IXmaps – Tracking your personal data through the NSA's warrantless wiretapping sites

Andrew Clement
Faculty of Information
University of Toronto
Toronto, Canada
andrew.clement@utoronto.ca

Abstract—**The National Security Agency's 'warrantless wiretapping program' is arguably one of the largest domestic surveillance operations in history. It also represents among the most serious contemporary challenges to democratic governance and civil liberties. This paper reports on the development of the IXmaps interactive mapping application designed to show internet users how their personal traffic may be intercepted by the NSA. Using crowdsourced data collection of thousands of individually generated traceroutes, IXmaps displays on a map of North America the path taken by user-initiated data packets in relation to the sites where the NSA has most likely established surveillance operations. We also discuss the potential for this mapping technique to serve as a tool for achieving better public understanding of surveillance in the internet core.**

Keywords-—NSA warrantless wiretapping; counter-surveillance; internet surveillance; crowdsourced veillance

I. INTRODUCTION

Mark Klein,[1] a retired AT&T technician, touched off a storm of controversy in 2006 when he revealed that the U.S. National Security Agency (NSA) had secretly installed surveillance equipment in AT&T's main San Francisco internet exchange point that was capable of copying and analyzing potentially all the internet traffic passing through that location. It soon became clear that this was the tip of a much bigger iceberg, covering all forms of telecommunications traffic and implicating most of the major telecommunications carriers across the U.S. What became known as the NSA warrantless surveillance controversy is arguably the largest single surveillance program conducted by a government over the communications of its citizens.[2] Congressional passage of special legislation in 2008 to protect telecommunications carriers against the dozens of lawsuits that ensued largely ended media attention to the controversy, but left the constitutional and civil liberties issues unresolved, and considerable mystery remaining about what actually happened. With no evidence to the contrary, it is highly likely that these surveillance activities continue, but with little public knowledge or understanding.

In addition to the obscurity that security agencies typically operate under, the challenge for citizen's comprehension of these widespread surveillance practices is compounded by the opacity of the inner workings of the internet and the popular imaginary of it as a "cloud". The IXmaps project seeks to address this lack of public understanding by rendering more visible the typically hidden workings of the internet core. In particular, this paper reports on how the IXmaps software operates, how its development grapples technically with the opacity of the internet "cloud", and how people can use IXmaps to reveal how the cloud is grounded in issues of policy, civic rights, and jurisdiction. The IXmaps project shares with a number of other internet mapping initiatives, such as Caida, the Cooperative Internet Association for Internet Data Analysis,[3] an interest in revealing and better understanding hidden features of internet structure and operation. Our primary focus on surveillance and privacy is distinctive.

We begin by describing the NSA's alleged internet surveillance activities. To better understand how this interception is done, as well as how we can detect where it is taking place, we then provide a brief overview of the technical aspects of internet routing. The heart of the paper is an explanation of the IXmaps traceroute generation, capture and mapping software that shows the path one's data packets take across the internet. A particular challenge is to determine the geographic location of the various core routers that packets pass through. We then discuss the rationale for siting the suspected NSA 'splitter' installations capable of intercepting one's communications in 18 key cities in the continental US. We close by identifying the current limitations and future directions of this approach to making the internet "cloud," and the activities within it, more publicly visible.

II. THE NSA'S WARRANTLESS WIRETAPPING PROGRAM

The *New York Times* first reported the interception of US domestic communications by the NSA in late 2005.[4] But it wasn't until Mark Klein, a recently retired AT&T technician, revealed the existence of a secret 'splitter' operation at 611 Folsom St in San Francisco that the scope and technical details of NSA surveillance came to public light. Klein reported that AT&T had spliced fiber-optic splitters into 16 'peering links' that connected its network with other major carriers and internet exchange points, directing an exact copy of all the traffic passing through these links into a 'secret room' on the 6th floor, Room 641A. Here a Narus STA 6400 analyzed all the packets passing by, providing "complete visibility for all Internet applications" according to its vendor. In other words, this operation enables the NSA to monitor not only who is

This research has been supported by the Canadian Office of the Privacy Commissioner and the Social Sciences and Humanities Research Council.

communicating with whom, but the entire contents of these communications as well.

