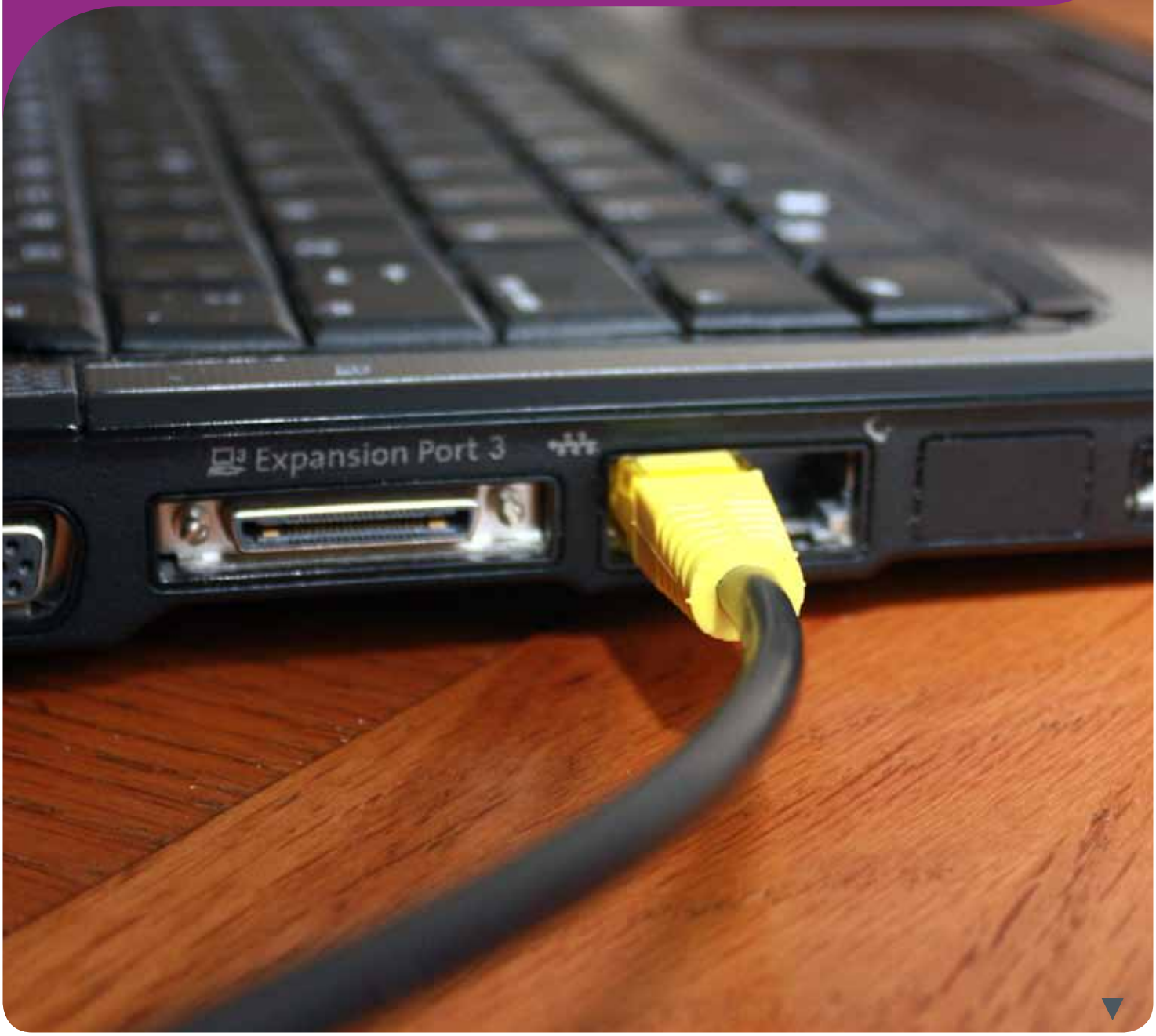
Consumer Focus
Campaigning for a fair deal

Online traceability: who did that?

Technical expert report on collecting robust evidence of copyright infringement through peer-to-peer filesharing

Dr Richard Clayton



About Consumer Focus

Consumer Focus is the statutory consumer champion for England, Wales, Scotland and (for postal consumers) Northern Ireland.

We operate across the whole of the economy, persuading businesses, public services and policy-makers to put consumers at the heart of what they do.

Consumer Focus tackles the issues that matter to consumers, and aims to give people a stronger voice. We don't just draw attention to problems – we work with consumers and with a range of organisations to champion creative solutions that make a difference to consumers' lives.

Following the Government's consumer advocacy reforms, we will continue to act in the consumer interest across a wide range of sectors until our general advocacy role passes to Citizens Advice in April 2013.

As part of the reforms, Consumer Focus will establish a new unit to identify and represent consumers' interests in complex, regulated sectors, including energy and postal issues and, in Scotland, water.

Our Annual Plan for 2012–13 is available online, **consumerfocus.org.uk**

For regular updates from Consumer Focus, sign up to our monthly e-newsletter by emailing enews@consumerfocus.org.uk or follow us on Twitter http://twitter.com/consumerfocus

About the author

Dr Richard Clayton, Computer Laboratory, University of Cambridge

Dr Richard Clayton is currently a Senior Research Assistant in the Computer Laboratory of the University of Cambridge (a 'post-doc researcher'). Dr Clayton has a particular research interest in 'traceability' – the determination of 'who did that' on the Internet.

Dr Clayton worked at Demon Internet, then the largest UK internet service provider, from 1995 until 2000. In October 2000 he took up the opportunity to go to Cambridge and study for a PhD. His doctorate was awarded in January 2006 for his thesis, 'Anonymity and Traceability in Cyberspace'. Substantial parts of this thesis dealt with the practical issues that arise when attempting to determine which internet user is responsible for a particular event.

Dr Clayton continued to work in the Computer Laboratory doing research into various aspects of computer security. He has also acted as specialist adviser to House of Lords and House of Commons Select Committees in matters to do with internet security and the security of Internet users. He has written, or co-written, over 40 peer-reviewed professional publications. Dr Clayton has acted as an expert witness in several criminal and civil court cases, being instructed on some occasions by the defendants and in others by plaintiffs or prosecutors. Some of these cases have been concerned with file sharing activity.

© Crown Copyright 2012

Open Government Licence: This report are subject to a worldwide nonexclusive Open Government Licence to enable and encourage the free use of public sector information covered by Crown copyright and database rights. The licence does not cover the stock photos included in this report. All existing Consumer Focus copyright statements are superseded by the Open Government Licence.

You are free to copy, publish, distribute and transmit the report; adapt the report; or use the report commercially. Under the terms of the Crown Copyright Licence you must, where you do any of the above, acknowledge the source of the report by attributing this report as follows:

Dr Richard Clayton (2012) Online traceability: who did that? Technical expert report on collecting robust evidence of copyright infringement through peer-to-peer filesharing. Consumer Focus.

The original report (**http://bit.ly/MFPLte**) as provided by Dr Richard Clayton to Consumer Focus in April 2012 is also subject to an Open Government Licence.

Foreword

Over the past decade, consumers have gained the technological means to carry out actions restricted by copyright with ease. Digital technologies allow consumers to copy and modify copyright protected works, and internet connectivity allows consumers to communicate copyright protected works to others. The Digital Economy Act was passed into law in 2010 in response to consumers infringing copyright by 'sharing' copyright protected works with each other on peer-to-peer filesharing networks.

Consumer Focus is working to ensure that any enforcement action against those who are alleged to have infringed copyright is fair and proportionate, respects their legitimate rights to privacy and follows due process. Moreover we are working to reform copyright exceptions and licensing to update copyright law for the digital age so that it supports economic growth and balances fairly the interests of consumers, copyright owners and creators. While there is a role for enforcement, addressing copyright infringement by consumers on a long-term basis requires copyright law reform so that it has legitimacy in the eyes of consumers. In turn, copyright licensing needs to support innovative legal markets which respond to technological developments to meet consumer demand in a timely manner at a reasonable price.

