Tracing misuse

Clients

Most cases of misuse will be reported as originating from one or more electronic identities, for example e-mail or IP addresses. Such identities are public, and can be seen by anyone on the Internet, but in most cases it will only be the local site that can relate these electronic identities to the individual responsible person. Ensuring that the actions of these electronic addresses can be assigned to responsible individuals is therefore fundamental to any attempt to track down network misuse. Since most electronic identities can be forged with various degrees of technical difficulty (e-mail addresses are easiest, IP addresses used for UDP packets slightly more difficult and IP addresses used in TCP connections significantly harder) it is also important to collect, as a matter of routine, sufficient reliable information to be able to prove when a forgery has taken place and thereby remove blame from an innocent individual.

Individuals are usually identified to computers by login name or e-mail address, which are not usually the identifiers that are used in complaints, so a conversion process will usually be needed to identify the individual who may be the subject of a complaint.

In the simplest case a reported IP address will be the fixed address of a workstation in a private office. The owner of the office will normally be responsible for activity by that IP address and only a record of the ownerships of allocated addresses will be needed.

The situation is more complicated where a number of different users may use the same computer, either because the computer is a system that supports multiple users simultaneously, or because it is a public workstation that may be used by different people at different times of day. In the latter case, it should be possible to identify a single login account that was logged in on the workstation at a particular time. To achieve this it is essential to have login records that can be searched by workstation address and time. A record of the ownership of login accounts and e-mail addresses is, of course, a basic management tool. It is recommended that all users be made formally responsible, and accountable, for the activities of accounts allocated to them.

The next level of complication arises when different client computers use a single IP address at different times. This occurs whenever a pool of IP addresses is shared between a number of computers, for example in dial-up, mobile or fixed networks where addresses are allocated temporarily to active computers (the DHCP protocol is commonly used to manage addresses in these situations). Here it is essential to have logs of which client computer was allocated each IP address and the times when the use of the addresses began and ended. Once the workstation has been identified, records of logins and times can once again be used to identify the responsible person.

Summary of Logs

Type of System	Logs required
Single-owner workstation, fixed IP	• IP -> owner
Shared workstation, fixed IP	• IP + time -> login login -> owner
Dynamically configured workstation	• IP + time -> workstation • workstation+time -> login • login -> owner

Federated Authentication Systems

Federated Authentication Systems allow an organisation providing a service to a user to rely on another organisation (typically the user's home organisation) to authenticate the user, rather than maintaining its own local database of usernames and passwords. Examples of federated authentication systems include the UK Access Management Federation (http://www.ukfederation.org/ [1]), used by publishers to authenticate access to online resources, and eduroam (http://www.eduroam.org [2]), used by education organisations across Europe to authenticate visitors from other organisations and provide them with network access.

Some federated systems are designed to preserve the anonymity of the user. In these cases the organisation providing a service is merely told 'yes, this is one of our users', accompanied, if necessary, by attributes that may be required by the service such as whether the user is a student or member of staff. Such systems add an additional step to the process of tracing access through logs, since the organisation that gives the authenticated user access to its service may not itself be able to link the service provided to the user it was provided to (the step resulting in a 'login' name in the previous table). Instead this requires two steps: the service provider needs to identify the home organisation and the home organisation needs to identify the user they authenticated.

Details of the logs that are needed will vary between different federated authentication systems and should form part of the federation agreement. In general, service providers need to record at least the home site that provided the authentication for a particular service request, together with other identifying material such as the service that was requested and any identifier that was provided by the home site. Home sites that are providing authentication to external organisations need to record at least the source and time of the authentication request, the same identifying information for the request, and the local user who was authenticated.

Intermediaries

The types of client logging discussed above deal with the situation where a direct TCP/IP connection exists between the client machine and the server. However there are also a number of services and configurations where some other machine is involved as an intermediary between the client and the server. Additional logs are needed in these cases as the end server will see the activity as originating with the intermediary, while the client logs will only show a connection to the intermediary and not to the end server where the alleged misuse occurred. Intermediary logs are required to link these two sets of information.

The most obvious examples of intermediaries are proxies and caches; store and forward systems such as e-mail servers also act as intermediaries. There are also other systems, particularly those that act as gateways between different Internet services, that may act as intermediaries and therefore need to record appropriate logs.

Some intermediary systems handle requests for very large numbers of clients and servers so that a simple timestamp may not be sufficient to identify a communication uniquely. Logs on these systems will often need to record additional details of each transaction, such as URLs for web requests or subjects of e-mail messages, to allow a particular communication to be identified. Complaints where these details are not included are likely to be very hard to investigate.

