Expediting the Digital Forensic Process through a Deduplicated Framework

UCD Forensics and Security Research Group

Xiaoyu Du, Mark Scanlon

School of Computer Science, University College Dublin, Ireland.

xiaoyu.du@ucdconnect.ie, mark.scanlon@ucd.ie

Introduction

In this age of big data, the sheer volume of cases anticipated to be encountered by digital evidence investigators is set to increase into the foreseeable future. Digital evidence backlogs gave become commonplace in local, national and international police forces throughout the globe. These backlogs often reach two years and can exceed four years in the extreme [1]. Addressing this backlog is crucial to ensure efficient investigation and prosecution. One promising solution is to redefine the traditional digital evidence processing model by moving much of the processing to a cloud-based environment.

Traditional process models specify a number of arduous steps for digital forensic investigation including identification, acquisition and storage, analysis, and reporting [2]. Digital Forensics as a Service (DFaaS) is a recent development capable of being integrated with the existing digital forensic process models leveraging the low-cost, always-on nature of cloud technologies. This research aims to design and implement a cloud-based deduplicated digital forensic system to improve the overall efficiency of the digital forensic investigation process.

Data Deduplication and DFaaS

Data deduplication is a data compression technique which aims to reduce the requirements of storage space and the network bandwidth during the data transmission. The premise behind data deduplication is outlined in Figure 1.

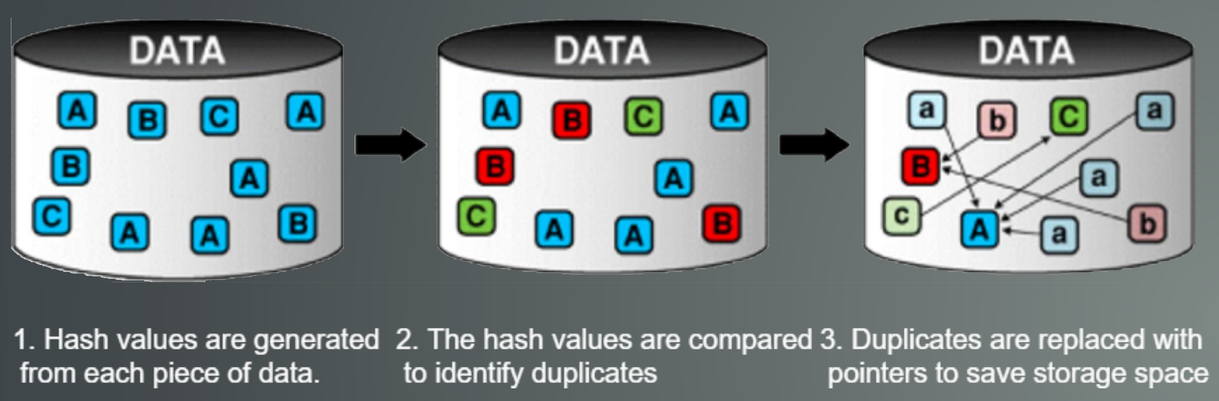

Figure 1. Data Deduplication

DFaaS is currently in the early stages of implementation for police forces with the largest operational system being implemented in the Netherlands [3, 4]. This implementation applies automated processing techniques to gather and process evidence. Its aim is to speed-up the investigation through providing case detectives with the access to query the evidence without needing to wait for expert manual analysis.

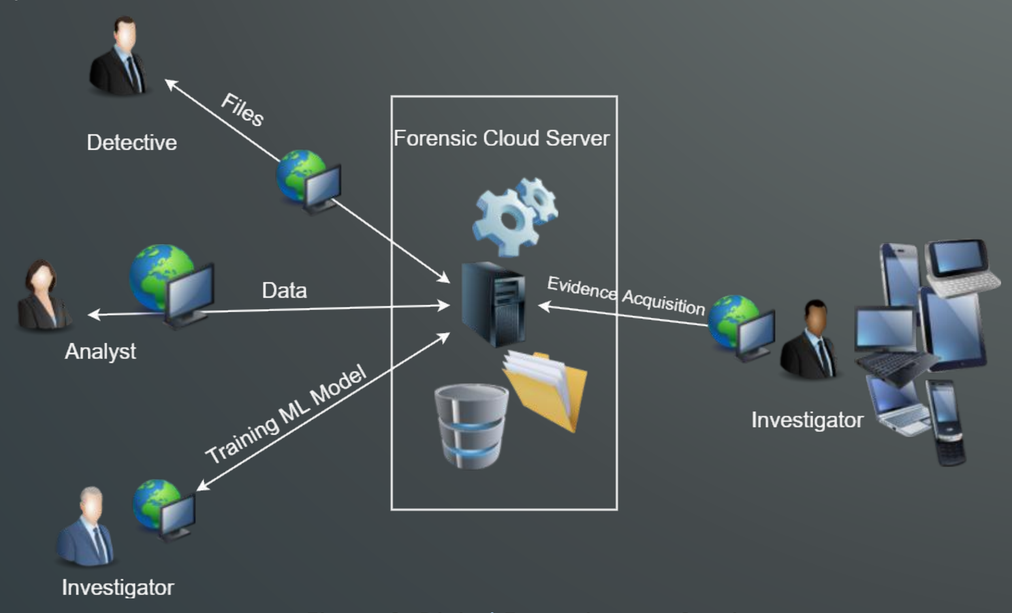

Figure 2. Digital Forensics as a Service

The cloud-based nature of this system can facilitate numerous additional benefits [2]:

- **Intelligent Forensics -** Manual expert evidence analysis and categorisation can be used for training machine learning based automated processing.
- **Resourcing -** Digital Forensics as a Service can offer sufficient storage space and powerful computing resources in an affordable manner.
- **Information Sharing -** Detectives, investigators, and expert analysts can work together in parallel on a case and their analysis can be shared across cases.
- **Efficient Management -** Easier management of both hardware and software resources ensuring the availability of the latest techniques to each investigation.

