Digital Evidence Regulation - an assessment of underlying issues in England and Wales

Angus M. Marshall BSc CEng FBCS CITP FRSA

PhD by publication

University of York

Computer Science

November 2022

Abstract

In the field of Digital Forensics, in England and Wales, the author has published a study of technical requirements found in Standard Operating Procedures and validation methods, evaluated potential mechanisms for producing evidence of verification as a means of reducing the validation and re-validation effort required, and examined the use of language in various documents produced, and referenced, by the Forensic Science Regulator. From this work, he argues that the current situation re validation may be giving a false sense of assurance that technical requirements are being satisfied, that it should be possible for evidence of verification to be made available to solve this problem, without requiring full disclosure of commercially sensitive or secret methods, and that the situation may have arisen through poor use of language in the Regulator's guides. He also suggests that the FSR's guides may have allowed, or caused, Digital Forensic Laboratories to ignore or misunderstand the importance of technical requirements in Standard Operating Procedure design and validation. Finally, having observed the lack of interest in the FSR's work and in method validation in court proceedings, he considers, from a lay perspective, the legal position relating to admissibility of computer-derived and computer-generated evidence. From this, he argues that the legal precedents are not entirely valid in the context of modern systems, and proposes a new classification of digital forensic systems which takes account of the increasingly automated analysis present in these tools.

Contents

List of Tables

Acknowledgments

My thanks go to Richard Paige (for letting me kick this off, providing feedback on drafts when required, and for wrangling the examiners), and Paul Cairns for encouraging me to complete in spite of my conviction that I'm doing this for all the wrong reasons.

But the biggest thanks go to Shirley, my wife, who wouldn't let me quit and tolerated me throughout the process.

Author's declaration

I declare that this thesis is a presentation of original work and I am the sole author of this narrative chapter.

This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1

Integrative Chapter

1.1 Foreword

This chapter considers the following documents and their contribution to improving understanding and application of digital forensic quality standards. In particular, it presents an argument, evidenced by the work presented in those papers, that the current regulatory regime for digital forensic work in England and Wales is based on a fundamentally flawed interpretation and implementation of the underlying standards chosen.

- ISO/IEC 27041:2015 Information technology — Security techniques — Guidance on assuring suitability and adequacy of incident investigative method (Marshall, AM (Ed – uncredited))[1] (Appendix B)

- ISO/IEC 27042:2015 Information technology — Security techniques — Guidelines for the analysis and interpretation of digital evidence (Marshall, AM (Ed – uncredited))[2] (Appendix C)

These two standards form part of a group of 4 under the umbrella of ISO/IEC JTC1 SC27 WG4's portfolio. The others being ISO/IEC 27037:2012 Information technology — Security techniques — Guidelines for identification, collection, acquisition and preservation of digital evidence and ISO/IEC 27043:2015 Information technology — Security techniques — Incident investigation principles and processes.

The following group of 3 papers, present results of attempts to understand why the principles of ISO/IEC 27041:2015 have not been adopted by the digital forensics community, and explore some of the issues relating to regulation of digital forensics, and reliance on apparently unproven tools.

- Requirements in digital forensics method definition: Observations from a UK study, AM Marshall, R Paige, Digital Investigation 27, 23-29, 2018 [3] (Appendix E)

- Digital forensic tool verification: An evaluation of options for establishing trustworthiness, AM Marshall, Forensic Science International: Digital Investigation 38, 301181, 2021 [4] (Appendix F)

- The unwanted effects of imprecise language in forensic science standards, AM Marshall, Forensic Science International: Digital Investigation 40, 301349, 2022 [5] (Appendix G)

The following papers are included purely as exemplars of how some of the issues discussed can be addressed through experimental verification and peer-review.

- SyncTriage: Using synchronisation artefacts to optimise acquisition order, C Hargreaves, A Marshall, Digital Investigation 28, S134-S140 [6] (Appendix H)

- CaseNote: Mobile phone call data obfuscation & techniques for call correlation, AM Marshall, P Miller, Digital Investigation 29, 82-90, 2019 [7] (Appendix I)

- WhatsApp server-side media persistence, AM Marshall, Digital Investigation 25, 114-115, 2018 [8] (Appendix J)

The work under consideration is based on a mixture of qualitative and quantitative research and the author's personal experiences as a digital forensics practitioner, since 2001, providing assistance to law-enforcement and the legal profession.

1.1.1 Note on terminology

As much of the argument presented in the author's work relates to "relaxed" use of terminology, it should be noted that two concepts in particular appear in this document and should be interpreted according to the ISO definitions as:

- Validation - Verification, where the specified requirements are fit for an intended use.

- Verification - Provision of objective evidence that a given item fulfils specified requirements.

Furthermore, a distinction is drawn between a tool, as defined by case law and something that this author calls a “Forensic Cyber Assistant”. The distinction is discussed in more detail, in relation to principles of admissibility, in 1.5, but can be summarised as follows:

- tool - something which assists a human being but still requires significant human input to control or operate it.

- Forensic Cyber-Assistant (FCA) - something which automates one or more modules in a digital forensic workflow. Having been configured, it proceeds to complete its defined task without further human intervention.

FCAs can be constructed from pipelines of tools and, may themselves, be used as tools if only a limited subset of their capabilities is being deployed under direct human control.

1.2 Aims and Objectives

1.2.1 Background

1.2.1.1 The Forensic Science Regulator and Applicable Standards

Since the creation of the role of Forensic Science Regulator, in 2008, forensic science providers (FSPs) in England and Wales have been required to consider implementing ISO/IEC 17025 as an overarching standard for their work. Although there was no statutory requirement for them to become accredited until the creation of The Accreditation of Forensic Service Providers Regulations 2018[9] and The Accreditation of Forensic Service Providers (Amendment) Regulations 2019[10], nor any particular interest in accreditation status during court proceedings (1.5 provides a discussion of why this may be the case for digital evidence in particular), some contractual arrangements have required it. With the passing of the Forensic Science Regulator Act (2021)[11], a statutory footing for the regulator has been established. This still does not require accreditation on the part of all FSPs, the accreditation regulations [9][10] being limited to those providing dactyloscopic and/or DNA services to designated law-enforcement authorities only, but does give the regulator the power to investigate and issue compliance notices against any FSP which does not meet an acceptable standard, following a complaint.

The legislation [11], activated in July 2022 [12], requires the Regulator to publish a Code of Practice with which compliance is not mandatory, but which can be used to inform admissibility of evidence in trials. (Until this code has been published, Version 7 of the Regulator's Codes are to be used [12]). The FSR act [11] also gives the Regulator an investigatory role in relation to any person believed not be acting in compliance with the Codes and the powers to issue Compliance notices in order to rectify defects in processes relating to forensic science activities.

This has the effect, therefore, of making the Regulator's Code the de facto standard for all forensic science activities in England and Wales although compliance will be enforced by exception rather than by statutory requirement.

At the time of writing, the Regulator's Statutory Code exists only in a draft form, but the latest version available [13] is similar to the previous Codes and Guidance notes which are based on the ISO/IEC 17025:2017 [14] standard. The same issues around overloading of concepts and misuse of terminology noted in this author's paper on misuse of language [5] appear to be present. For example, this draft continues to define verification as "confirmation, through the assessment of existing objective evidence or through experiment, that a method, process or device is fit (or remains fit) for the specific purpose intended." potentially causing more confusion by conflating the concepts of verification and confirmation. There is, however, a useful change in position in relation to the use of commercial software and it is now explicitly stated that "Commercial off-the-shelf software and software tools whose operation has an impact in obtaining results will require validation, or any existing validation to be verified, as laid out in section 28.3 - Validation of Methods." and there is also a change in the concept of the end user, which is embodied as "the requirements of all end users (e.g. other practitioners, investigators, prosecutors and the CJS) must be considered". [1]

Historically the ISO 17025 standard, although intended for "weights and

[1] It may, of course, be entirely coincidental that this wording appeared after this author submitted copies of his works to the Regulator for information.

measures" laboratory activities (formally, calibration and testing), has been successfully adopted by some forensic science disciplines, especially DNA. It was declared as the standard to be adopted by EU member states in 2011 [15], to enable the creation of a European Forensic Science Area by 2020 (a "single market" for forensic science services intended to remove inter-jurisdictional barriers by guaranteeing minimum standards for evidence production). It is this Council decision which led to the Accreditation Regulations [9][10], although the Council decision included digital evidence disciplines within its scope while the Regulations do not.

In brief, ISO/IEC 17025:2017 requires accredited organisations to have 4 key elements in place:

1. An overarching quality management system, similar to that required by ISO 9001.
2. Evidence of competence of staff involved in accredited processes.
3. Evidence of proficiency of the organisation, probably based on inter-laboratory proficiency comparison exercises.
4. Validation of methods - i.e. demonstration that methods used are "fit for their intended purpose".

It is the fourth of these "pillars of quality" and the relationship between validation of method and verification of the tools and FCAs involved in methods which led to the work under discussion in this chapter.

1.2.2 The author's motivation

The work presented here grew, at least in part, out of curiosity about misunderstandings, which appeared in published literature and private communications, relating to the intent, meaning and interpretation of standards.

These ranged from "you can't apply the same standards to digital as DNA because digital is different/fast moving/too complicated" [16][17] to "ISO 17025 won't work. We should be using 27037, 27041 and 27042"[17]. The latter of these is particularly interesting because this author was involved in the development of the 270xx group of standards and is very conscious that the teams involved took great care to ensure compatibility between that group and the ISO 17025 standard(s). The motivation, there, was to ensure that any organisation which adopted the 270xx group for internal investigations would be capable of demonstrating that their investigative methods had followed a similar QA regime to that required for law-enforcement, potentially reducing or eliminating the need for re-investigation or re-validation if their results were to be used in criminal cases.

These, coupled with a constant background of complaints about excessive cost, effort and risk of failure to gain accreditation, led to a desire to examine more closely how the relevant standards were being applied to digital forensic work and how some of the common issues could be resolved either by improving understanding or by modifying compliance methods.

1.2.3 Initial Aims

The initial aims, therefore, were

1. to explore the hypothesis that there is some fundamental difference between digital forensics and other forensic sciences which means that the extant quality standards cannot successfully be applied

    - including trying to understand why the 270xx group was seen as more directly applicable than the 1702x alternatives in spite of alignment and consistency striven for during their development[17]

2. to evaluate the potential for cost savings possible through adoption of the ISO/IEC 27041 recommendations about tool verification as an adjunct to method validation

3. to attempt to determine the potential effective mechanisms for tool vendors to provide support to customers seeking accreditation, through disclosure of tool verification information.

In practice, as will be seen later, the second and third of these aims proved particularly difficult to achieve and, although some progress has been made towards them, in the form of proposed mechanisms and explanations for problems encountered, much work remains to be done.

1.3 Methodology

1.3.1 Precursor work

Some of the concepts embodied in the 27041 and 27042 standards, of which this author was the editor, and hence primary author, grew out of work previously done for government agencies. This unpublished work resulted in the production of models of typical digital forensic processing workflows and role descriptors. These were based on information captured from questionnaires, workshops and direct observation of digital forensic service providers across England and Wales. Law-enforcement laboratories and private sector providers, of all sizes, were included. Some of this information also underpinned the EPS KTN Forensic Science Special Interest Group's "Digital Forensics Capability Review"[18] which this author led.

Although the full results of this work have not been published, the elements present in the KTN report [18] and the ISO/IEC 27041 [1] and 27042 [2] standards have been and, thanks to the processes involved in the production of all three of these documents, subjected to considerably more rigorous pre-publication review than for ”normal” academic work.

1.3.1.1 ISO standards

ISO/IEC JTC1 SC27 WG4 had, around the time of appointment of the first Forensic Science Regulator (2008), commenced development of ISO/IEC 27037 [19] as a standard for first response to information security incidents (covering the Identify, Capture, Acquire, Preserve phases). As this standard was likely to be adopted by businesses which might have to make their investigative results available to law-enforcement agencies, the author led the task of proposing and developing the complementary ISO/IEC 27041 [1] and 27042 [2]standards to provide guidance for the conduct of all stages of an investigation in a manner which would be compatible with organisations accredited to ISO/IEC 17025 [20][14]. This process was, initially, supported by the Forensic Science Regulator. The intention of the 270xx family, therefore, was to provide a compatible process, but one which was more easily applied to information technology specifically.

At that time, ISO 17025 (2005)[20] defined validation, but not verification. The 270xx standards adopted definitions of both validation and verification from ISO/IEC 27004 with minor wording changes for clarity. Similar definitions were adopted in the 2017 version of 17025[14], viz:

Validation Verification, where the specified requirements are fit for an intended use.

Verification Provision of objective evidence that a given item fulfils specified requirements.

and, via reference to ISO 17000 [21] defines these in terms of the following :

specified requirement need or expectation that is stated. Note 1 to entry: Specified requirements can be stated in normative documents such as regulations, standards and technical specifications. Note 2 to entry: Specified requirements can be detailed or general.

These are compatible with the concepts of verification and validation commonly used in software engineering or development - i.e. verification provides assurance that the product conforms to specification, while validation provides assurance that the user can use it to solve their particular problem.

ISO/IEC 27042 [2] leverages this in its recommendation that methods should be "atomic" - i.e. designed to perform single functions, to encourage modularity and re-use, while ISO/IEC 27041 [1] suggests that verification of "tools" (including FCAs) can be used as partial evidence of validation where a subset of a method's requirements can be mapped to a subset of a tool's verified requirements. (i.e. those elements of a method which rely on a tool operating in conformance with its specification can be considered validated because of the presence of evidence of verification.) This recommendation was included to deal with the issue of initial validation, but more importantly that of re-validation of methods when tools are updated. Given the frequency with which patches and new versions can be offered (in the case of at least one mobile phone tool, 5-6 new releases per annum is not unusual) in order to deal with errors and assist with the examination of new or updated devices, the requirement to potentially re-validate every method which involves the use of a particular tool could become overwhelming and all-consuming. The use of vendor or third-party provided assurance of verification would, the standard proposes, allow the re-validation to be reduced to an exercise in mapping previous requirements against the new verified specification, with user testing only required where there had been a significant change in specification of either the tool or the method.

1.3.2 The requirements study

The study on requirements [3] was an attempt to look more closely at how laboratories document and test methods with the intention of evaluating the ISO/IEC 27041 [1] "tool verification" as a part of method validation concept. The study examined 35 Standard Operating Procedure documents from 3 law-enforcement laboratories, and 7 validation plans from the same type of organisation. It also considered the SOPs provided as exemplars by the Scientific Working Group on Digital Evidence (SWGDE). In all cases, it was found that there were no explicit technical requirements in either SOPs or validation plans.

In parallel with this activity, attempts were made to engage with 5 tool/FCA providers to evaluate the potential for evidence of testing or verification to be made available, in order to allow the ISO/IEC 27041 [1] model to be applied through a mapping process. In 4 cases, although the providers were initially willing to discuss this concept, they chose to withdraw support for the study at the point where some sort of disclosure or inspection of internal development methods and/or testing became necessary. In the case of the fifth provider, it became apparent that their use of (self-described) "agile" development methods did not include any obvious form of requirements capture/extraction or verification testing in a form which would easily allow a detailed specification to be produced for the mapping exercise, although in-depth analysis of the information captured in their "user stories" might allow some form of requirements to be generated by inference or induction (resulting in a need to verify those requirements prior to use because they would depend on how the user stories were being interpreted by the analyst/developer responsible).

The conclusions from this exercise were, therefore,

- that (some) accredited organisations were not fully considering technical requirements in their validation exercises;

- that (some, likely to be a majority) tool/FCA providers were not able or willing to disclose specifications and evidence of verification; and

- that, therefore, evidence being produced via validated methods might well have an unjustified veneer of respectability as it could not be shown to be satisfying any particular scientific or engineering requirements.

1.3.2.1 Comparison with Metrology Laboratories

Following this, this author was concerned that his understanding of how the ISO 17025 standard should be applied was incorrect, and carried out a review of metrology laboratory accreditations, combined with a small telephone and email poll of accredited organisations.

This desk study, using the UKAS accreditation database [22] as source, suggested that each laboratory has a range of testing methods, designed to test samples of a particular product (e.g. metal alloy). These tests are verified and calibrated against known samples, and relatively simple in nature (in ISO/IEC 27042 [2] terms, "atomic") typically testing a single property of the sample at a time. When a customer approaches the lab. to have a sample tested, the lab. identifies the customer requirements (e.g. *show that the steel setscrews in this sample are good enough to be used as seatbelt anchors*). The lab. then maps these requirements onto the tests that it has accreditation for, making recommendations to the customer (e.g. *"seatbelt restraints need to pass the 8.8 tensile strength test and should be corrosion resistant to ASTM B117. We therefore propose tests for tensile strength and corrosion resistance to those standards"*). Once agreed with the customer, (i.e. confirmed that the laboratory's proposed standard tests satisfy the customer's requirements - meaning it is validated because the lab. has evidence that it can conduct those tests and the customer has agreed that those tests meet their needs) the tests are carried out and results reported to the customer.

For this type of testing, requirements are relatively easy to identify, whether as end-user (customer and customer's customer) requirements explicitly stated, or by implication through the nature of the request (i.e. the test laboratory has experience of similar requests and has suitable tests available). Validation is, effectively, a paper exercise providing confirmation that the agreed tests produce appropriate results to satisfy the customer's request.

The telephone and email poll provided confirmation that this is, indeed, the position in this type of laboratory and fits exactly with the original intentions behind the development of the ISO 17025 standard.

1.3.3 The verification options

Given the difficulties identified in the first study, the author turned to a more "Socratic" development exercise, considering how evidence of verification could be obtained or provided in a way which would allow vendors to retain commercial confidentiality, minimise exposure to liability and still provide sufficient levels of assurance that the regulatory regime and tests for admissibility could accept it. Several potential models were considered, including Capability Maturity Model Integration, IEEE730-2014 Standard for Software Quality Assurance Processes [23], ISO/IEC/IEEE 12207:2017 Software lifecycle processes [24] and ISO/IEC/IEEE 15289:2011 Content of life-cycle information products (documentation) [25] standards. Although these do provide relatively good degrees of assurance of adherence to internal procedures during software development, they do not inherently provide a simple mechanism for evidence of verification to be provided to third parties. In essence, they are designed for internal use rather than external consumption and not entirely appropriate for the sort of Commercial Off The Shelf (COTS) products that are found in the digital forensics tool/FCA marketplace. Therefore, a model based on the Trustworthy Software Foundation's trust levels model [26][27] was developed.

In the proposed model, various levels of disclosure about tool/FCA requirements and development were considered, ranging from complete openness (the Open Source Software model) to complete secrecy (the current default for commercially sensitive products). Estimates were made of potential liability, effort and exposure to other risks for the producer, user and trusted-third-party inspection body (where involved) and used to rate the various options in an attempt to identify appropriate levels of compromise which minimised exposure to potential harm to all parties.

Although highly theoretical in nature, because of the problem of lack of engagement with the providers, this exercise resulted in a successful evaluation of the options available and did produce a recommendation. At the time of writing, no producer, user, regulator or other interested party, has fully engaged with the suggested options, but some level of interest is being shown, at least in reading the published paper.

1.3.4 The language issue

This study involved a relatively simple, although time-consuming, exercise which required an in-depth comparison of key concepts as they were expressed in the various documents which make up the hierarchy of standards and implementation guidance used by the Regulator. At the root of this is the ISO 17025 standard [14], followed by the ILAC-G19 guidance [28] which attempts to show how the standard should be applied to all forensic sciences. These are followed by the FSR's Codes [29], the Annex for Digital Forensics [30] and the Guidance on Validation of Digital Forensic Methods[31].

Although a relatively simple method (examining definitions and use of defined terms) was used, the concept of consistency of language is something which editors of standards, including this author, know is important. One of the most common causes of comments and requests for corrections during the development of any standard is the incorrect use of a term which has a precise definition in the ISO dictionary as embodied in the Online Browsing Platform[2]. Failure to adhere to this "Humpty Dumptyism" leads to confusion and misapplication of standards. From the author's own perspective, native English-speakers tend to be more lax about their use of language and having the assistance of experts or co-editors who have English as a second or third language is invaluable in ensuring that terminology remains consistent.

1.4 Results

1.4.1 From the precursor work

The precursor studies, which led to the models used in ISO/IEC 27041 [1] and 27042 [2], showed that application of some simple software engineering principles allowed seemingly complex investigations to be decomposed into workflows composed of distinct modules, which could exhibit relatively low coupling and high cohesion[32]. The one obvious major barrier to this arose at the interface level where data from one module could not be ingested by another because of data format limitations within tools. This created a need to introduce transformation modules whose sole purpose was to reformat data output by one tool into a format appropriate for the next tool or process in the flow.

[2] https://www.iso.org/obp/ui

At the time in question (approx. 2011-2014) this was particularly important, and evident, in the use of early automated tools to identify common illegal images. Output from a general-purpose forensic examination tool had to be reformatted for the automated identification tool, whose output then had to be reformatted for re-ingestion into the general-purpose tool.

1.4.2 From the requirements study

Marshall and Paige [3] found that mapping tool or FCA specifications to user requirements, to aid validation, appeared to be impossible, firstly because the methods and validation plans examined did not contain technical requirements and secondly because no commercial FCA/tool provider was willing (or able) to disclose specifications, let alone provide evidence of verification.

For this first point, as the author's paper on language [5] in the various documents used, and produced, by the Regulator has highlighted, there is a lack of guidance on how requirements should be identified or articulated, and a possible over-emphasis on the potentially vague concept of "end-user" requirements.

There is also the issue of the concept of the Criminal Justice System (CJS) as end-user. This implies that the CJS can be viewed as some sort of homogeneous whole with simple and easily-identified requirements. In fact, by any common definition [33], the concept of the CJS embodies everything from detection (recognition, identification) of a potential crime, through first response, detailed investigation, preparation of case files, decision to prosecute, trial and punishment/rehabilitation. Rather than being, therefore, a single external customer entity with a single set of requirements based on the production of evidence for court use, the CJS could be more correctly viewed as a pipeline of stages, each with its own operatives, methods and goals, and thus different requirements at each stage influenced by those of the later stages.

This concept is important as its misuse can lead to cognitive bias in the investigative process. 1.4.2.1 explores this further.

1.4.2.1 The CJS Pipeline

The Regulator's April 2022 draft statutory code [13] acknowledges that the CJS itself is not necessarily the only end user and gives a list of exemplars.

Prior to seeing this revision, this author had considered the concept of end-user in the context of how a typical investigation might proceed and proposes the following pipeline model (Table 1.1), using the ISO/IEC 27037 [19] and 27042 [2] *Identify, Collect, Acquire, Preserve, Analyse, Interpret, Report (ICAPAIR)* phases as a means to identify the origin and likely nature of requirements for each phase. The operatives listed against each stage are those who carry out the processes and the end-user for each stage can be considered to be the operatives listed in the next stage.

In the pipeline, stages 7,8 and 9 are outwith the control of Forensic Science Providers (FSPs) and the FSPs' inputs to stages 8 and 9 are the same as their inputs to stage 7. Therefore, we can consider the ultimate end-user to be the the trial process, with the needs of the accused, legal representatives, judges, jurors and reporters as the final set to be satisfied.

Focusing on those needs alone, as this author highlighted in the language paper [5] could result in over-emphasis of those requirements earlier in the chain, potentially leading to the exclusion of other essential considerations.

In reality, the "end-user" of a process at any stage of the pipeline is the next process in the pipeline - e.g. imaging a device must satisfy the basic requirements of continuity and non-spoliation which are required of all processes, but must also result in an image which is suitable for processing in order to extract data. Extracted data must be presented in a form which is suitable for analysis, etc. Thus we have a set of legal/procedural requirements - i.e. those which must be satisfied in order for digital evidence (DE) to become legal digital evidence (LDE)[2], and technical requirements - i.e. those which must be satisfied for Potential Digital Evidence(PDE) to become DE[2].

Clearly, during the intermediate stages, technical requirements will mean that data intended for use for successor methods may not be in a format which satisfies the requirements of the trial itself (i.e. it has not been interpreted and presented for non-specialist human use). We should, therefore, expect to see a combination of overarching legal/procedural requirements and technical requirements defined for each method, and its associated validation, and that the requirements should develop away from the highly technical towards the legal/procedural as we progress towards the end of the pipeline.

Not only is the CJS pipeline somewhat longer than the Testing Laboratory pipeline, but it is less deterministic because of the evolutionary nature of criminal investigations and the additional, need to pass the CPS "gatekeeper" for charging decisions (the "public interest" and "realistic likelihood

Category	**Stage**	**ISO/IEC 270xx activity type**	**Operatives involved**	**Requirements categories**
Preparatory	-1 Response Preparation		DEFR, DES	Requirements from later stages necessary for specifications, verification and initial validation
	0 Crime happens	Incident Occurs		
Investigatory	1 Crime is noticed/reported		Inv.	
Investigatory	2 Digital First Response	ICAP	DEFR	Legal and Technical
Investigatory	3 Digital Investigation	AIR	DES	Legal and Technical
Reporting	4 Charging File Prepared	R	Inv	Legal
Decision	5 Charging Decision (Prosecute / NFA)		CPS	Legal
Reporting and Re-investigatory	6 Case File Prepared	AIR	CPS, Inv, DES	Legal
Reporting	7 Trial	IR	Legal representatives, accused, DEFR, DES, Judge, Jury, Reporters	Legal and Technical
	8 Verdict Reached		Jury, judge	Legal
	9 Release / Sentence			Legal

Table 1.1: CJS pipeline.

Abbreviations used in the table.
DEFR = Digital Evidence First Responder (See ISO/IEC 27037)
DES = Digital Evidence Specialist (see ISO/IEC 27037)
Inv = Investigator (usually a law-enforcement officer)
CPS = Crown Prosecution Service

of conviction" tests). Unlike the testing laboratory situation, where customer requirements can be identified and agreed prior to the tests being carried out, a criminal investigation is a living thing. Application of the ABC[3] principles and 5WH[4] method mean that new requirements are identified during the investigation, often as a result of specific questions being answered or results of tests becoming apparent. It is for this reason that ISO/IEC 27042 defines an investigation in terms of a group of linked analyses which are, themselves, composed of multiple processes.

In software engineering terms, a testing laboratory can operate a waterfall model, because it is not dealing with unknowns, while a criminal investigation is more likely to have an iterative spiral model in operation precisely because of the unknowns that are present at the start, but which must be discovered and addressed during the investigative process. Indeed, in the early stages of an investigation, an argument could be put forward that some sort of agile model is in operation because of the need to make rapid progress on potentially very incomplete information. Indeed, in the context of software development, agile models are known to trade off time against functionality or completeness, particularly where initial requirements definition is difficult or will consume too much of the time available for the project[34][35].

At a basic level, we can characterise the difference between the testing laboratory and the forensic laboratory, as the presence of a need to carry out some form of search in order to determine what might be testable or examinable, before moving on to formulate a forensic strategy for the case. In "wet" forensic sciences, this would typically be a manual/visual search of clothing etc, to find fibres, body fluids, and other trace evidence. In the digital realm, it starts with a search to determine the applications and data present on the device, before considering which might have relevance to the case. In the digital realm, this level of initial search will be conducted by the FCA, possibly with some initial filtering applied by the DEFR or DES based on available information or standard practice.

Notwithstanding these issues, although a digital forensic laboratory may have to start an investigation based on incomplete end-user requirements (i.e. the precise remit of the investigation may be unclear and may only be clarified through the recovery, analysis and interpretation of digital artefacts), it should still be possible to identify those processes which the laboratory is

[3] Assume Nothing, Believe Nothing, Challenge Everything

[4] Who, What, Why, When, Where, How

both capable and willing to perform, and to provide adequate specifications and requirements for those processes to be verified. Ideally, such processes would follow the ISO/IEC 27042 model of atomicity in order to encourage greater re-use within internal analytical pipelines through proper consideration of cohesion and coupling.

1.4.3 Verification options

The work on verification options [4], although somewhat theoretical in nature, is grounded in existing models which have been shown to work. The mechanisms it proposes are pragmatic and, as described in the paper, do show that it is possible to provide evidence of verification against requirements without having to disclose detail of how those requirements are satisfied. In other words, commercially sensitive methods do not have to be disclosed as long as the "black box" can be shown to satisfy stated requirements. If this could be accepted by the vendor, user and regulatory communities, it potentially reduces the validation burden dramatically and would align the digital forensic community's validation practices more closely with those found in "weights and measures" and other similar testing bodies which use the same standards.

1.4.4 Language Issues

This author has never been entirely comfortable with the way the FSR and ILAC have used certain terms. This is particularly true of the concept of "verification" which the FSR and ILAC use to mean *checking that a validation remains true*. This concept, in the ISO/IEC 270xx group, was declared as "confirmation" in order to distinguish it from the software-engineering definition of verification as conformance to specified requirements. Coupled with the knowledge that the 270xx group was often seen as more applicable than 17025 to the digital forensics domain, this author chose to explore the language used throughout the 17025 standards and its successor documents, from ILAC-G19 to the FSR's guidance notes on digital forensic method validation, in order to see if the language itself might be a source of confusion. Again, this was a somewhat "Socratic" exercise, but involved a detailed analysis of the way certain key concepts, including *validation*, *verification* and *requirements*, were being used and described in these documents.

The outcome of this is that there is an indication that words have their meanings changed as they progress through the hierarchy of documents. This starts in ILAC-G19 [28] where the concept of verification becomes overloaded, either meaning *satisfying specified requirements* or *confirmation that validation remains valid* depending upon context. The distinction is never clearly spelled out so it is up to the reader to interpret it. In the latest FSR Statutory Code draft[13] this overloading appears to be even worse, and the concepts of validation, verification and confirmation appear to be freely interchangeable to a large extent.

Furthermore, when it comes to understanding requirements, the FSR codes [29][36] and guidance [30][31] place emphasis solely on the CJS as end-user, ignoring the intermediate stages of the investigative pipeline. Technical requirements are defined in terms of personnel qualifications etc. and make no mention of investigations, tools or FCAs.

1.4.5 Conclusions

The major conclusion from the work undertaken is that we may have a flawed implementation of the ISO 17025 standard when applied to all forensic sciences. The documents which are supposed to clarify how it should be applied may actually obfuscate the principles and lead to confusion and misapplication. Since this starts with the ILAC-G19 [28] overarching guidance document, there is potential for errors and misconception to have propagated down to all disciplines.

As the metrology exercise (1.3.2.1) showed, although there is clearly a need to know error rates and applicable conditions for a method in advance, these are essentially part of a verification process. The validation itself should be an exercise in showing that the methods chosen satisfy the customer's requirements and that the laboratory can perform those methods reliably and reproducibly. If we accept that as a true statement of how the standard should be applied, then the initial proposition - that evidence of verification can be used to underpin validation - automatically becomes true, as it already is in other laboratories which use this standard [37]. The difficulty arises in obtaining the evidence of verification. Until vendors are prepared, or required, to be more open about their products, the onus is placed on the user to verify the products that they use in their methods.

The work discussed in this chapter does, however, show that it should be possible to overcome these issues and establish a more efficient regime which can provide assurance of investigative methods in a lower-cost and more efficient way, but it will require a "reset" on the part of the Regulator - probably starting with withdrawal of the existing annexes and guides, and more rigorous editing of the new statutory code to ensure that concepts are expressed clearly and consistently. There is, potentially, an argument that the Regulator's use of codes and guidance documents has, in fact, led to more confusion and lower standards than might have been the case if DFUs had been left to interpret the standards for themselves.

