HOW DOES THE ACCURACY OF GEO-LOCATION TECHNOLOGIES AFFECT THE LAW?[*]

by

DAN JERKER B. SVANTESSON[**]

Geo-location technologies allow website operators to identify the geographical location of those who visit their websites. Having knowledge of an access-seeker's geographical location means that they can provide content targeted to that location. This has several uses. For example, it enables the website operator to:
I. Provide advertisements relevant for the access-seeker's particular location;
II. Restrict access to content that may be unlawful in certain jurisdictions;
III. Restrict access to content that the website operator is licensed to provide only in a limited geographical area; and
IV. Avoid entering into transactions with people from locations known to be "fraud hot-spots".

The accuracy of geo-location technologies is obviously of fundamental importance for all these uses. However, the consequences of providing advertisements aimed at the wrong location may be far less serious than, for example, failing to restrict access to content that the website operator is licensed to provide only in a limited geographical area. In situations where a website operator seeks to rely on the use of a geo-location technology to argue that it has met its legal obligations, the accuracy levels of the geo-location technology used may indeed be determinative.

---

[*] This paper draws upon several other publications on the topic of geo-identification. See in particular: Svantesson, D. 2004 'Geo-location technologies and other means of placing borders on the 'borderless' Internet', John Marshall Journal of Computer & Information Law, Vol XXIII, No 1, pp. 101 – 139, and Svantesson, D. 2007 'Geo-identification and the Internet – A New Challenge for Australia's Internet Regulation', Murdoch E-Law Journal Vol 14, No 2, pp. 155 – 177.

[**] Associate Professor, Faculty of Law Bond University (Australia). E-mail: Dan_Svantesson@bond.edu.au, Website: www.svantesson.org.

Having provided a brief overview of how geo-location technologies work, this paper examines how accurate they are. As part of that discussion, recommendations are made for how courts ought to approach the use of geo-location technologies.

KEYWORDS
Private International law, geo-identification, geo-location, Cyberspace regulation

INTRODUCTION [1]

I first attended Masaryk University's Cyberspace conference in 2004. At that occasion I spoke about so-called geo-identification. Geo-identification is the practice of identifying Internet users' geographical location. This can be achieved in a number of ways. Most significantly, it can be done be technical means without the Internet users' knowledge.

Since my 2004 conference paper, I have continued to research legal aspects of geo-identification. In this paper, I will provide an update of some developments since my 2004 paper. Particular focus is placed on the accuracy of geo-location technologies.

HOW DOES IT WORK? [2]

My 2004 paper titled "Geo-location Technologies – A Brief Overview" gave an overview of how technical means for ascertaining Internet users' geographical locations work. Seeing how the rest of this paper requires some knowledge of this issue, I will provide a very brief repetition here.

Currently, the most relevant form of geo-location technology is based on the translation of Internet Protocol (IP) addresses into geographical locations, by the use of information stored by the provider of the geo-location service. As the access-seeker enters the appropriate Uniform Resource Locator (URL) into his/her browser, or clicks on the appropriate hyperlink, an access-request is sent to the server operating the requested website. As the server receives the access-request, it, in turn, sends a location request (e.g. forwards the access-seeker's IP address) to the provider of the geo-location service. The provider of the geo-location service has gathered information about the IP addresses in use, and built up a database of geo-location information. Based on the information in this database, the provider of the geo-location service gives the website server an educated guess as to the access-seeker's location. Armed with this information, the web server can provide the access-seeker with the information deemed suitable (e.g. a mes-

sage along the lines of: "Sorry. This website is intended for the people of Czech Republic only", or perhaps provide advertisement specifically targeted at people from the access-seeker's particular location). There are currently several products on the market utilising this type of systems[1]. This technology is not necessarily prohibitively expensive, nor is it particularly difficult to operate.[2]

THE ACCURACY [3]

The accuracy of sophisticated geo-location technologies is difficult to gauge. So far, there is a paucity both of independent studies, and of occasion where a court has evaluated the accuracy of geo-identification. One of the most important cases, in this context, is the French Court's judgment in the *Yahoo!* case.[3] There, experts concluded that "it may be estimated in practice that over 70% of the IP addresses of surfers residing in French territory can be identified as being French."[4]

In the *Nitke v Ashcroft* case,[5] the Court was again assisted by expert testimonies. One such expert, Seth Finkelstein concluded that:

