FP7 – SEC- 2011-284725

SURVEILLE

Surveillance: Ethical issues, legal limitations, and efficiency

Collaborative Project

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 284725

SURVEILLE Paper on Mass Surveillance by the National Security Agency (NSA) of the United States of America

Extract from SURVEILLE Deliverable D2.8: Update of D2.7 on the basis of input of other partners. Assessment of surveillance technologies and techniques applied in a terrorism prevention scenario.

Due date of deliverable: 31.07.2014

Actual submission date: 29.05.2014

Start date of project: 1.2.2012 Duration: 39 months

SURVEILLE Work Package number and lead: WP02 Prof. Tom Sorell

Author: Michelle Cayford (TU Delft)

SURVEILLE: Project co-funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Executive summary

- SURVEILLE deliverable D2.8 continues the approach pioneered in SURVEILLE deliverable D2.6 for combining technical, legal and ethical assessments for the use of surveillance technology in realistic serious crime scenarios. The new scenario considered is terrorism prevention by means of Internet monitoring, emulating what is known about signals intelligence agencies' methods of electronic mass surveillance. The technologies featured and assessed are: the use of a cable splitter off a fiber optic backbone; the use of 'Phantom Viewer' software; the use of social networking analysis and the use of 'Finspy' equipment installed on targeted computers. Non-technological surveillance techniques featured and assessed are the opening of baggage in an airport and the use of a covert surveillance team. The assessments are represented visually in a multidimensional matrix – a grid with numerical scores for fundamental rights risk and technical usability assessments, and colour coding for ethical risk assessment. Deliverable D2.8 was submitted to the European Commission on 29 May 2014. A separate SURVEILLE Paper, released parallel to this document, contains the assessments and the resulting matrix as produced in D2.8.

- This SURVEILLE paper contains another extract from deliverable D2.8, namely its chapter 7 that describes and assesses the surveillance methods applied by the National Security Agency (NSA) of the United States of America. Technical assessment of the techniques used by signals intelligence agencies for mass surveillance remains challenging because the programmes in question are classified. Nevertheless, educated guesswork is possible based on recent leaks in 2013-2014 including those attributed to NSA and CIA contractor Edward Snowden, previous revelations, and an understanding of what methods and devices are available. These suggest that the technological basis of mass surveillance is achieved by means of a combination of tapping fiber-optic cables, circumventing encryption, launching cyber attacks, gathering phone metadata, and utilizing traditional spying methods such as bugging embassies and tapping political leaders' phones.

Contents

1 Introduction

Many types of surveillance (phone tapping, collecting metadata, intercepting mail) used by the NSA are not new to the world of espionage. What is new in surveillance, thanks to huge advances in technology, is mass surveillance. In the digital age in which everyone's communication and information is stored and transmitted online it is now possible to gather massive amounts of information rather easily. What formerly would have required a laborious process of monitoring physical mail, tracking targets' movements, and entering homes or offices to gain information about a target's activities, contacts, financial information, etc., can now be done largely by monitoring their online activity. And in a post-9/11 era where terrorism is a primary concern of governments, all this information can be swept up not just about one person, but about thousands, potentially even everyone, in the event that this information may prove useful later, and in the interest of identifying and apprehending terrorists before they are able to act.

The objective of this section is to gather information about technical systems that the NSA uses to perform mass surveillance to be able to later analyze their functionality to determine how the balance between safety-gain and privacy-loss was set within NSA. The analysis was performed from the perspective of risk control. That is to say, the assessment focuses on an integral analysis of the system rather than on computer science aspects. The balance between costs and benefits about safety and security measures is natural in the risk sciences and thereby contributes to the global discussion. Before we can examine this balance, it has to be clear what technology is used. The technologies treated here have been chosen based on mass surveillance. That is, technologies that both demonstrate the NSA's mass surveillance as well as those that indicate its limitations in this arena.

What kind of surveillance technology the NSA uses and how it works are classified. Therefore much of our examination and explanation is founded on estimations based on recent leaks, previous revelations, and an understanding of what methods and devices are available. For that reason, paragraph 4 describes the structure of the Internet since it is the technological backbone on which mass surveillance hinges. Paragraph 5 discusses why so much of the world's Internet traffic passes through the U.S. And paragraph 6 analyzes NSA surveillance technologies, including fiber-optic cable tapping, the PRISM program, decryption, and analysis tools and databases.

2 Piecemeal information on NSA mass surveillance

Note that the analysis of mass surveillance technologies is frustrated by the fact that NSA information is classified and the information leaked by Edward Snowden has been distributed in a piecemeal fashion into the global public domain. How, when, and which documents are leaked are decisions in themselves mysterious, but no doubt strategically determined by the reporters with whom Snowden collaborates, notably Glen Greenwald, Laura Poitras, and Barton Gellman. Greenwald has partnered with different publications in various countries, resulting in reports and documents appearing

in one country while not in others.[1] While some websites try to keep track of all this information, none is fully complete. This manner of reporting makes is more difficult to follow the leaks and to combine potentially useful information into a meaningful ensemble.

The leaked information is itself also piecemeal. A document here, a screenshot there, on different subjects do not create a full picture of the NSA's activities under a given program. This makes the information open to misinterpretation. The headlines regarding the NSA tapping mobile phones in Europe was a case in point. It was reported that the agency had collected 70 million phone calls in France in a 30-day period, [2] as well as tracking 60 million phone calls in Spain and 46 million in Italy.[3] Subsequent publications in Norway and the Netherlands made similar claims regarding their respective countries. All these claims were based on screenshots from the BOUNDLESS INFORMANT database. However, it was later revealed that the gathered data was being collected by these respective countries' own intelligence agencies and then shared with the NSA. This data includes records collected in NATO areas of war, such as Afghanistan.[4] BOUNDLESS INFORMANT can sort data by aggregate (the overall figure for a country), DNR, or DNI. Further, it appears that it can also sort by SIGAD, that is, a particular collection facility. Therefore a screenshot showing calls collected under the heading of a given country's name can easily by misinterpreted, as Greenwald did in this case.[5] Thus, what was initially proclaimed to be a gross invasion of Europeans' privacy through mass surveillance by the NSA was found to be neither the NSA nor necessarily targeted against Europeans.

Further, as has been aptly pointed out by Daniel Soar in an article published in the London Review of Books, the slides that have been heavily relied upon in the coverage of this story are used for training analysts.[6] They could be considered as sales pitches for analysts to use a particular program. Analysts have a wide variety of tools to choose from to do their job and the PRISM collection manager, in his presentation, touts his program as the best for getting the job done. This adds to the misleading nature of the

[1] For example: Jacques Follorou and Glenn Greenwald, "Comment la NSA espionne la France," Le Monde.fr, October 21, 2013, http://www.lemonde.fr/technologies/article/2013/10/21/comment-la-nsa-espionne-la-france_3499758_651865.html; and
"Document Snowden: Nederland al sinds 1946 doelwit van NSA," nrc.nl, accessed January 8, 2014, http://www.nrc.nl/nieuws/2013/11/23/nederland-sinds-1946-doelwit-van-nsa/.

