Forensic Challenges on Multimedia Analytics, Big Data and the Internet of Things

Zeno Geradts
Netherlands Forensic Institute, Ministry of Security and Justice,
Laan van Ypenburg 6, 2497 GB, Den Haag, The Netherlands
z.geradts@nfi.minvenj.nl
Institute of Informatics and CLHC, University of Amsterdam, Amsterdam, The Netherlands
z.j.m.h.geradts@uva.nl

Keywords: Digital Evidence, IoT, Forensic Data Science, Multimedia, Video Analytics, Multimodal, Deep Learning.

Abstract: The speed of change in the digital world is challenging for forensic investigations. New devices are developed rapidly and the Internet of Things is also emerging. Getting access to devices is getting more complicated due to stronger encryption. The other issue is that the amount of multimedia data is expanding rapidly and finding relevant evidence is often a challenge. Several challenges can be handled by developing big data analysis platforms that are flexible in incorporating new methods and using artificial intelligence as well as deep learning. Since evidence must be used in court, the validation of the results is important to explain the possibilities and limitations of the forensic analysis.

1 INTRODUCTION

The field of digital forensics is transforming due to the ever growing computing power, bandwidth and the many devices that are developed for consumer use. The devices range from computers to smart phones, as well as storage devices, but also in cars and medical devices. In practice we see that most devices nowadays contain a chip and have digital information in them which could be of interest to forensic use. The Internet of Things is often discussed as a new possibility in forensic science

Nowadays we see also the amount of wearables grow and information on location and activity are stored online and on the device. After a crime happened, the information on these devices can be helpful as evidence in court.

In this paper we handle the different aspects in a forensic examination as well as the questions in court that are involved and he different research questions that are available.

Several guidelines exist in this field for instance from ENFSI (Geradts 2011), SWGDE(Sanders 2004, Casey 2011) and ASTM, and where useful these can be applied.

2 REPAIR, EXTRACT, ORGANIZE AND INTERPRET THE EVIDENCE

In digital evidence there are several challenges for a forensic case. (Oparnica 2016) First sometimes a digital device has been repaired, if it has been burned. The challenge is to extract the digital information from it. These efforts are more related to hardware investigation,

After this the analysis of the formats that are extracted are important to consider. From these files' information can be collected of data stored or processed by the user for instance whatsapp messages that someone wrote.

The final part is the interpretation of the evidence. A question asked by the court for instance did the suspect write this email or did someone else do this?

2.1 Repair

In forensic casework sometimes devices are damaged due to fire or water. In these cases it can be necessary to do a chip off extraction and sometimes a transplantation on another device.

Geradts Z.
Forensic Challenges on Multimedia Analytics, Big Data and the Internet of Things.
DOI: 10.5220/0007239400000000
In *Proceedings of the 15th International Joint Conference on e-Business and Telecommunications (ICETE 2018)*, pages 5-8
ISBN: 978-989-758-319-3

Also, a file can be unplayable since it has been partly erased, the field of recovery will try to make it readable again. An example is video recovery, where for example an open source tool defraser has been developed.(Gloe, Fischer and Kirchner 2014).

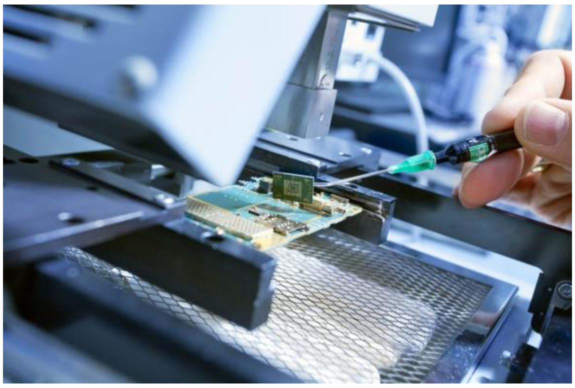

Figure 1: Chip off of a mobile phone at the Netherlands Forensic Institute (free of rights NFI).

2.2 Extract

The information must be extracted consequently. This can be a raw format which has to be interpreted later. For the extraction the verification if the information has been extracted correctly should be available. Also, the extraction should be complete, and if there are issues with partly deleted data it is important to make the user aware of this.(Souvignet 2010).

