101
Digital evidence

by

Dr Bradley Schatz BSc Computer Science, PhD Digital Forensics

Last reviewed October 2017

AUTHOR INFORMATION

[101.10] Author

Dr Bradley Schatz

Dr Bradley Schatz is the director of the independent digital forensics consultancy Schatz Forensic, and an adjunct Associate Professor at the Queensland University of Technology (QUT). He gained his Bachelor of Science degree in Computer Science at the University of Queensland in 1995, and subsequently practised software development, systems administration and computer security in the private sector. He completed a Doctor of Philosophy degree in Digital forensics at the Queensland University of Technology in 2007, and has practised digital forensics in private practice since. In 2008, Dr Schatz was appointed Adjunct Associate Professor at QUT, where he occasionally lectures and supervises doctoral students. He is a member of the Editorial Board of the journal "Digital Investigation: the International Journal of Digital Forensics"; the longest standing peer reviewed journal in the field. Dr Schatz additionally sits on the review panels of three international peer reviewed conferences related to digital forensics. He is the author or co-author of 13 peer reviewed academic papers and two book chapters all in the area of digital forensics.

The author would like to extend his gratitude to his reviewers, especially Eoghan Casey, whose feedback was invaluable.

Contact details

Dr Bradley Schatz

Director Schatz Forensic Pty Ltd

Email: bradley@schatzforensic.com

https://www.schatzforensic.com.au

[The next text page is 101-101]

CONTENTS

AUTHOR INFORMATION

GLOSSARY

SCOPE

INTRODUCTION

FOUNDATIONS OF DIGITAL EVIDENCE

PERSPECTIVES ON DIGITAL INFORMATION

DIGITAL FORENSIC PRACTICES

EXAMINATION

ANALYSIS

DIGITAL EVIDENCE SUB-DISCIPLINES

CHALLENGES TO DIGITAL EVIDENCE

VALIDATION, ERRORS AND RELIABILITY

ASSESSING CLAIMED PROFESSIONAL COMPETENCY

FUTURE

BIBLIOGRAPHY

[The next text page is 101-201]

GLOSSARY

Glossary

The following glossary provides definitions of relevance to this chapter. Some definitions are based on the more comprehensive SWGDE[1] and Sedona[2] glossaries.

Acquisition: The process of creating an ideally complete and perfect fidelity copy of the content of digitally stored or communicated information.

Algorithm: A detailed procedure for solving a problem.

Application: Software that enables the operator of a computer to enter, store, view, modify or extract information. Often used as a synonym for software or program. For example, common applications include spread sheets, email, word processors, and web browsers.

Allocated space: The portion of a File System currently in use for storing information and metadata.

ASCII (American Standard Code for Information Interchange): A format for representing English language text within computer systems.

Backup media: A type of removable storage media which stores information on magnetic tape. Backup tapes and other media are typically employed to hold a full copy of all of the data which a business considers to be required to resume business should a "disaster" occur.

Binary: A base 2 numbering system used by computers to represent numbers.

Bit: Smallest unit of information representation representing two states. The smallest unit of digital information.

Bit sequence: A contiguous sequence of bits of arbitrary length. A bit sequence is independent of the storage media in which it is transmitted or stored.

Browser: See web browser.

Byte: A sequence of 8 bits. A byte is typically enough information to store a single English language letter, number or punctuation mark, using the ASCII text encoding.

Collection: Preservation of digitally stored information via seizure of actual storage media.

Compact disk (CD): Storage media which stores information optically, originally designed for the digital storage and playback of music. Used in computing to store arbitrary digital information.

Computer forensics: See digital forensics.

Cryptographic hash: A compact and unique identifier (hash value) created by applying a complex mathematical algorithm to a sequence of bits, whether they be the content of a file, or all of the bits on a hard drive. Typically presented as a sequence of letters and numbers, hash values effectively identify content of digital information. The slightest change in such information, whether it be the introduction of a single punctuation character or the subtle retouching of a digital photograph, will result in a distinctly different hash value.

Dark web: Websites found in a restricted area of the internet, typically accessed using the Tor anonymity preserving protocol.

Database: A storage technology suited for efficiently storing, accessing and updating digital records. Databases are typically stored using a small number of files and are suited for storing records such as sales transactions and accounting data.

De-duplicated storage: A means of reducing the storage requirements of data by sharing common sub-portions of that data.

Digital evidence: Digitally stored or transmitted information which may be of probative value.

Digital forensics: The scientific examination, analysis and/or evaluation of computer evidence in legal matters.

Digital Versatile Disk (DVD): Storage media which stores information optically, originally designed for the digital storage and playback of video. Used in computing to store arbitrary digitally encoded data.

Domain name: A human understandable name (for example google.com) used to identify networked computers. Domain names are translated into computer understandable IP addresses in much the same way that humans use a telephone directory to identify the telephone number of a business or person.

Electronic evidence: See electronically stored information.

Electronically stored information (ESI): Electronically stored or transmitted information of potential probative value. Refers to the information content independent of the media in which it is stored.

Email: A document created or sent using an electronic mail system.

Email server: An application which forwards, stores, and delivers emails. Also used to refer to a computer on which an email server application is running.

Encoding: To change or translate the format of information into another form.

External metadata: Metadata stored in a location apart from the information it refers to. File metadata is typically stored in a separate location to the content of a file.

File: The fundamental unit for storage of information provided by operating systems, which associates a name and folder with stored bit sequence content. A file typically corresponds to a document. For example, documents such as Adobe PDFs, Microsoft Word documents and JPEG images are often stored as the content within a file. In much the same way that a book may contain chapters, a single file may contain multiple documents.

File share: A central file storage area accessible to multiple computers. Within business environments, information is typically stored in centralised shared “drives” which are accessible from every computer within the business.

File slack: A portion of the allocated space of a file system where remnants of the content of a prior file may remain.

File system: Somewhat akin to a filing cabinet in the physical world, a file system is generally the method of dividing up available storage space within a fixed capacity area of storage, providing a method of organising folders and files (and related information). Storage devices such as USB drives typically contain a single file system on which a user may store and organise multiple files and folders.

Flash drive: A storage media which stores data by electrically changing the physical properties of an integrated circuit. Includes USB drives and SD cards.

Floppy disk: A storage media which stores data magnetically on a single non-rigid disk. Primarily is of interest from a historic perspective. Devices which read this technology are today not commonly available, making reading the information embodied in such media somewhat difficult.

Folder (also called a directory): A fundamental grouping and organising unit provided by a file system and typically represented as an icon depicting a folder.

Forensic computing: See digital forensics.

Forensic image: A complete, identical, and authenticable copy of the content of a digital storage device.

Hack: The act of identifying and exploiting weaknesses in the design or operation of computer systems in order to gain access beyond that which is allowed by policy.

Hash: See cryptographic hash.

Hard disk: See hard drive.

Hard drive: The primary media for storage of digital information in 2011. Information is stored on one or more hard (rigid) magnetic disks within the drive. Hard drives have traditionally been housed within the computer, removal typically requiring disassembly of the computer.

Hardware: Hardware refers to the physical devices which comprise the computer and networks the operator of a computer interacts with. For example, this includes the keyboard, mouse, monitor or screen, and workstation.

Hexadecimal: A base 16 numbering scheme used to represent numbers.

Internal metadata: Information describing a document which is stored within the content of the document.

Internet: The collective of computers and digital devices connected together to facilitate communication between computers. The internet provides transit for numerous services, of which email and the world wide web are arguably the two most popular.

IP address: A unique number which identifies the address of a particular computer or digital device which participates in internet oriented communications. An IP address is to a computer in a similar manner as a mobile phone number is to a mobile phone.

Malware ("malicious software"): A class of software whose operation is to the detriment of a computer's operator or owner. Examples of malware include computer viruses, worms and spyware.

Media: Media refers to the location where digital information at rest resides. Such storage devices include floppy disks, hard disks, USB thumb drives, DVDs, and the flash cards found in digital cameras.

Metadata: Information describing other information. Typically describes the characteristics of subject information. For example, document metadata may describe the author and last editing time of a document.

MP3: A format for storing digital recordings of sound in a compressed encoding.

Native format: Electronic documents have a particular structure which is defined by the application which creates the original file. Information created and stored in such a file is said to be in its native format. Production of native format files is desirable in litigation, as such copies generally preserve both content and metadata. Viewing such a file may require the originating application (for example viewing a Microsoft Word document may require the Microsoft Word application). In some instances (for example due to unavailability of the originating application due to licensing), it may be more practical to transform a document into a more readily viewed format (for example, transformation of a Word Document to a PDF). In the context of disclosure of corporate email systems, “native format” production requests may be met by production of a derivative copy which reasonably preserves the content and metadata in an electronically usable manner. Such a production may be more correctly termed “quasi native format” or “near native format”.

Network: A collection of computers and computing devices interconnected by communications links.

Operating system: A software environment which facilitates the interaction of a user with the operation of multiple applications on a computing device. In 2017 the dominant operating systems are Microsoft Windows (Windows) and Apple Macintosh (MacOS) operating systems on laptop and desktop computers, and Android and iOS on mobile phones and tablets.

PDF (Portable Document Format): A species of file of a particular format suited for efficiently representing documents. Popularised by the free availability of the Adobe Reader application.

Persistent storage: Storage media which stores digital information beyond removal of power from a computing device. Includes hard drives, DVDs and flash memory.

Program: See software.

Random Access Memory (RAM): The short-term memory of a digital device. Information stored within the RAM of a computer is lost when power is removed from the device.

Server: A computer which serves a particular purpose and is shared by multiple users. Examples include file servers and mail servers.

Software: Information interpreted by a computer to process other information within a computing device. Applications and operating systems are both examples of software.

Timestamp: A record of a time and date.

Tor: A system for enabling anonymous communication via the internet.

Unallocated space: The portion of a storage device not currently assigned for storing information. Typically this is the area from which deleted files may be resurrected.

Uniform Resource Locator (URL): A textual address used to identify content and information stored on a website. Each webpage typically has a unique URL which, on entering the URL into a web browser, will result in the webpage being displayed. In general, URLs begin with “http://”.

USB drive (or thumb drive): Storage media favoured due to its small size, low cost and convenient ability to connect to most computers.

Virtual machine: A simulated surrogate of a computer.

Web browser: An application used to access services and applications which reside on remote computers (websites). As of 2011, common web browsers include Mozilla, Internet Explorer and Chrome.

Website: An information source which resides on the internet, accessible via a web browser. In 2017, services such as Google, webmail, Facebook, and eBay are all examples of well know websites.

World wide web (WWW): Refers to both the system of software which powers websites and web browsing, and the collective of all websites which exist on the internet.

Write blocker: A software or hardware function which enables storage media to be read, while preventing changes to the media.

[1] SWGDE (2009), SWGDE and SWGIT Digital & Multimedia Evidence Glossary, Scientific Working Group on Digital Evidence and Imaging Technology.

[2] Sedona Conference (2007), *The Sedona Conference Glossary: E-Discovery & Digital Information Management (2nd edition)*, The Sedona Conference.

[The next text page is 101-401]

SCOPE

[101.100] Scope

Digital forensic evidence is sourced from digital devices such as computers, mobile phones, and modern communications networks. Digital forensics is a discipline within the forensic sciences whose primary concern is the scientific examination, analysis and/or evaluation of digital evidence in legal matters[1]. The relative infancy of the field is demonstrated by the creation in 2008 of a new section within the American Academy of Forensic Sciences (AAFS) called “Digital & Multimedia Sciences”. This was the first new field to be acknowledged by the association in 27 years.

This chapter aims to inform the legal professional and fact finder as to the foundations, context, principles, practices, limitations and challenges of the field of digital forensics, in order that they may understand the field enough to effectively engage with the digital forensic expert. It is anticipated that this chapter will additionally be of interest to practitioners and researchers in the field.