[2] Angelique Chrisafis and Sam Jones, "Snowden Leaks: France Summons US Envoy over NSA Surveillance Claims," *The Guardian*, October 21, 2013, http://www.theguardian.com/world/2013/oct/21/snowden-leaks-france-us-envoy-nsa-surveillance.

[3] Harriet Alexander, "NSA 'Tracked 60 Million Phone Calls in Spain in a Month,'" *Telegraph.co.uk*, October 28, 2013, sec. worldnews, http://www.telegraph.co.uk/news/worldnews/europe/spain/10408572/NSA-tracked-60-million-phone-calls-in-Spain-in-a-month.html.

[4] Michael Vincent, "Top US Spy Says NSA Didn't Tap Millions of European Phones," *ABC News*, October 30, 2013, http://www.abc.net.au/news/2013-10-30/top-us-spy-says-nsa-didnt-tap-millions-of-european-phones/5056650.

[5] "Top Level Telecommunications: Screenshots from BOUNDLESSINFORMANT Can Be Misleading," November 23, 2013, http://electrospaces.blogspot.be/2013/11/screenshots-from-boundlessinformant-can.html.

[6] Daniel Soar, "How to Get Ahead at the NSA," *London Review of Books*, October 24, 2013.

information 'revealed' by these slides. In the case of PRISM, when news of the program hit the press it created a furor – this slide supposedly revealed that the NSA had direct access to the servers of major Internet companies. However, this turned out to not be the case. (Explained further in paragraph 6.2.)

Because the leaks are ongoing and for all appearances will continue for many months to come, this report will very soon be outdated. Specifics on NSA surveillance technology may come to light and information gathered and/or speculated on NSA surveillance technologies may be proven false. (On the other hand, further revelations may be more good media stories rather than solid operational information and/ or may fall under the technology categories already discussed here.)

3 A brief word on spying

This section of the deliverable focuses on the aspects of mass surveillance performed by the NSA. The leaks related to spying on other countries, including allies, such as bugging embassies and eavesdropping on heads of states' phones are not considered to be mass surveillance. Nation states, including friendly ones, spy on one another. This is not new, nor astonishing. Heads of state that have denounced this in the press have done so for public consumption. While this may be news to the general public, it is not news to governments or intelligence agencies.[7] For example, it is well known in intelligence circles (including courses at universities – i.e. public, non-classified information), that the DGSE (Direction Générale de la Sécurité Extérieure) in France is active in economic espionage.[8] Spying on political leaders to gain political advantage is as old as the sun, and arguably should be expected.[9] Therefore, we will not treat these more traditional forms of spying but will focus on mass surveillance.

Regarding the NSA's mass surveillance technology two points should be made. One is that not all of the agency's surveillance involves actual surveillance technology. For example, the NSA obtains call record metadata[10] from phone companies via court order (and not via surveillance technology). According to the court order served on Verizon, the phone company was required to provide the NSA with all domestic (with one or

[7] Former French Foreign Minister Bernard Kouchner said, "Let's be honest, we eavesdrop too. Everyone is listening to everyone else. But we don't have the same means as the United States, which makes us jealous." ("NSA spying threatens to undermine US foreign policy; Obama, Kerry try to quell furor abroad," by *Associated Press*, 26 Oct. 2013, http://www.foxnews.com/us/2013/10/26/nsa-spying-threatens-to-undermine-us-foreign-policy-obama-kerry-try-to-quell/)

[8] Mark Lowenthal in *Intelligence: From Secrets to Policy* (2008) states that the DGSE's economic spying includes American businesses as targets, focusing on companies in competition with large French businesses. J.M. Olsen in *Fair Play: the Moral Dilemmas of Spying* (2006) speaks of French moles discovered in IBM, Corning, and Texas Instruments in the 1980s. These references are sourced from the course "Les Politiques du Renseignement" taught at SciencesPo, Paris, France by Philippe Hayez, spring 2010.

[9] Madeleine Albright, former U.S. Secretary of State, recalled a conversation with the French ambassador in the 1990s, which indicated that the French had been intercepting her phone calls. ("Ex-US Diplomat: 'France spied on me'," *BBC*, 24 Oct. 2013)

[10] Metadata is the information about the phone call – the source and destination phone numbers, the time and duration of the call. It may also include location information. It does not include subscriber information or the content of the call.

both ends of the call being in the U.S.) phone metadata on a daily basis for a period of three months.[11] This court order was apparently renewed every time it expired, translating into Verizon continuously providing the NSA with phone metadata. U.S. Senators Dianne Feinstein and Saxby Chambliss stated that this three-month renewal had been going on for seven years prior to 2013.[12]

The second point is that various NSA surveillance programs fall under different NSA project code names, and target different aspects and areas of the Internet structure, but in fact, likely employ the same type of surveillance technology. Upstream, the collection of communications from fiber-optic cables, MUSCULAR, the operation that processes information gathered from internal cables linking Yahoo and Google data centers, and TEMPORA, a fiber-optic cable tapping operation attributed to UK signals intelligence,[13] may all use splitters to tap into the data flows.

4 Structure of the Internet

The Internet is a network of networks. There is not one over-arching network that is *the Internet*. Rather, the Internet is a collection of networks operated independently by Internet Service Providers (ISPs). Their independent networks are known as autonomous networks (ASN) in the industry. The large ISPs have what are called backbones, that is, large arteries that carry their Internet traffic.

Together the backbones of these large ISPs form the core of the Internet. The largest ISPs in the world are called Tier 1 companies.[14] Smaller ISPs connect into the backbones of the large companies and purchase transit traffic from them. In other words, smaller companies pay larger companies to transit their traffic for them. Tier 1 companies are also present at Internet Exchange Points.[15]

[11] United States Foreign Intelligence Surveillance Court, court order, docket number: BR 13-80, Apr. 25, 2013. Accessed from *The Guardian* website on Oct. 18, 2013.

[12] Ed O'Keefe, "Transcript: Dianne Feinstein, Saxby Chambliss Explain, Defend NSA Phone Records Program," *Washington Post*, accessed February 26, 2014, http://www.washingtonpost.com/blogs/post-politics/wp/2013/06/06/transcript-dianne-feinstein-saxby-chambliss-explain-defend-nsa-phone-records-program/.

[13] MacAskill, Ewen, Julian Borger, Nick Hopkins, Nick Davies, and James Ball. "GCHQ Taps Fibre-Optic Cables for Secret Access to World's Communications." *The Guardian*, June 21, 2013. http://www.theguardian.com/uk/2013/jun/21/gchq-cables-secret-world-communications-nsa.

