II: "Finding"
Traceability on the Internet

27th November 2002

Richard Clayton

These lecture notes were specially prepared for the Cambridge University Computer Science "Additional Topics" course, Michaelmas Term 2002.

richard.clayton@cl.cam.ac.uk

Outline

- TCP/IP refresher
- When IP addresses don't work
- When IP addresses do work
- Finding the source
- Dealing with dial-up
- Hiding on a LAN

The slides give the broad outline of the lectures and the notes ensure that the details are properly recorded, lest they be skipped over on the day. However, it is at least arguable that it will be far more interesting to take notice of what I say off-the-cuff rather than relying on this document as an accurate rendition of what the lecture was really about!

Further Reading

http://www.linx.net/noncore/bcp/traceability-bcp.html

written by UK ISP industry;
edited by Richard Clayton

http://www.cl.cam.ac.uk/~rnc1/The_Limits_of_Traceability.pdf

Richard Clayton

27th November 2002 Finding

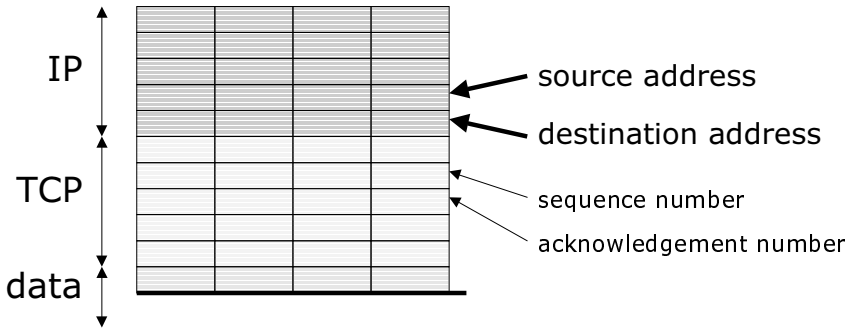
Besides the two documents quoted, there's not been a great deal of material collected together on traceability as a concept. However, these notes give some further references, as appropriate, on particular issues.

For a detailed set of PowerPoint slides on Denial of Service issues see:

"DDoS: Undeniably a global Internet problem looking for a solution", Yehuda Afek & Hank Nussbacher, RIPE-41 EOF Tutorial, 15/01/2002, Amsterdam.

http://www.ripe.net/ripe/meetings/archive/ripe-41/tutorials/eof-ddos.pdf

(Almost) All You Need to Know about TCP/IP



✶ TCP/IP is described in many textbooks. There are only a few important aspects of the protocol from the point of view of Traceability.

✶ The *destination IP address* says where the packet is to be sent. It is always, by definition, valid.

✶ The *source IP address* indicates where the packet came from. It can be forged (but may not then be allowed out of its originating network if the "firewalls" there (usually in fact just simple routers) are configured in accordance with RFC2267).

✶ When the packet reaches its destination, the source and destination addresses will be swapped over for the return journey.

✶ The *sequence number* indicates where the contents of the current packet fit in the notional buffer for the whole conversation. The *acknowledgement number* indicates how much of that buffer has been received so far. Both of these values start from a randomly chosen point in a 2^{32} byte buffer.

Are Addresses Valid ?