Summarising against the original objectives:

1. *to explore the hypothesis that there is some fundamental difference between digital forensics and other forensic sciences which means that the extant quality standards cannot successfully be applied – including trying to understand why the 270xx group was seen as more directly applicable than the 1702x alternatives* The major difference appears to lie in the potential scale and complexity of a digital investigation, coupled with the rate of change of technology [16]. However, complexity does not equate to complicated, and rate of change of technology, although a technical challenge, is not an insurmountable barrier. In practice, digital forensic investigations have much in common with other forensic sciences, the primary difference being that it is, for the most, part possible to preserve and revisit the digital crime scene far more easily than the physical. This results in the potential for more evidence to be recovered and reviewed, potentially using multiple methods, in the digital realm. As the work on the ISO/IEC 270xx (1.3.1.1)[3][5] group of standards has shown, the primary problem may lie in the language used in the implementation chosen by the regulator.

2. *to evaluate the potential for cost savings possible through adoption of the ISO/IEC 27041 recommendations about tool verification as an adjunct to method validation* Here, comparison with metrology laboratories (1.3.2.1) as well as consideration of potential models for providing evidence of verification[4] , without violating confidentiality, suggests that this is possible in the context of a properly modularised investigation made up of properly specified atomic processes[2].

3. *to determine the most effective mechanism for tool vendors to provide support to customers seeking accreditation through disclosure of tool verification information.* Again, the work done suggests that this can

be done, but it requires engagement with the tool providers to determine which of the proposed methods is acceptable and cost-effective[3][4].

Of course, all of this is rendered somewhat moot by the case-law assumption that computers always work correctly and that their results, as long as they are presented by an expert, can always be relied upon. This is explored further in the supplementary discussion below (1.5), which goes some way towards a possible explanation of why the CJS itself, and tool providers and users, can currently afford to have little interest in the issue of tool verification.

1.5 Supplementary Discussion - Legal Issues

1.5.1 Legal issues

Forensic Science sits at the interface between science and justice and, therefore, must take account of legal/judicial requirements as well as the purely scientific. This also, oftentimes, necessitates consideration of the particular legal system in which the work is to be used. For this reason, much of the work under discussion is constrained by the English and Welsh legal system, based on the common law and the Acts of the UK's Parliament which apply to England and Wales.

To start, it may be useful to consider how computer-originated evidence has been treated historically.

The Police and Criminal Evidence Act (1984) [38] contained the following text:

```
69
Evidence from computer records.
(1) In any proceedings, a statement in a document produced by a
    computer shall not be admissible as evidence of any fact
    stated therein unless it is shown |
(a) that there are no reasonable grounds for believing
            that the statement is inaccurate because of improper
            use of the computer;
(b) that at all material times the computer was operating
            properly, or if not, that any respect in which it was
            not operating properly or was out of operation was not
            such as to affect the production of the document or
            the accuracy of its contents; and
(c) that any relevant conditions specified in rules of
            court under subsection (2) below are satisfied.

(2) Provision may be made by rules of court requiring that in any
    proceedings where it is desired to give a statement in evidence
    by virtue of this section such information concerning the
    statement as may be required by the rules shall be provided in
    such form and at such time as may be so required.
```

This was repealed in 1999 by section 60 of the Youth Justice and Criminal Evidence Act [39] which simply states

```
60 Removal of restriction on use of evidence from computer records.
Section 69 of the Police and Criminal Evidence Act 1984 (evidence
from computer records inadmissible unless conditions relating to
proper use and operation of computer shown to be satisfied) shall
cease to have effect.
```

The effect of this is that, since 1999, there has been a rebuttable presumption that computer systems always operate correctly, as they are intended to, and thus that evidence produced from them can be relied upon unless the case to the contrary can be shown. This is stated in the Criminal Justice Act (2003)[40] Section 129 ("Representations other than by a person") in subsection 2 where it declares that "Subsection (1) does not affect the operation of the presumption that a mechanical device has been properly set or calibrated.".

This author would argue that this is a problematic assumption, not least because of the developments in the modes of operation of computer systems since 1999 or, rather, since the cases which established the precedents upon which we currently rely for admissibility of computer-derived evidence.

1.5.2 Comments on admissibility of computer-derived evidence and evolution of computer systems

1.5.2.1 Precedents in English law

In 1990, the year that R. v. Spiby [41] was established as a precedent for the admissibility of computer-based records, the ARPANET ceased to exist as a distinct network, the EFF was created and most users' experience of networking relied on dial-up connections to bulletin boards rather than connection to the Internet. Indeed, it could be argued that it was not until the release of Windows 95 that Internet, and particularly WWW, use started to become truly popular as that operating system included IP functionality and a usable web browser as pre-installed components. Hobbes' Internet Timeline [42] tends to support this theory, showing a marked acceleration in web host registration and presence starting in late 1995.

Spiby [41], in particular, related to the admissibility of printouts of records of telephone calls (times and numbers dialled) made in a hotel, where those records had been generated automatically by the hotel's telephone management system. The legal argument related to presentation of hearsay evidence (i.e. the computer was not giving evidence itself, but a human being was reporting what the computer had recorded) and the Court of Appeal upheld the original view that it was permissible for a human being to report information which had automatically been recorded by a computer.

Similarly, in R. v. Shepherd [43] the House of Lords held that it was acceptable for a store detective to give evidence about, and based on, till rolls without further evidence, from a computer expert for example, that the tills were working correctly. The ruling in this case also mentions that

> In Reg. v. Minors it is stated, at p. 446: ”to the extent to which a computer is merely used to perform functions of calculation, no question of hearsay is involved, and the requirements of sections 68 and 69 do not apply: Reg. v. Wood (1982) 76 Cr.App.R. 23 and Sophocleous v. Ringer [1988] R.T.R. 52.”

Lord Griffiths went on to state

> “Documents produced by computers are an increasingly common feature of all business and more and more people are becoming familiar with their uses and operation. Computers vary immensely in their complexity and in the operations they perform. The nature of the evidence to discharge the burden of showing that there has been no improper use of the computer and that it was operating properly will inevitably vary from case to case. The evidence must be tailored to suit the needs of the case. I suspect that it will very rarely be necessary to call an expert and that

> in the vast majority of cases it will be possible to discharge the burden by calling a witness who is familiar with the operation of the computer in the sense of knowing what the computer is required to do and who can say that it is doing it properly.
>
> The computer in this case was of the simplest kind printing limited basic information on each till roll. The store detective was able to describe how the tills operated, what the computer did, that there had been no trouble with the computer and how she had also examined all the till rolls which showed no evidence of malfunction either by the tills or the central computer. "

This author would argue that the critical phrases in this ruling are **"The computer in this case was of the simplest kind"** and **"a witness who is familiar with the operation of the computer..."**. A cash register is little more than a simple arithmetic calculator with a printing function and any errors in its operation would probably be fairly quickly spotted by either operator or customer, especially when such machines are in operation in multiple locations concurrently and not least because checking the output of such a simple computer is something which most people can carry out for themselves.

Even in the case of Sophocleous v. Ringer [44] where the debate related to the admissibility or use of a graph produced by a spreadsheet used to automate blood-alcohol calculations, the computer was performing a simple function which could be easily replicated by the human analyst in a reasonable amount of time.

None of the systems under consideration would have been connected to any form of permanent Internet connection, because such connections were rare, expensive and unnecessary at the time and none would have required any form of multi-tasking operating system to perform their functions.

1.5.3 Evolution and development of computer systems

If we accept the proposition that 1995 marks the start of a period of growth in Internet-connected systems, with the appearance of low-cost xDSL, WiFi and mobile IP data services continuing this trend to a point where, today, most software can reasonably expect to be able to "call home" periodically, or even rely upon a network connection for its functioning, we must also accept that this represents a fundamental shift in how software operates.

From being self-contained, with all resources required being present at time of installation and upgrades controlled entirely by the user (e.g. by running upgrade or patching programs from disk), we now have a situation where much software relies on resources being somewhere "in the cloud" with the process of upgrading or patching being under the control of the manufacturer or operating system rather than under the direct control of the user.

Growth in processor capability, memory and storage capacity, and a desire to do more on each device has led to widespread adoption of multi-tasking systems with even the simplest smartphone being capable of running multiple programs concurrently. This, of course means that computers have moved away from relying on simple executives such as MS-DOS to relying on full-fledged operating system which include more and more functionality to service the needs of the applications. (for example, Linux 1.0.0 from 1994 is claimed to have consisted of just 176250 lines of code (LOC)[45], which grew to 27.8 million by 2020[46]. Of course, some of this reflects the sheer number of device drivers which are required to support current and legacy hardware, but it also represents an increasingly complex and difficult to test and maintain product.) An operating system might be thought a somewhat extreme example, but it is rare for any software to become smaller over time - rather it is the norm, especially in the field of digital forensics, for popular products to be expected to offer new capabilities and new functions with each new release. Simply releasing a functionally equivalent, albeit more efficient or faster, product rarely seems to satisfy the market, or marketers.

We have also seen changes in fashion relating to development methods, moving away from rigid (and allegedly rigorous) waterfall-type models, where requirements capture was presumed possible before coding could be allowed to start, to more "flexible" and agile models where the use of prototypes, partial solutions, user stories, scenarios and experimental versions is used to allow customers/users and developers to, it is claimed, develop a better understanding of each others proposals, capabilities and needs[47].

1.5.4 The current situation

The Crown Prosecution Service (CPS) guidance [48] even goes so far as to note that "The repeal of section 69 of the Police and Criminal Evidence Act 1984 will be particularly helpful in cases where chains of linked computers are involved, for example internet fraud", implying that the challenge of proving correct operation of multiple interconnected systems could be extremely difficult within the constraints of the Criminal Justice System.

It is worth noting, also, that the CPS, in its guidance on Scientific Evidence [49], suggests that all Forensic Science Providers (FSPs) should comply with the Forensic Science Regulator's (FSR) codes, but then concentrates solely on the issue of DNA evidence, be it DNA17 or High Sensitivity DNA analysis, and the use of Streamlined Forensic Reporting.

The CPS guidance on the use of computer records as evidence, making reference to Archbold[50] and the relevant case law (mentioned above), highlights three particular scenarios:

1. Computer is used as a calculator to process information

2. Information that the computer has been programmed to record

3. Information recorded and processed by the computer which has been entered by a person, whether directly or indirectly

The third category remains hearsay as the computer simply reports what it has been told by a human being, but the first two categories are somewhat challenging in the context of current computer systems and, in particular, in relation to digital forensics and the Forensic Science Regulators codes.

The case law which establishes the first two situations dates back to 1980, 1983, 1988, 1990 and 2004 with the later dates referring to appeals which upheld the principles established in the earlier cases, and establishing them as precedents for future trials. Thus we are, in effect, dealing with legal principles founded on understanding of how computer systems worked and were applied in the 1980s as discussed above.

Notwithstanding, these principles weaken the FSR's requirement for forensic science methods to be validated. By allowing an assumption that computer systems always work correctly, and as intended, they remove the requirement for computer systems to be verified prior to use in a forensic science context – i.e. an unproven computer tool may be used in a validated method solely because of the assumption provided for by the common law. In digital forensics, especially, this seems to be somewhat inadvisable.

In order to understand why the author makes this assertion, it may be useful to consider some typical definitions of cybercrime categories and consider how they are paralleled in other situations.

Taking the UN ODC definitions[51], there are 3 main categories:

1. Cyber-dependent - crimes which require the presence of ICT infrastructure and which may target elements of it in order to have their desired effect

2. Cyber-enabled - crimes which can occur offline, but which can be facilitated by ICT

3. Child sexual exploitation and abuse

The first two of these categories also exist in Marshall, Moor and Tompsett's (MMT) taxonomy[52] which identifies the following roles of cyber-systems in criminal acts:

- Witness - observes but is not directly involved in the criminal act.

- Tool – provides assistance with the crime, but it can be carried out by other means.

- Accomplice – is essential to the successful conduct of the crime.

- Victim – is a direct target of the crime.

- Guardian – attempts to prevent the crime occurring.

At a minimum, the UN Cyber-dependent category matches MMT's Accomplice but may include elements of Witness, Guardian and Victim, while the Cyber-enabled category maps to MMT's Tool category.

The major difference between the Dependent/Accomplice definition and the Enabled/Tool definition is the degree to which the "cyber" or computer element is necessary for the commission of the crime.

If we consider this, in light of the situation in modern digital forensics, the sheer volume and complexity of systems which must be examined means that what we call *tools* are more or less essential assistants, or *accomplices* - i.e. although in theory we can replicate the processes carried out by those assistants through human/manual efforts, the time required to complete these would be so high that the delays incurred would be unacceptable, as would the resultant costs in the vast majority of cases. Therefore, throughout this document, a distinction has been drawn between forensic *tools*, which carry out simple tasks, and *"forensic cyber-assistants"* or *FCA*s, which complement their own tool functionality by adding analytical/interpretive capabilities which allow them to correlate and integrate data from multiple sources in order to produce reports.

The term, FCA, has been chosen because the commonly used digital forensic systems now far surpass the requirements of tools, as established by case law. They are no longer "used simply as a calculator to process information" [48] but now encapsulate, themselves, expert knowledge which is used to parse and present information, frequently pulling data from multiple sources (filesystem, database tables, browser cache etc.) in order to present it in a way which aids understanding. In this respect, therefore, they are carrying out multiple operations, based on reverse-engineering of software by some human expert who has provided input into their development processes. Furthermore, because these tools are capable of processing data in multiple different storage formats (filesystems, databases, file formats related to multiple operating systems and applications), they encapsulate the knowledge of multiple experts. This, this author would argue, is far from the original case which established the precedent - i.e. the use of a computer to perform blood-alcohol calculations[44].

Examining this in more detail, the "computer as tool" precedent is based upon the use of a more general piece of software, presumably a general-purpose calculator or spreadsheet, to carry out the type of calculation for which it was designed and for which many thousands or millions of other users had already successfully used it. Within this, there is an assumption that any errors in the calculator would have been identified, reported and corrected or publicised.

Ladkin, et al.[53], have discussed this issue in detail and highlighted the flaws in the assumptions in light of known estimates of defect rates in typical code and historic lack of identification of defects through normal use. Their assessment tends towards the position that evidence obtained from computers through their "normal" operation cannot be considered reliable, and that the second assumption, in particular is not valid. In light of developments[54] re the Post Office Horizon case[55], this may well be applicable and a review of the common-law and case-law position might result in a change to the position. Whether this would affect the "computer as tool" assumption is open to debate, and may depend on the complexity of the "tool" in question.

FCAs, however, are far more specialised and so less widely used, and their operation is far less well-understood by users. They are, therefore, subject to considerably less "field testing" than is the the case for any general purpose software product. For this reason, the scenario 1 presumption is, this author would argue, demonstrably inappropriate and outdated. Computer systems have evolved dramatically since the 1980s, as acknowledged by the

CPS comment about complex systems, and reliance on case precedents set in the time of 8-bit or even 16-bit uniprocessing systems is no longer “fit for purpose”.

If evidence is to be produced from any specialist or bespoke software, then it is unlikely that a FCA will have sufficient capability to progress much beyond the acquisition and preservation stages of the ISO/IEC 27037 and ISO/IEC 27042 ICAPAIR (Identify, Collect, Acquire , Preserve, Analyse, Interpret, Report) model, but for other systems it will be carrying out some analytical, interpretive and reporting functions.

For example, a modern forensic kiosk system may extract data from a mobile phone, then process the various databases present in the phone in order to associate contact details with communications records from several apps, in order to produce a timeline of communications, including deleted messages. As kiosks are deployed for use by non-specialists, the kiosk system itself takes on the role of the Digital Evidence Specialist (DES), as it is defined in ISO/IEC 27037:2016[19].

If we assume that the situation re the presumption will change, and in light of the Horizon case and potential challenges to computer evidence in Operation Venetic[56] trials it may be forced to, then the legislation relating to[11], and the role of, the Forensic Science Regulator become more important.

Venetic is included, here, as trials appear to be proceeding on the basis of the “normal operation” presumption. This author has suggested, in his capacity as an advisor to the legal profession, that this is unsafe because evidence has already been given that the systems in question had been modified, by non-UK law enforcement agencies, in order to extract data – i.e. the systems cannot have been operating normally because of the interference of external agencies. Furthermore, if a law-enforcement agency could “hack” those systems in order to obtain evidence, it is entirely feasible that some other parties could have interfered with the systems at some point in time. It is, therefore, suggested that additional evidence of correct operation or, at the very least, evidence of verification of the deployed “hack(s)” is required to render the resulting evidence admissible.

This situation is not unique to Venetic. Indeed, the ACPO Good Practice Guide[57] recognises, in Principle 2, that in some situations it may be necessary to alter the data stored on a device, potentially affecting its operation, in order to obtain evidence from it. This is particularly true in the case of personal devices such as smartphones and tablets where manufacturer techniques to protect user privacy have the effect of making access to data difficult. As a result, “tools” such as GrayShift’s GrayKey[58] are now routinely used in the digital forensic units’ laboratories but rarely mentioned or discussed in open court. One could suggest that this lack of disclosure is contrary to the principles of good justice, but that it is caused by two main factors :

1. Secrecy on the part of the provider – i.e. the method used is commercially sensitive and gives them a competitive edge in the market place

2. lack of understanding on the part of the user – i.e. because of manufacturer secrecy (and other factors), the users of these systems are unable to fully comply with ACPO Principle 2 in that they cannot provide a detailed explanation of the changes caused by the use of the tool and, therefore, cannot give evidence as to its reliability or completeness

The latter explanation may seem somewhat cynical but, as Marshall and Paige discovered[3] in their attempts to obtain information, about tool testing and specifications, from a range of providers, there is a dearth of evidence of fitness for purpose, or even real definitions of purpose, for FCAs generally.

References

[1] ISO/IEC, "ISO/IEC 27041:2016 guidance on assuring the suitability and adequacy of digital investigation method," 2016.

[2] ISO/IEC, "ISO/IEC 27042:2016 guidelines for the analysis and interpretation of digital evidence," 2016.

[3] A. M. Marshall and R. Paige, "Requirements in digital forensics method definition: Observations from a uk study," *Digital Investigation*, vol. 27, pp. 23–29, 2018.

[4] A. M. Marshall, "Digital forensic tool verification: An evaluation of options for establishing trustworthiness," *Forensic Science International: Digital Investigation*, vol. 38, p. 301181, 2021.

[5] A. M. Marshall, "The unwanted effects of imprecise language in forensic science standards," *Forensic Science International: Digital Investigation*, vol. 40, p. 301349, 2022.

[6] C. Hargreaves and A. Marshall, "Synctriage: Using synchronisation artefacts to optimise acquisition order," *Digital Investigation*, vol. 28, pp. S134–S140, 2019.

[7] A. M. Marshall and P. Miller, "Casenote: Mobile phone call data obfuscation & techniques for call correlation," *Digital Investigation*, vol. 29, pp. 82–90, 2019.

[8] A. M. Marshall, "Whatsapp server-side media persistence," *Digital Investigation*, vol. 25, pp. 114–115, 2018.

[9] "The Accreditation of Forensic Science Providers Regulations 2018." online at https://www.legislation.gov.uk/uksi/2018/1276/made, 2018, in force 2019. last accessed 6th September 2022.

[10] "The Accreditation of Forensic Service Providers (Amendment) Regulations 2019." online at https://www.legislation.gov.uk/uksi/2019/1384/made, 2019. last accessed 6th September 2022.

[11] "The Forensic Science Regulator Act 2021." online at https://www.legislation.gov.uk/ukpga/2021/14/contents/enacted, 2021. last accessed 6th September 2022.

[12] "The Forensic Science Regulator Act 2021 (Commencement No. 1 and Transitional Regulators 2022." online at https://www.legislation.gov.uk/uksi/2022/856/contents/made, 2022. last accessed 15th September 2022.

[13] Forensic Science Regulator, "Code of practice consultation draft (accessible version)." online at https://www.gov.uk/government/consultations/forensic-science-draft-statutory-code-of-practice/code-of-practice-consultation-draft-accessible-version, 2022. last accessed 14th September 2022, published 8th August 2022.

[14] ISO/IEC, "ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories," 2017.

[15] European Council, "Council Conclusions on the vision for European Forensic Science 2020 including the creation of a European Forensic Science Area and the development of forensic science infrastructure in Europe." online at https://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/jha/126875.pdf, 2011. last accessed 7th March 2022.

[16] P. Sommer, "Forensic science standards in fast-changing environments," *Science & Justice*, vol. 50, no. 1, pp. 12–17, 2010. Special Issue: 5th Triennial Conference of the European Academy of Forensic Science.

[17] "Iso 17025 - digital forensic struggle." online at https://www.forensicfocus.com/forums/general/iso-17025-digital-forensics-struggle/. last accessed 15th September 2022. This is one example of many similar discussions.

[18] A. M. Marshall, S. Higham, and T. Dyhouse, "Digital Forensics Capability Review," June 2013. available online at https://www.researchgate.net/publication/269332581_Digital_Forensics_Capability_Review.

[19] ISO/IEC, "ISO/IEC 27037:2016 guidelines for identification, collection, acquisition and preservation of digital evidence," 2016.

[20] ISO/IEC, "ISO/IEC 17025:2005 General requirements for the competence of testing and calibration laboratories," 2005.

[21] iISO/IEC, "ISO/IEC 17000:2020 Conformity assessment — Vocabulary and general principles," 2020.

[22] UKAS, "Directory of accredited organisations." online at https://www.ukas.com/services/other-services/directory-of-accredited-organisations/, 2018. last viewed 4th June 2018.

[23] IEEE, "IEEE 730-2104 Standard for Software Quality Assurance Processes," 2014.

[24] ISO/IEC, "ISO/IEC 12207:2008 Systems and software engineering — Software life cycle processes," 2008.

[25] ISO/IEC/IEEE, "ISO/IEC/IEEE 15289:2011 systems and software engineering — content of life-cycle information products (documentation)," 2011.

[26] Trustworthy Software Foundation, "Trustworthy software specification document." online at https://tsfdn.org/wp-content/uploads/2016/03/TS502-0-TS-Essentials-Specification-Issue-1.2-WHITE.pdf, 2016. last accessed, 24th February 2021.

[27] Trustworthy Software Foundation, "Trustworthy software guidance document." online at https://tsfdn.org/wp-content/uploads/2016/03/TS502-0-TS-Essentials-Guidance-Issue-1.2-WHITE.pdf, 2016. last accessed, 24th February 2021.

[28] International Laboratory Accreditation Cooperation, "ILAC G19:08/2014 Modules in a Forensic Science Process," 2014.

[29] Forensic Science Regulator, "Forensic Science Regulator: Codes of Practice and Conduct for forensic science providers and practitioners in the Criminal Justice System (FSR-C-100, Issue 5)," 2020.

[30] Forensic Science Regulator, "Codes of Practice and Conduct Appendix: Digital Forensic Services (FSR-C-107, Issue 2)," 2020.

[31] Forensic Science Regulator, "Forensic Science Regulator Guidance: Method Validation in Digital Forensics (FSR-G-218, Issue 2)," 2020.

[32] B. Du Bois, S. Demeyer, and J. Verelst, "Refactoring - improving coupling and cohesion of existing code," in *11th Working Conference on Reverse Engineering*, pp. 144–151, 2004.

[33] Crown Prosecution Service (CPS), "The Criminal Justice System." https://www.cps.gov.uk/about-cps/criminal-justice-system. last accessed 7th March 2022.

[34] J. Sutherland, "Future of scrum: parallel pipelining of sprints in complex projects," in *Agile Development Conference (ADC'05)*, pp. 90–99, 2005.

[35] P. Abrahamsson, O. Salo, J. Ronkainen, and J. Warsta, "Agile software development methods: Review and analysis," *CoRR*, vol. abs/1709.08439, 2017.

[36] Forensic Science Regulator, "Forensic Science Regulator: Codes of Practice and Conduct for forensic science providers and practitioners in the Criminal Justice System (FSR-C-100, Issue 7)," 2021.

[37] R. Trishch, O. Maletska, H. Hrinchenko, S. Artiukh, V. Burdeina, and N. Antonenko, "Development and validation of measurement techniques according to iso/iec 17025:2017," in *2019 IEEE 8th International Conference on Advanced Optoelectronics and Lasers (CAOL)*, pp. 1–6, 2019.

[38] "Police and Criminal Evidence Act (1984."

[39] "Youth Justice and Criminal Evidence Act(1999)."

[40] "Criminal Justice Act (2003)."

[41] "R. v. Spiby." available online at https://vlex.co.uk/vid/r-v-spiby-803092385, 1990. [1990] 91 Cr App R 186 CA.

[42] Robert HóbbesŹakon, "Hobbes' internet timeline 25." online at https://www.zakon.org/robert/internet/timeline. Last accessed 1/Jun/2022, last modified 1/Jan/2018.

[43] "R. v. Shepherd." available online at https://vlex.co.uk/vid/r-v-shepherd-792803861. [1992] HL 16 Dec 1992.

[44] "Sophocleous v. Ringer." available online at https://swarb.co.uk/sophocleous-v-ringer-1988/, 1988. [1988] R.T.R. 52.

[45] "History of the Linux Kernel." online at https://linuxhint.com/history-linux-kernel/. last accessed 15th September 2022.

[46] S. Bhartya, "Linux in 2020: 27.8 million lines of code in the kernel, 1.3 million in systemd." online at https://www.linux.com/news/linux-in-2020-27-8-million-lines-of-code-in-the-kernel-1-3-million-in-systemd/, 2020. last accessed 15th September 2022, published 7th January 2020.

[47] E.-M. Schön, J. Thomaschewski, and M. J. Escalon, "Agile requirements engineering: A systematic literature review," *Computer Standards and Interfaces*, vol. 49, pp. 79–91, 2017. https://doi.org/10.1016/j.csi.2016.08.011.

[48] Crown Prosecution Service (CPS), "Computer Records Evidence - Legal Guidance." https://www.cps.gov.uk/legal-guidance/computer-records-evidence. last accessed 13th September 2022, last reviewed by CPS 30th June 2017.

[49] Crown Prosecution Service (CPS), "Scientific Evidence - Legal Guidance." https://www.cps.gov.uk/legal-guidance/scientific-evidence. last accessed 13th September 2022, last updated 8th November 2019.

[50] *Archbold: criminal pleading, evidence and practice.* Sweet and Maxwell, 2022.

[51] United Nations Office on Drugs and Crime, "Cybercrime - global programme on cybercrime." online at https://www.unodc.org/unodc/en/cybercrime/global-programme-cybercrime.html. last accessed 3rd March 2022.

[52] A. M. Marshall, G. Moor, and B. Tompsett, "Criminalisation of the internet : an examination of illegal activity online," EAFS 4th Conference, 2006.

[53] P. B. Ladkin, B. Littlewood, H. Thimbleby, and M. Thomas, "The Law Commission presumption concerning the dependability of computer evidence," *Digital Evidence and Electronic Signature Law Review (DEESLR)*, vol. 17, 2020.

[54] P. Marshall, "Scandal at the Post Office: The Intersection of law, ethics and politics," *Digital Evidence and Electronic Signature Law Review (DEESLR)*, vol. 19, 2022. https://doi.org/10.14296/deeslr.v19i0.5395.

[55] T. McCormick, "The Post Office Horizon system and Seema Misra," *Digital Evidence and Electronic Signature Law Review (DEESLR)*, vol. 13, 2016. https://doi.org/10.14296/deeslr.v13i0.2.

[56] National Crime Agency (NCA). online at https://www.nationalcrimeagency.gov.uk/news/operation-venetic. last accessed 4th March 2022.

[57] ACPO, "ACPO Good Practice Guide for Digital Evidence." online at https://library.college.police.uk/docs/acpo/digital-evidence-2012.pdf, 2012. last accessed 24th February 2021.

[58] "GrayKey. Designed to make a difference.." online at https://www.grayshift.com/graykey/. last accessed 15th September 2022.

Appendix A

Confirmation of contribution to ISO/IEC standards

WARWICK
THE UNIVERSITY OF WARWICK

WMG
Innovative Solutions

Cyber Security Centre

2nd Floor Phase 3 IMC Building
University Road
Coventry
CV4 7AL
England

Department of Computer Science
University of York
Campus East
Deramore Lane
Heslington
York
YO10 5GH

Our Ref.: UCR/080391

Your Ref.:

Date: 8 February 2022

RECOGNITION OF EXPERT CONTRIBUTION TO STANDARDISATION

I am pleased to confirm that as part of his voluntary role with the Committee BSI IST/33/4 “Information Security – Controls and Services”. Mr. Marshall was both the Lead UK Expert, and ISO/IEC Project Editor, for the International Standards:

- ISO/IEC 27041 “Information technology — Security techniques — Guidance on assuring suitability and adequacy of incident investigative method”
- ISO/IEC 27042 “Information technology — Security techniques — Guidelines for the analysis and interpretation of digital evidence”

In those capacities, he drafted and co-ordinated production of the content of these projects, and I estimate his personal effort to be the predominant contribution to the final content of the published documents.

Should you require any further information, please feel free to contact me.

Professor I R C Bryant
Principal Investigator - Understanding Cyber Risk
& Chair, BSI Expert Committee IST/033/4 – Information Security – Controls and Services
e:ib@warwick.ac.uk
t:+44-24-769-51924

Appendix B

ISO/IEC 27041

Note: Because of licencing conditions it not possible to include a copy of this standard.

Appendix C

ISO/IEC 27042

Note: Because of licencing conditions it not possible to include a copy of this standard.

Appendix D

Confirmation of contribution to Requirements paper

McMaster University | ENGINEERING

Prof. Richard Paige,
Department of Computing and Software
1280 Main Street West
Hamilton, ON L8S 4L8

☎ (905) 525-9140 x 26627
🖷 (905) 525-9140
✉ paigeri@mcmaster.ca
🌐 www.cas.mcmaster.ca/paige

29 November 2022

To Whom It May Concern,

I acted as PI and budget holder for the University of York internally funded project 'Digital Forensic Tool Verification and its role in method validation' which produced the paper *Requirements in digital forensics method definition: Observations from a UK study* (published in *Digital Investigations*, volume 27, pages 23-29, 2018).

Mr. Marshall led the practical work, with my assistance as mentor and advisor. I confirm that he is primary author of the paper, and that he contributed at least 95% of the text with input from me in an editorial and advisory capacity once the initial drafting was complete.

Please don't hesitate to contact me if you need anything further.

Yours faithfully

Professor Richard Paige, FBCS, CEng
Joseph Ip Distinguished Engineering Professor
Director, McMaster Centre for Software Certification
Associate Chair (Research)
Department of Computing and Software
McMaster University
Hamilton, Ontario, Canada
paigeri@mcmaster.ca

Appendix E

Requirements in Digital Forensics

Contents lists available at ScienceDirect

Digital Investigation

journal homepage: www.elsevier.com/locate/diin

Requirements in digital forensics method definition: Observations from a UK study[☆]

Angus M. Marshall[*], Richard Paige

Dept. of Computer Science, University of York, UK

ARTICLE INFO

Article history:
Received 20 June 2018
Received in revised form
28 August 2018
Accepted 12 September 2018
Available online 20 September 2018

Keywords:
ISO 17025
ISO 27041
Quality standards
Method validation
Tool verification
Forensic tool development

ABSTRACT

During a project to examine the potential usefulness of evidence of tool verification as part of method validation for ISO 17025 accreditation, the authors have examined requirements statements in several digital forensic method descriptions and tools. They have identified that there is an absence of clear requirements statements in the methods and a reluctance or inability to disclose requirements on the part of tool producers. This leads to a break in evidence of correctness for both tools and methods, resulting in incomplete validation. They compare the digital forensics situation with other ISO 17025 accredited organisations, both forensic and non-forensic, and propose a means to close the gap and improve validation. They also review existing projects which may assist with their proposed solution.