The chapter broadly describes the context of the field before addressing the key concepts related to digital evidence, in terms of the digital environment, legal definitions, principles and varying perspectives. Following this, the scientific foundations and practices of the field are described. Current challenges in regard to utilising digital evidence are identified and the role of errors and validation examined. Finally, the state of the field in regards to professionalism is described.

[1] Based on the SWGDE definition in *SWGDE and SWGIT Glossary of Terms* (2015).

[The next text page is 101-601]

INTRODUCTION

INTRODUCTION

[101.200] Introduction

In the early days of computing, computer evidence often meant, "the regular print out from a computer". Today, the term "digital evidence" is used to inclusively refer to evidence stored within digital devices such as computers, tablets, mobile phones, GPS and captures of data transmitted over communications links, to name only a few.

[101.210] History

The earliest reported court cases related to computer evidence go back to the late 1960s, a time when big business was beginning to realise the benefits of automation of record keeping and information processing. On occasion, computer evidence would be presented by the maintainers of such computers. Such a person would typically have an intimate knowledge of the operation of the computer, and could accordingly attest to the reliability and interpret or explain the meaning of the evidence produced from it, and additionally act as an expert in interpreting evidence for the court.

As "personal" computers proliferated in the mid to late 1980s and early 1990s, the justice system was faced with criminals who used computers. This era saw the beginning of digital evidence being analysed and presented by third parties rather than the operators of computers. In criminal matters this typically fell to hobbyist police personnel with a particular interest in computers, and in civil matters, to independent experts from academia and information technology (IT) professionals. At a national level some specialised communities of interest and investigative units focusing on digital evidence were established.

The mid to late 1990s and early 2000s saw the widespread use of the internet, with communication, and in particular email, being the initial driver, with online commerce soon following. Crime involving computers increased accordingly, driving police forces to form organisational units whose focus was dedicated to digital evidence. It is during this period that the rapidly growing community of digital forensic practitioners began to organise, with the establishment of government sponsored bodies focusing on providing training and attempting to increase quality and consensus. This period saw corporate software tool vendors and academia increasingly take an interest in the area.

The mid to late 2000s saw the proliferation of computing into smaller and smaller devices, including mobile phones, MP3 players and tablet computers. This period saw significant growth in training and research through academic participation, with private practice service providers growing rapidly.

The era of the 2010s can be characterised by pervasive connectivity, and commodity computing devices. Desktop computers have been displaced within the home environment by laptops, mobile phones and tablets. Computing devices are now a commodity, with low powered computers embedded into household appliances, car consoles, and wristwatches to name a few. Whereas the former eras involved data primarily being stored on the computing device, information is now regularly stored remotely on third party computer systems, and often in different legal jurisdictions.

[101.220] Defining the field

Practitioners within the field initially referred to the field as "computer forensics" and "forensic computing"[1][2]. The term "digital forensics" emerged to take into account then emerging sources of evidence such as that sourced from network communications and mobile phones. The term "IT forensics" (as in information technology forensics) is infrequently used in reference to the same field; however, the term lacks general acceptance amongst practitioners.

From its beginnings in law enforcement, the field of digital forensics has grown to embrace a wide range of stakeholders with significant variance in terms of evidential standards and rigour. Accordingly, defining the nature of the field in a manner that satisfies all stakeholders is problematic. In the litigation and criminal prosecution and defence contexts, the following two definitions are illustrative of the general range of perspectives.

In 1999, McKemmish defined forensic computing as:

> The process of identifying, preserving, analysing, and presenting digital evidence in a manner that is legally acceptable.[3]

The "Scientific Working Group on Digital Evidence" in 2005 defined computer forensics as:

> The scientific examination, analysis and/or evaluation of digital evidence in legal matters.[4]

The first definition reflects the evolution of the field as a set of pragmatic techniques and procedures to bring digital evidence into court, with the only scrutiny applied being equal to the strength of scrutiny applied by the court or the opposing side in civil or criminal matters. The second definition reflects a growing pressure to build the field as a forensic science. This pressure comes from both within the practitioner community and from the courts in seeking reliable evidence from skilled practitioners.

[1] Sammes et al (2000)

[2] Kruse et. al (2001)

[3] McKemmish (1999)

[4] SWGDE (2015)

[101.230] The field in practice

Digital forensic practitioners are called upon in a wide variety of contexts, including without limit as:

- an expert witness providing opinion on digital evidence, or computing related matters;
- an expert witness providing technical evidence on computing related matters;
- a technical witness providing technical evidence on factual matters related to digital evidence;
- an independent computer expert;

- an investigator providing information in relation to claims where litigation may be anticipated; or
- a consultant providing support for litigation.

A common model in policing is for digital forensic analysts to work under instructions of investigators. In private practice, the role of investigator is typically taken on by the forensic expert, acting under instructions of a solicitor. In both of these cases the overriding duty of the expert is to the court.

Current practice directions in Federal and State Supreme Court jurisdictions related to the execution of search orders (formerly Anton Piller orders) provide for an "independent computer expert" to assist an independent solicitor in regard to digital evidence[1]. This role typically focuses on evidence preservation and search, rather than involving analysis or testimony.

Litigation support roles typically involve assisting litigators with managing the complexities of disclosure of digital information, or for providing strategic advice in relation to expert evidence.

[1] See for example, Federal Court of Australia Practice Note – Search Orders, as at 1 Nov 2016 -http://www.fedcourt.gov.au/law-and-practice/practice-documents/practice-notes/gpn-srch (accessed 13 October 2017).

[101.240] Related fields

The litigation support role of digital forensics has evolved into a somewhat distinct field known under a number of monikers, but primarily as electronic disclosure (e-Disclosure) and electronic discovery (e-Discovery). Electronic disclosure is concerned with managing the unique constraints and challenges which are a consequence of conducting disclosure over electronic documents. This principally includes, without limit:

- identifying sources of potentially relevant documents in complex digital environments;
- conversion of documents into formats which might be easily perused and read; and
- formulating and executing strategies for efficiently locating relevant documents using automated means.

In comparison to digital forensics, the field of ***electronic disclosure*** is focused on facilitating disclosure activities principally on electronic documents and records; information which typically has its origin in a human utterance or writing, then stored or communicated using digital technologies. Digital forensics generally has a wider purview, focusing on any information stored in digital form, and its relationship with attribution, provenance, authenticity, and events which may have occurred.

The field of ***incident response*** emerged from within the computer and information security fields, and is primarily concerned with responding to intrusions in computer systems. Priorities of resumption of service and prevention of re-occurrence historically drove the field, with forensic concerns such as attribution and prosecution traditionally taking a back seat. More recently, the incident response community has begun to adopt the principles, tools and methodologies of the field of digital forensics as businesses have become subject to regulatory requirements, such as notifying customers when security breaches occur, and the effects of security incidents have shifted to the criminal and litigious realm.

Within the wider information security field, practitioners and vendors have recently begun to adopt techniques and methodologies of the field of digital forensics towards auditing and interrogating IT environments with the goal of assuring that such environments remain secure.

The field of ***data recovery*** exists within the general information technology field. Its primary focus is finding and recovering information stored on faulty digital devices. For example, data recovery firms specialise in diagnosing which crucial part of a storage device has failed, sourcing replacement parts, and replacing such parts in order to recover information from a device.

[101.250] A note on precise terminology

This chapter employs the term "digital evidence" to refer to digitally stored or transmitted information which may be of probative value. This convention emerged from within the digital forensics field. The terms "electronically stored information" (ESI) and "electronic evidence" have emerged from within legal contexts to refer to electronically stored or transmitted information of potential probative value.

The subtle difference between electronically and digitally stored information lies in the technologies employed. The first generation of electronic technologies, known as "analog" technologies, include early generations of telephone, audio cassette and video cassette recorder (VCR). While still based on electronics, the subsequent generation of electronic technologies, which are referred to as digital technologies, store and transmit information in a fundamentally different way. Today the overwhelming majority of evidence admitted as ESI is, to be precise, digitally stored information.

The field of digital forensics is focused largely on the subset of electronic evidence which is digitally stored or transmitted. The remaining subset, "analog" electronically stored evidence, requires a wholly separate field of theory, method, and technique in regard to expert evidence.

[The next text page is 101-801]

FOUNDATIONS OF DIGITAL EVIDENCE

[101.400] Introduction

The term “digital evidence” generally refers to two distinct things at the same time: digital information and the media in which it is stored. As digital information is always stored using a physical form, it may be treated as a form of physical evidence. A consequence of this is that many of the existing rules of evidence and legal precedent may be applied equally as well to digital evidence as to regular physical evidence.

Digital information is, however, at its foundations independent of physical matter, and brings with it a unique “physics” of its own[1]. For example, the prevailing scientific theory is that physical matter is divisible, from material, to molecule, to atom and beyond. Digital information, however, has a finite granularity; the “bit”. This section describes the unique characteristics of digital evidence and outlines the implications of those characteristics.

[1] Cohen (2009)

[101.410] Fundamentals

The discriminating traits of digital evidence are the language and symbols used by digital devices to store and communicate information. In much the same way that the Braille writing system used by the blind relies on the presence or absence of a raised bump as the fundamental unit of information storage, digital evidence is composed of sequences of “bits”: distinct units of information which can only convey two distinct states. These states may be stored, for example, by two positions of an electrical switch, two directions of magnetisation, or two distinct levels of light intensity. When a bit is considered on its own, the two values are typically referred to as representing the states “on” or “off”, or as numbers, “1” or “0” (known as “binary numbers”).

[101.420] Perfect fidelity copies

Sequences of bits are discrete and independent of the medium on which they are stored or transmitted. They may be copied exactly, in such a manner that the original is left unaltered. A complete bit sequence copied from original storage media is indistinguishable from the original as far as its information content. Such a copy is called an “image” or “forensic image”.

[101.430] Latent

Digital information is encoded as bit sequences and then stored or communicated by converting the bit to a physical property, whether it be by exploiting magnetic fields in the form of hard disks, or variations in electromagnetic radiation in many communications technologies. Such encoded information is latent, ie it is not directly perceivable using unaided human senses, in the same way that one is unable to hear the music on an LP record merely by looking at the grooves on its surface.

Interpreting information from media almost always requires the use of some kind of automated process or tool to observe, characterise and represent such evidence. Accordingly, the failure modes and error types of such tools are of key significance when considering the validity of any conclusions which are based on the use of those tools. See [101.1600] for detailed coverage.

[101.440] Interpretation

The Braille language arranges dots into 2 x 3 chunks (*6 dots*) cells, with various combinations of dots by convention representing the English alphabet, punctuation marks, and in some instances, contracted common words. Similarly, in digital systems, bit sequences of *8 bits* (historically referred to as "bytes") or 16 bits, may represent the letters and punctuation of various languages. Depending on the context, exactly the same bit sequences may be interpreted as different meanings, for example as counting numbers or as text. This representation process is called "encoding".

Similar to the way that syntax, grammar, semantics, and publishing conventions are used to interpret sequences of letters and punctuation into words, sentences, paragraphs and readable books, bit sequences are interpreted based on their context into documents, images, music and the like.

An example of the way that bit sequences may be interpreted and represented follows. With reference to Table 1, three different interpretations of the same bit sequence are shown. A particular bit sequence is represented in row 2, encoded as binary numbers. Working with digital evidence at the bit sequence level is, however, tedious. Consequently bit sequences are commonly represented by a more concise numbering scheme based on a numbering system called hexadecimal. The hexadecimal numbering system employs a 16 symbol numbering scheme, using the symbols 0-9 to represent values zero through nine, and A-F (case insensitive) to represent 10 through 15. An alternative hexadecimal representation of the bit sequence at row 2 is presented at row 3. The bit sequence at row 2 may be interpreted as the text at row 1 by employing the ASCII text encoding scheme. The American Standard Code for Information Exchange (ASCII) standard originated in the 1960s and was originally used for communications between tele-printers. It remains pervasively used for representing English language text today.