[14] As of January 2011, according to Wikipedia, the Tier 1 networks were: AT&T, CenturyLink, Deutsche Telekom, XO Communications, GTT, Verizon Business, Sprint, TeliaSonera International, NTT Communications, Level 3 Communications, Tata Communications, and Zayo Group.

[15] Author's interview with Hadi Asghari, PhD candidate, Economics of Cybersecurity, TU Delft, November 18, 2013.

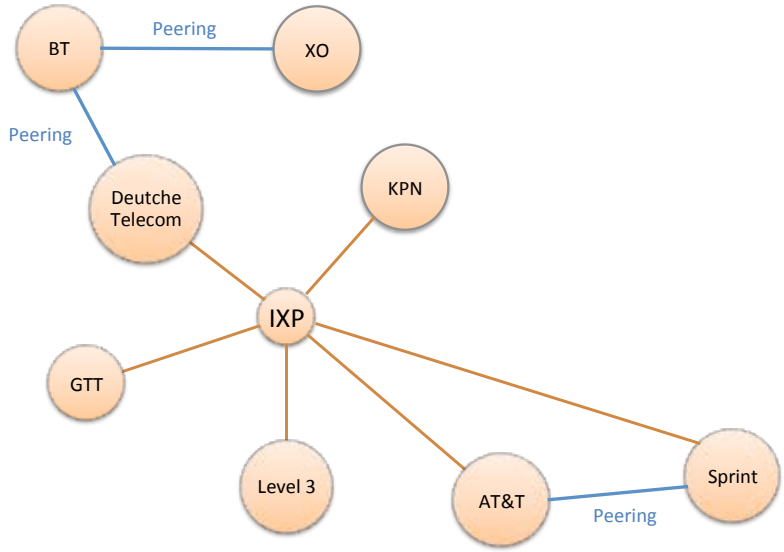

Figure 1: Internet as network of networks with IXPs and peering relationships. Each telecommunications company represents their own network, or ASN. (The companies shown here are purely for example. The diagram is not based on knowledge of specific peering agreements or IXPs where certain companies are known to exchange traffic with one another.)

There are two basic ways in which the ISP backbones connect with one another: through Internet Exchange Points (IXPs) and peering. "*Peering* is an agreement between ISPs to carry traffic for each other and for their respective customers. Peering is usually a bilateral business and technical arrangement, where two providers agree to accept traffic from one another, and from one another's customers."[16] Two companies in a peering agreement directly exchange traffic with one another. These are typically large Internet backbone providers who have enough traffic between each other to warrant the expenses of a separate circuit for private peering.[17] While peering can be expensive, it also means there is no transit cost involved. Google, for example, conducts private peering all over the world to avoid having to pay transit costs.[18]

When there is not enough traffic to warrant a peering arrangement, Internet Exchange Points are used. IXPs are physical locations where multiple ISPs exchange traffic with each other. These allow for all ISPs regardless of size to exchange traffic. Although ISPs are physically connected through IXPs, they do not have to directly exchange data. A data packet can cross multiple ASNs before reaching its destination. At IXPs providers exchange data packets and further them along towards their destinations.

[16] J. Scott Marcus, *Declaration of J. Scott Marcus in Support of Plaintiffs' Motion for Preliminary Action*, United States District Court for Northern District of California, Case of Hepting, Hicks, Jewel et al. vs. AT&T et al., 2006, 21–22, https://www.eff.org/node/55052.
[17] Ibid., 24.
[18] Asghari, interview.

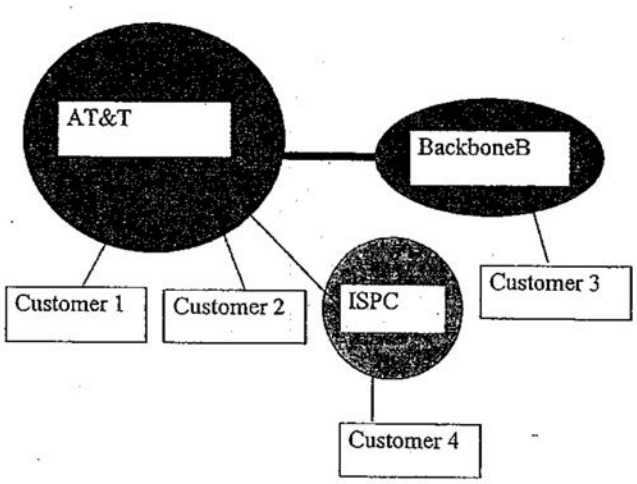

Figure 2: Peering relationship between AT&T and company Backbone B. ISPC, a small ISP, is a transit customer of AT&T and whose traffic is therefore also peered with Backbone B. (Source: Marcus, *Declaration of J. Scott Marcus*, p.23)

These data packets are how data is sent over the Internet. When a person uses email or performs a web search, their transaction is sent over the Internet in data packets. Each packet has a header containing the source and destination IP addresses, similar to the "from" and "to" address on a conventional letter. The packets are sent along their path via routers, with each router sending the packets closer to their final destination. Again, this is like conventional mail is routed by the postal system, according to country, city, postcode, street, and house number. The destination may return a response, such as a web page, file transfer, video, etc.[19]

5 U.S. as Internet Hub

Arguably, the fact that the vast majority of the world's Internet traffic passes through the U.S. (80% of the world's fiber-optic data,[20] which carries 99% of intercontinental data[21]), giving the NSA access to it, is one of the issues that has made the Snowden files world news and a global concern. Several factors contribute to this phenomenon. One is the history of the development of the Internet; another is the number of IXPs in a region; others include agreements between ISPs, Internet and company policies, and U.S. law.

[19] Andrew Clement, "IXmaps - Tracking Your Personal Data through the NSA's Warrantless Wiretapping Sites," vol. 6613122 (presented at the IEEE International Symposium on Technology and Society, Toronto, ON, Canada: ISTAS, 2013), 216–223;
Jonathan A. Obar and Andrew Clement, *Internet Surveillance and Boomerang Routing: A Call for Canadian Network Sovereignty*, SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, July 1, 2013), http://papers.ssrn.com/abstract=2311792.

[20] Olga Khazan, "The Creepy, Long-Standing Practice of Undersea Cable Tapping," *The Atlantic*, July 16, 2013, http://www.theatlantic.com/international/archive/2013/07/the-creepy-long-standing-practice-of-undersea-cable-tapping/277855/.

[21] Craig Timberg and Ellen Nakashima, "Agreements with Private Companies Protect U.S. Access to Cables' Data for Surveillance," *The Washington Post*, July 7, 2013, sec. Tech, http://www.washingtonpost.com/business/technology/agreements-with-private-companies-protect-us-access-to-cables-data-for-surveillance/2013/07/06/aa5d017a-df77-11e2-b2d4-ea6d8f477a01_story.html.

