Towards Forensic-Ready Software Systems

Liliana Pasquale
University College Dublin, Ireland

Dalal Alrajeh
Imperial College London, UK

Claudia Peersman
University of Bristol, UK

Thein Tun
The Open University, UK

Bashar Nuseibeh
The Open University, UK & Lero, Ireland

Awais Rashid
University of Bristol, UK

ABSTRACT
As software becomes more ubiquitous, and the risk of cyber-crimes increases, ensuring that software systems are forensic-ready (i.e., capable of supporting potential digital investigations) is critical. However, little or no attention has been given to how well-suited existing software engineering methodologies and practices are for the systematic development of such systems. In this paper, we consider the meaning of forensic readiness of software, define forensic readiness requirements, and highlight some of the open software engineering challenges in the face of forensic readiness. We use a real software system developed to investigate online sharing of child abuse media to illustrate the presented concepts.

CCS CONCEPTS
• Software and its engineering → Requirements analysis; • Applied computing → Evidence collection, storage and analysis;

KEYWORDS
Requirements, Forensic readiness, Forensic-ready Systems, Digital Forensics

ACM Reference Format:

1 INTRODUCTION
Forensic readiness represents the capability of an organization to support digital investigations proactively, i.e., before an incident occurs [24]. It is realized through the production of evidence that (i) facilitates the investigation and demonstration of compliance to organizational and regulatory policies, and (ii) can support legal proceedings [9]. To date, however, researchers’ attention has been geared towards the provision of general guidelines that could potentially enhance organizations’ operational and infrastructural capabilities to achieve forensic readiness. Little or no attention has been given to how the software systems deployed within these organizations can be designed to be themselves forensic-ready. With a rapid rise in cybercrime and cyber-enabled crime—number of identity theft incidents increased by 222% in 2016 [7] whilst online child sexual exploitation increased by 135% in 2016 [21]—there is an urgent need to consider what forensic readiness means for software systems and how such readiness can be incorporated as part of software development processes. Work in this area is either very preliminary or has been limited to specific aspects of forensic readiness, such as ensuring that evidence relevant to potential incidents is preserved [3, 17, 19] or that evidence integrity is maintained [16]. No work to date has considered the wider potential or implication of rigorous software engineering on the development of forensic-ready systems.

Our vision is to investigate the notion of forensic readiness in software systems and understand how forensic-ready software systems can be developed systematically. To achieve this vision we investigate forensic readiness requirements over software systems and assumptions over their encompassing environment. Requirements and assumptions can be used to derive implementable software specifications that achieve forensic readiness. Some of these requirements are data-centred, aimed to ensure availability, relevance, minimalism, non-repudiation, completeness, and linkability of data. Others are process-centred, aimed to ensure that the process through which the software system performs digital forensic activities is sound. We elicit forensic readiness requirements by reviewing existing literature and examining a real world investigative toolkit, iCOP [20], which was designed with the purpose of facilitating investigations of online child abuse media shared through P2P networks. Finally, we present open research challenges that relate to different aspects of engineering forensic-ready software systems and that consider how such systems can operate within emerging cyber-physical environments.

2 SOFTWARE FORENSIC READINESS
Although forensic readiness is a notion that is not new in the context of digital forensics, what it means and how it is conceptualised, differs amongst researchers, e.g., [11, 22, 24, 27]. In this paper, we are interested in what forensic readiness means for software engineering practices. At the heart of digital forensic readiness is the digital data, including the media and the activity logs available within an organization’s information system network, or on users’ devices. These data could hold valuable information about how a particular incident occurred and by whom, potentially resulting in a successful prosecution of the perpetrator [5, 13]. In this context, forensic readiness of system network or devices implies maximal usefulness of the data held as potential digital evidence...
admissible in court. Such usefulness can only be attained if the data and the process through which they are acquired, analysed and stored is forensically sound. (A forensically sound process is one that maximizes the evidentiary weight of digital evidence [18], whilst forensically sound evidence is one that can endure legal scrutiny in a court of law [10].) We take the view that forensic readiness in the context of software engineering is a property that encapsulates the capabilities of software to: (1) conduct digital forensic processes in a forensically sound way; and (2) produce forensically sound evidence. As we are concerned with developing software that ensures forensic soundness of data and processes, prior to the occurrence of an incident or attack, such capabilities must be proactive. Furthermore, we consider forensic readiness to be a property that is achievable either partially or fully, depending on the software capabilities, and is with respect to a set of speculated incidents that an organization has identified and assessed as critical.