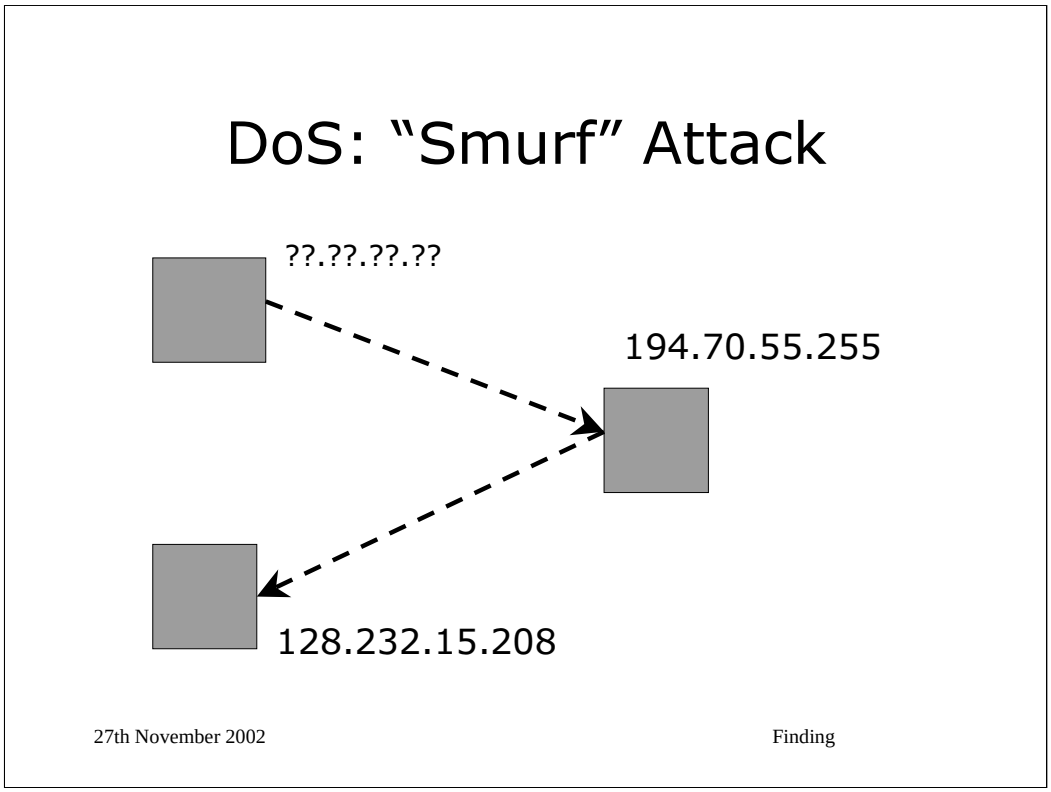
- Destination address is always valid
- Source address is valid for 2-way traffic
- Can send single bad packets with 1-way traffic
- Can do denial of service with 1-way traffic
- Filters can be useful in ensuring validity; but beware of source routing
- Also, can spoof addresses if the stack is poorly written and can predict responses (see later)

27th November 2002 Finding

✶ If you are not interested in getting packets back from a remote machine then the validity of the source address is irrelevant. If you wish to avoid being traced then you might set an invalid address. There are a number of attacks that are possible with 1-way traffic such as denial of service attacks and the sending of malformed packets that crash the remote system.

eg Teardrop (invalid fragments)
Land (connection to self)
"Ping of Death" (extra long packets)
WinNuke (buffer overflow on 139)
SYN (only one handshake packet, so consumes resources)
Reflector (repeated SYN-ACK responding to forged SYN)
Looping UDP (connects echo(7) to chargen(19))
etc etc

✶ It is possible to filter packets to ensure they are valid (spotting insider addresses coming from outside and vice versa). However, IP does have a concept of "source routing" which causes packets to go via particular intermediate addresses first. In practice, however, source routed packets may well get dropped because they're prima facie evidence of wickedness !



✶ The "smurf" attack was known for some time before it was integrated into a simple-to-use tool in the summer of 1997. Some claim that the attack was used the previous year to "persuade" AGIS to drop Cyber Promotions (Sanford Wallace's bulk email sending company – see Friday's lecture).

✶ The basic idea of a smurf is to send an ICMP echo request packet to the broadcast address of a network. All machines on that network will then respond with an ICMP echo reply. If you forge the source address then those echo replies will be sent to the machine you are attacking.

✶ At one time you could find addresses with amplification factors of 10,000 or more, however these days there are few networks with >10. Nevertheless, smurf is still a popular technique for the bad guys, often launched these days as part of a "distributed denial of service (DDoS)" attack.