> *A provider of content via the Internet cannot reasonably be expected to know the location of readers, if the context is one in which location would lead to a denial of the ability to read the content.*
>
> *This is because material can be read on the Internet through many alternate geographic routes, where the content can intentionally be relayed through third party intermediaries which act to mask and obscure location. Further, intrinsic inaccuracies such as changes in address assignment and proxying*

---

1 See e.g. http://www.quova.com/, and http://www.digitalenvoy.net/.

2 The author does not have sufficient information, and is anyhow not qualified, to independently assess the accuracy of these products. But, for example Geobytes' product is available from $500 US per annum, appears easy to operate (see demo: http://www.geobytes.com/demo.htm), and the producers argue that the product is accurate to 97% on country-level and 75-80% on city-level. See further: http://www.geobytes.com/.

3 *International League Against Racism & Anti-Semitism (LICRA) and the Union of French Jewish Students (UEJF) v. Yahoo! Inc.* County Court of Paris, interim court order of 20th of November 2000.

4 *International League Against Racism & Anti-Semitism (LICRA) and the Union of French Jewish Students (UEJF) v. Yahoo! Inc.* County Court of Paris, interim court order of 20th of November 2000. However, it would seem that one of the experts, Ben Laurie, later felt a need to explain his statement. (Laurie, B. '*An Expert's Apology*', at http://www.apache-ssl.org/apology.html at 14 August 2007.

5 *Nitke v. Ashcroft*, 253 F.Supp.2d 587 (S.D.N.Y. Mar 24, 2003) (NO. 01 CIV. 11476 (RMB).

by such large providers as America Online (AOL) means many users cannot be reliably located.[6]

Despite this, providers indicate the potential accuracy to be very high. For example, Digital Element claims that their product NetAcuity is over 99.9% accurate at a country level and over 94% accurate at a city-level.[7] Quova – another leading provider of geo-location technologies – state that: "In audited tests [carried out by PricewaterhouseCoopers] using large, independent third party data sets of actual web users, Quova's country-level accuracy was measured at 99.9% and its US state-level accuracy at 95.0%."[8]

Such impressive statistical figures have been criticised.

The way the vendors arrive at their accuracy statistics is to cross-check the physical location of sampling of Internet users (as determined by their software) against customer provided locational information already in the possession of the software vendors. There is no way to independently verify whether the software could provide the claimed levels of accuracy if the software vendors didn't first have other customer location information which their software may be using to determine customer location. Put somewhat differently, it is as if a "psychic" claimed to be able to accurately know what card a customer held in their hand 99.5% of the time, and to prove it, the psychic would ask to see the cards in the hands of a sampling of customers before announcing that indeed those were the same cards he knew the customers to possess.[9]

Furthermore, some producers of geo-location technologies appear to base their accuracy rates on the risk that their positive guesses are incorrect. Imagine, for example, that the operator of a Swedish website wished to make the website accessible to people located in Sweden only. A provider of a geo-location technology may state that its product is accurate to 99%, referring to the fact that, of the people the technology identifies as located in Sweden, 99% of them will actually be located in Sweden (i.e. the number of

---

6 Expert testimony by Seth Finkelstein (10 November, 2003) http://sethf.com/nitke/ashcroft.php at 14 August 2007.

7 http://www.digital-element.com/ip_intelligence/ip_intelligence.html at 10 August 2007.

8 http://www.quova.com/page.php?id=132 at 10 August 2007.

9 Information Technology Association of America, '*ECommerce Taxation and the Limitations of Geolocation Tools*', at http://www.itaa.org/taxfinance/docs/geolocationpaper.pdf 5 February 2007. While the details of PricewaterhouseCoopers' audit are not publicly available, it is possible that this criticism does not apply to the statistics presented by Quova, if the reference information was completely disconnected from Quova's data collection.

false positives is 1 in 100). However, this figure does not say anything about the rate of false negatives; that is, it does not reveal how many people, actually located in Sweden, will be refused access. It is, thus, similar to claiming to be able to tell if a person is male or not with 99% accuracy, and then only nominate people with extensive facial hair as males – the number of false positives is likely to be very low, while the number of false negatives may be high. Consequently, courts must be wary of accuracy rates presented by manufacturers of geo-location technologies.