TABLE 1 Interpretation and representation of bit sequences

1	Interpreted meaning	The fat *cat* sat on the mat.
2	Encoded Bit sequence	01010100 01101000 01100101 00100000 01100110 01100001 *01110100 00100000 01100011* 01100001 01110100 00100000 01110011 01100001 01110100 00100000 01101111 01101110 00100000 01110100 01101000 01100101 00100000 01101101 01100001 01110100 00101110
3	Bit sequence represented in hexadecimal	54 68 65 20 66 61 74 20 *63 61 74* 20 73 61 74 20 6f 6e 20 74 68 65 20 6d 61 74 2e

[101.450] Volume and duplicability

Traditional communication of written information in the terrestrial world relied on moving physical information containing things from place to place (ie postal mail). In contrast, communication in the digital world involves transmission and receipt of perfect copies of information. Such processes tend to produce numerous copies dispersed into multiple locations. For example, a report may be stored on a worker's desktop as a file, then when it is emailed to a colleague, a copy may be stored in her sent items within her email, another copy stored on the mail server of her company, and finally another copy stored in the inbox of the email recipient.

A consequence of this is that the economics of information production has changed, with the cost of copying information now effectively nil. This, in conjunction with the identical nature of copies, has produced marked effects in the way we produce and retain information. Significant amounts of information are now produced by copying pre-existing information. For example, carbon copying (CC) documents, saving versions and using precedents as the basis of new documents all exploit the cost-free nature of copying.

Accordingly, the rate of production and retention of information has accelerated dramatically. On a yearly basis, the capacity of storage devices in general grows almost by double, while the speed at which we are able to read the contained information only grows at around 10%. This complicates the speed with which digital evidence may be preserved and searched. This has marked effects on the way we collect, analyse and search for evidence.

[101.460] Fragility

A second aspect of the principle of perfect copying is that digital evidence is fragile, being subject to inadvertent modification or deliberate forgery. At the level where bits are encoded as a physical property, modifications are, in the vast majority of cases, undetectable.

Outwardly simple acts such as powering on, or connecting a digital device to a computer, have far reaching effects on the bit sequences involved, producing unrecoverable changes. For example, simply logging into a system to "have a peek" at the workspace of a suspect may result in deletion or the overwriting of key evidence related to the operator's interactions with the system, such as which files the operator recently opened and which websites the operator visited.

Accordingly, digital forensic practice aims to prevent, and if not able to prevent, then to minimise, the amount of change to the bit sequences within digital media and devices considered as potential evidence. In order to identify any changes in bit sequences, mathematical signatures called cryptographic hashes (described in a later section) are employed. Despite the fragility of bit sequences, significant amounts of redundant information are typically stored in digital devices, which in some cases enables detection of modifications.

[101.470] Trace evidence

The relationship between actions and physical evidence in the terrestrial world is generally understood as adhering to a principle of transfer of discrete pieces of matter between things that come near each other (Locard's exchange principle)[1]. A similar principle applies in the digital realm. When a user interacts with a digital device, or when a digital device performs some action, traces of that action remain as bit sequences. For most actions, the resultant traces will be overwritten and not persist; however, for some actions traces will remain. By knowing which actions create which traces, we may search for relevant traces, and where such traces are found, place such actions as a potential cause.

For example, all documents, and files in general within the common computing environments, will have associated with them time and date stamps ("timestamps") that are automatically updated based on a number of actions which occur in a file's lifetime.

Such traces are typically produced by the operating system of a computer based on a complex set of rules related to movement, modification, creation and copying of files, and may be interpreted as consistent with a number of actions which may have occurred in the history of the file. However, like all bit sequences, such records are subject to forgery. In this case, the records may be faked by direct forgery of their bit sequences or via setting of the computer clock (upon which their values are usually based) to times other than the correct one.

Much of the practice of computer forensics relies on identifying and evaluating such trace evidence, and relating the potential causes with the particular case theories.

[1] "… any time two or more surfaces come into contact with one another there is a mutual exchange of trace matter between those surfaces."

[101.480] Fast moving

Digital information, and the environment in which it exists, is subject to rapid evolution. This results in the raw material of the field changing at a rate challenging to the rigorous application of regular scientific processes. For example, a significant portion of the mobile phone using community replace their phones every two years. Replacements typically involve entirely new technologies, or variations significantly different from their fore-bearers. Consequently, forensic techniques must continually be adapted to address these types of changes.

[The next text page is 101-1001]

PERSPECTIVES ON DIGITAL INFORMATION

[101.600] Introduction

Just as one can conceptualise a human being from various perspectives, ranging from a complex arrangement of molecules, to a system of organs, to a being capable of thought and emotion, so too can one view the bit sequences contained within computing systems from multiple perspectives. This section identifies and describes perspectives on data of relevance to discussing the forensic aspects of digital evidence.

[101.610] Bit sequence vs information

The latent nature of digital evidence was identified in the prior section. Information only occurs in context, encoded as bit sequences, and must be materialised by some set of rules in order to be interpreted as information.

[101.620] Computer stored vs computer generated

In regard to admissibility and weight, a significant discriminator of information embodied within digital evidence is the ultimate means of production of such information.

Computer-stored records or documents typically contain the expression of a person. Common examples generally include the content of emails or word processing documents.

Computer-generated records contain the results of some automated process, untouched by human expression. Examples include digital photographs and logs of activity maintained by computing devices (event logs).

Finally, there is the category of records which are both computer stored and computer generated. For example, a spread sheet may contain rows and columns of figures which are entered by a person (computer stored) and calculation results based on those figures (computer generated).

[101.630] Content vs metadata

Another discriminator is between the generally perceivable portion of a digital record or document and other related information.

Content is a term used to identify information as opposed to its container. The content of a PDF file is typically a textual document which may be displayed on a computer screen or printed to paper.

Metadata is ancillary information related to a document or other entity. Metadata is pervasive within digital systems and is at the foundations of the organisation and maintenance of information. Examples of metadata in the physical world include post marks used within the postal system, and call numbers attached to books in libraries.

Internal metadata is information relevant to a document which is stored within the content of the file. For example, most PDF documents contain internal metadata which documents the time and date at which the document was created. Some mobile phones store the location where a photograph was taken as latitude and longitude metadata within the same file as the digital photograph.

External metadata is information relevant to a document or file which is stored in a location apart from the content of a file. For example, in all common operating systems, each file has associated with it metadata regarding its organisation with respect to folders, the size of the content of the file, and the date and time the content of the file was last modified, to name a few.

As the majority of metadata artefacts are generated by the computer as a matter of course in general operation and users are generally unaware of the full range of metadata maintained, metadata features significantly in questions of provenance and authenticity. For example, claims of document backdating and manufactured emails will typically involve an examination of metadata related to such documents.

[101.640] Allocated vs unallocated

This distinction between allocated and unallocated storage is relevant to the recovery of deleted information. Files, documents, records, and the artefacts used to organise them are said to be allocated from the available storage area of storage media. The rest of the space is said to be unallocated. When a file is deleted, the storage area which was allocated to the storage of that file is returned to the pool of unallocated storage, ready to be allocated to the storage of another file in the future. If those areas are not overwritten by new content, the file continues to exist and is said to exist in unallocated space. The content and metadata are recoverable from unallocated space via a number of techniques.

This distinction is of relevance to the completeness of preservation of evidence and matters where possession, custody or control of content is of significance. In the English case, *R v Porter*[1] the Court found that Porter was not in possession of child exploitation material found in the unallocated space of a storage device. In certain electronic disclosure matters, it may be argued that it is unreasonable to require search and recovery of documents potentially found in unallocated space; however, if the allegation is that of deliberate deletion, then it is reasonable to require acquisition and search of the unallocated space.

[1] [2006] EWCA Crim 560.

[101.650] Part vs whole

The physical nature of terrestrial property and goods imposes implicit constraints which do not necessarily have equivalents in the digital environment. For example, while crime scene investigators may collect items which are small and discrete, it is generally impossible to remove an entire room from a house, walls included. In the digital environment, it is generally feasible to do the digital equivalent of collecting everything, from the small and discrete to the entire house, without damage or contamination.

The ability to preserve evidence more widely than in physical crime scenes has led to disputes in regard to interpretation of the scope allowed by search orders and legislative provisions granting third party access to digital information. The crux of such disputes is whether an entire storage device (or a forensic image of a storage device) is to be considered a document. For example, in a 2003 Queensland case, *TLC Consulting Services Pty Ltd v White*[1] the Court found that while a computer (a server) was a "repository of records", the entire contents of a computer's storage was also to be considered a record.

Resolving such boundary issues in the digital environment is not straightforward, requiring a refined view of possession and control inside the digital environment. To draw a parallel, in the terrestrial world, orders allowing one to search for records in a building at a street address may be overly broad in the instance where the building is an apartment building. Similarly, in the digital environment, a single computer may contain records in separate areas for unrelated legal entities, with a single server containing multiple "apartments", each separately owned and controlled. In 2017, a common method for dividing single computer servers into multiple apartments is through the use of a "virtual machine", a simulated surrogate of a computer.

[1] [2003] QCA 131 (unreported, de Jersey CJ, Davies JA and Atkinson J, 21 March 2003, BC200301227).

[101.660] Volatile vs persistent

The lifetime of information both within and between humans varies: irrelevant events quickly dissipate from short-term memory, while important events become memories, and perhaps are made material by writings or recordings. Within the digital environment there is a similar variability, where information held in the short-term memory (the Random Access Memory (RAM)) is highly volatile, and more important events persist within storage devices.

Traditionally, only information which persisted after electrical power was removed from the computer was acquired in computer forensic investigations. However, it is possible to copy both the storage devices and the volatile memory of a computer. By omitting preservation of the volatile memory of the computer, the completeness of the preservation could be argued to be deficient.

Without such evidence, arguments that exculpatory evidence found only in the volatile memory has been lost are made. For example, the "trojan defence" argues that the owner of the computer was not responsible for the illicit content found on her computer. The true culprit was a piece of software which infected the computer, downloaded the illicit content, then deleted all traces of itself. Such a defence has been successfully argued in the past. Where volatile memory evidence has been preserved, it may be analysed for traces of such malicious software, and go some way towards countering or supporting such arguments. In 2017 the value of volatile information is increasingly appreciated within the field, and acquisition of the short-term memory (RAM) of computers is commonplace in investigations involving malware, and regularly collected in criminal contexts.

Broadening focus from the volatile short-term memory of digital devices to long-term (persistent) storage, the volatility of such stored data is still a significant concern. Automated processes within the digital environment regularly cause data to be destroyed through their designed operation. For example, backup systems are typically designed to regularly overwrite prior backups, and while emails in the Microsoft Exchange system are "believed" to be deleted when removed from the deleted items folder, they typically persist for an additional 30 days if not more, before a regular process finally deletes them. The court may grant search orders to preserve such volatile evidence, even in the absence of evidence that destruction will be deliberate[1].

[1] *Universal Music Australia Pty Ltd v Sharman License Holdings Ltd* (2004) 205 ALR 319.

[101.670] Local vs remote

For much of the history of computer forensics, digital evidence has generally been stored locally in the computer. Much of the technique of the field focused on gaining access to such digitally stored information in a manner that preserved its integrity.

Most smart phones, tablets, and traditional computers have constant access to the internet via wireless means. This constant network connectivity has enabled information to be stored remotely rather than locally, a shift that has been exemplified by applications such as webmail, social networking, and cloud storage systems.

This shift has created significant challenges in regard to obtaining appropriate authorisation and gaining access to evidence, due to such remotely stored information regularly residing in jurisdictions other than the primary investigation site. In the business sphere, many organisations have shifted their information and applications into “the cloud”, which often means their data may be held in data centres in multiple jurisdictions. In the personal sphere, this is particularly evident where social media, webmail and other online publishing methods are used.

In the criminal prosecution context, policing entities gain a varying level of assistance from the larger application providers, depending on whether the information requested is to be used as evidence or for intelligence. For the former, the Mutual Legal Assistance Treaty (MLAT) is observed to be effective in only limited circumstances due to the time and organisational overheads involved. In the civil litigation context, the court orders of one country are generally not honoured in another country. For example, litigants in Australia must initiate separate legal action in the jurisdiction in which the application provider is domiciled, often the USA.