✶ BTW: The usual reason for attacking a system is that it is an IRC server. If you can hose its connections then you get a "net split" and you can then become operator and "own" a channel (and kick off the original operator who kicked you off& etc etc) There are entire FAQs on this sort of thing ☹

Smurf Protection

- Ingress filtering (RFC2267)
- Change directed broadcast rules (RFC2644)
- "Name and shame" lists for amplifiers
  http://www.netscan.org
  http://www.powertech.no/smurf/
- Detection of flows on border routers or at exchanges (use interface #s or MACs)
- Low probability responses for tracking
- Traceback

27th November 2002 Finding

✶ If systems didn't let forged packets out onto the Net then all would be wonderful (see RFC2267). Unfortunately it is seldom simple for complex networks to install suitable filtering.

✶ If routers didn't respond to ICMP echo requests to the broadcast address then there would be no amplification (though you can still hide your tracks).

✶ Some people believe that if there is a public list of the "best" amplifiers then lazy "script kiddies" will use those and the owners of the amplifiers will have to fix their problem in order to get their bandwidth back. The appearance of the lists coincided with a reduction in amplifiers. Cause and effect is hard to confirm.

✶ If you can find where the flow is coming from on your network then you can use low level info (Ethernet MACs or router interface info) to trace it back to the relevant peer. They can then track back across their network and block the flow nearer to the source. This is what is done in practice today.

Tracing Schemes

- ICMP Traceback (Bellovin, Leech, Taylor) 2000
  - ICMP packet accompanies packets at p(1/20000)
  - gives MAC addresses or flow info
  - directed form improves chances for specific traffic
- Logging (Snoeren et al) 2001
  - routers record packets as they pass
  - ORs a single bit into several Bloom filters
  - regular dump of pattern onto disk for interrogation
- Auditing (eg: Wanwall, Mazu)
  - effectively an IDS looking for attack patterns
  - can be done out-of-band by sniffing the traffic

27th November 2002 Finding

✶ If routers sent valid info about their identity and the nature of the flow along with 1 in 20,000 of the packets they were handling then a machine suffering from a DoS would get information about where the packets were coming from. Unfortunately, this won't work well until more routers on the Internet are upgraded.

See: http://www.ietf.org/html.charters/itrace-charter.html

There are practical issues (traffic rises 0.1% or so), bad guys may forge information to mislead, and firewalls may drop ICMP traffic anyway!

✶ "Hash-Based IP Traceback", Snoeren, Partridge, Sanchez, Jones, Tchakountio, Kent, Strayer. SIGCOMM'01 San Diego California, 2001.
http://www.acm.org/sigs/sigcomm/sigcomm2001/p1.html

Process packets through a hash to produce a digest. Use that digest value to set a single bit in a "Bloom filter". With several independent hashes can get good differentiation between packets. Lots of detail is to do with expected variations in packet contents! Blocks of bits are regularly recorded and can then be interrogated out of band to determine if a given packet passed through the router. Powerful technique that can, in principle, spot single packets.

Spoofing

- 3-way handshake

-->	SYN	client offset
<--	SYN-ACK	server offset
-->	ACK	

- If offset (and other info) is predictable don't need to see the return traffic to have a successful conversation
- Described by Morris (85) and CERT (95)
- Fix by making sequence numbers random and perhaps by suitable packet filtering at borders

27th November 2002 Finding

✶ Spoofing connections was first described in:

"A Weakness in the 4.2BSD UNIX TCP/IP Software", Robert T. Morris, Computing Science Technical Report No. 117, AT&T Bell Laboratories, Murray Hill, New Jersey. 1985

online at: http://www.pdos.lcs.mit.edu/~rtm/papers/117.pdf

The paper is particularly concerned with systems that trust other local machines (through hosts.allow mechanisms permitting rlogin &c). If you can successfully pretend to be local then you will have unauthorised access.