In addition, when assessing the accuracy of methods for geo-identification, it is important to avoid placing the focus on the marketing-driven *average* accuracy-rates presented by the companies behind the method in question, and instead pay attention to the *context-specific* accuracy rate.

> *If a company were to assert that its method is, for example, '98% accurate' on average across all its applications involving analysis of locations throughout the world, it is likely that the accuracy rate for Canadian and American location distinctions alone is lower than 98%, given the unique difficulties in this context.*[10]

Placing the focus on context-specific accuracy rates will inevitably complicate and increase the cost of court proceedings in that expert evidence may be required in each individual case. However, the importance of ensuring that the courts base their decisions on the accuracy rate that is relevant in the particular case at hand cannot be overstated. In other words, the first advice I provide to judges faced with disputes involving the use of geo-identification is to encourage the use of expert witnesses testifying as to the context-specific accuracy rate for the case at hand.

There is a range of factors affecting the accuracy of geo-location technologies. Due to the dual nature of the geo-location process, these factors can be divided into two categories: 'source problems' and 'circumvention problems'.

---

[10] Edelman, B. '*Shortcomings and Challenges in the Restriction of Internet Retransmissions of Over-the-air Television Content to Canadian Internet Users*', at http://cyber.law.harvard.edu/people/edelman/pubs/jump-091701.pdf at 14 August 2007, at 6. The "unique difficulties" Mr Edelman speaks of are multiple. First, a number of ISPs offer their services in both US and Canada. Second, the proximity and economic ties between the two countries means that many companies have offices in both countries. Third, the widespread use of intranets with a single access point to the Internet. Fourth, communication between Canada and the US is not particularly likely to pass through well-known "peering points" or contain the telltale transoceanic time delays.

The source problems are the problems associated with building up and/or collecting accurate geo-location data. In relation to IP addresses, there is currently no real equivalent to the address registers listing physical addresses, or the phone registers listing phone numbers. Consequently, those engaged in creating databases of geo-location information must rely on other, less straightforward, methods. Obviously, the accuracy of the material in the geo-location databases depend on, and can never be better than, the accuracy of the collected data. Common methods of collecting relevant material include, for example, gathering data from registration databases,[11] network routing information, DNS systems, host name translations, ISP information and Web content.[12] As discussed in detail by Edelman, all of these sources may provide inaccurate information.[13]

Turning to circumvention problems, it can be noted that, while some circumvention techniques are technologically advanced (e.g. deep linking to streaming video content without accessing the HTTP server),[14] others are easy enough to be used by virtually anyone (e.g. anonymising techniques)[15] or even inherent in the system-structure ("tunnelling methods").[16] With this in mind, it will presumably always be possible to circumvent geo-location technologies.

Arguably the easiest way to circumvent the type of geo-location technologies described above, is through the use of so-called anonymisers. Anonymisers are applications designed to allow web-users to visit websites anonymously. Anonymisers act as an added layer – a buffer – between the web-surfer and the websites she/he visits. When a web-surfer uses an anonymiser, her/his IP number is only transmitted to the provider of the anonymiser. She/he is then assigned a new IP number by the anonymiser in

---

[11] I.e. Réseaux IP Européens Network Coordination Centre (http://www.ripe.net at 5 February 2007, American Registry for Internet Numbers (http://www.arin.net at 5 February 2007, Asia Pacific Network Information Centre (http://www.apnic.net at 5 February 2007 and Latin American and Caribbean IP address Regional Registry (http://lacnic.net at 5 February 2007.

[12] *See e.g. Internet Geography Guide – A NetGeo White Paper* (can be requested from: http://www.netgeo.com/ at 5 February 2007.

[13] Edelman, B. '*Shortcomings and Challenges in the Restriction of Internet Retransmissions of Over-the-air Television Content to Canadian Internet Users'*, at http://cyber.law.harvard.edu/people/edelman/pubs/jump-091701.pdf at 14 August 2007, at 3-7.

[14] Ibid at 10.

[15] Ibid, at 8. For some examples of anonymising services, see e.g.: EPIC Online Guide to Practical Privacy Tools http://www.epic.org/privacy/tools.html at 5 February 2007.

