Downloads, Logs and Captures: Evidence from Cyberspace

Peter Sommer
Computer Security Research Centre,
London School of Economics & Political Science

ABSTRACT: *The growing use of the Internet, online hosts, electronic banking and bulletin board systems means that with increasing frequency evidence needs to be collected from remote computers for use in legal proceedings. Issues of the evaluation of weight still need to be addressed even if strict rules of admissibility are removed. The background processes involved need to be understood if courts are to be able to test evidential quality The controls that should be in place are discussed and a series of tests of provenance and reliability are suggested. However such tests will never be more than decision aids.*

A frequent and increasing requirement is the acquisition of reliable evidence of the existence, content and provenance of data held on a remote computer. Any prediction about the growth of electronic commerce has to include a forecast that some of these transactions will go wrong - what evidence can we hope to find of the intentions and actions of the participants?

It is still true that the vast majority of computer-derived exhibits offered in legal proceedings consist of relatively normal documents which happen to have been produced by a computer. Usually this is in the form of print-out of various kinds such as accounts, statements, invoices, correspondence, reports, memoranda, or copies thereof. A further common type of computer evidence is read-out from single-purpose devices such as alcohol level meters and telephone call-loggers. [1] Nearly all of the cases on the problems of computer evidence - and most of the articles in academic legal journals - are about these two broad classes. Over the last few years the courts have also been offered evidence derived from data media. Typically these have been files derived from the hard-disks installed within computers, the removable floppy disks used for temporary storage and the tapes and optical disks used for back-up or archive.[2] The courts have had little difficulty in principle in accepting such files, even if they include graphics as well as conventional text. [3]

[1] Eg, among many English cases, Burditt v Roberts [1986] RTR 391, Castle v Cross [1985] 1 All ER 87, [1984]1 WLR 1327, McKeown v DPP (1995) Crim LR 69, R v Burke (1990) Crim LR 288, R v Pettigrew [1968]2 All ER 195

[2] A brief introduction to the field is provided in: Sommer, P 'Clueless: Computer Forensics' *Secure Computing*, November 1995, pp 15-21 and a more detailed explanation of the specialist technology needed in disk imaging can be found in Wilding, E, *Computer Evidence: a Forensic Investigations Handbook*, Sweet & Maxwell, London, 1997 pp 124-128

[3] R v Fellows, *Times* 03 Oct 96 (Court of Appeal)

But there are many circumstances in which normal print-out straight from the computer which generated it does not exist or is not immediately available. Data media may only be obtainable with the consent of the computer owner or by formal discovery/disclosure - and by the time it is delivered it may not contain the precise files required. In the alternative it may be possible to secure seizure under warrant or order - but such warrants and orders are only issued against precise criteria[4] and again, files and file contents may alter between the time at which it is first decided that a file is worthwhile potential evidence and the point at which a warrant or order can be executed.[5]

Since the early 1970s computers accessible via the public telephone system, leased lines or the Internet have offered a variety of electronic products: text, structured data, application programs and, increasingly, the opportunity to buy goods and services. These remote access computers can be bulletin board systems, conferencing systems, online publishing systems and a whole range of Internet-based facilities, above all the World Wide Web. For more than fifteen years various forms of electronic banking have been available for small businesses and individuals[6]; financial institutions and large companies have been making payments and settlements since the 1960s [7] [8], initially via the telex system. We will visit these shortly.

As the range and use of these connected computers develops and the contexts in which they operate reflect an ever widening set of commercial and social circumstances, the files and data that are held become relevant in an increasing range of states of affairs. However many investigators and regulators seem so far broadly unaware of the problems of acquiring evidence sufficiently robust against hostile criticism in court[9].

In criminal law we can single out allegations claiming:

- infringed copyright materials offered in the course of a business [10]
- evidence of fraudulent offers to deliver goods [11]
- evidence of fraudulent offers to provide services[12]

[4] In English Criminal Law general powers of seizure are defined in ss 8-16, Police and Criminal Evidence Act, 1984 and Orders deriving therefrom but there are many additional powers for specific instances, eg s 7 Bankers Books Evidence Act, 1879; in respect of Computer Misuse - s 14, Computer Misuse Act, 1990 and in respect of criminal breach of copyright - s 109, Copyright, Design and Patents Act, 1988. There is also a civil *ex parte* procedure available using an *Anton Piller* Order; however these have to be carried out under the supervision of an independent solicitor and there is extensive protection for the interests of innocent victims of such searches through cross-undertakings.

[5] Obtaining files from remote computers without authority is a criminal offence: s1 Computer Misuse Act, 1990 and evidence obtained thereby runs the risk of discretionary judicial exclusion under s 78 Police and Criminal Evidence Act, 1984. However there is protection for "law enforcement officers" in s 162, Criminal Justice and Public Order Act, 1994

[6] The first seems to have been Verbraucherbank on the German videotex system Bildschirmtext in 1981; the Bank of Scotland launched HOBS in 1984

[7] Cedel, which provides electronic settlement in Eurosecurities, was founded in 1970, for example.

[8] Electronic transactions between banks and other financial institutions present fewer evidential problems as these take place on closed networks such as SWIFT and use specialised protocols which incorporate, at both the technical and legal level, extensive authentication provisions as well as clearly defined rules declaring at what points obligations are transferred.

[9] For example: they are not discussed at all in the report *Information Technologies for the Control of Money Laundering*, US Office of Technology Assessment OTA-ITC-630, September 1995, despite its extensive review of the forms of bank reporting, formats of electronic funds transfer messages, and possibilities of computer-aided surveillance.

