Chapter 10 Opportunities and challenges for the future

Book Section

Chapter 10 Opportunities and challenges for the future

Robin Bryant, Ed Day and Ian Kennedy

Introduction

In this chapter we look at some of the opportunities and challenges for the policing of digital crime in the future. This includes developments in digital criminal investigation, (digital forensics and anti-forensics) and some of the problems with the use of expert witnesses in court. The challenges and opportunities often reside in the same phenomenon, for example cloud computing offers the opportunity for distributed approaches to dealing with large amounts of data from mobile phones (see below) but also provides challenges in terms of the recovery of evidence.

Developments in digital forensic investigation

In this section we discuss some of the recent developments in digital forensic investigation, with reference to both the opportunities offered and the challenges presented. Garfinkel (2010, p. 64) argues that we are now reaching the end of the ‘Golden Age of Digital Forensics’ as we encounter volumes and types of data that defy analysis with existing tools. It is certainly true that the processing of digital data using automated tools such as FTK or EnCase is a lengthy process. It can take hours or sometimes days to pre-process the various pieces of digital evidence for a single case, and this must be done before analysts even consider what aspects require further examination (Cantrell et al, 2012). The explosion in the sheer quantity of digital data and vastly increased device storage capacities have put increased pressure on digital investigators. However there are various ways the workload of digital forensics labs can be mitigated: outsourcing, triage, workload management and automation (Jones & Valli, 2008).

The rise in the amount of information to be processed in a digital investigation requires a concomitant increase in processing power, which can potentially be achieved either via faster or multiple computers, or a combination of the two. The use of multiple machines working on different aspects of a task at the same time is known as ‘parallel processing’, and this is often performed by using distributed resources that are situated and maintained remotely (distributed computing, using a ‘Grid’).

There has long been a call for distributing forensic digital processing (Richard & Roussev, 2004), but so called ‘first generation’ forensic tools such as EnCase and Forensic ToolKit were limited in terms of both automation of tasks and their ability to exploit parallel or distributed processing (Ayers, 2009). A ‘second generation’ of organisations and individuals began using such tools with increased computing power. For example AccessData in partnership with Dell claim to provide enhanced processing speeds through the use of a ‘data centre’ with ‘high performance servers’ to run FTK processing (Dell, 2009). Multiple core machines such as a Beowulf cluster (50-100 cores), and a supercomputer (4096 cores) have been used to reduce processing times (Ayers, 2009), but the disadvantage of such approaches is the need for expensive hardware.

Grid computing however provides an alternative means of providing the necessary increase in processing power. It reduces the need for new hardware by utilizing existing resources such as spare desktop capacity, and provides a far more powerful (parallel) computing resource than would normally be available to an individual investigator. In essence this type of computing links a number of resources together in a ‘grid’ and enables a user to run a program using vast but distributed resources that are remotely situated and maintained. Grids allow a number of nodes to process the data in parallel with potentially substantial time benefits (Voss, Vander Meer & Fergusson, 2009). For example the FBI uses an internal grid computing network, the ‘Grid Computing Initiative’ which utilizes unused capacity from their users’ PCs (FBI, 2010). One of the contributors to this book is currently involved in a research project to determine the viability of grid computing for mobile phone forensic investigation utilising the National e-Infrastructure Service (formerly the National Grid Service). The National e-Infrastructure Service (NES) is a computing grid which has already been successfully used for a number of research applications where massive computing power is needed, for example to model criminal behaviour of individuals within a city (Malleson et al, 2009).

Increasing the number of processors or machines to complete an investigative task reduces the time spent on each task but can be expensive, and is best used in conjunction with triaging. Digital investigation triaging is a heuristic (experience based) method which identifies a hierarchy of importance of digital tasks. Some work has been done on formalizing this process, for example Rogers et al, (2006), but they say very little about triaging other than reiterating it requires prioritization of tasks. Nor does the model identify how best to triage between cases, an important task for efficient caseload management for digital forensic labs. Instead focus is given to triaging within a case, for example on deciding whether the analysis of chat logs is worth the time and monetary expense for that case. ADF Solutions have recently developed software that claims to triage between cases (ADF, 2010).

Many aspects of the digital forensic investigation of PCs can also be automated using commercial tools such as X-ways' WinHex, Guidance Software's Encase, and AccessData's FTK. There are also free and open source tools which perform a similar function, although more of these are available for Linux than for Windows. Ideally such tools should be forensically sound – the Computer Forensic Tool Testing (CFTT) project at the United States National Institute of Standards and Technology (NIST) tests forensic tools using a standardised methodology. The UK lags far behind in tool evaluation (Sommer, 2010) and it remains to be seen how the anticipated granting of regulatory powers for the Forensic Science Regulator will affect this situation, if at all (Adetunji, 2011).

