Tracking the Location of Internet Addresses

Nigel Carson, Ferrier Hodgson

Nigel Carson is the Director, Forensic IT at Ferrier Hodgson and has broad experience in Computer Forensics and Information Security.

The Internet is a sprawling beast of electronic pathways and resources that has spawned a frenzy of online business development, legitimate or otherwise, by breaking the traditional cost barriers of time and distance in information transportation.

This new and sometimes faceless world presents a rapidly evolving set of challenges for litigators. This article hopes to demystify the vagaries of the digital revolution as they relate to litigation matters.

Importance of Internet Locality

One of the key concerns is the confusion in tracking the location of potential parties in litigation through their connection with an Internet address or domain name. Typical situations where this occurs include the following:

- an incriminating email is received and then repudiated by the apparent sender. The sender's name is suspected to be forged and on closer inspection of the email, information technology (**IT**) staff confirm that the originating Internet address does not match the sender's known location. Further investigation is required to locate the origins of the email;
- a website provides services that infringe copyright or the intellectual property of another company and court orders are sought to be executed at the website's physical location. It is uncertain exactly where this location is, since the website appears to be run by a customer of a large Internet service provider (**ISP**) with national coverage. Further investigation is required to perform execution at the ISP's office closest to the target website to identify the customer address;
- a commercial website is defaced, made unavailable or otherwise interrupted from performing business. IT staff examine firewall logs and determine the suspected attacker's Internet address. Further investigation is required to determine if the attacker is within local legal jurisdiction and whether there may be some prospect of prosecution and reclamation of damages; and
- a software house releases a new version of a popular commercial tool. Within days the anti-piracy product activation component of the product is compromised and a patch is released to enable free use of the tool. An individual within a pirate group boasts of the exploit and the only investigative lead is an Internet protocol address (**IP Address**) and a nickname. The next step of the investigation is to track the IP Address to a physical site.

This article, which is mostly founded on my personal experience in investigating Internet matters over the last 10 years, briefly describes some of the technologies and methods that can be used to determine, with a high degree of confidence, the geographic location of an Internet entity, such as a website or personal computer.

There are two primary types of information associated with Internet communications that can identify a person or business, the IP Address or the domain name (leaving aside any obvious information, such as the contents of an email).

IP Address

Each and every device or computer that actively participates in Internet based communications is identified by an IP Address which at any given time should identify the device or computer as a unique entity on the Internet. Currently, most IP Addresses are expressed as four dotted decimal numbers in the range 0 to 255, eg 207.46.250.119.

It is hard to draw a good physical analogy to an IP Address. Comparing an IP Address with a street address works in the sense that information delivery on the Internet is somewhat like the postal system, with data being encapsulated in delivery 'envelopes' and containing a source and destination address. The problem is that once allocated to an entity (being a house or primary occupant), street addresses don't suddenly get allocated to your neighbour or somebody in the next suburb. IP Addresses on the other hand are, for the most part, dynamically allocated, usually by the upstream ISP and are quite transient in nature. The dynamic allocation of IP Addresses allows ISPs to service more clients than they normally could if each customer was assigned a static or unchanging IP Address. Even those of us that enjoy the stability and speed of broadband services, such as those provided by cable modems, are still prone to fluctuations of their IP Address. This is transparent to the user and the change will mostly go unnoticed.

An IP Address is usually assigned to a customer by an ISP which has been allocated a range of IP Address numbers by the relevant Internet authority or Regional Internet Registry (**RIR**). The Australian Pacific Network Information Centre (**APNIC**) is one of 5 RIRs operating in the world that allocates IP Address space.

The basic mechanisms of IP Address allocation and assignment are quite simple. APNIC have a very large number of IP Addresses which they allocate in ranges to top tier ISPs or similar entities, who in turn may sub-allocate smaller ranges of numbers to downstream ISPs or assign them

directly to customers, such as residential subscribers.

Companies will often request an assignment of IP Addresses from their ISP to cater for their Internet exposure. Blocks of 256 sequential numbers are very common. Residential users will draw a single IP Address out of a pool that their ISP has reserved for such use.

APNIC and other Internet bodies provide online services to query the allocation of IP Address blocks through what is known as **WHOIS** queries, and this is often the first step used in the investigation process.

A sample of a typical WHOIS query of an ISP customer's IP Address is included in Figure 1.