Proxies and Caches

Proxies are systems explicitly designed to act as intermediaries. Clients make requests to a proxy and the proxy may send that request to a server on behalf of the client. Proxies are usually designed to support one or a small number of protocols, for example HTTP and FTP, and, unlike gateways, use the same protocol for the requests they receive and send. A caching proxy may respond directly to some requests as an alternative to passing them on. A filtering proxy may determine that either the request or response is not appropriate according to a set of rules and may either block it or replace the response with a warning.

As the server’s records will show the request coming from the IP address of the proxy, the proxy must itself retain a log of the client IP address or authenticated user on behalf of which each request was made. A busy proxy will often make many requests each second to a popular server so the time of the request and identity of the server will not be sufficient to identify an individual client request uniquely. It is therefore normal for this type of proxy to retain additional information about each request, for example the web URL, that will allow the responsible client to be linked to the information recorded by the server. Where the protocol and local policy permit, a great deal of investigation time can be saved if the proxy includes the client address in each request it passes on. If this is visible to the end-user, or recorded in the logs on the final server, then there will be no need to search through large volumes of proxy logs.

E-mail and News

Store and forward systems, such as news and electronic mail, differ from proxies in that the transaction with the client is completed before the message is passed to another server.

However they still act as intermediaries so should retain a log of the client from which each message was received and the server or other destination to which it was passed. In theory each mail or news message includes as part of its content a full record of its origin and path: however, this information is relatively easy for a malicious user to forge. Trustworthy logs kept by servers are important tools in detecting this kind of forgery. This may be especially important if messages are forged so as to appear the responsibility of an innocent party. As with proxies it is common to record the message subject or ID to ensure correct identification of a particular message.

Network Address Translation

Network Address Translation (NAT) is another type of intermediary, but one that works at the network rather than the application layer. Clients of an address translation system usually have private IP addresses, as defined in RFC1918. Clients send any packets for external destinations to the NAT system, which rewrites them with a public source IP address and forwards them to their destination. The NAT system must of course remember the state of each communication so that when it receives response packets it can rewrite their destination addresses and send them to the correct client.

NAT systems can use a variety of strategies to allocate external addresses to internal clients. Some create static mappings between external and internal addresses for each client; more commonly the system will behave like a proxy, using one or a small pool of external addresses for all communications. Address translation systems have the same basic logging requirement as other intermediaries: to be able to relate a request made by the translation system to the client that invoked it. However the complexity of mapping used by some systems can make this a challenge. Attention to traceability requirements must therefore be included at the design and implementation stages of any address translation system.

Gateway Servers

The final class of intermediaries is gateway systems that take a request from a client in one form and use it to generate a request in another form to a server. The most familiar gateway at present is probably a webmail server, which takes an HTTP form submission and uses it to generate an SMTP request. The command sent by the gateway will not necessarily contain any information originating from the client, so it is particularly important that gateway servers keep reliable records of their activities. Some gateways may add client information to their output – for example webmail servers often include the client IP address as a header in the SMTP mail – but it is important to know how much of this information can be relied upon, and how much is under the control of a potentially malicious user. Of course if the gateway itself is compromised then nothing about either its output or its logs can be trusted. Some systems can be made to act as accidental gateways, for example a badly configured web cache may allow e-mail forgery. Such systems, and others with inadequate logging, are a hazard to the Internet as they provide abusers with complete anonymity for unauthorised or illegal activity.

Outsourced Servers

It is increasingly common for organisations to outsource some of these services (for example e-mail or filtering proxies) to third parties. Where this is done, it is in the interests of both parties to ensure that adequate logs are kept and can be used to investigate problems. As

with federated access management, this is likely to involve collaboration, since the logs of what happened and who was involved are likely to be held by different organisations. Outsourcing agreements should ensure that the different logs contain the information necessary to link them and that responsibilities for doing this are clearly defined.

Summary of Logs

Type of System	**Logs required**
Proxy server	• destination IP address + time + reque client/user
Mail or news server	• destination IP address + time + reque client/user
NAT/PAT server/SOCKS proxy	• source & destination IP address (+po client/user
Gateway server	• server IP address + time + request ->

Source URL: https://community.jisc.ac.uk/library/janet-services-documentation/tracing-misuse

Links
[1] http://www.ukfederation.org/
[2] http://www.eduroam.org/