This report, written by Dr Richard Clayton, outlines how copyright owners can collect robust evidence of copyright infringement through peer-to-peer filesharing. Consumer Focus commissioned this report to assist Ofcom in the implementation of the Digital Economy Act 2010 through a statutory Initial Obligations Code. When it comes to taking action against people accused of infringement, the standards of evidence are critical. The Digital Economy Act 2010 requires that the Initial Obligations Code makes provisions on the 'means of obtaining evidence' and the 'standard of evidence' for copyright owners who want to lodge 'copyright infringement reports' against consumers with their Internet Service Provider (ISP).

UK courts have yet to fully test evidence of copyright infringement through peer-to-peer filesharing by consumers. So far the cases which have been brought against consumers in courts have either settled or been determined at the summary judgment stage. Therefore there is currently no authoritative guidance on how copyright owners can collect evidence of online copyright infringement which on the balance of probability prove that an infringement has been committed on an internet connection.

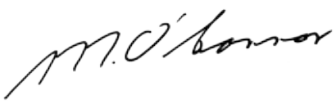
Internet access is commonly shared within a household, typically through WiFi. Increasingly, the same internet connection is used by multiple individuals using multiple computers or internet enabled devices. In the two years since the Digital Economy Act 2010 passed into law, mobile broadband coverage and usage has increased dramatically. Large-scale public commercial WiFi providers are moving to provide internet access in city centres and public transport networks. Private businesses such as hotels and pubs increasingly provide internet access as standard. Indeed public bodies and private businesses such as libraries and internet cafes provide essential internet access to consumers who live in the 20 per cent of UK households which do not have internet access at home.

This report provides advice on standards and procedures which should be adopted to ensure that copyright owners can reliably identify an internet connection which has been used to infringe copyright through peer-to-peer filesharing. Dr Clayton then describes how ISPs can robustly match internet subscriber details to IP addresses, which are dynamically allocated to domestic internet connections. Under the Digital Economy Act 2010 subscribers, who are the bill payers for an internet connection, can appeal a notification of alleged copyright infringement if they can show that they did not commit the alleged infringement, and took 'reasonable steps' to prevent others from infringing. Dr Clayton therefore concludes his expert report on traceability by assessing how subscribers to an internet connection could identify who may have used their connection to infringe copyright.

Dr Clayton's finding that the subscribers will not be able, on a technical level, to determine which computer in a household was used to infringe copyright, or identify the individual at the keyboard, raises serious questions about whether the Digital Economy Act appeals process can operate fairly.

In order to assist the ongoing technical and legal debate on traceability online, and in particular the detection of online copyright infringement by consumers, we publish this report with only slight editorial amendments under an Open Government Licence. The un-amended expert report by Dr Clayton as provided by Consumer Focus to Ofcom is also available online.[1]

I would like to thank Dr Clayton for providing Consumer Focus with expert advice on online traceability. Consumer Focus has greatly benefitted from his technical expertise, as well as his patience. I believe that the reader will benefit from Dr Clayton's ability to explain complex technical processes to the uninitiated and I hope this report is a useful addition to the debate in this important area at a time when it seems that more and more of our life is lived online.



Mike O'Connor
Chief Executive

[1] **http://bit.ly/MFPLte**

Contents

Abbreviations

ADSL – Asymmetric Digital Subscriber Line
BST – British Summer Time
CGN – Carrier-grade NAT
DMCA – Digital Millennium Copyright Act
EC – European Commission
IP – Internet Protocol
IPv4 – Internet Protocol version 4
IPv6 – Internet Protocol version 6
ISP – Internet Service Provider
MAC - Media Access Control
MB – Megabit
NAT – Network Address Translation
NTP – Network Time Protocol
Ofcom – Office of Communications
P2P – peer-to-peer
PDT – Pacific Daylight Saving Time
RCS – Revision Control System
RIRs – Regional Internet Registries
SI – Statutory Instrument, also known as order, regulation or secondary legislation
TCP – Transmission Control Protocol
UTC – Coordinated Universal Time
WEP – Wired Equivalent Privacy
Wi-Fi – a wireless local area network, also known as WLAN
WPA – Wi-Fi Protected Access
WPA2 – Wi-Fi Protected Access II

Overall recommendations

The following is a summary of Dr Richard Clayton's recommendations as to the standards that Ofcom should set in relation to monitoring systems to ensure that the collection of the IP addresses of uploaders of copyright infringing material on peer-to-peer networks is robust and error-free.

Justifications and explanations can be found in #57 to #83 of the main report.

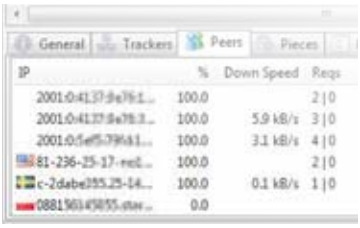
Preparation

- The monitoring machine must not be used for any other purpose than monitoring.
- The monitoring machine needs to be secured against unauthorised use. Its software should be configured to prevent unauthorised access to the machine and it should be kept up to date with all security-relevant patches.
- System logs should be regularly inspected to ensure that there is no evidence of intrusion.
- If an intrusion is discovered then no monitoring result can be relied upon.

- The monitoring machine should be running a Network Time Protocol (NTP) daemon synchronised to reputable time sources and must be capable of providing timestamps of events which are accurate to one second or less.
- The monitoring machine should provide timestamps in Coordinated Universal Time (UTC), ie +0000.
- Dates should be specified in ISO format, ie 2012-12-25.
- Before and after each monitoring run the operator should ask the machine what it thinks the time is – and compare that with the speaking clock, or some other independent and reliable source of time information. If monitoring is to run continuously, then the time should be checked on a daily basis. A specific contemporaneous note should be made that these checks have been made.
- The monitoring software that is to be used should be developed according to current best practice.
- The source code of the monitoring software should be held in a software revision control system (RCS).

- A contemporaneous note should be made whenever monitoring occurs as to precisely which version of the software is in use.
- A testing suite should be developed for the monitoring software that demonstrates that it is functioning correctly. This suite should be rerun whenever changes are made to the software (or the software is run on a new machine for the first time) to ensure that no inadvertent flaws have been introduced.
- The test suite output should be preserved in the RCS.
- When flaws are detected the test suite should be enhanced so as to check that the flaw has been fixed and is not reintroduced at any later stage.

Monitoring uploading

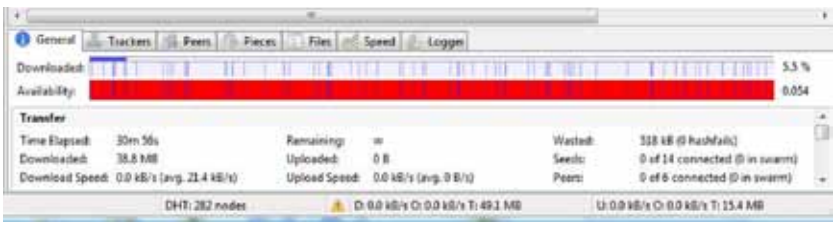


- The monitoring system should obtain a torrent file (or equivalent in other peer-to-peer systems).
- A specific contemporaneous note should be made of the details (so far as they can be ascertained) of the peer-to-peer system that is being monitored. This would include protocol version numbers, a copy of the tracker file (if Bit Torrent is involved) along with a record of where it was fetched from. These details should be placed into an RCS.
- The monitoring system should then proceed to use the peer-to-peer system to download a complete copy of the shared-file.
- It should be established that the shared-file that was downloaded is a copyright protected work and that, for example, the material has not been mislabelled.
- Cryptographic hashes of the various pieces of the shared-file should be calculated for use in checking that further downloads contain matching data.
- A specific contemporaneous note should be made of the identity of the shared-file that is to be monitored, how it was established that it was a copyrighted work, and a record made of the cryptographic hashes. The copyright protected file, and all the other information should be preserved within an RCS.

- Whenever the monitoring system fetches a piece of the copyright protected shared-file from an uploader, a contemporaneous record should be made of the time of the start and end of the Transmission Control Protocol (TCP) connection that was used. The remote IP address and remote TCP port number must also be recorded.
- Whenever a piece of the shared-file is downloaded, a cryptographic hash value should be calculated and – if it is correct – should be recorded along with the other details.
- If the tracker machine indicates that a piece of the copyright protected shared-file is available from a particular peer, but it proves impossible to make contact, or the data transferred does not exactly match the copyright protected shared-file, then no record should be made.
- All of the monitoring data, along with any event messages from the monitoring software that indicate its operation, should be written to a logging file. This logging file should be preserved within the RCS.
- The copyright infringement report given to the Internet Service Provider (ISP) should include the details of the peer-to-peer system, the tracker file (when relevant), the identity of the copyright work and the details of the piece that was downloaded. The time, IP address and TCP port number of the uploader should also be provided.