[101.680] Locked vs unlocked

The use of encryption is a fundamental building block of modern computing, and essential for maintaining privacy over, and authenticating the identity of, the participants of online transactions. Regularly mobile phones and tablets, and to a lesser extent, laptops, are locked via the use of a password or PIN, and the information stored on those devices protected by encryption. Those protections are sufficiently robust on devices manufactured by Apple as to regularly render their content (as of 2017) inaccessible via reproducible, peer reviewed forensic techniques.

[The next text page is 101-1201]

DIGITAL FORENSIC PRACTICES

[101.800] Introduction

It is common to describe digital forensic methods in terms of discrete activities which are related to digital evidence. This section describes these overarching activities in brief.[1]

[1] For a more in-depth treatment of digital forensic related methodology see Casey E, *Digital evidence and computer crime*, 3rd ed, s 1 (Elseveir, 2011).

[101.810] Authorisation

Digital investigations rarely occur outside of considerations of authority and ownership. In the criminal justice system and the regulatory environment, authority to search and seize third party digital evidence is typically pursuant to legislative provision. In civil litigation, authority to search third party information may be by discovery (disclosure), preliminary discovery and search (Anton Piller) orders.

In employment matters, for example investigations into inappropriate usage within the workplace, or computer intrusions, the co-mingling of personal or private information with company records, and reasonable expectations of privacy, are of significant import in regard to the employer’s authority.

[101.820] Survey

Survey refers to the act of identifying potential sources of digital evidence. For example, this could involve identifying computers, mobile phones, and USB storage devices at a physical crime scene or location and identifying internet based services under the control of a suspect, such as webmail, social media (ie Facebook), cloud storage (ie Dropbox) and iCloud.

[101.830] Acquisition and collection

Once potential sources of evidence are identified, the next concern is that of preserving the evidence as much as possible in the state in which it was found. Acquisition and collection are activities related to preserving the bit sequence content of digital evidence media. In criminal matters it is common to collect a computer as evidence and keep it in secure storage, then acquire the storage of the computer just prior to examination. In civil matters, it is more common to acquire images of evidence than to collect them.

Collection refers to the act of physically taking into possession an item containing digital evidence, whereas acquisition generally refers to the creation of a perfect fidelity copy (referred to as a "forensic image") of the evidentiary material. In the more mature area of disk forensics, storage media typically has standardised interfaces which enable a complete[1] and perfect fidelity copy which is both authenticable and complete, while preserving the integrity of the original source of evidence. Numerous techniques may be employed to effect such a copy.

The most common acquisition technique typically involves the following:

1. The suspect computer is powered off.
2. The suspect storage device(s) within the computer (typically a hard drive) are identified, removed and it/their identifying and relevant features documented.
3. The suspect storage device is attached to a device which in general prevents any modification of the original evidentiary material (a "write blocker").
4. All of the accessible content of the hard drive is read through the write blocker and stored within a series of files which form a perfect and authenticable surrogate for the data stored within the original device (a forensic image). In doing this, a cryptographic hash is calculated over the data as it is read and, on completion, the calculated hash is stored in or with the copy.
5. A cryptographic hash is then calculated from the forensic image and compared with the earlier calculated hash to ensure that the stored copy is exactly the same as the original.

The above acquisition procedure produces a complete copy of the bit sequence content of the original storage media. In contrast, acquiring copies of files alone results in the omission of numerous potentially relevant evidentiary artefacts, including deleted files in unallocated space, file metadata, and alternate data streams. It is important to ensure that the method of acquisition results in a complete and accurate copy of the original in order to counter later claims of missing exculpatory evidence and incompetence, and to prevent the consequent effects on the admissibility and weight of evidence. In the Colorado case, *Gates Rubber Co v Bando Chemical Industries Ltd*[2], a matter where claims of intentional destruction of evidence were at issue, a technician attempting to acquire evidence overwrote 7-8% of the original evidential storage media and only made a "file by file backup" of the storage media, omitting deleted files and certain file metadata. The failure of the technician to acquire a full image of the original storage weighed heavily against the plaintiff, with the Court agreeing that in this instance the technician had "a duty to utilize the method which would yield the most complete and accurate results."

Much of the accepted theory underlying forensic acquisition methods and procedures addresses the above benchmark. However, as the field has grown, acquisition techniques have come to vary both between and within the digital forensic sub-disciplines, based on the suspect device which is being acquired, and operational concerns. The acquisition techniques may be categorised as follows:

- **Physical acquisition**: A complete copy of all data on the device is taken.
- **Logical acquisition**: A copy of only part of the accessible data is taken.
- **Live acquisition**: The copy operation is taken while the device is powered on.
- **Dead acquisition**: The copy operation is undertaken with the device powered off.
- **Physical collection:**The actual hardware device is seized (for later acquisition).

In civil matters, operational concerns and commercial considerations related to loss of productivity or revenue motivate alternative approaches to powering off devices in order to copy them. For example, in corporate computing environments requiring non-stop operation, live physical acquisition is increasingly adopted to mitigate the expense associated with powering off computers.

In contrast to the storage forensics specialisation, a mobile phone or tablet may only support an interface which enables one to economically extract active SMS and phone book entries from a powered on phone (ie live logical acquisition). This technique typically has the effect of modifying to some extent the evidentiary material on the suspect phone. Mobile phones and tablets (in particular iPhones and iPads as of 2017) are regularly locked to the extent that forensic acquisition is not economically or even practically possible.

It is important to note that historically, good practice guides related to digital evidence held to the principle that copies be taken of evidence without the alteration of the original evidence. For example, the UK-based Association of Chief Police Officers (ACPO) stated:

No action taken by law enforcement agencies or their agents should change data held on a computer or storage media which may subsequently be relied upon in court.

This is at best an ideal to strive for, as unobservable changes regularly occur outside the purview of the forensic examiner. For example, the act of powering up a hard drive may cause certain systems management related storage areas of the drive to be modified, reflecting the number of hours the drive has been powered on; and the act of acquiring volatile memory involves interacting with and changing the state of the computer. At its worst, this ideal holds the potential to create unwarranted doubt and confusion based on overly broad claims of contamination. With reference to the “part vs whole” issues identified at [101.650], claims of contamination should demonstrate contamination of specific artefacts within the media of direct relevance to the matters of forensic significance, rather than the media as a whole.

The ACPO acknowledged the boundaries in applicability of the above principle, stating:

> In circumstances where a person finds it necessary to access original data held on a computer or on storage media, that person must be competent to do so and be able to give evidence explaining the relevance and the implications of their actions.[3]

The challenge in forensic acquisition lies in quantifying the changes which may have been made to the digital evidence during acquisition in such a manner that claims of contamination or overwriting of exculpatory data are sufficiently addressed.

[1] This is not strictly true, as a digital device often contain storage areas which are not accessible to the operator or forensic investigator in an economical manner.

[2] 167 FRD 90, 112 (District of Colorado, 1996).

[3] ACPO (2012)

[101.840] Preservation

Preservation refers to activities which are undertaken to assure the provenance, integrity and authenticity of evidence. As in the physical evidence related disciplines, continuity of custody and provenance is established and maintained, and access to the evidence is controlled.

As digital evidence may be copied with perfect fidelity, there exists the risk of undetected modification; intentionally, accidentally, or merely through natural degradation of the materials which store the digital information. This risk is mitigated by the use of *cryptographic hashes* as a means of quickly determining whether apparent copies of bit sequences are the same or differ. A cryptographic hash is a compact and unique numeric identifier (hash value) created by applying a complex mathematical algorithm to a sequence of bits, whether they be content of a file, or all of the bits on a hard drive. By comparing the hash values of two sequences of bits, one can quickly tell, to a very high statistical probability, that the two sequences of bits are exactly the same, or different.

Table 2 provides an example which illustrates the differences which occur in hash values when two similar but subtly different inputs are used. The hash algorithm used is the Message Digest 5 (MD5) algorithm, and is used in this instance for conciseness. The hash value presented in row 1 is of the bit sequence which corresponds to the sentence to the left when using an ASCII encoding. In the same way that the sum “1 + 1 = 2” is true regardless of how it is calculated, under the ASCII encoding scheme, the hash value which is calculated from the sentence at row 1 will always be that as presented, regardless of how or where the calculation takes place.

TABLE 1 MD5 Hash values vary based on the bit sequence they are calculated from

	Text	MD5 Hash	Notes
1	The fat cat sat on the mat.	3b38f4c62adb8f6b497a6243ebcb9aa6	ASCII encoding
2	The fat cat sat on the mat.	28a0b8326d16aa834c1d4d1bdbe8a846	ASCII encoding

In view of this fact, at first glance one would expect that the text at row 2 would have the same hash value as for row 1; however, it is immediately apparent that the two hash values are not equal. On close examination, it is apparent that the text at row 2 is subtly different from the text at row 1, with an extra space inserted between the words “fat” and “cat”. This is the cause for the different hash value.

In practice, a hash is typically calculated from the source data during acquisition, and followed by calculation of a second hash from the result of the copy operation, the forensic image. The two calculated hashes being the same generally indicates that the copied bit sequence is an exact copy of the source data. Hash values are small enough (for example, a commonly employed hash, MD5, is represented as 32 written characters) that they may be written in case notes or otherwise recorded, in order that at a later date, a hash value may be re-calculated and compared to demonstrate that the copy remains unaltered.

The application of hashing in digital forensics relies on the statistical likelihood that no two different bit sequences will have the same hash value, and poses two potential errors:

- two different files (or forensic images) having the same hash; or
- two identical files (or forensic images) having different hashes.

The theoretical likelihood of the former outcome (called by cryptographers a “collision”) is $2^{n-1}$where n is the length in bits of the hash. For example, for the SHA-256 hash algorithm, which has a hash length of 256 bits, the likelihood of a collision (and hence an erroneous result) would be one in 2^{255} (or 5.7 x 10^{76}).

The latter potential error, that of two identical bit sequences having different hashes, is not possible and has an error rate of zero.

It is noteworthy, however, that for the two most commonly used hash algorithms in digital forensics (MD5 and SHA1) that the practical likelihood of collision is considerably lower than the ideal likelihood due to flaws in their design. Despite widespread reporting of these two algorithms being “broken”, in the context of their usage in authenticating forensic images, the likelihood remains sufficiently low to remain in 2017 reasonably assured[1][2].

[1] Thompson (2005)

[2] Stevens et al (2009)

[The next text page is 101-1401]

EXAMINATION

[101.1000] Introduction

Examination refers to activities and techniques which make the latent information embodied in digital evidence available for human perception. At the foundations, examination techniques rely on recognising context, structural relationships and patterns (formats) in bit sequences, to interpret those bit sequences into computing artefacts such as volumes and file systems and user-perceivable artefacts such as files, folders, emails or deleted records. At higher levels, examination techniques include indexing, searching, sorting and filtering methods which enable identifying evidentiary artefacts which match criteria such as keywords, times and date ranges.

The structural relationships and patterns employed are to a large extent based on the rules and conventions implicit in the designs and implementations of operating systems and applications. For example, early operating systems in the 1980s adopted a specific format for storing data in the first storage address of a hard disk, in order to both allow the operating system to load from an arbitrary position on a disk and locate where the file system was located. This structural format has over time become a convention which, followed on a computer of today, still allows one to find the file system present on a hard disk.

Due to the proprietary and commercial context in which computing has evolved, much of these conventions, rules and structural relationships which are necessary to interpret information from raw data are held closely by software manufacturers and not shared. Consequently, many of these rules and relationships have been identified through experimentation and reverse engineering, leading to variability in completeness and limits in understanding. Forensic tools based on such limitations may present incorrect interpretations of raw data. In this context, the information interpreted and presented by forensic tools must be validated in order that one may have reasonable confidence in the reliability of the information being examined. See [101.1800] for more detail on validation.