Introduction

ISO/IEC 27041 (ISO/IEC, 2016), as part of a group of standards dealing with digital investigations, is the standard which describes a process by which a method can be shown to be fit for its intended purpose. To achieve this, it proposes a process for the validation of methods used in a digital investigation. Within the description of validation it suggests that evidence of a tool's verification against a declared set of requirements can be used as means to reduce the amount of validation required for processes in which the tool participates. i.e. it suggests that those process requirements which are wholly satisfied by the tool, and for which evidence of verification exists, need not be subjected to further testing.

Note: in this project we have concentrated solely on the validation and verification issue. The other standards in the group propose models of evidence gathering and processing which. although useful, are not considered core issues for this work.

From the perspective of software engineering the proposal in ISO/IEC 27041 (ISO/IEC, 2016) is entirely acceptable. However, for such a mechanism to succeed, the tool and the process in which it participates must be specified in terms of requirements which can be mapped against each other to show how the tool conforms to, or partially fulfills, the requirements of the process.

In effect, the proposal is that there is some degree of overlap between tool requirements and method requirements, ranging from the possibility that a tool's requirements are a complete subset of a method's requirements (Fig. 1) to the, potentially, less likely situation where a method's requirements are a subset of a tool's (Fig. 2).

In practice, because some of the requirements for a method with an investigative context will be non-technical in nature, it is believed that the most common situation will be that shown in Fig. 3, where a tool's requirements intersect with those of a method, and only those tool requirements lying in the intersection are relevant to the validation of the method.

During research into how this mechanism could be applied in practice, particularly to allow producers of tools for digital forensic processes to support their customers' compliance with ISO 17025's[1] validation requirement (ISO, 2005a), through disclosure of evidence of testing and without compromising commercially sensitive information such as details of test data, the authors have found that such a mapping appears, at the time of writing, to be impossible to

[☆] This work was supported by the University of York Research Priming Fund.

[*] Corresponding author.
E-mail address: angus.marshall@york.ac.uk (A.M. Marshall).

[1] In this document we concentrate on the use of ISO 17025:2005 as the currently deployed standard. We consider the implications of transition/update to the 2017 version in the Conclusions of this document.

https://doi.org/10.1016/j.diin.2018.09.004
1742-2876/© 2018 Elsevier Ltd. All rights reserved.

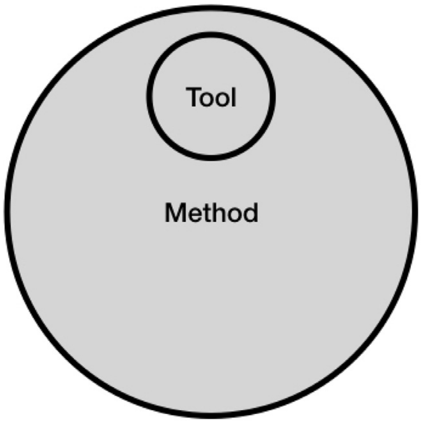


Fig. 1. Tool requirements are a subset of method. Typical of specialist tools or small tools produced to assist with part of a method(Shaded area = the set of requirements which much be satisfied for validation.)

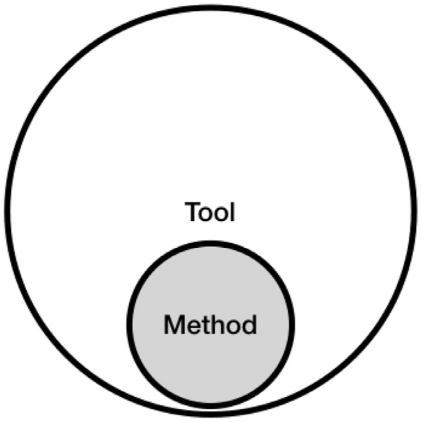


Fig. 2. Method requirements are a subset of tool. Considered rare, but possible where a method exactly follows a process defined by the tool producer and uses only a subset of the tool functionality(Shaded area = the set of requirements which much be satisfied for validation.)

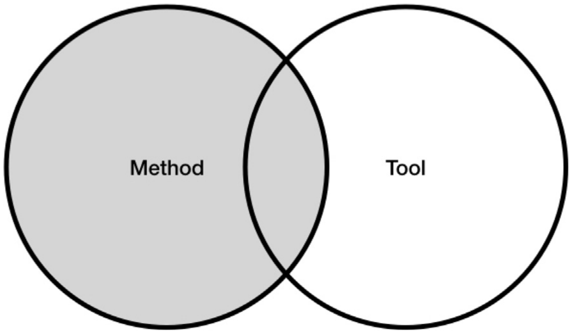


Fig. 3. Tool requirements intersect with the method. Common where the tool fulfils some or all of the technical requirements, but there are other non-technical requirements to be satisfied(Shaded area = the set of requirements which much be satisfied for validation.)

perform. This is because it has proved impossible to obtain the necessary levels of information about requirements from any of the participants in the study. Two main factors appear to affect this:

- Firstly, the process definitions examined in our study do not contain any technical requirements which can be mapped. Rather, they contain primarily non-technical requirements aligned to the needs of the Criminal Justice System.
- Secondly, the tool producers are either unable (in the case of most small providers) or unwilling (in the case of most larger providers) to provide information about how they capture customer requirements, let alone disclose what those requirements are.

Some even went as far as responding to the request for information with statements such as “The information you seek is commercially sensitive as we operate in a very competitive landscape. Unfortunately, we can't give out any specifics on our product development techniques to third parties.” The authors struggle to understand this type of response as our questions related to high-level development models and requirements capture methods rather than specific details of implementation of tools or tests. We can only surmise that the tool providers who responded in this way either lack confidence in their own products or believe that they are using innovative development techniques which no other developer has considered.

Principles of ISO 17025

Before examining the concept of validation more closely, it may be helpful to review some of the principles underpinning ISO 17025 which are embodied in the earlier version and which have influenced its use in “non-forensic” organisations such as those carrying out calbration of tools or testing of chemical compounds or metal alloys.

Gravel (2002), writing in 2002 about the 1999 version of ISO 17025 described 8 principles which were embodied within the standard as:

Capacity: Concept that a laboratory has the resources (people with the required skills and knowledge, the environment with the required facilities and equipment, the quality control, and the procedures) in order to undertake the work and produce competent results.

Exercise of responsibility: Concept that persons in the organisation have the authority to execute specific functions within the overall scope of work and that the organisation can demonstrate accountability for the results of the work.

Scientific method: Concept that the work carried out by the organisation is based on accepted scientific approaches, preferably consensus-based, and that any deviations from accepted scientific approaches can be substantiated in a manner considered generally acceptable by experts in that field.

Objectivity of results:

1. Concept that the results produced within the scope of work of the organisation, are mainly based on measurable or derived quantities.
2. Concept that subjective test results are produced only by persons deemed qualified to do so and that such results are noted as being subjective, or are known by experts in that field of testing to be mainly subjective.

Impartiality of conduct: Concept that the pursuit of competent results through the use of generally accepted scientific approaches is the primary and overriding influence on the work of persons executing tests - all other influences being considered secondary and not permitted to take precedence.

Traceability of measurement:

1. Concept that the results produced, within the scope of work of the laboratory, are based on a recognised system of measurement that derives from accepted, known quantities (SI system) or other intrinsic or well-characterised devices or quantities.

2. Concept that the chain of comparison of measurement between these accepted, known quantities or intrinsic devices or quantities, and the device providing the objective result, is unbroken for the transfer of measurement characteristics, including uncertainty, for the whole of the measurement chain.

   Repeatability of test: Concept that the test which produced the objective results, will produce the same results, within accepted deviations during subsequent testing, and within the constraints of using the same procedures, equipment and persons used during a previous execution of the test.

   Transparency of process: Concept that the processes existent within the laboratory producing the objective results, are open to internal and external scrutiny, so that factors which may adversely affect the laboratory's pursuit of objective results based on scientific method, can be readily identified and mitigated.

With the exceptions of Capacity and Exercise of responsibility, these principles establish a need to show, not just that a chosen method satisfies requirements for an intended use, but that the method is fundamentally correct or sound, and satisfies broader ranging technical requirements.

From our reviews of both the 2005 and 2017 versions of ISO 17025, it appears that these principles have been retained in the most recent versions of the standard.

Application of ISO 17025:2005 to “non-forensic” disciplines

A regularly voiced criticism of ISO 17025 is that it is, as its title suggests, intended for Testing and Calibration laboratories. In order to understand how ISO 17025 is applied in these “non-forensic” organisations, and to determine if or how it is applied differently in a forensic context, the authors carried out a review of publicly available accreditation records.

The United Kingdom Accreditation Service (UKAS) maintains a register of accredited bodies (UKAS, 2018) which is open for public inspection. The entries in this register include detail of each test for which a body has been accredited, giving a brief description of the method used where appropriate or necessary.

Examination of a sample of 100 accredited organisations in a range of “non-forensic” and “non-medical” areas reveals that these organisations apply two approaches to defining the requirements for their accredited process:

> **Physical properties**: Where precise measurement of physical properties is possible (e.g. for volumetric, force, torque, acoustics), the schedules of accreditations specify, using SI units, the range of measurement possible and tolerances (uncertainty) allowed for that measurement.
>
> **External standards**: In other circumstances, where an industry has defined its own standards, the accreditation is based on implementation of the published standard which either defines the range and uncertainty for the measurement, or defines the method itself.

In both of these cases, the requirements for the method, and thus its validation, are available in published form (either directly in the schedule of accreditation or in the published standard) and thus can be subjected to independent scrutiny and adopted by others practicing in the same technical field. In fact, the published requirements allow an independent verification of the method to show correctness in the form of conformance to a general set of standardised requirements rather than just conformance to the requirements for a particular use-case.

Moreover, the presence of these published criteria allow customers to identify those testing bodies whose methods may satisfy their needs before entering into discussions with the testing body. In effect, the listed requirements and associated tests become a menu from which the customer and test body can choose the most appropriate way of meeting the customer's particular needs.

A discussion of validation

In many discussions of accreditation against the standard, the concept of “validation of the tool” or even “tool accreditation” is raised by users and vendors as a means to shortening or eliminating the process. To the authors, this hints that there may be some either confusion about the meanings of these terms, or a different use of language in effect. It is, therefore, instructive to consider the software engineering distinction between verification and validation and contrast it with the ISO 17025 view.

ISO 17025:2005 approach to validation

ISO 17025:2005 (ISO, 2005a) contains no direct definition of validation but, in accordance with ISO practice, refers the reader to ISO 17000 and ISO 9000 for inheritance of relevant definitions. This practice, of relying on definitions found in other standards, is common with the ISO range of standards, but can cause problems for some users as they may perceive a requirement to have access to the defining standard as well as the standard they are trying to implement, or they may rely solely on common usage of the word as opposed to ISO's stipulative definitions (aka the “Humpty Dumpty” rule[2]). In practice, ISO provides an Online Browsing Platform (ISO, 2018) (OBP) which allows access to definitions and some other text without further expenditure.

Using the OBP, the authors have found that ISO 17000 contains no definition of validation. Thus the ISO 9000:2005 (ISO, 2005b) definition should be used as this is the most recently published version prior to the publication of ISO 17025:2005. This gives the following definition of validation:

> "Confirmation, through the provision of objective evidence, that requirements for a specific intended use or application have been fulfilled.
> NOTE 1 The term validated is used to designate the corresponding status.
> NOTE 2 The use conditions for validation can be real or simulated."
>
> and defines objective evidence as
>
> "Data supporting the existence or verity of something
> NOTE: Objective evidence may be obtained through observation, measurement, test, or other means."
>
> with requirement as
>
> "need or expectation that is stated, generally implied or obligatory
> Note 1 to entry: Generally implied means that it is custom or common practice for the organization (3.3.1), its customers (3.3.5) and other interested parties (3.3.7), that the need or expectation under consideration is implied.

[2] "When I use a word", Humpty Dumpty said in rather a scornful tone, "it means just what I choose it to mean" (Rev. Charles Dodgson, 1872).

Note 2 to entry: A qualifier can be used to denote a specific type of requirement, e.g. product requirement, quality management requirement, customer requirement.
Note 3 to entry: A specified requirement is one that is stated, for example in a document (3.7.2).
Note 4 to entry: Requirements can be generated by different interested parties (3.3.7).
Note 5 to entry: This definition differs from that provided in 3.12.1 of ISO/IEC Directives, Part 2:2004.3.12.1 requirement expression in the content of a document conveying criteria to be fulfilled if compliance with the document is to be claimed and from which no deviation is permitted."

This suggests that validation is a demonstration of suitability for a particular use-case, that the requirements for a validated process should be derived from the intended use-case and that validation should be the process of obtaining data which shows that a method or process meets those specific requirements.

Software engineering approach to verification and validation

In the world of digital forensics we tend to rely on third-party tools which we trust have been produced in accordance with good engineering practices. For the most common analytical tools, this is software which we trust has been correctly specified, implemented and tested. However, the responses to our questions about development models suggest that there is some disconnect between the tool producers and the way end-users are expected to provide evidence of fitness for purpose. In order to understand how this may have arisen, we turned to a consideration of Software Engineering terminology to discover if there is a fundamental conceptual difference.

In Software Engineering, we commonly paraphrase Verification as “are we building the product right?” and validation as “are we building the right product?” (Boehm, 1984). i.e. verification is a demonstration of the correctness of the product whereas validation is a demonstration of suitability for a particular use. More formally the IEEE Standard Glossary of Software Engineering Terminology (IEEE, 1990), states these as.

Verification

(1) The process of evaluating a system or component to determine whether the products of a given development phase satisfy the conditions imposed at the start of that phase.
(2) Formal proof of program correctness.

Validation

The process of evaluating a system or component during or at the end of the development process to determine whether it satisfies specified requirements.

For completeness (IEEE, 1990), also defines a requirement as.

(1) A condition or capability needed by a user to solve a problem or achieve an objective.
(2) A condition or capability that must be met or possessed by a system or system component to satisfy a contract, standard, specification, or other formally imposed documents.
(3) A documented representation of a condition or capability as in (1) or (2).

These definitions are completely consistent with those found in the ISO and ISO/IEC standards under consideration.

Software products should, therefore, be subjected to verification during development - to show that they are correct and complete, and validation post-development to show that they meet the requirements for their intended use-cases. In more common terms, the validation test can be considered to be an acceptance test.

In the case of custom software, produced in response to a particular problem, the process of verification could result in validation for that problem. In the case of off the shelf software (e.g. word processors, spreadsheets, common forensic tools), however, verification during the development phases is based on a generic statement of requirements which meets the needs of a perceived customer or a group of idealised customers. It is the responsibility of the customer to ensure that the verified tool provides a valid solution to their problem as part of the procurement and pre-deployment process.

It is, thus, entirely possible to verify a product which cannot be validated as it does not provide a suitable solution to the problem under consideration (e.g. a custom-built spreadsheet may be completely correctly built but unusable as a presentation package) and it is also possible to validate an unverified product by showing that, despite its inherent flaws, the product satisfies a particular case-specific set of requirements. For example, a calculator which always states that $2 + 2 = 5$ is unlikely to be verifiable, but can participate in a validated method where the requirement is to calculate that $3 + 3 = 6$. Similarly a tool, designed to parse FAT filesystems only, will not parse NTFS. It is therefore, not verifiable for NTFS but can participate in methods which are validated for examination of a FAT formatted filesystem.

In the latter case the unverified product cannot be shown to have any utility beyond the limited circumstances for which it is validated.

In the former case, however, the verified product may be useful in other situations and the presence of evidence of verification can be used to assist the process of choosing it as a potential solution - i.e. the evidence of verification may show that some, or all, of the validation requirements have already been met during the development process.

This depends entirely on the existence of suitable statements of requirements for both the tool as it was developed and the situation in which it is to be used, and satisfactory evidence that those requirements have been satisfied.

Implications for method validation

Given that the definitions and usage of validation and verification, as outlined above, appear to be consistent it should, therefore, be possible to use software engineering evidence of verification, as suggested in ISO/IEC 27041 (ISO/IEC, 2016) as part of the validation of a suitably documented method.

Our study

Laboratory documentation

In our study, we examined a small randomly chosen set of Standard Operating Procedures (SOPs) and Validation plans and records from two accredited digital forensic laboratories. The SOPs were written in a format which appears to be based on the SWGDE Model (Scientific Working Group on Digital Evidence (SWGDE), 2012) and be consistent with the accepted standard format within forensic science laboratories in the UK. These contain sections detailing Purpose, Scope, Equipment, Limitations, Procedure, Processing, Success/Failure Criteria and References. None of these SOPs contained any obvious definitions of technical requirements. Rather they tend to define success in terms of processing completing without any errors being reported, and give a broad area of application in the Scope statement.

Validation plans contained some identified requirements, but these were arranged as End User (the Criminal Justice System), Legal (including compliance with ISO 17025), Compatibility (output format only) and Ethical. No obvious low-level technical requirements were specified in any of the plans.

Validation records showed that validation processes tended to consist of evidence that the process under test produced the same results as the same process run on other equipment or that it produced expected results from a particular test case.

The testing thus satisfied the letter of the ISO 17025:2005 description of validation, but may not have achieved the level suggested by the principles in (Gravel, 2002), particularly in respect of Traceability and Transparency.

This apparent failing is not thought, by the authors, to be a problem for other forensic disciplines whose roots lie in other sciences such as chemistry, physics or biology, where the methods used in forensic laboratories are specific adaptations of well-known methods which are used for other purposes and which have been subjected to rigorous peer-review through publication and extensive use in other work.

Digital Forensics, however, has its roots in engineering and is highly reliant on reverse-engineering of decisions and implementations made by others. Many of these implementations (e.g. hard disc firmware, filesystem implementations, data caching) are not published or reviewed as they are commercially sensitive and/or there is no need for the majority of users/customers to have any particular interest in the low-level implementational detail which is of particular interest to a digital forensic examiner or analyst. As a result, it may be considered to be difficult for producers or users of forensic tools to show that the tools are actually correct except by potentially lengthy and costly empirical methods.

This is compounded by a fundamental difference in the nature of the way in which off the shelf software (OTSS) is used. In a non-forensic context, OTSS is typically intended to process inputs provided by a user in order to generate a particular output. In this situation, the inputs are known, or can be examined, before the output is seen and thus detection of incorrect results can be simple. In the forensic context, however, examinations start with a source of potential evidence whose contents are unknown. Thus the inputs to the whole forensic process are unknown. Although the user may have some experience of what abnormal outputs look like, this depends entirely on the tool actually producing abnormal outputs or indications of errors. It is entirely possible for a tool to process inputs incorrectly and produce something which still appears to be consistent with correct operation. In the absence of objective verification evidence, assessment of the correctness, or otherwise, of any results produced by a tool relies solely on the experience of the operator.

It should also be borne in mind that updates to hardware and software may have no apparent effect on system behaviour as far as a typical user is concerned, but may dramatically change the way in which internal processing is carried out and data is stored. This impacts both on the ability to recover and interpret data and on the behaviour of the tools used to perform these operations.

Vendor evidence of verification

Our study circulated a questionnaire and received 14 responses from tool providers. Of these, 2 could be considered major providers although one is more focussed on e-Discovery than criminal investigations.

The 12 small providers seemed confused about what was meant by customer requirements with responses including “I'm my own customer”, “Sorry, I don't understand the question’, “Forums, social media”, “I do not - many potential customers seem utterly bemused why they should be interested at all”. Of the complete set of 14, 3 identified the use of JIRA/Confluence/Github as a means of deriving requirements and three others identified Meetings and Communications with end users as the mechanisms used.

When asked how they demonstrated that their tool satisfied user requirements, responses include use of NIST test disc images, use within ISO 17025 accredited laboratories, and meetings. Only one of the survey group mentioned compliance testing.

We also, as noted in the introduction, met with considerable resistance from some of the better-known providers when we asked for information about this topic. As a result, we cannot provide objective evidence for any degree of confidence that tool providers are meeting the genuine requirements of the digital forensic laboratories.

Customers for the tools have little incentive to consider the technical requirements as it seems possible to obtain accreditation to ISO 17025:2005 without them, and most tool providers are either unable or unwilling to provide evidence that they have verified their tools against any customer or technical requirements.

Transition to ISO 17025:2017.

The position in respect of accreditation to ISO 17025:2017 (ISO, 2017) may be somewhat different as this now contains definitions of validation and verification which are very similar to those used in ISO 27041 and the software engineering world, viz:

Validation Verification, where the specified requirements are fit for an intended use

Verification Provision of objective evidence that a given item fulfils specified requirements

Thus validation appears, in the newer version, to be reliant on verification against specified requirements and comparison of those requirements with the requirements of the intended use-case.

Conclusion

Contrary to previous arguments that ISO 17025 (Sommer, 2018) is an unwieldy standard for digital forensics because of the complexity of validation, we believe that it can be applied if certain preconditions are met.

For ISO 17025 to be successfully applied, the existing understanding of requirements needs to be reconsidered. Rather than relying on the concepts of “customer requirements” (International Laboratory Accreditation Cooperation, 2014), where the customer is the customer of the laboratory (i.e. law enforcement agents, lawyers, the criminal justice system etc.) to provide the baseline for method validation, forensic science providers should consider the technical requirements for their own processes and use the customer requirements as a means of selecting the most appropriate processes to deploy. This would be consistent with the way other “non-forensic” accredited testing and calibration organisations operate.

Within forensic science disciplines we suggest that all labs will have the same common core technical requirements for generic method types (e.g. in digital forensics, hard disc imaging is a core process, as is extraction of data from devices running specific iOS versions etc.), that these should be established by technical working groups from within each discipline, and documented in agreed international standards which can be maintained for use and development by the community.

The requirements contained in these standards can then form the basis of a specification mechanism for methods. Clear identification of the technical requirements vs. the non-technical would allow producers and users to identify priority areas for new tool development.

Publication, and public maintenance, of this common set of requirements would also allow transparency in the verification and validation process. Rather than relying on “commercially sensitive” information, which may or may not be correct, it would become possible for all those involved to use the disclosed information and make claims (with appropriate substantiating evidence) based upon it.

Furthermore, if the suggestion of ISO/IEC 27041:2015 (ISO/IEC, 2016) that processes should be designed to be atomic in nature (i.e. small, single purpose with low coupling and high cohesion to other processes) can be followed, the set of requirements for any one process can be kept to a minimum, resulting in a better defined set of conditions for validation and an elimination of revalidation being triggered by changes elsewhere in the process. All the methods which were volunteered for our study were monolithic in nature and contained a high degree of repetition of tightly coupled (by virtue of being included in each SOP) initial process stages (e.g. retrieval of physical items from an evidence store) before progressing to the unique elements of the process.

Existing related work

Introduction

Since starting the original project, we have been made aware of some projects which may provide, at least in part, some of the missing requirements, specifications and evidence of correctness. A brief review of two of these, in the context of our analysis and proposals, is given below.

NIST/DHS computer forensics tool testing

The National Institute for Science and Technology (NIST) and the Dept. of Homeland Security (DHS) have started some of this work in their Computer Forensics Tool Testing programme (National Institute for Science and Technology (NIST), 2018) (CFTT). In this project, a steering group defines the requirements for particular tool functions and NIST then tests tools against the resulting specifications. At the time of writing, the coverage is somewhat limited, concentrating on a few areas which may be particularly common in investigations, but a good range of tools has been considered and an online catalogue of tools and results has been produced.

The Federated Tool Testing project as a sub-project of this initiative may be a particularly useful model as it makes available a test suite which can be used by anyone who wishes to test tools against the requirements already defined by the project and share their results.

It is unclear, however, how the programme's priority areas are established or how the requirements are, themselves, validated at as this part of the process does not appear to be documented. It is also noteworthy that the requirements are purely at the tool level rather than the broader method level. This may result in an undue emphasis on producing requirements for existing tools, at the expense of producing requirements which have not yet been satisfied but which should be considered high priority as they reflect an emerging real problem area.

We also suggest that a broader consideration could create opportunities for better tool integration (i.e. improved exchange of data between tools and better cohesion for improved process flows) as well as improved concordance with external requirements such as legal issues.

SWGDE guidance on testing and validation

The Scientific Working Group on Digital Evidence (SWGDE) has issued a number of documents which are intended to assist in the design, implementation and validation of methods for digital forensic processes. Of these, the two which appear to have most direct application to the area we are investigating are.

- SWGDE Recommended Guidelines for Validation Testing (Scientific Working Group on Digital Evidence (SWGDE), 2014)
- SWGDE Minimum Requirements for Testing Tools used in Digital and Multimedia Forensics (Scientific Working Group on Digital Evidence (SWGDE), 2018) (At the time of writing, this document was in draft form and had been issued for consultation).

The SWGDE validation guidance (Scientific Working Group on Digital Evidence (SWGDE), 2014) states that

> Validation testing should be applied to all tools, techniques and procedures

and further that

> Tools, techniques and procedures, which, by virtue of their widespread use, duration of use, and acceptability by the larger information technology community, are generally acknowledged as reliable and trustworthy. Consideration may be given to the general acceptance of a tool, technique, or procedure in the determination of whether validation is required.

The latter paragraph appears, to some extent, to contradict the former. In our experience, it seems that this is generally interpreted to mean that something which is in widespread use may be considered reliable.

We argue that this is not the intent of the “general acceptance” statement. In part, this is because of the presence of the phrase “larger information technology community” which is a clear indication that the tools, techniques and procedures under consideration are of a more general-purpose nature than the specialist tools deployed in an investigative context. Spreadsheets, word processors, email programs etc. may generally be considered acceptably reliable because they have minimal impact on evidential product and, should they prove to have an error, the sheer number of users worldwide means that it is likely to be detected and documented relatively quickly.

More importantly, however, if this general acceptance principle is allowed to apply to commonly adopted “forensic” tools, techniques and procedures it has the potential to result in bad evidence. If the tool, technique or procedure has not been subjected to independent scrutiny (e.g. through peer-reviewed publication or properly evidenced validation testing) there is insufficient evidence that it does work correctly. As we note above, digital forensics relies heavily on reverse engineering in order to process and interpret data. At the level that most users operate, it does not have sufficient foundational scientific principles to allow a reversion to first principles to be applied in order to demonstrate correctness. There is always likely to be some doubt or uncertainty about the way the data is being processed and interpreted. This can be reduced only through production of evidence of correctness and adequacy through appropriate software engineering methods, such as testing.

Note: we do not see this as a flaw in the SWGDE guidance, but rather in the way that a large part of the community has chosen to interpret this particular recommendation. It should be noted that similar phrases appear in other guidance and, in our experience, are similarly (mis)interpreted.

The remainder of this document gives a high-level overview of the development of a testing procedure which, if underpinned by well-defined requirements which allow the identification of appropriate test cases could result in good evidence of validation and identification of boundary cases for methods.

The tool testing guide (Scientific Working Group on Digital Evidence (SWGDE), 2018) is more detailed in its recommendations and gives advice about specific tool types and the conditions which should be considered for their testing. Again, however, it makes little reference to using a well-defined set of requirements to assist in the identification of test cases. It does acknowledge that the testing proposed is purely a minimum and that organisations should consider their own particular requirements.

It is our view that evidence of testing, produced in the recommended way, could be applied as an adjunct to method validation, providing the requirements are properly defined and documented. It should be remembered, however, that tool testing alone is unlikely to be produce the evidence of validation required by either ISO 17025 (ISO, 2005a), (ISO, 2017) or ISO/IEC 27041 (ISO/IEC, 2016), unless it can be clearly shown that the method is wholly and solely implemented by the tool (see Fig. 2).

Final thoughts

While the NIST and SWGDE projects outlined above may start to provide the type of evidence that is necessary to demonstrate that a method is valid, the potential lack of transparency in the requirements definition processes introduces another element of uncertainty. i.e. if the requirements cannot be shown to be correct, can tests based on those requirements show correctness? This can, to a large extent, be addressed by adopting the “non-forensic” accredited organisation model of using publicly available agreed standard specifications/requirements and/or methods which can be subjected to external independent scrutiny.

It also be useful to engage in a more open process, similar to those proposed for use in the specification and testing of safety-critical systems (Martins and Gorschek, 2016).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.diin.2018.09.004.

References

Boehm, B.W., 1984. Verifying and validating software requirements and design specifications. IEEE Softw 1, 75–88.

Gravel, J., 2002. Principles behind the Requirements of ISO 17025 online at. http://www.cala.ca/ISO-IEC17025Principals.pdf. (Accessed 25 April 2018).

IEEE, 1990. IEEE standard glossary of software engineering terminology. IEEE Std 610.12-1990, 1–84.

International Laboratory Accreditation Cooperation, 2014. ILAC G19:08/2014 Modules in a Forensic Science Process.

ISO, 2005. ISO 17025:2005 General Requirements for the Competence of Testing and Calibration Laboratories.

ISO, 2005. ISO 9000:2005 Quality Management Systems — Fundamentals and Vocabulary.

ISO, 2017. ISO 17025:2017 General Requirements for the Competence of Testing and Calibration Laboratories.

ISO, 2018. ISO Online Browsing Platform online at. https://www.iso.org/obp/ui/. (Accessed 13 August 2018).

ISO/IEC, 2016. ISO/IEC 27041:2016 Guidance on Assuring the Suitability and Adequacy of Digital Investigation Method.

Martins, L.E.G., Gorschek, T., 2016. Requirements engineering for safety-critical systems: a systematic literature review. Inf. Software Technol. 75, 71–89.

National Institute for Science and Technology (NIST), 2018. Computer Forensics Tool Testing Programme online at. https://www.nist.gov/itl/ssd/software-quality-group/computer-forensics-tool-testing-program-cftt. (Accessed 13 August 2018).

Rev. Charles Dodgson (Lewis Carroll), 1872. Through the Looking Glass.

Scientific Working Group on Digital Evidence (SWGDE), 2012. SWGDE model standard operation procedures for computer forensics online at. https://www.swgde.org/documents/Current Documents/SWGDE QAM and SOP Manuals/SWGDE Model SOP for Computer Forensics. (Accessed 5 June 2018).

Scientific Working Group on Digital Evidence (SWGDE), 2014. SWGDE Recommended Guidelines for Validation Testing (Version 2.0). Last accessed 13 August 2018.

Scientific Working Group on Digital Evidence (SWGDE), 2018. SWGDE Minimum Requirements for Testing Tools Used in Digital and Multimedia Forensics. Draft Version 1.0 dated 9th July 2018. (Accessed 13 August 2018).

Sommer, P., 2018. Accrediting digital forensics - what are the choices? Digit. Invest. 25, 116–120.

UKAS, 2018. Directory of Accredited Organisations online at. https://www.ukas.com/services/other-services/directory-of-accredited-organisations/. (Accessed 4 June 2018).