[10] For example, under s 107, Copyright Design and Patents Act, 1988

[11] Under ss 15(1) and 16 , Theft Act, 1968

- evidence of fraudulent or non-compliant investment offers[13]
- holding or offering pornographic files and images[14]
- incitements to racial hatred, terrorism and other offences[15]
- conspiracies[16]
- computer misuse[17]

In civil proceedings we can suggest:

- evidence of the existence and terms of a contract which is alleged to have been breached[18]
- evidence whether a document or e-mail or file was sent or received as dated
- breach of copyright
- defamation[19]
- evidence of negligence

For ordinary, non-legal purposes the act of viewing or acquiring a file from a remote computer is relatively trivial, a question of learning simple sequences of commands in application programs. But when it is desired to produce reliable evidence all manner of difficulties appear:

- how can we show that the file acquired is what was on the remote computer?
- what do we need to do to show that the process of acquisition was free from error?
- how do we preserve the file once it has been acquired and be able to show that any subsequent copying processes have not introduced contamination?

---

[12] Under s 1, Theft Act, 1978

[13] See, for example, Drinkhall, J 'Internet Fraud', (1997) 4 JFC 3, p 258-261

[14] An extensive review of US and English law appears in Akdeniz, Y 'Computer Pornography; a Comparative Study of the US and English Obscenity Laws and Child Pornography Laws in Relation to the Internet' in *International Review of Law, Computers and Technology* Vol 10 No 2, 1996 pp 235-261

[15] Racial hatred offences include ss 18,19,21 and 23, Public Order Act, 1986, terrorism is addressed in the Prevention of Terrorism (Temporary Provisions) , Act, 1989. Simple incitement of another to commit an offence is a common law misdemeanour. Blasphemy also remains a common law misdemeanour - Whitehouse v Gay News Ltd and Lemon [1979] AC 617.

[16] In English law, a distinct statutory offence under s 1(1) Criminal Law Act, 1977

[17] That is, legislation specifically directed at unauthorised computer intrusion or unauthorised data modification such as the UK computer Misuse Act, 1990. An overview of international laws in this area appears in Schjolberg, S 'The Legal Framework - Unauthorised Access to Computer Systems', presentation at the International Information Integrity Institute Conference, Oslo, June 18-20, 1996. Available at http://www.mossbyrett.of.no/info/legal.html

[18] A useful introductory comparative study of US and European electronic commerce laws can be found in Hance, O, *Business and Law on the Internet*, McGraw Hill, New York, 1996 pp 147-175

[19] A review of the possibilities of online defamation is provided in Waelde, C and Edwards, L 'Defamation and the Internet: a Case Study of Anomalies and Difficulties in the Information Age' *International Review of Law, Computers and Technology* Vol 10 No 2, 1996 pp 263-294. See also Hance, O *op cit* pp 105-111 for a comparative study of US and European laws

And these are merely provisional issues of weight in relation to the mechanics of evidence acquisition; in most real life situations courts will also need to consider the weight of the purported content of the files, how they were handled by the computer which has produced them and what inferences may reasonably be drawn therefrom. In addition, most common law jurisdictions also have substantial and complex tests of admissibility of "computer records". [20] We need to step back a little.

Computer Evidence in General

Evidence derived from computers has many of the features of other types of evidence. Miller [21]provides a useful broad explanation:

> Evidence is information used to decide whether disputed propositions are true. A court cannot normally obtain evidence directly (first hand). A source is relied on, such as a document or human witness. The reliability of the information is assessed directly by testing the reliability of the source. If witnesses are used, they are cross-examined; if documents are used, a human witness is often asked to verify that a document is authentic and to give oral testimony about its content.

A similar definition comes from Wright[22]:

> Evidence is anything that demonstrates, clarifies, or shows the truth of a fact or point in question. A proponent may offer many types of evidence: documents, objects, witness testimony, and the results of practical demonstrations and scientific procedures. To be useful, evidence must persuade.

The following are Miller's general tests for the reliability of a computer-derived Exhibit [23]: it should be possible to show that evidence is:

- **authentic** - specifically linked to the circumstances and persons alleged
- **accurate** - free from any reasonable doubt about the quality of procedures used to collect the material, analyse the material if that is appropriate and necessary and finally to introduce it into court - and produced by someone who can explain what has been done. In the case of Exhibits which themselves contain statements - a letter or other document, for example - "accuracy" must also encompass accuracy of

[20] In English criminal law the primary computer-specific admissibility test appears in s 69, Police and Criminal Evidence Act, 1984 where a certificate of proper working of the computer is required. The law is widely regarded as unnecessarily complex as it has had to interact with the documentary hearsay provisions of ss 23 and 24 of the Criminal Justice Act 1988 and there have also been difficulties of practical interpretation. See Reed, C, (1990-91) 2 CLSR 13-16 and two contrasting cases: R v Shephard [1993] AC 380 and R v Cochrane [1993] Crim LR 48. S 69 is likely to go as recommended in the Law Commission's Consultation Paper No 138 "Evidence in Criminal Proceedings: Hearsay and Related Topics (para 14.32) and in the way that its civil equivalent, s 5, Civil Evidence Act, 1968 has been replaced by a simpler approach to "business records" in any form in ss 8-9 Civil Evidence Act, 1995. In the US the main problems of admissibility arise from the problems of searching and seizing while conforming to the Fourth Amendment. See *Federal Guidelines for Searching and Seizing Computers*, Computer Crime Unit, General Litigation Section, Department of Justice

[21] *Electronic Evidence - Can You Prove the Transaction Took Place?* Miller, C Computer Lawyer Vol 9 No5, pp 21-33, May 1992

[22] *The Law of Electronic Commerce* 2nd edition, Wright, B , Little Brown, Boston, 1995, p 7.1

[23] adapted and developed from Miller, *op cit*

content; and that normally requires the document's originator to be make a Witness Statement and be available for cross-examination [24]

- **complete** - tells within its own terms a complete story of particular set of circumstances or events

A number of writers use the terms "reliable" and "authentic" almost interchangeably and the definitions given above are the more useful for being more restricted.

Reed tries to explain authentication thus [25]:

> Authentication means satisfying the court that (a) the contents of the record have remained unchanged, (b) that the information in the record does in fact originate from its purported source, whether human or machine, and (c) that extraneous information such as the apparent date of the record is accurate. As with paper records, the necessary degree of authentication may be proved through oral and circumstantial evidence, if available, or via technological features in the system or the record

How computer evidence is different

But there are also many elements which set computer-derived evidence apart.