More mature or widely used structures, such as file systems, are in general comprehensively understood. The New Technology File System (NTFS) found on most computers running the Microsoft Windows environment has received significant scrutiny to the point that it is understood well enough that independently developed functional clones of the NTFS have been achieved. Such activities have provided the forensic community with a comprehensive theory of operation upon which analysis may be undertaken[1].

Many other structures with very specific uses tend to be only partially understood. For example, earlier versions of Microsoft Windows stored records related to the websites a user had visited in a particularly formatted file called "index.dat". Records interpreted from this file are routinely tendered as evidence of a user's activities in regards to searching and viewing of websites. The format of the structures in this file has never been publicly released by its maker. Only the experimentally derived interpretation of a small subset of the format has been documented and subjected to third party verification.

An example of the potential impact of misinterpreted browsing history lies in the Florida case of *Anthony*[2]. The accused was charged with drugging, suffocating and then dumping the body of her 2-year-old daughter. In support of the claim of Ms Anthony having drugged her daughter, deleted web browsing records, interpreted by a forensic tool called "Cacheback" were produced in support of claims of 84 visits to a webpage related to chloroform. Analysis with a competing tool, and subsequent analysis by the developer of the "Cacheback" tool, both repudiated the original interpretation, with only one visit to the chloroform-related webpage being interpreted.

A subtle point in relation to this experimental approach to identifying the meanings of storage formats, and consequently the materialisation of information from bit sequences, is that such information should not be viewed as based on a hard fact; rather, it is based on theory. The strength of such theories is, as in any science, dependent on the rigour of their testing.

The theories of operation described above are all related to digital artefacts which are accessible to some extent via the operating system which produced them. Atop the theory of operation of these formats and structures rests a related set of theories regarding useful side effects of the operation of the system, mostly related to areas of storage which are not accessible to the operating system (*unallocated space*).

Of the techniques related to unallocated space, the ability to find traces of deleted files is of common interest. The process of deletion of files in both the NTFS and FAT file systems popularised by Microsoft, results in the areas which were being used to store the file being flagged as not in use. Such remnants will remain unmodified until they are allocated to a new file and overwritten. Identifying structures which are flagged as deleted affords a straightforward means for recovering deleted files which have not been overwritten. In contrast to this, the file system employed by Apple MacOS based computers adopts a scheme which regularly results in the overwriting of the records which specify a file's name, size, and location (file metadata), making file recovery by the aforementioned means near impossible.

Storage devices often contain considerable quantities of intact file content not accessible to the operating system through structural means. This is due to the overwriting of the corresponding file metadata, which originally recorded the existence of and storage locations assigned to contain the pieces of the file. In such cases a technique known as *carving* may be employed to attempt to salvage deleted files from their constituent pieces. Carving relies on the internal characteristics and consistency of an assembled file in a similar way as one relies on the characteristics and consistency of a picture when putting together a jigsaw puzzle. While yielding good successes, such techniques are, however, error prone and still an open research problem.

All of the techniques described so far have focused on identifying and interpreting whole files; however, in many instances portions of file content may have been overwritten. Carving may be applied to finding subparts of files, for example, for identifying valid frames from an otherwise deleted video file.

When considering the techniques which involve re-interpreting information from bit sequences within digital evidence, it is important to be aware that it is common for portions of digital evidence to remain latent and obscure, with their underlying theory of operation and structure still unknown within the wider forensic community. The volume shadow copy component of the Windows operating systems remains only partially understood by the wider forensic community, and without doubt contains information of merit to investigations.

[1] This is true for versions of this file system adhering to pre-2003 versions of the operating system. Despite the comprehensive nature of this theory of operation, in 2017 portions (ie de-duplicated storage) of the more recent evolutions of this file system still remain without a clear theory of operation.

[2] *State of Florida v Casey Marie Anthony* 2008-CF-15606-A-O.

[101.1010] Finding, sorting and filtering

The numbers of individual files found on personal computing devices regularly exceeds a million. Accordingly, methods of both reducing the number of files to be considered, and finding files based on specific criteria, are – in nearly all cases – a necessity.

While search based on text is commonplace and a powerful tool for finding documents, it carries with it a number of constraints. Just as searching for relevant themes within a book benefits from a comprehensive index, so too does searching of digital evidence. Typically this requires as a preparatory stage that an index be generated of the terms within the digital evidence to be searched.

The comprehensiveness of such an index depends on effective recognition of bit sequences as containing text, which in turn relies on the lower level examination results described earlier in this section. This is not straightforward. File formats must be understood for effective extraction of text, and there are often files for which this is not possible. For any English word, there are typically at least two bit sequences to which it may correspond. When one adds variations based on upper and lower case the number of permutations which must be indexed grows further.

The ability to filter and sort based on criteria such as time and date, and type of file (ie video or document) are important in reducing evidence quantities. File hashes provide a convenient method for concisely identifying or "fingerprinting" the content of a file, a property which is used to detect duplicate copies of files. Pre-categorised collections of hashes (hash sets) such as the National Software Reference Library (NSRL)[1] hash-set maintained by the USA-based National Institute of Science and Technology (NIST), provide a means to rapidly identify and exclude large numbers of files which are commonly found on computers, such as the files which comprise the Windows operating system or the files which comprise a popular application.

[1] http://www.nsrl.nist.gov.

[The next text page is 101-1601]

ANALYSIS

[101.1200] Introduction

Analysis refers to the critical thinking which enables investigative questions to be answered through identification and interpretation of trace evidence. Products of forensic analysis may include:

- timelines illustrating the chronology of relevant events;
- link analysis charts showing interactions and relationships between people and things of relevance; or
- reconstruction of process to demonstrate what may have occurred in the past.

The primary theory underlying analysis of digital evidence is similar to Locard's theory of Exchange[1]; when a user interacts with a digital device, or when a digital device performs some action, traces of that action remain as bit sequences. For most actions, the resultant traces will be overwritten and not persist; however, for some actions traces will remain. By knowing which actions create which traces, we may search for relevant traces and, where such traces are found, place such actions as a potential cause.

In working backwards from traces to potential actions, one must be aware of the often wide range of other potential actions which could have caused such traces. Consider the naïve example that while rain leaves water on the ground, finding water on the ground does not mean that it just rained.

The fundamental methodology underlying a rigorous analysis is the scientific method:

- **Observation**: Observations regarding occurrences in either the physical or digital world lead to claims which form the point of departure for a digital investigation.
- **Hypothesis**: The claims are translated into hypotheses about events which may have occurred in the digital environment as a result of, or as a precursor of, the claims. Alternative explanations are considered, and hypotheses are formed about events which may have occurred in relation to those alternative explanations.
- **Prediction**: Based on the hypotheses about events, predictions are made as to what traces may have been left behind, and where and how they may be found.
- **Experimentation and testing**: Forensic investigators then employ analytic techniques in order to identify predicted trace evidence and inconsistencies not predicted by the operational hypotheses. A hypothesis is strengthened by finding evidence supporting multiple corroborating traces consistent with the hypothesis. The strength of hypothesis lies, however, in the extent to which attempts have been made to falsify

the hypothesis. Where there is one or more alternate hypotheses then the strength of each and all the hypotheses must be evaluated.

- **Conclusion**: Conclusions are then drawn based on the outcomes of the testing. Typically this results in the evidence supporting claims, being inconclusive, or refuting the claims.

The generation of hypotheses from claims typically involves an iterative process of generating sub-hypotheses until a point is reached where a prediction may be made regarding where a discrete piece of trace evidence may be located. The following example in Table 3 demonstrates how a single claim may be decomposed into multiple sub-hypotheses.

TABLE 3 Claims translated to hypotheses and subordinate hypotheses

Identifier	**Claim, hypothesis, sub-hypothesis**
Initial claim:	Key employee stole proprietary information while exiting the business.
H0:	Proprietary information was emailed out of the business –or–
H1:	Proprietary information was copied to a USB thumb drive and taken out of the business –or–
H2:	Other (eg photocopied).
H0.1:	Proprietary information was emailed by regular work email.
H0.2:	Proprietary information was emailed by private webmail.
H0.2.1:	Records of webmail related to proprietary information will exist as webmail fragments in the file system of the employee's laptop.
H0.2.2:	Records of webmail related to proprietary information will exist as webmail fragments in the volume shadow copy of the file system of the employee's laptop.

[1] "… any time two or more surfaces come into contact with one another there is a mutual exchange of trace matter between those surfaces.", Lee H, *Henry Lee's Crime Scene Handbook* (Elseveir, 2001).

[101.1210] Interpretation

The above methodology generally involves a broad understanding and consideration of:

- how the operators of computers generally behave and interact with computing devices to achieve their goals;
- how particular interactions between people and computers, and computers and other computers, leave behind traces;
- how other unrelated activities may leave behind equivalent traces which might be misinterpreted; and
- the limits involved in interpreting information sourced from bit sequences.

Failures of knowledge or consideration in relation to the above have the potential to significantly alter the conclusions drawn by the investigator. For example, a common claim is that of document fabrication, where it is claimed a document has been created at a time and date after its purported date of authorship. Such claims are commonly based on file and document internal metadata which can appear to the lay person as inconsistent, yet on analysis are explained by multiple differing historical narratives.

All conclusions based on analysis of digital evidence are in essence interpretive, as they are fundamentally based on sequences of bits, which may have been created by any number of means. For example, the presence of a particular time and date stamp within a Microsoft Word formatted document may generally be assumed to have been produced by the Microsoft Word application on a computer with a reliable clock and stored on reliable media. However, in specific instances, analysis and interpretation of the real world significance of such a time and date warrants a careful consideration of the potential ways in which a time and date stamp inconsistent with expectation could potentially have come into existence, which include without limit;

- deliberate forgery at the bit sequence level;
- a malfunctioning or manipulated clock;
- a misunderstanding of the full range of user activities which might result in the time and date stamp; or
- production by an application other than that assumed (for example due to editing the application using a competing product such as OpenOffice writer).

[101.1220] Attribution

Attribution is a special case of interpretation. Where a case requires attribution to a particular individual or entity, multiple corroborating sources of evidence, including evidence sourced from the terrestrial world, are preferable. For example, evidential records taken from a work computer shared by multiple people are subject to uncertainty as to the operator. Unrelated records of a personal nature, for example webmail or online banking, may go some way towards providing sufficient corroboration for attribution where the records happened around the same time.

[101.1230] Reporting and testimony

Final reports aim to provide conclusions to the primary investigative questions in a manner that allows the non-technical reader to grasp the conclusions, while providing sufficient detail in support of the following principles:

- Transparent – Are the analysis and interpretations capable of being independently verified in their entirety?[1]
- Substantiated – Conclusions should be written with a thorough explanation of the supporting evidence and reasoning[2], and explicitly declare any assumptions relied upon.
- Relevant – Conclusions and supporting facts should omit records or information which are irrelevant or misleading
- Unambiguous – Any facts or evidence relied upon should be precisely identified.

All evidence, materials and assumptions relied upon, tests undertaken, and evidence found should be identified, where relevant.

Common faults with forensic reporting include failing to adequately describe where evidence is found, failure to provide substantiation of opinions, presentation of opinion as fact, production of information subject to interpretation as fact, failure to declare assumptions, and omission of information which would otherwise give a different interpretation.

[1] Casey E, *Digital evidence and computer crime*, 3rd ed (Elseveir, 2011), p 219.

[2] Casey E, *Digital evidence and computer crime*, 3rd ed (Elseveir, 2011), p 219.

[The next text page is 101-1801]

DIGITAL EVIDENCE SUB-DISCIPLINES

[101.1400] Introduction

Various sub-disciplines have emerged to address the markedly different families of digital devices and their related environments. This section identifies and describes those primary sub-disciplines.

[101.1410] Storage forensics

The most mature and pervasive of digital evidence sub-specialisations is that which considers data at rest on storage devices. In the early days of digital forensics, this typically meant floppy disks and hard drives. Now it includes devices ranging from USB thumb drives to storage composed of multiple disk drives.