✶ To run the attack successfully you have to predict the sequence numbers (either by knowing them a priori, or by knowing an offset from another (non-spoofed) connection made first). Since the spoofed host will notice the unexpected SYN-ACK traffic it may also be necessary to run a "denial of service" on it to keep it from issuing a RST for the "connection".

✶ Morris suggests filtering out packets coming in from the outside that have internal source addresses (this is related to the RFC2267 filtering) and also ensuring that the sequence numbers are truly random.

✶ In 1995 there were enough systems being compromised for CERT to issue an advisory (CA-1995-01), and as late as October 2000 FreeBSD was being fixed to use something better than a simple PRNG to create "random" sequence numbers!

Who "Owns" an Address ?

- Regional registries issue numbers
  ARIN, APNIC & RIPE
- ISPs reallocate within their blocks
- Hence "whois" will yield owner
- Reverse DNS should also yield name

  eg: for 100.101.102.103:
  103.102.101.100.in-addr.arpa

27th November 2002 Finding

✶ IANA "owns" the IP address space, but it is managed by three "regional" registries:

ARIN	North & South America, sub-Saharan Africa http://www.arin.net/
APNIC	Asia-Pacific (ie Far East & the Antipodes) http://www.apnic.net/
RIPE	Europe, Middle East, parts of Africa http://www.ripe.net/

✶ The registries provide IP address registration services. They maintain databases of IP address "ownership" and AS (Autonomous System) numbers (routable blocks of IP space).

✶ Other systems and registries provide the "forward" mapping from domain names to IP addresses, but the regional registries maintain the framework for the "reverse" mapping from IP address to "machine name". The actual "reverse DNS" entries are of course held in a distributed database in the normal manner.

If the Owner is Unclear ?

- Traceroute may give a clue

```
 5    59 ms    61 ms    64 ms
                tele-border-12-168.router.demon.net
 6    65 ms    66 ms    63 ms  linx.u-net.net
 7    64 ms    61 ms    63 ms  194.119.177.228
 8   179 ms    66 ms    62 ms  213.2.253.5
 9    62 ms    61 ms    63 ms  212.188.191.1
10     *        *        *     Request timed out.
```

- ie: try to identify upstream providers

✶ *traceroute* works by sending ICMP echo requests (or sometimes UDP packets) with a small “hop count”. As the packets traverse each machine along the way to their destination the “hop count” is progressively reduced. When it expires, an ICMP “hop count exceeded” packet will be generated and sent back to the source machine. By progressively increasing the hop count it is possible to map the outgoing path.

✶ When the ICMP packet arrives, the IP address can be translated (via reverse DNS) to give the name of the machines on the outgoing path. These names are usually a good guide as to the ownership of the routers involved.

✶ *NB: The traceroute example in the slide is quite old, the U-Net customer involved has changed and there is now some proper reverse DNS for the machines that are currently on the path. The example has been retained because it shows how even if the remote machine’s identity is unclear, an organisation “upstream” from them can be easily identified. The upstream provider is likely to have suitable records for identifying the machine that is of interest.*

Traceability of Email

```
Received: from pop3.demon.co.uk by rnc1.al.cl.cam.ac.uk with POP3
 id <"happyday.1009968986:20:22479:12".happyday@pop3.demon.co.uk>
 for <happyday@pop3.demon.co.uk> ; Wed, 2 Jan 2002 10:56:39 +0000
Return-Path: <mvcic@caramail.com>
Received: from punt-2.mail.demon.net by mailstore for richard@happyday.demon.co.uk
          id 1009968986:20:22479:12; Wed, 02 Jan 2002 10:56:26 GMT
Received: from servovalle.ipvcov.cl ([164.77.204.218]) by punt-2.mail.demon.net
           id aa2022374; 2 Jan 2002 10:56 GMT
Received: from mx2.mortgageloanfast.com (slip-12-64-210-233.mis.prserv.net [12.64.210.233])
    by servovalle.ipvcov.cl (8.9.3/8.8.7) with SMTP id HAA18642;
    Wed, 2 Jan 2002 07:13:59 -0300
From: mvcic@caramail.com
Date: Wed, 02 Jan 2002 03:55:22 -0700
To: yearned@internetz.com
Message-Id: <31gb2y88su1gmy.7gaa6vrr2gt@mx2.mortgageloanfast.com>
Subject: Save Money on Your Mortgage Payment!
```


✶ As email passes through a mail system a “Received:” header line will be added to the top of the existing message. Inspecting the headers will therefore provide a trace of where the email has come from.