Solid-state Drives (SSDs)

Currently most PCs use a hard disk drive (HDD) for storage of data and in forensic terms HDDs are reasonably well-understood and documented. Increasingly PCs are being fitted with alternative solid-state drives (SSDs) instead of, or in addition to HDDs. The SSDs have a number of advantages over HDDs: they are smaller, quicker (increased 'latency'), quieter, more energy efficient and more robust (able to withstand sudden shocks). They are however more expensive, but as with much digital technology the price is steadily falling. From the outset netbooks were often fitted with SSDs. By default most tablets are now fitted with SSDs; for example the new Microsoft Surface Pro tablet contains a 64GB or 128GB SSD according to customer choice.

However**,** the increasing use of SSDs has significant implications for digital forensic investigation. Digital information is represented as binary digits as an aid to human understanding and manipulation, but the real data has to be stored in some physical manner. For example RAM consists of memory cells filled with capacitive charge, optical media such as CDs and DVDs contain lands and pits on a reflective surface, and magnetic hard drive storage involves small magnets oriented north or south (Bell & Boddington, 2010). The conceptual basis of the magnetic storage system used in HDDs has changed little since the 1950s (Hughes, 2002) and a drive's controller chip acts predictably: any processes it carries out do not generally interfere with the forensic investigation. The investigation of HDDs follows a well-established set of procedures that enable law enforcement to, amongst other things, extract data that a user had deleted. This is reasonably straightforward because HDDs merely mark deleted files as deleted, but the file content remains on the drive until the space is needed for new data (see the section on the forensic analysis of PCs in Chapter 8 for more detail).

Modern SSDs store data using Flash NAND memory chips which retain data on power-off (in contrast to RAM which requires an electric current to retain data). At the physical level the NAND within the SSDs consists of transistors (technically, 'floating gate MOSFETs', see OCZ, 2013) that are in one of two electronic conditions (even when power is removed) which in effect represent the binary states of 0 or 1 (we have simplified the description). At the logical level SSDs can be thought of as being divided into blocks, a typical block being about 512KB. Each block is divided into pages of about 4KB in size, it follows that each block contains approximately 128 pages. As data is written to the drive, it occupies a page. While each page can be read and written to, it is only possible to delete an entire block. Also, pages cannot be overwritten; they must be empty first to write data to them.

An SSD (unlike a USB pen drive, which also uses Flash NAND and is not considered here) has a controller chip that performs more complex processes than an HDD controller (Chen et al., 2009). The SSD controller runs an application known as the Flash Translation Layer (FTL) which acts as an intermediary between the operating system (OS) and the actual data storage within the blocks. To write data, an OS sends a request to the FTL via a standard hard drive interface (such as SATA) and the FTL translates this logical request and maps it to actual data on the SSD NAND chips (Bell & Boddington, 2010). The number of such writes to SSDs chips is limited, with chips failing after as few as 10,000 writes (Olson & Langlois, 2008). With modern SSDs, in order to avoid failure a process known as wear levelling is performed by the FTL (older SSDs behave somewhat differently and are not covered here). This involves directing writes to storage areas that have been written to less (rather than to where the OS requests), thus levelling out the number of writes per storage area. The FTL 'tells' the OS that it wrote to the correct place logically even though it may have physically written to a very different area from that stated by the original logical address (Bell & Boddington, 2010). If an OS deletes a file then tries to write new data to the logical location of the deleted file the SSD cannot overwrite the deleted file since it must erase data before re-writing. (Unlike HDDs, SSDs cannot overwrite existing data: they must first erase old data before writing new data to the same location). Therefore the FTL also performs a process known as 'Garbage Collection' which moves used data pages to other data blocks in preparation for future writes to that block. Garbage Collection occurs independently of the operating system – it is a feature of the internal firmware SSD – and occurs when the SSD is powered on but not otherwise busy reading or writing data.

An ATA command 'TRIM' was introduced (first available in Windows 7), that allows OSs to tell the SSD that the file is deleted thus reducing the amount of data moving required by the SSD (King & Vidas, 2011). Unfortunately (for the digital forensic investigator) when TRIM is enabled with an SSD and Garbage Collection occurs, the file becomes effectively irrecoverable (unlike with HDDs where deleted files are usually recoverable). This process of permanent deletion is sometimes known as 'SSD self-corrosion' (Bell & Boddington, 2010) or 'SSD self-contamination'. This has significant forensic implications as it means that as soon after the SSD is powered on there is the possibility that permanent data deletion will occur and the certainty that original data will be altered before an exact image can be made (and even if a write-blocker is used; Sheward, 2012). This would be a serious breach of most current digital forensic guidelines (for example, ACPO, 2012). Indeed Bednar and Katos (2012, p. 6), are clear that the information and advice contained in the 2010 ACPO guidelines 'cannot be used for [the] handling of SSD devices'. The 2012 ACPO guidelines do not appear to deal explicitly with SSDs (although NAND memory chips are discussed in the NPIA mobile phone Standard Operating Procedures). The only known way to avoid SSD self-contamination is by removing the controller, detaching the memory chips from the SSD and by using specialist hardware to access and read the data content (Gubanov & Afonin, 2012) – a time-consuming process and one not guaranteed to succeed.