2.3 Make Data Readable

After extraction the data should be made readable. Knowledge on file properties should be handled. For instance one might first decrypt the data by using a brute force dictionary attack. Nowadays encryption is stronger, so it is more complicated to get access to the data.(2016 FBI). Although, sometimes it is easier to get access by court order from the cloud service that makes a backup of the data of the smartphone (Walden Ian 2013).

Furthermore, if a file is read out, knowledge on the format should be handled. For instance if there is a mp4-file with H264 CODEC, one should have an appropriate player for the file (Kamenicky Jan et al. 2016). Also, the presentation of the data to the user in an understandable form is important.

2.4 Organize the Data

Most devices contain many different files, and log files etc which might be of use. Also, date and time stamps of the files could be considered. Since the amount of data is huge, is necessary to summarize the data for later analysis. For this it is necessary to have some knowledge on the case and what type of questions are asked by the investigator or the court.

2.5 Interpretation

Finally the interpretation task is most challenging and is often discussed in court. For instance if we have a phone with location information, the question might was the suspect at this location or somewhere else. Was the evidence planted on the device or not, has there been tampered with the evidence.

3 BIG DATA AND MULTIMEDIA

The amount of video and other data that have to be investigated is in most cases several terabytes. This means that either a triage is necessary before analyzing all the evidence, or that the data needs to be summarized.

The current methods for video summarization can help to search through video(Snoek and Worring 2007, Koppen and Worring 2009) information faster. Interpretation of what happens in a video is still difficult, as well as to detect when and what happened during a crime. Multiple interpretations are possible and should be considered (Habibian, Mensink and Snoek 2017).

3.1 Deep Learning

For the analysis of video we see that deep learning has made much progress and for good quality photographs it equals or wins from the human vision system.(Sudars 2017).

Much Research is conducted in this field, and one tries to train systems with example data. Also, with computer systems it is important to know that the training sets can create bias of the system due to the nature of this set. For example if face comparison systems are mostly trained with white Caucasian males, the system will perform less on other persons.(Tripathi 2017).

3.2 Multimodal Forensics

In video there is often has a combination of biometric and image properties. In Video one might find a face, clothing, gait, audio and other features of the body that might relate to the suspect.

When multiple cameras are setup for instance in an airport, it is possible to track the persons with these features. From a forensic perspective the likelihood ratio (LR) is used and can also be used to combine evidence (Tripathi 2017, Haraksim 2014). Current research is focused on how to estimate the LR and how using this in practice on video. In court one should be aware of possible pitfalls when using this. (Dawid 2017, Morrison 2017)

Also the current research in facial biometrics is aware of issues with facial recognition system as well as experts. The combination of experts and AI improves the quality of the results (Phillips P Jonathon et al. 2018).

4 INTERNET OF THINGS

With the internet of things it is thought that all devices of a person are connected to the internet and communicate to each other. For instance the fridge might order milk automatically if needed, Since we have ipv6 the idea is that every device or object has it own ip-address (An Overview of IPv6 2012). For lawful data interception of the Internet of Things this is ideal (Layer 3 Connectivity: IPv6 Technologies for the IoT 2013) (Dihulia and Farooqui 2017), however for privacy protection one should protect the privacy of the user, and might consider privacy enhancing technologies(Lee 2015).

The idea of all sensors around persons also deviating events including a murder can be detected, and more digital evidence can be collected after the crime happened (Lund 2014). Ideally of course the system would warn the police that a crime is going to happen based on prediction models, to prevent the crime from happening at all(Schaefer and Mazerolle 2017)

With the internet of things it is expected that data is only temporary available due to the high amount. One might think of a driverless car that processes 2000 Terbytes per year, however only limited data is stored. For this reason live forensic investigation should be researched further.

5 ANTI FORENSICS

Since people become more aware of digital traces, it is also for privacy purposes necessary to wipe the information or alter the information to mislead investigation(Moon 2013, Moon 2015).

It is important in forensic casework to determine if anti forensics tools have been used in a crime.

6 CONCLUSION AND DISCUSSION

The field of digital evidence and forensic big data analysis has grown quick in the last decades.