[16] Ibid at 9.

relation to any websites she/he visits while applying the anonymiser. It needs to be stressed that these applications were not developed for the purpose of circumventing geo-location technologies. However, by identifying the location of the anonymiser (or, more specifically, the location with which the IP numbers assigned by the anonymiser are associated), one may be able to find anonymisers from the country one wishes to appear to be located in. For example, when using an anonymiser called *The Cloak,*[17] I was assigned an IP number (216.127.72.7) indicating my location as being the US, while when using an anonymiser called *Anonymouse,*[18] I was assigned an IP number (82.96.100.100) indicating my location as being Germany.

The number of anonymisers available is limited, and thus there are only a limited number of countries one can appear to be located in, using such applications. However, the use of so-called proxy servers opens up further possibilities. A bit simplified, a proxy server is a server that sits between the web-browser and the server being accessed. Thus, just like the anonymisers discussed above, a proxy server acts as a buffer between the web-surfer and the websites visited. The main difference is that while the anonymisers are web-applications, the use of proxy servers is determined by the settings in the web-browser. Using a proxy server to circumvent geo-location technologies, involves two easy steps. First it is necessary to obtain the address (with its port number) of a proxy server from the country you wish to appear to be located in. Then the browser settings must be changed to the obtained proxy address (with its port number). For example, users of Microsoft's Internet Explorer can change their proxy server setting by first clicking on *Internet Options* under *Tools*, and then clicking on *LAN Settings* under *Connections*.

A few words of warning must, however, be said in this context. Some proxy servers, and anonymisers, can log all information that passes through them. In other words, all the web-surfer's traffic can be accessed by the operator of the anonymiser or proxy server. Thus, it is not advisable to send passwords or credit card details through a proxy server, or anonymiser. Furthermore, it is to be noted that, due to the common usage of corporate firewalls, people connecting to the Internet using a computer connected to

[17] http://www.the-cloak.com/login.html at 5 February 2007).
[18] http://anonymouse.org/anonwww_de.html at 5 February 2007).

the network of a larger institution, such as a university or a company, may not be able to use proxy servers in the manner outlined above.

When discussing how the effectiveness of IP based geo-location technologies is affected by the availability of anonymisers and proxy servers, it is to be noted that the producers of IP based geo-location technologies are working to identify the servers providing the anonymising services.[19] Once identified, the value of the anonymising tool for circumventing geo-location technologies is obviously limited.

Furthermore, proxy servers are frequently victims of their own success. Once a proxy server becomes widely used, it may be exposed to heavy traffic loads slowing it down, or even causing it to stop functioning. Once slowed down, it usefulness is lowered and its users may turn to other proxies instead. For this, and other reasons, proxy servers are generally speaking not very reliable. A proxy server available one day may be gone the next.

It is also important to note that there are different kinds of proxies. For our purposes, the most important distinction is between so-called transparent and non-transparent proxy servers. While the latter type caters for a degree of anonymity, the former does not as it "does not modify the request or response beyond what is required for proxy authentication and identification".[20]

One last observation must be made when assessing the usefulness of proxy servers. While it is relatively easy to use proxy servers, the average Internet user is unlikely to have the technical knowledge to do so. However, in 2001, Edelman speculated as to the future of geo-identification and noted that: "geographic analysis tools are likely to suffer in effectiveness due to the increasing availability of automated tools and generally-known methods for bypassing security systems."[21] As a result of a research project funded by a *Bond University Vice Chancellor's Research Grant*, one such automated tool is now available for free on the website on which I present my research findings – www.svantesson.org. It takes the form of a downloadable toolbar which allows the user to select to surf through a range of third-

[19] See e.g. http://www.quova.com/page.php?id=43 at 5 February 2007.

[20] RFC 2616, http://tools.ietf.org/html/rfc2616#page-46 at 14 August 2007.

[21] Benjamin Edelman, *Shortcomings and Challenges in the Restriction of Internet Retransmissions of Over-the-air Television Content to Canadian Internet Users,* at http://cyber.law.harvard.edu/people/edelman/pubs/jump-091701.pdf at 14 August 2007, at 11.

party proxy servers. The list of proxy servers the user can chose from is updated regularly and it includes information about which country the user will appear to be located in.

The second advice I offer to judges faced with a dispute involving the use of geo-identification is that they need to take great care in relation to the issue of 'leakage'. It will presumably always be possible to circumvent geo-identification and it is therefore unreasonable to demand that e.g. website operators use technologies that are leakage-free. The courts must focus on what is reasonable in the circumstances.