General

- A detailed description of every monitoring system design should be made available to the public. Ofcom should not consider 'secret' designs to be capable of creating reliable results.
- Ofcom should have a right to audit monitoring systems in order to ensure that any standards that they set out are being adhered to.
- A new monitoring system design should be audited before it is first used to collect data that will form the basis of a copyright infringement report.



Introduction

1 I was instructed by Consumer Focus, which wishes to provide advice to Ofcom as to what standards it could set for the collection of evidence of copyright infringement occurring in peer-to-peer file sharing.

2 Sections 3 to 16 of the Digital Economy Act 2010 establish a regime wherein online copyright infringement is detected by a copyright owner, or more likely a third party acting on their behalf. The copyright owner then submits a "copyright infringement report" to an Internet Service Provider (ISP).

3 The copyright infringement report states that there appears to have been an infringement of copyright and "includes evidence of the apparent infringement that shows the subscriber's IP address and the time at which the evidence was gathered". The ISP will use the IP address and time to identify the subscriber account and will notify the subscriber of the report.

4 Should the number of copyright infringement reports reach a threshold to be prescribed in the regulations that will instantiate the Digital Economy Act 2010 regime then the subscriber will be placed on a "copyright infringement list". It is anticipated that under the Initial Obligations Code the copyright owners will apply for a court order, known as Norwich Pharmacal Order,[2] to reveal the identities of subscribers on the copyright infringement list and this will lead to civil action against the person who has infringed upon their copyright.

5 Additionally, under the Technical Obligations Code, subscribers on the "copyright infringement list" are "relevant subscribers" for the purpose of "technical measures". The Digital Economy Act 2010 definition of technical measures includes a measure which "limits the speed or other capacity of the service provided to a subscriber" or one which "suspends the service provided to a subscriber".

6 In this report I set out the theoretical background to the detection, by third parties, of the IP addresses of computers that are engaged in copyright infringing peer-to-peer file sharing. I then give a detailed account of how such a detection system should be operated in practice so that it will be able to provide robust and unambiguous results. I also discuss how incorrect results can arise.

[2] A Norwich Pharmacal Order, named after the first case in which such an order was granted, is a disclosure order which compels an innocent third party, such as an ISP, to provide information (usually personal data which identifies an individual) relating to unlawful conduct so that legal action can be taken against a wrongdoer. The process by which an ISP matches an IP address to a subscriber account is outlined in #90 to #103.

7 Establishing relevant IP addresses is only the first step in the overall Digital Economy Act 2010 process. I go on to discuss what can go wrong when the ISP attempts to identify the subscriber and the problems the subscriber may face in identifying the individual who is actually responsible for the copyright infringement.

8 I also mention how file sharing systems have already evolved to counter interference by ISPs and how they evolve in future to mask the IP address of the connection being used for copyright infringement. I go on to suggest how future changes to their design might affect the reliability of the results that third party monitoring systems can obtain. If such changes occur, the monitoring procedures I describe are unlikely to still be appropriate.

9 This report only considers monitoring systems, situated elsewhere on the Internet, that monitor file sharing by the customers of many ISPs and then track specific instances of that activity back to customers of particular ISPs.

Photo courtesy of Gates Foundation

10 The report considers the issues as follows:

- The technical background: what IP addresses are for and why they can be used to reliably identify the other end of many Internet connections.
- The basic process by which we achieve 'traceability': the determination of 'who did that?' on the Internet.
- The architectural design of file sharing protocols.
- How file sharing activity can be monitored in order to record the IP addresses of the Internet connections being used.
- How monitoring systems should be configured so that they are capable of producing consistently reliable results and what records need to be kept in order to permit disputed results to be double-checked.
- The details that must be given to the ISP so that they can identify the subscriber who must be notified of a copyright infringement report.
- What problems the ISP may encounter in linking an IP address generated by a properly operated monitoring system with a subscriber account.
- Whether the subscriber will be in a position to know the identity of the individual whose file sharing activities were detected.
- How file sharing protocols could evolve to prevent third party monitoring systems from obtaining the IP address information needed to generate a copyright infringement report under the Digital Economy Act 2010.

Digital Economy Act 2010

IP addresses

11 The Internet is a packet switching network which operates using the Internet Protocol (IP). Data that is to be sent from one place to another is placed into packets and the packet is labelled with the destination address. The packet is then routed across the Internet by devices that check the address and send it onwards towards its final destination.

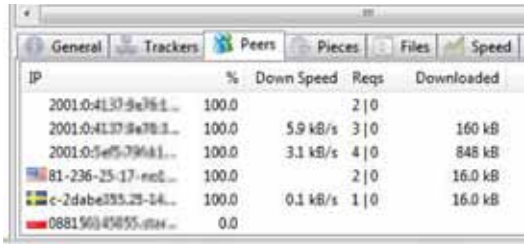
12 In the current 'version 4' of IP (IPv4), the addresses used are 32 bits long, and are conventionally written as four decimal numbers. For example, the IP address '01010001100000011111100000100000' is almost invariably written as 81.129.248.32, since 81 is 01010001 in binary, 129 is 10000001, 248 is 11111000 and 32 is 00100000.

13 From the point of view of the rest of the Internet, IP addresses are unique – for if not, then it would not be possible to know where to send a particular data packet.

14 An IP packet not only contains the destination IP address, as described above, but also a source IP address which indicates where it has come from. When a response is made to an incoming packet, the two addresses are swapped over so that the response will return to the packet's originator.

15 Websites and file sharing protocols generally use the Transmission Control Protocol (TCP) for communications. The TCP protocol is layered on top of the IP protocol and uses IP addresses for the endpoints.

16 When a TCP connection is first made, three special 'handshake' packets are exchanged between the two endpoints. These handshake packets contain two 'initial sequence numbers', one of which is randomly chosen by each end of the connection. These initial sequence numbers – provided that they are truly random, which they will be on modern systems – ensure that third parties, elsewhere on the Internet, are not able to 'spoof' a connection and fool a machine into believing that it is swapping traffic with a different IP address than is actually the case.



17 The consequence of the way in which TCP connections are made is that each end of such a connection is able to log the IP address of the other end of the connection, and these IP addresses can be relied upon as an accurate record. However, as will become apparent, this is just one aspect of traceability, since the other end of the connection may be an intermediate device, such as an asymmetric digital subscriber line (ADSL) router with network address translation (NAT) functionality.

18 In order to allow many simultaneous connections between a pair of endpoints, TCP adds a 16-bit 'port number' to the both the source and destination addresses. At the destination this port number will be used during the opening of the connection to select between different services (such as HTTP (a web server), POP3 (an email mailbox), or FTP (a file transfer system)), each of which conventionally listens for connections on a different port (80, 110, 21, etc.). At the source of the connection, the port numbers permit the same service to be used in parallel by several different programs or users without any confusion arising as to which of these parallel connections a particular data packet belongs to.

19 In IPv6, which is beginning to be deployed, addresses are 128 bits long (and written out for humans in a different style). There are differences between IPv4 and IPv6, but none that are relevant to the present discussion, so henceforward I will just write IP to mean both IPv4 and IPv6.

The basic method of tracing the use of an IP address

20 IP address space is allocated by Regional Internet Registries (RIRs) in a hierarchical manner to individual ISPs. Each ISP will be allocated one or more blocks of contiguous IP addresses for the use of their customers.

21 To determine the subscriber a particular IP address was allocated to at a particular time, it is first necessary to look up the IP address in the public databases maintained by the RIRs. This will determine which block the IP address is in, and which ISP has been allocated that block. Contact details for the ISP will be available from the RIR's public records.

22 The ISP is solely responsible for allocations of IP addresses within their block of address space. They almost never publish any data about which subscriber was allocated which address and over what time period. Hence, in almost every case, to establish which subscriber was using a particular IP address at a particular time it will be necessary to correspond with the ISP.