5.1 History

The Internet began in the U.S. It subsequently has a very developed network. Looking at a map of submarine fiber-optic cables highlights this point. The U.S. has the highest concentration of these expensive cables, with 57 connected to its shores. (The UK is number two with 49 cables; the number drops significantly for subsequent countries.)[22] By default, Internet traffic takes the fastest route; the fastest route is often that which goes through the U.S. with its highly developed network. For example, the fastest route for traffic between Europe and China would be through the U.S. rather than a circuitous route across Europe and Asia.

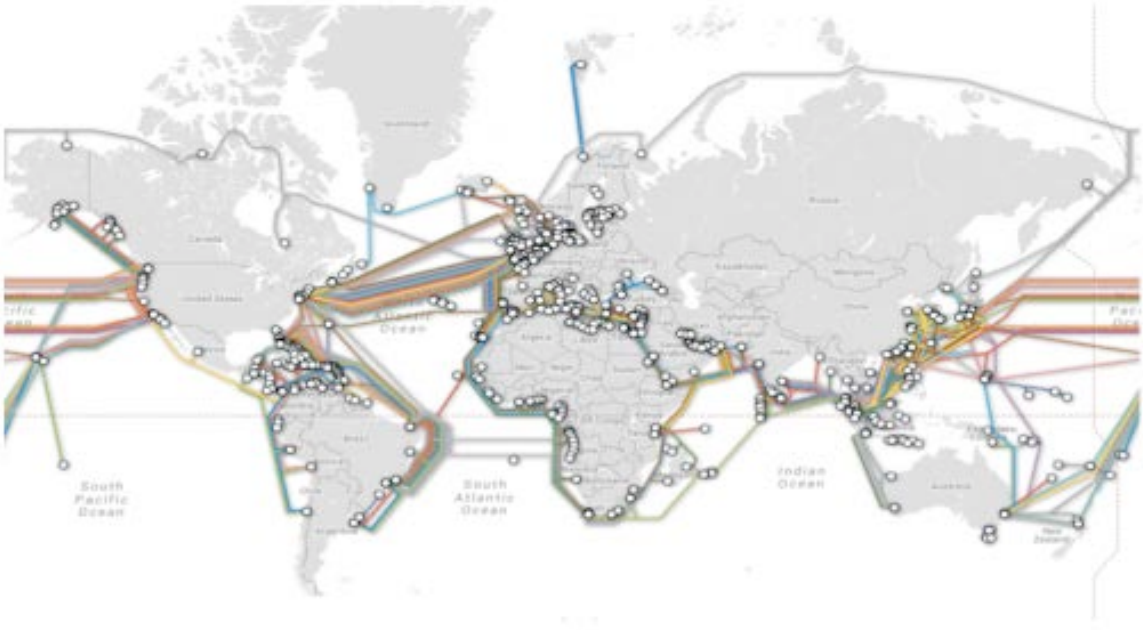

Figure 3: Submarine Cable Map (Source: TeleGeography, 2013, http://www.telegeography.com/telecom-resources/submarine-cable-map/index.html)

5.2 IXPs

In certain parts of the world, a lack of IXPs can contribute to Internet traffic passing through the U.S.

Researchers at the University of Toronto, Canada have investigated Internet routing, particularly as it pertains to Canadian traffic passing through the U.S. Their conclusion is that due to Canada's lack of IXPs, great amounts of the country's traffic pass through the U.S. Even traffic originating and terminating in one city in Canada, such as Toronto, takes a "boomerang route" through the U.S. As of 2012 Canada had only three IXPs while the U.S. had 86. Therefore, increasing the number of IXPs in Canada would reduce

[22] TeleGeography, "Submarine Cable Map" (PriMetrica Inc, 2013), http://www.telegeography.com/telecom-resources/submarine-cable-map/index.html.

the dependency on foreign routing.[23] Elsewhere in the world, Europe had 141 IXPs, Latin America 34, Africa 21, and Asia-Pacific 77.[24]

Figure 4: A Canadian boomerang route based in Toronto (Source: Clement, "IXmaps," p.222)

5.3 Internet content location

Another simple reason the U.S. sees so much Internet traffic is that the majority of Internet content is located in the U.S. Many of the web pages and files that people the world over access, reside on servers in the U.S. The Internet traffic must go there to retrieve the requested files.

5.4 Policies, agreements, and law

In one of the papers by the researchers at the University of Toronto, the author states, "It is well known… 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."[25] Another paper says, "Political economic relationships between carriers contribute to routing patterns."[26] While Internet traffic is set to take the fastest route, a network administrator can override this and set the traffic to take the route that is cheapest for the company or that is according to other network policies.[27]

Due to the globalization of the telecommunications industry, many purely international calls are routed through the U.S. A communications privacy expert who previously worked at the NSA claimed that in the early 2000s the U.S. government was "quietly encouraging the communications industry to increase the amount of international traffic that is routed through American-based switches."[28]

---

[23] Obar and Clement, *Internet Surveillance and Boomerang Routing*.
[24] Patrick S. Ryan and Jason Gerson, *A Primer on Internet Exchange Points for Policymakers and Non-Engineers*, SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, August 11, 2012), 21, http://papers.ssrn.com/abstract=2128103.
[25] Clement, "IXmaps - Tracking Your Personal Data through the NSA's Warrantless Wiretapping Sites," 222.
[26] Obar and Clement, *Internet Surveillance and Boomerang Routing*.
[27] Author's email exchange with Hadi Asghari, November 21, 2013.
[28] Marcus, *Declaration of J. Scott Marcus,* Exhibit B.

Company policy can also contribute to Internet data being located in the U.S. Google, for example, replicates data for reliability purposes. If a person created a Gmail account in Europe, that account may be located on a European server. To ensure that the customer still has access to his or her email even if the server goes down, Google replicates that account 3 to 10 times and stores copies on servers around the world, including in the U.S. So while the customer has a European email account, a copy is stored in the U.S. Further, it is unclear where companies like Google put a customer's mailbox or how they decide where to store it. If the customer created it while in the U.S. it may have been stored on a U.S. server. If the customer subsequently relocates to Europe and is accessing his email from Europe, the mailbox may remain on the U.S. server.[29]

Another contributing factor to NSA's access to Internet data is U.S. law. Under the Patriot Act, any company with an office in the States is required to comply with U.S. law and therefore with NSA requests. Furthermore, because it is a U.S. company, the NSA can ask for data from their servers elsewhere in the world. Google is headquartered in the U.S. and therefore subject to U.S. law; therefore, if the NSA requests information from its servers in Europe, it is required to give it. Even if a company is European but has offices or employees located in the U.S. it must comply with U.S. law.[30]

As far as Europe-to-Europe Internet traffic, normal traffic would not pass through the U.S. This includes peer-to-peer (P2P) file sharing – a decentralized network with no central server, and in which network users share tasks (e.g. streaming audio/ video) and resources (processing power, disk storage). A portion of each user's resources is made available to all the other users.[31] Another example of traffic that would stay within Europe is emails sent from one university account to another. A TU Delft email account to a TU Delft email account would stay in Europe; however, a TU Delft email account to a Hotmail account may go through the U.S.[32]

6 NSA surveillance technologies

6.1 Cable tapping

Taps on fiber-optic cables are known to take place under the TEMPORA program attributed to UK signals intelligence, and may also be the means used in NSA's "Upstream" and MUSCULAR operations. According to an NSA slide, Upstream is "the collection of communications on fiber cables and infrastructure as data flows past." (See figure 7.) MUSCULAR is the collection of data from Google and Yahoo's internal networks.