- **computer data can change moment by moment within a computer and along a transmission line** Many forms of conventional evidence are claimed to be a "snapshot" of a particular sequence of circumstances, but the problems are particularly acute with computers. This can create considerable difficulties over authentication as to content and time of creation
- **computer data can be easily altered without leaving any obvious trace that such alteration has taken place** Alterations in hand-written and typed documents are usually self-evident; log and account books are designed so that it is easy to detect if an entry or page has been omitted. [26] It can be argued that there are plenty of examples of forgery based on typed and hand-written originals, but computer-based documents can be forged with an ease and freedom from detection which is of a quite different order. It is of course entirely possible to design a computer system which thwarts certain forms of unacknowledged alteration[27]. But, in contrast say, to paper-based books of account, there are few obvious "standards" which set a measure of what to expect.
- **computer material can be easily changed as a result of the process of collecting it as evidence** Many forms of forensic examination run the risk of contamination. Biological samples from a subject can be inter-mingled with those of the examiner. [28] But the problems with some computer-derived material are intense - the very act

[24] This is the general common law position; there have been a number of statutory exceptions starting with the Bankers' Books Evidence Act 1879

[25] Reed , *op cit* 1990-91 CLSR 13-16

[26] Indeed, s 722(2) of the Companies Acts 1985/89 states that if registers, indexes, books or accounting records are not in a bound book then adequate precautions against falsification and to facilitate its discovery must be taken

[27] For example, by the use of audit trails or "digital fingerprinting" - adding checksums to files.

[28] Some of the problems are discussed in Steventon, B 'The Ability to Challenge DNA Evidence' Royal Commission on Criminal Justice, Research Study No 9, HMSO, 1993 and Alldridge, P 'Recognising Novel Scientific Techniques: DNA as a test case' [1992] Crim LR 687 at pp 689-691

of opening an application or file, even if there is no intention to alter anything, often in fact creates changes although they may not be immediately visible.

- **much immediate computer evidence is not obviously readable by humans** Actual Exhibits are often derived, manipulated and "presented" away from their point of origin. This becomes apparent as soon as one moves from the limited vision of "computer evidence" as being simply a "record or document produced by a computer". There is nothing wholly unique about this; the typical DNA trace Exhibit is not DNA itself but a purported representation in a form which aids analysis. [29] The particular problem in relation to computer evidence is the large number of possible and potentially "accurate" representations of original computer data that can exist. What is seized may be a computer disk which in turn contains large numbers of directories of files of various kinds, while what is put immediately before the Court may be any of a number of purportedly accurate print-outs or screen dumps. The large variety of possible representations of original material makes difficult the evolution of "standards" such as tend to exist with, say, DNA charts. And the possibilities for inaccurate representation are very much greater. Nearly always, computer-derived exhibits require that the court make a chain of inference before reaching a conclusion.
- **computers create evidence as well as record and produce it** The traditional manually-maintained "books of account" consisted of sheets of paper into which hand-written or typed entries had been made; subsequent calculations were also substantially manual, even if a simple calculator was employed for some of the stages. But in the computerised equivalent, it is only the original entries that are manually input and all the other "records" are produced by the computer. There are many examples where computers "assemble" documents, etc; and only do so at the point at which a request is made for the document to be created. This can be true of online requests as well as conventional print-out or on-screen reports.

These questions extend well beyond the particular problems of evidence from remote computers but a specific examination of them yields some answers to the broader issues.

Key Tests

To understand how remotely-acquired computer files might meet the conventional tests of evidential reliability - authentication, accuracy and completeness - we need to identify the various stages typically involved:

- **Remote Computer's Correct Working Test** Can we show that the remote computer was behaving "correctly" or "normally" at the time? This requirement may be related to tests of admissibility[30] but even if a particular jurisdiction does not impose such a test, it is still necessary to show a court that the output of the remote computer can be relied on. If all that the remote computer is doing is holding a series of files placed there by human beings - the equivalent of a filing cabinet - that test may be relatively easy to satisfy. But more typically computers assemble or produce "records" from databases, generated against various criteria. When a computer says it has, say, seven "matches" and no more, or that an account balance is so many pounds, or that a particular magazine extract is identical to what appeared in the print original and the given date of publication is correct, can it to

[29] See Alldridge, *ibid*
[30] For example, in English law, under s 69, PACE, 1984, see above

be relied on? Or when a computer invites a customer to view pictures and details of goods for sale, creates an order form for customer completion, accepts the order and authority for payment via credit card and says the "goods are on their way" - again, why should we rely on it?[31] The answers here will include judgements about the accuracy of the data input process as well as the reliability of the functionality of the computer hardware and software. In other words, this is the same proto-test[32] as would be applied if there were only one computer involved and what was in question was simply straight-forward print-out. There is a difference, though: normally a certificate of normal or proper working would be issued by someone familiar with the operation of the computer in the sense of knowing what the computer is required to do[33]. It is not at all clear how far a remote user of a computer could be said to be "familiar" with it.

- **Provenance of Computer Source Test** How can we show that data has been obtained from a specific computer and nowhere else?
- **Content / Party Authentication Test** Can we link the material from the remote computer to the person accused or party to the civil proceedings and the events that are the subject of the legal proceedings? Given the volatility of computer files, acquisition also usually needs to be linked to a specific day and time. This level of authentication cannot be done within a purely technological / computer context but will require other forms of evidence such as witness statements, exhibits indicating ownership of or access to the computer and/or data media, or the possibility of inference from the nature of the content of the files.
- **Acquisition Process Test** Can we give a full and believable explanation of the processes by which the file was acquired from the remote computer to the user's machine to show that the result is accurate, free from contamination and complete?
- **Continuity of Evidence / Chain of Custody Test** Can we give a full account to explain what has subsequently happened to the material retrieved? Typically it will have been acquired on a PC, subsequently viewed, perhaps analysed and probably copied several times, if only to provide copies for counter-parties and experts. How was the evidence frozen and rendered tamper-proof?
- **Quality of Forensic Presentation Test** Once the material has been obtained by investigators: Has there been any subsequent processing, such as retrieval from archive formats or examination via an application program? Have any special analytic tools been used? If files have had to be decrypted - how was that achieved? On the basis that print-out is being offered in evidence, how was the print-out obtained? How accurate and objective were all these methods - and how far can they be said to be generally accepted?