23 Most ISPs allocate IP addresses to subscribers in a dynamic manner, so that a particular IP address is used by a subscriber for a few hours, days, or possibly weeks, and it is then freed (so that no-one is using it) and it can be allocated to another subscriber thereafter. If the subscriber disconnects from the Internet then they may well get a different IP address when they reconnect.

24 The alternative to dynamic allocation is static allocation, in which a customer always has the same IP address throughout the time that they are buying service from the ISP. This arrangement is unusual in the UK for consumer connections to the Internet, but common for business services.

25 The ISP's records of past IP address allocations will generally be preserved for a few weeks or months, but will then be discarded since this data has no long term business significance to an ISP. Since these records will be personal data, as defined by the Data Protection Directive, the ISP is obliged to delete the records when they no longer serve a business purpose.

26 However, to assist law enforcement, the UK implemented the Data Retention Directive in 2009 to create a framework under which ISPs can be required to retain these records for 12 months. This 'data retention' obligation overrides the provisions of the Data Protection Directive, but it only applies to ISPs which have been served with a relevant statutory notice.[3]

27 The Home Office does not disclose which ISPs have been served with statutory notices to compel the retention of data. Although the legislation requires the serving of statutory notices on all ISPs, it is not currently Home Office policy to do this. It is generally understood that the Home Office have been dealing only with the very largest ISPs.

28 It should be carefully noted that once the ISP has discarded their records of IP address allocation, either at the end of 12 month period required by a data retention notice, or when they have no further business need for the information, then the ISP will be unable to ascertain which subscriber was using an IP address at a particular time.

[3] See SI 2009:859 The Data Retention (EC Directive) Regulations 2009, section 10(1)

Complications caused by NAT

29 Network Address Translation (NAT) systems are extremely common. Almost all consumers now connect to the Internet through a NAT system because this allows them to run two or more computers over a single Internet connection with a single publicly facing IP address. They are also used by many businesses to either run more machines than they have been allocated IP addresses, or merely for the 'firewall' security properties that a NAT system can provide.

30 A NAT translates between internal (the business or household) and external (the global Internet) IP addresses, changing the source address of outgoing packets (to the Internet) to be the Internet facing IP address and then when packets return from the Internet the destination IP address is set to be internal IP address of the relevant computer on the internal network.

31 In order to prevent clashes when two internal machines access the same external machine – and to simplify its book-keeping – the NAT device will generally also rewrite the source port number of outgoing packets and correspondingly fix up the value for incoming packets.

32 When a NAT is in use, logging at remote sites will record the relevant Internet facing (public) IP address, but the internal (private) IP address will not be disclosed to the remote site.

33 If the NAT kept logs of the connections made through it then the NAT owner could identify the internal machine which made a particular connection. In practice, very few business NAT systems and practically no consumer NAT systems will keep any records at all once the connection is closed. This means that to all intents and purposes, traceability will cease at the NAT.

34 There is a further related difficulty when considering file sharing activity by mobile phones, or the so-called 'dongles' that use mobile phone technology to allow standard computers to access the Internet. These types of Internet access have become popular at a time when we are running out of IPv4 addresses, so it has not been possible to give a unique IPv4 address to every individual user.

35 The solution being employed by the mobile phone ISPs is to share addresses using so-called “Carrier-grade NAT” (CGN). These systems can be thought of as NAT systems that are operated by the ISP. On the Internet facing side they have a pool of a few tens of thousands of IP addresses. On the customer facing side they provide service to many millions of customers. Hence any particular IP address is being shared by perhaps a thousand different customers at any particular time and, unlike the NAT systems described above (in #29 – #33), these customers are complete strangers to one another.

36 The CGN systems are capable of creating logs of connections, so that if you approach the ISP knowing not only the date/time and IP address but also the ‘source port number’ that was used to make the connection (a further TCP connection identifier, see #18 above) then the ISP can, in principle, identify the customer who made that particular connection.

37 I say “in principle” because the logs generated by CGNs are extremely voluminous and so they are only retained for short periods. They fall outside the current statutory data retention regime, and I understand that it would be unusual for these logs to be kept for more than a few days even at the largest ISPs. I have also been told that at times of high traffic these logs may not even be created in the first place to avoid overloading the system.

38 What this means is that when a connection is made through a CGN system it is only possible to trace the customer who made the connection if a report contains an IP address, a source port number and an accurate timestamp. Reports must be made within a few days if traceability is to succeed, and in many cases it must be accepted that tracing will not be practicable.

How file sharing systems work

39 There are a number of different peer-to-peer systems in popular use, but by far the most popular at the present time is Bit Torrent. Someone who wishes to share a file creates a 'torrent file'. The torrent file provides details of the 'shared-file' and specifies a 'tracker', a machine that will coordinate the file sharing activity.

40 Although the focus of this document is on the sharing of copyright protected files without appropriate permission, Bit Torrent and other peer-to-peer systems are regularly used to share files in an entirely lawful manner. The actual mechanics of the sharing are identical in each case.

41 The shared-file is treated as if it was split into a number of equal sized 'pieces' (the size of which will be chosen to be 32 kB, 64 kB, 128 kB etc up to 16 MB). A checksum is calculated for each piece, using a cryptographic hash function, and the checksum value is placed in the torrent file. The checksum ensures that errors in transmission can be detected and the use of the hash function makes it computationally impossible to substitute a fake replacement piece.

42 The torrent file can be freely distributed by any method, such as just emailing it to people, or placing it on a website. There are specialist search engines, sometimes called indexers, for locating torrent files.

43 Once a copy of the torrent file has been obtained, the file sharing software will use the details within it to make a connection to the tracker machine to indicate interest in downloading the shared-file.

44 The tracker will provide details of the IP addresses to which connections should be made to request copies of some or all of the shared-file content. The file sharing software will make these connections, often in parallel, and will ask for particular pieces of the shared-file to be sent back.

45 In the jargon, the shared-file is 'downloaded' from 'peers' elsewhere on the Internet who are making the content available. Since these peers are sending the pieces of the shared-file 'to the Internet' they are called 'uploaders'.

46 The shared-file will be downloaded piece by piece from one or more of the uploaders. The downloaded data can be checksummed and the result compared with the value recorded in the torrent file to ensure that no problems have occurred. Once all the pieces have been successfully obtained the original shared-file is reassembled.

47 The file sharing software will inform the tracker as each piece is successfully obtained and the tracker will then tell others who want the shared-file which pieces are now available from the downloader and will provide them with the relevant IP address so that they can make contact. So although a downloader sets off merely to fetch a copy for their own use, their system will also be automatically prepared to upload pieces of the shared-file to others.

48 File sharing software contains various configuration parameters that can limit the amount of uploading, but the system as a whole relies on people uploading as well as downloading, so there is strong encouragement (through better performance for example) to permit higher rates of uploading, and the software will not generally prevent uploading altogether.

49 When the shared-file is complete it is considered good manners to keep it available for access by others for a while. To encourage people to display these good manners, the tracker may keep statistics on the relative amounts of downloading and uploading, and will discriminate against peers in future if they have downloaded far more data than they have allowed to be uploaded.

50 Different peer-to-peer systems work in different ways. Systems such as Ares use distributed systems to hold indexes (rather than using a specific tracker machine). However, once the search phase is complete and a list of peers has been compiled to indicate where pieces of the shared-file are located, then downloading (and apprising the system of which pieces of the shared-file can be uploaded to others) is very much the same.

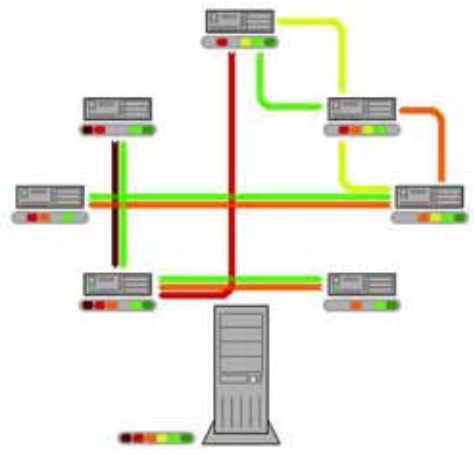


Image courtesy of Wikiadd, Wikimedia Commons

How to approach the traceability of file sharing activity

51 All currently operating mainstream peer-to-peer file sharing systems pass pieces of the shared-files directly between participants using the TCP protocol. This means that a monitoring system that joins in the file sharing activity can in principle make a reliable record of the IP addresses at the other end of the connections used for upload or download.