Appendix F

Tool verification

Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

journal homepage: www.elsevier.com/locate/fsidi

Digital forensic tool verification: An evaluation of options for establishing trustworthiness

Angus M. Marshall

n-gate ltd., N. Yorkshire, UK

ARTICLE INFO

Article history:
Received 25 February 2021
Received in revised form
14 May 2021
Accepted 15 May 2021
Available online xxx

Keywords:
ISO 17025
ISO/IEC 27037
ISO/IEC 27041
ISO/IEC 27042
Quality standards
Method validation
Forensic tool verification
Forensic tool development
Forensic tool testing
Software trustworthiness

ABSTRACT

Marshall and Paige (2018)2018) reported a reluctance, on the part of tool producers, to allow scrutiny of their internal testing processes and results. The absence of evidence of tool verification increases the work required for method validation to comply with regulatory and international standards requirements. The author presents a range of models for tool verification and evaluates them in terms of estimated risk and cost to end-user and provider. A preferred option is identified.

1. Introduction

Marshall and Paige (2018) describe the potential use of evidence of tool verification in support of method validation to achieve compliance with the requirements of ISO 17025 (ISO, 2017) and/or ISO/IEC 27041 (ISO/IEC, 2016a). They described three scenarios to show how tools may participate in methods:

1. **Tool is a subset of method.** All the requirements of the tool are present in a larger set of requirements for the method in which the tool is to participate. This is typical of equipment used in "wet forensic science" processes where each device fulfils a single function.
2. **Method is a subset of tool.** All the requirements of the method are present in a larger set of requirements which the tool satisfies. This is potentially the case for some general purpose digital forensic tools, and certainly the case for software such as word processors where studies have shown that most users only use a fraction of the capabilities of such tools.
3. **Tool intersects with Method.** The tool and the method have different sets of requirements, but there is an overlap between them.

In all 3 cases, however, only those requirements which are common to both method and tool would require evidence of verification in order to support validation of the method under consideration.

For the purposes of this exercise, it is assumed that requirements for methods and tools have been properly identified. Marshall and Paige's work (Marshall and Paige, 2018) suggests that is not demonstrably true, but addressing this topic is beyond the scope of this paper. The author is preparing another paper to address this issue in more detail.[1]

This paper, therefore, addresses the issues of disclosure of requirements and disclosure of evidence of verification (i.e. compliance with the disclosed requirements).

E-mail address: angus@n-gate.net.

[1] At this point the author wishes to make it clear that he is not suggesting that any existing tools do not satisfy the requirements of their users, nor that users are, necessarily, using tools which are inappropriate for their requirements. The underlying issue is that, in the circumstances described, it is not possible to provide objective evidence that a tool is "fit for purpose" without expending considerable effort during (re)validation to show that a tool performs its role in a method appropriately. Historically, there are known cases of certain tools failing to read the last sector of a disc, or tools which supposedly perform the same file extraction producing different numbers of files from the same evidence source. In these situations, the old technique of "dual-tooling" often failed to resolve the key question of "which tool is wrong - if any?", leaving us only with some form of verification as a means to proving tool correctness.

https://doi.org/10.1016/j.fsidi.2021.301181
2666-2817/© 2021 Elsevier Ltd. All rights reserved.

2. A comment on requirements

Over the years many attempts have been made to produce standardised models for digital forensic workflows and processes. Not all have been published, but most seem to support the proposition in ISO/IEC 27042 (ISO/IEC, 2016b) that, after the PDE (Potential Digital Evidence) acquisition stage (ISO/IEC, 2016c), any examination can be broken down into a collection of lower–level processes, each of which performs a single function, dealing with a particular source, type, or class of artefact.

The acquisition stage itself falls into one of three categories which can be defined in terms of the ACPO principles for electronic evidence (ACPO, 2012) as Principle 1, Principle 2 or Remote Source devices, i.e.

- **Principle 1** These are physically accessible PDE sources which inherently support the principle that the extraction of PDE from them should not cause changes to data on the source. i.e. they are classical “dead box” type storage devices where data is held semi-permanently and can be protected through the use of write-blockers or similar methods. It is usually possible to repeat the PDE acquisition process,[2] when necessary, and achieve identical results to the first acquisition. Both physical and logical acquisitions are possible.
- **Principle 2** These are physically accessible PDE sources which require some modification to their means of operation, by implantation of additional software or subversion of in-built security (“rooting” or “jailbreaking”) in some way in order to gain access. Because this is a modification, there is a degree of risk associated with the method and a higher degree of competence is required on the part of the operator. Once the acquisition has been achieved, however, further acquisitions can be carried out and should achieve substantially similar results to previous acquisitions. Physical and logical extractions are possible, but the PDE extracted may be incomplete because of limitations inherent in the way the modification is allowed to access the internal storage.
- **Live/Remote Sources** These are sources which are not amenable to physical access (e.g. cloud storage, messaging servers, social media systems, corporate servers) and from which PDE can most easily be acquired through the same interface which is available to the normal user. The examiner is, at best, presented with a logical view of data held on the remote source and probably only obtains access to live data rather than deleted historic data. This category includes PDE sources which may appear to be physically accessible, but cannot be taken offline or out of service for examination. Thus their contents are likely to change during PDE extraction, resulting in a “smear”. Such sources may even require a fully online or live examination instead of an offline, “dead box” or static approach.

Regardless of which category the PDE source falls into, the processing that follows acquisition, i.e. the processing of the image, then follows the model given in ISO/IEC 27042 (ISO/IEC, 2016b) as an iterative process of analysis and interpretation of artefacts in order to develop understanding of what was present on the source and what it means to the investigation.

The net effect of this is that there is likely to be a high degree of commonality in requirements at the process level, particularly in the functional requirements (ISO/IEC, 2016a), between end-user organisations. This arises because there is a finite, although growing, set of mechanisms for representing and storing data (i.e. coding mechanisms, filesystems, file formats and protocols) and all processes, or sequences thereof, must, at some point, make use of, or interpret, one or more of these.

3. Verification

We have, therefore, a situation where processes, specified by SOPs (Standard Operating Procedures) at the end-user level, will have common requirements dictated by the PDE that they have to work with. These could, therefore, be documented in the SOPs and used as part of the acceptance criteria for adoption of a tool to participate in the methods captured in the SOPs.

Marshall and Paige (2018) have explored this in more detail and ISO/IEC 27041 (ISO/IEC, 2016a) contains recommendations for documenting and using these requirements.

Thus the potential for evidence of tool verification against disclosed requirements to be used in support of validation is real. The challenge seems to be to find a mechanism which allows tools to be verified, and for evidence of verification to be disclosed, without exposing any of the parties involved to undue risk of liability or of disclosure of commercially sensitive information.

In fact, the situation is similar to that found in safety-sensitive situations where it may be necessary to establish trust in a subsystem or software product without revealing excessive detail about the product itself. Indeed, the author would argue that any forensic tool is a safety-sensitive product because its results can lead to life-changing effects.

3.1. Software trustworthiness levels

For specialist software and bespoke development projects, the application of rigorous development and quality assurance models such as the Capability Maturity Management Integration (Paulk, 2009), IEEE 730–2014 (IEEE, 2014) Standard for Software Quality Assurance Processes, or ISO/IEC 12207:2017 (ISO/IEC, 2008) Software life cycle processes standards and ISO/IEC/IEEE 15289:2011 (ISO/IEC/IEEE, 2011) Content of life-cycle information products (documentation) standards should result in inherently trustworthy products. However, these models do not really consider the issue of disclosure of evidence of verification, and thus trustworthiness, to parties other than customers who have engaged the development organisation.

In considering forensic tools we are, on the whole, dealing with commercial off-the-shelf (COTS) products, albeit of a specialised nature. Some of these started as personal projects, almost at the hobbyist level, to solve specific problems. Of these, some proved particularly useful and evolved into more general purpose packages which can deal with multiple types and sources of evidence through the addition of elements of case management functionality. These tools and their development teams tend to evolve organically at first before becoming large enough to consider the adoption of rigorous development models. As COTS products, though, their direct customers are themselves - i.e. they have not been commissioned by any external entity to produce the tools and have neither moral nor contractual obligation to meet any external requirements in current regulatory and market conditions.[3] At the time of writing, even a cursory inspection of material available from

[2] There are obvious exceptions such as discs which are damaged or beginning to fail where consistent reads of some sectors are not possible - but these are special cases.

[3] This is not dissimilar to the situation arising from accreditation to the ISO 17025 standard - the accreditation is for the organisation, people and process by which a product is produced, but not for the product itself. There is implied trust which can be difficult to measure or guarantee.

Table 1
TSF trustworthiness levels.

TL	Software Audience	Control Set
TL 0	No requirement for Trustworthy Software	No requirement
TL 1	Mass Market with Implicit Need (M/I)	TS Essentials (TSE)
TL 2	Mass Market with Implicit Need (M/I)	Baseline TS controls forming a sub-set of the TS Framework (TSF)
TL 3	Mass Market with Explicit Need (M/E)	TS Framework (TSF)
TL 4	Niche Market with Explicit Need (N/E)	Comprehensive TS controls utilising the full TS Framework (TSF)

the more well-known forensic tool providers will show that they do not publicly disclose detail of specifications or limitations of their products except in the broadest terms. As Marshall and Paige noted (Marshall and Paige, 2018), they are also unwilling to allow third-party inspection of processes, even if NDAs are offered.

Models similar to CMMI, and others mentioned above are, of course, suitable for consideration when selecting a third party to carry out the certified or accredited diligence models described below, as the implicit trust which results from their application allows us to (mostly) avoid the need for verification of the output of the certified or accredited process.

Regardless of the history of the tool, or its current development methods, mechanisms for the production and disclosure of trustworthy evidence of verification (i.e. conformance to specification), which are independent of other considerations, can be devised.

In order to achieve this, we need to consider what is meant by trustworthiness in the context of a software or hardware product. Ultimately, it is a means of measuring our confidence or assuredness that the results produced by the product are reliable, reproducible and repeatable and that we know the limitations of the product (e.g. the conditions under which it is known to fail in some way).

3.1.1. Trustworthy Software Foundation trustworthiness levels

The Trustworthy Software Foundation has proposed 5 levels of software trustworthiness (Trustworthy Software Foundation, 2016a), based on audience (end users), development methods (production processes) and testing (verification). These are summarised in Table 1.

For the purposes of this discussion, the main feature of the TSF model, is the identification of 5 levels of explicit trustworthiness in a product. These range from 0 (no need for trust) to 4 (need for maximum possible trust).

From the TSF levels, digital forensic tools would seem to fall into TL3 or TL2 which demand, in TSF terms, that controls (i.e. rigorous development methods, requirements definitions and testing (Trustworthy Software Foundation, 2016b)), are in place during development. In a few cases, TL4 might be necessary.

As noted above, because tools are COTS, we cannot easily mandate development methods and, not least because of the rapid development cycles involved in updating existing tools to cope with new PDE source and with the need for new tools to deal with new PDE types, probably should not attempt to control the whole development process in this way. We can, however, encourage tool providers to facilitate more rapid adoption of upgrades and new products by providing evidence of the trustworthiness of claims about functionality and thus trustworthiness of their tools.

It should also be noted that, although the TSF concerns itself with software, the principles embodied in its guidance could be adapted for application to hardware.

3.2. Options for establishing trustworthiness of digital forensic tools

Trustworthiness is related to verification - i.e. the producer needs to establish that their product performs as they claim it does and needs to disclose sufficient information for their customers to be able to satisfy themselves that an appropriate TL has been achieved. Thus something about the verification of the product against a publicly available specification must be made available to customers. The question becomes one of disclosure - i.e. “How we can reveal sufficient information about verification without disclosing trade secrets and/or creating undesirable liabilities and risks to the producer?”.

The following list describes some options for performing verification of digital forensic tools, and disclosing sufficient information about verification, with comments about how each works and an approximate mapping to Trustworthiness Levels.

The disclosure mechanisms themselves are derived from practices known to already exist (i.e. observed in practice amongst vendors and users) with the addition of mechanisms based on the concepts of certification and accreditation as they are used in the application of standards such as ISO 17025, ISO/IEC 27037 et al. The list itself may not be comprehensive, but the mechanisms described are intended to offer sufficient granularity that, when applied with appropriate third parties where necessary, the levels of liability, risk and trust associated with each can be estimated for comparison purposes. This evaluation and comparison is explored further in section 3.3.

3.2.1. List of possible methods for establishing trustworthiness through verification

Claim-free (CF) Producer makes no claims. Software is released “as-is”. User has full responsibility for ensuring compliance with requirements. Disclosure is not possible because evidence does not exist.

N.B. This is not the Open Source default although many free tools and scripts may fall into this category.

Spec. sheet only (SSO) producer publishes list of claimed functionality, cannot disclose results of internal testing because it either hasn't been done or hasn't been recorded. End-user assumes all responsibility for compliance.

Internal verification (IV) producer publishes list of claimed functionality, has carried out internal testing but does not disclose unless forced to. End-user responsible for compliance.

External verification (EV) producer publishes list of claimed functionality, a third party tests against their own requirements (subset of, or intersection with requirements which map to producer's claims), often as part of a ”group test” (e.g. NIST tool testing). End-user can use third party results to assist compliance, but must establish trust in third party and may still have further work to do (e.g. where third-party has not tested a claim). This is often proposed as the answer to the disclosure problem as it provides an improvement over CF, SSO and IV, but results may lag releases by a considerable amount of time and some claims (e.g. unique functionality) may never be tested.

Certified internal diligence (CID) producer is willing to disclose more detailed requirements and carries out appropriate testing. Third party examines testing and, optionally, development regime (NOT detail of tests or implementations) and certifies producer as competent.

Sampled accredited internal diligence (SAID) As CID, but third party can, and does, examine/observe a sample of the producer's work directly to check adequacy and competence.

Full accredited internal diligence (FAID) As SAID, but ALL relevant producer processes are inspected and checked by the third party. Some may be subject to further independent testing by the third party.

Open diligence (OD) producer discloses requirements or specification, test plans, test data and test results publicly for peer-review and adoption by others. End-user may carry out their own confirmation (see ISO/IEC 27041 (ISO/IEC, 2016a)) or engage a third-party. Other producers can adopt the published information and use it to demonstrate equivalence.

Currently, it seems that CF, SSO and IV are the common “disclosure” (or, more accurately non-disclosure) methods used for digital forensic tools.

EV includes the NIST Computer Forensics Tool Testing Program (National Institute for Science and Technology(NIST), 2017) (which includes a variant of OD through publication of specifications and test suites), while the related Federated Testing (FT) project can be viewed as a combination of IV, EV and OD. Neither scheme is true OD, however, as the published material is derived from end-user requirements and thus may concentrate on only a subset of commonly available features across all tools and cases, rather than giving full coverage of specialist features in any particular tool or unusual case requirements.

Federated Test results are produced by a crowdsourced third party (or a consortium of volunteers running publicly available tests and thus acting as third parties) rather than being commissioned by a tool producer. The distinction between CFTT as EV + OD and FT as EV + IV + OD may be considered somewhat artificial because both CFTT and FT use the same tests and data. The major difference lies in the fact that volunteers conduct the Federated Testing, have access to copies of all the tests in the image files provided by NIST, and have fewer controls (if any) on the way they conduct or modify and report tests and test results. Thus, there is an increased requirement for those third parties to be trustworthy in order for their results to be verifiably trustworthy. A single rogue tester could have a hugely negative impact on the trustworthiness of results of CFTT.

CID, SAID and FAID are similar to the methods currently used to certify and accredit organisations to the ISO/IEC 2700x or ISO 17025/ISO 17020 standards.

Table 2 provides a summary of estimated liability, risk and effort required to implement each of the methods described above. Although the ratings are estimates, they are based on the author's experience of reviewing practices in a range of laboratories whilst undertaking commercial projects related to the production of test scenarios and SOPs.

Table 2
Comparison of the disclosure models based on risk, liability and effort.

Model	PL	PE	PR	UL	UE	UR	TL	TPTL
CF	L	L	L	VH	VH	VH	0	n/a
SSO	M	M	L	H	VH	H	0	n/a
IV	M	H	L	H	VH	H	0	n/a
EV	L	VL	L	H	H	H	1	2–3
CID	L	H	M	M	H	M	2–3	2–3
SAID	L	H	L	M	M	M	2–3	2–3
FAID	L	VH	M	L	L	L	3–4	2–3
OD	H	VH	H	M	H	M	3–4	2-4 if used

PL = Producer's potential liability (exposure to potential lawsuits and compensation claims).
PE = Producer's effort requirement (required over and above existing development effort).
PR = Producer exposure to other risks (non-financial, e.g. reputational, IPR etc.).
UL = User's potential liability (exposure to potential lawsuits and compensation claims).
UE = User's effort requirement (required over and above existing effort).
UR = User exposure to other risks (non-financial, e.g. reputational, IPR etc.).
TL = potential achievable Trust Level for this option.
TPTL = Third-Party Trust Level (minimum required for third-party to participate and produce trustworthy results).
All ratings are estimated using a scale of Very Low (VL), Low (L), Medium (M), High (H) to Very High (VH). Liability is a measure of potential cost of dealing with legal action, and/or carrying out remediation, in the event that flaws are found. Other risks include potential for IPR breaches, exposure of security problems, reputational harm etc. Effort is additional effort required to implement the method.

3.3. *Evaluation*

From this, it can be seen that the current CF, SSO, IV and EV system pushes costs towards the end-user and does little to increase the TL of any particular tool. EV has some impact, but only where requirements are common for a “mass market”.

CID, SAID and FAID achieve a redistribution of costs from user to producer and, in practice, this would equate to a reduction in costs across the whole sector because the current replicated costs (i.e. incurred by all users) would be eliminated and replaced by a single cost point at the producer. This is, of course, likely to result in an increase in pricing which has the effect of distributing the cost across all users, resulting in a reduced total cost.

Based on the estimated ratings, SAID appears to offer the best return for all parties. Producer costs are increased in effort, but not elsewhere, user costs are reduced in all three categories and a TL of 2 or 3 is achievable using this method.

FAID offers further reductions for the end-user, but at the expense of significantly higher effort on the part of the producer and the resulting potential for TL4 is not thought to be necessary for forensic tools.

4. Recommendation

The author recommends, therefore, that the industry should explore the SAID (Sampled Accredited Internal Diligence) option as a means for disclosing tool verification in a way which does not violate commercial confidentiality in development methods, allows users to reduce their method validation costs, spreads the cost of verification more equitably, and allows more rapid adoption of upgraded or enhanced tools by potentially removing the need for some or all re-validation. Ideally it should be compared with the CID and FAID models as well as the status-quo to produce empirical values for the various ratings used in Table 2 and thus provide a real evaluation of the cost implications for the user of tool verification information as an aid to method validation.

References

ACPO, 2012. Acpo Good Practice Guide for Digital Evidence. https://library.college.police.uk/docs/acpo/digital-evidence-2012.pdf. (Accessed 24 February 2021).

IEEE, 2014. IEEE 730-2104 Standard for Software Quality Assurance Processes.

ISO, 2017. ISO 17025:2017 General Requirements for the Competence of Testing and Calibration Laboratories.

ISO/IEC, 2008. ISO/IEC 12207:2008 Systems and Software Engineering — Software Life Cycle Processes.

ISO/IEC, 2016a. ISO/IEC 27041:2016 Guidance on Assuring the Suitability and Adequacy of Digital Investigation Method.

ISO/IEC, 2016b. ISO/IEC 27042:2016 Guidelines for the Analysis and Interpretation of Digital Evidence.

ISO/IEC, 2016c. ISO/IEC 27037:2016 Guidelines for Identification, Collection, Acquisition and Preservation of Digital Evidence.

ISO/IEC/IEEE, 2011. ISO/IEC/IEEE 15289:2011 Systems and Software Engineering — Content of Life-Cycle Information Products (Documentation).

Marshall, A., Paige, R., 2018. Requirements in digital forensics method definition: observations from a UK study. Digit. Invest. 27, 23—29.

National Institute for Science and Technology NIST, 2017. Computer forensics tool testing programm(cftt), 2018,2019,2020. https://www.nist.gov/itl/ssd/software-quality-group/computer-forensics-tool-testing-program-cftt. (Accessed 25 February 2021).

Paulk, M.C., 2009. A history of the capability maturity model for software. ASQ Software Quality Professional 12 (1), 5—19.

Trustworthy Software Foundation, 2016a. Trustworthy software specification document. https://tsfdn.org/wp-content/uploads/2016/03/TS502-0-TS-Essentials-Specification-Issue-1.2-WHITE.pdf. (Accessed February 2021).

Trustworthy Software Foundation, 2016b. Trustworthy software guidance document. https://tsfdn.org/wp-content/uploads/2016/03/TS502-0-TS-Essentials-Guidance-Issue-1.2-WHITE.pdf. (Accessed February 2021).

Appendix G

Language issues

Forensic Science International: Digital Investigation 40 (2022) 301349

Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

journal homepage: www.elsevier.com/locate/fsidi

The unwanted effects of imprecise language in forensic science standards

Angus M. Marshall

N-gate Ltd., N. Yorkshire, UK

ARTICLE INFO

Article history:
Received 11 June 2021
Accepted 15 January 2022

Keywords:
ISO/IEC 17 025
ISO/IEC 27 037
ISO/IEC 27 041
ISO/IEC 27 042
Quality standards
Method validation
Requirements definitions

ABSTRACT

The author has previously contributed to work on requirements definitions in digital forensic methods, and identified a potential gap which could not be fully explained at the time (Marshall and Paige, 2018). The former Forensic Science Regulator (FSR), with others, commented on this and challenged the finding (Tully et al., 2020). This paper re-addresses this issue and explores the issue of language used in the various standards from ISO/IEC 17 025 (2017) through to the FSR's own guidance on digital forensic method (Forensic Science Regulator, 2020), comparing it with language used in other related standards and in software engineering standards. From this, the author proposes that the language used by the FSR may cause an over-emphasis on establishing requirements for the ultimate end-user, to the detriment of requirements for purely internal use of processes. This can also result in overly complex methods, which are inherently difficult to fully validate, being produced. Furthermore, the use of overloaded terminology may also lead to confusion about some key concepts in the various stages of method validation and re-validation.

1. Introduction

Tully et al. (2020) stated, of Marshall and Paige (2018) that

> “They observed an absence of clear technical requirements within digital forensics service providers and a reluctance to disclose customer requirements by tool providers. They also described a lack of technical requirements in validation plans. This lack of technical requirements contravenes the validation requirements set out in the Codes; we are unable to determine whether the organisations included in the Marshall and Paige study also sought accreditation to the Codes, but it is standard UKAS practice to raise non-conformities if the validation is not in line with the requirements of the Codes. ”

Since that work, this author has been studying the issue of technical requirements more closely and believes that there may be an explanation, for the observed lack of technical requirements, in the use of language in the various documents used and provided by the Forensic Science Regulator.

It is also interesting to note the reference to the validation being in line with the requirements of the “Codes”, rather than the root ISO/IEC 17025 or 17020 standards. This appears to contradict Tully et al.‘s assertion that

> “it is not within the gift of the Regulator to unilaterally change an international standard or convert a guidance document to an accreditation standard”.

The regulator's own foreword to the latest version of the codes available at the time of writing declares

> “ Some have equated the current lack of statutory enforcement powers for the Regulator with an assumption that compliance is voluntary. That is not the case and any non-compliance with the Code of Conduct must be declared in statements”.

From this it is obvious that the Code is intended to be far more than guidance.

Since the FSR's Codes and guidance documents are reliant on the ILAC G19 (International Laboratory Accreditation Cooperation, 2014) “Modules in a Forensic Science Process” document as a

E-mail address: angus@n-gate.net.

https://doi.org/10.1016/j.fsidi.2022.301349
2666-2817/© 2022 Elsevier Ltd. All rights reserved.

normative reference, it appears that the FSR's documents are, indeed, intended to convert guidance documents into standards.[1] It is also interesting to note that the UKAS register (UKAS, 2021) lists accreditations with adherence to the FSR's Codes, rather than to the ISO/IEC standard alone, further reinforcing the view that the FSR's Codes are being applied as more than simply guidance.

2. Hierarchy of documents

Within the discipline of digital forensics, the FSR has adopted ISO/IEC 17025 (ISO/IEC, 2017) as the base standard for all laboratory-based forensic science disciplines. ILAC G19 (International Laboratory Accreditation Cooperation, 2014) has been adopted as guidance on how to apply ISO/IEC 17025 to forensic sciences and the FSR's own Codes of Practice and Conduct (Forensic Science Regulator, 2021) provide further information about how to apply these in England and Wales. Various guidance notes then address specifics of implementation for particular disciplines, with the FSR guidance on validation of digital forensic methods (Forensic Science Regulato, 2020) being applicable for this discussion.

3. Response to Tully et al

As one of the authors of the referenced work (Marshall and Paige, 2018), this author stands by the observation that there is an absence of what a computer scientist might consider to be technical requirements in the documents considered in that study, and would like to clarify that all of the documents considered were provided (in confidence) by organisations which had been accredited. The only exceptions to this are the sample SOPs (Standard Operating Procedures) mentioned in the paper (Marshall and Paige, 2018). In the documents provided by accredited organisations, the SOPs contained no clear statements of requirements, whilst associated validation plans contained sections titled “End User Requirements” and listed mainly Criminal Justice System (CJS) requirements.

The absence of requirements in the SOPs may be explained by a perceived need to produce SOPs quickly, because of pressure to become accredited, coupled with a natural tendency to document what was being done at the time, rather than reviewing existing procedures, identifying requirements and producing new SOPs to meet those - i.e. the SOPs which contain no requirements may represent “custom and practice” or legacy procedures rather than embodying justified “best practice”.

The language used to describe, and the nature of, requirements in the validation plans are, perhaps, more interesting. A clear emphasis is placed on satisfying the customer's needs. This is no bad thing, but the customer is explicitly taken to be the CJS.

This leads to two potential problems. Firstly, the primary concern is about the output of any process - i.e. the report produced, with considerations of inputs and processing stages tending to be considered only where they affect risk to the end user. Secondly, it encourages the production of monolithic processes - i.e. potentially multi-stage methods which start with the original source of potential evidence and result in a report or partial report. This second proposition is borne out by the fact that most, if not all, of the SOPs considered included decision points where alternative tools or processes could be applied depending on the results of earlier stages. Further evidence for this can be found in the UKAS register (UKAS, 2021) where it is clear that many accredited organisations have only one or two SOPs which handle multiple evidence sources and types, using a variety of tools.

Such methods are inherently complex and difficult to validate because of the potential difficulty in establishing suitable tests for all possible combinations of branches. It is for this reason that ISO/IEC 27042 (ISO/IEC, 2016a) recommends the use of “atomic” methods with decisions made between the methods rather than inside them. In the ISO/IEC 27042 model, these atomic methods are combined to form an analysis, one or more analyses combine to form an examination, and multiple examinations are combined to form an overall investigation.

From the end-user or customer perspective, their requirements should be satisfied by the investigation and the examinations, analyses and processes should be chosen to construct an appropriate investigation. The end-user requirements, however, need not propagate to the lower levels of the ISO/IEC 27042 model (ISO/IEC, 2016a).

It is also interesting to note that Tully et al., whilst accepting the potential utility of ISO/IEC 27037 (ISO/IEC, 2016b) as useful to assist in implementing ISO/IEC 17025 (ISO/IEC, 2017) for digital forensics, do not make significant mention of the other standards in that group.

ISO/IEC 27037 describes the initial four stages of a full forensic process, dealing only with the initial collection and processing of evidence sources to the point where copies have been produced for storage and further processing. ISO/IEC 27042 completes this process model by adding analytical, interpretive and reporting phases, and ISO/IEC 27041 deals with quality issues. The group of 3 standards was designed, by this author amongst others, to provide a comprehensive end to end guide, which is complementary to, and compatible with, the principles of the FSR's standards, but only when all 3 standards are combined.

4. The Forensic Science Regulator's standards

The only conformity standard, and thus the standard against which any organisation can be accredited is ISO/IEC 17025 (ISO/IEC, 2017)[2] (In England and Wales, this is extended by the FSR's Codes of Practice and Conduct, which are declared mandatory for accreditation of forensic science organisations, by the regulator.)

As Tully et al. (2020) noted, there has been considerable resistance to this standard's application to digital forensics, not least because it is difficult to interpret how concepts such as error rate and calibration can be applied to digital processes which, essentially, either succeed or fail. In practice, issues such as false positive and false negative rates, operating limits (e.g. file sizes, storage device sizes etc.) should probably fall under these concepts, but are difficult to measure or estimate because of the vast range of conditions and devices which may need to be considered to get near realistic useable values. Further discussion of these issues is beyond the scope of this document. Instead, the author wishes to concentrate on the concept of “technical requirements” and why they may be lacking, in his view, from SOPs and validation plans.

5. Definitions

In order to progress the analysis, several definitions need to be considered.

[1] ILAC G19 states “This document is intended to provide **guidance** for laboratories, scene of crime investigation units and other entities, hereafter called forensic units, involved in examination and testing in the forensic science process by providing **guidance** for the application of ISO/IEC 17020 and ISO/IEC 17025.” In spite of this, it uses words such as “shall” to indicate mandatory requirements - something which is not normally permitted in guidance standards published by ISO/IEC.

[2] In this discussion the author is concentrating on lab. based activity and thus ISO/IEC 17020 is not considered, but a similar situation applies to its use.

5.1. Software engineering

Because many, if not most, digital forensic processes have a high reliance on software, three definitions from the realm of software engineering, embodied in ISO/IEC/IEEE 29148:2018 (ISO/IEC/IEEE, 2018), may have relevance.

Condition measurable qualitative or quantitative attribute that is stipulated for a requirement and that indicates a circumstance or event under which a requirement applies.
Constraint externally imposed limitation on the system, its design, or implementation or on the process used to develop or modify a system. Note 1 to entry: A constraint is a factor that is imposed on the solution by force or compulsion and may limit or modify the design.
Requirement statement which translates or expresses a need and its associated constraints and conditions. Note 1 to entry: Requirements exist at different levels in the system structure. Note 2 to entry: A requirement is an expression of one or more particular needs in a very specific, precise and unambiguous manner. Note 3 to entry: A requirement always relates to a system, software or service, or other item of interest.

5.2. Calibration and testing

ISO/IEC 17025:2017 (ISO/IEC, 2017) contains

Validation Verification, where the specified requirements are fit for an intended use.
Verification Provision of objective evidence that a given item fulfils specified requirements.

This document contains no specific definition of a requirement, but refers the reader to ISO 17000 (ISO/IEC, 2020) which defines

specified requirement need or expectation that is stated. Note 1 to entry: Specified requirements can be stated in normative documents such as regulations, standards and technical specifications. Note 2 to entry: Specified requirements can be detailed or general.

5.3. Comment

There is already an element of conflict between the Software Engineering and Calibration and Testing definitions of requirements. In the former, it is a statement of need with constraints and conditions applicable, while the latter couches it in more general terms. The concept of the “technical requirement”, discussed above, is aligned with the Software Engineering definition and technical requirements form a subset of the total set of requirements for the method. It is also interesting to note that the definition as “specific requirement” may contain an implication that there are unspecified, i.e. unstated or undocumented requirements and the definition of validation allows these to be ignored.

5.3.1. Requirements

ISO/IEC 27041 (ISO/IEC, 2016c) provides 5 categories of requirement in its description of validation of digital forensic methods as functional, performance, interface, process and non-functional. Of these, the functional and performance requirements are purely technical - i.e. relate to what the method does and its operating conditions. The interface requirements can be technical or non-technical depending on what interfacing is needed (e.g. provision of output in a specified format for ingestion by another tool is a technical requirement, while an indication that data is stored in an unspecified format may be non-technical as the detail of how data should be stored is left as an implementation issue).