Figure 1

Query:

Queried **whois.apnic.net** with "**144.136.68.xxx**"...
% [whois.apnic.net node-2]
% Whois data copyright terms http://www.apnic.net/db/dbcopyright.html

Response:

inetnum: 144.136.0.0 - 144.136.255.255
netname: [*ISP's netname*]
descr: [*ISP name*]
descr: [*ISP postal address*]
descr: [*ISP location*]
Domain Name: [*ISP domain name*]

The allocation range in Figure 1 contains about 65,000 IP Addresses. Each ISP would have many of these ranges to assign to customers.

A sample of a typical WHOIS query of an IP Address from a business (**Business**) that has been assigned a block of IP addresses is included in Figure 2.

Figure 2

Query:

Queried **whois.apnic.net** with "**210.8.xxx.253**"...
% [whois.apnic.net node-2]
% Whois data copyright terms http://www.apnic.net/db/dbcopyright.html

Response:

inetnum: 210.8.0.0 - 210.11.255.255
netname: [Business A netname]
descr: [Business A name]
descr: [Business A address]
descr: [Business A address]
country: AU
inetnum: 210.8.xxx.0 - 210.8.xxx.255
netname: [Business B netname]
descr: [Business B name]
descr: [Business B address]
country: AU

Notice in Figure 2 that additional records have been included to describe the sub-assignment of 256 IP Addresses to Business B by Business A. Business A in this case is a local Internet authority and Business B is the current owner of the specific IP Address being searched for. Not all WHOIS queries or services will return the same set of records as there are numerous options and flags that can be used to gather more complete data. WHOIS services are also used to return detail about domain names, as seen in the next section.

An IP Address can be used to reveal details of location through a WHOIS query. A residential user WHOIS query is likely to provide the address of an ISP while a business user WHOIS query may provide the business address, depending on the detail available in the WHOIS records.

This information is not reliable by itself, and additional investigation is required to give confidence about any geographic addresses revealed. As an example, if a company is administered by a Melbourne based ISP, communications that it sends from a Sydney office may appear to be coming from Melbourne when an IP Address WHOIS query is used.

Domain Name

IP Addresses make sense in the mathematical world that computers are based in but they are not easy for people to remember and not easy to communicate. For this and other reasons, a system of name resolution was devised whereby an Internet resource, such as a web server, could be referenced by a friendly name called a domain name. Instead of typing http://207.46.250.119 into your web browser, the domain name system allows you to type http://www.microsoft.com. The computer will then use a domain name server (**DNS**), usually provided by the ISP, to lookup the IP Address that is matched to the www.microsoft.com domain name and allow the communication. Conversely, to determine if an IP Address is associated with a domain name, a reverse domain search can be performed.

It is important that domain names be regulated to avoid conflicts, duplicates and other anomalies. The Internet Corporation for Assigned Names and Numbers (**ICANN**) has assumed responsibility for the tasks of globally managing IP Address and domain name space. ICANN has delegated authority to local Internet authorities known as Registrars to manage much of the day to day administration of registering domain names. For example, Melbourne IT is one of many Registrars accredited by ICANN to register the Top Level Domains of '.com', '.net' and '.org'. It also has authority to manage Australia's commercial domain space – 'com.au'.

There is no requirement for a website to be registered with a domain name and many personal websites that run out of residential premises are not so registered. It is also possible to register a domain name and never have an Internet presence. Many simple dot com names were bought many years ago and 'parked' in the hope of finding a buyer, possibly a business, where the name has some importance.

Location specific information can be found in domain name records using WHOIS queries. This concerns investigating who has registered a given domain name rather than what entity a range of IP Addresses has been assigned to.

These records may not indicate any street address, and in any case, due to the size of some companies, such information might be of limited value. If an address is present, it is usually the head office of the company or the centre of administration.

There is a relatively new extension to domain name resource records that

allows the inclusion of longitude and latitude data that is attached to a domain name. This provision may one day be used to assist in geographical tracing but it will take many years to populate and verify such records.

Websites

One needs to be wary when seeking locality information based on a website of a business because the website often resolves to a location well away from the business itself, often at a remote data centre of an outsource provider. Even large ISPs outsource hosting of websites and where an investigation is based on the contents of a website it may be that the only feasible way to obtain evidence from the website is through remote access, provided by the site administrator.

Web servers are often rack mounted inside secure cages of large data centre sites and are not readily accessible. An ability to remotely access and download data from a remote site should be included in any search orders.

Geographic Certainty

There is no certainty of the physical location of an entity or device bound to an IP Address. Unlike a street address, an IP Address does not inherently contain information about location. The best that can be done is to establish a degree of certainty or confidence in the location of such an address through a process of investigation.