The methods and techniques of storing files and folders on storage media have changed at a slower rate than other aspects of the field, primarily due to the purpose of such storage – the long-term storage of information.

Storage forensics primarily focuses on the content and structure of operator perceivable artefacts such as files and folders, the substructure which is used by the system to manage those artefacts (the file system), and artefacts of significance which have emerged as by-products of the design of the system (deleted files and file slack).

[101.1420] Small scale digital device forensics

With the transition of mobile phones from voice communications devices to miniature computing environments, encompassing such functionality as text messaging, photography, and email, evidence from such devices is often relevant. These devices, along with other small, task specific digital devices such as PDAs, GPS devices, digital cameras and music players, form a class of devices collectively referred to as "small scale digital devices".

These devices were not designed with interoperability in mind, and accordingly, a wide variation occurred in their architectural and structural underpinnings. A consequence of this variation was that the techniques for merely extracting text messages from a mobile phone varied not only between mobile phones of different manufacturers, but additionally between different models of mobile phones from the one manufacturer.

In 2017, much of this diversity has disappeared, replaced largely with a dual monoculture of devices manufactured by Apple and a wide range of phones running the Android operating system. While examining the data stored in these devices has become far less of a challenge due to the pervasive adoption of a common database format for storing information, gaining access to that information has become the primary challenge. This is due to the use of security measures which effectively prevent unauthorised access to information. An example of their effectiveness was demonstrated by the inability of the US Federal Bureau of Investigation to extract information from an iPhone involved in the 2016 San Bernardino shootings.

This situation has resulted in this subfield adopting practices and techniques which significantly depart from traditional principles adopted in the storage forensics field. For example, many evidence extraction techniques fail to meet the completeness principle by failing to extract deleted data, and the principle of maintaining evidence immutability is violated by overwriting portions of the data while acquiring a larger copy of the system.

This “weakening” of principles is based on pragmatism and is reflective of a shift in the wider digital forensic discipline to acknowledge completeness and immutability as ideals rather than requirements.

There is now significant variability in the depth of evidence recovered and presentation of evidence between forensic tools. This highlights the pivotal role of the expert in identifying the limitations of techniques and tools employed and interpreting the significance of evidence in light of those limitations.

[101.1430] Network and cloud forensics

Storage forensics and small scale device forensics both generally consider data at rest. Network forensics considers data in motion, providing the capability to intercept, capture and interpret the communications between computers or digital devices. Network forensics is a niche field not widely practised in the civil space both due to criminal provisions related to the private interception of telecommunications and due to a lack of general means to intercept the communications of arbitrary computers. Network forensics is more widely practised in the policing sphere, where telecommunications interception powers are used to compel internet service providers (ISPs) to intercept communications.

The pervasive connecting of computers (and increasingly small scale devices) to the internet has resulted in significant changes in the way that data is stored. Whereas traditionally one would hold on to their data on a storage medium such as a hard drive in a computer or on a floppy disk, increasingly data is stored on third party services via the internet. Examples of such services (currently referred to as “cloud services”) include webmail (such as “Hotmail”) and social networking (such as “Facebook”).

Much information is now stored “in the cloud”. Many corporates have shifted their email mailboxes and servers into the cloud, and non-commercial email users regularly do not possess complete copies of their email mailboxes but rather only have recent portions residing on their mobile phone devices.

Challenges with this shift include identifying, accessing and interpreting the actual physical storage location of such data, along with distance, authority and jurisdictional concerns. For example, accessing email held in webmail services presents significant challenges when those services are held beyond the resident jurisdiction of a “user of interest to investigative authorities”.

[101.1440] “Internet of things” forensics

The “internet of things” is a buzzword for discussing the embedding of computers into a wide range of otherwise common non-computing devices, such as cars, light bulbs and watches. This is an emerging area of forensics, the most common and relevant being evidence sourced from car entertainment systems. These systems regularly record the location of the car at regular time intervals, and events such as the opening and closing of doors.

[101.1450] Volatile memory forensics

Volatile memory forensics considers the information content of the random access memory (RAM) of running computers. Somewhat akin to the short-term memory of humans, the RAM of computers typically keeps the most frequently used information close to hand, and is unique in that it quickly (ie within minutes) dissipates to an unrecoverable state once a computer is switched off. The content of RAM is increasingly of forensic interest, for example:

- obtaining keys and passwords not otherwise available;
- finding remnants of transient activities such as online chat; and
- identifying whether a computer was infected by a virus or other malicious software (valuable in countering "the Trojan defence").

[101.1460] Software forensics

In the Australian case of *Computer Edge Pty Ltd v Apple Computer Inc*[1], Gibbs CJ described a computer program (another term for software) as:

> … **a set of instructions designed to cause a computer to perform a particular function or to produce a particular result**. The instructions are … expressed in a computer language – either in a source code (which is not far removed from ordinary language, and is hence called a high level language) or in an assembly code (a low level language, which is further removed from ordinary language than a source code), or successively in both.
>
> The source code or assembly code cannot be used directly in the computer, and must be converted into an object code, which is "machine readable", ie which can be directly used in the computer. The conversion is effected by a computer, itself properly programmed.

Software forensics is the analysis of software artefacts, including source code and object code, towards issues including authorship attribution, provenance, and the behaviour of computers, in matters related to copyright violation, IP theft, fraud and ascribing liability due to software failures.

[1] (1986) 161 CLR 171 at 178-179.

[101.1470] Malware forensics

Malware, a truncation of the term "malicious software", refers to a particular class of software whose operation is to the detriment of a computer's operator. Categories of malware include computer viruses, worms, Trojans and spyware to name a few. Malware forensics focuses analysing evidence found on computers compromised by such software, with goals such as attribution of the software's author and identification of the effects of the malware.

[The next text page is 101-2001]

CHALLENGES TO DIGITAL EVIDENCE

[101.1600] Introduction

Any conclusion which has its origins in digital evidence may be subject to failure, if not properly founded in fact and scientific testing. False positives are conclusions based on hypotheses which should have been refuted but were not. False negatives are conclusions which are contrary to those that should have been reached but were not. Such failures may be caused by faults at any stage of the forensic investigation.

Faults are the making or missing of evidentiary sources, content, relationships, context, timing, ordering, location, consistencies and inconsistencies, and can occur at any stage of the investigative process[1]. Faults do not always result in failures; however, where faults are relevant to the operant hypotheses in a matter, the effects of such faults must be considered to ensure failures do not result.

[1] Cohen (2009)

[101.1610] Miss faults

Consider a fault such as follows. In the field, a copy (forensic image) of a computer's storage device is acquired and re-read to verify that the copy is a perfect fidelity copy. On returning to the laboratory, as a part of standard preservation and archiving activities, the examiner attempts to copy the forensic image to separate media. A small portion of the copy operation fails due to a hitherto unexperienced problem with the storage media on which the forensic image is stored.

The fault here is that a particular piece of content has been missed. Such a fault obviously gives rise to the potential of failure due to missing exculpatory evidence; however, it does not give rise to failures if the matters at hand are only relevant to the information stored in the intact portion of the forensic image. The claim of failure due to missing exculpatory evidence could be countered as irrelevant where:

- the missed content was in a location not accessible to the original computer; or
- the apparent type of the file or file(s) containing the missed content was not relevant to the claims.

Other examples of miss faults include:

- not identifying and collecting relevant digital devices and evidence sources;
- missed content due to poor collection methods;
- search results missing information where it is stored in encoding formats which are unable to be interpreted;

- content missing from keyword searches due to font formatting or encoding related interpretation errors;
- overly broad filtering criteria;
- overly specific search terms;
- inaccessible content due to encryption;
- natural degradation of evidence media;
- production of imperfect reproductions of evidence, for example production of the visible content of an email without the metadata which describes its path and provenance;
- conclusions that no virus or malware exists based on a single antivirus test.

[101.1620] Make faults

All evidence presented is an interpretation, and dependent on the correct operation of complex systems of computer software and hardware. Faults in such interpretation can arise from malfunctioning computer hardware and software, incorrect configuration and implementation errors.

One of the most common "make faults" occurs in relation to the interpretation of time and date stamps associated with computer records. Time and date stamps are routinely misinterpreted due to misunderstandings about their meanings, the complex and inconsistent set of rules generally governing their production by software, changes to those rules between updates of software versions, inaccurate clocks within computers, and idiosyncratic document production conventions of computer operators.

For example, in matters where possession is at issue, the practice of producing summaries of files which include a "last accessed time" should, on Windows-based computers, and in absence of interpretation, be avoided. The last accessed time is generally unrelated to access involving a human; rather, it often represents the last time the computer was scanned for a virus infection, or the last time a file was updated.

Interpreting trace evidence is fraught with unknown assumptions. Consider the following true example. During the course of a teaching day, a number of pornographic images "popped up" onto the screen of a computer in partial view of a full class of students. The prosecution prevailed in their contention that the defendant, a substitute teacher, was intentionally browsing for pornography while the class was busy at work. This conclusion was largely based on evidence of the prosecution expert, who claimed that a link on a webpage being the colour red meant that the link had been clicked on. An independent analysis of the evidence indicated that the link was red in colour because it was designed that way[1], and hence that there was no evidence to support the contention that the link had been clicked on.

Other examples of make faults include:

- implementation errors related to interpreting time zones and daylight savings;
- assumptions related to time zone and location;
- contamination with evidence from prior cases, for example from poor evidence management practices;
- misinterpretation of the meaning of evidence;
- document forgeries;

- assumptions that software behaves as it is documented.

[1] http://www.sunbelt-software.com/ihs/alex/julieamerosummary.pdf.

[The next text page is 101-2201]

VALIDATION, ERRORS AND RELIABILITY

[101.1800] Introduction

The fundamental concern associated with expert evidence of a scientific or technical nature is whether the underlying theory and methods are reliable. In the field of digital forensics, reliability should be evaluated by considering multiple levels:

- Algorithm/technique: Is it scientifically valid, repeatable, reproducible?
- Implementation: Does the software tool correctly implement the algorithm or technique?
- Application: Was the tool used correctly?
- Interpretation: Was the result understood and communicated correctly by the examiner?

This section identifies a variety of factors relevant to assessing evidential reliability and admissibility in the digital evidence field, by outlining the current approaches assuring techniques are scientifically valid and the risk of errors is mitigated.

[101.1810] Examples of tool errors

A natural consequence of the latent nature of digital evidence is that all digital evidence must, in the overwhelming majority of instances, be interpreted by an automated process (ie by software running on a computer) to be perceived by the expert or finder of fact. Such processes, and the theories of operation on which they are based, must be accurate and free of faults to the extent that faults produced are not of direct relevance to the claims.

The theories of operation embodied in forensic tools are regularly based on incomplete understandings of the bit sequences they interpret. Moreover, the extent and effect of implementation errors and malfunction within these forensic tools is an unknown. It is not uncommon for forensic tools to fail, in both readily apparent ways (such as when a tool "freezes" and becomes unresponsive) and in ways which are not readily apparent (such as when a tool misreports a count of items).

The reported errors in forensic tools include:

- subtle differences in the appearance of documents;

- emails unable to be read;
- search not indexing the full text of documents;
- times and dates being off by plus or minus 13 hours for link files;
- counts of visits to websites being off by significant amounts[1].

[1] *State of Florida v Casey Marie Anthony* 2008-CF-15606-A-O.

[101.1820] Error types in digital forensics

While the methods and tools of the field are generally considered by practitioners to be reliable, the question has arisen as to the adoption of methods other forensic disciplines use to report on the reliability of techniques[1]. Fields such as DNA analysis use statistical approaches to describe their reliability in terms of error rate, so why doesn't digital forensics?

The reason lies in the types of error being considered. Experimental science typically distinguishes between two types of observational errors: random errors and systematic errors. Random errors in measurements are variations in measurement due to unpredictable and unknown changes in the subject of an experiment, while systematic errors are variations in measurement due to faults in the measuring tool or by the operator of the tool. In fields such as DNA analysis, the principal type of potential error is the random error. Both types of errors are possible in the context of digital forensics, with systematic errors being the most relevant.