Process and non-functional relate to administrative and procedural aspects, and human factors and quality issues respectively. Neither of these normally contains technical requirements although the concepts of portability and maintainability in the non-functional category may imply certain technical requirements (e.g. particular implementation languages or data formats).

Regardless of the variations in language, it is clear that the process of validation is intended to ensure that methods satisfy declared needs, stated as requirements. The question thus becomes one of how to identify those requirements.

ISO/IEC 17025 (ISO/IEC, 2017) is silent on this matter, so the document hierarchy should be referred to in order to obtain guidance. No advice is given in the higher-level vocabulary documents, so we must look further down the chain, into the guidance documents which are intended to aid implementation for forensic sciences.

5.4. ILAC G19

ILAC G19 (International Laboratory Accreditation Cooperation, 2014) provides the following:

Customer The customer is normally the organization and/or a person asking the forensic unit to perform all or a specific part of the forensic science process. This also includes the term ‘client’. This may be an internal customer. If work is requested via legal mandate (e.g. court order) or if the results of examination/testing are to be provided to a member of the judicial system, then the judicial system may be considered to be the customer.
Validation Validation is the confirmation by examination and the provision of objective evidence that the particular requirements for a specific intended use are fulfilled.

The definition of validation is largely consistent with that used in the root standard, but the introduction of the concept of customer is important as this term is used liberally throughout this guidance. It contains no specific advice about method specification or design, but does make the following recommendations:

Examination and testing strategy In defining the examination and testing strategy the forensic unit should consider, where appropriate, the following:

- Customer requirements
- The ability of the forensic science examinations to help address the identified issues
- Urgency and priority of customer requirements
- Appropriate background information
- Alternatives to the propositions which have already been provided by the customer
- Resources available to the forensic unit
- Experts that may need to be consulted prior to examination or testing
- The examinations or testing that has the potential to provide the most information in response to the various propositions and alternatives
- Issues that could affect the integrity of the items under examination or testing
- Constraints that may exist e.g. the need to preserve material for other purposes, cost

- Examination/tests or other activities that may have a destructive effect on subsequent examination/testing
- Co-ordination of multiple disciplinary examination/testing to determine the sample(s) that need to be taken and the sequence of performing sampling or examination/testing
- Examination/testing services that are currently available in laboratories
- Consideration of anti-contamination precautions appropriate for the examinations/testing under discussion and all evidence types that potentially may be available
- On-going review of examination strategy and testing in light of new and significant information
- What is technically possible and worthwhile to meet the customer's requirement, including the defence.

Other elements may also be considered in the examination and testing strategy.

This mirrors, to a large extent, the principle established in ISO/IEC 27042 (ISO/IEC, 2016a) about the construction of an investigation from examinations made up of analyses, composed of atomic processes.

ILAC G-19 (International Laboratory Accreditation Cooperation, 2014) also contains considerable further advice about how validation should be conducted and how infrequently used or externally validated methods should be approved for use through a process of “verification”. This use of the term “verification” is not consistent with the definition in ISO/IEC17025 (ISO/IEC, 2017). In context, it is akin to the “confirmation” concept embodied in ISO/IEC 27041 (ISO/IEC, 2016c) as

> **confirmation** formal assessment of existing objective evidence that a process is fit (or remains fit) for a specified purpose

Throughout the guidance, ILAC G19 is clear that customer requirements are one consideration only, and that scientific rigour, understanding of limitations, accuracy, precision etc. all need to be included in considering how to validate a method. Some mention is made of the concept of an internal customer - i.e. acknowledgement of the fact that some processes may produce outputs which are used only as inputs to other processes and never given to the ultimate “end-user” customer who commissioned the overall work.

There is, however, as noted above, considerable emphasis on meeting customer (“end-user”) requirements as the goal of the overall investigation process.

5.5. FSR's codes

The FSR's codes of practice and conduct (Forensic Science Regulator, 2021) contain 11 pages of guidance on validation of methods. Within these pages, there are 64 sub-clauses. Of these,

- 5 describe selection of methods - concentrating on validation as the requirement for selection
- 6 describe the concept of validation stating that “The validation procedure shall include, where relevant, but is not limited to: a. determining the end-users’ requirements; b. determining the specification; … d. a review of the end-user's requirements and specification …”
- 6 relate to determining “end-user” (i.e. customer) requirements
- 3 relate to specification of the method stating that “A detailed specification shall be written for the method, product or service, and shall include the technical quality standards. It may be an extension of the end-user requirement document or a separate document.” and refers the reader to the ILAC G19 clause 3.10, mentioned above, for issues to consider.
- 5 clauses relate to risk assessment, defining risk as something which applies to the criminal justice system (CJS) (i.e. the ultimate customer)
- 4 relate to reviewing end-users’ requirements stating that “The forensic unit shall review the end-user's requirement to ensure that requirements considered essential/mandatory have been translated correctly into the specification and the specification is fit for purpose.”
- 2 clauses deal with acceptance criteria
- 6 deal with the validation plan
- 2 are for validation of measurement-based methods, and therefore likely to be viewed as not applicable to digital work.
- 2 are for validation of interpretive methods and address competence and proficiency rather than validation of methods per se.
- 6 relate to “verification” (in the ILAC G19 sense) of adopted methods
- 2 deal with minor changes to methods, including the need for re-validation
- 4 deal with infrequently used methods, again using the term “verification” in the ILAC G19 sense
- 1 is for validation outcomes
- 4 deal with assessment of acceptance criteria compliance
- 5 describe the validation report
- 3 describe the statement of validation completion
- 4 deal with the validation library (a document repository) and
- the final 1 details the need for an implementation plan and any constraints relating to implementation of methods, products or service.

There are two issues in this section, from this author's perspective:

> **Overloading of verification** In software terms, the concept of verification is overloaded in both this and the ILAC G19 document. The term is used in a way which is inconsistent with accepted use in another relevant domain. While this may be acceptable in other forensic sciences (which may not have the same concept of verification), the overload is likely to lead to confusion about exact meaning. It is for this reason that the concept of “confirmation” was introduced in ISO/IEC 27041 to express the ILAC G19 concept.
>
> **Over-emphasis of end-user requirements** There are hints, within the FSR's codes that the concept of an end-user is not restricted to the CJS, but the number of clauses which explicitly deal with CJS needs tends to give the impression that CJS requirements are the primary concern.

It is also noteworthy that the section on “Technical Requirements” in the regulator's codes deals only with personnel, including qualifications, education and training of personnel, but excluding competence, which is given its own clause. Again, this is an example of an overloaded term. Any reasonable computer scientist or engineer would expect to find information about product, method or tool requirements under this heading - i.e. material which can be used to generate a technical specification.

The Digital Forensic Services appendix (Forensic Science Regulator, 2020) to the Regulator's Codes restates the recommendations of the Codes and directs the reader to a specific guidance document on validation of digital forensic methods in addition to the general guidance on validation.

5.6. FSR guidance on validation for digital forensic methods

The FSR's guidance on digital forensic method validation (Forensic Science Regulato, 2020) contains the following advice:

4.1.1 The end goal of validation is for the user of the method (the forensic unit), and the user of any information derived from it (the end user), to be confident about whether the method is fit for purpose as well as understanding any limitations. The ability to assess if a method is fit for purpose depends on first defining what the forensic unit needs the method to reliably do, as well as identifying who are the end users of the method and subsequent results.

4.1.2 The requirements, in their simplest form, capture what aspects of the method the expert will rely on for their critical findings, i.e. what the expert needs to provide in a statement or report.

4.1.4 If the method is being adopted or adapted from elsewhere, the end-user requirements will need creating from scratch. Rather than including all the functional and non-functional aspects, the requirements ought to focus on features that affect the ability to give reliable results.

and further defines

End user The end user of forensic science is the Criminal Justice System, essentially the courts. A method or tool may not be directly used by the courts, but it is assumed that the results will be. Anything that may prove or disprove an assumption to be true, for example, an exhibit or the lack of expected findings.

This tends to allow the functional and performance requirements defined in ISO/IEC 27041 (ISO/IEC, 2016c) to be ignored, for the purposes of validation according to the FSR's guidance. Although clause 4.1.1 clearly mentions internal considerations of reliability and functionality, the emphasis is very much on the “customer” throughout.

6. Argument

Based on the definitions and descriptions in the applicable documents, this author argues that the use of overloaded terms and the, probably unintended, emphasis on “end-user’ CJS customer requirements has created a situation where it is possible, as evidenced by the documents seen in the Marshall and Paige (2018) study, for an organisation to be accredited but for the accredited processes to fall below what might be considered the best standard possible. By allowing the creation of monolithic processes whose design is driven by particular customer needs, sometimes to the exclusion (or neglect) of technical requirements, the existing regime has created the potential for inefficient and non-validtable methods (from a software engineering perspective) to give the impression of being “fit for purpose”.

Furthermore, this author would argue that there is a conflict between the Regulator's concept of the CJS as customer and the ILAC G19 definition of a customer as the person or organisation commissioning the forensic examination. In practice, although the CJS is the potential end-user of any reports (assuming the case goes to trial), the instructing person or organisation is likely to be from the law-enforcement community and will probably phrase the instruction to meet the needs of that community first and foremost. This creates a risk of confirmation bias, as identified by Sunde and Dror (2021).

This does not mean that all accredited organisations are wrong, nor that all methods currently in use are inherently problematic, but it does mean that there is probably a need to take a step back, to review the guidance documents, in particular, for consistency of language and meaning and for organisations to look more closely at the concept of “customer” or “end-user” in context. Organisations being commissioned to undertake work should review the nature of the instructions received from the customer and, where necessary, convert them into something compatible with the actual end-user (CJS) requirements, in order to avoid potential confirmation bias. This is particularly important for in-house law-enforcement laboratories where the high-level of integration into the investigative and prosecutorial culture may further exacerbate the problem of confirmation bias. They should also reconsider how they define requirements for various processing stages, and explore the potential for the use of simpler processes, which are easier to (re-) validate and maintain.

Monolithic SOPs may be adequate for volume casework, where the same type of processing is required for most jobs, but they limit the ability to rapidly develop new methods, or adapt existing methods to new technologies. Smaller, modular, processes with high cohesion and low coupling (Du Bois et al., Verelst) may offer better support for these situations.

References

B. Du Bois, S. Demeyer, J. Verelst, Refactoring - improving coupling and cohesion of existing code, in: 11th Working Conference on Reverse Engineering, pp. 144–151.

Forensic Science Regulator, Forensic Science Regulator Guidance: Method Validation in Digital Forensics, 2020. FSR-G-218, Issue 2.

Forensic Science Regulator, 2020. Codes of Practice and Conduct Appendix: Digital Forensic Services. FSR-C-107, Issue 2.

Forensic Science Regulator, 2021. Forensic Science Regulator: Codes of Practice and Conduct for Forensic Science Providers and Practitioners in the Criminal Justice System. FSR-C-100, Issue 7.

International Laboratory Accreditation Cooperation, 2014. ILAC G19:08/2014 Modules in a Forensic Science Process.

ISO/IEC, 2016a. ISO/IEC 27042:2016 Guidelines for the Analysis and Interpretation of Digital Evidence.

ISO/IEC, 2016b. ISO/IEC 27037:2016 Guidelines for Identification, Collection, Acquisition and Preservation of Digital Evidence.

ISO/IEC, 2016c. ISO/IEC 27041:2016 Guidance on Assuring the Suitability and Adequacy of Digital Investigation Method.

ISO/IEC, 2017. ISO/IEC 17025:2017 General Requirements for the Competence of Testing and Calibration Laboratories.

ISO/IEC, 2020. ISO/IEC 17000:2020 Conformity Assessment — Vocabulary and General Principles. ISO/IEC.

ISO/IEC/IEEE, 2018. ISO/IEC/IEEE International Standard - Systems and Software Engineering — Life Cycle Processes — Requirements Engineering, ISO/IEC/IEEE 29148:2018(E), pp. 1–104.

Marshall, A., Paige, R., 2018. Requirements in digital forensics method definition: observations from a UK study. Digit. Invest. 27, 23–29.

Sunde, N., Dror, I.E., 2021. A hierarchy of expert performance (hep) applied to digital forensics: reliability and biasability in digital forensics decision making. Forensic Sci. Int.: Digit. Invest. 37, 301175.

Tully, G., Cohen, N., Compton, D., Davies, G., Isbell, R., Watson, T., 2020. Quality standards for digital forensics: learning from experience in england & wales. Forensic Sci. Int.: Digit. Invest. 32, 200905.

UKAS, 2021. Who's Accredited? online at. https://www.ukas.com/find-an-organisation/. Last viewed 7th June 2021.

Appendix H

Synctriage

Digital Investigation 28 (2019) S134–S140

Contents lists available at ScienceDirect

Digital Investigation

journal homepage: www.elsevier.com/locate/diin

SyncTriage: Using synchronisation artefacts to optimise acquisition order

Christopher Hargreaves [a], [*], Angus Marshall [b]

[a] Department of Computer Science, University of Oxford, UK
[b] Department of Computer Science, University of York, UK

ARTICLE INFO

Article history:

Keywords:
Digital forensics
Triage
Synchronisation

ABSTRACT

While the number and variety of devices can be problematic in a digital investigation, it is also a problem for consumers. As a result, software developers have implemented synchronisation features to assist customers handle the multitude of devices that they now use. This paper describes how these synchronisation features can be exploited as part of a digital investigation to use the results from the examination of one device to infer content of other devices. This extracted information is potentially useful in determining the devices that should have the most resources expended during an investigation to obtain the most actionable evidence in the quickest and most efficient manner.

Introduction

Garfinkel (2010) discusses changes in the computer industry that can create challenges for digital forensics. While we are coming to the end of the ten-year period discussed in the paper, challenges such as the growing size of storage devices, the increasing need to analyse and correlate data from multiple devices, and pervasive encryption are still highly relevant. For example, Luck (2016) reported that in 2016 the Metropolitan Police conducted an estimated 49,036 digital forensic examinations. Also, another source (UK Parliament, 2016) describes a 2006 investigation that involved 274 computers and 1785 external storage devices, and Hall (2018) describes the ongoing problem with encrypted evidence.

In order to effectively handle large numbers of cases containing large numbers of exhibits using limited resources it is necessary to prioritise which devices to examine first. Casey et al. (2009) presents three levels of digital forensic examination: *survey/triage forensic inspection*, *preliminary forensic examination*, *in-depth forensic examination*. For the second two approaches, various models of digital forensics can be applied that contain differing levels of detail that extend Carrier (2003): *acquisition*, *analysis*, *presentation*. For the *survey/triage forensic inspection* level, which involves a "targeted review of all available media to determine which items contain the most useful evidence and require additional processing", this process is expanded in Overill et al. (2013) to encompass the following stages: i) pre-seizure: generating a list of anticipated digital devices, ii) search and seizure of devices, iii) post-seizure – screening of the seized devices for the likely existence of relevant evidence in a prioritised manner.

This last stage is described in the literature, for example, Rogers et al. (2006) introduced the Computer Forensics Field Triage Process Model (CFFTPM), which includes stages that include a review of user profile content, a timeline, and an internet artefact review, followed by case specific examination. Variations and extensions of this process are also implemented in a variety of commercial tools that can be used for triage, e.g. *ADF*, *Axiom*, *SPEKTOR* etc.

However, despite recognition of the benefit of a triage stage of digital investigations and known methods for extracting data that provides the best overview of a device to inform triage decisions, there is still a bottleneck in the process. Fig. 1 shows that in order to make a triage decision, data is needed, and therefore first access must be gained to devices, and then some basic extraction of data performed. As the number of devices in a case expands, and given additional protection mechanisms on devices, combined with the volume of data, this approach does not scale.

On initial consideration it appears that this bottleneck is impossible to overcome, but this paper provides an example of where in certain situations there is a work-around to allow a partial "examination" of devices to take place without necessarily needing to gain access to them or to extract any data from the devices themselves. The approach involves exploiting synchronisation

[*] Corresponding author.
E-mail addresses: chris@hargs.co.uk, christopher.hargreaves@cs.ox.ac.uk (C. Hargreaves), angus.marshall@york.ac.uk (A. Marshall).

https://doi.org/10.1016/j.diin.2019.01.022

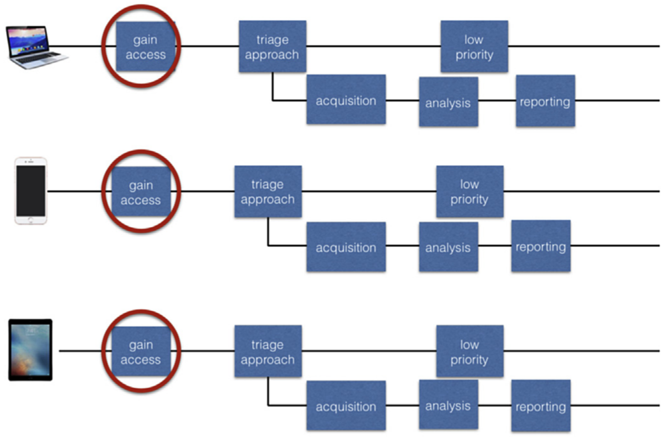


Fig. 1. This shows that existing triage approaches have an inherent bottleneck; at a minimum requires access to the device and inspection of some data, albeit not necessarily a full acquisition.

artefacts to infer the content of devices that have not been accessed, or indeed may not even have been identified and seized.

It should be noted that there is a challenge in reporting this work as it was conducted in 2016. Since then, there have been several papers about synchronisation artefacts that cover some of the components of the work. These are discussed in the related work section, but they are either focused on specific synchronisation artefacts, or problems in digital forensics other than triage. Therefore, despite the work published in this area since 2016, this paper is still able to make the following contributions:

- It provides a summary of some artefacts that can be used to extract synchronisation information,
- It presents and evaluates a new overall approach for inferring the existence of, and partial content of other devices.

The remainder of the paper is structured as follows: section 2 provides a summary of the previous work in this area, section 3 discusses the methodology for the research and sections 4 and 5 presents the results. The work is evaluated in section 6 and section 7 provides the conclusions and further work.

Related work

The need for triage in modern digital investigations and details of triage approaches have already been discussed in the introduction. Therefore, this section focuses on the existing work on synchronisation artefacts.

Several papers discuss cloud-based storage. For example Chung et al. (2012) provides an overall method for investigating such services. Several services are considered including *Amazon S3*, *Dropbox*, *Evernote*, and *Google Docs*. The paper understandably focuses on content, access records, and times of activity, but there is mention of *Evernote* storing references to the type of smartphone OS that created a note. Farina and Kechadi (2014) discusses artefacts left by *BTSync* (now *Resilio Sync*) but there were no artefacts reported that could be used to identify other devices that have synchronised content.

Operating system level synchronisation is also discussed in some previous work, for example, Friedman et al. (2012) discusses *iCloud* data and its synchronisation to *Apple* devices. The work provides information to determine if *iCloud* was enabled on a device, and also identified the same content present on multiple synchronised devices, including the same calendar web addresses, but “there was little evidence showing the two devices were connected to each other through *iCloud*” i.e. it was not possible to identify one device from the other. Further work in this area was reported in a later work Bubbins (2015), which specifically examined the *FindMyiPhone* feature of *iCloud* and was able to retrieve a list of devices connected to the account and their properties e.g. model, battery level etc. This was achieved using *MacOS* cached browser data and data from *iOS* devices prior to iOS 8.3.

Browser synchronisation work includes Wright (2015) which examined the synchronisation artefacts in Google Chrome and provided a means to determine that Chrome Synchronisation was enabled. It also showed that visits to web pages were synchronised across multiple devices, and provided a method to determine which URLs in the history were conducted on another device. It also discusses that the SyncData.sqlite3 database includes references to the other devices included in the synchronisation process. Boucher and Le-Khac (2018) provides a framework to address the problem of determining whether artefacts found on a device really originated on that device or if they were synchronised from somewhere else. Again, the research in the paper supports the existence of synchronisation artefacts, but the focus is very different to the research aim described in this paper since Boucher and Le-Khac (2018) treats synchronisation as a problem for an examination rather than exploiting it to assist investigations.

There are two obvious examples of this approach that are already performed. First, the examination of *iPhone* backups that are stored on a PC. In this case the examination of one device (the PC) results in the indirect examination of another device (the iPhone) that may not necessarily be in the possession of the examiner. The second example is the examination of *Windows* shortcut files and similar artefacts, which if they reference removable storage or network storage, provide information about the content of secondary devices that again, may not have even been identified.

Both of these examples are extremely useful techniques, but the overall concept, which is that data from one device can be used to infer the content of other devices has not been explored in terms of its generalisability as an approach.

Methodology

Overview

The overall aim of the research is to determine if it is possible to mitigate the “gain access” bottleneck of the triage process in order to determine the priority of examination. Fig. 2 illustrates the overall approach, which can be compared with Fig. 1. The method is to gain access to a subset of the total number of devices and focus on extracting artefacts that can be used both to determine the existence of other devices, and also to determine content and events on those other devices. This information can then be used to inform the overall triage process.

To address this aim, the research in this paper was carried out in several stages. Firstly, a review of apps was performed to determine the likely categories of apps that have some sort of synchronisation capability. To achieve this, the ‘app categories’ on both iOS and Android app stores were examined. A simple feasibility study was conducted, installing several apps from each category and reviewing the features from a user perspective.

Secondly, candidate apps were selected based on the features identified during the feasibility study, and previous work documented in previous work. For each of these apps, digital forensic artefact research was carried out, with a focus on obtaining information about other devices that were part of the synchronisation set. The methodology for this stage is discussed in more detail in

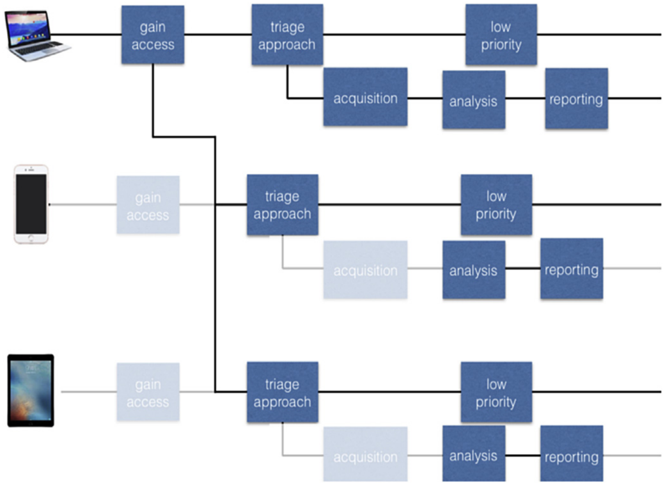


Fig. 2. This illustrates the new approach where access to, and analysis of one device can produce data that can be used to inform the triage process with data about other devices.

Section 3.2

Finally, a software prototype was designed and implemented that practically applied the artefact knowledge identified during the research phase.

Synchronisation artefact research

Experiments were conducted with a range of different apps (discussed in section 4.2). The experiments for each of these followed a typical pattern for digital forensic artefact research: experimental setup, data generation, data analysis.

Experimental setup: Several devices were obtained and set up. These included: iPhone 4S (iOS 9.3), iPhone 4S (iOS9 8.4 jailbroken), Nexus 5 (Android 6.0.2), Vodafone Prime 6 (Android 5.0.2 rooted), PC (Windows 10), Mac (OS X El Capitan). The devices were selected based on availability, and to provide coverage across a variety of operating systems, and for the mobile devices, variations of rooted/jailbroken or standard. The PC/Mac based devices were virtualized using *VMware*.

Data generation: Each app was installed on a subset of the devices, and accounts created. Data was then incrementally added on each device, with device acquisition/imaging taking place after each addition. Imaging of the PC/Mac was achieved by duplicating the VMDK files, and data from the mobile devices was acquired using *Magnet Acquire*. The mobile data was also processed using *pymobilesupport* (Hargreaves, 2016) to export the data into a format that was easier to analyse, e.g. mapping hash-based filenames to their original path on the device. The precise data generated depended on the nature of the application under test, but ranged from web visits for browsers, messages being sent for messaging apps, photographs being taken for photo-based apps, etc. For text-based data generation, unique and easily searchable data was used and test URLs related to the *SyncTriage* project were set up to make keyword searching during the analysis stage more effective.

Data analysis: The data analysis consisted in some cases of manual inspection of all the files within app file system containers, plus known keyword searching. This was mostly performed using *X-Ways Forensics* and various bespoke searching tools, for example to expand *plist* data stored within SQLite database fields, or to interpret *NSKeyedArchiver* formats.

In some cases, several iterations of this data generation/data analysis cycle were performed. For some apps with negative results, only a single iteration was performed, but for apps with complex data formats, additional iterations of this process were performed to confirm the formats, and interpret the data stored.

Results: experimental

App review

The app stores for Android and iOS were examined and lists of app categories extracted. For brevity the app categories are not listed in full here, but the points below discuss some of the categories that were determined to have the most potential in terms of synchronisation artefacts, either from common knowledge of app features, or artefacts discussed in the previous work section.

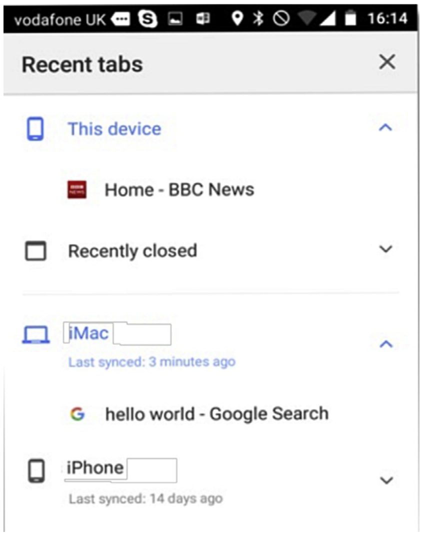
Browsers: The major browsers now all implement some form of synchronisation, from bookmarks to history. Fig. 3 shows one of the examples from a live device of what *SyncTriage* should aim to recover, where the examination of one device shows the tabs open on other devices.

Communication Apps: Many of the chat applications (*Telegram*, *Hangouts* etc.) are designed such that conversations can be carried out across multiple devices. There are many types of investigation where communication is critical and if possible, knowing the origin device of specific messages may provide insight into which device(s) should be prioritised.

Social Networking: Similar to the communication apps, social media apps present the same content on multiple devices. Either knowing that a specific device was used to share content at a particular time, or even that a device was in general use at a particular time could assist in prioritising the examination of a specific device.

Media & Video: Several apps e.g. *VLC* or *Plex* allow media to be viewed on remote storage on the local network. Certain investigation categories for example, indecent images of children, may benefit from the ability to determine the main source device of such media. This may be particularly true in cases where a network storage device has been physically concealed by the suspect and potentially not recovered during a seizure.

Note taking apps: Notes are used for a variety of purposes, from storing URLs, contacts, to-do lists etc. This app category has been included since if it were possible to identify the source device of a



vodafone UK 16:14

Recent tabs ×

This device

Home - BBC News

Recently closed

iMac

Last synced: 3 minutes ago

hello world - Google Search

iPhone

Last synced: 14 days ago

Fig. 3. An example of early indications of the type of synchronisation data that can be used to infer existence/content of other devices.

note of interest, it may provide insight into what a device was being used for at a specific time. That information could be used to prioritise further examinations of specific devices.

Photos: Photographs are a common media type to be synchronised over multiple devices. In addition, they are known to contain metadata including the device type, dates and times, and potentially geo-location information, on the originating device at least. Despite being a relatively simple and well understood artefact they have not yet been explored in the context of the *SyncTriage* process, i.e. can synchronised copies be used to identify the presence and nature of the devices from which they have originated.

Cloud Storage: Files stored in the cloud that are synchronised to other devices may provide the opportunity to determine the existence of other devices, or if metadata and the origin device is recoverable, may indicate that a specific device was in use at a particular time.

App selection for testing and results

After noting the categories above, considering known features of several common apps, and taking into account previous literature on synchronisation artefacts, the following 'candidate apps' were selected for more detailed examination: *Chrome*, *Firefox*, *Facebook Messenger*, *WhatsApp*, *Google Hangouts*, *Telegram*, *Viber*, *Skype*, *VLC*, *YouTube*, *Evernote*, *Google Photos*, *Instagram*, *Facebook*, *Twitter*, and *Dropbox*. This is far from an exhaustive list of applications that showed potential for synchronisation artefact recovery. However, the focus of this paper is to provide a proof of concept of the use of synchronisation artefacts for digital forensic triage, rather than an exhaustive artefact research piece.

There is insufficient space in this paper to provide full details on all the artefact results. Nevertheless, a summary of some of the key results for several of the applications studied are shown in Table 1.

Results: software prototype

Overall design

The overall design of the *sync_triage* tool is a plug-in based framework written in *Python* 3. It currently supports mounted disk images (tested using FTK Imager for mounting), iOS backup folders, and Android ADB backup files. The tool is currently command line only. The overall design of the software is shown in Fig. 4, and each of the stages are discussed in the subsequent sections.

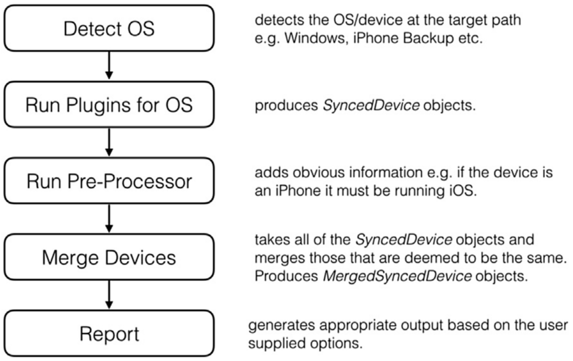

Detect OS
detects the OS/device at the target path e.g. Windows, iPhone Backup etc.
Run Plugins for OS
produces *SyncedDevice* objects.
Run Pre-Processor
adds obvious information e.g. if the device is an iPhone it must be running iOS.
Merge Devices
takes all of the *SyncedDevice* objects and merges those that are deemed to be the same. Produces *MergedSyncedDevice* objects.
Report
generates appropriate output based on the user supplied options.

Fig. 4. This shows the overall flow of the *sync_triage* software.

Operating system specific plugins

At time of writing, plugins have been written for *Windows*, *iOS*, and *Android* in order to demonstrate the feasibility of the approach on multiple platforms. The plugins implemented are listed in Table 2.

All the plugins produce lists of *SyncedDevices*, which include details about the inferred device, but also the provenance, i.e. the source device and forensic artefact from which it was inferred.

Preprocessor

This stage processes the list of devices and applies rules to add in any information that can be obviously inferred. For example, if the device is an *iPhone*, then the operating system is known to be iOS, even if the specific version of the operating system is not known. An example of before and after processing is shown in Fig. 5. Currently, these are hard coded rules and therefore need to be manually updated as new devices and operating systems are released. If information is inferred from external general knowledge, rather than taken directly from information found in the analysed data, then the value is enclosed in square brackets in order to distinguish it.