[31] And, in the case of such forms of electronic commerce, where is the trail of evidence showing offer, acceptance and consideration and where exactly do the various obligations pass the point of non-repudiation?

[32] A proto-test because within it there are a number of further subsidiary criteria to be tested

[33] In R v Shephard [1993] 2 WLR 102, Lord Griffiths laid down the UK test for admissibility: "Computers vary immensely in their complexity and in the operations they perform. The nature of the evidence to discharge the burden of showing that there has been no improper use of the computer and that it was operating properly will inevitably vary from case to case. The evidence must be tailored to suit the needs of the case. I suspect that it will very rarely be necessary to call an expert and that in the vast majority of cases it will be possible to discharge the burden by calling a witness who is familiar with the operation of the computer in the sense of knowing what the computer is required to do and who can say that it is doing it properly… The computer in this case was of the simplest kind printing limited basic information on each till roll."

Services from Remote Computers

The first and third of these tests - remote computer's correct working and content/party authentication - raise difficult but scarcely new issues. We can see the problems of the others by returning to the practicalities of what happens when a customer - "user" - interacts with a remote computer and how investigators typically operate. First, an indication of the range of services typically available and the vocabulary employed:

- **Bulletin Board Systems** (BBSs) first appeared informally on mainframes at the beginning of the 1970s as a way of providing an electronic noticeboard where messages could be left for public or private consumption and as a repository for program files that people might find useful. By the end of the 1970s BBSs were also being hosted on PCs (personal computers) and accessed via the public phone network (PTSN) using modems. A substantial hobbyist movement flourished until 1995 or so, when much of the activity transferred to the Internet. BBSs were also used in business and professional environments and survive as part of workgroup products such as IBM/Lotus Notes and on Intranets, which are networks internal to individual organisations but using Internet-like protocols, tools and software.
- **Conferencing Systems**, of which CompuServe, AOL, the Well and CiX, are examples, are BBSs grown, to one extent or another, large and commercial. At the time of writing, CompuServe has 4m members world-wide and AOL 7.5m. Payment is usually based on time-connected. Two important features are the numbers of individual forums or meeting rooms given over to a vast range of highly specialist topics and the extent to which commerce, in the form of electronic shopping malls, is well established.
- **Online Hosts** provide facilities for the publishing of electronic archive material, initially scientific bibliographies which could be searched with boolean operators, but now feature the full texts of many newspapers, magazines and journals as well as archived law reports, market price data, company returns and investment analytic reports. The oldest and largest is Dialog, originally owned by Lockheed, now part of Knight Ridder; Profile is part of the Pearson / *Financial Times* group. Dialog in particular carries the databases of several hundred smaller individual publishers. Payment is usually on a combination of subscription, quantity of data downloaded and connect-time.
- The **Internet** is essentially a network - a means of communication - but computers connected to it can and do publish, these days using the **World Wide Web** of pages of text and images and hypertext links. But it is also possible for whole files to be transmitted without being immediately viewed or used - normally by the File Transfer Protocol - **FTP**.
- **Online Banking** is typically available to small businesses and individuals via the public phone system[34]: the user's PC connects to the bank's computer and is able to interrogate it about transactions and also authorise payments.

[34] And also, in some cases, via the Internet

Acquisition Process Test

In nearly all instances the user[35] is operating from a PC; typically the services are being accessed via the phone system[36]. The user's PC has a modem and there is another, to answer it, on the remote computer. [37] The user's PC also needs a piece of software to mediate the connection. For the more sophisticated and commercial services specially written specific software is provided; the user sees a single on-screen interface and follows a series of menu'd instructions to acquire the desired remote services.[38] The software conceals the background technical processes of making phone calls and sending instructions to the remote computer. Online banking software also needs specific dedicated software, which additionally handles matters of authentication and the encryption necessary to keep the transactions confidential. To a certain extent this masking of computer mechanics is also effected for users of the Internet World Wide Web through "browser" products like *Netscape* and *Microsoft Internet Explorer*.

It is easier to understand what is going on - and we need this explanation if we are to understand some of the problems of evidential quality - by looking at the more generic forms of "comms" (computer communications) software. Such software has several tasks:

- it has to be in a position to send and receive data through an appropriate port on the user's computer, usually the serial (RS232) port
- it has to be able to send instructions to the modem[39] that is providing the link to the telephone line
- in the more traditional types of software it has to provide a **terminal emulation**, that is, make the user's computer screen look as much as possible like a traditional computer terminal directly connected to a mainframe or mini-computer. Typically the user's comms software presents a substantially blank screen, ready to receive messages, etc, with a small area for the user to input commands and check the status of the communications link. Before the wide-spread availability of PCs, users interacted with larger computers via so-called "dumb terminals" - devices with keyboards and screens but no processing power of their own. Each major computer manufacturer produced its own range; the simplest displayed lines of characters one after each other, like an old-fashioned teletype, but by the 1970s terminals able to drop characters into the middle of on-screen forms ("cursor-addressing") and to handle simple graphics were common. PC software now provides the same facilities and can usually support many of the historically popular terminal emulations such as VT-52 and VT-100.[40] Under terminal emulation, what the user at the end of the telephone line sees is a series of "screens" consisting of data sent to it by the remote computer. When the user sends instructions, what appears on his

[35] For this purpose the "user" is the person collecting the evidence, either as part of a normal business activity or for some specific investigatory purpose

[36] Some business and academic users of the Internet have direct connections via leased lines but this does not invalidate the arguments made here though there may be an impact in terms of the drawing up of warrants for interception; in the UK the Interception of Communications Act, 1985, is limited to public telecommunications services.