The formal format of the Received: lines is documented in RFC2821 & 2822, though in practice a fair amount of variability will be encountered. In principle you will be told the name of the machine generating the Received: line, when the line was added, where the email came from and who it was addressed to at that stage.

✶ It is not uncommon to see three different names presented for the machine from which the email came:

the name claimed (in the SMTP “HELO” line)

the remote IP address

the reverse DNS lookup for the remote IP address

The two names may differ for legitimate reasons, but in the example given, a sender of bulk unsolicited email has been (vainly) attempting to hide their tracks [they were using an AT&T dial-up in Dallas and exploiting an “open mail relay” in Chile].

✶ See a FAQ, eg: http://www.stopspam.org/email/headers/headers.html for more about reading email headers.

Traceability on USENET

```
Xref: news.demon.co.uk demon.service:134733
Path: news.demon.co.uk!demon!happyday.demon.co.uk!highwayman.com!richard
From: Richard Clayton <richard@highwayman.com>
Newsgroups: demon.service
Subject: Re: Duplicated Email again
Date: Wed, 2 Jan 2002 10:18:32 +0000
Organization: Highwayman Associates Ltd
Message-ID: <$K5ckpH45tM8EAuH@highwayman.com>
References: <32k53u8ddn500leo1b94g0l490djnpgb5a@4ax.com>
 <gpk53u87a813ubnsoslb02qqa3mjhehhlb@4ax.com>
NNTP-Posting-Host: happyday.demon.co.uk
X-NNTP-Posting-Host: happyday.demon.co.uk:158.152.30.53
X-Trace: news.demon.co.uk 1009967118 nnrp-07:19862 NO-IDENT happyday.demon.co.uk:158.152.30.53
X-Complaints-To: abuse@demon.net
X-Newsreader: Turnpike Integrated Version 5.01 M <7fPN0OjtGdv6AX1irDkRphaT6+>
```


✶ Almost everything in a Usenet article header can be forged by the sender. In the example given the only trustworthy headers are parts of the Path, the X-Trace line and the X-NNTP-Posting-Host header. For some news servers, even these can be suspect!

✶ The X-Trace line is usually the most reliable header. It will indicate which “proper” news machine first accepted the article. The format is not standardised. In this case the time of posting and the posting machine (a Demon dial-up) are identified.

✶ The Path gives the path taken by the article across Usenet (rather like the Received lines in email). Each new machine adds their identity to the front of the path. Note that some of the path (in this case “highwayman.com!richard”) may be “preloaded” before the article reaches a trustworthy server.

✶ Forged Paths can be detected by collecting articles from several servers and comparing them. With experience, the true injection point can be located.

Traceability on IRC

- Need to map nickname to server to IP address
- May be intentionally untraceable

- Different policy aims may be envisaged
  children should be anonymous
  dirty old men should not be anonymous

✶ There are a number of IRC networks and they differ in their policies. Some are actively hostile to any form of long-term traceability being present by refusing to keep any logs.

✶ However, abuse of the servers usually leads to “K:Lining” whereby an IP address (or address block) is barred from connecting – and this obviously requires an instantaneous mapping from abuser to IP address to be possible. ie: the lack of logs is not quite the same as a lack of traceability.