[29] Asghari, interview.
[30] Ibid.
[31] Schollmeier, Rudiger, "A Definition of Peer-to-Peer Networking for the Classification of Peer-to-Peer Architectures and Applications," in *First International Conference on Peer-to-Peer Computing, 27-29 August 2001, Linköping, Sweden: Proceedings* (Los Alamitos, Calif: IEEE Computer Society, 2002).
[32] Asghari, interview.

In the early 2000s there was a court case against AT&T, which disclosed that the NSA was wire-tapping AT&T's fiber-optic cables in San Francisco.[33] It is likely that this is a technology the NSA still uses as part of its surveillance program. The following sections examine the technology involved in wire-tapping.

6.1.a Fiber-optics technology interception

Fiber-optic cables carry the world's communications across the globe. More than 550,000 miles of these undersea cables connect our world,[34] carrying 99 percent of the world's intercontinental data.[35] Eighty percent of this fiber-optic data flows through the U.S.[36]

Standard fiber-optic cables over land consist of 144 individual glass fibers; those undersea consist of only 8 individual glass fibers. Each fiber carries 40-160 light wave signals and each of these signals handles 10-40 gigabits of traffic; this translates into a massive flow of data.[37] The data is turned into "ultra-short flashes of light. These flashes represent the zeros and ones that all digital information is comprised of. A photodiode at the end of the cable turns the light flashes back into electrical signals."[38] The signals, however, have to be re-amplified about every 80 kilometers otherwise they will drop. The signal is re-amplified with a regenerator. Each fiber optic has to be to re-amplified separately, so the fiber optics must be laid out separately rather than bundled together. This is the weak point at which the cable can be more easily tapped into.[39]

There are various methods that the NSA has been known or suspected of using for tapping communications cables. One method was used in the 1970s and 80s against the Soviets. It involved submarines and divers dropping waterproof recording pods on submarine cables off the eastern coast of the USSR. The tapes were gathered every few weeks and listened to by the NSA.[40] In 2005 it was reported that the submarine USS Jimmy Carter was being used to transport technicians to the bottom of the ocean to tap fiber-optic cables.[41] This would indicate that the NSA continued the practice of mid-ocean cable taps until at least as recently as nine years ago. What undoubtedly changed was the manner of transmitting the information back to the NSA.[42]

---

[33] In 2008 Congress gave legal immunity to telecom companies participating in the NSA's wire tapping program. The AT&T case was subsequently dropped.

[34] David A. Fulghum, "Pressure Mounts to Find New Science to Meet Cyber-Intel Needs," *Aviation Week & Space Technology*, Printed Headline: Electronic Blitz, March 29, 2010, 58.

[35] Craig Timberg and Ellen Nakashima, "Agreements with Private Companies Protect U.S. Access to Cables' Data for Surveillance."

[36] Olga Khazan, "The Creepy, Long-Standing Practice of Undersea Cable Tapping."

[37] Fulghum, "Pressure Mounts to Find New Science to Meet Cyber-Intel Needs."

[38] Fabian Schmidt, "Tapping the World's Fiber Optic Cables | Sci-Tech | DW.DE | 30.06.2013," *DW.DE*, June 30, 2013, http://www.dw.de/tapping-the-worlds-fiber-optic-cables/a-16916476.

[39] Ibid.

[40] Khazan, "The Creepy, Long-Standing Practice of Undersea Cable Tapping."

[41] Ibid.

[42] Based on the case against the Soviets, we can infer that these mid-ocean taps are not literally in the middle of the ocean, but rather off the coasts of countries. In this case the taps were off the USSR shoreline in the Sea of Okhotsk and the Barents Sea. ("Tapping Into Soviet Secrets," *Sun Sentinel*, accessed

An alternative to mid-ocean taps is tapping the cables when they land ashore as has been revealed to be the case in TEMPORA. In this case the following tapping methods could be used: using "intercept probes" to capture the light sent across the cable; slightly bending the cables and capturing the content with a receiver; installing splitters.[43] (This alternative does not negate the possibility of the NSA doing both ocean taps and taps ashore.)

Using intercept probes requires gaining access to the landing station. The probes "capture the light being sent across the cable," bounce it through a prism, make a copy of it, and turn "it into binary data without disrupting the flow of the original Internet traffic."[44]

To put a bend in the cable so that light can be extracted out of it and the data captured out of the light, the cable is run around a small cylindrical device. This causes the cable to emit a certain amount of light, which is captured by a receiver along with the data it contains.[45]

However, the easiest and most reliable way is to have the consent of the transit company and split the wires.[46] Given the apparent relationship the NSA has with telecommunications companies and its history of tapping AT&T's cables, it seems likely that it has gained companies' consent and access to landing stations or otherwise to install splitters.

6.1.b Fiber optic splitters

In 2006 a retired AT&T technician, Mark Klein, revealed that the NSA had installed surveillance equipment at AT&T's Internet exchange point in San Francisco. There was a court case against AT&T based on these revelations, and the documents related to this case are the most extensive related to how the NSA installs and uses this type of surveillance technology. Looking at this case can give a good indication of how the NSA most probably still employs this kind of surveillance technology.

To intercept AT&T's signal at 611 Folsom St. in San Francisco, the NSA installed splitters. A splitter is a piece of equipment that physically splits the fiber/cable and sends the signal in two different directions. The simplest form of splitter is a "T" which has one signal coming in and two signals going out. The NSA splitters used in this case were 50/50 splitters – of the signal that came in, 50% of it went out one fiber and 50% went out the other. This does not refer to 50% of the data, but to 50% of the signal. That is, the signal is split in half, making it weaker, but 100% of the data is sent through each of the two signals. Essentially this means that a copy of the data is being made. The data is

---

March 11, 2014, http://articles.sun-sentinel.com/1998-11-08/news/9811070278_1_soviet-plans-spy-submarines-book-reports.)

