Keynote:
Battling the Digital Forensic Backlog

DR MARK SCANLON

School of Computer Science
University College Dublin

International Workshop on Cloud Security and Forensics;
International Conference on Innovative Computing Technology

Introduction

- The early 21st century has seen a dramatic increase in new and ever-evolving technologies available to consumers and industry alike.
- Generally, the consumer-level user base is now more adept and knowledgeable about what technologies they employ in their day-to-day lives.
- The number of cases where digital evidence is relevant to an investigation is ever increasing and it is envisioned that the existing backlog for law enforcement will balloon in the coming years as the prevalence of digital devices increases.

Digital Forensic Backlog: Example Impact on Prosecution

- This backlog is having a significant impact on the ideal legal process.
- 12-18 months backlog is commonplace worldwide
- According to a report by the Irish National Police, delays of up to four years
  - ``Seriously impacted on the timeliness of criminal investigations'' in recent years.
  - In some cases, these delays have resulted in prosecutions being dismissed in courts.

Digital Forensic Backlog: Example Impact on Prosecution

- This issue regarding the digital evidence backlog is further compounded due to the cross-border, intra-agency cooperation required by many forensic investigations.
- If a given country has an especially low digital investigative capacity, it can have a significant knock-on effect in an international context

Digital Forensic Challenges

- The complexity problem
  - Arises from data being acquired at the lowest (i.e. binary) format with increasing volume and heterogeneity, which calls for sophisticated data reduction techniques prior to analysis.

Digital Forensic Challenges

- The diversity problem
  - Results from ever-increasing volumes of data
  - Lack of standard techniques to examine and analyse the increasing numbers and types of sources, which bring a plurality of operating systems, file formats, etc.
  - The lack of standardisation of digital evidence storage and the formatting of associated metadata also unnecessarily adds to the complexity of sharing digital evidence between national and international law enforcement agencies

Digital Forensic Challenges

- The volume problem
  - Increasing storage capacities
  - Increasing variety of device
  - Lack of sufficient automation

- The amount of data per case at the FBI's 15 regional computer forensic laboratories has grown 6.65 times between 2003-2011, from 84GB to 559GB

Digital Forensic Challenges

- The consistency and correlation problem
  - Results from the fact that existing tools are designed to find fragments of evidence, but not to otherwise assist in investigations.

Digital Forensic Challenges

- The unified time lining problem
  - Multiple sources present different time zone references, timestamp interpretations, clock skew/drift issues, and the syntax aspects involved in generating a unified timeline.

Rising Trends: Internet-of-Things Forensics

- Over 13.4bn IoT devices in existence in 2015
- Estimated to grow to 38.5bn by 2020
- Consumer Domain
  - Smart Home, Connected Vehicles, Digital Healthcare
- Industrial Domain
  - Retail, Connected Buildings, Agriculture

Rising Trends: Internet-of-Things Forensics

- Compared to traditional digital forensics, there is less certainty in where data originated from, and where it is stored.
- Data persistence may be a problem.
  - IoT devices themselves typically have limited memory (and may have no persistent data storage).
  - Thus any data that is stored for longer periods may be stored in some in-network hub, or sent to the cloud for more persistent storage.
- This means that each challenge related to cloud forensics will likely apply in the IoT domain also.

Rising Trends: Cloud Computing Forensics

- Usage of cloud services such as Amazon Cloud Drive, Office 365, Google Drive and Dropbox are now commonplace amongst the majority of Internet users.
- From a digital forensics point of view, these services present a number of unique challenges
- Typically, data in the cloud is distributed over a number of distinct nodes - unlike more traditional forensic scenarios where data is stored on a single machine.

Rising Trends: Cloud Computing Forensics

- Due to the distributed nature of cloud services, data can potentially reside in multiple legal jurisdictions, leading to investigators relying on local laws and regulations regarding the collection of evidence.
- This can potentially increase the time, cost and difficulty associated with a forensic investigation.
- From a technical standpoint, the fact that a single file can be split into a number of data blocks that are then stored on different remote nodes adds another layer of complexity thereby making traditional digital forensic tools redundant.