52 If the monitoring system provides pieces of a shared-file for others to download then it can make a record of downloaders. If the monitoring system downloads pieces of a shared-file from others it can make a record of uploaders.

53 If downloaders are to be monitored then the monitoring system will need to make one or more pieces of the shared-file available – it must be an uploader. This means that the monitoring system will need to actively participate in the peer-to-peer system by making a tracker (in Bit Torrent) or a distributed database (other systems) aware that it has some or all of the pieces of the shared-file. It will then, upon request, need to transfer pieces of the shared-file to others.

54 When a download occurs, then the monitoring system can record the details of the TCP connection. I am not qualified to assess whether copyright infringement occurs as soon as just one byte is transferred over that connection, or whether it is necessary for the entire piece to be transferred, or indeed whether the remote machine must succeed in fetching all of the pieces of the shared-file.

55 If it is important, from a legal perspective, to be entirely sure that a particular piece of the shared-file has been transferred successfully then this might be deduced by checking whether the downloader has started to advertise the availability of the particular piece that it was sent.

56 If Ofcom believes, from legal advice that it obtains, that it is essential to check that transfers have been successful, then I would expect it to find considerable difficulties in setting out the detail of a sound procedure to do this for all commonly used peer-to-peer systems. Hence, if it came to the conclusion that only uploading was to be monitored then, in my view, that would be a sensible line to take.

57 Monitoring of uploaders is entirely straightforward. In this case the monitoring system will be playing the part of a downloader. It will need to contact the tracker (etc.) to obtain the list of the IP addresses of available uploaders for each of the individual pieces of the shared-file. Direct contact can then be made with these computers and a request made for pieces to be transferred.

58 Whilst the transfer (the download) of a piece is occurring the monitoring system can record the details of the TCP connection. The monitoring system will be able directly confirm that the entire piece has been transferred, and will be able to compare it the data received with the original copyrighted material to show that it matches.

59 As noted above in the discussions of NATs and dynamic IP address allocation, when recording the details of a TCP connection, it is not sufficient to merely record the IP address of the other end of the file transfer, but it will also be essential to make a note of the source port number and to provide an accurate timestamp.

60 Finally in this section, it might be wondered why the description of the monitoring of uploaders has described a process which involves the transfer of a piece of the shared-file. The reason is that this is a robust mechanism that establishes without doubt that the uploader can actually supply valid pieces of the copyright protected shared-file.

61 In 2008 three researchers from the University of Washington, Michael Piatek, Tadayoshi Kohno and Arvind Krishnamurthy published an academic paper which showed that some monitoring companies were incorrectly identifying uploaders in Bit Torrent peer-to-peer file sharing.[4]

62 What the University of Washington researchers established was that some of the monitoring systems were making only perfunctory efforts to contact the peer-to-peer nodes. Instead, the tracking systems were relying solely on the metadata maintained within the system that documented the IP addresses of peers that (apparently) had pieces of the shared-file available for download. This metadata could be spoofed and the researchers amused themselves by arranging for one of their laser printers to be accused of uploading the film 'Iron Man'.

[4] Michael Piatek, Tadayoshi Kohon & Arvind Krishnamurthy, Challenges and Directions for Monitoring P2P File Sharing Networks –or– Why My printer Received a DMCA Takedown Notice, (**http://dmca.cs.washington.edu/dmca_hotsec08.pdf**) Department of Computer Science & Engineering University of Washington, Washington, University of Washington, 2008

Detailed evidentiary requirements for recording file sharing activity

63 The previous section sets out the general mechanism by which file sharing activity can be monitored. This section sets out specific things that need to be done in order to ensure that the monitoring is as accurate and error free as possible.

64 I will describe only the monitoring of uploaders (i.e. where the monitoring system is downloading a piece of a shared-file). Extending the description to cover the collection of information about multiple pieces is straightforward and so it needs no further detailed discussion by me. If downloaders are to be monitored then much of the detail will be very similar, but considerable complications arise if it is necessary to check that transfers are successful (#55) and I have not attempted to provide any detail about how that should be done.

65 To start with the basics. The monitoring machine needs to be secured against unauthorised use. Its software should be configured to prevent unauthorised access to the machine and it should be kept up to date with all security-relevant patches. System logs should be regularly inspected to ensure that there is no evidence of intrusion. If an intrusion is discovered then all of the results gathered by the machine can no longer be trusted.

66 In my opinion, it would be entirely inappropriate for the monitoring machine to be used for any other purpose than monitoring. If it was also being used as a web server, email server, or worse as a general purpose workstation, then this markedly increases the 'attack surface' and makes it considerably more difficult to ensure that it remains secure.

67 The monitoring machine should be running a NTP daemon synchronised to reputable time sources so that it is capable of providing timestamps of events which are accurate to one second or less. Using inaccurate timestamps could mean that the ISP identifies the wrong subscriber as having been allocated the IP address.

68 The monitoring machine should be configured in such a way that all the timestamps it provides are always in Coordinated Universal Time (UTC), ie +0000, no matter what season of year it may be. Dates should be specified in ISO format, ie 2012-12-25, to avoid any confusion between, say, 2nd March and 3rd February.

69 NTP daemons can fail, or the machine time can drift too far or too fast to be corrected. Therefore, before and after each monitoring run the operator should ask the machine what it thinks the time is – and compare that with the speaking clock, or some other independent and reliable source of time information. This will confirm that the NTP daemon is operating correctly. If monitoring is to run continuously, then the time should be checked on a daily basis. A specific contemporaneous note should be made that these checks have been made.

70 The monitoring software that is to be used should be developed according to current best practice. The source code should be held in a software revision control system. A contemporaneous note should be made whenever monitoring occurs as to precisely which version of the software is in use. Should flaws in the software later come to light it will then be possible to establish which previous results are now suspect.

71 A testing suite should be developed for the monitoring software that demonstrates that it is functioning correctly. This suite should be rerun whenever changes are made to the software (or the software is run on a new machine for the first time) to ensure that no inadvertent flaws have been introduced. The test suite output should be preserved in the revision control system. In accordance with best practice, when flaws are detected the test suite should be enhanced so as to check that the flaw has been fixed and is not reintroduced at any later stage.

72 In order to monitor the uploaders of a specific piece of copyrighted material, the first step will be for the monitoring system to obtain a torrent file (or equivalent in other peer-to-peer systems), contact the tracker and then proceed to download a complete copy of the shared-file.

73 This initial download of the entire shared-file serves two purposes. The first is to be able to establish that the shared-file is indeed the copyrighted work that it claims to be and is not some other material that has been mislabelled – as is often the case. The second purpose is to be able to calculate cryptographic hashes of the various pieces of the shared-file, because these can then be used to quickly determine that future downloads contain matching data.

74 A specific contemporaneous note should be made of the identity of the shared-file that is to be monitored, how it was established that it was a copyrighted work, and a record made of the cryptographic hashes. The copyrighted material, and all the other information should be preserved within a revision control system to allow them to be inspected if a dispute arises.

75 A specific contemporaneous note should be made of the details (so far as they can be ascertained) of the peer-to-peer system that is being monitored. This would include protocol version numbers, a copy of the tracker file (if Bit Torrent is involved) along with a record of where it was fetched from. Some of this information will be required for the copyright infringement reports that are to be sent to ISPs, but it should all be recorded to assist in resolving disputes and to address technical complications that might arise should peer-to-peer systems change in the future without appropriate changes having been made to the monitoring software.

76 When the monitoring system fetches a piece of the copyright protected shared-file from an uploader, a contemporaneous record should be made of the time of the start and end of the TCP connection that was used. The remote IP address and remote TCP port number must also be recorded.

77 Once each piece of the shared-file has been downloaded, a cryptographic hash value should be calculated and – if it is correct – should be recorded along with the other details.

78 Quite clearly, if the tracker machine indicates that a piece of the copyright protected shared-file is available from a particular peer, but the monitoring system cannot make contact, or the data transferred does not exactly match the copyright protected shared-file, then no record should be made.