Further, as well as the garbage collection, wear-levelling and TRIM problems, the way an FTL behaves is also not necessarily consistent between devices, and its algorithms are protected proprietary data. Thus investigators are faced with FTLs that could work in a number of different ways (Bell & Boddington, 2010) which also makes the analysis of SSDs much more difficult for forensic investigators. However the outlook is not entirely bleak since magnetic drives are likely to still be encountered for many years and there are also hybrid drives: part SSD and part magnetic drive which may yet provide other forensic opportunities.

Developments in cloud computing

In 2012 the university sector employer of four contributors to this book (in Canterbury in south East England) decided to replace their previous email, calendar and contacts software with the new Microsoft Office 365. Now, instead of the webmail and other services being physically provided by the university they are instead provided through the 'Microsoft Online Portal', a data server based somewhere in the European Union, probably several hundred miles away. The services are in 'the Cloud'. Many other organisations, particularly small and medium sized businesses, are also migrating to cloud computing due to the obvious economic benefits and flexibility this provides compared with traditional in-house IT infrastructure and services. Individuals too are increasingly using cloud computing, in many cases with no great knowledge of doing so (or the need to know). For example a user of an ipad with ios6 has the option of turning on 'iCloud'; 'Gmail' is a popular cloud-based form of email; Google Apps is a convenient way of cloud-sharing documents with others; many users of android-based smartphones would have installed 'Dropbox' for online storage; and Microsoft offer free online cloud storage with their 'SkyDrive' service. These services have developed in part because of the greater availability and speed of broadband services and the advent of Web 2.0.

Traditionally a user's own machine (for example, the PC on their desk) would store data and run software applications, or alternatively a machine or server centrally located within the same organisation would be used. But now the data and services can be located elsewhere and accessed through the internet, and through a variety of devices (PC, tablet, smartphone). The metaphor is a 'Cloud' floating somewhere above the user, somewhat ephemeral and delocalised, and not within the control of those on the ground. However, Vaquero et al. and others have proposed a more formal definition of a cloud computing service (a CCS) as: "[…] a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services)'. An even more detailed definition is offered by Grance and Mell (2009): 'Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction'. There is a general distinction between a 'public' cloud computing (the internet used for data transfer) and 'private' cloud computing (where an organisation's intranet is employed). Sometimes however the two are linked, with private clouds off-loading some tasks and data to a public cloud. Within the field of digital crime policing, attention is currently more directed at public cloud computing services as these provide a greater challenge to law enforcement.

Svantesson and Clarke, (2010) outline five characteristics of a CCS:

- the CCS is delivered over a telecommunications network;
- users of the CCS rely on it for access to and/or processing of data;

- the data is under the legal control of the user of the CCS;
- the CCS is ‘virtualised’, so the user is unaware of which server is running or which host is delivering the service, or where the host is located; and
- the CCS is delivered through a relatively flexible contractual arrangement in terms of the volume of data the CCS stores for a client.

A CCS can operate in three main ways, and supply different types of services as shown in the table, although it should be noted that these can be combined with a single provider (Taylor et al. 2010, p. 304).

CCS	**Description**	**Example**
Platform as a Service (PaaS)	CCS provides the operating system, applications, data storage etc. for the end user through the internet	Oracle Cloud
Infrastructure as a Service (IaaS)		
Software as a Service (SaaS)	Software applications are run on CCS provider’s system, accessed by the end user through the internet	Google Apps

Security of data is an issue for all providers of cloud computing services, and by implication, for their clients. As noted elsewhere in this book, personal data is a valuable commodity and is likely to be a subject of interest for many people intent on its criminal exploitation. The attempted malicious hacking of cloud services is almost certainly a regular occurrence. Some CCSs are also hosted in countries where privacy laws are inadequate, and security is perhaps less of a priority. There is also some indication that the Cloud is being utilised by those with criminal intentions, for example allegations include individuals using cloud computing services to ‘run command and control services for botnets, to launch spam campaigns and to host phishing websites’ (Wolpe, 2010 citing Rik

Ferguson, Trend Micro). For investigators, indications of the use of CSSs can often be found on a digital device, for example, Dropbox leaves traces in the Windows system (McClain, 2011).

There are a number of forensic opportunities for the investigator in cases where it seems likely that potential evidence is located in the Cloud. For example a CCS could be automatically recording storage usage, number of log-ins, data displayed and accessed, and perhaps most importantly the IP address of a user together with date and time of access, with some of this information being retained for significant periods (Taylor et al. 2010,p. 305). However, it is probably correct to argue that in terms the forensic capture and analysis of digital data, the current investigator mind-set is still 'off-line', and the guidelines provided in many good practice manuals for example seem to implicitly assume that the investigator has control of the storage medium. Part of the problem is that evidence 'is more ethereal and dynamic in the cloud environment with non-or semi-permanent data' (Taylor et al, 2010, p. 304). Some of the information which is normally written to the operating system (e.g. date stamps) will instead be written to the cloud with the attendant difficulties in recovery. The legal basis for the seizure of equipment is also somewhat more uncertain (Mason and George, 2011). Data is also moved around in the Cloud from server to server and this might pose difficulties in terms of a digital investigator tracking the data down, and there will then be the added problem of gaining the legal authority to require the Cloud supplier to provide the data. Finally, there are likely to be increased difficulties with the audit trail the Cloud is likely to provide (continuity logs in particular) when compared with, say, hard drive forensics.