Merging devices

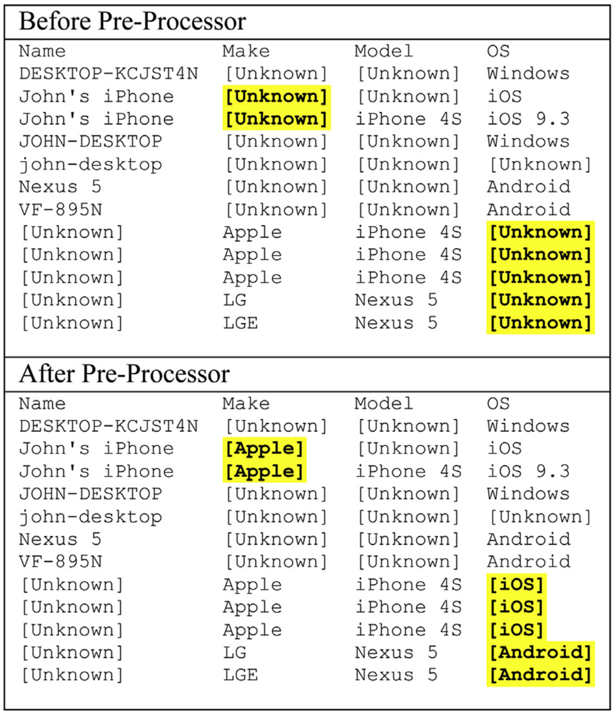
Once the pre-processing has been completed it is necessary to de-duplicate the results. The reason duplicates exist is that plugins simply extract device information from the various sources in the data being processed, and report them to the main program. As a result, a device which runs several apps that independently

Table 1
A summary of some of the artefacts recovered for some of the applications examined.

Application	Key Results
Chrome (Windows)	• Device names of linked synchronised devices • A subset of URLs visited on those devices
Windows 10 Mail (Windows)	• Notifications of service use on other devices can be extracted from emails stored on disk.
Evernote (Windows)	• The type of device used to create each note can be recovered, along with note creation time, and possibly GPS location.
Dropbox (Windows)	• EXIF data from synced photos reveal the make and model of devices, along with times of activity and possible GPS data,
Firefox (Windows)	• Open tabs on other devices
Firefox (iOS)	• List of synced devices, their names, types and operating system • Open tabs on those devices
Google Photos (Android)	• Information about photos stored on remote devices and the Google's servers. Provides make and model of other devices as well as timestamps of usage and possibly GPS data.
iCloud Sync (Windows)	• Using EXIF scanning: make, model, timestamps and possibly GPS data. • Using client.db: times that an iOS device was used to take pictures.
Skype (Windows)	• Content from other devices is recoverable, but it is not possible to determine the origin device.
Messages (iOS)	• Contains authorisation codes from service providers so the use of an app can be determined, although the device in use cannot be determined.

Table 2
A list of plugins currently implemented in the *sync_triage* prototype.

Target Device	Plugins Implemented
Windows	chrome – *extracts devices from SyncData.sqlite3.* dropbox_photos – *scans exif data in photos for devices.* email_scanner – *scans emails for "account in use on device" emails.* evernote – *scans note_attr table for references to devices.* icloud_photos – *scans client.db for photos in server_items table.* viber – *extracts phone number of device used for service.*
Mac OS	–
iOS	chrome – *extracts devices from SyncData.sqlite3.* firefox – *scans browser.db for clients.* sms – *searches for references to service authentication messages.*
Android	google_photos – *extracts exif data from remote_media table.*



```
Before Pre-Processor
Name              Make        Model       OS
DESKTOP-KCJST4N   [Unknown]   [Unknown]   Windows
John's iPhone     [Unknown]   [Unknown]   iOS
John's iPhone     [Unknown]   iPhone 4S   iOS 9.3
JOHN-DESKTOP      [Unknown]   [Unknown]   Windows
john-desktop      [Unknown]   [Unknown]   [Unknown]
Nexus 5           [Unknown]   [Unknown]   Android
VF-895N           [Unknown]   [Unknown]   Android
[Unknown]         Apple       iPhone 4S   [Unknown]
[Unknown]         Apple       iPhone 4S   [Unknown]
[Unknown]         Apple       iPhone 4S   [Unknown]
[Unknown]         LG          Nexus 5     [Unknown]
[Unknown]         LGE         Nexus 5     [Unknown]

After Pre-Processor
Name              Make        Model       OS
DESKTOP-KCJST4N   [Unknown]   [Unknown]   Windows
John's iPhone     [Apple]     [Unknown]   iOS
John's iPhone     [Apple]     iPhone 4S   iOS 9.3
JOHN-DESKTOP      [Unknown]   [Unknown]   Windows
john-desktop      [Unknown]   [Unknown]   [Unknown]
Nexus 5           [Unknown]   [Unknown]   Android
VF-895N           [Unknown]   [Unknown]   Android
[Unknown]         Apple       iPhone 4S   [iOS]
[Unknown]         Apple       iPhone 4S   [iOS]
[Unknown]         Apple       iPhone 4S   [iOS]
[Unknown]         LG          Nexus 5     [Android]
[Unknown]         LGE         Nexus 5     [Android]
```

Fig. 5. An example of the pre-processing of results, showing before and after inferencing of missing data.

synchronise will be reported by multiple plugins. The deduplication process attempts to eliminate all obvious duplications of a device.

Devices are merged as a result of several relatively simplistic rules, for example:

- Merge if has the same name
- Merge if same make and model

When merging does occur, the rules used are preserved in the log file and the individual devices that were combined are recorded within the newly created merged device so that full provenance of results can be inspected. This can be seen in the 'Refs' column in Fig. 6. These are very simple rules at present, but in future, more complex logic could be substituted in for this phase.

Reporting

The results shown earlier in Fig. 6 demonstrated the default output from the tool. There are two other modes that provide more details. The –details option displays additional information about the discovered devices, shown in Fig. 8. You can see that in addition to the name, make, model and operating system that was shown in the summary view, much more information has been recovered. You can see software that is known to have been installed on the device, information about web visits conducted on the device, a basic timeline of activity (in this case just reporting pictures taken, time and location). The details view also reports the original *synced_device* objects that were merged to infer the existence and information about this device, which in turn provide the original file path from which that information was extracted.

The other display option that has been implemented is the 'Universal Timeline' view, invoked with the –timeline option and shown in Fig. 7. This extracts the events from each inferred device and presents them all in a timeline. The use case for this feature is to assist with decision making about which device to examine, particularly in cases where the time of the alleged offence is known. It may be possible to identify the device that was in use closest to the time of the incident.

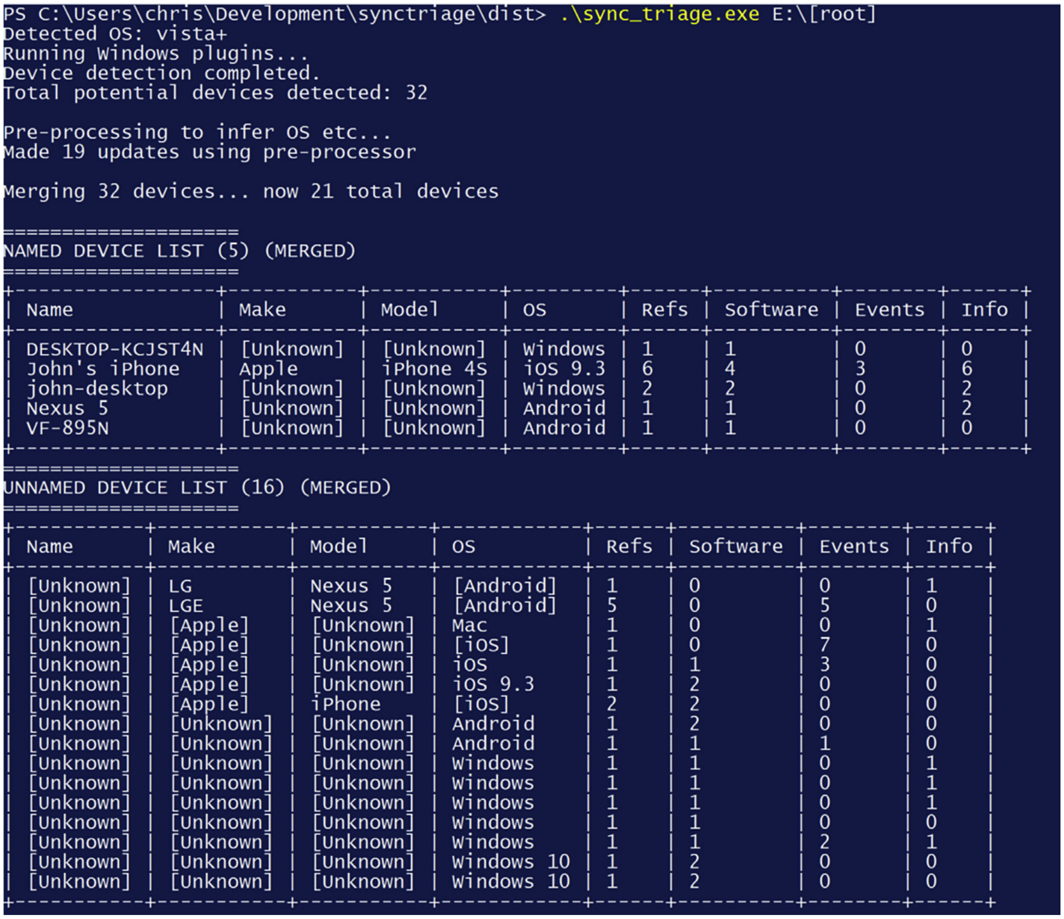
Evaluation

Overall the *SyncTriage* has been a successful proof of concept. In Fig. 8, many details can be seen about the use of an iPhone that has not been accessed at all. Also, in Fig. 7, only 2 of the 21 timeline events shown occurred on the device that is being examined.

In terms of use cases for this approach, *SyncTriage* should help with: detecting devices that have not been seized, determining which device was in use at the time of an offence, inferring content on devices that have not yet been forensically processed. All of these use cases ultimately will help in prioritising the devices to examine first and retrieve actionable evidence as quickly and efficiently as possible.

There are however limitations to this research. For example, the review of apps was far from systematic, although as a proof of concept piece of work, this is not a major concern as the apps selected have allowed the approach to be demonstrated. However, what it does not show is the scale of the effectiveness of the approach. For example, a number of plugins have been produced for a *Windows* examination, but far fewer for *Android* and *iOS*. Even in the case of the existing plugins they are likely to be highly sensitive to the version of the application or operating system. This was the rationale of the plugin-based architecture, but that does not reduce the overhead of conducting the research and software development to keep this artefact extraction and processing up-to-date.

In terms of performance, the program runs on the sample *Windows 10* image and an example 'real world' system in less than a second, since it precisely targets specific artefacts. This could be further cut down as the program is currently single threaded, but



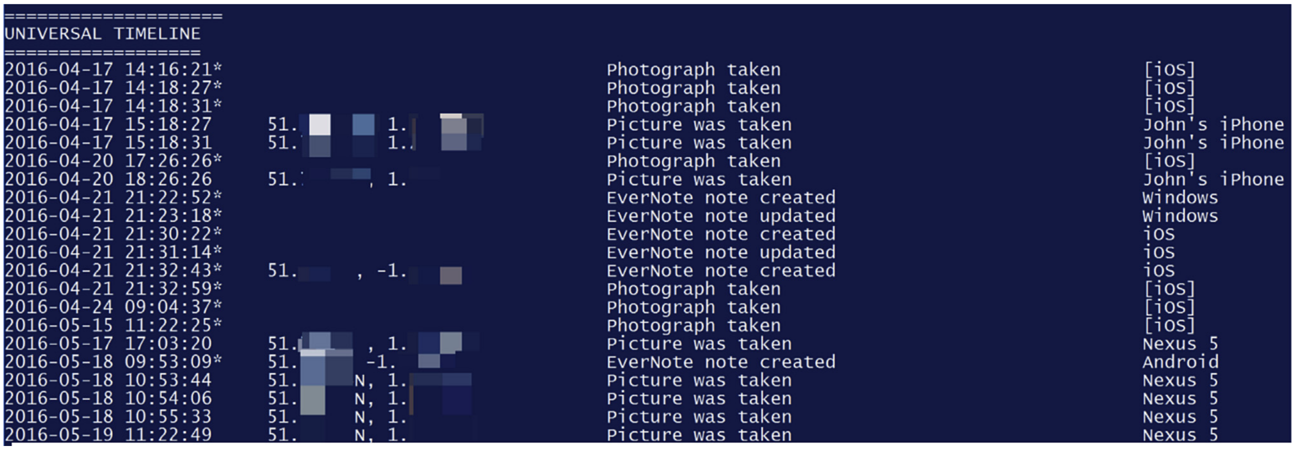
```
PS C:\Users\chris\Development\synctriage\dist> .\sync_triage.exe E:\[root]
Detected OS: vista+
Running Windows plugins...
Device detection completed.
Total potential devices detected: 32

Pre-processing to infer OS etc...
Made 19 updates using pre-processor

Merging 32 devices... now 21 total devices

====================
NAMED DEVICE LIST (5) (MERGED)
====================
+------------------+-----------+-----------+----------+------+----------+--------+------+
| Name             | Make      | Model     | OS       | Refs | Software | Events | Info |
+------------------+-----------+-----------+----------+------+----------+--------+------+
| DESKTOP-KCJST4N  | [Unknown] | [Unknown] | Windows  | 1    | 1        | 0      | 0    |
| John's iPhone    | Apple     | iPhone 4S | iOS 9.3  | 6    | 4        | 3      | 6    |
| john-desktop     | [Unknown] | [Unknown] | Windows  | 2    | 2        | 0      | 2    |
| Nexus 5          | [Unknown] | [Unknown] | Android  | 1    | 1        | 0      | 2    |
| VF-895N          | [Unknown] | [Unknown] | Android  | 1    | 1        | 0      | 0    |
+------------------+-----------+-----------+----------+------+----------+--------+------+
====================
UNNAMED DEVICE LIST (16) (MERGED)
====================
+-----------+-----------+-----------+------------+------+----------+--------+------+
| Name      | Make      | Model     | OS         | Refs | Software | Events | Info |
+-----------+-----------+-----------+------------+------+----------+--------+------+
| [Unknown] | LG        | Nexus 5   | [Android]  | 1    | 0        | 0      | 1    |
| [Unknown] | LGE       | Nexus 5   | [Android]  | 5    | 0        | 5      | 0    |
| [Unknown] | [Apple]   | [Unknown] | Mac        | 1    | 0        | 0      | 1    |
| [Unknown] | [Apple]   | [Unknown] | [iOS]      | 1    | 0        | 7      | 0    |
| [Unknown] | [Apple]   | [Unknown] | iOS        | 1    | 1        | 3      | 0    |
| [Unknown] | [Apple]   | [Unknown] | iOS 9.3    | 1    | 2        | 0      | 0    |
| [Unknown] | [Apple]   | iPhone    | [iOS]      | 2    | 2        | 0      | 0    |
| [Unknown] | [Unknown] | [Unknown] | Android    | 1    | 2        | 0      | 0    |
| [Unknown] | [Unknown] | [Unknown] | Android    | 1    | 1        | 1      | 0    |
| [Unknown] | [Unknown] | [Unknown] | Windows    | 1    | 1        | 0      | 1    |
| [Unknown] | [Unknown] | [Unknown] | Windows    | 1    | 1        | 0      | 1    |
| [Unknown] | [Unknown] | [Unknown] | Windows    | 1    | 1        | 0      | 1    |
| [Unknown] | [Unknown] | [Unknown] | Windows    | 1    | 1        | 0      | 0    |
| [Unknown] | [Unknown] | [Unknown] | Windows    | 1    | 1        | 2      | 1    |
| [Unknown] | [Unknown] | [Unknown] | Windows 10 | 1    | 2        | 0      | 0    |
| [Unknown] | [Unknown] | [Unknown] | Windows 10 | 1    | 2        | 0      | 0    |
+-----------+-----------+-----------+------------+------+----------+--------+------+
```

Fig. 6. Example output from *sync_triage* after examining a disk image, showing four other devices detected and 16 other entries that could not be automatically merged.



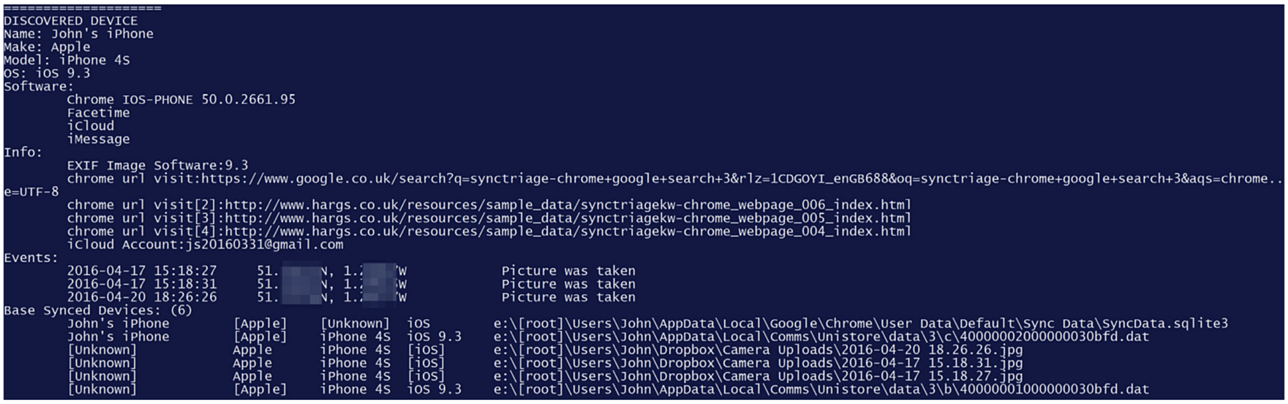
```
====================
UNIVERSAL TIMELINE
====================
2016-04-17 14:16:21*                                    Photograph taken             [iOS]
2016-04-17 14:18:27*                                    Photograph taken             [iOS]
2016-04-17 14:18:31*                                    Photograph taken             [iOS]
2016-04-17 15:18:27    51.[redacted] 1.[redacted]       Picture was taken            John's iPhone
2016-04-17 15:18:31    51.[redacted] 1.[redacted]       Picture was taken            John's iPhone
2016-04-20 17:26:26*                                    Photograph taken             [iOS]
2016-04-20 18:26:26    51.[redacted], 1.[redacted]      Picture was taken            John's iPhone
2016-04-21 21:22:52*                                    EverNote note created        Windows
2016-04-21 21:23:18*                                    EverNote note updated        Windows
2016-04-21 21:30:22*                                    EverNote note created        iOS
2016-04-21 21:31:14*                                    EverNote note updated        iOS
2016-04-21 21:32:43*   51.[redacted] , -1.[redacted]    EverNote note created        iOS
2016-04-21 21:32:59*                                    Photograph taken             [iOS]
2016-04-24 09:04:37*                                    Photograph taken             [iOS]
2016-05-15 11:22:25*                                    Photograph taken             [iOS]
2016-05-17 17:03:20    51.[redacted] , 1.[redacted]     Picture was taken            Nexus 5
2016-05-18 09:53:09*   51.[redacted] -1.[redacted]      EverNote note created        Android
2016-05-18 10:53:44    51.[redacted] N, 1.[redacted]    Picture was taken            Nexus 5
2016-05-18 10:54:06    51.[redacted] N, 1.[redacted]    Picture was taken            Nexus 5
2016-05-18 10:55:33    51.[redacted] N, 1.[redacted]    Picture was taken            Nexus 5
2016-05-19 11:22:49    51.[redacted] N, 1.[redacted]    Picture was taken            Nexus 5
```

Fig. 7. This shows the 'universal timeline' view of *sync_triage* which combines all of the identified events from all the extrapolated devices into a single timeline. This has significant potential if the time of an offense is known and it can be correlated with a particular device being in use at that time.

with the current runtimes this is not a priority.

It has been shown that a significant amount of automation is possible, certainly for low level device detail extraction, for some inference of missing information, and the merging of some devices. However, even with enhancements to the merging process there will be limits to what is possible for an automated process to do. Therefore, part of the development of this approach must include a more interactive user interface that allows the devices to be explored, manually merged, and the automatic merging reapplied in light of the user suppled information.

Nevertheless, more automation in the merging process may be possible, including information from the events, e.g. if an event is recorded for an unnamed *iOS* device, and an event was recorded for a named *iOS* device at a very similar time, then this could be



```
====================
DISCOVERED DEVICE
Name: John's iPhone
Make: Apple
Model: iPhone 4S
OS: iOS 9.3
Software:
        Chrome IOS-PHONE 50.0.2661.95
        Facetime
        iCloud
        iMessage
Info:
        EXIF Image Software:9.3
        chrome url visit:https://www.google.co.uk/search?q=synctriage-chrome+google+search+3&rlz=1CDGOYI_enGB688&oq=synctriage-chrome+google+search+3&aqs=chrome..
e=UTF-8
        chrome url visit[2]:http://www.hargs.co.uk/resources/sample_data/synctriagekw-chrome_webpage_006_index.html
        chrome url visit[3]:http://www.hargs.co.uk/resources/sample_data/synctriagekw-chrome_webpage_005_index.html
        chrome url visit[4]:http://www.hargs.co.uk/resources/sample_data/synctriagekw-chrome_webpage_004_index.html
        iCloud Account:js20160331@gmail.com
Events:
        2016-04-17 15:18:27     51.[redacted]N, 1.[redacted]W           Picture was taken
        2016-04-17 15:18:31     51.[redacted]N, 1.[redacted]W           Picture was taken
        2016-04-20 18:26:26     51.[redacted]N, 1.[redacted]W           Picture was taken
Base Synced Devices: (6)
        John's iPhone        [Apple]    [Unknown]  iOS        e:\[root]\Users\John\AppData\Local\Google\Chrome\User Data\Default\Sync Data\SyncData.sqlite3
        John's iPhone        [Apple]    iPhone 4S  iOS 9.3    e:\[root]\Users\John\AppData\Local\Comms\Unistore\data\3\c\40000002000000030bfd.dat
        [Unknown]            Apple      iPhone 4S  [iOS]      e:\[root]\Users\John\Dropbox\Camera Uploads\2016-04-20 18.26.26.jpg
        [Unknown]            Apple      iPhone 4S  [iOS]      e:\[root]\Users\John\Dropbox\Camera Uploads\2016-04-17 15.18.31.jpg
        [Unknown]            Apple      iPhone 4S  [iOS]      e:\[root]\Users\John\Dropbox\Camera Uploads\2016-04-17 15.18.27.jpg
        [Unknown]            [Apple]    iPhone 4S  iOS 9.3    e:\[root]\Users\John\AppData\Local\Comms\Unistore\data\3\b\40000001000000030bfd.dat
```

Fig. 8. This shows the 'details' view of an inferred device. It can be seen that an iPhone 4S exists running iOS 9.3, that it is known to contain several apps, and was in use at specific times in 2016, and visited several websites using *Google Chrome*.

considered to be a 'session' and the devices could be merged.

Furthermore, the events/timeline feature would benefit from expansion, so that the times that particular devices were in used can be more easily and reliably determined.

Finally, at present, this tool analyses only one device at a time. It would be beneficial for additional devices to be added to the set from different sources as this is needed to explore the idea of 'acquisition order optimisation'.

Conclusions and future work

This research has tested the concept of exploiting synchronisation artefacts on one device to extrapolate the existence and content of other devices for the purposes of digital forensic triage. The approach shows promise and further work involves expanding the range of plugins to test the extent to which artefacts exist that can be used for this device inference. There is also additional work to do on the concept of merging the inferred devices in a more sophisticated manner, or providing a user interface that allows the investigator to easily manually merge devices together. Finally, process-based research also needs to be conducted on how this approach can be integrated into digital forensic workflows and used to improve the acquisition order of devices.

Acknowledgements

This work was funded under a Home Office Innovation call through SBRI/Innovate UK.

References

Boucher, Le-Khac, 2018. Forensic framework to identify local vs synced artefacts. Digit. Invest. 24 (Suppl.), 68–75.

Bubbins, 2015. Identification of Devices Connected to a Suspect's iCloud Account when Using the Application Find My iPhone. MSc Thesis. Cranfield University.

Carrier, 2003. Defining digital forensic examination and analysis tools using abstraction layers. International Journal of Digital Evidence 1 (4).

Casey, et al., 2009. Investigation Delayed Is Justice Denied: Proposals for Expediting Forensic Examinations of Digital Evidence.

Chung, H., Park, J., Lee, S., Kang, C., November 2012. Digital forensic investigation of cloud storage services. Dig. Invest. 9 (2), 81–95.

Farina, Scanlon, Kechadi, 2014. BitTorrent sync: first impressions and digital forensic implications. Digit. Invest. 11 (Suppl. 1), 77–86.

Friedman, Brunty, Fenger, 2012. A Digital Forensic Analysis on the iCloud® and its Synchronization to Apple® Devices. http://www.marshall.edu/forensics/files/FRIEDMANRACHEL-Research-Paper-08242012.pdf. (Accessed 23 October 2018).

Garfinkel, 2010. Digital Forensics Research: The Next 10 Years. DFRWS 2010.

Hall, 2018. The Reg Visits London Met Police's Digital and Electronics Forensics Labs. The Resister. https://www.theregister.co.uk/2018/01/22/digital_forensics/. (Accessed 23 October 2018).

Hargreaves, 2016. Pymobilesupport. https://bitbucket.org/chrishargreaves/pymobilesupport.

Luck, 2016. Challenges and Opportunities for Statistics in Digital Forensics. https://www.turing-gateway.cam.ac.uk/sites/default/files/asset/doc/1612/Luck.pdf. (Accessed 28 October 2018).

Overill, Silomon, Roscoe, 2013. Triage template pipelines in digital forensic investigations. Digit. Invest. 10 (2), 168–174.

Rogers, 2006. Computer Forensics Field Triage Process Model. J. Digital Forensics Secur. Law 1 (2).

UK Government, 2016. Annex: Visits to the Metropolitan Police Forensic Services and LGC Forensics. https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/501/50110.htm. (Accessed 23 October 2018).

Wright, 2015. Forensic Artefacts Related to Google's Chrome Synchronisation Feature. MSc Thesis. Cranfield University.

Appendix I

Call data obfuscation

Digital Investigation 29 (2019) 82–90

Contents lists available at ScienceDirect

Digital Investigation

journal homepage: www.elsevier.com/locate/diin

CaseNote: Mobile phone call data obfuscation & techniques for call correlation☆

Angus M. Marshall [a, *, 1], Peter Miller [b, 1]

[a] n-gate ltd., UK
[b] West Midlands Police, UK

ARTICLE INFO

Article history:
Received 20 November 2018
Received in revised form
12 March 2019
Accepted 12 March 2019
Available online 20 March 2019

ABSTRACT

The use of call data records (CDRs) to establish links between suspects is well known and understood. In a number of major enquiries in the UK, however, it was found that CDRs contained apparently erroneous or nonsensical data which prevented the use of well-established techniques based on caller IDs contained within CDRs. Further analysis suggested that some form of number "spoofing" was being used and techniques were developed to associate calls made using one of the two mechanisms which appeared to be in operation. A working hypothesis about how the calls are being handled in order to implement obfuscation has been developed, based on additional data leaked by some of the mobile network providers.

Introduction

In 2017, law enforcement agents investigating organised crime gangs, in the UK, became aware that usual techniques used to map contact between members of the gangs and their associates, based on call data records (CDR), were not working properly. Normally, it has been possible to identify associates from caller ID and IMSI/IMEI data contained in the call records of another person.

In recent cases, however, the CDR data directed investigators towards obviously unconnected innocent third parties, while in others the data contained nonsensical unissued numbers which could not be linked to any real subscriber. As a result, investigation of the gangs was becoming difficult and much time and effort was being wasted.

The first named author was contacted and put in touch with specialists within UK law enforcement agencies who had been working on the problem. It was apparent that the suspects in question were using standard mobile phones, albeit of the "feature phone" class with deliberately limited functionality to prevent leakage of identifying data, but with customised SIMs which were traced to a UK commercial reseller. The SIMs themselves had been sold by one of the reseller's agents in a country where it is difficult to obtain data. Material found at some of the suspect's premises and on their computers suggested that they had been researching the use of "stealth", "spy" or "spoofer" SIMs which promised higher levels of encryption than normal, coupled with obfuscation of called and calling numbers.

Following analysis of CDRs from several cases, consultation with SIM providers and examination of web sites advertising "stealth" SIMS, the following description has been produced.

"Stealth" SIMs

A "stealth" SIM is a SIM which can be used to access the mobile phone network, but which makes use of network management features (USSD (European Telecommunication, 1996) for user interaction/control and CAMEL (ETSI, 2014; Ghadialy, 2004–2019) for call routing and handling) to mask its location and true identity from anyone receiving calls from it. It appears that stealth SIMs can also be used to allow calls to be made to false numbers (either full numbers or short codes (ETSI, 1996), in order to hide the identity of the called party. The diagrams below show four calling situations - a normal mobile to mobile call, a roaming mobile to mobile call, and two different call mechanisms involving the use of "stealth" SIMs.

☆ The content of this paper has been developed from evidential reports which have been presented in court. We are grateful to the reviewers who asked questions which allowed us to investigate the obfuscation mechanisms more fully and produce a better explanation.

* Corresponding author.
E-mail address: angus@n-gate.net (A.M. Marshall).

[1] Note: Although Mr. Marshall and DC Miller are the named authors of this article, the work presented includes the results of efforts by law enforcement investigators who cannot be named. Their contribution is gratefully acknowledged.

https://doi.org/10.1016/j.diin.2019.03.004

Background: mobile phone network "stealth" and call routing manipulation features

A modern mobile phone network can be considered to have 3 channels associated with it:

1. A control channel, used to establish calls, carry text messages (SMS) and switch network features on and off at the request of users
2. A voice channel, used when a call has been established to carry audible data
3. A data channel (sometimes known as GPRS) used for Internet data sessions.

In the case of "stealth" SIMs, the use of the control channel is of particular interest.

Modern networks can make use of a specification known as CAMEL (*Customised Applications for Mobile networks Enhanced Logic*) (ETSI, 2014; Ghadialy, 2004–2019) to provide a way for SIM's home networks to manage calls and costs from wherever they are in the world. Because CAMEL interactions happen between networks it is rare for them to be recorded in customer call data records or be made available through billing data, but we have seen such data in a small number of cases.

Typical functions include:

- Arranging for the handset to register with a company automated switchboard so that calls made to an extension within the company can be automatically routed to the mobile handset (Sweden used a system similar to this for their "Call a Swede" campaign in 2016. A single phone number was published and anyone calling it would be connected to a random volunteer whose number was unknown to the caller. The volunteers registered their numbers with the central service in order to allow it to route calls to them). This method is also used to provide "anonymous" temporary numbers for use in online advertisements and dating sites.
- Arranging for calls made from the mobile handset to be sent via the company switchboard so that they can be logged and/or sent over a more cost-effective route, such as VOIP.

In these cases, the direct dial number for the handset would not be apparent to either callers or called parties as they would be interacting with a number assigned by the switchboard. Typically, the switchboard can present any number as the CLID when a call is routed through it. This is a method often used by cold callers (unsolicited callers) and scam callers to make identification more difficult.

Some network providers use this functionality to provide additional features on their network such as:

- Voice changing - the call is routed through a system that manipulates the audio to disguise the caller
- Roaming call back - in order to avoid roaming charges, the initial caller can request that the network calls the intended party and then calls back to the mobile handset using a cheaper route.