In the previous sections it has been demonstrated that locality information can be determined through the use of domain name records either by a reverse WHOIS query on an IP Address or a domain name WHOIS query with a more user friendly domain name. If the information provided is not a location, it will often give phone numbers, email addresses or other information that could lead to a location.

The information that is provided cannot be relied upon as to date there has been very little regulation of the accuracy of names that are submitted to the relevant authorities at the time of allocation or registration. The Australian domain space is more reliable than the '.com' domain space because it is more onerous on the registrant to provide justification for the use of the name. The domain name space is full of false information and the smaller the Internet presence, the less accurate the information is likely to be.

Other limitations with WHOIS records are that often a very large block of IP Addresses may have a single entry for the entire block despite the owning company having hundreds of offices around the world. Different WHOIS servers may contain inconsistencies, with some records becoming stale.

The investigation of an IP Address location is more likely to succeed where the investigation is not based on purely historical data but where the IP Address is still active. A web site under investigation will likely remain active and relatively stable. While the IP Address may change, the web site domain name will remain stable and this can be used to continue checking the state of the site.

Trace Route

After determining whatever information can be derived from the IP Address or domain name, a much better way of determining geographic location in the case of a live IP address is to conduct a trace to the target IP Address or domain name with the use of a trace route tool.

Trace route tools use a feature of Internet traffic known as Time To Live to map out all the individual routers or hops between routers between the investigator's computer and the target IP Address.

Trace route software comes preloaded with almost every operating system. Windows users can type "tracert [IP Address or domain name]" into a command line console to access this function. There are many excellent commercial products that will also perform this task, perform WHOIS searches and possibly provide a visual reference on a world map. The visual maps tend to be based on the reliability of domain name records so they suffer the issues already mentioned and tend to be of novelty value only.

Information derived from a trace route command will provide a number of valuable geographic indicators in the form of:

- number of hops to the target computer;
- domain names associated with routers along the way, that often provide geographic references; and
- time of transit of information for each step of the way.

A sample of trace route information using the Windows tracert command is included in Figure 3.

Figure 3

3 4 ms 4 ms 5 ms atm1-0-603.cor4.hay.connect.com.au [210.9.210.14]

4 6 ms 9 ms 9 ms ge-0-1-1.bdr5.hay.connect.com.au [203.63.217.82]

5 6 ms 8 ms 9 ms hay-telstra.gw.connect.com.au [203.63.130.249]

6 9 ms 5 ms 8 ms Port-Channel201.chw38.Sydney.telstra.net [203.50.19.

7 7 ms 9 ms 5 ms telstr196.lnk.telstra.net [139.130.35.10]

8 6 ms 9 ms 6 ms CPE-61-9-195-10.nsw.bigpond.net.au [61.9.195.10]

9 9 ms 9 ms 10 ms 61.9.207.210

Each line in the figure above represents a 'hop', or step, towards the target. Each hop represents an Internet router device encountered along the way. This trace represents output from a trace route starting at hop 3.

The number of hops to the target is an indication of geographic distance. If user A traces to a computer within A's own network there will be a single hop to the target with almost no delay. If user A traces to an address on another ISP network there will be several hops to the target, often several hops just navigating out of the originating ISP network. While there is no direct formula that can be applied, there is a strong relationship between the number of hops and the geographical distance.

Forwarding routers and devices that perform the "traffic warden" job of directing communications across the Internet are often registered with meaningful domain name entries which provide further valuable information. The city/country/state, the function of the router or the type of network topology in use may be indicated. In the example in Figure 3, at hop count 8, the name of the device at bigpond suggests that it is based in New South Wales and at hop 6 there is a device appearing to be in Sydney, further corroborating the location. Sometimes there are indications as to what suburb the connection is in through an abbreviation or whether the connection is through cable modem, dial-up or digital subscriber line. This additional information may allow further queries to be made that can pinpoint the exact suburb or even a housing development. There may, for instance, be a new cable rollout that is indicated in the domain name by a specific abbreviation. Contact with the ISP may provide a list of useful abbreviations.

The latency time information has not been considered to be very reliable due to the fact it may be influenced more by network congestion than the distance of the media run. In general it has been shown to have a direct relationship to geographic distance but several samples may need to be taken at different times of the day.

After conducting a trace route and analysing information from domain names of routers along the way, it is likely that the investigator has determined, as far as possible, the geographic location of a given IP address. Sometimes a standard trace route will fail as it gets close to the source since a network device along the path, often at the perimeter of the target network, is dropping the trace route queries and not allowing a response. There are techniques to circumvent this problem.