However, cloud computing provides a number of opportunities for the policing of digital crime. For example, a distributed cluster of PCs on the cloud can provide the power to break a password used for encryption. Cloud computing also provides new forensic opportunities because even if the data is securely erased at one location there might still be a copy at another. The Cloud also provides the facility to enable an organisation to back up and utilise data whilst digital evidence is being collected in situ.

Policing the Deep Web

Most users interface with the internet through search engines such as Google. To 'find something' on the web a person will simply type a description into the conveniently provided Google search box which will generally produce an impressive number of 'hits'. The user will then explore the so-called 'visible' web, often jumping from one website to the next. Occasionally a bookmark will be used to access a website or the website address will be typed in for direct navigation to the site. Many people will also be familiar with other visible parts of the internet, for example portals to SNSs such as Facebook or Hyves, or through apps such as Apple 'Maps'. However, beneath this readily accessible surface there is a less visible region of the internet that is more difficult to access; the 'deep web'(sometimes known as the 'hidden web' or 'invisible web'). Most users will be aware that access to some parts of the internet is controlled in some way, through gatekeeping such as the use of usernames and passwords, and that they may have to manually enter the web address of the site to accessing resources. These include web-based email services to special interest forums, commercial services (such as the 'Sky Go' portal for subscribers to view subscription satellite broadcasts online), and databases of academic journals available to students and others.

The deep web is undoubtedly very much larger (in terms of volume of data) than the visible web. In 2001 Michael Bergman (widely acknowledged as the originator of the term deep web and citing a 'Bright Planet' survey (Bergman, 2001)) estimated that deep web contained some 7,500 terabytes of information and 550 billion individual documents, compared to nineteen terabytes of information and one billion documents in the visible web. Most of the deep web consists of entirely legitimate content and is not connected with criminal activity. It is also utilised by people, such as political dissidents within repressive regimes, who need to keep their identities secret for reasons of personal safety: indeed this was one of the original intentions when the 'Freenet' part of deep web was first developed.

However, the deep web is inevitably also used for criminal activities and these 'sub nets' are sometimes collectively known as the 'dark web' or the 'dark net'. Examples of dark web activities include invite-only and password-protected chatrooms and forums used by cybercriminals for sharing information or trading. Some attempt will generally have been made to conceal their existence, for example by using obscure naming systems, and casual users of the internet will not generally access the criminally oriented parts of deep web or indeed even be aware of its existence. Deeper still are the parts of the deep web which can only be accessed using specialist software such as ‘I2P’ or ‘Tor’ (the latter also is available as a plugin for some internet web browsers). The open source software available to access the deep web varies, but most of it can supposedly conceal the identity (IP address) of a user, for example by using random pathways to communicate between parts of the network, and encryption. (The anonymity provided by Tor is sometimes employed as a tool by investigators to hide their surveillance of suspects.)

Furthermore, a suspect using the dark web site ‘The Armoury’ for example (which purports to specialise in the supply of firearms and other weapons) might not only employ the inbuilt anonymity provided by Tor and similar software, but also conduct business through ‘stealth listings’ of weapons which are not visible unless the specific URL is known. The transaction is also conducted using a number of ‘tumbler boxes’ which further complicate attempts to trace and demonstrate payment.

As with the visible surface web, internet services such as forums, data warehouses and wikis are all available within the dark web. These hidden parts are not simply stored as single items on a particular server but instead are distributed through networks. Lurid claims are made for the supposed availability of nefarious services and items on the dark web including hyperbole surrounding child abuse imagery, the supply of illegal drugs and firearms and even offers to conduct murder (for example, see the popular English newspaper, the ‘Daily Mirror’, (2012)). Claims are also sometimes made for the terrorist use of the deep web. However, the very nature of deep web (for example the need for particular software and the anonymity) might make it less attractive to those seeking to profit from crime. For example, organised criminal groups attempting to profit from the supply of child sexual abuse imagery need to market their services, and this is much more easily conducted on the surface web. Similarly, radical groups attempting to recruit new members from disaffected communities are likely to wish to display their message in locations which are most likely to be seen by potential recruits.

However, although claims of serious criminal activity within the deep web are usually unsubstantiated there is clear evidence for the availability of a number of illegal products and services, and there are also forums for sharing information on how to perpetrate certain illegal activities. For example, the 'Silk Road' website claims to offer illegal drugs by mail order. In 2012 it was offering for sale prescription anti-depressants and opiates, steroid and illegal street drugs including marijuana, ecstasy, heroin and cocaine, together with 'date rape' drugs such as GHB. There are a number of indications that the Silk Road illegal offer is 'for real' and not an elaborate scam, for example, a man from Melbourne was arrested in 2012 for importing narcotics he had allegedly sourced from the Silk Road (AFP, 2012). The Silk Road uses a form of 'anonymised' currency called 'bitcoins' (often referred to as 'BTCs'). Potential purchasers first exchange traditional forms of money (for example, using a debit card and an online trading site such as Mt. Gox) for bitcoins, and these are then used for transactions and trading. In mid-2012 it was estimated that approximately 1.9 million US dollars was being generated through the Silk Road site (Christin, 2012, p. 3). The site is currently the subject of attention from the LEAs around the world, including the US Drug Enforcement Agency. The users and content of the dark web is therefore of clear and legitimate interest to law enforcement agencies.