Identifying Dial-up Users

- Dynamic IP is commonplace
- RADIUS logs connect and disconnect
- Hence from time **+** IP can deduce account
- Various "gotchas"
  - UDP means logs incomplete
  - Time may be inaccurate
  - Logs are large and only kept short-term

27th November 2002 Finding

✶ Dial-up connections to the Internet are usually given a "dynamic IP address". ie: the IP address depends upon the modem port used or is taken from a small pool of addresses rather than being dependent upon the account being used for connection.

✶ One of the most common systems for authorising dial-up connections is called RADIUS (Remote Authentication Dial In User Service). The basics of this system are documented in RFC2138. The logs from a RADIUS server will typically contain the information about which IP address was allocated to each particular dial-up connection.

✶ Since IP addresses are re-used, if it is necessary to trace which account was using the address an accurate timestamp will be needed both at the remote machine where the IP address has performed an action and also for the RADIUS server logs. Accurate timestamps are straightforward to achieve by use of NTP (Network Time Protocol).

✶ RADIUS logs are non-trivial in size and so they are seldom kept for long periods. Therefore if it is necessary to trace an account it is important to do this relatively promptly.

More Practical Problems

- RADIUS and IP allocation may be done by different organisations
- Account may be generic (sales promotion)
- Remote machine may only have DNS record (and hence IP address is deduced)

✶ The modems and RADIUS servers may be operated by different organisations, both of whom may be different from the customer facing organisation that the customer believes to be their ISP. This can lead to some complexity if you are trying to establish the mapping between an IP address and an account name.

For example, in the early 1990s Pipex operated banks of dial-up modems for a large number of reseller ISPs (“pipettes”). Today BT operates “SurfPort” on behalf of a number of ISPs – delivering customer calls over their “Colossus” IP network. In the USA disintermediation is far more widespread with many local ISPs having a national reach through deals with the operators of the “modem banks”.

✶ Tracing to an account may not get you very far if the account is generic. ISPs issue CDs (it used to be diskettes!) with trial accounts on, all of which are identical. These accounts usually give very limited access just to a sign-up web site. However, if the accounts can access the open Internet then clearly traceability will be more problematic than otherwise.

✶ Traceability works with IP addresses. If the remote site has not recorded an IP address but a name then this can cause problems. If the name was provided by the connecting software then it may be forged. If it was recorded by the remote site instead of an IP address then one has to assume the address mapping has not changed in the meantime. This is not always the case.

Identifying the User

- Ask them for name and address
- Credit card info
- Telephone callback
- Other relationship (store card, account no)
- Caller Line Identification (CLI)

27th November 2002 Finding

✶ Having established which account used the IP address that "did something" then it is usually desirable to determine who was operating the account. This is not always the case – sometimes just knowing the account is sufficient; if it is an abuse incident (unsolicited bulk email perhaps) then the account will be suspended. The identity of the user is not relevant in such a case. However, a police officer seeking the poster of paedophile material will be interested in establishing who the user was.

✶ Most ISPs will wish to know your name and address before letting you open an account. They will probably check its internal consistency (does this postcode apply to this town?) to try and screen out grossly inaccurate responses. Online postcode databases make this check easy to evade.

✶ If you are paying for the account then it is likely that you'll be using a credit or debit card. This provides, through the banking system, traceability to a particular person.

✶ Free ISPs also like to identify their customers, both for marketing purposes or to prevent abuse. They may collect information like your Tesco card number in order to identify you. It would be unusual for a free ISP to allow dial-up connection without Caller Line Identification (CLI).

CLI

- CLI travels to all the telco switches
- At the user level CLI can be withheld (141)
- ISPs will be allowed (as 999 operators currently can) to see suppressed CLI "soon"

- **BUT** CLI tends to fail:
  on international calls
  at telco boundaries
  when using bulk carriers

27th November 2002 Finding

✶ Every phone line has an 'A' Number which is the "real" number of the phone line (that appears on the phone bill). A line can have several 'presentation' numbers as well, which may be selectable by the user when placing calls. A line may even be flagged to allow callers to present any number they wish – so called class C (or type 3) presentation numbers – OFTEL have to approve the use of this facility on a per customer basis.