In the past most of the evidence was not encrypted and easy to collect. Nowadays strong encryption is used on the mobile phones, and the cost to extract the information is higher.

Cloud storage still gives opportunities of accessing the data by law enforcement. The real challenge is processing multiple data source nearly real time, and also use predictive tools to prevent crime from happening. A good balance between privacy and predictive policing is necessary if those methods become feasible.

For digital investigation it is necessary in most cases to do a triage of the information available, since it is not feasible and economical to analyze all. Digital forensic software can help in the analysis of the traces, though one should know anti forensic methods. Since software and hardware change, it is important to cooperate internationally and do continuous validation tests and research on new development.

REFERENCES

An Overview of IPv6. 2012. *IN: Linear and Nonlinear Video and TV Applications: Using IPv6 and IPv6 Multicast.* John Wiley & Sons, Inc., pp. 45–94.

Casey, E. 2011. *Digital evidence and computer crime: Forensic science, computers, and the internet.* Academic press.

Dawid, A.P. 2017. Forensic likelihood ratio: Statistical problems and pitfalls. *Science & justice : journal of the Forensic Science Society,* 57(1), pp.73–75.

Dihulia, S. and Farooqui, T. 2017. ITCP based Security Enhancement for IoT Devices in IPV6 Protocol. *International Journal of Computer Applications,* 178(2).

Geradts, Z. 2011. ENFSI Forensic IT Working group. *Digital Investigation,* 8(2).

Gloe, T., Fischer, A. and Kirchner, M. 2014. Forensic analysis of video file formats. *Digital Investigation,* 11.

Habibian, A., Mensink, T. and Snoek, C.G.M. 2017. Video2vec Embeddings Recognize Events When Examples Are Scarce. *IEEE transactions on pattern* *analysis and machine intelligence*, 39(10), pp.2089–2103.

Haraksim, R. 2014. *Validation of likelihood ratio methods used for forensic evidence evaluation.* University of Twente, Enschede, The Netherlands.

Koppen, P. and Worring, M. 2009. Multi-target tracking in time-lapse video forensics. *IN: Proceedings of the First ACM workshop on Multimedia in forensics - MiFor '09.* ACM Press, p. 61.

Layer 3 Connectivity: IPv6 Technologies for the IoT. 2013. *IN: Building the Internet of Things with IPv6 and MIPv6: The Evolving World of M2M Communications.* John Wiley & Sons, Inc., pp. 220–256.

Lee, J.-H. 2015. IPv6 Address Configuration for Privacy Protection in the IoT. *Journal of Security Engineering,* 12(3).

Lund, P. 2014. An investigator's approach to digital evidence. *Digital Evidence and Electronic Signature Law Review,* 6(0).

Moon, P.-J. 2013. On the Availability of Anti-Forensic Tools for Android Smartphones. *The Journal of the Korea institute of electronic communication sciences,* 8(6).

Moon, P.-J. 2015. The Development of Anti-Forensic Tools for Android Smartphones. *The Journal of the Korea institute of electronic communication sciences,* 10(1).

Morrison, G.S. 2017. What should a forensic practitioner's likelihood ratio be? II. *Science & justice : journal of the Forensic Science Society,* 57(6), pp.472–476.

Oparnica, G. 2016. Digital evidence and digital forensic education. *Digital Evidence and Electronic Signature Law Review,* 13(0).

Sanders, J. 2004. Review of: Case Studies in Forensic Epidemiology. *Journal of Forensic Sciences,* 49(4).

Schaefer, L. and Mazerolle, L. 2017. Predicting perceptions of crime: Community residents' recognition and classification of local crime problems. *Australian & New Zealand Journal of Criminology,* p.000486581772159.

Snoek, C.G.M. and Worring, M. 2007. Concept-Based Video Retrieval. *Foundations and Trends® in Information Retrieval*, 2(4).

Souvignet, T.R. 2010. Digital forensics: Introducing the 'Extract–Analyse' model. *Science & Justice,* 50(1).

Sudars, K. 2017. Face recognition Face2vec based on deep learning: Small database case. *Automatic Control and Computer Sciences,* 51(1).

Tripathi, B.K. 2017. On the complex domain deep machine learning for face recognition. *Applied Intelligence,* 47(2).