Klein's revelations ignited a storm of controversy resulting in over four dozen court cases against U.S. telecom carriers and the federal government. These cases allege that the carriers illegally complied with multiple surveillance requests from the NSA during the Bush Administration to provide without warrants specific information about US citizens.[1] They specify three main types of request, for:

- Detailed phone records (i.e. customer calling records revealing who called whom, when, but not revealing any of the content of the calls)
- Monitoring of telephone calls (i.e. eavesdropping on the actual phone conversations, either live, or more usually, recording them for later analysis and listening - this has long been possible with conventional phone technology.
- Installation of internet "splitter rooms", so that NSA can directly copy and monitor all traffic at that point - email, web access, VoIP conversations, etc.

Though all three forms of request raise serious issues about the legality of the U.S. government's actions and the encroachment on civil liberties,[2] in this paper we are concerned primarily with the last of these – the mass replication of internet traffic at internet exchange points. This is the most novel form of surveillance as well as the one with the greatest scope, depth and potential for incursion into an individual's activities.

The main charges were against AT&T as well as Verizon/MCI, BellSouth, Sprint, and Cingular.[3] However, the FISA Amendments Act (FISAA) of 2008, popularly known as the "Telecom Immunity Act", has rendered these cases moot, as this legislation "allow[s] federal judges to waive lawsuits if the telecom firms can prove that they were authorized by the president and assured that the program was legal."[5] There were several more court cases against the federal government, with the federal government seeking to have each dismissed on "national security" grounds. A few are still pending.[4]

The secrecy that pervades this topic makes it difficult to determine whether the NSA surveillance program is continuing or not, but recent reports strongly suggest that not only is it on-going, but is expanding during the Obama Administration. James Bamford's article in the March 2012 issue of *Wired* details the construction of an enormous data centre in Bluffdale Utah capable of storing and analyzing the complete record of interpersonal internet traffic.[6] In July 2012, three whistleblowers, William E. Binney, Thomas A. Drake, and J. Kirk Wiebe, all former NSA employees, gave evidence in the Electronic Frontier Foundation's (EFF's) lawsuit against the government's illegal mass surveillance program, *Jewel v. NSA*. confirming the surveillance allegations.[7] In particular, Binney, a former NSA technical director, claims the current program, known as Stellar Wind, is capable of intercepting virtually all email in the US and much else.[8]

Given that the NSA's internet surveillance is on-going but its details a closely guarded secret, how can we determine where it is being conducted, and whose traffic is capable of being intercepted? These are the central questions this paper examines. We will focus our investigation on AT&T, and the splitter installation at 611 Folsom Street, as this is the best documented case and provides a model for the interception of internet traffic at other major internet exchange points in the U.S. To understand the role that sites like 611 Folsom Street play, we turn now to a quick review of internet routing.

III. INTERNET ROUTING BASICS

Data sent over the internet is typically transmitted in a series of small packets, each with a header, containing, among other items, source and destination IP addresses, much like the *return* and *to* addresses on a conventional piece of mail. Each packet is passed through data traffic routers, with each router passing the packets closer to the intended destination, again much like the conventional postal service routes mail. The destination, when reached, may return a response, whether it is a web page, video, file transfer, etc.

Each router is usually associated with three key attributes – an IP number, a hostname, and an AS number. An IP number uniquely identifies the address of each router and is used in the protocol controlling traffic between routers. A hostname is a more mnemonic identifier, usually associated with a particular router or server, and useful for humans referring to or addressing the device. URLs, for example, contain hostnames. The ASN, or autonomous system number, corresponds to the

[1] While the Bush Administration initially denied the role of telecommunications carriers, it subsequently confirmed this in general terms. Lichtblau, Eric. 2007. "Role of Telecom Firms in Wiretaps Is Confirmed" New York Times (August 24); http://www.nytimes.com/2007/08/24/washington/24nsa.html?ex=1345608000&en=4e8428cf3d46306c&ei=5090&partner=rssuserland&emc=rss

[2] These three different activities are often conflated in the press coverage. Reporters aren't always careful to distinguish between the various forms of "eavesdropping" and "wiretapping" and use the terms loosely.

[3] Unlike the five telecom carriers facing lawsuits, Qwest reportedly did not comply with the NSA's request to turn in customers' telephone records. Moreover, Qwest CEO claims that this request was made in February 2001, well before the attacks of September 11. In addition to requests for phone records, Qwest was also approached by unnamed "clandestine agencies" about allowing the latter the use of Qwest's "fiber-optic communications network for government purposes." Qwest says that it did not comply.