In some instances, the techniques used in digital forensics can be characterised as random errors, and described in terms of an error rate. In [101.840] the usage of cryptographic hashes for authenticating forensic images was introduced, and the error rates associated with using the technique described. The error rates cited are statistical properties of the underlying algorithm, and based on information and complexity theory.

Systematic errors are possible both in the implementation of forensic techniques (typically as software based tools), in the application of the tool, and in the interpretation of the results produced by the tool. In implementations, systematic errors are caused by logic errors and incomplete assumptions (bugs) that are reliably triggered by particular circumstances in the data that they read. An example of such a systematic error might be an implementation of a crytographic hash algorithm in a particular tool, whereby a bug causes it to produce the same hash value regardless of the information that is fed into it. The output of such a broken implementation is perfectly predictable and measurable and is not random.

Current approaches to addressing systematic errors in digital forensics apply methods drawn from software engineering regarding testing and validation of software (tool validation) at the implementation layer, and training and quality management at the tool usage and interpretation layer[2].

[1] Lyle (2017)

[2] SWGDE (2017)

[101.1830] Tool validation

Given the necessity of reliable tools for interpreting digital evidence, identifying methods to assure the reliability of such tools has long been on the agenda. The US National Institute of Science and Technology (NIST) has been undertaking work in this area since 2002, leading to a modest amount of testing having occurred in regard to:

- devices which prevent modification to digital evidence (write blockers);

- applications for copying the contents of hard disks (imaging applications);
- string searching;
- deleted file recovery;
- file carving;
- mobile device acquisition.

In practice, the current approaches to addressing the issue of potential tool error are to identify corroborating evidence via separate techniques and to validate interpretations made by one tool with the interpretations of another independently constructed tool. Where an interpretation is central to a forensic finding, it is good practice to manually examine at the lowest levels the data being interpreted, as such an examination may reveal additional relevant information or information which has been misinterpreted by tools.

[101.1840] Scientific controls and standards

Scientific controls and standards play a significant part in testing within the physical forensic sciences. In the context of testing, a standard is "a prepared sample that has known properties that is used as a control during forensic analyses"[1]. In the physical forensic sciences, a scientific standard is a benchmark against which measurements are made. For example, a carefully maintained object of a known and precise weight is used to calibrate a scale, or a sealed glass vial of a known composition is used to compare with a sample of unknown composition. A control is "a test performed in parallel with experimental samples that is designed to demonstrate that a procedure is working correctly and the results are valid."

While the utility of utilising prepared scientific standards as a basis on which to base testing of the functional performance of techniques and tools has been recognised for some time, there currently exist no commercial digital forensic science standards for this purpose. The principal efforts which have been undertaken towards this goal include the Digital Forensics Tool Testing project[2], the Standardised Forensic Corpora[3], and test standards focused on deleted file recovery produced by the National Institute of Science and Technology (NIST). The methods adopted for testing within the digital forensic tool manufacturers are in general not currently in the public record.

The role of scientific controls in digital forensics is currently an area of some confusion within the field. The inclusion of controls within the Daubert test, ASCLD-LAB and ISO/IEC 17025 laboratory quality management standards have led to attempts by practitioners to attempt to address scientific controls in their practice; however, there exists little in the literature or body of knowledge in the field that positively outlines where controls have any place in the field. In the negative, the Scientific Working Group on Digital Evidence (SWGDE) stated in 2008: "… controls are not applicable in the computer forensics sub-discipline"[4], a recommendation that has yet to be refuted in the field's literature.

[1] Barbara (2007)

[2] http://www.dftt.sourceforge.net (accessed 13 October 2017).

[3] Garfinkel et al (2009)

[4] SWGDE (2008)

[101.1850] Consensus of experts

The digital ecosystem, and the trace evidence left behind as digital evidence, changes so quickly that achieving general acceptance on techniques and theories of operation is possible in only the most mature and mainstream of areas. The complexity and rapid development of digital evidence similarly mean that "cutting edge" analytic techniques are routinely employed, despite never having achieved general acceptance. Finally, the significant variation in technical understanding, skills, and education within the field of practitioners makes consensus a difficult undertaking in even the simplest areas.

The areas of greatest consensus are in the abstract and general, with the concepts of transparency (freely testable by a third party) and repeatability (successive tests should yield the same result) commonly cited. Despite the consensus that these are desirable attributes, the field is only in the early days of translating these fundamental scientific principles into practice. For example, an expert may claim that a file was deleted at a certain time based on the output of a storage forensics tool, without documenting or understanding how she, or the tool, came to that conclusion. Or, for example, it is not uncommon for one tool to count a different number of items found during a search than another tool.

While not a widespread problem, finding consensus amongst the opinions of experts does yield inconsistent outcomes. For example, the Scientific Working Group on Digital Evidence (SWGDE) stated in their 2008 position paper on standards and controls in computer forensics that:

> In computer forensics, however, false positives are non-existent. If the forensic hardware or software used fails for any reason, the examination will not produce erroneous data. The tools and processes might fail to find existing data, producing a false negative, but they will never find non-existent data.[1]

This is not correct, with counter examples including misrepresentation of timestamps and carved files.

Achieving consensus is problematic in more concrete and mature areas of practice. For example, an empirical study of good practice in acquisition techniques in storage forensics (the most mature area of the discipline) identified significant conflict amongst a panel of experts[2].

[1] ibid

[2] Carlton et al (2009)

[101.1860] Peer review and publication

Publication of research in the area occurs primarily through conferences, journals, and online via blogs. Academic peer review occurs in a range of journals and conferences, the focus of which ranges from broad, focusing on computer security or forensic science, to specific, focusing exclusively on digital forensics. The following venues are recognised as the leading academic forums focused specifically on digital forensics:

- The Journal of Digital Investigation (est 2004);
- The DFRWS Conference[1] (est 2001);
- International Federation of Information Processing (IFIP) Working Group 11.9 Conference[2] (est 2005).

[1] http://www.dfrws.org (accessed 13 October 2017).

[2] http://www.ifip119.org (accessed 13 October 2017).

[101.1870] Refutability

Within the scientific method a hypothesis or assertion is only considered "scientific" if it is falsifiable – hypotheses are generated and then tested by experimentation or observation to determine if they are false. Even a single refutation is considered to disprove the hypothesis.

Despite early calls to turn the discipline into a forensic science[1], it has been only in recent times that the field has begun to identify how to apply the scientific method within the context of existing methodology and techniques[2][3][4].

[1] Palmer (2001)

[2] Carrier (2006)

[3] Casey (2011)

[4] Cohen (2009)

[101.1880] Addressing criticisms of the wider forensic disciplines

The value and reliability of forensic evidence in general is increasingly being questioned, with the most recent review of significance being the President's Council of Advisors on Science and Technology (PCAST) report[1], the focus of which was the scientific validity of feature-comparison methods. The report:

- identified important gaps in clarity regarding scientific standards for assessing validity and reliability of forensic methods;
- identified the need to evaluate whether the validity of a range of feature comparison methods can be scientifically established;
- found that for assertions about the probative value of matches to be scientifically valid, the underlying methods needed to have established false positive rates.

PCAST sought and received feedback from leading communities within the field of digital forensics in the effort underlying the report. It is important to note that the feedback sought was focused on "feature-comparison" methods, which form only a small proportion of the range of techniques used in finding and analysing digital evidence.

[1] PCAST (2016)

[The next text page is 101-2401]

ASSESSING CLAIMED PROFESSIONAL COMPETENCY

[101.2000] Introduction

Due to the short history and rapid development of the field of digital forensics, it is difficult for the finder of fact or solicitor to evaluate claimed formal and practical competency or expertise in the area. Unlike professions such as accountancy and the wider forensic sciences, mandated competency standards, defined either by governmental bodies or recognised professional associations, do not exist, nor are likely to emerge in the short term. In attempting to assess competence and expertise, aside from word-of-mouth testimonials and the public record, one is left with experience, and a wide range of qualifications, accreditations, certifications and post-nominals of uncertain merit.

This section provides background on a number of aspects relevant to assessing the credentials and practices of practitioners.

[101.2010] Experience

The experience and accumulated knowledge which is accrued through practice can be a significant indicator of expertise.

The pioneers of the field of digital forensics in the 1980s and 1990s included police officers with no credentials in computing other than an interest in tinkering with computers. A number of these pioneers continue to practise digital forensics today in the private and public sector, and academia. Some remain without formal qualifications, while a number have pursued formal qualifications in computer science and information technology.

[101.2020] Core body of knowledge

The foundational body of knowledge relevant to digital evidence practice is that of "information technology" ("IT"), of which "information systems" and "computer science" are subspecialities. Graduate level degrees in the fields of IT are widely available, and while the fields are fast moving, there exists significant agreement on the bodies of knowledge at their common foundations.

The body of knowledge underlying the field of digital forensics is a distinct specialisation which depends and builds upon the formerly mentioned body of knowledge related to IT and computing. There is no consensus on the body of knowledge for digital forensics.

[101.2030] Qualifications

In the wider forensic sciences, the general way to be recognised as competent is to demonstrate a significant external validation of the skill level. This is typically founded on formal education in the relevant discipline, with the bachelor's degree being a minimum. Subsequent specialisation within that discipline through study and validated research is then typically demonstrated before recognition of expertise[1].

Graduate level subjects in digital forensics have been offered in Australia since at least 2003, primarily within computer science and information technology degrees within computer security and information security specialisations. Two Australian universities, Edith Cowan University and the University of Western Sydney, began offering bachelor's level degrees with computer forensic specialisations in 2005. In 2008 educational institutions offering courses specifically in digital forensics ranged from TAFE advanced diplomas up to masters by coursework. To the author's knowledge, masters by research and doctoral degrees have been offered since 2003.

Identifying the relevance of such qualifications to competence in the field holds the following challenges. While qualifications from universities may assure that a candidate has undertaken a course of study or research, and that the course has undergone internal and independent review, the extent to which the course syllabus is relevant to the forensic expertise required may be difficult to ascertain. Furthermore, in a fast moving field, the knowledge and skills related to such qualifications quickly become stale without practice and ongoing professional education.

[1] Jones & Valli (2009)

[101.2040] Certifications

The most contentious area of credentialing within the field is in the area of certifications. Bodies offering certifications include digital evidence tool providers, for-profit companies offering (or closely allied with companies offering) training or licensed training materials, professional organisations and other not-for-profit entities. Table 4 provides a summary of the most commonly produced credentials, the granting body, and a categorisation.

TABLE 4 Common certifications related to digital forensics

Post-nominal	Certification	Granting body	Category
ACE	AccessData Certified Examiner	AccessData	Tool vendor
CCCI	Certified Computer Crime Investigator	High Tech Crime Network HTCN	For profit
CCFT	Certified Computer Forensic Technician		
CCE	Certified Computer Examiner	The International Society of Forensic Computer Examiners	For profit / training

Post-nominal	Certification	Granting body	Category
CCFE	Certified Computer Forensics Examiner	Information Assurance Certification Review Board	Not for profit
CCFP	Certified Cyber Forensics Professional	International Information System Security Certification Consortium	For profit / training
CFCE	Certified Forensic Computer Examiner	International Association of Computer Investigative Specialists (IACIS)	Professional organisation / Law enforcement only
CHFI	Certified Hacking Forensic Investigator	International Council of Electronic Commerce Consultants	For profit / training
DFCP, DFCA	Digital Forensic Certified Practitioner, Digital Forensic Certified Associate	Digital Forensics Certification Board	Not for profit
EnCE	Encase Certified Examiner (EnCE)	Guidance Software	Tool vendor
GCFA	GIAC Certified Forensic Analyst	Global Information Assurance Certification	For profit / training

Assessing the relationship between the possession of a particular certification and a practitioner's capacity to competently perform digital evidence related tasks is difficult for a number of reasons:

- Few of the certifying bodies provide the transparency to identify the body of knowledge underlying their certification, with training typically only measured in weeks.
- Few of the certifying bodies provide the transparency to determine the criteria by which they judge competency in the field.
- Tool vendor provided certifications typically focus on the usage of the particular vendor's software tool, and thus address only the small proportion of the body of knowledge which the tool is designed to assist in addressing.