79 The initial set of records will come from the fetching of the complete copy of the copyright protected shared file (see #72). Thereafter the monitoring system will be fetching pieces of the shared-file from uploaders as it learns of their existence from the tracker machine.

80 All of the monitoring data, along with any event messages from the monitoring software that indicate its operation, should be written to a logging file. This logging file should be preserved for the foreseeable future within a revision control system so as to assist with dispute resolution.

81 If subscribers are taken to court for copyright infringement there are likely to be applications to the courts to have independent experts review the monitoring systems to assess whether their design, implementation and operation can be relied upon to produce valid results.

82 Therefore, monitoring systems should be designed in such a way as to clearly distinguish between:

- information that needs to be kept entirely secret – such as the IP addresses from which the monitoring is done, which if ever disclosed would render the monitoring ineffective;
- information that is merely proprietary – such as the system source code, whose disclosure could assist unscrupulous competitors, but that court appointed experts might reasonably be permitted to inspect;
- and, information which provides part of the trail of evidence that demonstrates that monitoring has been correctly performed. There should be no objection to providing this information to subscribers who are notified that their Internet connection is believed to have been used for copyright infringement through peer-to-peer file sharing.

83 Appropriate design choices in creating the monitoring systems will simplify independent review and will also allow the creation of detailed descriptions of their operation – essential for public confidence – without compromising their effectiveness.

84 In my view, Ofcom should specify that detailed descriptions of monitoring systems – at the same sort of general level as this document – should published before they start to be deployed, and that 'secret' designs should not be considered capable of creating reliable results.

85 It goes almost without saying that Ofcom should have the right to audit the logging information from the monitoring systems in order to ensure that any standards that they set out in the Initial Obligations Code are being adhered to.

86 Ofcom should make a point of exercising its right to perform an audit when incorrect identifications are reported to it as I discuss in #105 below.

87 I would also recommend that a new monitoring system design is audited before it starts being used to collect the data that will form the basis for copyright infringement reports. This audit could be done by an independent third party, but to prevent 'shopping around' for more tolerant auditors, it would be more appropriate for Ofcom to be responsible for the audit.

88 This initial auditing requirement should not be over-burdensome because copyright holders are likely to purchase monitoring services from a small number of specialist companies – so only a few audits will be necessary.

89 The publishing of detailed descriptions of monitoring systems along with the audit results will considerably improve public understanding and confidence in the whole concept of monitoring, which may have an impact on the number of appeals that are made.

Failures in subscriber account identification by ISPs

90 Having set out the issues that arise with monitoring systems both in broad terms and then in detail, I now turn to the situation at an ISP that receives a copyright infringement report and which must notify the appropriate subscriber. Considerable care needs to be taken by the ISP with the technical details in order to avoid errors occurring during its part of the process.

91 I have explained how the monitoring system is to be operated so that the timestamps it provides are accurate. The ISP also needs to have accurate timestamps in its logging data as well. If an inaccurate timestamp is used, even one that is just a few seconds out, then an erroneous identification may be made of the previous or next customer to be dynamically allocated the particular IP address.

92 Although automated clock synchronisation is 'best practice' and machines will initially be set up correctly, in my experience synchronisation mechanisms can quietly fail causing ISP machine clocks to regularly drift away from 'wall-clock time'. Because accurate timestamps are seldom relevant to the ISP's day to day operations, it may be many months before this is noticed.

93 The next problem is that logging of connection data is inherently unreliable. Records are sent from the machines that handle the connections to the logging machines using the User Datagram Protocol (UDP). Any error on the network will cause the UDP packets to be irretrievably lost, and this loss will not be detected. These losses are always assumed to be rare, but once again the ISPs have no pressing business need to monitor and in practice will only investigate long after the start of the problem when someone notices inconsistent results.

94 When a subscriber connects to the Internet, perhaps by switching on their ADSL router after it has been turned off overnight, the router will use the account and password credentials it has stored in order to authenticate itself to the ISP. This generates a 'START' record in the ISP logging systems. When the user disconnects (and the ISP systems notice), a 'STOP' record is created. These START and STOP records delimit the time during which an IP address is in use by a particular subscriber.

95 Internet connections may only last a few minutes, until the user switches off the ADSL router or until poor line quality causes the link to drop and be re-established. However, in some cases connections will last for weeks or even sometimes months.

96 When the ISP tries to determine which subscriber was allocated an IP address at a particular time it is best practice to locate both the START and STOP records and check for consistency. If either record is missing, then misidentification can occur, and failure to apply best practice (e.g. by only checking START records) can easily cause the wrong subscriber to be identified.

97 Large ISPs generally use several machines to record the START/STOP records. If one of these machines fails, then until the system is reconfigured, a proportion (a fifth or an eighth perhaps) of the records will be lost. Monitoring ought to detect such outages, but in my experience, subtle failure (that is, failures that affect the logging but leave other functions of the machine operating normally) can be overlooked for several days at a time.

98 Since I have made it clear that reports of file transfer events should include both start and stop times, the ISP should be checking that the IP address was allocated to the same subscriber at both of these times. If not, or if there is no record of the IP address being in use at the relevant time then it is clear that an error has occurred either in the monitoring system that has generated the copyright infringement report or within the ISP's logging systems. If this occurs then I discuss what should then happen in #104 below.

99 A different type of error can arise when ISPs use manual processes to consult the logs to establish which subscriber was using an IP address at a particular time. If the ISP's data lookup operations are not completely automated then there is considerable risk of human error in doing the lookups.

100 Of particular relevance here is a 2009 UK murder case, tried at the Old Bailey. I was employed by the Crown Prosecution Service after flaws in the Internet traceability evidence came to light during cross-examination of a witness in a retrial. The ISP was asked to identify eight IP addresses. Initially they were only able to identify four of these eight – in some cases because they incorrectly translated times in Pacific Daylight Saving Time (PDT) into Greenwich Mean Time (GMT). When a more senior employee was asked to re-examine their records he avoided these errors and identified seven of the eight events.

101 The last event was reported by the police to have occurred at 18:50 GMT whereas it actually occurred an hour earlier at 18:50 British Summer Time (BST). This final error was entirely the police's – they guessed at a time-zone for a timestamp provided to them by one of the High Street banks. Because, by chance, the senior employee had kept a record of all of the START/STOP records for the relevant IP address, it was possible, over a year later, to identify the issue and provide a corrected result.

102 The police applied a 'Doctrine of Perfection' to the forensic data. When the results were initially inconsistent they asked the ISP to check their working and thereafter all but one of the identifications was correct. They erroneously assumed the eighth had differed because an open wireless connection had been used, but in the retrial this explanation was shown to be false. I then assisted the ISP and police in determining what the final error was – so that perfect results were finally established.

103 It is my view that the types of errors I have discussed in the preceding paragraphs are widespread. Although they are sometimes detected because of the meticulous attention to detail that a murder trial entails, there are few other 'feedback mechanisms' that ensure that standards are maintained and systemic errors avoided. In my view these types of errors will explain a fair proportion of the misidentifications that are regularly reported in the press in file sharing cases. Automation will reduce some types of errors, but automation does not of itself eliminate the unreliability of many of the system components.

104 It is my very strong recommendation to Ofcom that they should bear the possibility of error in mind as they specify protocols and mechanisms for identifying file sharing activity. In particular, if the ISP receives a batch of data and detects just a single error in it, then (unless the cause can be promptly identified and corrected) the whole batch should be discarded. The reasoning for this is that the error is very likely to be systemic. Perhaps one event has been mapped to an unallocated IP address and hence detected, whereas all of the others may also have been mapped to the wrong customers, but the ISP will not be in a position to determine that these attributions are incorrect.

105 The discarding of the batch by the ISP should also be reported to Ofcom, since it should be investigating whether the underlying problem is on the monitoring side (in which case all of its data, no matter which ISP it was sent to) must be considered unreliable; or it will be at the ISP, in which case its other identifications will be unreliable.

106 When Ofcom has identified where the error was made, it should inform any other ISPs who received copyright infringement reports that were based on erroneous IP address information and make the details of the error public. Knowledge of the details of the error will clearly have an impact on any appeals that are being made by individuals claiming they have been falsely accused.