The investigation of both the deep web and the dark web can inevitably pose problems for investigators given its closed and particularly in the case of the dark web, its often bogus nature. In terms of the 'legitimate' parts of deep web, a number of search tools are being developed including some from Google. Although these will not necessarily allow access to password-controlled resources they will at least reveal the existence of some of these resources. Proactive investigations of the dark web will often employ software such as Tor, and could include investigators posing as a customer for an illegal service or establishing a connection with individuals or organisations involved in the sharing of illegal imagery. Reactive investigations in the dark web (for example, attempting to collect intelligence against an identified suspect) are perhaps more problematic, but use of the dark web does not guarantee anonymity for a suspect, largely because of the naivety of most users of Tor and other similar software. An indication that Tor had been used to access dark web would be an SSL request to the user's ISP to become the first Tor node in the communication pathway, together with the use of encryption. However, if users employ Tor through a web browser they might still have cookies enabled, and may also be running javascript, both of which will assist a digital forensic investigator in testing hypotheses in terms of a suspect accessing illegal material.

Anti-forensics

The term 'anti-forensics' is used to describe a range of methods, techniques and actions that are intended to thwart the efficacy of a forensic examination of a digital artefact. At the outset, it should be noted that 'anti-forensic' methods are often employed by individuals for what we would regard as entirely legitimate and non-criminal reasons – for example, for reasons of privacy, to protect themselves against the actions of oppressive regimes or simply on grounds of principle and ideology. The use of anti-forensic tool by an individual should not necessarily be seen as a 'smoking gun' that indicates guilt. That said, organised crime groups in particular are likely to be fully cognisant of the advantages of adopting an 'anti-forensic' mind-set.

We can identify five main categories of anti-forensics activities and these are described in the table. Note that the categories are not mutually exclusive.

Category	**Explanation**	**Example**
Source elimination	Use of methods to attempt to stop any evidence accruing, particular at source	Using an 'incognito' window with the Firefox brows Using Trojan commands
Destruction	Attempting to completely erase data.	Using software that employs algorithms to repeatedl overwrite data. Physically destroying the medium that held the data.
Obfuscation	Covering tracks, changing file attributes to confuse time of origination etc.	Use of 'log cleaners' Spoofing IP addresses The 'Zeus Botnet Crime-ware toolkit'
Hiding	Altering data in such a way that it is not visible or not 'readable'	Steganography. Encrypting files.

		Hiding data in Windows ‘slack space’
Counterfeiting	Attempting to present one form of data as another innocuous form	Changing a file extension on an image file (e.g. .gif) another extension that is less ‘suspicious’ (e.g. .xls)

(Derived in part from Harris (2006) and Rogers (2005)).

Cryptography, the science of encryption, is an important part of anti-forensics, and presents particular challenges to the investigation of digital crime. It involves changing data into a form which is unreadable (encrypted) and has been used for thousands of years, often as a means of retaining the confidentiality of information, particularly important in times of war. A well-known example was the Enigma code used by the German military in World War 2 and its decryption at Bletchley Part in the UK by a group of workers that included Alan Turing. Encryption can be used for text and data, not just for data representing text, and also includes data conveyed over a communication network such as the internet. All data, including images, videos and audio files (if not already in in numerical form) can be converted into a digital format.

Modern forms of encryption use a mathematical set of instructions (an algorithm) and a ‘key’, or set of keys. The encrypted data can be converted back to its original intelligible form, using the algorithm and a key, but not usually the same key as was originally used to perform the encryption. The lengthy and complex processing work of encryption is performed by software, much of which is easily available via the internet. Examples include, PGP (‘Pretty Good Privacy’) and its derivatives (such as GNU Privacy Guard), and other software such as TrueCrypt and AxCrypt. A number of popular applications, such as the Microsoft Office suite also provide encryption as an option for the user. Encryption can be used for a complete hard drive, for individual files, or for network traffic. (Many readers will be familiar with the existence of the latter when using secure online payment methods).

The algorithms used for encryption are derived from theorems that originated in number theory (a branch of pure mathematics), and particularly ‘trap door’ mathematics which provides algorithms that in practice will only work in one direction. This makes modern forms of encryption very powerful, and almost impossible to ‘crack’ if correctly implemented.