- Experience requirements for attaining certifications are variable, with uncertain standards of background checking of experience claims.
- Few certifications have requirements of meeting any code of ethics or standards of professional conduct.

[101.2050] Accreditation

In the laboratories of the wider forensic sciences, accreditation has been seen as a means by which laboratories may demonstrate that their quality management systems and procedures adhere to an established set of standards. Since the early 1980s, the United States based American Society of Crime Laboratory Directors (ASCLD), Laboratory Accreditation Board (LAB) has offered accreditation of crime labs. In 2003, ASCLD/LAB, with the assistance of the Scientific Working Group on Digital Evidence (SWGDE), for the first time approved digital evidence as a part of its accreditation for crime laboratories. Since 2004 ASCLD/LAB has additionally offered an accreditation based on the ISO international standard: *General requirements for the competence of testing and calibration laboratories* (ISO/IEC 17025: 2005). The stated objective of the standard is that it:

> specifies the general requirements for the competence to carry out tests and/or calibrations, including sampling. It covers testing and calibration performed using standard methods, non-standard methods, and laboratory-developed methods.

In the United Kingdom, laboratory managers were encouraged by the Association of Chief Police Officers (ACPO) to attain the more general ISO international standard: *Quality management systems* (ISO/IEC 9001) in the medium term, with attainment of ISO/IEC 17025 indicated as a potential long-term goal[1]. The UK Forensic Science Regulator is mandating accreditation requirements for all providers of forensic science services, with digital forensic service providers required to be accredited to ISO/IEC 17025 by October 2017.

In Australia, the National Association of Testing Authorities (NATA) offers accreditation to laboratories based on ISO/IEC 17025 (see [26.420]), supplemented by an additional set of requirements[2]. The Australian Federal Police have adopted this framework as an accreditation standard.

The goals of such quality management frameworks: quality control, repeatable procedures, monitoring of activities, response to deviations in practice and external assessment, are all relevant to attaining and maintaining quality within digital evidence related investigations. The effect of such frameworks within the context of digital forensic evidence matters is, however, of questionable merit. Such quality management systems add most value in acquisition and evidence management activities, where loss or contamination of evidence is best prevented.

The field has yet to achieve consensus around adopting ISO/IEC 17025. The mismatch between the physical sciences based bias of the frameworks and the unique nature of digital evidence leaves the measurement oriented focus of the ISO/IEC 17050 standard not directly relevant to the majority of activities undertaken in relation to examination and analysis of evidence. The financial and personnel overheads associated with administering such a system carry with them a substantial burden[3].

[1] ACPO (2009)

[2] NATA (2008)

[3] SWGDE (2009)

[101.2060] Standards and principles of good practice

There currently exists no digital evidence specific standard by which one may benchmark the activities of a practitioner. This is due to the fast moving and evolving nature of the field; any standard would quickly become stale, necessitating deviation from "standard" or "best" practices. Such deviations would inevitably invite criticism, for little benefit. Consequently the field has adopted a principles and guidelines based approach, identifying elements of good practice which lead to flexibility for the practitioner in adapting to new challenges.

A number of guides published by governmental and standards bodies are of relevance to good practice in the area of digital forensic evidence. These include:

- *ISO/IEC 27037:2012 – Information technology – Security techniques – Guidelines for identification, collection, acquisition, and preservation of digital evidence*;
- *ISO/IEC 27042:2015 – Information technology – Security techniques – Guidelines for the analysis and interpretation of digital evidence*;
- *Searching and Seizing Computers and Obtaining Electronic Evidence in Criminal Investigations* (2009) US Department of Justice;
- *Good practice guide for Digital Evidence* (2012) UK Association of Chief Police Officers;
- *Forensic Examination of Digital Evidence – A guide for law enforcement* (2004) US National Institute of Justice;
- *Digital Evidence: Standards & Principles* (2000) Scientific Working Group on Digital Evidence (SWGDE) and International Organisation on Digital Evidence (IOCE); and
- *HB-171-2003 Guidelines for the management of IT Evidence* (2003) Standards Australia International.

While many of these guides have been subjected to peer-review by practitioners, the recommendations and practices contained within them are by no means considered definitive. There is limited consensus as to what constitutes standard or even good practice[1].

[1] Carlton & Worthley (2009)

[101.2070] Internal structure

Within the profession, the following categories are emerging[1] for discriminating between the capabilities of digital evidence practitioners:

- **Digital evidence technicians**: possess the skill level to perform basic laboratory tasks related to digital evidence to a defined standard under the supervision of a suitably qualified manager. Such assistants have a skill level for acquiring data from storage devices using defined processes.
- **Digital evidence examiners**: possess a competence in the body of knowledge of digital forensics and information technology. Such technicians have a skill level for undertaking examination of digital evidence for the purpose of providing technical evidence (for example finding files and fragments of information) but without giving interpretation (expert opinion).
- **Digital evidence experts:** in addition to the capabilities above, possess capabilities for analysing and interpreting of evidence for the purpose of giving opinion oriented expert evidence.

[1] UK Council for the Registration of Forensic Practitioners (now defunct) http://www.computerevidence.co.uk/Papers/ComputersandLaw/RegisteredForensicPractitioner.htm (accessed 13 October 2017).

[The next text page is 101-2601]

FUTURE

[101.2200] General

Identifying, acquiring and preserving evidence from online sources will become increasingly challenging, due to cross jurisdictional issues and a shift to storing information online rather than locally. Encrypted and locked devices will increasingly present challenges to analysis in gaining access to the data stored within.

In the policing sphere, case backlogs are a persistent problem and will remain so. With certain classes of crime becoming commonplace in digital evidence units (for example online child exploitation) and the analytic tasks related to those crimes being well understood, some degree of automation in the forensic evidence management and analysis process will assist in relieving much of the drudgery currently involved in investigating these classes of cases. A step towards this direction is the ongoing establishment of the child exploitation hash databases such as Project VIC, which, through sharing of hash "fingerprints" of known child exploitation material, enable the rapid identification of such material in suspect evidence, eliminating much of the manual visual classification effort required.

So far the courts have taken an optimistic view of the reliability of computers and the information contained within. Such a stance is understandable given that in the early days of computing, the actions of computers were generally deterministic, based on the information given to the computer, and the operations which the computer was configured to perform. However, computers, and the ecosystem in which they exist, are now sufficiently complex that the presumption of reliability in receiving and applying expert opinions will face challenges. While information does not randomly blink into existence, it is commonplace for faults to occur due to the complex and unobserved interactions between IT components. More problematic is the commonplace occurrence of computer break-ins and infections by malicious software (which in the majority of cases acts as an agent of a human actor).

This presumption of reliability will come under increasing pressure, specifically in regard to the reliability and meaning of interpretations generated by forensic tools, given their often incomplete and opaque models of operation. The commerciality of the field currently discourages the open dialog required to bring scientific rigour to these interpretations.

A combination of our now critical reliance on digital information infrastructures, the pervasive nature of digital devices in the personal and work spheres, and the increasing rate of online and computer related crime, is leading to an increasing demand for digital forensic investigations performed by competent, credible experts. The future will see the establishment within Australia of a credible professional credentialing authority with transparent and inclusive approaches to assessing practitioner capabilities and background.

[The next text page is 101-2801]

BIBLIOGRAPHY

[101.2300] Bibliography

Whitcomb C, *The Future of Professionalism in Digital Forensics* (2008), http://www.dfcb.org/New_DFCB_Site/docs/2008_AAFS_FNL_The_Future_of_Professionalism_in_Digital_Forensicscmw.pdf (accessed 13 October 2017).

Sammes T and Jenkinson B, *Forensic computing: a practitioner's guide* (Springer, London, 2000).

Kruse WG and Heiser JG, *Computer forensics: incident response essentials* (Addison-Wesley, Boston, 2001).

McKemmish R, "What is forensic computing?" (1999) *Trends and issues in Crime and Criminal Justice*.

SWGDE, *SWGDE and SWGIT Glossary of Terms* (2015), https://www.swgit.org/pdf/SWGDE%20and%20SWGIT%20Digital%20and%20Multimedia%20Evidence%20Glossary?docID=60 (accessed 13 October 2017).

Cohen F, *Digital forensic evidence examination* (Fred Cohen & Associates, 2009).

ACPO, *Good Practice Guide for Digital Evidence* (The Association of Chief Police Officers of England, Wales and Northern Ireland, 2012).

Stevens M, Lenstra A and Weger B, *Chosen-prefix Collisions for MD5 and Applications* (2009) Cryptology and Information Security Group (PNA5), Centrum Wiskunde & Informatica, https://homepages.cwi.nl/~stevens/papers/stJOC-SLdW.pdf (accessed 13 October 2017).

Thompson E, *MD5 Collisions and the impact on computer forensics*, 1, (2005) Digital investigation, Vol 2.

Jansen W and Ayres R, *Guidelines on Cell Phone Forensics*, s l (NIST, 2007).

SWGDE, *SWGDE Position on the National Research Council Report to Congress* (2009).

Boyd C and Forster P, *Time and date issues in forensic computing – a case study* (2004) Digital investigation, Vol 1, pp 18-23.

Wood, Justice, *Forensic Sciences from the Judicial Perspective* (16th International Symposium on Forensic Sciences, Canberra, 2002).

SWGDE, *Standards and Controls* (2008).

Carlton G and Worthley R, *An evaluation of agreement and conflict among computer forensics experts* (HICCS Conference, 2009).

Palmer G (ed), *A road map for digital forensics research* (Digital Forensics Research Workshop, 2001).

Carrier B, *A Hypothesis-based Approach to Digital Forensic Investigations* (Purdue University, West Lafayette, 2006).

Casey E, *Digital evidence and computer crime*, 3rd ed, s l (Elseveir, 2011).

Barbara J, *Computer Forensics Standards and Controls* (2007), https://www.forensicmag.com/article/2008/01/computer-forensics-standards-and-controls (accessed 13 October 2017).

Garfinkel S, Farrell P, Roussev V and Dinolt G, *Bringing Science to Digital Forensics with Standardized Forensic Corpora* (Digital forensics workshop, 2009).

Jones A and Valli C, *Building a Digital Forensic Laboratory*, s l (Elseveir, 2009).

ACPO, *ACPO Managers Guide Good Practice and Advice Guide for Managers of e-Crime Investigations*, s l (The Association of Chief Police Officers of England, Wales and Northern Ireland, 2009).

NATA, *Technical Circular 9*, s l (National Association of Testing Authority (NATA), 2008).

SWGDE, *SWGDE Position on the National Research Council Report to Congress* (2009).

Tichy W, *Should computer scientists experiment more?* (1998) Computer (IEEE Computer Society), Vol 31, 5.

Denning P, "Is Computer Science Science?" (2005) Communications of the ACM, Vol 48, 4.

Lee H, *Henry Lee's Crime Scene Handbook* (Elseveir, 2001).

PCAST, "Report to the President" (2016) *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods.*

SWGDE, *SWGDE Establishing Confidence in Digital Forensic Results by Error Mitigation Analysis* (2017), https://www.swgde.org/documents/Current%20Documents/SWGDE%20Establishing%20Confidence%20in%20Digital%20Forensic%20Results%20by%20Error%20Mitigation%20Analysis (accessed 13 October 2017).

Lyle J, *Defining, Measuring, and Mitigating Errors for Digital Forensic Tools* (NIST, 2017), https://www.cftt.nist.gov/presentations/aafs-2016-lyle.pptx (accessed 13 October 2017).

[101.2310] Further reading

Casey E, *Digital evidence and computer crime*, 3rd ed (Elsevier, 2011).

Cohen F, *Challenges to digital forensic evidence* (Fred Cohen & Associates, 2008).

[The next text page is 110-1]