume of data available, which brings challenges for digital investigation. For example, challenges in data collection, data analysis, data recovery and many more (Harshish and Feng, 2011). The problem solving is complicated and is still in progress. (Alessandro, 2015)

At the Smart City project of the research, Big data are collected by the IoT automatically, which is a system combined information system communications technology which makes use of radio frequency identification and
Many digital Forensics evidence acquisition issues are depending on the case are related to the CSP (Cloud Service Provider)'s support or co-operation. Zawoad (2013) has listed some updated cloud issues. The distributed data centre may cross several national borders in storage. That might make the Chain of Custody (CoC) as an important part of the document extremely difficult. When Big data in the cloud, the investigator officer has no physical control of device, the only possibility is rely on the CSP. If the related CSP is not technically competent or without to be trained sound forensically, there is no guarantee that the audit trail could be put into the Chain of Custody forms and could be completed appropriately for the future testimony.

Multi-state legal issue could be affected to obtain permissions for authentication of evidence search as well. For instance, if the case related data centre located at a country which does not have a Data Protection Act, trying to get the authorised data will be difficult.

A summary of Big Data digital forensics challenges in the cloud is reviewed and the consequence is as Table 1 shown; where the issues marked a yes (a “Y”) shown problem outstanding; means further development work is still required currently; while the issues have a solution has been marked a no (i.e., a “N”).

IV A CASE STUDY

Since 2014, a world-leading project in autonomous transportation systems and intelligent mobility (Sant, 2015) has started. The Milton Keynes Smart City project focuses on the application of the next-generation of information technology to all walks of life, thereby embedding sensors and equipment to vehicles on roads and railways, bridges and tunnels, mobile communications systems, and others in every corner of the place, thereby forming the internet of things through the internet. This will enable us integrate the Internet of things (IoTS) through powerful computer clusters and cloud computing. This will enable people to manage productivity and life more meticulously and in a more dynamic manner, leading to a state of global intelligence.

It is a typical Big Data collected by the IoTS and stored in the computer cloud. Here, we will discuss about Big Data in Smart City Project where IoT and raspberry pies gathered information; i.e., Big Data in the cloud and their access security issues emerged, as well as the Digital Forensics cloud challenge solution with the impact on PII (ISO, 2014).

In this case study, the smartphone as an IoTS terminal device to collect customer’s information data, the data volume of which is consistently increasing; i.e., Big Data formed. Then the substantial various data are transmitted to the cloud service provider to come up as Big Data in the cloud. The Smart City control centre according to these data to provide local information and suggest the best service, in order to make the customer has a very enjoyable stay at Milton Keynes. Therefore, Big information and interactive data generated. And in order to get cyber security, data acquisition, processing and storage become the consequence correspondingly.
In this case study, the smartphone as an IoT terminal device to collect customer’s information data, the data volume of which is consistently increasing; i.e., Big Data formed. Then the substantial various data are transmitted to the cloud service provider to come up as Big Data in the cloud. The Smart City control centre according to these data to provide local information and suggest the best service, in order to make the customer has a very enjoyable stay at Milton Keynes.

However, if these Big data are breached, the provided information could be false, which maybe leads to serious crime. Then the digital forensics practitioner will walk in and acquire evidence to locate the suspect.

Nevertheless, in the cloud, Big Data of the Smart City are not the only data exist in that cloud. It might be distributed at several locations and at different sectors of the storage. If these data are hacked, the driverless vehicle is crashed. It would be extremely difficult to get a warrant, visit the crime scene, acquire the possible evidence, image the Big Data, filling-in the CoC (Chain of Custody) forms, preserve the acquired evidence (as Table 1 shown). Then examine recovered data and work out the analysis to report to the law enforcement.

The investigation officer needs to bear in mind; he does not own the cloud utility, no matter SaaS, PaaS or IaaS as the following table shown.

Many tasks need to be done by the CSP’s help. Even under the CSP’s support not all the investigation works could be complete as an individual digital/computer system experienced (Feng, 2015). The SaaS is the best developed service. However, apart from access control, authorized investigation officer cannot control the applications, data load, operating systems, servers and network system.

IaaS is the most closet to hardware scenario when a digital forensics cloud investigation is carried out. However, an IaaS user cannot get control of the servers and network system of the IaaS fully for the authorized investigation officer.

### V. DISCUSSIONS

Due to the volume issue, the investigation officer require an adequate bandwidth to image the virtual machine of the Big data in the cloud.

Ruan (2015) did a survey on cloud forensic capacity, the results shown, the majority lies on CSPs. With support from CSP, a read only API provided by the case relevant CSP for network, process and access logs to the representative of the customer to acquire some data by read only. Marty (2011) proposed an Ajax library for logging checking or imply log management in the cloud by the CSPs.

Hegarty et al (2011) suggested for each time uploading/downloading, checking data integrity as one of the problem solving solution; if anything suspicious going on, isolate the cloud immediately. Nevertheless it seems to complicate the cloud activities.

Another problem is potential multi location raised multi state-laws. In order to specify CSPs’ responsibility in the cloud cyber security as well as their role in a digital forensic investigation, an updated Service Level Agreement (SLA) needs to be published and put into operation. Globally collaboration is required, all the states laws should apply.