[37] A modem converts data signalling for transmission on telephone voice-circuits; most large remote computer services have banks of modems attached to lines so that they can handle many simultaneous callers.

[38] For example, CompuServe customers use *WinCIM* (Windows CompuServe Information Manager), FT Profile offers *FreeWay.*

[39] Or ISDN terminal adapter

[40] These two were originally terminals designed by Digital (DEC).

screen is not, as might be thought, the immediate result of typing on the keyboard but an "echo back" from the remote computer

- various forms of **error correction** are in place, to avoid the effects of noise on telecommunications lines. At its least complicated, when the user is sending and receiving simple text-based material, error-correction is on a per-character basis - "parity" checking. In reality true error-correction doesn't happen; the scheme is only good enough to make errors manifest, so that surrounding material is not relied on. At higher communication speeds, which in turn involve more sophisticated modems, the modem hardware may also provide an element of error-correction without the intervention of the user or the comms software.[41] More complex error-correction is possible in file transfer and on the Internet - see below.
- an important feature, especially significant to the issues of evidence collection, is **logging**[42] This opens up a file on the user's computer which records all characters received by the terminal. It is often possible to "play back" a previous online session, hence the potential value as an audit trail. There are some limitations, though:
  - the starting and ending of any log-file is wholly under the control of the PC user so that the provenance of any log file (and indeed its admissibility) have to be separately established[43]
  - such files are usually very simple in format so that subsequent alteration, say by using a basic word-processor, is both easy to carry out and difficult to detect
  - since the log captures *all* characters a certain amount of "junk" will appear. At the very least all typing mistakes made by the user (plus his attempts at correction) will be seen. Most log files are unable to reproduce the more complex terminal emulations; the cursor addressing instructions are captured in the file but cannot be made to play-back properly - junk characters are interpolated with the original on-screen text. It is obviously possible to clean the junk characters out by post-event editing, but in so doing you run the risk of tampering with the evidence.
  - a log file can also be sent direct to a printer, and this may have some value in terms of preserving the evidence (considered further below). But here the problems of junk characters are even more acute. In addition, at a purely operational level, if the printer log is truly simultaneous the overall speed of receiving characters from the remote computer will be reduced to that of the printer, which will inevitably be slower than when characters are merely displayed on screen and recorded to disk.
  - larger computers owned by end-users may have further, independent facilities for logging activity, often associated with the computer's operating system[44]. These logs operate outside the comms software and tend to record all *user*

[41] For example, V42 / V42bis, common in modems operating above 9600 bps, features both error correction and data compression within the hardware

[42] This sometimes referred to as "capturing"; the absence of standardised terminology within the computer industry is a nuisance, particularly as "capture" can also refer to "screen dump", considered below.

[43] Operating systems usually provide and day-and-time stamp on files, signifying the last occasion on which the file was "opened", that is altered in some way. However the day-and-time stamp in turn depends on the accuracy of the system clock; in any event there are several relatively simple ways to alter a day-and-time stamp

[44] This is more typical of systems which support several simultaneous users and local area networks rather than stand-alone personal computers

activity rather than characters received at the screen. The evidential value of this type of internal log depends on what precisely it was recording

- the other important feature is **file upload and download**, the facility for computers to exchange files direct to each other's hard-disks. Uploading is sending a file, downloading is receiving a file. To initiate the process, the user has to get both the remote computer and his own machine into the right states by following a series of on-screen menus. Simultaneous commands on both machines lock them together so that the file is transferred.[45] The actual transfer is handle by an error-correcting protocol. There are several in popular use and they are all capable of genuine error correction as faulty "blocks" or sections are identified and requests for re-sends generated. The various protocols vary slightly in their internal method and transfer efficiency. One benefit from an evidential point-of-view is that the protocols are robust - a transfer may fail if the transmission line is poor, but incorrect though plausible transfer of a file is almost impossible.

The sophisticated software referred to at the beginning of this section has much the same features, except that the user can't see them and that logging may not be possible.

As can be seen, the immediate process of acquisition of files from a remote computer is usually reliable, because errors are either detected or corrected. The main problem comes from log files, which are often used to record text-based materials, say from online services.

There are three further routes available to the user who wishes to record what he has seen on his screen.

- A **screen dump,** as its name implies, involves the saving of the contents of whatever is on a computer screen at a specific time. On older, simpler computers where the main output is text, there is often a "print-screen" key or facility: the screen contents are sent to a printer. More recent computer operating systems (for example, Windows and MAC) tend to be graphics-based and screen dumps require an additional program operating "in the background" which produces a graphics file.[46] The limitations are as follows:
  - all that is captured is what was on the screen at one specific time; the procedure cannot show a sequence of events; provenance as to circumstance and time will usually have to be established extrinsically, usually by way of statement
  - the whole of the screen is captured so that, where a comms program is being used, what is printed out or saved is both the information coming from the remote computer but also any status messages created by the comms program itself
  - in the case of "print-screen", the process may be imperfect and some characters seen on a screen may look different, or corrupt, when they emerge from the printer
  - in the case of graphics dumps saved to file, the resulting file is itself capable of subsequent editing so that the provenance of that file will also have to be separately established

[45] Some protocols, such as Zmodem, have the capacity to auto-start on the receiving computer once an appropriate data stream is detected.

[46] Paint Show Pro is one such example for Windows - the main uses of screen dumps are in computer instruction manuals and where particular graphics images cannot otherwise be transferred from one format to another.