[4] The two most prominent of these cases are *Jewel v. NSA* and *Clapper v. Amnesty,* both dating back to 2008. In *Jewel v. NSA*, EFF is suing the NSA and other government agencies on behalf of AT&T customers to stop the illegal, unconstitutional and ongoing dragnet surveillance of their communications and communications records. In *Clapper v. Amnesty et al* the Supreme Court has recently denied the ACLU's challenge to the constitutionality of FISAA based on 'lack of standing.'

particular network operator responsible for the router. For the purposes of this paper, it is important to note only that the AS number (ASN) associated with each router or IP address is usually public. We will use these ASNs to determine which carriers are involved in routing at each hop. In particular, we are most immediately interested in detecting AT&T's (ASN=7018) core routers. We'll discuss how to determine the geographic location of each router in the following section, which when combined with the carrier information helps pinpoint the location of NSA splitter operations. But first we explain how IXmaps generates a database of traceroutes results from widely distributed locations.

IV. MAPPING PACKETS ACROSS THE INTERNET USING IXMAPS

A. *Generating and collecting traceroutes*

The IXmaps platform consists of three major parts. The first is a software program called TRgen, which traces the route packets take from the users local machine to a given device on the internet. Based on the widely available 'traceroute' program mentioned above,[9] TRgen automatically sends out a series of test packets to a given destination address. From the Contribute page of the IXmaps website, users can download and install the TRgen appropriate to their operating system (i.e. Windows, Linux or MacOS). TRgen offers the user the option of various sets of pre-defined destination URLs or target destination of their choosing. When run, TRgen generates a sequence of IP addresses and associated hostnames, starting with an anonymized local address, together with the round trip times it takes packets to reach each of the intermediate routers on the way to the destination site. TRgen immediately stores these route sequences in a database on the IXmaps server. TRgen then augments the database by adding the AS number and latitude and longitude of each router in the route sequence. We obtain the AS numbers from the *whois* function of the American Registry for Internet Numbers (ARIN)[10]. The latitude and longitude information we acquire via MaxMind's free GeoLite database.[11] See Table 1 for an example based on a traceroute between a home computer in Toronto Canada to the San Francisco Art Institute, (http://sfai.edu) performed December 13, 2009.

Table 1: TRgen output for traceroute 1859

Hop	IP Address	Round Trip Times				AS#	Latitude	Longitude	Hostname
1	206.248.154.0	0	0	0	15	5645	42.4	-82.1833	206.248.154.0
2	69.196.136.66	0	*	0	0	5645	43.8667	-79.4333	2120.ae0.bdr02.tor.packetflow.ca
3	64.34.236.121	0	0	16	0	13768	42.9833	-81.25	64.34.236.121
4	216.187.114.145	0	0	0	0	3303	40.6888	-74.0203	10ge.xe-2-0-0.tor-151f-cor-1.peer1.net
5	216.187.114.133	0	0	0	0	3303	40.6888	-74.0203	10ge.xe-0-0-0.tor-1yg-cor-1.peer1.net
6	216.187.114.141	15	16	15	16	3303	40.6888	-74.0203	oc48-po5-0.chi-eqx-dis-1.peer1.net
7	206.223.119.79	16	15	16	15	293	37.555	-122.269	ex1-g1-0.eqchil.sbcglobal.net
8	151.164.99.110	16	16	15	16	7132	38.0	-97.0	151.164.99.110
9	151.164.99.129	15	15	16	15	7132	38.0	-97.0	151.164.99.129
10	12.122.79.85	16	16	15	16	7018	38.0	-97.0	gar3.cgcil.ip.att.net
11	12.122.133.218	63	62	63	62	7018	38.0	-97.0	cr1.cgcil.ip.att.net
12	12.122.4.121	62	63	62	63	7018	38.0	-97.0	cr1.sffca.ip.att.net
13	12.123.15.110	63	62	63	62	7018	38.0	-97.0	cr83.sffca.ip.att.net
14	12.122.110.113	62	63	62	63	7018	38.0	-97.0	gar26.sffca.ip.att.net
15	12.91.92.250	62	62	63	62	7018	38.0	-97.0	12.91.92.250
16	63.197.251.33	79	63	62	63	7132	37.8033	-122.411	63.197.251.33

Where:

- **hop** = the number of the hop from the originating node (12.231.120.0)
- **IP Address** = IP address of the router that corresponds to that particular hop
- **Round trip times** are the times in milliseconds that packets take to go from the originating device to the router at that hop, and back again. The four columns correspond to successive rounds of measuring the round trip times.
- **AS# r**efers to the Autonomous System Number (ASN) of the particular network carrier for the packets. 7018 and 7132 are both ASNs of AT&T.[5]
- **Latitude and longitude** represent the geographic coordinates of the device.
- **Hostname** is the network name for the device, such as found in the ARIN WhoIs database.