3 MOTIVATING EXAMPLE

To provide a sense of our envisaged research direction and challenges to overcome, we discuss the iCOP toolkit [20]. This has the main purpose of identifying, preserving and analysing new or previously unknown child sexual abuse (CSA) media shared by suspects on peer-to-peer (P2P) networks. As shown in Figure 1, iCOP has two major components: the P2P Engine and the Analysis Engine. The P2P engine monitors information (e.g., IP addresses, filenames and hash values of files) together with metadata (e.g., when a user was last seen sharing a file) from public traffic on P2P networks. This information is passed on to the Analysis Engine, which compares the monitored hash values to a list of known hashes of CSA media seized by law enforcement. This allows the system to disregard CSA media already known to law enforcement. The file names that do not occur in the known hash lists are then analysed to assess their likelihood of containing CSA media. File names flagged as suspicious are passed back to the P2P Engine for downloading. The content of the downloaded files is subsequently analysed automatically by a Media Analysis module to determine whether the files contain child abuse material. Finally, the resulting list of suspicious new or previously unknown files is examined by investigators to confirm whether they contain child abuse images and videos. Once confirmed, these items are fed back into the hash database as known CSA media in future searches.

![Figure 1: Overview of the iCOP toolkit](image)

We aim to explore new software engineering methodologies and techniques for the design of software systems capable of supporting digital investigations proactively. In what follows, we consider the key requirements that software systems must satisfy and illustrate such requirements using the iCOP toolkit.

4 FORENSIC READINESS REQUIREMENTS

In this section, we describe a preliminary set of requirements for forensic-ready systems that were elicited by reviewing existing literature on forensic readiness. We discuss the requirements in the context of the iCOP toolkit. We also distinguish between requirements that are data-centred and others that are process-centred.

Availability. Data that may be useful for investigating potential incidents should be available [24, 27]. To achieve availability, data that may provide investigative clues must be preserved and retrievable by law enforcement agencies or individuals who are in charge of conducting an investigation. As data may not be kept in non-volatile memories (e.g., network traffic) and physical devices can have limitations (e.g., damaged hard drives), the capability to preserve data proactively must be in place. Preservation can be triggered by changes in the data to be collected [25], can be performed periodically [30] or for a limited amount of time [14], in order not to consume resources (e.g., battery power in a mobile device). To facilitate retrieval of preserved data, metadata should also be stored. In the context of the iCOP toolkit, data that can be useful to investigate incidents can be video and image files indicating a new child abuse and information about the users sharing CSA media (e.g., IP address, client ID). However, not all files can be preserved successfully because downloads are often slow and they stall if the computers sharing the file go off-line. To facilitate retrieval of such data, CSA related material, P2P network users and victims should be identified unambiguously. This is challenging because often the same CSA material is shared under different filenames and victims can appear in different files.

Relevance. Data preserved proactively should be relevant to potential incident cases [28]. Relevance of data means whether data is able to support or refute hypotheses explaining how incidents occurred [3]. Ensuring relevance of preserved data allows an organization to have the data preservation activities more targeted on the risks to the business [24]. Relevance can be subject to the judgement of an investigator [26] and is typically determined by the files and data types available [12, 23] (e.g., email addresses, message information, date and time information, cookies, social security and credit card numbers from a computer hard disk image). In our example, to satisfy the relevance requirement, iCOP should ensure that stored CSA media is of new material or previously unknown. This is challenging because analysing the content of any file shared on the P2P network is not computationally tractable. The iCOP toolkit uses textual features of the filenames and characteristics of the users sharing the files (e.g., users sharing the greatest number of suspected files or sharing the greatest number of files) to identify CSA related media. However, files that do not contain any textual clues to their illegal content - which otherwise is relevant data - or that are shared by new users may not be preserved.