[43] Khazan, "The Creepy, Long-Standing Practice of Undersea Cable Tapping."

[44] Ibid.

[45] Shona Ghosh, "How Spies Could Tap Fibre Cables," *PCPro*, June 27, 2013, http://www.pcpro.co.uk/news/security/382666/how-spies-could-tap-fibre-cables#ixzz2ZE2mKAJ1.

[46] Ibid.

then sent on its way to its original destination through one signal, and a copy of this data is sent into another cable owned and operated by the NSA.

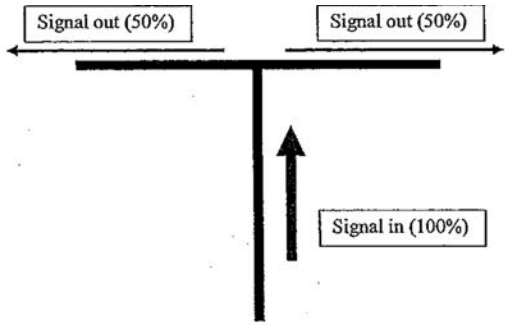

Figure 5: 50/50 Splitter (Source: *Declaration of J. Scott Marcus*, p.13)

Effectively, these splitters were installed at an Internet Exchange Point. At the San Francisco facility AT&T exchanged Internet traffic with 16 companies[47] with which it had peering arrangements. The splitters diverted traffic related to AT&T's Common Backbone, the network that provides Internet access (as opposed to telephone traffic). It is believed that all, or substantially all of AT&T's peering traffic was diverted through the splitters, but that its own traffic was not. The splitters were placed on the fiber-optic cables belonging to these 16 ISPs before they reached the AT&T Common Backbone. Thus, their traffic was copied, but AT&T's own traffic was not. See figure 6 below.[48]

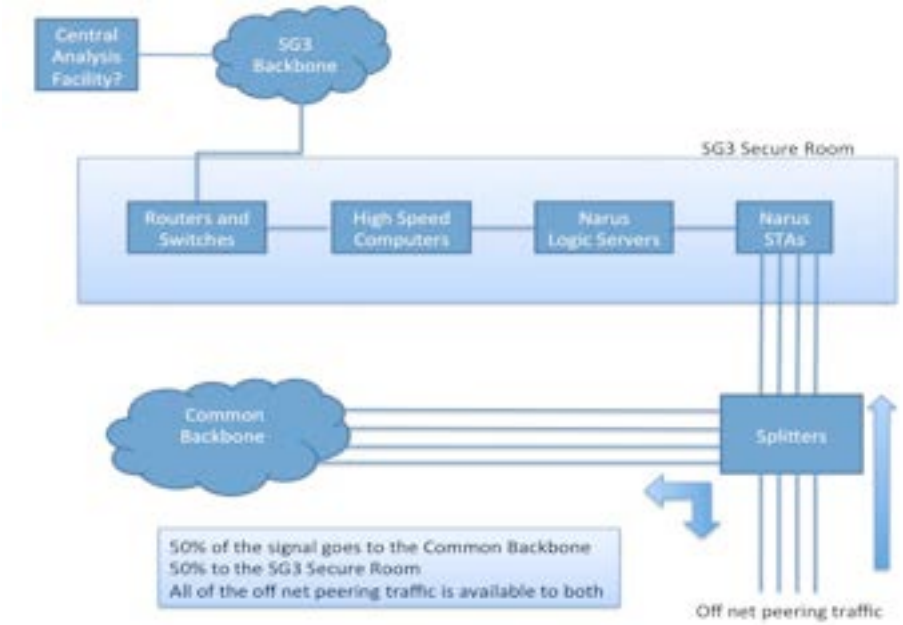

Figure 6: Configuration of splitters and SG3 Secure Room at AT&T (After source: *Declaration of J. Scott Marcus*, p.16)

[47] Mark Klein, *Declaration of Mark Klein in Support of Plaintiffs' Motion for Preliminary Injunction*, United States District Court Northern District of California, Case of Hepting, Hicks, Jewel et al. v. AT&T et al., 2006, B-20, https://www.eff.org/node/55051.
[48] Marcus, *Declaration of J. Scott Marcus*, 15–16.

6.1.c Deep packet inspection

After passing through the splitters, the copied data was sent into a secure room at the San Francisco facility where it was processed by Narus equipment.

Narus is deep packet inspection (DPI) technology. DPI is the inspection and analysis of Internet traffic in real time. It extracts not just basic protocol information, such as source and destination IP addresses (this is referred to as shallow packet inspection), but also the deeper layers of the traffic, which consist of the actual content of the traffic.[49]

DPI was designed by engineers for Internet Service Providers to optimize their networks. "The primary technical capability underlying DPI is the ability to recognize. DPI has been developed to detect, for example, applications, protocols, media content, viruses or data in a specific format, such as credit card numbers."[50] Once data is recognized the ISP can manipulate it, for example, by blocking or prioritizing certain traffic. It can also carry out notification actions, such as "generating reports, alarms or billing incidents."[51] Using DPI, ISPs can prioritize applications like VoIP to improve service or identify illegal downloads.[52] In other words, ISPs use DPI to look at the Internet activity of their customers and to act upon it. Are their customers using HTTP for web streaming; what kinds of files are they sending?

The following key functions of DPI use have been identified by a 2009 study: network security, bandwidth management, government surveillance, content regulation, copyright enforcement, and ad-injection.[53] (Government surveillance here refers to lawful interception by police and – not necessarily lawful – surveillance performed by the intelligence community; it is in the context of governments pressuring ISPs to use DPI.) Bandwidth management has been found to be the primary application for DPI.[54]

A basic finding of a 2012 study by Asghari, van Eeten, and Mueller is that deep packet inspection is widely used by ISPs globally. In 2011, more than two-thirds of the 75 countries examined showed DPI use, with half showing some level of DPI use and the other half showing "pervasive use."[55] China is one of the countries with pervasive DPI

[49] Alcatel-Lucent, "Mobile Application Assurance on the Alcatel-Lucent 7750 Service Router Mobile Gateway," 2011, 2.
[50] Hadi Asghari, Michel van Eeten, and Milton Mueller, "Unravelling the Economic and Political Drivers of Deep Packet Inspection" (presented at the GigaNet 7th Annual Symposium, Baku, Azerbaijan, 2012), 3.
[51] Ibid.
[52] Alex Wawro, "What Is Deep Packet Inspection?," *PCWorld*, February 1, 2012, http://www.pcworld.com/article/249137/what_is_deep_packet_inspection_.html.
[53] Ralf Bendrath, "Global technology trends and national regulation: Explaining Variation in the Governance of Deep Packet Inspection," 2009, as cited in Asghari, van Eeten, and Mueller, "Unravelling the Economic and Political Drivers of Deep Packet Inspection," 5–6.
[54] Body of European Regulators for Electronic Communications, "A view of traffic management and other practices resulting in restrictions to the open Internet in Europe, 2012, as cited in Asghari, van Eeten, and Mueller, "Unravelling the Economic and Political Drivers of Deep Packet Inspection," 5.
[55] Hadi Asghari, Michel van Eeten, and Milton Mueller, "Unravelling the Economic and Political Drivers of Deep Packet Inspection," 9.