107 It will not be appropriate to apply this Doctrine of Perfection (one failure spoils the whole batch) in the case where an ISP does not hold any of the required records to identify a customer. In particular, if the ISP has discarded the relevant records from a CGN system the allocation of the IP address/port may no longer be known. This would not hint at any systemic problems (such as clock drift) and so the rest of the batch can be processed as usual.

Identification of file sharing individuals by a subscriber

108 Notwithstanding all the discussion about possible errors, provided that all of the systems are operating correctly, file sharing activity can be traced from a monitoring system, through the ISP's records of IP address allocation to the subscriber, the person who has contracted with the ISP to purchase Internet access. However, the ISP is inherently unable to identify the individual has been infringing copyright, whether it is the subscriber or not.

109 The subscriber may not be the individual who has infringed copyright, and until they are notified of the copyright infringement report they may not have even been previously aware that others were using the connection they have paid for to engage in file sharing. Depending on the circumstances, the subscriber may find it entirely impossible to determine who the copyright infringer was.

110 There are four main reasons why the subscriber may not be the person who has infringed copyright – and may be unable to determine who was:

a there may be many people within a household;

b there may intentionally be many strangers using an Internet connection;

c there may unintentionally be many strangers using a connection;

d and, finally, the account credentials may be used, probably fraudulently, somewhere else altogether.

111 The most common (indeed very likely) situation is (a) above: that a household is comprised of several people, one of whom formally pays the bills for Internet access (and is known to the ISP as the subscriber). The other people share the connection, and one of them is file sharing and infringes copyright. The simplest model would be a family, where it might be relatively easy to establish which individual had been infringing.

112 However, the same doubt as to responsibility arises when adults 'flat-share' or lodgers rent rooms, or if it is agreed that close neighbours may piggy-back on the connection, or indeed when casual visitors use the Internet whilst they are on the premises. It is far less clear that the subscriber will ever be aware of what Internet activity is occurring; or that they can reasonably prevent file sharing of copyrighted material over the connection; or that there is any practical way, after the fact, of identifying which individual was responsible for a particular file sharing incident.

113 When the police encounter this type of problem, perhaps when tracking down someone who has been handling child sexual abuse images, they will seize all the computers at the property and employ highly skilled experts to perform forensic examinations to determine which computers show evidence of illegal acts. The police also use their powers to interview all the individuals concerned, ask about visitors, and combine all this information with the forensic evidence to determine who should be prosecuted.

114 The tracking of copyright infringement back to an IP address does not generally give any clues as to which computer was involved, and the type of NAT system that is typically used by households will not provide any logging that would be of assistance (see #29 to #33).

115 Even if there is only a single computer within the household, the copyright infringement report will not provide any clues as to who was at the keyboard.

116 Subscribers will seldom, if ever, be in a position to take the police approach that I outlined in #113. In practice, this means that when a subscriber is told by their ISP that a copyright infringement report has been received (and they were not culpable) then that subscriber will never, unless a confession is forthcoming, be in a position to be in the least bit sure as to whose activities were detected.

117 Case (b) extends (a) by considering the intentional use of an Internet connection by strangers. This arises, for example, when a coffee shop, a public library, or a municipality provides 'free Internet access'. This type of Internet access invariably employs NAT devices, and once again it would be unusual for logs of connections to be created.

118 Even if logs were kept, they will only identify connecting devices (laptops, smartphones, etc.) by their Media Access Control (MAC) address. Since no record will be kept of the MAC addresses of the devices used by passing strangers, the log will be of very limited value. Furthermore, specialist software can be used to allow laptops and smartphones to spoof any MAC address that they wish.

119 Thus, in case (b) as well, if copyright infringement does take place, then the subscriber will almost inevitably find it completely impossible to determine the computer that was used and identify the individual at the keyboard.

120 It has sometimes been suggested that restrictions should be placed on this type of public access, perhaps by limiting the speed or the volume of data transferred; but these will only slow down file sharing not prevent it, and could significantly impact other forms of Internet usage. Since all file sharing would be slowed, including lawful sharing of files where no copyright infringement occurs, the public will not be well served by such an approach.

121 Even if an attempt is made to block file sharing, perhaps by restricting connections to just web traffic, it may not succeed if the file sharing software does a good enough impersonation of a standard web connection to be allowed through. Should infringement occur, not only may it be hard to identify the infringer, but it may also be very difficult to determine how the restrictions were circumvented.

122 Case (c) is where strangers are using the connection without the knowledge of the subscriber and this almost always arises in the context of wireless access to the Internet. Wireless access points are widely installed by both businesses and consumers,[5] and a great many of the devices provide 'open' access. This openness may arise because subscribers are unaware that wireless access points can be secured or are unable to make the configuration changes to make them secure. However, many businesses and consumers choose to leave their wireless access points 'open' because the use of a security system would cause too much inconvenience or prevent them being used as the subscriber intends.

123 Wireless access points are often secured by the use of encryption schemes that require knowledge of a passphrase. These may still not be secure in that the devices may be using Wired Equivalent Privacy (WEP), a deprecated scheme that can be cracked by an eavesdropper in minutes, or a weak passphrase may have been chosen for a Wi-Fi Protected Access (WPA) or Wi-Fi Protected Access II (WPA2) scheme.

[5] According to the Ofcom 2011 Communications Market Report (**http://stakeholders.ofcom.org.uk/binaries/research/cmr/cmr11/UK_CMR_2011_FINAL.pdf**) Wi-Fi routers were used by 75 per cent of broadband households in Q1 2011, up from 66 per cent in Q1 2010

124 In a case from 2004, which I discuss in my PhD thesis, UK police raided a flat looking for the sender of 'phishing' emails that were being used as part of a fraudulent scheme to steal money from bank accounts. The flat's occupant was a bank employee, which had persuaded the police that they had identified the right man, although he seemed to be using a different name for the purposes of the scam. As it turned out, their suspect had a wireless access point with no security enabled, and it rapidly became apparent that this was likely to be the source of the emails and the bank employee was innocent. Officers remained on the scene, waiting for carpenters to arrive to fix the smashed-down front door. Whilst they were still there, by a lucky chance, someone walked down the street and – to avoid walking upstairs – called up to attract the attention of his mate, who lived in another flat opposite… since the shouted name matched the one the police were seeking, they were able to arrest the right man after all!

125 It would also be possible for an attacker to place malware (malicious software) onto a victim's computer and this program would allow other people to use their machine without permission. These programs, are sometimes called 'trojans' (after the Greek's Trojan horse), or 'viruses' (by analogy with biological viruses).

126 Malware is extremely commonly found on end-user machines; the number of affected machines world-wide is generally estimated to be between 2 per cent and 5 per cent. At any given time in the UK there are tens of thousands of machines running malware of some kind.

127 Malware which relays web traffic to obscure the location of criminal websites is regularly encountered and between 2007 and 2009 around two-thirds of all 'phishing' involved fake bank websites whose real locations were hidden by the use of relays via malware on compromised end-user machines. That said, I am not at present aware of any widely distributed malware whose purpose is to facilitate untraceable file sharing.

128 It is not generally possible to determine whether malware is implicated in a particular set of events without a forensic examination of the relevant computer, which will either find the malware itself, or will throw up secondary evidence that shows it was present in the past. This type of investigation is well beyond the capabilities of most subscribers, and employing an expert to do this type of investigation would not be cheap.

129 The final case (d) involves the theft of credentials. It is obvious how a dialup connection (using a fixed line or a mobile) can be used with a stolen login name and password. It is also possible to do the same with some broadband connections. The ISPs (and in particular the cable companies) disallow connections over phone lines where no-one has purchased broadband service – but they have little incentive to check whether an incorrect set of credentials have been used on a broadband-enabled connection.

130 In another case discussed in my PhD thesis, the police tracked down a broadband user and eliminated them from their inquiry because they had a convincing alibi. This led the police to deduce that the Internet access credentials were being used by someone else. However the BT machine that handled the credentials (and hence ought to have known which phone line they were used from) was not, at that time in 2003, keeping any logs and the tracing was unsuccessful.

131 On cable networks, “cloned modems” can permit the theft of service by criminals. The cable company delivers a service to two different places, where the connection devices authenticate with the same credentials. Since the cable company is losing revenue it has a significant incentive to detect the fraud, but in the meantime its records will assign two lots of activity to a single account.