Since 2009 we have recruited over 100 database contributors, who have installed the TRgen software and

[5] See AT&T Global IP Network Settlement-Free Peering Policy: http://www.corp.att.com/peering/

generated traceroutes to target destinations we have provided as well as those of their own choosing. By May 2013, we had accumulated a database consisting of 24,747 individual traceroutes, from 173 distinct originating addresses and 2,573 distinct destinations. The database contained 19,032 distinct IP addresses, belonging to over 800 different ASNs, including all the major North American carriers. Given the bias in favour of Toronto as an origin, the home city of the research team which has initiated about half of all the traceroutes, this database hardly offers a representative sample of North American routing. However, its overall size and diversity of traceroute endpoints (~90 different originating cities) give good reasons to believe it offers adequately reliable insights into NSA surveillance locations.

B. Geo-locating routers

It would be tempting to rely exclusively on the geographic coordinates provided by Maxmind for mapping the traceroutes produced by TRgen. After all, Maxmind is among the most prominent on-line services to provide location information for IP addresses, and is widely used in the internet industry.[6] Maxmind claims that it is "83% accurate for cities in the US within a 40 kilometer radius."[12] This may be fine for Maxmind's principal target use, which is to locate end users' devices to the accuracy of a postal code or ZIP code to aid in marketing, digital rights management and electronic commerce. However, Maxmind's business model does not appear oriented to locating core routers. While numerically much fewer than edge routers, these core routers are our main focus and we've found for them that Maxmind is far from accurate.

Determining the exact physical location of a particular IP address, or rather its associated router, is a notoriously difficult task.[13][14] This is especially challenging in the case of internet core facilities as it relies on ISPs regularly reporting publically precise, accurate and up to date location information about their routers. There is little incentive for the carriers to do this, while incurring various risks.

In particular, we have found that most of the IP addresses associated with internet core routers are given generic country locations or associated with corporate head offices rather than the actual switching centers. We can see this in the example above. Maxmind reports that almost every IP address associated with AT&T (i.e. hops 8 to 16) has a lat/long of 38.0 -97.0. The rounding of these figures to a full degree, and the corresponding lack of trailing decimal digits, provides an obvious clue that these do not correspond to actual physical locations. In this case, 38, -97 corresponds to a vacant corn field in Kansas near the geographic centre of the continental US.[7] Furthermore, three of the apparently more precise locations mentioned above represent the official business addresses of the carriers, as listed in the ARIN WhoIs database, but are no more accurate in terms of actual location of the routers.

While the Maxmind location data for the routers at the edges of the internet, or the initial and final hops in an individual traceroute, are usually accurate enough for our purposes, we need to perform our own geo-location calculations for core and other intermediate routers. There are two principal methods reported in the literature for locating routers based on their IP addresses – hostname parsing and landmark/latency analysis – and like other recent IP geo-location projects, [15] we use both in combination, together with other custom heuristics.

Hostname parsing is the more straightforward. Fortunately for us, unlike their reticence about making public the location of particular IP addresses, carriers are relatively forthcoming about the hostnames they associate with core router IP addresses. We can see this in the example above as most hops have apparently meaningful hostnames. This is likely because these names are helpful for network management and diagnostics. Each carrier has its own conventions for incorporating an indication of location within the hostname. In some cases, even the street address is coded. For example, the Peer 1 router at hop 4 above, with its hostname including `tor-151f` alerts us that is in Toronto, at 151 Front Street, site of Canada's largest public internet exchange point. AT&T doesn't go this far, only including city and state codes. In particular, in the case above, we can see 5 hops that are handled by AT&T routers in Chicago and San Francisco, e.g. `cr1.cgcil.ip.att.net` and `cr83.sffca.ip.att.net` respectively. Given that the facility at 611 Folsom Street is AT&T's principal internet routing centre in San Francisco, we can attribute the hops between one of the pairs of routers with **`sffca`** in the host name as the site of an NSA splitter. We confirmed this by examining additional traceroutes destined for the San Francisco area and found a consistent pattern of them passing through AT&T's **`sffca`**`.ip.att.net` routers.

Analysis of latencies, based on the Round Trip Times generated by TRgen, also provides important clues as the physical location, at least to the city region level of precision. Since light travels approximately 200,000 km/sec in fiber-optic cable, minimum latencies circumscribe the area within a determinate radius of originating routers with a known location. While there are considerable variations, due to router delays and congestion, this is helpful in determining between plausible alternatives the city that a router is located in.