- A program called **Lotus ScreenCam** overcomes some of these problems. It is set to operate in the background and can be used on computers with the Windows family of operating systems. In effect it is a very superior form of log file, in that it can capture *all* on-screen activity including exact mouse movements, opening and selection of menu items, and so on. [47] As with the more normal sort of log file, the starting and ending of a ScreenCam file is wholly under the control of the PC user so that the provenance of any file upon which it is intended to rely (and indeed its admissibility) has to be separately established. Subsequent editing does not appear to be that easy though misrepresenting the circumstances in which one was obtained, is relatively easy.
- Finally, it is also possible to use actual **video recording**; devices are available to convert the ordinary video monitor output of a computer (known as VGA on PCs) into the electrical requirements for display on a conventional television. These devices can also be fed into domestic or semi-professional video recorders and indeed are principally sold to enable the production of video-based sales aids and other presentations for medium-sized audiences. In an evidence-collecting situation, the video tape has almost the same advantages and disadvantages as the VideoCam program except that:
    - the tape is recorded on a device wholly separate from any computer and may therefore avoid the problems of admissibility and computer reliability
    - post-event editing is possible
    - the quality and detail may be relatively poor so that smaller type fonts maybe difficult to read; freezing an individual video frame may result in very disappointing results

In all of these circumstances, the "proper", "normal" or "correct" working of the user's computer will have to be asserted or demonstrated, if not for reasons of admissibility[48] then in order to give confidence in the results obtained.

Internet-specific issues

The foregoing descriptions apply to computers that talk to each other direct along telephone lines. Since the mid-1990s, though, for reasons of convenience and cost, the Internet has become the most favoured transmission medium and many of the commercial and non-commercial services referred to above have now appeared in Internet form.[49] In addition many wholly new services and facilities have appeared and proved themselves significant. All of these give rise to the potential for legal proceedings and we need to know about possible sources of evidence.

As we will shortly see, the real problems relate to authentication of computer source, but first we need to assess the impact on the other issues of evidence quality arising from the ways in

[47] Lotus ScreenCam operates in a similar fashion to video recorder. Its main use is to train people in the running of specific computer programs: the trainer starts the screencam process, follows a sequence of activities which are then recorded and then stops the process. It can also separately record a voice commentary. The result is a file which can be given to pupils for their own repeated personal training sessions.

[48] Eg s 69 PACE, 1984. The user provides a certificate upon which he can be cross-examined.

[49] Even before the Internet broke through from its academic and military roots and became easily available to businesses and homes, companies have, since the early 1980s, been making extensive use of private networks, usually operating to the standards known as X.25.

which Internet-based communication diverges from the telephone-based direct computer-to-computer link.

The technical transmission protocol underlying the Internet and the services which mimic it is called TCP/IP and the principles if not the detail are similar for other types of wide area network. Data transmissions between computers are divided into a series of small segments or packets as this makes the most economical use of the physical cables that link the various constituent computers together. The packets contain the actual electronic digits of the transaction or the message the computers wish to exchange, facilities for error-correction, and a numbering scheme to make sure that if packets arrive out-of-order (one of the strengths of the Internet is how, if one physical route has failed or become congested another route is automatically selected), the receiving computer knows how to reconstitute the original.[50] In addition the packets carry addressing information - so that each packet knows its destination and each computer knows which packets are meant for it and it alone.[51] If one were to place a surveillance device across a major Internet cable or "pipe" one would be presented with fragments of several hundred thousand simultaneous computer-to-computer conversations - it is the addressing data in each packet that identifies each individual "conversation".

The presence of error correction means that potential evidence in the form of files received via Internet communications can nearly always be shown to have passed the immediate Acquisition Process Test. If an Internet link fails, the first thing that happens is that the Internet automatically looks for an alternative route; the second is that the speed of transmission drops, a phenomenon alas too familiar to regular Internet users. Eventually the transmission may stop prematurely. But the protocol effectively prevents the transmission of corrupted data.

The immediate analogues for the telephone-based services we have been considering are:

- **Telnet,** the facility to allow a user on one computer to use another directly, as though they were sitting at a terminal local to that computer. This facility is relatively little used outside the research community and the computer industry
- **FTP**, File Transfer Protocol, which provides the Uploading and Downloading described earlier, but in a specific Internet context

A computer user can also turn on logging, so that critical sessions are recorded to a file.

But the Internet has also spawned the **World Wide Web** (WWW)and this now provides the interface through which most developing commercial and domestic Internet activity takes place. It has moved on from being a mere publishing vehicle to providing facilities for the sale of a wide variety of goods and services; increasingly, individuals and companies can carry out bank transactions and trade investment instruments. To understand where evidence of activities might exist, we need to know a little of what is going on beneath its glossy surface.

The user of the World Wide Web views it through a piece of software called a browser; at the time of writing the two leading products are Netscape and Microsoft Internet Explorer. The Web appears as a series of colourful text-and-image pages[52]. Certain elements on the pages are highlighted and offer the user more specific, or related information. The user clicks

[50] The error-correction is similar to the various protocols used for file transfer between computers talking "point-to-point" on the PTSN

[51] In fact the packets contains the address of the source computer as well as the destination and also other types of "status" data

[52] It is also possible to include moving pictures and sound

his mouse on these elements and a further page appears. The process of moving around in this way is called navigating and the linking scheme is known generically as *hypertext*. Any page on the web can be hyper-linked to any other page on any of the several million computers connected to the Internet.

The pages exist as files on various computers and are prepared using a "mark-up" language which includes both the text and instructions where any pictures are to be inserted (and where the pictures are to be found); the browser software interprets the mark-up language to produce the screens the user actually sees. The hyper-text links also form part of the mark-up language; they include the specific instructions required to route the user's computer to the precise location on the Internet-linked computer where the next set of pages are to be found, and to load them into the browser. Thus the WWW user, having found a suitable start point, does not have manually to type in too many complex instructions.