132 In a case where credentials have been misappropriated, if the co-operation of the ISP is not forthcoming, and possibly even if it is, will there any likelihood that a subscriber who has been notified of copyright infringing file sharing will be able to show that their credentials were in use by someone else.

133 In his judgment on the Norwich Pharmacal Order application by Golden Eye International,[6] Mr Justice Arnold set out five possible reasons that the subscriber would not be the person who had been doing the file sharing. These are related to my categories as follows:

- The IP address identifies a computer and someone else in the same household (whether a resident or visitor) was using the computer at the relevant time (which might be with or without the knowledge of the subscriber).

  This in a special case of the (a) category, discussed in #111 to #114 above, where there is only a single computer in the household, or no use of NAT.

- The IP address identifies a router and someone else in the same household (whether a resident or visitor) was using a computer communicating via the same router (which might be with or without the knowledge of the subscriber).

  This is the general version of my (a) category, discussed in #111 to #114 above.

- The IP address identifies a wireless router with an insecure (either open or weakly encrypted) connection and someone outside the household was accessing the internet via that router (in all probability, without the knowledge of the subscriber).

  This is a specific instance of my (c) category, discussed in #122 to #124 above.

- The IP address identifies a computer or router, the computer or a computer connected to the router has been infected by a trojan and someone outside the household was using the computer to access the internet (almost certainly, without the knowledge of the subscriber).

  This is another version of my (c) category; see #125 to #128.

- The IP address identifies a computer which is open to public use, for example in an internet café or library.

  This is my category (b); see #117 to #120 above.

134 My category (d) was not considered in the Golden Eye judgment.

[6] Golden Eye (International) Ltd & Anor v Telefonica UK Ltd [2012] EWHC 723 (Ch) 26 March 2012 (**http://www.bailii.org/ew/cases/EWHC/Ch/2012/723.html**)

Peer-to-peer developments

135 Peer-to-peer systems all face much the same design challenges – scaling to handle millions of users; dealing with users that wish to 'free ride' by downloading and not uploading; and in avoiding the unwanted (by them) attention of ISPs and copyright owners.

136 When one system adds an innovation that is effective in addressing any of these issues, then other systems have the opportunity, and the motivation, to adopt it in short order. For example 'swarming' – the ability to fetch different parts of a shared file from different peers – was introduced by Bit Torrent, but quickly adopted by other peer-to-peer designs.

137 Nowadays, file sharing connections are often encrypted. This has been widely implemented in response to attempts by ISP to discourage the use of peer-to-peer protocols by 'traffic shaping'. This is a euphemism for the blocking or slowing of file sharing traffic (of both lawful and unlawful content) that the ISP is able to detect. The encrypted flows are more difficult for the equipment used by the ISP to identify, although there is something of an arms race occurring.

138 Because the files are decrypted at the end points, encryption makes no difference to the approach that I have outlined above for the monitoring of copyright infringement occurring on peer-to-peer networks.

139 What would make a difference, would be for file sharing systems to cease to have direct connections between nodes, or to cease to use TCP. If this was to happen then Ofcom would need to revisit the whole topic of monitoring because the approach I have outlined could no longer be considered to be reliable.

140 Direct connections are an efficient way of moving data around, but are by no means compulsory. Andrei Serjantov[7] describes a system where all connections are proxied and stored files encrypted. When his system is in use, it is not possible to learn where files are fetched from, and the locations that store the files have no idea what content they are storing and they will be entirely unable to determine whether it infringes copyright or not.

[7] Andrei Serjantov, Anonymizing Censorship Resistant Systems, (**http://www.iptps.org/papers-2002/120.pdf**) University of Cambridge Computer Laboratory, Cambridge, 1 March 2002

141 More mundanely than Serjantov's design, file sharing connections can be made over general purpose anonymising systems such as Tor (although the Tor developers discourage use of their system for this purpose, and naïve attempts to use Tor may be unsuccessful). The other end of the TCP connection will be a 'Tor exit node' – and many people operate these to help people living under repressive regimes to communicate. When Tor is in use, activity cannot be traced back beyond the exit node, even when well-resourced entities such as national governments attempt to do so.

142 TCP is not the only protocol that could be used. Bit Torrent has used UDP for a 'trackerless' variant[8] and adaptations of Safeweb's 'TriangleBoy'[9] could be used to obscure the IP addresses of the other end of peer-to-peer connections. If peer-to-peer systems do not use TCP then the theoretical underpinning of the accuracy of monitoring system results described in this report is inapplicable; and it is most likely that redesign would be needed.

Photo courtesy of philipcampbell's photostream

[8] Andrew Loewenstern, DHT protocol, (**http://www.bittorrent.org/beps/bep_0005.html**)BitTorrent.org, last modified 2008-02-08
[9] TriangleBoy Whitepaper, (**http://www.webrant.com/safeweb_site/html/www/tboy_whitepaper.html**) SafeWeb

Summary

143 In #11 to #62 I have set out the relevant background and have explained the theoretical basis for monitoring file sharing activity and in #63 to #89 I have provided detailed guidance as to how it should be done in practice so that valid results are obtained.

144 I have also explained, in #90 to #107, the problems that may occur at ISPs when they process the copyright infringement reports, and I have set out a Doctrine of Perfection that should be applied to reduce the risk of systemic failures causing widespread incorrect identification of customer accounts.

145 I have explained that although the largest ISPs are required to keep records of IP address allocation for a year, this statutory requirement does not in practice apply to smaller ISPs at the present time. Additionally, where the larger ISPs are using Carrier Grade NAT systems, they may not be able to identify customers unless requests are made very promptly, and possibly not even then.

146 I have explained, in #108 to #134, that for many different reasons a subscriber may be entirely unable to identify who has been using their Internet connection for file sharing.

147 In #135 to #142 I have considered various aspects of peer-to-peer design and development. I have drawn attention to the rapid adoption of encryption in the face of 'evolutionary pressure' from ISP traffic shaping. I predict that peer-to-peer systems will evolve to evade the type of monitoring system I have been considering in this document. Ofcom must keep this firmly in mind to ensure that it does not give the impression that any rules they set out for peer-to-peer monitoring are to be blindly applied when systems change and they become technically inappropriate.

148 Finally, in the last few paragraphs below (#149 to #154), I set out an 'executive summary' of the key 'take-aways'.

What overall view should be taken of the Digital Economy Act 2010 process ?

149 When peer-to-peer file sharing is used to infringe copyright, third parties can join in and determine the IP addresses of participants with whom they communicate. This monitoring has a sound theoretical basis, but it is necessary to get the practical details correct.

150 Good record keeping will ensure that the monitoring systems can be audited and if a systemic error should occur it will be straightforward to rescind previous mistaken reports.

151 ISPs will usually be able to identify the subscriber who was allocated an IP address at the relevant time. Errors can and do occur in this process, so a Doctrine of Perfection needs to be applied to provide the best possible chance of detecting systemic errors.

152 When the subscriber learns of a copyright infringement report the subscriber may be the infringer, or they may know who it was. However, in many cases the subscriber will not know who was at fault and they will have no practical way of determining whose file sharing activity has been detected.

153 When subscribers are notified, an outline of the working of the monitoring system should be provided to them. Provided that Ofcom insists upon appropriate standards and auditing shows that they have been adhered to, it would be appropriate for notifications to indicate that identification of subscribers was usually accurate. Nevertheless, it is important that the text should reflect the possibility of errors occurring during the process.

154 Subscribers should also be told of the full range of scenarios which can lead to file sharing occurring without their knowledge. This should reduce complaints about errors, whilst ensuring that a subscriber who can rule out alternative explanations will be well placed to draw attention to undetected errors in procedures or infrastructure at the ISP or within the monitoring system.

Published: July 2012

If you have any questions or would like further information about this report, please contact Saskia Walzel, by telephone on 020 7799 7977 or via email saskia.walzel@consumerfocus.org.uk

For regular updates from Consumer Focus, sign up to our monthly e-newsletter by emailing enews@consumerfocus.org.uk or follow us on Twitter **http://twitter.com/consumerfocus**

If you require this publication in Braille, large print or on audio CD please contact us.

Deaf, hard of hearing or speech-impaired consumers can contact Consumer Focus via Text Relay:

From a textphone, call 18001 020 7799 7900

From a telephone, call 18002 020 7799 7900