---

<table>
<thead>
<tr>
<th>Digital forensics challenges to Big Data in the Cloud</th>
<th>Cloud services</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Access control/accessibility</td>
<td>SaaS, PaaS, IaaS</td>
<td>Logs</td>
</tr>
<tr>
<td>CSP dependent</td>
<td>Y, Y, Y</td>
<td></td>
</tr>
<tr>
<td>Volatile data</td>
<td>Y, N, N</td>
<td></td>
</tr>
<tr>
<td>Trustiness</td>
<td>Y, Y, Y</td>
<td></td>
</tr>
<tr>
<td>Bandwidth</td>
<td>Y, N, N</td>
<td></td>
</tr>
<tr>
<td>Multi-tenancy</td>
<td>Y, Y, Y</td>
<td></td>
</tr>
<tr>
<td>Distributed logs</td>
<td>Y, Y, Y</td>
<td></td>
</tr>
<tr>
<td>Volatility of logs</td>
<td>Y, N, N</td>
<td></td>
</tr>
<tr>
<td>Logs in multiple tiers and layers</td>
<td>Y, Y, Y</td>
<td></td>
</tr>
<tr>
<td>Logs accessibility</td>
<td>Y, Y, Y</td>
<td></td>
</tr>
<tr>
<td>Logs lack of critical information</td>
<td>Y, Y, Y</td>
<td></td>
</tr>
<tr>
<td>Chain of Custody</td>
<td>Y, Y, Y</td>
<td></td>
</tr>
<tr>
<td>Issues on existing forensic tools</td>
<td>Y, Y, Y</td>
<td></td>
</tr>
<tr>
<td>Crime scene recognition</td>
<td>Y, Y, Y</td>
<td></td>
</tr>
<tr>
<td>Crime scene reconstruction</td>
<td>Y, Y, N</td>
<td></td>
</tr>
<tr>
<td>Multi-state laws</td>
<td>Y, Y, Y</td>
<td></td>
</tr>
<tr>
<td>Report</td>
<td>Y, Y, Y</td>
<td></td>
</tr>
<tr>
<td>Compliance</td>
<td>Y, Y, N</td>
<td></td>
</tr>
<tr>
<td>Presentation</td>
<td>Y, Y, Y</td>
<td></td>
</tr>
<tr>
<td>Combination issue</td>
<td>Y, Y, Y</td>
<td></td>
</tr>
<tr>
<td>Integrity</td>
<td>Y, Y, Y</td>
<td></td>
</tr>
<tr>
<td>Warrant</td>
<td>Y, Y, Y</td>
<td></td>
</tr>
<tr>
<td>Localization</td>
<td>Y, Y, Y</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Digital Forensics Challenges Big Data in the Cloud

---

### VI. EVALUATION AND CONCLUSIONS

In this paper, we have explored the main Digital Forensics challenges about the Big Data with Cloud Computing PII in the cloud services. (Rezendes, 2015) (ISO, 2014). Updated impact has been discussed and problem solving solutions have been explored and critically analyzed. A serial of design and implementation have been carried out and there is still new development going on at the NCCR (National Centre of Cyberstalking Research) Institute, University of Bedfordshire, currently.

When the recent development finishes at the end of June, we can do a few decision making, according to further experiment results comparison and analysis (Feng, 2016), to recommend a more appropriate approach to sort out the Big data at Cloud Computing service environment issues for digital forensics investigations.
This digital forensics challenges with Big Data is crucial. Almost all the traditional way of investigation is not appropriate any longer, as cloud users lost control on many aspects. Here, a problem solving research with Big Data in the Cloud environment is significant for the upcoming cyber security applications in data science, such as the Smart City project. An impact on future cloud information governance, risk management, and compliance, trustworthiness (Liu, 2014) as well as human factor in Big Data cyber security would play an important role.

In particular, in Smart city projects, PII could be a cutting-edge element for citizen’s trustiness (Petit, 2015) (Hoppe, 2008) and leads the consequence on the development and applications (Feng, 2015).

Since being created in 2004, European Network and Information Security Agency (ENISA) has done many works for the EU states. The Big Data and cloud development Big Data & Smart Sustainable Society Workshop -2016 pushes ENISA to produce more guideline and regulations to protect Big Data in the cloud. To date, we have investigated thoroughly about the outstanding issues, but there is still no perfect solution to sort all of these. There are more adventure work needed.

VII. SUGGESTED FUTURE DEVELOPMENTS

Upon some of the Big Data digital forensics challenges we have summarized today, there are plenty of development need to be carried out to compromise the best trade-off between cloud application services and digital forensics in the Big Data cyber security; such as, analyze social media Big Data for cyber psychology issues; or analyze Big Data in the cloud for national citizen’s health or sports activity record, data service for centralized national healthcare as well.

To date, there is still neither an appropriate professional digital forensic toolkit to carry out computing cloud investigations, nor for the Big Data cases. Guidance software and their peers need to speed up their developments.

Nowadays application research on cyber security field is getting more and more attention. Cyber psychology is one of the newly explorations. Cyber psychology is research on anything human psychology with Cyber technology related.

As a new cross-subject area, its aim and objectives are using the conventional psychology view to analyze the newly developed cyber techniques. Then provide user guide to make use of cyber resource and worked out efficient network management. Currently University of Bedfordshire NCCR is working on social media chat messages, which is another Big Data application. Its security is not only in technical side, but also in social psychology category.

So, using digital forensics technique to acquire social network messages and analyze the content becomes another Big Data challenge.

With Big Data and Cloud computing technology development, digital forensics will face more and more challenges in the near future. There are plenty of exploration for us to discover and work out the reasonable problem solving solutions.

REFERENCES


Feng X. and Zhao Y. (2016): Cyber Security Data Mining with the Mobile Internet of Things (IoT) -- Personally Identifiable Information on Smart Technologies and proceeding of the DaMiS 2016 -- International Workshop on Data Mining on Internet of Things Systems, U.K. 2016


Feng, X (2011) “Computer Law in UK”, UCC Data Retriever, Digital Library Workshop, Ireland


IoTBDH-2017