Rising Trends: Cloud Computing Forensics

- Cloud Service Providers (CSP) and their user base must be taken into consideration.
- Investigators are reliant on the willingness of CSPs to allow for the acquisition and reproduction of data.
- The lack of standardisation among the varying CSPs, differing levels of data security and their Service Level Agreements are obstacles to both cloud forensic investigators and researchers.

Rising Trends: Cloud Computing Forensics

- The multi-tenancy of many cloud systems poses three significant challenges to digital forensic investigations.

1. In the majority of cases the privacy and confidentiality of legitimate users must be taken into account by investigators due to the shared infrastructures that support cloud systems.
2. The distributed nature of cloud systems along with multi-tenancy can require the acquisition of vast volumes of data leading to amplification of the storage problem.
3. Finally, the use of IP anonymity and the easy-to-use features of many cloud systems, such as requiring minimal information when signing up for a service, can lead to situations where identifying a criminal is near impossible.

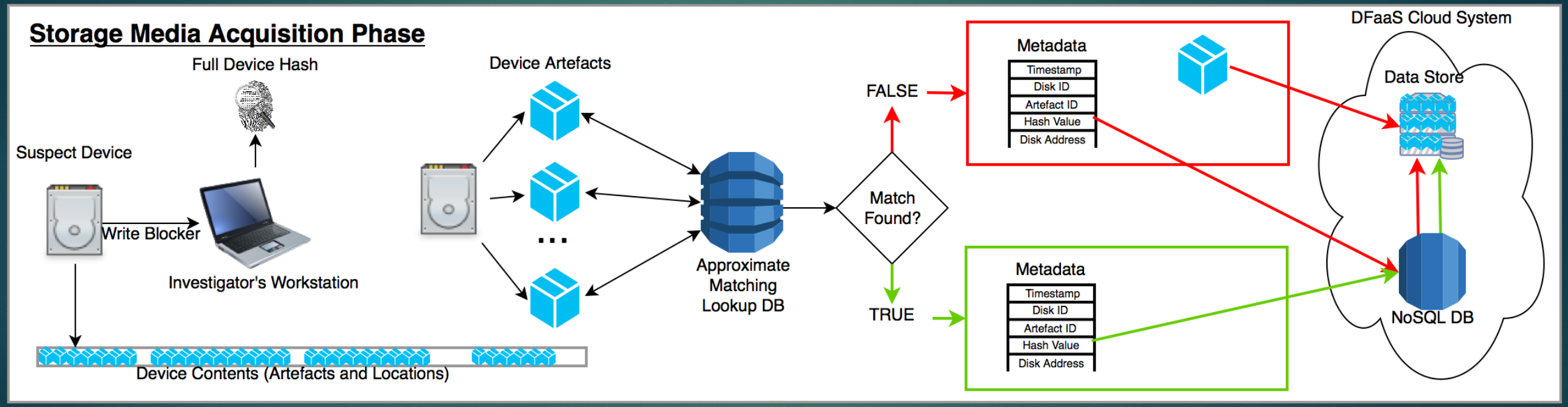
Future Research Areas: DFaaS

- Digital Forensics as a Service (DFaaS) is a modern extension of the traditional digital forensic process.
- Since 2010, the Netherlands Forensic Institute (NFI) have implemented a DFaaS solution in order to combat the volume of backlogged cases
- This DFaaS solution takes care of much of the storage, automation, investigator enquiry in the cases it manages.