A commonly used form of encryption uses public and private keys, and in this context a key is a number or several numbers. A 'public key' is potentially accessible to anyone and a 'private key' is kept secret by an individual. A public key is shared with others who may then use it to encrypt messages. The encrypted messages can be decrypted by the recipients, each using his or her private key. To illustrate both the process, and just how difficult encryption can be to crack we will take a simple example (adapted from Bryant and Bryant, 2008, pp. 100-101) in which Brian wants to encrypt the message 1234 and send it to Anneka using the RSA method of encryption (in practice the encryption is done automatically using software). She chooses two prime numbers such as 26,863 and 102,001 (a prime number is any number that is not divisible by any whole number other than one and itself). Next she multiplies these two prime numbers together: 26,863 × 102,001 = 2,740,052,863. She then reduces each of the chosen prime numbers by exactly one, and multiplies the answers: 26,862 × 102,000 = 2,739,924,000. Next she needs to find two other numbers that must have a particular mathematical connection with the second result (that is, with 2,739,924,000). These two numbers (a) and (b) have to be chosen so that:

(a) shares no common divisor (other than 1) with the number 2,739,924,000;

(b) shares no common divisor (other than 1) with 2,739,924,000), but also when multiplied by the number (a) and then divided by 2,739,924,000 leaves a remainder of exactly 1.

Using the mathematical process 'Euclid's Extended Algorithm', Anneka finds two numbers that satisfy the criteria, namely (a) = 103, and (b) = 1,143,851,767.

(To check this, calculate 103 × 1,143,851,767 = 117,816,732,001. Then check that 2,739,924,000 divides into this (43 times) leaving a remainder of exactly 1).

The two numbers 103 and 2,740,052,863 are Anneka's public key. (It would in fact be more appropriate to describe this as Anneka's public lock, but the term 'public key' is the one normally used). To help the explanation we will designate 103 as the first part of her public key and 2,740,052,863 as the second part of her public key. She sends both parts of her public key to Brian.

The number 1,143,851,767 ((b)) is Anneka's private key, and she keeps this to herself.

Brian now sets about encrypting his plaintext message 1234, to be sent to Anneka. First he raises 1234 to the power of the first part of Anneka's public key, 103 (that is 1234^{103}; a very large number), and then divides this result by the second part of her public key, (2,740,052,863) and notes the remainder. The remainder is the encrypted message. In this case the remainder is 1,063,268,443 and hence the encrypted version of 1234 that Brian sends to Anneka is 1,063,268,443. Note that Brian has not used any information that has been kept secret – he has only used Anneka's public key.

Anneka now takes Brian's encrypted message and raises it to the power of her private key, 1,143,851,767, and then divides by the second part of her public key, 2,740,052,863. Without the private key this part of the decryption process is impossible to achieve by 'trial and error'. The remainder in this case is exactly 1234; Anneka has successfully decrypted Brian's message.

This form of encryption is certainly difficult to decipher, but there are other advantages; by not using a secret key in the transmission process the danger of compromising the security of the message is practically eliminated. Public key systems can also be used to communicate securely with groups of people. All of this has obvious advantages to those who wish to share data securely, and inevitably this will include those with criminal intent, such as individuals sharing child abuse imagery. Over the past 20 years or so there have been frequent examples of offenders (including terrorist suspects), employing encryption as an anti-forensics method. For example, in 1995 members of the Aum Shinri Kyo sect released Sarin gas on the Tokyo underground, killing 12 people and injuring thousands more. Members of the cult had used RSA encryption (see above) on computer files in an attempt to hide their plans to deploy weapons of mass destruction.

The existence of an encrypted artefact will be reasonably clear to the investigator: for example, an encrypted word file will appear as a 'nonsense' string of letters and symbols when viewed with forensic software, and will not contain intelligible text (although the file might contain readable information concerning the software used for encryption). Encryption poses significant problems for digital criminal investigations (Reyes et al., 2007), as without the password a robustly encrypted digital artefact is near-impossible to 'crack'. We provide an overview in Chapter 3 of some methods open to the investigator (within the law) to secure passwords.

It is best to capture information using live forensics when possible, perhaps using covert means (e.g. surveillance), using surreptitious installation of key logging software through a back door method, physical key loggers, electromagnetic capture of emanations from a suspect's PC etc (but see the discussion in Chapter 4 on the legal considerations in the UK surrounding covert methods

Expert evidence

The expert's opinion, findings and associated methodologies in a range of specialist fields are now subject to an increasing level of scrutiny. This is in part due to recent problems with expert evidence, for example high profile miscarriages of justice with flawed expert evidence (Law Commission, 2011). The Solicitors Journal (2011) cites the Law Commissioner, Professor David Ormerod as saying that judges are 'in the unsatisfactory position of having no real test to gauge the unreliability of expert evidence'. In the case of *R* v *Clark* [1999], Professor Sir Roy Meadows made claims regarding infant cot death which had with 'no statistical basis' (Royal Statistical Society, 2001). As a paediatrician (not a statistician), Meadows was testifying outside of his area of expertise, and this occurred in other cases too (*R* v *Cannings* [2002] and *R* v *Patel* [2003]). All these convictions were quashed on appeal, and the Law Commission subsequently reviewed the admissibility of expert evidence for use in criminal trails (Law Commission, 2011). Their report calls for a move to incorporate a test of the level of reliability of the opinion of an expert witness to ensure that the evidence is based on sound scientific principles, techniques and assumptions.