A line also has a default CLI Presentation Restriction (CLIPR) state for an outgoing call, which will usually be one of [True | False]. This may be over-ridden on a per call basis (by dialling 141 in front of the called number).

When a call passes across a telco interconnect, an additional flag comes into play, which can be called 'CLIPR-Trusted'. This flag tells the terminating network whether the CLIPR flag can be trusted to be accurate.

The terminating telco (or a subscriber with the 'presentation override facility' eg a 999 operator – though they also have C7 level access) will always see the presentation number (or if none the A number). If CLIPR-Trusted is set then a subscriber will see the same number as the operator if CLIPR is false, otherwise they will see "withheld". If CLIPR-Trusted is not set then the subscriber will always see "unavailable".

✶ Note that at the C7 level the telco can see both the A number and the presentation number but only one can be passed across the Q931 interface to a NAS (Network Access System, the "modems" at the ISP).

Passwords

- Passwords are poor identifiers
  - ISP staff
  - household
  - post-it notes
  - Usenet
  - social engineering
- Accounts may be legitimately used by many people; so spotting extra use can be hard

27th November 2002 Finding

✶ Tracing an event via its IP address to an ISP account is not the same as locating the person who "did it". The account may have been in use by someone other than its owner. Account ownership is usually demonstrated by providing a password – and that password can be compromised in many ways.

✶ The ISP staff may be aware of customer password settings. Others in the same household or office may know the password. The password may have been inadvertently posted to Usenet (along with some other debugging information relating to a dial-up problem) or indeed the password may have been disclosed to someone plausible who just asked for it (a process usually known as "social engineering").

✶ Alternatively, the account may have multiple legitimate users and there may be insufficient records to demonstrate which of the users was responsible for a particular event. This may not be a problem to the ISP, who will close an account no matter which individual perpetrated some abuse, but it will be a problem to a police officer who needs to arrest the correct person.

Traceability on LANs

- A LAN is a broadcast medium
- Naïve to think MAC addresses are fixed
- Possible to steal MAC & IP addresses
- Hard to locate senders
  - big practical problem for DHCP
  - bridges know direction
  - can fingerprint the NICs

27th November 2002 Finding

✶ Ethernet LANs are a broadcast medium (using CSMA/CD to share the channel capacity). Nodes are identified by a Media Access Control (MAC) addresses which are supposed to be unique [the first 24 bits is the Organizationally Unique Identifier (OUI) assigned to a vendor and the second 24 bits is a serial number].

✶ In practice, nodes have to have assignable MAC addresses because 'DECnet' requires soft numbers (it does not have an ARP equivalent) and because modern driver software assembles complete packets (including the MAC address) before passing them to the hardware.

✶ It is possible to impersonate MAC and IP addresses on a LAN (though the machines that are really using them will tend to get upset, which may require DoS techniques as was the case when spoofing connections).

✶ Locating a node on an Ethernet is not trivial. If the network uses bridges or switches then they may provide a "management interface" that allows traffic to be localised to a particular LAN segment. Otherwise, it may be necessary to inspect all machines to see if they are transmitting. Intel say that NICs have sufficient variation in manufacture that unique "fingerprints" are possible – but this still supposes that you can actually locate all the machines that are attached to your wiring.

More Complications

- Network Address Translation
  - used to preserve IP address space
  - used to hide network architecture
  - unlikely to be logged
- DHCP
  - dynamic allocation of addresses
  - logging can be problematic

27th November 2002 Finding

✶ Network Address Translation (NAT) is widely used to conserve address space, to allow the operation of several machines on a single dial-up connection and for security reasons by ensuring that machines are not visible to the open Internet. The IP address recorded at a remote site is likely to be the address of the kit doing the NAT. Mapping this to a particular machine “behind the NAT” is unlikely to be possible since it is rare to record NAT assignments in logs.