Other corrections are made on an individual case by case basis, as anomalies come to our attention. Having several traceroutes available in the database for testing against enables a triangulation approach to establishing greater confidence in the accurate assigning of locations. Once a reliable pattern is

[6] Some of the other similar geo-location services are:
Quova: http://www.quova.com/
IP2Location: http://www.ip2location.com/
GeoBytes: http://www.geobytes.com/ipLocator.htm
InfoSniper: http://www.infosniper.net/index.php?lang=1
IPGlobalPosition: http://www.ipglobalposition.com

[7] The official Geographical Center of the United States is Latitude 39 degrees 50 minutes Longitude 98 degrees 35 minutes. See: http://www.kansastravel.org/geographicalcenter.htm

ascertained for particular IP addresses or hostnames, the rules are incorporated into a script that routinely and automatically reviews IP locations and revises them accordingly. Altogether, using these various techniques, we have 'corrected' at least to the city level, about half of the more than 18,000 distinct IP addresses in the database. Because we have prioritized for attention the most commonly used core routers, this translates into over 57% of the more than 250,000 hops having both endpoints 'corrected' in this manner. These are concentrated in the internet core, with most of the remaining 'uncorrected' hops close to the edges, where the Maxmind data is much more accurate. While this still leaves many individual hops with questionable locations, their number is relatively small and so they do not throw off significantly the coarse statistics we cite below.

Table 2 shows the details for TR1859 after the geo-location corrections and related revisions have been incorporated. The first column of symbols following the IP address indicates the country in which the router is located, and the second column indicates whether an NSA splitter operation is suspected in this location.

C. Mapping traceroute data

Once a traceroute has been added to the database and the locations of its constitutive routers determined, while keeping in mind the limitations of the geo-location process, we can turn to the third and final major component of the IXmaps program, the cartographic visualization of individual traceroutes. With the latitude and longitude information for each hop, we depict routes geographically using Google Maps and Google Earth. The former is the default on the Explore page as it doesn't require prior software downloading, and more easily displays multiple routes, while the latter is more suitable for detailed exploration of selected routes, using its 3-D rendering and fly-through features.

See Fig. 1 for a continent-wide view of traceroute #1859, and Fig. 2 for a close-up of this route in the vicinity of 611 Folsom Street, both using Google Earth.

Table 2: "Corrected" TRgen output for traceroute 1859

Traceroute detail

Traceroute id: **1859**
origin: **M5S2M8** destination: **sfai.edu** [63.197.251.33]
submitted by: AndrewC submitted on: 2009-12-13 12:06:51-05

Hop	IP Address	Min. Latency	Carrier	Location	GeoPrecision	Hostname
0	206.248.154.0	0	TekSavvy	Toronto ON	city level	206.248.154.0
1	69.196.136.66	0	TekSavvy	Toronto ON	city level	2120.ae0.bdr02.tor.packetflow.ca
2	64.34.236.121	0	Peer 1	Toronto ON	city level	64.34.236.121
3	216.187.114.145	0	Peer 1	Toronto ON	building level	10ge.xe-2-0-0.tor-151f-cor-1.peer1.net
4	216.187.114.133	0	Peer 1	Toronto ON	building level	10ge.xe-0-0-0.tor-1yg-cor-1.peer1.net
5	216.187.114.141	15	Peer 1	Chicago IL	building level	oc48-po5-0.chi-eqx-dis-1.peer1.net
6	206.223.119.79	15	ESNET - ESnet	Chicago IL	building level	ex1-g1-0.eqchil.sbcglobal.net
7	151.164.99.110	15	AT&T Internet Services	Chicago IL	city level	151.164.99.110
8	151.164.99.129	15	AT&T Internet Services	Chicago IL	city level	151.164.99.129
9	12.122.79.85	15	AT&T WorldNet Services	Chicago IL	city level	gar3.cgcil.ip.att.net
10	12.122.133.218	62	AT&T WorldNet Services	Chicago IL	city level	cr1.cgcil.ip.att.net
11	12.122.4.121	62	AT&T WorldNet Services	San Francisco CA	building level	cr1.sffca.ip.att.net
12	12.123.15.110	62	AT&T WorldNet Services	San Francisco CA	building level	cr83.sffca.ip.att.net
13	12.122.110.113	62	AT&T WorldNet Services	San Francisco CA	building level	gar26.sffca.ip.att.net
14	12.91.92.250	62	AT&T WorldNet Services	San Francisco CA	building level	12.91.92.250
15	63.197.251.33	62	AT&T Internet Services	San Francisco CA	Maxmind	63.197.251.33