use[56] where ISPs use DPI to block access to certain websites such as YouTube.[57] Another observation of the Asghari et al. study is that over the three-year period studied from 2009 to 2011, ISPs seemed to be moving toward using DPI "for a smaller portion of their subscriber base or only during peak hours."[58]

Returning to the Narus system in the AT&T San Francisco facility, the Narus equipment in the secure room consisted of two parts – the Narus Semantic Traffic Analyzer (STA) 6400 and the Narus Logic Server. Steve Bannerman, Vice President of Narus Marketing explained that Narus works by monitoring all Internet traffic, looking at the 1s and 0s that make up the data generated by Internet activity. It looks at the different layers in data packets, including the first layer, which is the "to" and "from" addresses and the deeper layers, referred to as the "payload," composing the actual content of what is being sent.[59]

Part one of the Narus system – the STA 6400 – monitors data packets for metadata that matches "key pairs," such as "a specific IP address or a range of IP addresses, a keyword within a Web browser request, or a pattern identifying a certain type of traffic such as a VPN or Tor connection."[60] The matching packets are put into the second part of the system, the Narus Logic Server. This second part consists of analytic processing systems that re-assemble the network sessions of the matching packets, mine them for "metadata, file attachments, and other application data" then index and deposit the data into a database.[61] Narus is capable of processing huge volumes of data and of storing the traffic that it captures.

Also present at the AT&T facility was what was referred to as the "SG3 backbone." This was a private backbone network, connected to the secure room, which was separate from AT&T's backbone. J. Scott Marcus, former Chief Technology Officer for the telephone company GTE and former adviser to the Federal Communications Commissions, believed that this indicated that after the data was collected and processed to identify data of interest by the Narus equipment, it was sent via the SG3 backbone to central locations for further analysis. (Additional aspects, as well as limitations of the Narus system are examined in paragraph 4.6.4.a.)

6.1.d Possible additional locations of NSA splitters

AT&T staff revealed that besides San Francisco, NSA surveillance sites also existed in Atlanta, Los Angeles, San Diego, San Jose, and Seattle. Using criteria based on the assumption that the NSA would want to intercept the most data with the fewest

[56] Ibid.
[57] Wawro, "What Is Deep Packet Inspection?".
[58] Asghari, van Eeten, and Mueller, "Unravelling the Economic and Political Drivers of Deep Packet Inspection," 10.
[59] PBS Frontline, "Spying on the Home Front," May 15, 2007, http://www.pbs.org/wgbh/pages/frontline/homefront/.
[60] Sean Gallagher, "Building a Panopticon: The Evolution of the NSA's XKEYSCORE," *Ars Technica*, August 9, 2013, http://arstechnica.com/information-technology/2013/08/building-a-panopticon-the-evolution-of-the-nsas-XKEYSCORE/.
[61] Ibid.

number of splitter sites, researchers at the University of Toronto have identified another 12 cities that likely host NSA splitters. Of 1,319 U.S.-only Internet data traffic routes in the researchers' database, only seven did not pass through one of these 18 cities suspected to have NSA surveillance sites. That means that with splitters at these locations the NSA could intercept 99% of U.S.-only trace routes.[62]

6.1.e Concluding remarks on fiber-optic cable tapping

The splitters at the AT&T San Francisco location were never known to have been removed. These splitters and those at other U.S. cities could be what the Upstream code names in the PRISM slides refer to – FAIRVIEW, STORMBREW, BLARNEY, OAKSTAR. (See figure 7.) In any case it seems likely that the NSA surveillance technology being employed in Upstream are splitters placed on the fiber-optic cables.

As in the San Francisco case, the TEMPORA operation, attributed to UK signals intelligence, involves agreements with telecom companies. It is known that this operation is tapping into fiber-optic cables and that it is working with the following companies: BT, Verizon Business, Vodafone Cable, Global Crossing, Level 3, Viatel, and Interoute.[63] The set-up seems very similar to that in the AT&T case and the surveillance technology used may be as well.

The program, MUSCULAR, in which the NSA exploits the data links connecting Yahoo and Google data centers, could also be using splitters and DPI to tap the fiber-optic cables connecting the data centers.

6.2 PRISM

The NSA program, PRISM, seems, particularly in Europe, to have become synonymous with all NSA leaks. Many talks, conferences, etc. entitled "PRISM," in fact, refer to the NSA leaks and mass surveillance as a whole.[64] For this reason, as well as examining this program as it relates to mass surveillance, we discuss it here. The Guardian article revealing the PRISM program reported that this program gave the NSA direct access to the servers of major Internet providers such as Google, Apple, Skype, and Yahoo. This 'revelation' was based on a slide for training analysts about PRISM (see figure 7 below). The slide states that there are two types of collection in FAA702 Operations. One is 'Upstream,' collecting communications on fiber cables, and the other is PRISM, "collection directly from the servers" of nine U.S. Internet services providers. The interpretation by *The Guardian* and *The Washington Post* was that this meant these companies were collaborating with the NSA to give it a direct connection to their

[62] Clement, "IXmaps - Tracking Your Personal Data through the NSA's Warrantless Wiretapping Sites."
[63] James Ball, Luke Harding, and Juliette Garside, "BT and Vodafone among Telecoms Companies Passing Details to GCHQ," *The Guardian*, August 2, 2013, sec. Business, http://www.theguardian.com/business/2013/aug/02/telecoms-bt-vodafone-cables-gchq.
[64] For example, the Computers, Privacy and Data Protection conference, Jan. 2014 had a session titled, "The EU Response to PRISM," http://www.cpdpconferences.org/Programme.html#top, and the Annual Conference on Data Protection in the EU 2014 has a half day of talks entitled "The Aftermath of 'PRISM,'" https://www.era.int/upload/dokumente/15735.pdf

servers, to "unilaterally seize" all manner of communications from them.[65] This proved, however, to be erroneous.