Minimality. Data preserved proactively should not include any information that is unnecessary for the purpose of an investigation. Satisfaction of this goal can have the side-effect of reducing the amount of resources that are spent looking for digital evidence and, therefore, the costs of an investigation [27]. In our case study, to satisfy the minimality goal, the iCOP toolkit should not preserve files that do not refer to a child abuse case (false positives). A typical false positive error can be when webcam videos showing a
which connections are related via a common IP address or GUID. Additionally, all files that are confirmed to contain CSA content should be accessible by police investigators in possession of login credentials. For example, iCOP should ensure that preserved CSA content can only be accessed by police investigators to identify unknown victims.

Legal compliance. The process adopted to preserve data should ensure compliance with existing regulations, which may vary depending on the jurisdiction(s) in which an incident may occur. Identification of what regulations apply to a specific system depends on the nationality and the physical location of the data subject, as well as the physical location of the organization collecting data. For example, the General Data Protection Regulation (GDPR)2 in Europe and the Fourth Amendment in USA [1] regulate what data can be preserved and under which conditions (privacy). The EU Data Retention Directive, can prescribe for how long data should be retained (retention). The GDPR also prescribes for how long data can be accessed and by whom (access to retained data). For the iCOP case study, media files can be preserved in UK because they are ‘voluntarily’ shared to third parties but this would not apply in other European countries, such as Belgium. Any monitoring and downloading of CSA media can only take place at suitable law enforcement premises and access to CSA material is only given to police investigators, in order to ensure privacy of victims’ identities.

5 SOFTWARE ENGINEERING CHALLENGES

We elicit a number of open software engineering challenges.

1. Representing and reasoning about forensic-ready systems. We have presented a first conceptualization of forensic readiness requirements of software systems. There is a need to build a consensus around the key characteristics of forensic-ready software systems. We can divide the implications and challenges into three sub-categories: (i) concept (how to represent and reason about forensic-ready systems and their properties), (ii) method (how to design and implement forensic-ready systems), and (iii) tools (how to analyse and support the development of forensic-ready systems). Existing work on concepts and taxonomy of dependability [4] could be a useful reference model to extend to forensic readiness. Such characterization would facilitate a better understanding of the potential relationship between forensic-ready requirements and other types of requirements such as security, privacy and safety. Furthermore, there is a need to characterise forensic-ready systems formally. This requires identifying formal languages —if any—that are best suited to express forensic readiness requirements, and will allow us to understand the extent to which existing representation and reasoning techniques are applicable to forensic-ready systems.

2. Methods for engineering forensic-ready software systems. The notion of forensic readiness poses challenging questions for software engineering methods and particularly how should existing methods adapt to account for forensic readiness requirements. Research is needed to answer a number of fundamental questions related to how requirements for forensic-ready systems should be implemented and whether these requirements are solely about data preservation activities. Architectural patterns (similar to security patterns [29]) could also be investigated to design forensic-ready systems. Additional challenges relate to managing trade-offs between forensic readiness requirements such as privacy and availability given that some of these could manifest at runtime. For

---

1https://www.darpa.mil/program/transparent-computing

2Regulation (EU) 2016/679
ACKNOWLEDGEMENTS

This work is supported by EPSRC Grant: EP/N028112/1, EU Safer Internet Programme project SI 2601002, and SFI Grants 10/CE/11855, 13/RC/2094 and 15/SIRG/3501.

REFERENCES

[22] A. Poe and L. Labuschagne. 2012. A conceptual model for digital forensic readiness in software systems and the requirements that support its attainment, highlighting some of the open software engineering challenges. For future work we plan to provide a formal characterization of forensic readiness requirements. We will also explore techniques for analyzing tradeoffs between conflicting requirements. Finally we will investigate aspects related to the implementation of a forensic-ready system, such as the generation of specification for such systems or assessment of relevance of preserved data.

6 CONCLUSION

In this paper we investigated the notion of forensic readiness in software systems and the requirements that support its attainment, highlighting some of the open software engineering challenges. For future work we plan to provide a formal characterization of forensic readiness requirements. We will also explore techniques for analyzing tradeoffs between conflicting requirements. Finally we will investigate aspects related to the implementation of a forensic-ready system, such as the generation of specification for such systems or assessment of relevance of preserved data.