What happens if we need to obtain a page for evidential purposes? [53] The immediate process is even easier than using the file download procedures described earlier and the error-correction facilities inherent in TCP/IP ensure freedom from immediate error. Facilities within the browser software also offer to allow one to save the result to disk as well as view immediately on screen. Unhappily, the reality is more complex. The designers of the World Wide Web - and the concept is in a state of rapid development - have built in various features which while convenient for ordinary use create problems if it is desired to collect reliable evidence:

- some of the facilities within the browsers to save WWW pages to disk are imperfect; text may be saved but not associated images; again, with some very complex pages[54], what is seen on screen and what is saved to disk may be quite different
- the method used to save a file to disk may not carry any individual labelling which shows where and when it was obtained
- such saved files are very easily modified or forged; accidental alteration is also a substantial hazard
- browsers also have a facility called a *cache*, a temporary holding area for recently accessed pages so that a user can view such pages without having to wait for the delays involved in going back each time to the actual source computer[55]. The cache itself is saved to disk, which means that it can be viewed later[56]; but there is no immediate, bullet-proof way of telling when a specific page was last acquired. Thus if a whole series of cached pages are examined an entirely false picture could be built up - the pages are almost certainly not contemporaneous. Moreover, during a live session, what is produced on screen could be a mixture of pages immediately acquired from the remote computer and others acquired earlier on. Thus even the recording of a live session (say by using Lotus ScreenCam) could be misleading.
- caches exist elsewhere. Many Internet Service Providers (ISPs), those who provide the links between individual users and the major high-speed communications cables, have facilities are known as proxy servers to speed up the delivery of popular pages and overcome congestion problems on the network. Thus a customer of such an ISP may not be able to be sure that what he has received on his computer is the latest

[53] Graham Smith analyses the "The Web - the Challenge of the Virtual Document" from the perspective of copyright law in Graham J H Smith, 'The Future of Intellectual Property in an Online World' *Computers and Law*, Vol 7 Iss 2, June/July 1996 pp33-37
[54] eg involving "frames" and "templates"
[55] eg Index pages are frequently re-called as users navigate around a particular site
[56] eg via such programs as *MSIE Cache Explorer* and *Secret Agent*

version from the source computer as opposed to an earlier cached version held by his ISP.

When the WWW is used as an interface for electronic commerce further problems appear. The pages of instruction which are converted into the pages the user actually sees have often been created on-the-fly by the remote computer, which itself may be linked to a further conventional "accounts", "catalogue", "sales/invoice" or "retail bank" computer. Thus no immediate complete record of what the user actually saw may exist at the remote computer either, though presumably for many purposes the owner of that computer would like to be able convincingly to demonstrate to others the terms and facts of a transaction. [57]

Provenance of Computer Source Test

To return now the tests of evidential reliability: When a remote computer has been reached on the PTSN by dialling a specific telephone number it is a reasonable assumption that the computer which answers is what it purports to be: the phone number can usually be linked to the address of an organisation or individual[58]. The issue of the authentication of the computer source in these circumstances is entirely straight-forward, even more so if the communications link is a dedicated leased line. But the whole idea of the public Internet, and indeed the private "packet-switched" or "X.25" services is that large numbers of computers, tens of millions in the case of the Internet, co-exist on the same network. Provenance of computer source, therefore, has to be shown by some element in the procedures by which two individual computers interact.

On the Internet, addressing information in the packets provides, alas, only a provisional degree of proof of source. There are two problems, and they apply equally to FTP and WWW activities:

- **genuine site attacked, false information created by another** Hackers have been able to attack websites, having acquired usernames and passwords by means of "packet-sniffing". That has enabled them to introduce "alternative" pages to the official ones; the phenomenon has now become common-place.[59]
- **addressing scheme within packets faked, information appears to be coming from one source but in fact comes from another** Here, there is an attack on the packets themselves means that addressing and other status information cannot be relied on[60].

In commercial transactions reliance is usually placed on additional controls, which can vary from the same simple credit-card-authorisation-plus-insistence-on-delivery-to-credit-card-

[57] The trend is towards special Internet-specific protocols such as SET but these are directed towards authenticating the transaction, not to showing the totality of the contract - what specifically was offered, and on what terms.

[58] Eg by identifying who pays the telephone bill and/or who occupies the premises

[59] Eg The Labour Party, *Guardian* 10 Dec 1996; CIA, 30 Sep 1996, USAF 30 Dec 1996, etc The weakness that is exploited arises from the fact that most websites are updated remotely: the HTML pages are created on computers other than the one hosting the site and are FTP'd over the Internet. Packet sniffing enables hackers to identify packets carrying FTP requests destined for the website and which carry sequences associated with log-ons and passwords.

[60] For an extended discussion see Cheswick, WR and Bellovin, SM *Firewalls and Internet Security: Repelling the Wily Hacker*, Addison-Wesley, Reading, Mass, 1994, *passim* but especially pp 163-164

holder's-registered-address which is used in telephone-based credit card transactions all the way through to multilevel protocols such as SET. Formal EDIs[61] and schemes for securities trading also create large volumes of extrinsic audit trails.

Continuity of Evidence / Chain of Custody Test

Thus far we have been seeing how data arrives from the remote computer to the user's computer, but more has to be done before the result can become an Exhibit. Except in those cases where an immediate record has been sent to a printer, what now exists on the investigator's computer is a disc file, or perhaps several. In any subsequent printing out or file copying there are considerable dangers of alteration, inadvertent or otherwise. As for any other types of physical evidence, what is needed is continuity right up to the point where an Exhibit is presented in court; sometimes this is referred to as "chain of custody". Continuity of regular evidence - devices, weapons, physical and biological specimens etc - is normally addressed by way of statements from officers at a scene of crime or raid, by "bagging and tagging" and the maintenance of custody records which show each time an item has been examined.

As before, the issues are more readily understood if one views the processes stage-by-stage:

- **the investigator's computer** ideally should have some of the sterile qualities of a test-tube or sample bag, except that it can't: at the minimum the computer will need to have an operating system and comms software of various kinds. The emerging practice seems to be to save any downloaded files into a newly created directory or "folder" on the hard-disk[62]; most forms of external storage media are too slow for convenience when downloading. The only control that exists at this stage is the detail and probity of the investigator's accompanying statement, if it exists
- thereafter, almost **immediate copying to external storage** is obviously desirable as a way of "freezing" the evidence and releasing the user's computer for other purposes. The commonest form of external storage currently is the floppy disk, write protected.[63] However in the case of 3½ inch disks, this is simply a small shutter which can be opened and closed at will and without audit trail. For larger quantities, CD-ROM, which is a write-once-read-many device (WORM), is used and also offers greater protection[64]
- Nearly always further copies will be needed for other investigators, experts, prosecutors and the defence team: it makes good sense to assign the earliest copy as the "control"[65].