Legend
- NSA: Known NSA listening facility in the city
- NSA: Suspected NSA listening facility in the city

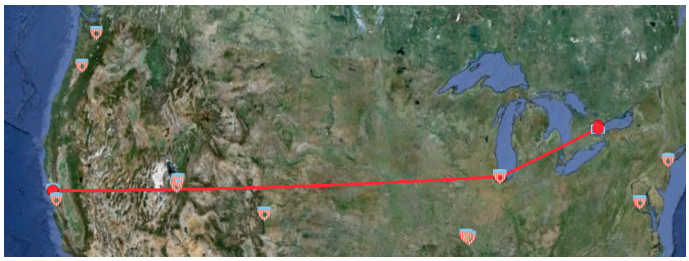


Figure 1: IXmaps rendering of traceroute #1859

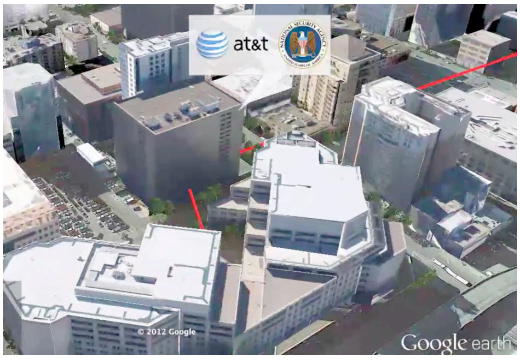


Figure 2: IXmaps rendering of traceroute #1859 at 611 Folsom Street

V. MAPPING NSA SURVEILLANCE SITES AND INTERNET TRAFFIC THROUGH THEM

A. Where are the NSA splitter sites?

While we know of the NSA splitter site at 611 Folsom Street, and can map packet routes that go through it, what about the other suspected sites? Based on his conversations and meeting with other AT&T technical staff, Klein reports that similar installations were installed in five other locations – Seattle, San Jose, Los Angeles, San Diego and Atlanta.[1] However, these 6 sites would not be sufficient to comprehensively intercept US internet traffic, as there are other, more important routing centres that would be much more attractive for interception purposes. Scott Marcus, a former Federal Communications Commission expert, estimates that AT&T has 15-20 splitter sites.[16] However, he isn't able to identify any sites in particular without further specific evidence. Presuming that the NSA's goal was to be able to intercept the largest proportion of US internet traffic with the fewest possible sites, we developed a crude schema for scoring cities based on how much internet traffic was likely to pass through them. Using only our personal own estimates of 3 determinants of internet prominence, with crude relative weightings: telecom infrastructure (10); city size (pop) (5); and geographic location (4), we developed an ordered ranking of the top 18 US cities in terms of their likelihood of hosting an NSA splitter installation. See Table 3.[8]

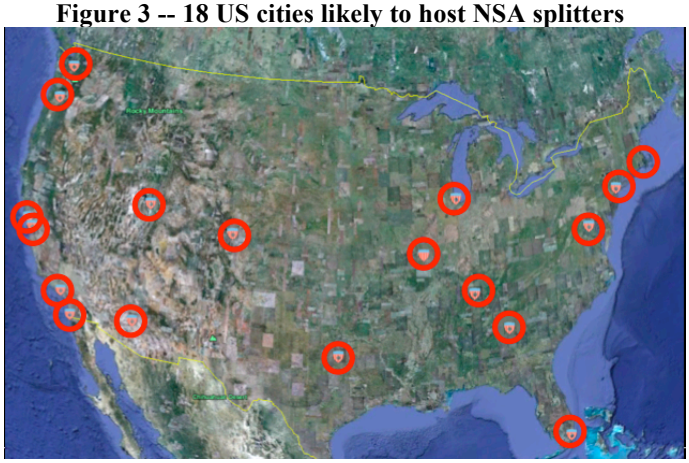
[8] For a more detailed view of this table and the calculations behind it, see the Google Docs spreadsheet at: http://bit.ly/SxOiJn

Table 3 -- Ranking of top 18 US cities by likelihood of NSA splitter installation

Rank	City	Score
1	New York, NY	95
2	Chicago, IL	95
3	Los Angeles, CA	95
4	Atlanta, GA	87
5	Dallas, TX	82
6	San Francisco, CA (611 Folsom St)	77
7	Washington (area), DC	77
8	Seattle, WA	75
9	San Diego, CA	62
10	Boston/Cambridge (area), MA	61
11	Miami, FL	60
12	San Jose, CA	57
13	Phoenix, AZ	54
14	St Louis, MO	48
15	Portland, OR	47
16	Denver, CO	39
17	Nashville, TN	39
18	Salt Lake City, UT	29

To test our hypothesis, we examined all the US-only routes in the IXmaps database, which numbered 1,319 at the time. Of these, only 7 did not pass through any of the 18 cities. In other words, installing splitters in the internet exchange points in just these cities would be sufficient for the NSA to intercept over 99% of our US only traceroutes! These are shown in Fig. 3.