Requiring greater levels of scientific rigour as the basis for expert evidence derived from a forensic science presents a number of challenges, not least of which is that the term 'forensic science' is itself a misnomer. It is claimed that forensic science lacks the scientific principles that underpin more traditional scientific disciplines (Kennedy, 2003). With little formal research and no research agenda, there is a correlation between 'dubious forensic science and wrongful convictions' (Cooley, 2004). The absence of a body of knowledge, established through accepted scientific methodologies has led to criticism of practitioners, and suggestions that they are rhetorical in their application of substance or methodology (Saks & Faigman, 2008). Saks & Faigman go on to state that scientific principles such as rigorous empirical testing, inductive methodologies and reporting of error rates are all absent from many of the 'non-science forensic science' disciplines.

Without this scientific pedigree, many of the specialties within forensic science face the risk of being labelled as 'junk science' in court (Huber, 1993). Epstein (2009) cites fingerprints, handwriting and firearms as three examples of such science, and goes on to suggest that evidence from such sources should be excluded from trials. Other examples of problematic forensic science include voice identification, footprints, ear-prints, bite marks, tool marks, blood spatter and hair comparison (Edmond, Biber, Kemp & Porter, 2009). Broadly speaking, all of these specialties concern themselves with applying individualization to link an artifact to a suspect. Disciplines such as computer forensics and malware forensics likewise utilise the automated record-keeping nature of computer software and operating systems to apply provenance to identified artefacts.

It is not uncommon to discover malware in the course of any forensic computer investigation, and suspects frequently offer the 'Trojan Defence', arguing that the contentious actions were performed as a result of some form of malware (or cyber-criminal) having gained control of his or her computer (see Chapter 8). Both civil and criminal forensic practitioners of course have a duty to consider such a defence (in the UK this would be under the Civil Procedure Rules (Ministry of Justice, 2010) and the Criminal Procedure Rules (Ministry of Justice, 2011)). Forensic practitioners are reliant on their tools, skills and knowledge of malware to detect, identify and study the behaviour of any identified malware, and they need to form an opinion on the impact any identified malware has had. But the lack of a scientific footing for malware forensics has a greater impact for the discipline than it does for computer forensics. The availability of both undergraduate and post-graduate qualifications in computer forensics provides an opportunity for practitioners to engage with their discipline on an academic and scientific footing, but although included as modules on some courses, there are no such equivalent academic qualifications specifically dedicated to malware forensics. Malware is designed to obfuscate its true intentions and hinder attempts to analyse it, and there is therefore already a level of uncertainty associated with any conclusions drawn from malware analysis, and this can be used in a case to raise ‘reasonable doubt’ about the true nature and intentions of a particular piece of malware. The complexity of the subject matter and the specialist skills required to study it (eg: reverse engineering & assembly language) may also make the specialty less accessible to practitioners.

Lawyers seeking to undermine evidence produced from malware analysis currently have a rich choice of methods of attack for introducing reasonable doubt. Even one of the most fundamental requirements of digital evidence, the ability to verify and hence corroborate the findings of the expert, is open to challenge. An established tenet of science is that hypotheses are supported by reproducible experiments (Beckett, 2010) and that hypotheses are tested through falsification, or refutability (Popper, 1968). Although there are some examples of the use of falsifiability in forensic hypothesis testing (see Willassen, 2008, in the case of the investigation of digital timestamps) these are few and far between.

Practitioners intending to present digital evidence must expect to be required to defend their findings, and disclose enough detail to enable an opposing expert to verify and possibly provide an alternative explanation for an artefact. ‘Dual-tool validation’ is often promoted as a tenet of a scientific approach to forensic computing and hence good practice (NIST, 2001; SWGDE, 2004) and is explained in Chapter 8 – in essence it means using one forensic software tool to check the results achieved by another. But there are problems with this, for example one forensic provider states ‘Dual-tool verification can confirm result integrity during analysis’ (Forensic Control, 2011), but this is a bold claim, and open to challenge if a third tool or manual inspection of the raw data identify a discrepency. Another provider makes the less radical claim that the forensic software products EnCase and FTK ‘allow for a dual-tool approach for the verification of findings’ (Cy4or, 2009). However, as in the previous example, no scientific studies or supporting evidence are cited. A third example is a freelance forensic investigator who states on his website (in relation to tool validation), that ‘I don't validate my tools - I validate my results. Generally I do this with dual tool verification’ (Drinkwater, 2009). This statement is contradictory, as a second tool is used to check the results of another.

Dual-tool ‘verification’ thus falls short of the standard scientific criterion for of verification: that of the need of falsification as the test of hypotheses. Put simply, both software tools could arrive at the same wrong conclusion either because both are using the same erroneous method or by coincidence. Arriving at the same result is not a scientific ‘proof’ is that the outcome is true. Although it is accepted that there are no documented examples of two tools making the same error (Beckett & Slay, 2007), this none-the–less could arise, for example, by the use of the same underlying Windows API call, and in such circumstances both tools are making the same erroneous assumption (Sommer, 2010). The scientific approach is not to repeatedly look for confirmation of the hypothesis but instead to formulate ways in which it can be falsified and test for this.