In their excellent paper, *Internet Geolocation and Evasion,*[22] Muir and van Oorschot provide a framework for understanding the various attempts at evaluating the accuracy of geo-location technologies. They note that we must separate three different problems. The first problem is to "[d]etermine the geographical location of an Internet user, given a connection attempt or content request initiated by that user".[23] This is a representation of what we what to gain from geo-identification, and when sceptics state that geo-location technologies simply do not work, they focus on this question. In contrast, when the providers of geo-location technologies discuss their accuracy levels, they refer to what Muir and van Oorschot introduce as Problem 2; that is, the problem of determining "the geographical location of the Internet device using a given IP address".[24] This problem is different since it is disconnected from the question of the location of the user as it focuses exclusively on the location of the device which may not necessarily be the computer from which the user is accessing the Internet. To link the two problems, Muir and van Oorschot introduce Problem 3 – determining "the IP address of an Internet end-user's device, given a content request initiated by that user."[25] An example can usefully illustrate the significance of these distinctions.

Imagine that you are located in Tokyo, and access the Internet from a computer located there. Using a proxy located in Copenhagen you connect to a web server located in Istanbul. Should that web server be using a geo-

---

[22] Muir, J. A. and van Oorschot, P. C. *'Internet Geolocation and Evasion'* (10 April 2006), at 13. http://www.ccsl.carleton.ca/~jamuir/papers/TR-06-05.pdf at 7 August 2007.

[23] Ibid.

[24] Ibid.

[25] Ibid.

location technology, we can link the scenario to the three problems outlined above. Problem 2 relates to the geo-location technology's ability to connect the IP address, provided by the proxy, to a geographical location. In our example, that location – Istanbul – clearly has nothing to do with the end-user's actual location. If, however, the geo-location technology can overcome the third problem, it will have ascertained the IP address of your computer located in Tokyo. It can then proceed to the task of dealing with problem one; that is, linking the IP address of the end-user device (i.e. your computer in Tokyo) with a geographical location.

This brings me to the third advice I offer to judges faced with disputes involving the use of geo-identification. It is crucial that the courts take care to properly appreciate the technical nuances involved. This is of course true in any context, but it seems that the rapid development of technology causes particular difficulties.

CONCLUDING REMARKS [4]

Since 2003, I have been talking about how geo-identification will change the Internet. I have repeatedly suggested that geo-identification will become widely adopted, and as more and more content becomes geographically restricted, the Internet will lose one of its most important and valuable characteristics – its 'borderlessness'. In other words, geo-identification may have the consequence of making the Internet more similar to the 'real world' divided by so many borders of different kinds. Perhaps it could be said that we are currently witnessing the concept of sovereign nation states being imposed on the Internet, at the expense of the Internet's global nature. This is a classic example of how the undefined interest of the masses is overrun by the defined interest of a few – it is in most peoples' general interest that the Internet remains borderless, but a small group of advertisers, broadcasters, intellectual property owners and security personnel have a very specific interest in a geographically divided Internet.

However, as of yet, there seems to be no evidence of a widespread adoption of geo-identification, and certainly no widespread adoption of geo-identification that risks having a severe impact on the Internet's borderless nature. Consequently, I accept that my credibility as a 'doomsday prophet' is waning. At the same time, I note that several main Internet actors have started using geo-identification, that investment in geo-identification con-

tinues, that new geo-identification products are being developed, that geo-identification can be easily and cost effectively done, that aspects of the law is structured in a manner that encourages the use of geo-identification, that the law has started to take account of geo-identification and that geographical location always matters in trade and other human interaction. Thus, I am still convinced that we have only seen the tip of the iceberg so far – geo-identification will increase in use, with the inevitable consequence that larger and larger parts of the Internet becomes less borderless. In this context I offer my fourth and last advice to judges faced with disputes involving geo-identification. The courts must bear in mind the policy implications of their judgments. A judgment to the effect that a website operator is protected against a lawsuit in a particular jurisdiction, if she/he has used appropriate geo-identification tools to avoid visitors from that jurisdiction, will clearly make the use of geo-identification more attractive. As highlighted above, the increased use of geo-identification comes at a great cost to the usefulness of the Internet.