Figure 3 -- 18 US cities likely to host NSA splitters



While this result of course does not prove that these cities actually have NSA splitter operations, nor that the NSA has access to all the IXPs in them, it is powerful confirmation that it is technically feasible for the NSA to install splitters in

relatively few strategic internet choke points from where it could intercept a very large proportion of internet traffic. This high percentage helps justify our claim that these cities are strongly suspected of hosting NSA warrantless surveillance facilities.

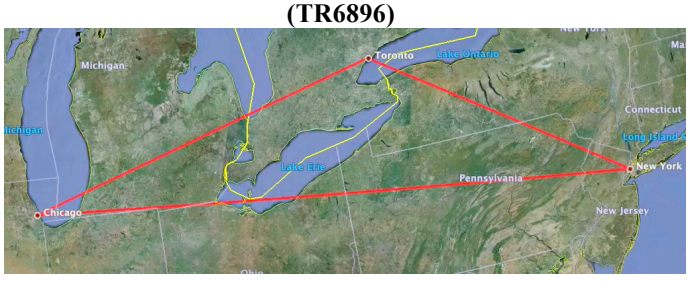
B. Does my internet traffic pass through NSA splitter sites?

With the suspected NSA cities identified, we are in the position to give individual internet users a reasoned indication of whether their particular communications are likely to be subject to warrantless interception. Exploiting the feature of TRgen to target any user-provided URL, individuals can produce traceroutes customized to their own internet activities. IXmaps renders both the tabular and map views of these traceroutes with distinctive icons to highlight those hops most susceptible to NSA splitting. See Table 1 (above) shows TR 1859 with these hops in the AT&T facilities in both San Francisco and Chicago flagged.

C. Non-US traffic may also be exposed to NSA splitters

So far we have concentrated on traffic that explicitly travels via US routing centres, i.e. originating or terminating in the US, or both. It is well known, at least in internet engineering circles, that traffic that neither originates nor terminates in the US may nevertheless transit via the US, mainly due to the interconnection arrangements of the major international carriers.[17] However, the extent of this practice and its surveillance implications are less well known. Canadian traffic, largely due to its proximity to the US as well as the structure of the North American internet service industry, is especially prone to routing via the US. We refer to traffic that originates and terminates in the same country, but transits another, as "boomerang traffic." Analysis of IXmaps data reveals that approximately one third of the 2,500 Canadian routes follow a boomerang pattern. That long distance Canadian routes may be routed via the US is not surprising, but we were struck by the number of routes that start and end in the same Canadian city, but are routed via the US. We have found over 100 such boomerang routes based in Toronto alone. Figure 4 shows one example that transits New York and Chicago, both cities strongly suspected of hosting NSA splitters. Whether crossing the continent, or returning to the same city, Canadian boomerang traffic is almost entirely exposed to NSA surveillance.

Figure 4 -- A Canadian boomerang route based in Toronto (TR6896)



VI. DISCUSSION

When a government conducts surveillance on its citizens and acts outside of established legal bounds, as the US government has demonstrably done in the case of the NSA's warrantless wiretapping program, the norms of liberal democratic governance are seriously violated and demand public accountability. However, the usual mechanisms for such accountability have not so far succeeded. Congress has passed legislation that retroactively grants immunity from prosecution to the implicated telecom carriers and the executive branch has largely stymied court challenges by invoking a blanket "state secrets" exemption. While there have been several notable journalistic exposés[2] and brave whistleblowers from both the NSA and AT&T have brought damning information to light, so far there has been no sustained public outrage or even debate of the serious issues raised.