Although dual-tool verification cannot scientifically ‘confirm’ a result, it can provide corroboration, if only at the level of probability rather than certainty. However, the main benefit in applying a dual-tool approach arises where there is a discrepancy in results (Turner, 2008), thereby highlighting the need for closer analysis. An example of this was during the trial of Casey Anthony who was charged with the murder of Caylee Marie Anthony in Orlando, Florida. During this trial a discrepancy was identified between two internet history tools used to produce expert testimony. As a result of this discovery, the developer of one of the tools corroborated the tool’s output by reverting to the underlying raw data and interpreting the data manually (Wilson, 2011). Ideally, an independent party unaware of the expected outcome should have undertaken this step.

The acceptance of a tool or methodology sanctioned by others is common practice in both legal and scientific circles. In judicial processes, legal precedent can be cited from prior cases where particular techniques have been admitted into proceedings, but scientific work advances by citing and carefully extending a previously established body of knowledge through hypothesis testing. The difference arises in how these precedents are determined and hence accepted. Kritzer (2009) argues scientific and legal inquiry differ in how they persuade and hence accept propositions. He explains that the scientific tenet of general acceptance and peer review is advanced through repeated attempts to falsify a hypothesis, and that truth in a scientific context is complex and elusive, and can only be approached by a process of eliminating falsehoods. This is very different from the concept of truth as applied within the legal context, which is revealed through the adversarial process. In accepting a given truth, the legal enquirer values certainty, whilst the scientist values doubt and scepticism, argues Marsico (2004). He goes on to state that if justice is blind, then it will 'blindly follow evidence presented as truth'. The role of judges, he continues, should be limited to evaluating the admissibility of evidence, rather than to evaluate the credibility of scientific evidence.

The Daubert test (or Daubert standard) in the USA, developed in 1993 from the Supreme Court decision in the case of *Daubert v Merrell Dew Pharmaceuticals,* provides a framework to assist the judiciary in evaluating scientific evidence and the admissibility of this evidence from expert witnesses under Federal Rule of Evidence 702. As explained in Chapter 4, one of the four considerations of the Daubert test is the acceptability (to the relevant scientific community) of the theory or technique used to derive the evidence. However, within the field of digital forensics there is no established 'scientific community' to draw upon (Marsico, 2004). Instead justification for the scientific basis of a particular digital forensic method or software application is often based upon one of two forms of fallacious argument *(ad populum* and *consensus gentium*), that a conclusion is true simply on the basis it is believed by a large number of people; that is by 'common consent'. We can see clear *consensus gentium* arguments in a number of examples in relation to the validity of particular software and hardware for informing expert scientific testimony. For example Guidance Software (2011) state that they have evaluated their EnCase software product against the Daubert test. In addressing the general acceptance criteria of this test, they argue that with in excess of 30,000 licensed users their product is generally accepted. However, Carrier (2002) points out that a forensic tool will probably be chosen on the basis of interface and support, and that the size of the user community is not a valid measure of procedural acceptance.

Van Buskirk & Liu (2006) have observed that statements such as those by Guidance Software lead to a tendency within the judicial system to presume that forensic software is reliable. They identify EnCase reliability issues, but Limongelli (2008) of Guidance Software defends its reliability, citing Williford v State [Texas, 2004] in which the court concluded that EnCase software is reliable. However this conclusion was based on the anecdotal testimony of a single police officer. Limongelli however goes on to cite Sanders v State [Texas, 2006] in which it was concluded that once the scientific reliability of a specific methodology is determined, 'other courts may take judicial notice' of the result. No consideration is given to the possible effects new versions of the software, or to bugs and/or errors that might arise in particular environments.

Sommer (2010) identifies how (in the UK) through the application of Part 33.6 of the Criminal Procedure Rules, at a pre-trial hearing just two individuals (the opposing experts in a case) can accept the validity of digital forensic evidence derived using a novel approach as 'sufficient' for the purposes of the case under consideration (but without implying a more general acceptance of the technique). Beach (2010) suggests that the scientific approach of falsification of hypotheses is not an issue for practitioners operating within the legal arena, as the concept of 'truth' differs between the science and legal profession – the law's objective is to resolve conflicts rather than to increase knowledge. Within the bounds of a single case, truth is deemed 'static' and not open to re-evaluation. Denning (2005) argues this acceptance of untested theories is a wider problem within the computer science community as a whole.

In the UK, partly in response to a number of miscarriages of justice, the Forensic Regulator was formed in 2008 with a remit to 'establish and monitor compliance with quality standards for the provision of forensic science services to the police and wider criminal justice system' (Forensic Science Regulator 2009). The Regulator's 'Codes of Practice and Conduct' (Home Office 2011) are aligned to BS EN ISO/IEC 17025:2005 (ISO 2005). Forensic service providers now therefore face an emerging regulatory requirement to demonstrate their working practices meet with minimum standards. In anticipation of this trend, some police forces are beginning to award contracts for outsourced work partly on the condition that the provider is ISO 17025 accredited. Given the appointment of a Forensic Regulator and emerging regulatory standards, it can be argued that the issues identified currently undermine the trust that can be placed in findings tendered in criminal proceedings. The production of electronic evidence therefore requires the use of reliable tools and competent operators, both currently areas of active research.