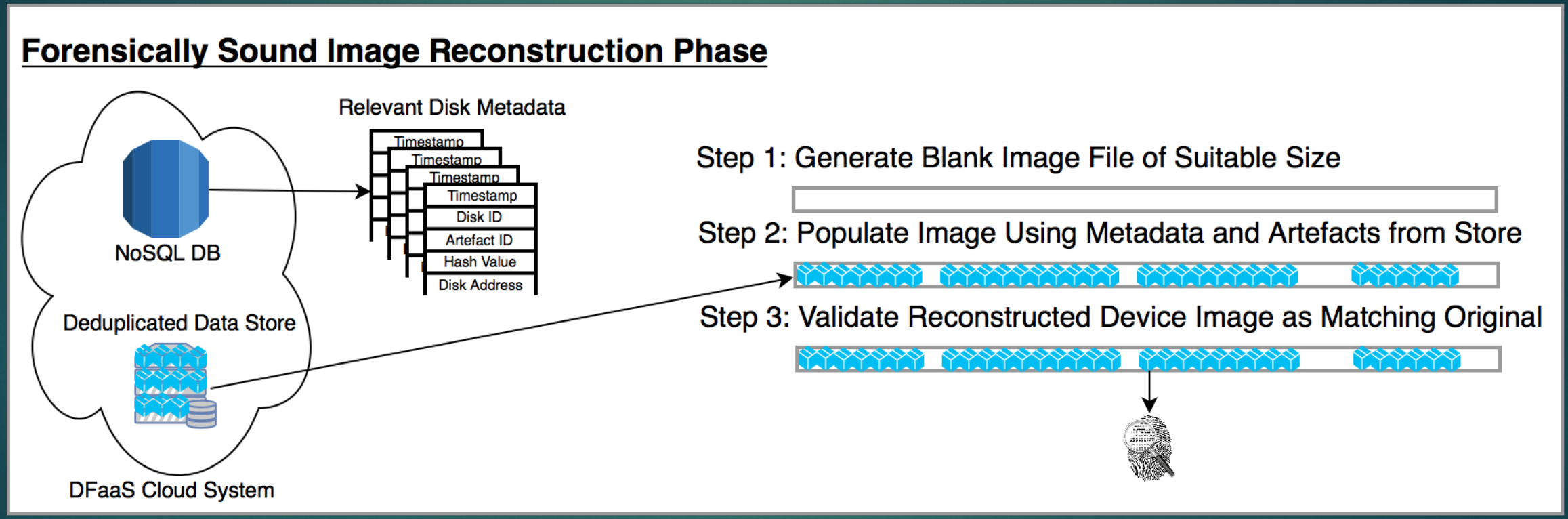
Future Research Areas: DFaaS

- Advantages:
  - Efficient resource management
  - Enabling detectives to directly query the data
  - Improving the turn around time between forming a hypothesis in an investigation its confirmation based on the evidence
  - Facilitating easier collaboration between detectives working on the same case through annotation and shared knowledge.

Future Research Areas: Data Deduplication



Future Research Areas: Data Deduplication



Complementary Research Areas

- Current investigation practice involves the analysis of data on standalone workstations.
- As such, the sophistication of the techniques that can be practically employed are limited.
- Much research has been conducted in a variety of areas that has theoretical relevance to digital forensics, but has been impractical to apply to date.

Complementary Research Areas

- A movement towards DFaaS and high-performance computing offers advantages beyond merely expediting the techniques currently used in forensics investigations, which remain reliant on manual input.
- It also promises a situation where this complementary research may practically be brought to bear on digital forensic investigations.

Complementary Research Areas: Information Retrieval Example

Complementary Research Areas: Information Retrieval Example

- Traditionally, IR researchers have been faced with the trade-off between the competing goals of precision (retrieving only relevant documents) and recall (retrieving all the relevant documents)
- Improving on one of these metrics typically results in a reduction in the other.

Complementary Research Areas: Information Retrieval Example

- IR for digital forensics can be seen as a typical example of legal information retrieval.
- Although this is certainly true at the point a case is being built for court, it could be argued that the level of recall required at the triage stage can be sacrificed somewhat for greater precision
- Thus there is the potential for configurable IR systems to be utilised in forensics investigations, whose focus will change depending on the stage of the investigation.

Conclusion

- Many issues affecting investigative efficiency
  - Combined, the negative effect can amplify
- The prediction is that the backlog will increase into the future
- A wealth of research directions are available to be taken
  - Cues can be taken from complementary CS research areas

ECCWS 2017: Call for Papers

- 16th European Conference on Cyber Warfare and Security
- Abstract Deadline: 8th December 2016
- Full Paper Deadline: 26th January 2017
- Conference Dates: 29th-30th June 2017

- Indexed by:
  - Thomson Reuters ISI (WOS) Conference Proceedings Citation Index
  - EBSCO database of Conference Proceedings
  - ProQuest database
  - Institution of Engineering and Technology in the UK
  - Google Scholar
  - Google Book search
  - Elsevier SCOPUS

MARK.SCANLON@UCD.IE

HTTP://WWW.MARKSCANLON.CO

@MRKSCN