Compounding the usual difficulties in holding powerful players responsible for their actions is the intrinsically invisible character of internet surveillance. Beyond the notorious secrecy of the NSA, the surveillance is conducted out of sight and leaves no discernible trace. For the great majority of the population, the workings of the internet, especially in its core, are dauntingly complex and inscrutable. These factors may help explain why far lesser scandals, but ones that people can relate to more easily, generate much greater public attention and political response.[9]

We have developed the IXmaps internet mapping application mainly to overcome these obstacles by promoting greater transparency and visibility of the NSA surveillance activities. Within the limits of the available information, we have been able to reveal the likely sites of NSA surveillance operations and show interested individuals where the NSA may intercept their own data packets. By using interactive maps and graphic images, we hope to make the surveillance more vivid, discussable and a matter of public concern.

The mapping application we have developed may also be appealing and useful to people who are not already concerned about internet surveillance, but simply and more generally curious about how the internet works and where their own activities fit within the wider phenomena. In the course of this we expect they will encounter the surveillance and privacy aspects, and we hope learn more about these issues.

But more than just serving concerned citizens and curious explorers, IXmaps encourages and relies on its users to contribute to building its database of traceroutes. This crowdsourcing is necessary to ensure a good geographic distribution of originating points, so that the North American internet core is well surveyed, but also provides the means for people to view the internet from their own personal perspectives. Perhaps more importantly, integrating these contributions in an open and publicly visible manner constitutes a form of collective counter-surveillance with the potential to empower participants in holding the US government and its national security agency to account for its warrantless surveillance.

---

[9] A good example of this at the time of writing is the horsemeat scandal gripping the UK. See: http://www.guardian.co.uk/uk/horsemeat-scandal

VII. ACKNOWLEDGMENT

The IXmaps project is the work of a research team that currently includes Nancy Paterson, Colin McCann, Antonio Gamba, Jonathan Obar and Lauren Dimonte. David Phillips, Steve Harvey, Gabby Resch and Erik Stewart made invaluable contributions to earlier versions of the software. We are grateful to those individuals, largely anonymous, who have contributed to the database by installing and running TRgen, or have provided feedback that has helped improve the application. We also appreciate the careful and helpful comments of an anonymous reviewer.

REFERENCES

[1] M. Klein, Wiring up the Big Brother Machine… and fighting it, Charleston, SC: BookSurge, 2009

[2] J.Bamford, The Shadow Factory: The UltraSecret NSA from 9/11 to the Eavesdropping on America, New York: Doubleday, 2008.

[3] J. Risen and E. Lichtblau, Bush Lets U.S. Spy on Callers Without Courts, New York Times, December 16, 2005. http://www.nytimes.com/2005/12/16/politics/16program.html?ex=1145419200&en=87817a067833b164&ei=5070

[4] Klein 2009 [1], pp. 31-40.

[5] M. Soraghan, "House passes FISA overhaul" The Hill, June 20, 2008; and http://www.sourcewatch.org/index.php?title=FISA_Amendments_Act_of_2008

[6] J. Bamford, "The NSA Is Building the Country's Biggest Spy Center (Watch What You Say)". Wired. March 15, 2012. http://www.wired.com/threatlevel/2012/03/ff_nsadatacenter/all/1

[7] Electronic Frontier Foundation, Three NSA Whistleblowers Back EFF's Lawsuit Over Government's Massive Spying Program, July 2, 0212. https://www.eff.org/press/releases/three-nsa-whistleblowers-back-effs-lawsuit-over-governments-massive-spying-program

[8] P. Harris, US data whistleblower: 'It's a violation of everybody's constitutional rights', Guardian, Sept. 15, 2013, http://www.guardian.co.uk/technology/2012/sep/15/data-whistleblower-constitutional-rights

[9] http://en.wikipedia.org/wiki/Traceroute

[10] https://ws.arin.net/whois

[11] http://dev.maxmind.com/geoip/geolite

[12] http://www.maxmind.com/app/faq#accurate

[13] Dodge. M. and Kitchen, R. "New Cartographies to Chart Cyberspace" Geoinformatics (April/May):1. 2002

[14] E. Katz-Bassett, J. John, A. Krishnamurthy, D. Wetherall, T. Anderson, and Y. Chawathe, Towards IP Geolocation Using Delay and Topology Measurements," in ACM IMC '06, October 2006.

[15] Y. Wang, D. Burgener, M. Flores, A. Kuzmanovic, and C. Huang, Towards Street-Level Client-Independent IP Geolocation, in Proceedings of USENIX NSDI 2011, Boston, MA, March 2011.

[16] PBS Frontline. Spying on the Home Front, May 15, 2007.

[17] William B. Norton, The Internet Peering Playbook: Connecting to the Core of the Internet, DrPeering Press, 2012, p. 71.