✶ Even where machines are on the open Internet, their IP addresses may not be fixed, but may be dynamically allocated using a protocol such as DHCP (Dynamic Host Configuration Protocol). This means that an individual machine may change IP address from day to day. Keeping logs would be unusual. Keeping them for long periods would be more unusual still.

Authenticity

- Logs need to be authentic & correctly timed
- DNS needs to be trustworthy
- IP Allocations need to be documented
- Machines need to be secure
- Staff need to be trustworthy
  nightmare scenarios :
  chasing a sysadmin or ISP staff

27th November 2002 Finding

✶ Traceability is the process of following a chain of data, from IP address to ISP, to customer account, to end user. If any part of this chain contains dud data, whether through accident or design, then it will not lead to the correct account, let alone the correct person. Authenticity is therefore essential.

✶ The risks of relying on DNS remaining the same between when a log is created and when it consulted have already been mentioned. Further problems arise in assessing the authenticity of logs if the local provision of DNS can be subverted, perhaps by “cache poisoning” attacks. It is usually considered best practice to record raw IP addresses alongside any DNS results.

Retention & Preservation

- Data Retention is a matter for Data Protection legislation; have to show a business need
- Data Preservation is at the request of Law Enforcement to prevent auto-erase. It is covered in the Cybercrime Convention.
  - Work is going on within the G8 to provide trans-border requests and some form of fast divulge to allow multi-hop traceability. The principle is simple, but the details are complex and have yet to be worked out.

27th November 2002 Finding

✶ In the UK, retention of logging data is governed by the *Data Protection Act 1998* and *The Telecommunications (Data Protection and Privacy) Regulations 1999*. In general terms, under the DPA you may not keep data unless you have a business need to do so. The regulations set specific requirements for information relating to a "call". It is generally accepted that even where logs are not required for billing purposes, they can still be kept for a month (or six) in order to prevent "abuse" by customers. Thereafter they must be destroyed or anonymised.

✶ Keeping logs in case the police need them is not a business need. However, the *Anti-Terrorism, Crime and Security Act 2001* envisages a voluntary Code of Practice on keeping logs to prevent terrorism. If a voluntary code fails then the Secretary of State has powers to make it compulsory. The Act has encountered considerable problems and it is not yet clear what type of regime we will have.

✶ The Cybercrime Convention (first signed in 2001, yet to be ratified or come into force) contains provisions for data preservation (ie the storage of logs so that they are not destroyed) for up to 90 days and for "expeditious disclosure" of information that indicates the source of traffic.

http://conventions.coe.int/Treaty/en/Treaties/Html/185.htm

Review

- 2-way traffic makes an IP address trustworthy
- Registries and traceroute will locate ISP
- ISP logging will locate the account
- Account details will reveal user
- CLI will reveal dial-up user
- Local records (NAT/DHCP) will reveal a LAN user
  - BUT the last hop may not lead you to exactly the right person, especially if looking for a skilled adversary who can "frame" an innocent bystander

27th November 2002 Finding

✶ It should probably not be surprising that traceability over the "last hop" from an account to a user is poorly supported. Most of the traceability mechanisms are provided by ISPs and they discharge their obligations to the network by being able to locate a miscreant account and disable it. They have limited interest in locating a specific individual.

✶ For a discussion of how "last hop" traceability breaks down in a number of different Internet access technologies see my "Limits of Traceability" paper.

http://www.cl.cam.ac.uk/~rnc1/The_Limits_of_Traceability.pdf

"Practical Anonymity"

- Steal a password
- Use a free account and withhold your CLI
- Use a pre-paid WAP phone
- Use a cybercafe
- Use a LAN (maybe steal a MAC/IP address)
- Multiple jurisdictions will slow tracing down
- NB: Best Practice is far from universal

27th November 2002 Finding

✶ The Anonymity systems discussed in the first lecture prevent the secret police from knowing which of *n* people sent some traffic. When *n* is small they may lock them all up anyway.

One might reasonably take the view that where traceability fails, as in the slide, then there is some practical anonymity.