Looking Glass / Remote Servers

Looking Glass servers are resources on the Internet, often provided as diagnostic tools by ISPs, to determine congestion at various nodes around the world. Many of these Looking Glass servers can be accessed publicly and are useful tools for performing a more detailed tabulation of hop latency. Trace routes can be conducted from several servers at different world sites to help gauge the accuracy of an estimated location. If a server exists in the city the target is expected to be in, then a quick response and relatively few hops when conducting a trace route from that Looking Glass would be expected. Comparisons can be made from different world sites to give a greater degree of confidence than obtained with a single location trace route.

Investigators working in large international organisations can often ask colleagues to run traces from their remote computers and forward the results for consolidation.

Additional Research

Additional information can be derived from IP Addresses to geographic mapping databases maintained around the world. These databases appear to be mostly derived from domain name information matched to known or assumed physical information and some guessing based on topological routing clusters. IP2Geo is a technology based on several mapping techniques involving DNS names, network latency, host-to-location and router prefix information to infer the location of a target host.[1] This appears to be a relatively new and untested technology that makes informed guesses about location based on the 3 underlying technologies of:

- GeoTrack (based on domain names of target and nearby hosts);
- GeoPing (based on network latency measured from geographically distributed locations); and
- GeoCluster (based on IP Address to location mapping information from Internet routers using Border Gateway Protocol and other sources).

Another method of determining location is to identify and derive information from computer services running on the target host. It may be a computer running a time synchronisation function or other service advertising the local time zone of that computer. The target computer may run other services, such as a web server or a file transfer protocol service that contains banners or other information about the location of the computer.

Fundamental Limitations

There is a fundamental limitation to the exact determination of a given computer on the Internet because of two primary factors:

- the IP Address of a communication may be hidden through the use of proxy computers, network address translating routers and other technology that will hide the final destination at a network perimeter; and
- the IP Address may be dynamically allocated and shared over a period of time amongst unrelated users.

A scenario where the former might occur is where a residential Internet user shares his broadband connection internally within the residence to several other computers. There may be a wireless access point set up on the internal network that additionally shares the Internet connection with the neighbour across the road. While this may be a breach of the network services agreement with the ISP it is not unheard of. The customer's broadband router connects to the ISP and is assigned a valid world routable IP Address. All the computers that connect to the broadband router through its wireless access point or hard wired network ports will be given a private IP Address that will never be used on the Internet and the router maps the internal addresses to outgoing and incoming communications, with everything appearing to come through the single point with a single IP Address. Through a process of investigation, the address of the residence may be provided by the customer's ISP and an order served to search computers. The originating computer from which a communication was sent would then need to be located by a physical site search in addition to keyword or other searching on the computer itself. If the communication came from the neighbour's computer through the

wireless access point, then the terms of the order would probably restrict the ability to proceed further with the search.

A similar scenario would be a university site where, due to the complexity and segregation of internal networks, there may be some substantial work to be done once the campus site is located.

The latter case, where IP Address allocation is dynamic, will often be an issue when tracking down residential ISP customers, particularly dial-up users who may use several IP Addresses a week. By recording the time and date of communications, the investigator can request that the ISP checks their allocation logs that will map the IP Address to the customer account. ISPs can do this through call charge records or dynamic host configuration logs. The customer account will normally contain street address details. Search execution will be a two step process, firstly at the ISP premises and secondly at the address to be supplied by the ISP, based on the IP Address corresponding to the time and date details previously collected.

In the case of dynamic IP Address allocation, the more recent the time and date details collected, the more likely the chances are of an ISP mapping these details to an account. ISPs only maintain mapping records for a certain period before discarding them and it is advisable for an investigator to confirm the ISP's retention policy.

Conclusion

Domain names and underlying IP Addresses do not in themselves provide reliable location specific information. However, in conjunction with the use of tools and investigation techniques they are likely to provide a geographic address of a site within reasonable proximity to the target. Some important tools, such as traceroute, will require that the site is still active and the investigator has knowledge of the current IP Address or domain name.

The last hop in a trace route of an IP Address or domain name may be the final and desired destination or it may be a stepping stone in the process of locating the relevant computer. Therefore any search orders sought should allow for additional exploration and remote access to data. Advice should be sought from technology experts where there is any doubt about the assumed location of Internet communications.

---

1 ACM Digital Library technical whitepaper "An Investigation of Geographic Mapping Techniques for Internet Hosts". Available at http://www.acm.org viewed 4 may 2005.