In order for these features to be available, the SIM provider needs to have a way to route calls via their own switchboard or a rented virtual switchboard. Such services are readily available and can make use of low cost Voice Over Internet Protocol (VOIP) technology which uses the Internet for call distribution. Systems which provide call routing can be programmed by remote administrators, using dedicated software or via web-based interfaces. Use of software to control the system could allow CLIDs and routes to be changed whenever a call is made or completed.

Most mobile phone users are familiar with the use of SMS (Short Message Service) which is commonly used to send text message from one handset to another using spare capacity on the network's control channel. A second messaging system exists in modern networks, known as USSD (European Telecommunication, 1996) (*Unstructured Supplementary Service Data*), although it is somewhat more limited in its mode of operation as far as the user is concerned. While SMS allows handset to handset communications, with an element of store-and-forward behaviours (i.e. messages will be stored by the network until the receiving handset is available), USSD allows handset to application communications during a live session. i.e., while the handset is active. USSD commands can be sent to programs running on the networks' "servers" in order to use special services or change configuration.

Of particular interest, in the situation we are considering, are commands which take the form of special codes, typically beginning with an asterisk (*) or hash (#) and ending with a hash (#) which are instructions to use services offered by the SIM's home network. Common usages include checking account balances, obtaining sports scores, news & weather forecasts, blocking certain numbers, and sending a message to another number to request a call back.

USSD functions can be implemented, at will, by network providers and can be used to control the operation of "stealth" features including, commonly, changing numbers used for redirection or CLID spoofing.

Some network operators offer "stealth" services to allow users to make "prank" calls without fear of being traced, or to allow anonymous dating/chat conversations to be carried without enabling stalking. However, several providers offer a "stealth SIM solution" or "anonymous SIM" which is advertised as being able to allow users to appear to be located in other parts of the world, present falsified (spoofed) CLIDs, prevent cell-site geolocation, and change the caller's voice. Such features would allow users to prevent government and law enforcement agencies from obtaining data about their locations and calls. The mechanism offered by at least one provider also prevents the return of calls, as the CLID shown to call recipients is false. SIMs of this type cost around $300 (USA) per month to rent and operate.

Normal mobile to mobile calling using standard SIMs

For a call to be connected, both handsets must be switched on and registered with the mobile network via a nearby cell mast. Each cell has a unique identifier, and each handset is identified by its IMEI and the IMSI of the SIM in it. When a user dials a number, the network looks up the IMSI identified with it and routes the call via the cell with which that IMSI is registered. An overview of the process is shown in Fig. 1.

Typically, the cells in which the call starts and ends for each party are recorded by the network and can be disclosed when required. In this way, the approximate location of each handset involved in the call can be determined by finding the location of the cell(s) to which it was connected.

This assumes that both handsets are using SIMs issued by the network to which they are connected. If the either SIM is roaming onto another network, there will be additional interactions between the network to which it is connected the SIM's home network, to check data held in the Home Location Register (HLR) in order to determine if the SIM is permitted to connect to the other network and thus be given an entry in the Visitor Location Register (VLR), as well as further checks to determine if calls of a given type can be made or received, and how they should be handled (Figs. 2 and 3.). The interaction between the visited and home networks makes use of the CAMEL protocols in order to exchange details of the call type and desired call handling/routing.

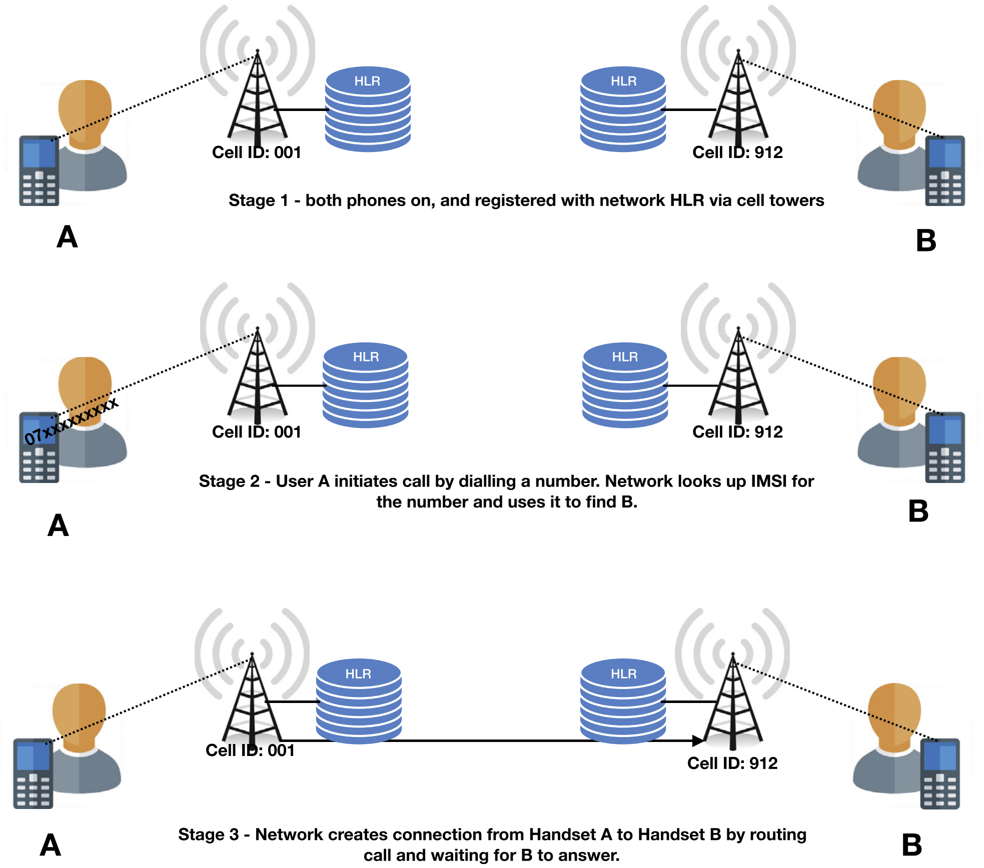

HLR
Cell ID: 001
HLR
Cell ID: 912
Stage 1 - both phones on, and registered with network HLR via cell towers
A
B
07xxxxxxxxxx
HLR
Cell ID: 001
HLR
Cell ID: 912
Stage 2 - User A initiates call by dialling a number. Network looks up IMSI for the number and uses it to find B.
A
B
HLR
Cell ID: 001
HLR
Cell ID: 912
Stage 3 - Network creates connection from Handset A to Handset B by routing call and waiting for B to answer.
A
B

Fig. 1. Normal mobile to mobile calling - SIMs on own home networks.

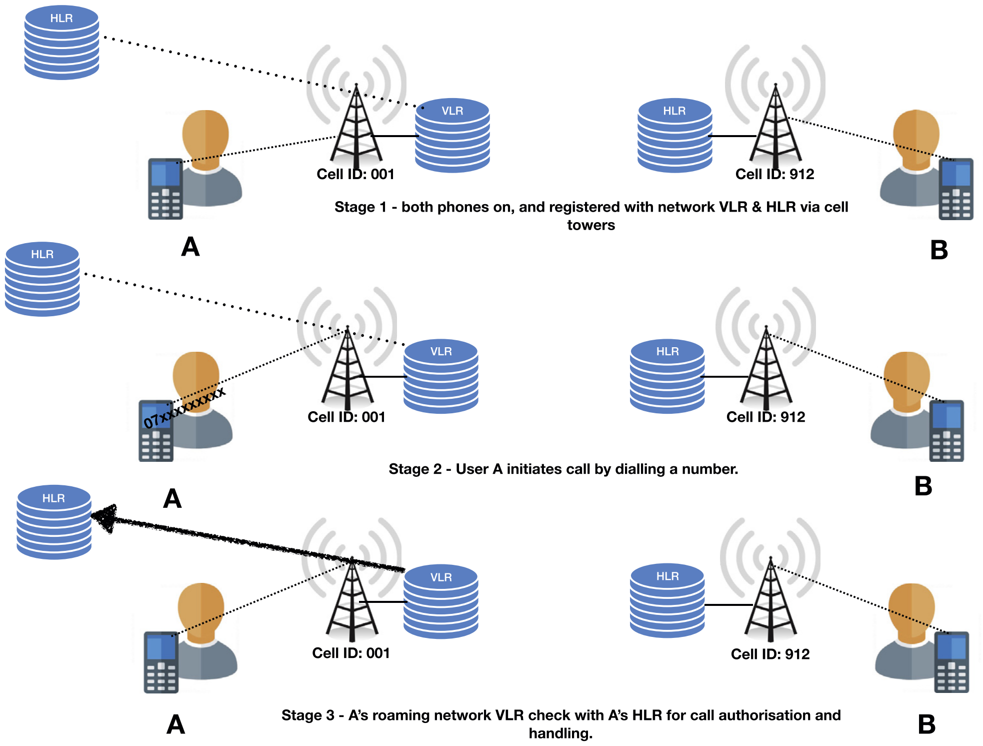

HLR
VLR
Cell ID: 001
HLR
Cell ID: 912
Stage 1 - both phones on, and registered with network VLR & HLR via cell towers
A
B
HLR
07xxxxxxxxxx
VLR
Cell ID: 001
HLR
Cell ID: 912
Stage 2 - User A initiates call by dialling a number.
A
B
HLR
VLR
Cell ID: 001
HLR
Cell ID: 912
Stage 3 - A's roaming network VLR check with A's HLR for call authorisation and handling.
A
B

Fig. 2. Roaming SIM requiring interaction between roaming network VLR and home network HLR. Initial stages. No “spoofing” involved.

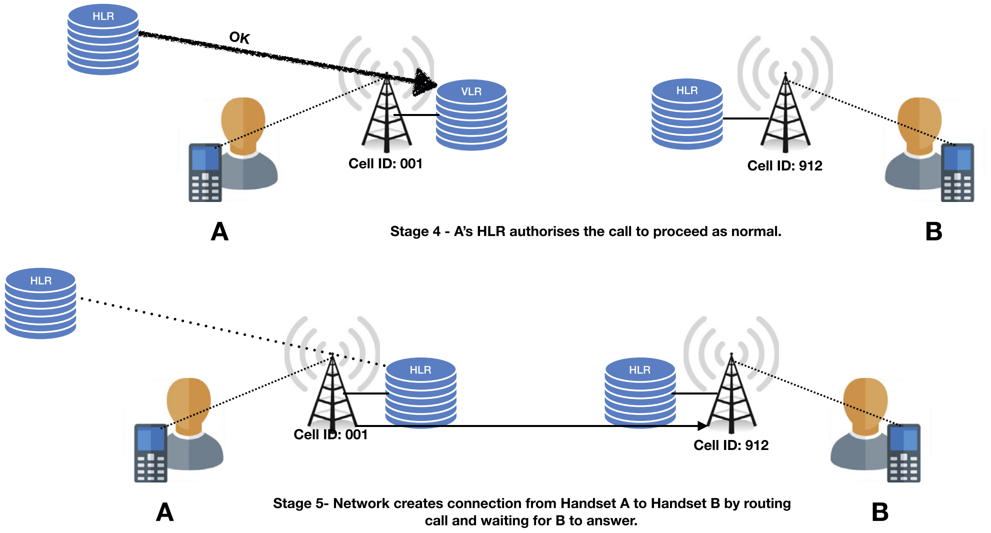

HLR
OK
VLR
Cell ID: 001
HLR
Cell ID: 912
A
Stage 4 - A's HLR authorises the call to proceed as normal.
B
HLR
HLR
HLR
Cell ID: 001
Cell ID: 912
A
Stage 5- Network creates connection from Handset A to Handset B by routing call and waiting for B to answer.
B

Fig. 3. Roaming SIM requiring interaction between roaming network VLR and home network HLR. Call completion without spoofing mechanism. Interaction between the networks uses the CAMEL protocols to exchange call handling data and instructions.

Stealth" SIM calling another "stealth" SIM

Stealth SIMs make use of network features which exist to allow for call costs to be managed and reduced by allowing redirection to an alternative provider or route. As far as we can ascertain, the organisations behind the stealth SIMs have created their own Mobile Virtual Network Operators (MVNOs) and use SIMs from legitimate providers which are re-programmed to use these MVNOs as their home networks. Two mechanisms exist for call handling - either through call redirection (below) or roaming callback (below).

Call redirection

Because the SIM is roaming, the VLR with which it is registered will interact with the HLR in order to check authorisation to make a call, and to obtain any special handling for the call. The HLR will, through the use of appropriate CAMEL programming and protocols, return a redirect request which causes the roaming network to pass the call to an alternative number. In the cases we have examined, the CAMEL transaction data shows that the alternative number is that of a SIP (Session Initiation Protocol) gateway, in the country in which the SIM is roaming, in order to hand the call over to a Voice Over Internet Protocol (VOIP) carrier for onward transmission.

The outbound VOIP gateway is also programmed to send a false calling line identifier (CLID - the claimed number calling the final destination), which can be changed to prevent patterns of calls from particular numbers being identified, to the receiving phone. The mechanism involved in this is similar to that used by many "cold calling" companies such as those who make unsolicited calls about PPI (payment protection insurance), boiler upgrades, or accident compensation. Falsifying the CLID makes it effectively impossible for the called party, or anyone viewing their CDR, to identify who has called them. In some cases, the stealth SIM subscriber can exert some control over the CLID by sending appropriate USSD codes to the home network to change the CLID. We have observed that some of the VOIP carriers use "lazy" numbers (see below) by default whilst others appear to change the presented number periodically, using a set of unallocated numbers. We have also noted that some users of the stealth SIMs appear to deliberately use number which are known to them (e.g local taxi firms) and it is thought that this is done to further confuse investigators. An examination of manuals and websites for various stealth SIM providers confirms that a USSD mechanism is usually provided to allow users to change their presented "spoof" CLID at will.

The calling process is shown in the diagrams below (Fig. 4 & Fig. 5).

In cases where this mechanism is involved, we could expect the billing data for the originating handset to show the call being redirected, although the redirected number is not usually recorded in the CDR, and the call being connected as a normal call. If the network stores CAMEL data, this data may contain more information about the exact nature of the redirection involved (see Tables 1 and 2 and section Correlating data, below).

The billing data for the called handset would show an incoming call at the same time as the original call was placed, usually with a delay of a few (typically no more than 2) seconds because of the interactions required between caller and network, from a "random" number which is not obviously related to the caller and the same data will appear in the receiving handset log.

Where the called and false incoming CLID appear to be mobile numbers it will be seen that no IMSI, IMEI or cell data for those numbers is available in the CDRs for the caller or recipient. This is an indicator that a stealth mechanism may have been used.

In a few cases, where CAMEL data has been present, we have been able to identify the SIP gateway and VOIP carrier, confirming that the call is being transmitted across a VOIP network rather than any other option.

As far as we can ascertain, there is no obvious direct relationship between the MVNO and the VOIP carrier beyond customer and service provider, although we have found that some VOIP carriers may be un-cooperative with law-enforcement requests and believe that they alerted suspects to ongoing investiations, as the SIMs stopped being used shortly after the law-enforcement requests for data were made. We suspect, although for operational reasons have been unable to confirm, that the MVNO is simply choosing from existing commercial providers or resellers in each country in order to support their network, but may have a particular profile of provider that they prefer to work with.

Roaming callback

An alternative mechanism is shown in Fig. 6 and Fig. 7.

In this, the call is placed as before, with the HLR controlling call handling. In this case, however, the home network causes the call to

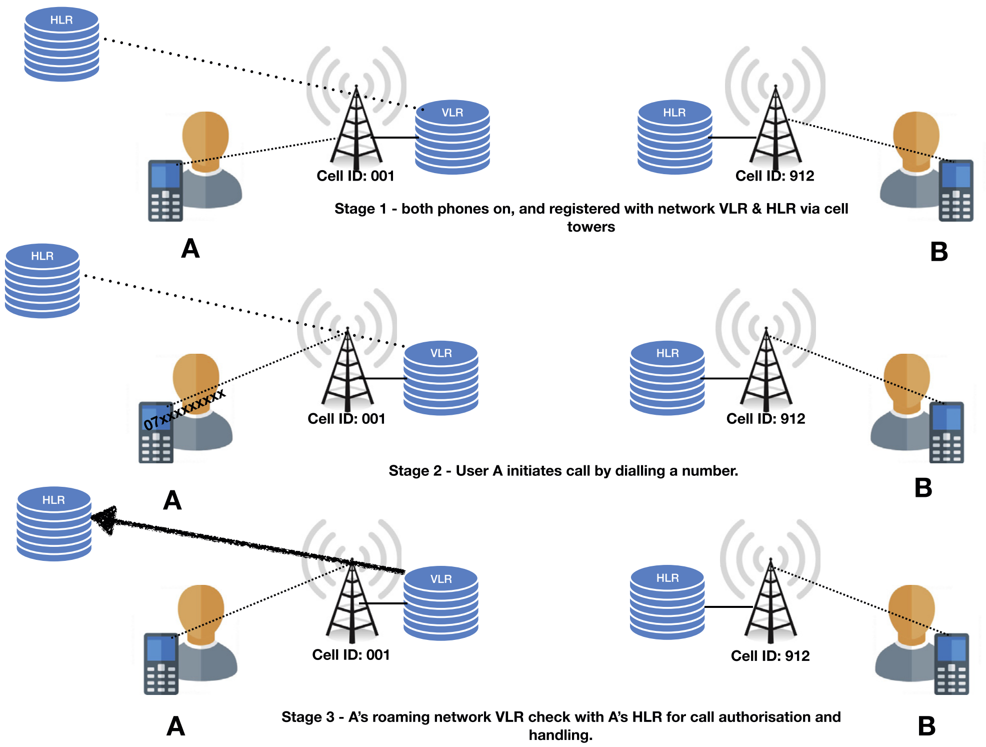

HLR
VLR
HLR
Cell ID: 001
Cell ID: 912
Stage 1 - both phones on, and registered with network VLR & HLR via cell towers
A
B
HLR
VLR
HLR
07xxxxxxxxx
Cell ID: 001
Cell ID: 912
Stage 2 - User A initiates call by dialling a number.
A
B
HLR
VLR
HLR
Cell ID: 001
Cell ID: 912
Stage 3 - A's roaming network VLR check with A's HLR for call authorisation and handling.
A
B

Fig. 4. Initial calling stages.

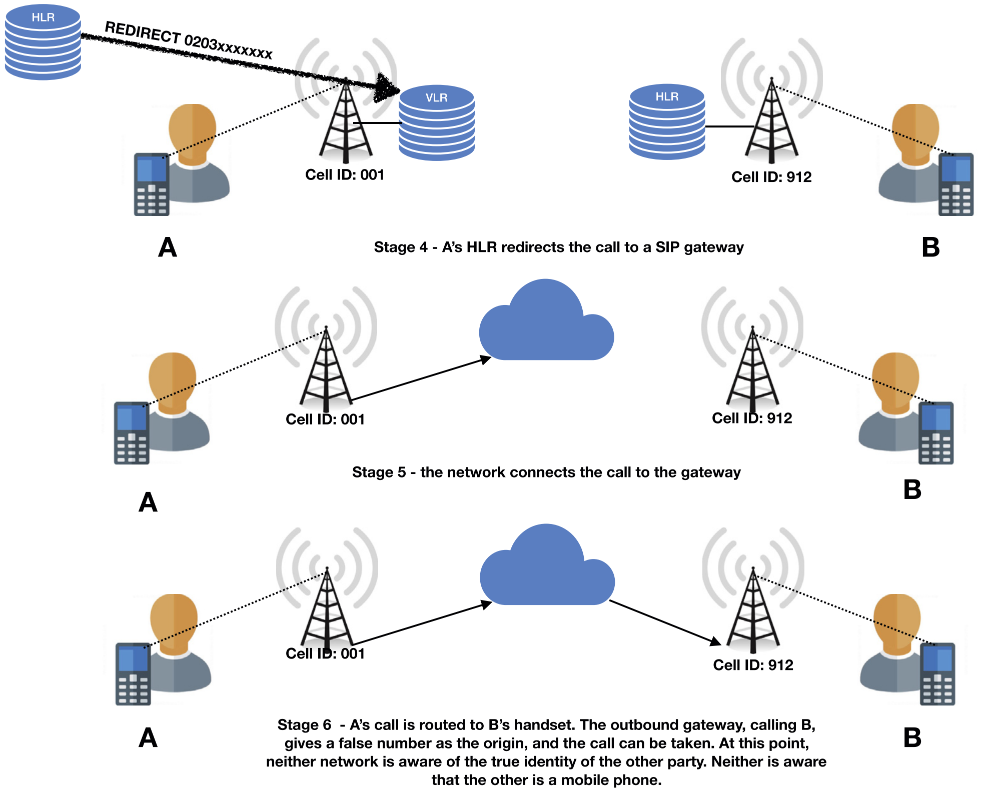

HLR
REDIRECT 0203xxxxxxx
VLR
HLR
Cell ID: 001
Cell ID: 912
A
Stage 4 - A's HLR redirects the call to a SIP gateway
B
Cell ID: 001
Cell ID: 912
Stage 5 - the network connects the call to the gateway
A
B
Cell ID: 001
Cell ID: 912
A
Stage 6 - A's call is routed to B's handset. The outbound gateway, calling B, gives a false number as the origin, and the call can be taken. At this point, neither network is aware of the true identity of the other party. Neither is aware that the other is a mobile phone.
B

Fig. 5. A's home network redirects call via CLID spoofing VOIP gateway.

be dropped before it has been connected. The home network then causes the VOIP system to call the original caller (A) from a different number. The VOIP system then makes a second call to the intended recipient (B). Once both parties have answered they are effectively participating in a conference call.

In both cases, the CLID (from the network system to the

Table 1
Simplified example of call redirect evidence in billing records from one provider. The call shows as 3 events with forwarding and network transit identified. The redirected number is not shown. Not all networks provide this level of detail in the billing data.

ID	Event type	Start date & time of call	End date and time of call	Calling number	Called number
1	Roaming Call Forward	14/04/2018 13:00:13	14/04/2018 13:12:14	07700900001	07700900002
2	Transit	14/04/2018 13:00:13	14/04/2018 13:12:14	07700900001	07700900002
3	Mobile Subscriber Originating	14/04/2018 13:00:13	14/04/2018 13:12:14	07700900001	07700900002

Table 2
Simplified example of CAMEL record confirming the redirect mechanism is operating. The destination number is the landline access number for the VOIP network.

ID	Start date & time of call	End date and time of call	Called number	CAMEL Destination Number
1	14/04/2018 13:00:13	14/04/2018 13:12:14	07700900002	02079460999

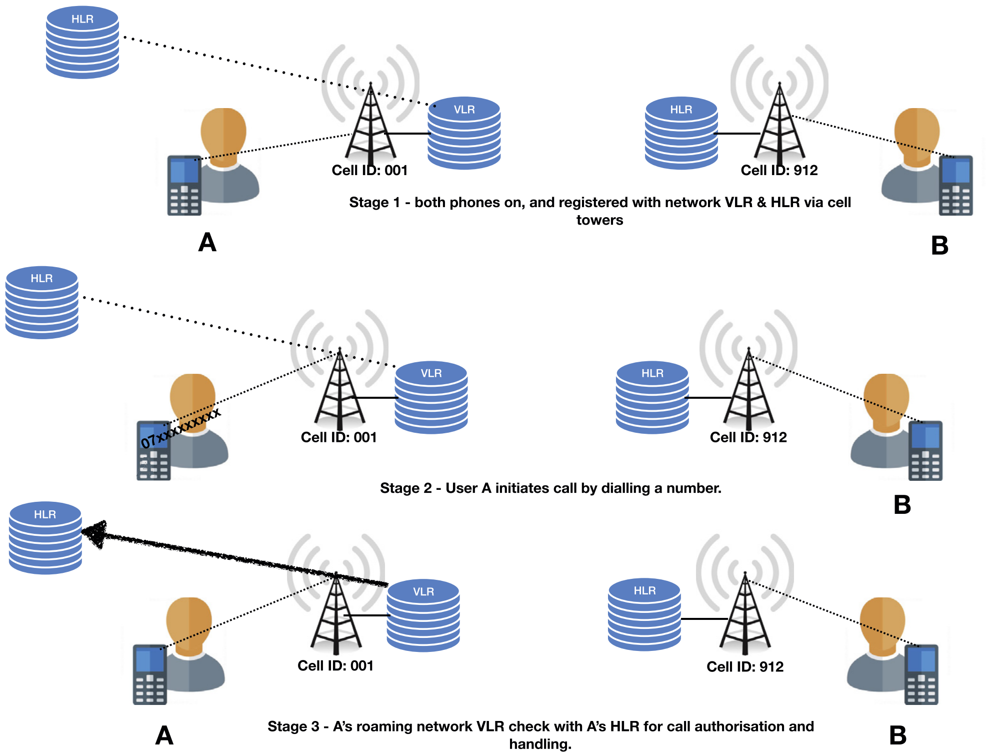

HLR
VLR
Cell ID: 001
HLR
Cell ID: 912
Stage 1 - both phones on, and registered with network VLR & HLR via cell towers
A
B
HLR
VLR
07xxxxxxxxx
Cell ID: 001
HLR
Cell ID: 912
Stage 2 - User A initiates call by dialling a number.
B
HLR
VLR
Cell ID: 001
HLR
Cell ID: 912
A
Stage 3 - A's roaming network VLR check with A's HLR for call authorisation and handling.
B

Fig. 6. Initial calling stages.

handset) can be falsified, with both parties seeing a different calling number which may change with every callback. Again, if the false CLID appears to be a mobile number there will be no IMEI, IMSI or cell data for the apparent caller in the CDRs.

Involvement of "normal" SIM

Where a standard "non-stealth" SIM needs to call a "stealth" SIM, neither the call redirect nor callback process will work as the SIM is not issued by the stealth MVNO. In this case, the stealth SIM user is typically provided with one or more "access" numbers which can be used to contact them, and may be changed regularly. The user of the normal SIM will dial one of these numbers, which will show in their billing data, but the call will still be handled via the SIP/VOIP system and the recipient will still receive a call from a spoofed CLID.

If a stealth SIM calls a "normal" SIM, the redirect or callback mechanism will proceed as described above.

Effect

The overall effect of both mechanisms, except where a stealth SIM is used to call a "normal" SIM, is that the caller has no way of identifying the recipient and vice-versa as both "see" false numbers which are used by various parts of the network to route the call to its real destination. From an investigative perspective, if only one of the handsets is recovered, or call data obtained for only side of the conversation, it is not possible to determine who or where the other party was.

Correlating data

Where billing data for both caller and recipient are available, it is possible to correlate calls by examining the end time and duration in particular (start times may be several seconds apart depending on which of the mechanisms is involved - typically 2 seconds for redirection and up to 15 seconds for callback). Because the SIMs are configured to allow roaming across multiple networks, it may also

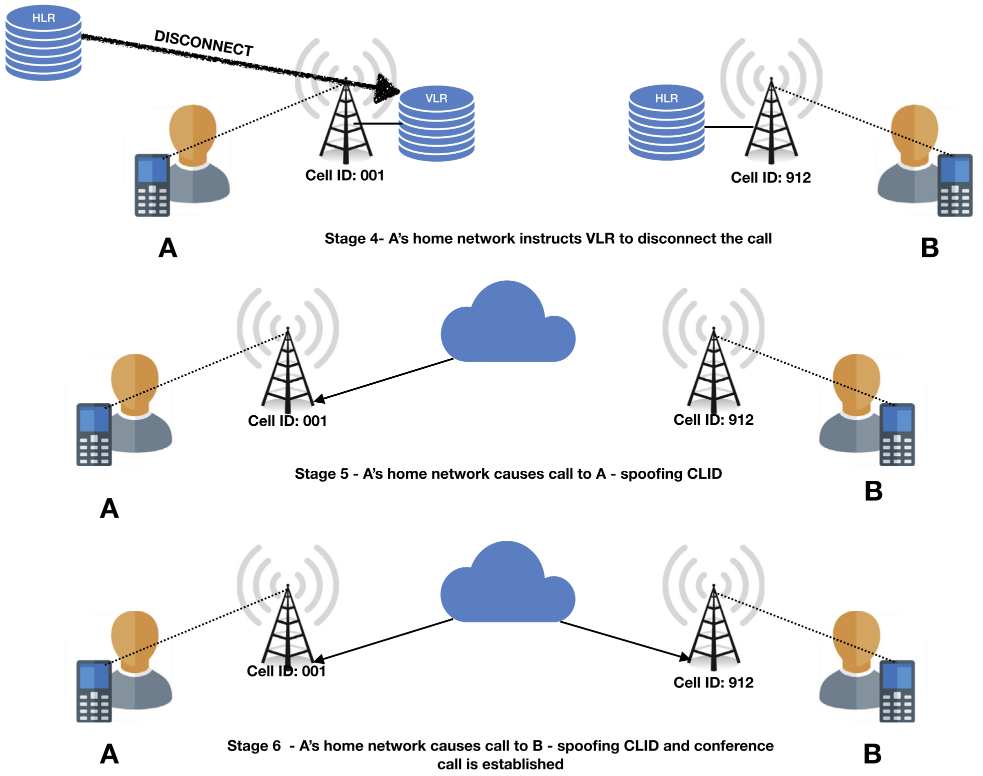

HLR
DISCONNECT
VLR
Cell ID: 001
HLR
Cell ID: 912
A
Stage 4- A's home network instructs VLR to disconnect the call
B
Cell ID: 001
Cell ID: 912
Stage 5 - A's home network causes call to A - spoofing CLID
A
B
Cell ID: 001
Cell ID: 912
A
Stage 6 - A's home network causes call to B - spoofing CLID and conference call is established
B

Fig. 7. Call completion via the callback mechanism.

be necessary to obtain billing data from several network providers in order to establish a complete pattern of activity.

If handset call logs are available, the presence of 0 duration outgoing calls which are immediately followed by incoming calls would suggest the use of a callback mechanism.

In the case of a redirect mechanism, the handset logs and call data show a dialled number, but the logs for that number, if it exists, neither show corresponding incoming calls nor redirection to a voicemail service.

In some cases, the billing data is sufficiently detailed (Table 1) to show that the call has been redirected and, in the early days of data collection, some CAMEL data (See Table 2) showing redirection being requested was also obtained.

Correlation of data - general principles

The billing data held for each account varies depending on the network provider, but the minimum information required to allow for correlation to be completed is:

Start date & time of call	End date and time of call	Type of call	Calling number	Called number

The times may be presented in local time (e.g. BST, CET etc.) or in UTC (Universal Time Co-ordinated, which is effectively the same as GMT). All times should be converted to the same standard prior to correlation. Many billing records include the call duration as a separate record.

The type of call is typically MT (Mobile Terminated = incoming call), MO (Mobile Originated = outgoing call from the subscribed SIM), or PSTN (Public Switched Telephone Network, aka landline originated = incoming call). The calling number is that of the subscribed SIM making the call (in the case of outgoing), or the number presented by the network (for incoming calls). The called number will be the number of the subscribed SIM (for incoming calls), or it will be the number dialled by the handset/SIM (for outgoing calls). Where a "stealth" SIM is in use, the called number recorded may not match the number dialled by the user due to the redirection mechanism employed by the network, but can be the number of the gateway to which the call has been redirected, particularly where the network records transitions as the call is handled. Several different numbers may match the same gateway.

If a callback mechanism is in use, the billing data will not record the original 0 duration outbound call, but will only show the resulting incoming call.

Additional data may be included in the billing record, including a record of call redirection in CAMEL data fields (Table 2, above), call starting and end cell site data etc., but this is dependent on the mobile network provider.