[61] Electronic Data Interchange

[62] If the material is downloaded into an existing directory there is the danger that files from several different sessions, perhaps even from different computers in different cases, become confused.

[63] A physical barrier prevents further writing to the disk.

[64] An additional safeguard, discussed but not so far used, is the use of public key cryptography to add the investigator's digital signature to the captured file; the aim would be both to authenticate the file and provide a technical guarantee of the absence of subsequent alteration. However there are as yet no standards or products to do this, so that the proposed solution may create problems of its own. Some of the issues of freezing and subsequently authenticating documents in general are considered in *Code of Practice for Legal Admissibility of Information Stored on Electronic Document Management Systems* British Standards Institution DISC PD0008, 1996

[65] As can be seen, there are analogies to be made with what happens in tape-recorded police interviews under the Police and Criminal Evidence Act, 1984 - Code of Practice, Section E. See also: Bevan, V

The extent to which these computer files are vulnerable to unauditable alteration prompts one to suggest that very high standards of record-keeping are required. [66]

Quality of Forensic Presentation Test

At this point the evidence is still in electronic form and will usually need to be converted into something which a court can handle, which typically will be print-out. In addition, further processing may be regarded as desirable:

- log files may contain "junk" characters which impede understanding of the content - how can one demonstrate the neutrality of any clean-up exercise?
- audit trails may be extremely lengthy and detailed - some form of editing may help a court understand a sequence of events, but can the editing ever be "objective"?
- many files available for download from BBSs and Internet sites are saved in a compressed format - usually to save space on the host computer and to reduce download times. The best-known format in the PC world in ZIP. ZIP files need to be uncompressed before use; this is a simple process but not always a completely mechanical one[67] and the integrity of each such event really should be proved
- World Wide Web files saved to disk do not necessarily automatically print-out correctly and some post-event editing may be needed

In some instances it may be appropriate to offer *two* exhibits - "raw" version and then an edited one, together with some narrative to explain what was done and why. This could then be subject to cross-examination. Perhaps the simple answer to these problems is to recognise that all these exhibits, even the very simplest print-out, are no more than purported derivations of what was originally obtained in electronic form and they need to proved by detailed statement[68].

Conclusions

The need for acquisition of reliable evidence of the existence, content and provenance of data from cyberspace prompts the development of a highly disciplined approach by investigators, auditors, lawyers and computer mangers. Formal electronic commerce will undoubtedly seek to resolve some of the problems by devising protocols which bind together computer-based counter-party authentication[69], audit trails and precisely defined contractual frameworks. But

---

and Lidstone, K, *The Investigation of Crime: A Guide to Police Powers,* Butterworths, London, 1991 at pp 446-454

[66] In the English courts, it would appear that certificates under s 69 PACE, 1984 would be required in respect of computers involved in each of any copying processes

[67] For example, a ZIP file can itself contain many files and also preserve an original directory structure: detailed knowledge of the directory and file structure of the computer upon which the uncompressing takes place is therefore essential if there is to be confidence

[68] In practice there are many other issues of the forensic examination of computer-derived materials, including the handling of disk exhibits, the retrieval of apparently deleted material and how to handle encrypted files, but these are beyond the scope of the present article

[69] For example, by using public key cryptography

a great deal of computer usage will remain informal and in these circumstances perhaps all that can be offered are something based on the series of tests examined in this article:

- Remote Computer's Correct Working Test
- Provenance of Computer Source Test
- Content / Party Authentication Test
- Acquisition Process Test
- Continuity of Evidence / Chain of Custody Test
- Quality of Forensic Presentation Test

We can also present a more pragmatic and informal set of questions:

- Where did the exhibit originally come from?
- What intermediate stages were required to produce the exhibit as it is now being shown?
- What computers were involved at each stage? Which was the source computer, the investigator's computer, what computers were used for intermediate storage and processing
- Who was involved in these intermediate stages and what did they do?
- How detailed and plausible are the investigator's statements in support?
- Are there additional items of evidence which provide corroboration?

The speed of change in the use of computer technology tends to operate against the possibility of devising formal or "standard" evidence collection and presentation protocols. By the time such protocols have been through a "general acceptance" [70]test, the technology in actual use will have moved on. It may be that these relatively loose tests are the best we can reasonably expect. On the horizon are yet more technical tools for investigators. It is possible to eavesdrop activities across the Internet: a technician at one terminal can monitor all the keystrokes between two other computers and, it is claimed, re-assemble them. One US case has already come to public attention but is unlikely to be tested in the courts[71]. Keystroke monitoring was also used in the UK case of R v Pryce and Bevan, where two young men were accused of unauthorised access into large numbers of computers owned, *inter alia*, by the United States Air Force and the defence company Lockheed.[72] What further tests of provenance are we likely to need as such methods become more common? And what further powers will law enforcement authorities require in order to carry out some of the newer kinds of surveillance?[73]

[70] eg under rules similar to the *Frye* procedures, Frye v US 293 F 1013 (DC Cir. 1923)

[71] The case of Juan Cesar Ardita, a 20-year-old Argentinian accused of unauthorised access into a number of sensitive US computers and whose activities were tracked using a software tool called iWatch. There is currently no relevant extradition treaty between Argentina and the US. *See* UPI: 29 March 1996; Washington Post 30 March 1996, San Jose Mercury News 30 March and 8 April 1996 and Newsbytes 5 April 1996.

[72] The events took place in 1994 and the monitoring tools were called Stethoscope, Network Security Monitor and Pathfinder. The cases finally came before the UK courts in 1997; the author acted as defence expert.

[73] At a conference held by the National Criminal Intelligence Service in London on May 28 1997, NCIS speakers indicated that they would be lobbying for new types of warrant which would enable them to require Internet Service Providers to provide access to client traffic.