Digital Evidence Acquisition Methodology

As shown in Figure 3, the client can read each artefact and its metadata from the suspect drive, calculate the artefact's hash value, and then send this information to the cloud-based system. The system can then check the database to see if the file already exists, and if not, can send an artefact request to the client. In the database, information is stored for each artefact encountered, its metadata for each time it is found, and acquisition specific metadata later used to ensure forensically sound reconstruction.

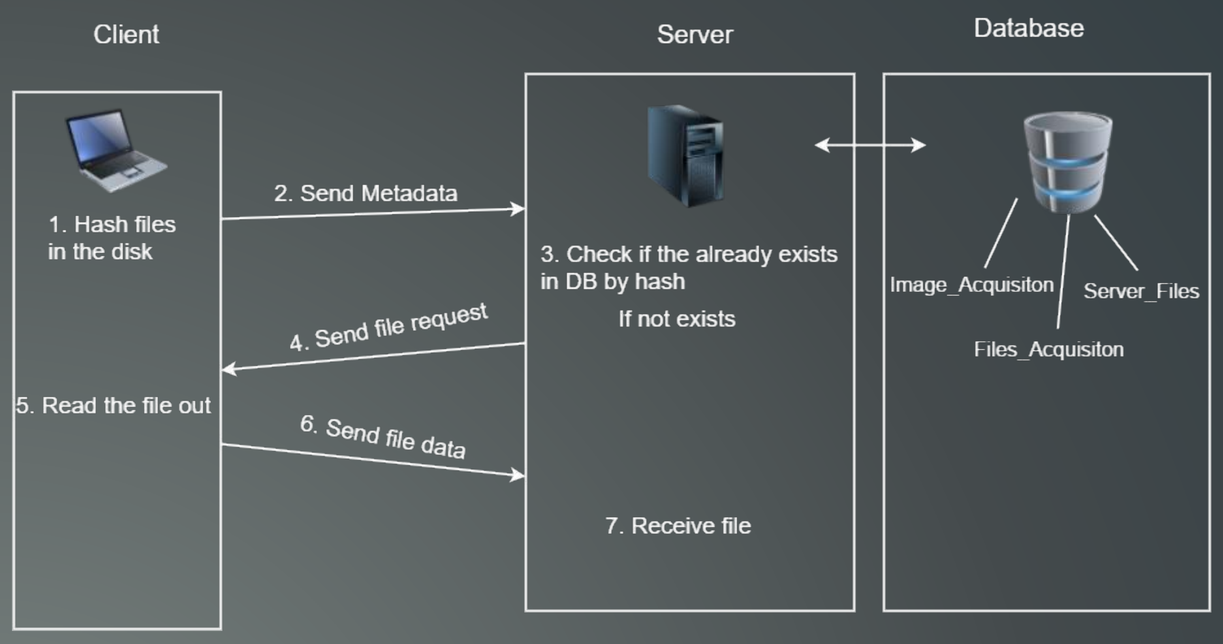

Figure 3. Digital Evidence Acquisition

The artefact data acquisition step includes: 1). Collecting all files, with their metadata stored in the database. 2). Slack space and unallocated space where data might remain in the file system. For each device seized in an investigation, every bit of the original data will be copied and transferred to the cloud. The database stores the image acquisition information and the artefacts' metadata, and the artefact is stored only once on the system.

During the process of disk image reconstruction, a disk staging area is first created with the same size as the acquisition. Each artefact, one-by-one, is written to the precise location in this staging area as it was discovered in the original disk image. For the unallocated space, its position in the disk is treated as a set of continuous blocks, and is treated like any other artefact in the database. Of course, an artefact might not be saved in a continuous stream on the disk, e.g., file fragmentation. As such, the starting offset and the length of each fragment is stored in the database, and for reconstruction, each file fragment is written to its corresponding location on the disk staging area.

Conclusion

DFaaS can provide a suite of benefits over traditional digital forensic process models. While combining cloud technologies with digital forensics is currently in its infancy [3], the existing service-based forensic system, XIRAF, which has been implemented in the Netherlands shows great promise for this technique [3, 4]. This research will continue on building the techniques and tools needed for a more intelligent cloud-based system for digital forensic processing, expediting the acquisition, analysis and reporting steps of the traditional process.

References

[1] Scanlon M. *Battling the Digital Forensic Backlog through Data Deduplication*. In: Proceedings of the 6th IEEE International Conference on Innovative Computing Technologies (INTECH 2016), Dublin, Ireland, 2016.

[2] Du X, Le-Khac NA, Scanlon M. *Evaluation of Digital Forensic Process Models with Respect to Digital Forensics as a Service*. 16th European Conference on Cyber Warfare and Security (ECCWS 2017), Dublin, Ireland 2017.

[3] van Baar R, van Beek H, van Eijk E. Digital Forensics as a Service: A Game Changer. Digital Investigation 2014;11:S54–S62.

[4] van Beek H, van Eijk E, van Baar R, Ugen M, Bodde J, Siemelink A. Digital Forensics as a Service: Game On. Digital Investigation 2015;15:20–38.

Acknowledgements

This work is funded by the School of Computer Science, University College Dublin.