Correlation of data - redirected calls

In order to correlate records which involve a "stealth" call redirection system, therefore, we need to start by checking the end times of calls on the two accounts which are believed to have communicated with each other. Where one handset has placed an outgoing call, and the other has received one, and the end times are the same (to within 1 or 2 seconds) and the call durations are the same (to within 1 or 2 seconds) and there is a pattern of such calls appearing, it is reasonable to infer that the two accounts have used a redirection system to communicate with each other.

If the calling number associated with the incoming calls is obviously false (e.g. 07777777777, 07111111111 or similar “lazy” numbers, an unassigned number, or a number which is clearly in use by an innocent third party), this implies that the number has been “spoofed” for some reason.

Examples of the correlation mechanisms are given below. The number of calls in each example is small and probably insufficient to allow a proper inference of regular communication to be drawn. However, in a real case, a larger number of such co-incident calls would, we believe, be sufficient for such an inference to be drawn and considered reasonable and correct.

Example[2]

ID	Start date & time of call	End date and time of call	Type of call	Calling number	Called number
Account A (07700 900001)					
1	14/04/2018 12:00:13	14/04/2018 12:02:14	MO	07700900001	07700900915
2	14/04/2018 12:32:58	14/04/2018 12:35:15	MO	07700900001	02079460091
3	14/04/2018 12:45:56	14/04/2018 12:54:02	MT	07777777777	07700900001
4	14/04/2018 13:01:33	14/04/2018 13:17:09	MO	07700900001	02079460831
5	14/04/2018 13:31:12	14/04/2018 13:45:11	MO	07700900001	07700900915
Account B (07700 900901)					
1	14/04/2018 12:00:14	14/04/2018 12:02:13	MT	07700900836	07700900901
2	14/04/2018 12:45:57	14/04/2018 12:54:01	MO	07700900901	07700900934
3	14/04/2018 13:31:11	14/04/2018 13:45:12	MT	07777777777	07700900901

In this case, we can see that Account B has received 2 calls (1 and 3) and that these coincide with 2 of Handset A's outgoing calls (1 and 5). Account B also made a call (2) and this coincides with Handset A's incoming call (3). “Lazy” numbers are shown as incoming in call 3 on both accounts, probably indicating that number “spoofing” was in use. The inference is that 07700900915 is a virtual “access number” which B has issued to A in order to allow A to call B without knowing B's real number, and that 07700900934 is similarly A's virtual access number.

If CAMEL data corresponding to these calls could be obtained it would show redirection to a landline which is the entry point to a SIP/VOIP network as described above.

Correlation of data - callback system

In the callback system we are, again, looking for patterns of calls where both accounts terminate incoming calls of similar durations (typically 10–15 seconds different due to the way the callback systems need to connect to one handset and then call the other) at the same time (usually to within 1 or 2 seconds again). The incoming numbers are, again, likely to be “spoofed” and show as unassigned or assigned to an uninvolved third party.

Example

ID	Start date & time of call	End date and time of call	Type of call	Calling number	Called number
Account A (07700 990001)					
1	15/04/2018 19:10:10	15/04/2018 19:31:07	MT	07700900815	07700990001
2	15/04/2018 20:03:21	15/04/2018 20:05:37	MT	07700980836	07700990001
3	15/04/2018 20:07:11	15/04/2018 20:09:01	MT	07700904336	07700990001
Account B (07700 990030)					
1	15/04/2018 20:03:36	15/04/2018 20:05:36	MT	07700905006	07700990030
2	15/04/2018 20:07:24	15/04/2018 20:09:02	MT	07700905320	07700990030

Here we can see that Account B's incoming calls terminate at the same times as Account A's incoming calls 2 and 3. The start times of these calls are quite close to each other, and within the time that we would expect to see in the case of a callback system being employed. The calling numbers are “spoofed".

Conclusion

The methods outlined above have been tested “in the field” and results have been accepted as evidence without challenge. The process is, however, still almost entirely manual and there is a need for further work to be carried out to develop a suitable robust automated process which can deal with the different formats in which CDRs are presented as well as accounting for timezone and daylight savings time changes.

Further work is also required on analysis of the VOIP leg of the route. The authors have noted that, where VOIP call data has been obtained, the clocks used in the VOIP system appear to contain a certain amount of “random” drift which results in the data appearing less reliable than that provided by the mobile networks. Since the randomisation present in the VOIP clock drift creates uncertainty in the call correlation process, work to establish the reliability of any correlation involving this clock is essential before the VOIP data is relied on to produce evidence.

We are also very conscious that no statistical analysis has been carried out. We would welcome suggestions for how controlled experiments to produce suitable data can be carried out, given that it either requires the co-operation of untrusted MVNOs and SIP/VOIP providers or the creation of private MVNOs and SIP/VOIP networks based on the hypothesis presented in this paper.

Annex - glossary and abbreviations

CAMEL *Customised Applications for Mobile networks Enhanced Logic.* This is a standard which allows mobile network operators to specify and operate additional features & functions on their network. Typically, it is used to allow the handset's home network to control roaming calls.

[2] Note - in all of the examples shown, we have used only numbers which OfCOM recommends for use in TV and radio productions. These are non-allocated numbers and there is a limited range of them. In real cases, there is likely to be greater variation in the prefixes, particular for incoming numbers, than we have been able to show.

CAMEL features allow control of call routing and provision of additional services by the network.

CLID *Calling Line IDentity.* The phone number disclosed as the calling number to a handset which is receiving a call. This is not necessarily real phone number of the calling party as it can be changed by using appropriate network technology (e.g. through the use of private switchboard systems).

GSM *Global System for Mobile communications.* The first generation mobile phone network using digital technology.

GPRS *Generalised Packet Radio Service.* An addition to GSM, often known as 2G networking, which adds functions to carry computer data alongside audio, as well as provide an alternative higher speed channel for SMS.

HLR *Home Location Register.* A database of active IMSIs which are authorised to use the network. Each mobile network operator maintains their own HLR for their issued IMSIs. The HLR also contains the phone numbers which are associated with each IMSI.

ICCID *Integrated Circuit Card IDentifier.* A unique number, found in SmartCards (of which a SIM is one example) which identifies the IC application type (e.g. 89 is the code for telecommunications cards), the country of origin, the issuer identification and a unique account number.

IMEI *International Mobile Equipment Identifier.* A unique number which is used to identify the handset to the network. This can be used to identify the type of handset being used and its capabilities so that configuration can be sent to it, when required. It can also be checked against databases of stolen or lost handsets to enable networks to block access.

IMSI *International Mobile Subscriber Identity.* A unique number which the SIM uses to identify itself to the network. The network uses this to establish the SIM's authority to use the network, associate it with the subscriber and the callable number associated with the SIM. The IMSI includes components which identify the home country and network for the SIM.

MO (in call data logs) - Mobile Originated call. i.e. the call was made FROM the handset & SIM for which the data has been retained

MT (in call data records) - Mobile Terminated call. i.e. the call was made TO the handset & SIM for which the data has been retained.

PBX *Private Branch Exchange.* A private exchange or switchboard, such as those used by large companies. PBXs can carry out call routing operations and can also be programmed to present a different CLID to the network for calls which are routed through the PBX.

SIM *Subscriber Identity Module* A type of smart card used to allow mobile phones to operate on GSM/UMTS networks. A SIM may contain several IMSIs in order to allow it to access the most appropriate network when roaming.

Smart Card a card, typically credit card or smaller in size, containing an Integrated Circuit (IC) or “chip” which provides specific processing or data functions on demand.

SMS Short message service - the official name for text messaging. This refers to the use of the mobile phone network “control” or GPRS channel to send short text messages.

UMTS *Universal Mobile Telecommunications System.* The 3rd generation mobile phone network technology, based on GSM.

USSD *Unstructured Supplementary Service Data.* A data transmission and receipt service offered by GSM networks. Normally it is used to communicate with the network to retrieve account data or set particular options. USSD messages handled by the home network normally start with an asterisk (*) or hash (#) and are terminated by a hash (#).

VLR *Visitor Location Register.* Similar to HLR, this is a register of IMSIs which have been authorised, via their HLRs, to roam onto this network.

VOIP *Voice Over Internet Protocol.* A mechanism which allows telephone calls to be carried over the Internet. This is typically used to reduce call costs.

References

ETSI, 1996. Digital Cellular Telecommunications System; Interworking between the Public Land Mobile Network (PLMN) and the Packet Switched Public Data Network (PSPDN) for Packet Assembly/Disassembly (PAD) Facility Access available online at: https://www.etsi.org/deliver/etsi_gts/09/0905/05.00.00_60/gsmts_0905v050000p.pdf. (Accessed 20 November 2018).

ETSI, 2014. Digital Cellular Telecommunications System (Phase 2+); Universal Mobile Telecommunications System (UMTS); Customised Applications for Mobile Network Enhanced Logic (CAMEL) Phase X; CAMEL Application Part (CAP) Specification available online at: https://www.etsi.org/deliver/etsi_ts/129000_129099/129078/12.00.00_60/ts_129078v120000p.pdf. (Accessed 20 November 2018).

European Telecommunications Standards Institute (ETSI), 1996. UnStructured Supplementary Data Services available online at: https://www.etsi.org/deliver/etsi_gts/03/0390/05.00.00_60/gsmts_0390v050000p.pdf last. (Accessed 20 November 2018).

Ghadialy, Zahid, 2004-2019. CAMEL: an Introduction available online at: https://www.3g4g.co.uk/Tutorial/ZG/zg_camel.html. (Accessed 12 March 2019).

Appendix J

Whatsapp media persistence

Digital Investigation 25 (2018) 114–115

Contents lists available at ScienceDirect

Digital Investigation

journal homepage: www.elsevier.com/locate/diin

Case Note

WhatsApp server-side media persistence

Angus M. Marshall

University of York, UK

Case context

In early 2015, the author was approached to assist with the investigation of an alleged case of possession and potential distribution of images of child abuse. The investigating officer had received a report that a worker had been showing video clips to residents of a care home, using a mobile phone. Two complainants had given evidence that they had been shown the video clip in question and that they believed that it was being shown via the WhatsApp messaging app.

The officer had seized the mobile phone and it had been processed by his organisation's digital forensics unit. They had successfully extracted data from the handset, including the WhatsApp live database and 5 backups using conventional commercial tools, and had decrypted the backups.

Examination of the WhatsApp data files around the date & time in question produced no useful results. There were no cached video clips which matched the description and no preview frames relating to the type of video described were found.

Investigation goal/technical challenge

The officer remained convinced that the complainants were telling the truth and requested external assistance with the investigation. The author was contacted and asked to carry out an independent examination. A copy of the data downloaded from the handset, with all associated data from the standard tool, was made available.

Digital evidence involved

Although the cached media contained no files which matched those described by the complainants, there was some indication that the handset had received a small amount of other "adult" and potentially offensive/illegal material via WhatsApp. The original laboratory processing had recovered the WhatsApp encrypted backup files and the live WhatsApp database.

Processes and/or tools used

WhatsApp version 2.11.399 was present on the handset. This version uses a common key to encrypt all backups for all users (Ibrahim, 2014). Having decrypted the backups using Ibrahim's method (Ibrahim, 2014), the author extracted all data from the databases using Sangiacomo & Weidner's WhatsApp Xtract (WAX) tool version 2.1 (Sangiacomo and Weidner, 2012).

None of the images present in the database matched the description given by the officer, but the author noted that there were some thumbnails which might indicate that obscene material had been both received and sent by the application. It was also noted that there were URLs present which appeared to be links to video files (i.e. ending in. mp4,.mpg and. avi extensions).

Simple tests were run, in accordance with the author's own recommendations (Marshall, 2003), to determine if these URLs would allow direct download of the media files and their associated meta-data. It was found that the URLs did allow direct download of the files with no obvious intervention of proxies or redirects and no apparent requirement for authentication.

Therefore, the resulting message data were processed, using a simple bash script with wget, to retrieve any and all media files still present at the URL's contained in the database.

In order to reduce the number of media files to be visually assessed, a further script was run to determine which messages in the backup files were no longer present in the live database, and also which pre-dated the time of the activity which was the subject of the complaints.

Result

At the time of the original investigation, the video files were successfully downloaded several months after the associated messages/chat session had been deleted from the user's WhatsApp database and were found to be accessible, in an unencrypted format, via a publicly accessible URL.

The lifespan of the files on the server appeared to be considerably longer than normal server-side "housekeeping" practices would allow. The author's hypothesis, therefore, is that messages of this type are likely to be forwarded from user to user, circulating amongst a group of users with similar interest. As a result, each time the file is forwarded, the server may reset the clock which is

E-mail address: angus@n-gate.net.

https://doi.org/10.1016/j.diin.2018.04.005

used to determine which files can be deleted during normal garbage collection activity.

This is further supported by the fact that other non-offensive and legal media files, mainly of a humorous or political nature, were also retrieved during the processing stage.

Lessons learned

The actual process took several iterations as the author started by relying on the cached data and thumbnail images. The video in question was represented by a thumbnail within the backup databases but, because the thumbnail image was of a completely black frame, it was overlooked.

Following this investigation, the author changed his method from reliance on thumbnails to carrying out a proper full media retrieval using the simple wget script and reviewing the actual video files. Although this resulted in longer processing times, it produced more complete results.

The experience serves as a good reminder that thumbnail still images selected from video files may not be the most useful representation of the content of the video for an investigator to review.

Future work

More recent tests (November 2017) carried out using Google's Chrome web browser with Developer extensions, and WhatsApp's Web interface to their system, allowed the complete URLs for media files to be observed as part of the normal network traffic. At this time, WhatsApp appeared to be operating a similar scheme, where video (and other large) files were held on their servers, although the original servers found in the case appear to have been taken offline. Again, it was found that there was no obvious need for authentication when making the HTTP GET request for the file, but that it was necessary to set a known web browser User-Agent in the wget fetch in order to obtain the files.

The recent files were, however, delivered in an encrypted form. The exact nature of the encryption key was not known, but simple tests suggest that the same persistence of data on the server occurs, albeit with encrypted files. This was verified by sending a test message and then deleting it at both sender and receiver. Several days after deletion, the media file was retrievable.

This is an area which the author is currently exploring to determine if the persistence behaviour is still present and whether the media files are encrypted once only (at time of first upload) or re-encrypted for each new user who receives them. It is proposed that this study will be extended to other common messaging and social media apps. In an attempt to determine which factors affect server-side media persistence and how accurate are vendors' claims about when media files can be guaranteed to have been removed from their servers.

References

Ibrahim, M., 16th Feb. 2014. How to Decrypt WhatsApp Database Messages. https://stackpointer.io/security/decrypt-whatsapp-database-messages/261/. (Last accessed 15 March 2017).

Marshall, A.M., 2003. An improved protocol for the examination of rogue WWW sites. Sci. Justice 43 (4), 237–248. Forensic Science Society, Harrogate, UK.

Sangiacomo, F., Weidner, M., 5th April 2012. WhatsApp Xtract. https://forum.xda-developers.com/showthread.php?p=24603294. (Last accessed 15 March 2017).

Appendix K

The author's other published works

Note: this excludes quarterly "IRQ" column in Digital Forensics Magazine from 2010 to date.

K.1 Books and book chapters

- Marshall, AM “Digital Forensics”, Wiley, 2008.
- Marshall, AM & Stephens, P “Identity and Identity Theft” in Bryant, R (Ed.), “Investigating Digital Crime” , Wiley, May 2008.
- Bryant, RP & Marshall, AM “Criminological and Motivational Perspectives” in Bryant, R (Ed.), “Investigating Digital Crime” , Wiley, May 2008.

K.2 Refereed journal articles, excluding those included in appendices

- Marshall, AM “Standards, Regulation & Quality in digital investigations : the state we are in”, Digital Investigation, Elsevier 2011
- Marshall, AM “Quality standards and regulation : challenges for digital forensics” Measurement & Control, (68) Institute of Measurement & Control, London, 2010.
- Marshall, AM & Clarkson, A “Future Crime & Detection in Cyberspace” Measurement & Control, (66, 248-251) Institute of Measurement & Control, London, 2008.
- Gwynne B & Marshall AM “The application of ’Crime Lites’ to examination of computer components ’in situ”’, Science and Justice, 2005.
- Marshall, AM “Digital Evidence”, Measurement & Control, (38, pp79-82), Institute of Measurement & Control, London, 2005.
- Marshall, AM & Tompsett, BC “Identity theft in an online world”, Computer Law & Security Report, (21, pp128-137), Elsevier, 2005.
- Marshall, AM “An improved protocol for the investigation of rogue WWW sites” , Science & Justice (43, pp237-248), Forensic Science Society, Harrogate, 2003.

K.3 Papers in Conference Proceedings, excluding those included in appendices

- Moor, GN & Marshall, AM "Corporate ID Theft : an examination of means and opportunities", Proc. ECCE 2006
- Zeus-Brown, AM & Marshall, AM "Street crime detection and monitoring methods : can these be transferred to the internet in order to give a set of guide lines for cyber crime investigation?" Proc. ECCE 2006
- Marshall, AM & Tompsett, BC "Silicon Pathology - the future of forensic computing ?" , Science & Justice, (44, pp43-50), Forensic Science Society, Harrogate, 2004.
- Marshall, AM & Tompsett, BC "Spam 'n' chips – a discussion of internet crime", Science & Justice (42, pp 117-122), Forensic Science Society, Harrogate, 2002
- Boldyreff C, Gaskell C, Marshall AM & Warren PJ "Establishing a Measurement Programme for the World Wide Web", Proc. 2001 Symposium on Applications and the Internet (SAINT-2001), IEEE Computer Press, January 2001.
- Boldyreff C, Gaskell C, Marshall AM & Warren PJ "Web-SEM Project: Establishing Effective Web Site Evaluation Metrics" Proc. 2nd International Workshop on Web Site Evolution WSE'2000, Zurich. pp WSE17-WSE20, online at http://www.cs.ucr.edu/ stilley/wse2000
- Marshall AM, Ellison D, Samson WB & Swanston MT, 'A Technique for Adding Global Generalisation to CMAC' in 'Neural Computing - Research & Applications III' Keating JG (Ed.), St. Patrick's College, Maynooth, Co. Kildare, Ireland 1995

K.4 Refereed Conference Presentations and Workshops

- Marshall, AM & Tompsett, BC "Further results from an observational study of crime an online auction site", presented at British Society of

Criminology Conference, Huddersfield, July 2008

- Zeus-Brown, AM & Marshall, AM "Social Engineering in a digital environment", presented at British Society of Criminology Conference, Huddersfield, July 2008

- Chlapoutakis, G & Marshall, AM "The development of a modular software framework for a distributed attack profiling network", presented at British Society of Criminology Conference, Huddersfield, July 2008

- Chlapoutakis, G, Marshall, AM, Tompsett, BC, & Zeus-Brown, AM "To catch a Cyberthief : Legal and Technical Issues Affecting Network Monitoring", presented at BILETA 2008, Glasgow, April 2008

- Marshall, AM, Tompsett, BC & Semmens, NC "Towards an automated online detection and profiling system", presented at British Society of Criminology Conference 2007, London, Sept. 2007

- Marshall, AM, Clarkstone, S & Tompsett, BC "Results from an e-bay observational study", presented at British Society of Criminology Conference 2007, London, Sept. 2007

- Zeus-Brown, AM, Marshall, AM & Tompsett, BC "Remote victim support in the digital world : profiling of computer systems and attacks" presented at British Society of Criminology Conference, London, 2007

- Marshall, AM, Moor, GN & Tompsett, BC, "Criminalising the Internet", presented at EAFS2006, Helsinki, June 2006

- Clarkson, A & Marshall, AM "Digital Evidence Reporting and Presentation : A Transatlantic Perspective", Forensic Science Society, 9th July 2006, Leeds.

- Marshall AM "Hackers, crackers, spammer and scammers", Workshop on internet tracing, Forensic Science Society spring meeting, Derby Apr. 23rd 2004.

- Marshall AM "Identity Theft Online" (presented at BAHID meeting Nov. 2003, Sheffield.)

- Marshall AM, Ellison D, Samson WB & Swanston MT, 'Adapting CMAC for improved Adaptive Control'. Presented at NACT-I workshop, Glasgow 1995

K.5 Invited Conference Presentations and Workshops

- Marshall, AM "The regulation gap: validation, verification and the reality of digital forensic standards", Digital Justice: Ethical, Policy and Practice Issues, Keele University 9th January 2019.

- Marshall, AM "Digital Forensics: Are we ready for it?", SUAC 2016 conference, College of Policing, Ryton, 3rd November 2016

- Marshall, AM "Maximising evidential potential of digital devices" The Investigator : Computer & Mobile Phone Conference, Wyboston, 17th March 2010.

- Marshall, AM "Software Validation : challenges and Options" Forensic Science Regulator's Conference, Birmingham, 9th February 2010.

- Marshall, AM "Emerging future technical challenges facing investigators" The Investigator : Computer & Mobile Phone Conference, Towcester, 29th September 2009.

- Marshall, AM "Digital Devices, Traces and Regulation", Forensic Science Society spring meeting, Nottingham 25th April 2009.

- Marshall, AM "Reliable Forensics?" Forensic Science Regulator's Conference, Birmingham, 31st March 2009

- Marshall, AM "Wireless Devices : opportunities and threats at crime scenes" , workshop at CSI Conference, University of Teesside, April 2008

- Marshall, AM "Moore's Law: Miniaturisation vs. Identification" , presented at CSI conference, University of Teesside, April 2008.

- Marshall, AM & Tompsett, BC, "Working with external partners. The Forensic Science Perspective", presentation at HEA workshop, University of Northumbria, 18th November 2005.

- Marshall, AM & Tompsett, BC "Digital Evidence Evaluation", workshop at FIRN/HEA Forensic Science Conference, 7th July 2005.

K.6 Refereed Conference Posters

- Tompsett, BC, Marshall, AM & Semmens, NC, "Cybercrime Terminology", research poster presented at EAFS2006, Helsinki, June 2006. (shorlisted for poster prize)

- Marshall, AM, "Steganographic Opportunities in Modern Digital Files", research poster presented at EAFS2006, Helsinki, 2006 (shortlisted for poster prize)

- Marshall, AM & Zeus-Brown AM, "Remote Covert Investigation of FTP for Forensic Purposes". Research poster presented at EAFS2006, Helsinki, June 2006. (shortlisted for poster prize)

- Tompsett BC, Marshall AM & Semmens NC "Cyberprofiling", research poster at American Society of Criminology Conference, Toronto, November 2005.

- Tompsett BC, Marshall AM & Semmens NC "Cyberprofiling : Offender Profiling and Geographic Profiling of Crime on the Internet", presented at Computer Network Forensics Workshop 2005 (IEEE/CreateNet SecureComm Athens 2005). Published in proceedings

- Tompsett BC, Semmens NC & Marshall AM "Digital Recidivism – cyber-criminal career development", presented at British Society of Criminology Conference, Leeds June 2005

- Watson, K & Marshall, AM "Identity_theft.con" (poster presented at Forensic Science Society Summer meeting, June 2003) (runner-up for "best poster" prize)

- Marshall AM, & Ellison, DE, 'A CMAC Based Broom-Balancer on Transputers' (Poster) WOTUG-15,1992

K.7 Other Presentations, Articles and research-based output

- Marshall, AM - oral evidence to House of Lords Science & Technology Committee inquiry into Forensic Science 27th November 2018 (viewable at https://parliamentlive.tv/Event/Index/6afc9b5a-9c59-4c5f-af84-16fff5ddb26b)
- Marshall, AM - written evidence to House of Lords Science & Technology Committee inquiry into Forensic Science (available at http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-committee-lords/forensic-science/written/89341.pdf), October 2018.
- BBC World service "Click" programme. Interview about the "Snowden" Haven app. 2018
- BBC News 24. Interviews (2) about NHS "Wannacry" outbreak. 2017/2018
- National Trading Standards e-crime guide, 2014
- Guest lectures/seminars on quality standards at College of Policing, 2014
- Marshall, AM, Higham, S & Dyhouse, T "Digital Forensics Capability Review", Electronics, Sensors & Photonics KTN for Forensic Science Special Interest Group, 2014
- Marshall, AM, Home Office CAST research reports on digital forensic processes, 2012 (Internal reports, restricted circulation)
- Digital Forensics Magazine "IRQ" quarterly column/page on current issues in digital forensics. 2010-date
- BBC Radio Scotland "Scenes of Crime" (radio series) – major contributor to episode 3 : Digital Crime (invited by the host -Val McDermid). May 2009.
- Acklam Library (Middlesbrough) "Murder, myths and mayhem" - crime writers and forensic scientists presentation & panel session. October 2008

- Marshall, AM “Digital devices at crime scenes”, Durham Constabulary CSI conference / briefing, April 2008

- Marshall, AM “Crime on the Internet”, invited session at Birkbeck Forensic Science Dayschool, Birkbeck, University of London, 26th November 2005.

- Marshall, AM “Watch out for wireless”, Police Professional, November 2005.

- Berrett, R, Marshall, AM, Marshall, S & Sutherland, C “Where did they go wrong?” Forensic Science Panel – Harrogate Crime Writing Festival, July 2005.

- Marshall, AM “Identity Crises on the Cards”, Yorkshire Post, 24th June 2005.

- Marshall, A “Conning Nets” (book review), Science & Justice (45, pp55-56), Forensic Science Society, 2005

- Marshall, AM “More Spooks than CSI” invited paper – presented at the Harrogate Crime Writing Festival, Jul. 24th 2004

- Marshall, AM “Lies, damned lies and deception in cyberspace”. Yorkshire Post, 26 April 2004.

- Marshall AM, “Digital Evidence” - seminar at University of Derby, Feb. 19th 2004

- Marshall AM, “Online Identity” - seminar at Centre for Internet Computing, February 18th 2004

- Marshall AM, “Forensic Computing – an introduction”, seminar at University of Abertay Dundee, Nov. 2003.

- Marshall, AM, Editorials on Year 2000 problem, Hull Daily Mail, October & November 1998.

- Marshall, AM “Reader’s Response” (Letter), International Journal of Forensic Computing (4), Computer Forensic Services Ltd., April 1997

- Marshall AM, 'Usenet and NetNews', Computer Bulletin (OUP) April 1995
- Marshall AM, 'You Little Minix!' (Review of the Minix Operating System) Amiga Shopper, January 1993

K.8 Public Commercial Training Material Developed

- Marshall, AM & Manson, P "Forensic Readiness & Response" 4-day training course developed for Office for National Statistics. n-gate ltd., March 2013.
- Marshall, AM "Computer evidence for Trading Standards Officers" , University of Teesside, 2008
- Marshall, AM "Internet 101" - basic Introduction to WWW browsing and searching, e-mail, Usenet News, University of Hull, March 2000
- Marshall, AM "PostgreSQL for the Web", May 2000
- Marshall, AM "Introduction to PHP", May 2000
- Marshall, AM "Introduction to HTML 3.2", October 1999
- Marshall, AM "Year 2000 Compliance in PC Applications." atecc, University of Lincolnshire & Humberside, March 1999
- Marshall, AM "Surviving Millennium Meltdown - a self-assessment workshop". Year 2000 audit workshop for SMEs. Humberside Y2k/University of Lincolnshire & Humberside, November 1998
- Marshall, AM "UK 'Y2k BugBuster' supplement",Learning Tree International Ltd., July 1998
- Marshall, AM "Intranet & WWW authoring", University of Lincolnshire & Humberside, Jan. 1998
- Marshall, AM "Introduction to C for programmers" 1996
- Marshall, AM "C programming for beginners" 1995

K.9 Membership of editorial boards, conference & publication committees and review panels

- 2020 to date - Associate Editor, FSI: Digital Investigation (Elsevier)

- 2018 - Member of review panel for IEEE Transactions on Information Forensics and Security

- 2016 - Member of review panel for International Journal of Digital Forensics & Law.

- 2013 - Member of editorial board of Digital Investigation journal (Elesevier) (now FSI: Digital Investigation)

- 2007 to date – Reviewer for Journal of Digital Forensic Practice (invited by editor : M. Rogers)

- 2007 - Committee member, Advances in Computer Security & Forensics Conference (Liverpool)

- 2006 – Chairman/convener, e-crime and Computer Evidence (ECCE) 2006 conference (Nottingham)

- 2006 to 2015 – Reviewer for Science & Justice (Forensic Science Society/Elsevier)

- 2006 – Committee member, Advances in Computer Security & Forensics Conference (Liverpool)

- 2006 to date – Reviewer for IEEE Transactions on Information Security & Forensics

- 2005 – Chairman/convener, e-crime and Computer Evidence (ECCE) 2005 conference (Monaco)

K.10 Grant applications and awards

- 2021-22 - assisted in preparation of Horizon Europe bids “OCTANE” and “RITHMS”. “RITHMS” bid successful in the Fight against trafficking in cultural goods, kick-off in October 2022. Project lead is L'Istituto Italiano di Tecnologia (IIT). I have been asked to participate as lead on AI regulation and standards for evidence in criminal investigations/prosecution.

- 2017 - led bid for EPSRC Digital Economy network “CIDRA - Cyber Investigation, Data Retention and Authentication” to explore issues common to cyber-investigations and archiving. Partners : Open Univ., Bodleian Library (Oxford Univ.), Univ. Hull, Bristol Univ. Value: £1.3m. Unsuccessful as panel didn't consider it fitted call well. Currently exploring ways to rework and submit via other routes.

- 2015 - led DEVCE bid for OSCT/SBRI digital forensics funding. £25k (approx.) awarded for technology demonstrator project in partnership with HARGS Ltd. Successfully completed and invited to apply for stage 2 funding.

- 2013 - member of successful project consortium for ISEC “ECENTRE” project.

- 2008 – University of Teesside Research Fund – awarded £2500 for joint work on “Reliable Forensics” with Dr. P. Brookes of School of Computing.

- 2007 - “Scambusters” awarded £3000 for projects relevant to NE Trading Standards Scambusters initiative.

- 2006 – EPSRC - Applications for Cyberprofiling project follow-on projects (Rapid Evaluation of Electronic & Digital Evidence (REEDE) and Applicability & Interpretation of Cyberprofiling in Real Environments (AICRE)) outline proposals : approx £1m each. Unsuccessful, group elected not to resubmit. Digital Evidence Research Network (9 Universities, 7 external partners, 3 year project, £130k). Unsusccesful, group elected not to resubmit.

- 2005 – University of Teesside Research Fund – awarded £5000 to investigate the use of laser scanning for evidence recording and reconstruction (joint application with colleagues in Crime Scene Science)

- 2005 – University of Teesside Research Fund – awarded £2500 for equipment to investigate neural network based steganography detection.

- 2005 – Peter Berg Fund (University of Teesside) £1000 awarded for creation and support of Forensic Research Incubator Journal – to encourage and support new researchers

- 2004 – EPSRC "Think Crime" programme : "Cyberprofiling" (with N.C. Semmens (Sheffield) and B.C. Tompsett (Hull)) £170k grant awarded, March 2005 (1-year project in association with Computer Associates, Humberside Police, North Yorkshire Trading Standards, Information Commissioner & C. Spencer Ltd.)

- Oct. 2002 - Oct. 2004 – TCS project "A web-based document management system for civil engineering" with C. Spencer Ltd. Graded "2" by KTP management (Momenta). (two follow-on projects agreed and awarded.) Project graded "2" by independent evaluation panel. Invited to submit report to awards panel for consideration for 2005 awards (company elected not to submit).