Figure 7: NSA PRISM slide (Source: NSA PRISM Collection Manager's Presentation on PRISM, http://www.washingtonpost.com/wp-srv/special/politics/prism-collection-documents/)

The NSA routinely and emphatically denied that it had direct access to company servers (in contrast to the leak regarding Verizon phone records which was admitted to by Washington officials), as did the companies themselves.[66] Director of National Intelligence James Clapper, released a statement, which said, "PRISM is not an undisclosed collection or data mining program. It is an internal government computer system used to facilitate" the collection of foreign intelligence gathered under section 702 of FISA.[67]

From a business point of view, Daniel Soar makes the point that the idea that this kind of collaboration was going on was "at some level outlandish: in most cases… the data the company possesses is what generates its phenomenal value, and it was hard to

[65] Glenn Greenwald and Ewen MacAskill, "NSA Prism Program Taps in to User Data of Apple, Google and Others," *The Guardian*, June 7, 2013, sec. World news, http://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data.
[66] Declan McCullagh, "No Evidence of NSA's 'Direct Access' to Tech Companies," *CNET*, June 7, 2013, http://news.cnet.com/8301-13578_3-57588337-38/no-evidence-of-nsas-direct-access-to-tech-companies/.
[67] James Clapper, "Facts on the Collection of Intelligence Pursuant to Section 702 of FISA," June 8, 2013, http://www.dni.gov/files/documents/Facts%20on%20the%20Collection%20of%20Intelligence%20Pursuant%20to%20Section%20702.pdf.

imagine that this commercially priceless property would be freely shared with anyone, let alone with the government."[68]

A PowerPoint presentation is by nature a simplification of information to convey the main ideas. In addition to this lack of detail, the PRISM slides are somewhat of a sales-pitch by the PRISM Collections Manager to convince new analysts of the superiority of PRISM over other possible programs they could be using. These two factors result in a phrase such as 'Collection directly from servers,' which does not in actuality tell us much about the program and leaves various interpretations possible.

The "direct access" described here is access to a particular account through a court order for that particular account, not a wholesale sucking up of all the information on the company's users. The court order gives the NSA access to the targeted account as well as to the accounts it is in contact with. This follows in the same principle of a court order on a phone number yielding the phone numbers the targeted phone has communicated with (with the obvious difference of a Facebook page containing content, while phone metadata does not). The NSA and the attorney general serve a court order on one of these companies for one or more foreign accounts, say for example, several Facebook accounts in Pakistan. Those accounts are then monitored and their activity is sent back to the NSA. This is where "direct access" fits in – the NSA has access to the account in real time.[69] Marc Ambinder, writing in The Week, speaks of a 'mirror' of the accounts that the company somehow creates and only the NSA has access to. When the selected account is updated the Facebook server and the mirrored server are both updated in real-time. PRISM is the tool that allows the analyst to monitor and analyze this data and all the data on foreign targets provided to the NSA by Internet companies in the U.S.[70]

Another leaked slide shows a screenshot of how to search PRISM entries (see figure 8 below). The total count of records given on the slide is 117,675. This is a relatively small number given how many Gmail, Facebook, etc. accounts exist across the globe. That suggests that in April 2013 when the screenshot was taken, the NSA was monitoring no more than 0.01% of the billion Facebook users on the planet.[71] Granted, there are many other unknowns, such as, are these only entries under the category 'Counterterrorism' with other categories containing more entries. Alternatively, this screenshot suggests that the "entire list" could include other agencies than the NSA using PRISM, like the FBI, and therefore the NSA list would be even smaller. In any case, this number seems to support the point that PRISM is used for more targeted investigation. Some may argue that it is still mass surveillance (depending on how one defines this term), but it is indisputably further along the spectrum toward "targeted" than splitting cables. (And therefore, in the authors' opinion is a misguided choice of a title for discussions on NSA mass surveillance.) The procedures under PRISM for gaining information from company

[68] Soar, "How to Get Ahead at the NSA."
[69] Ibid.
[70] Marc Ambinder, "Solving the Mystery of PRISM," *The Week*, June 7, 2013, http://theweek.com/article/index/245360/solving-the-mystery-of-prism.
[71] Soar, "How to Get Ahead at the NSA."

servers follows that of obtaining phone records and in this sense, is a more traditional form of targeted surveillance used in a new medium.

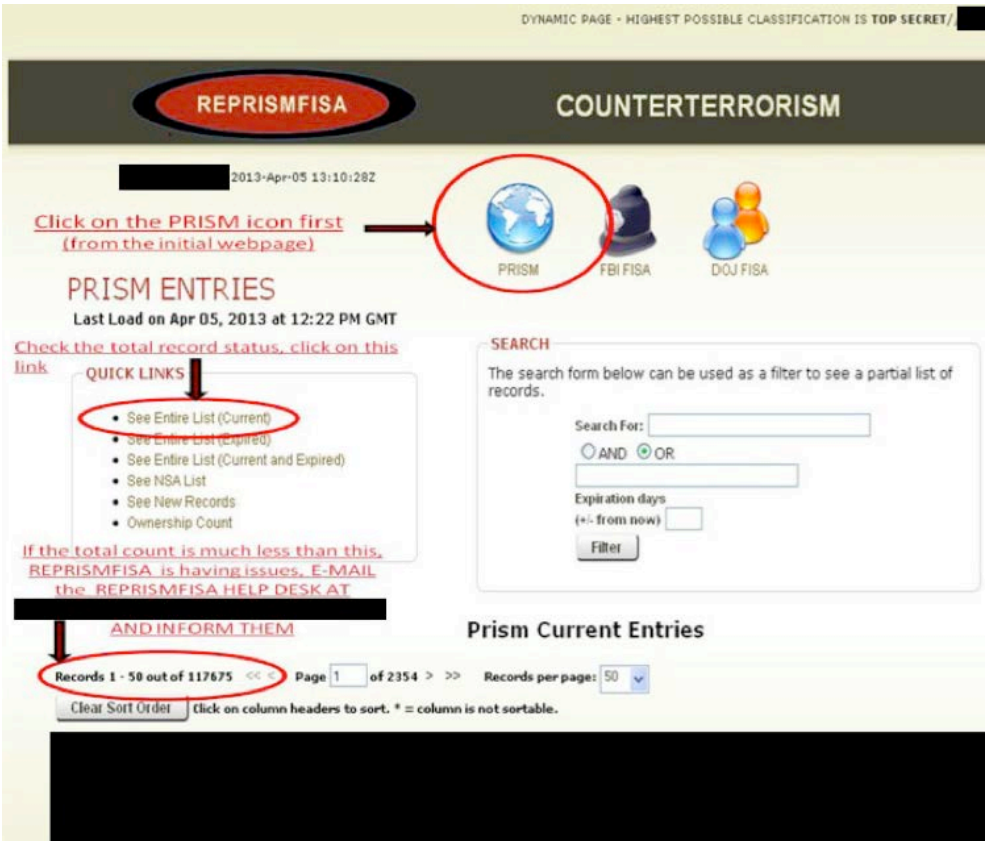

Figure 8: NSA PRISM slide (Source: NSA PRISM Collection Manager's Presentation on PRISM, http://www.washingtonpost.com/wp-srv/special/politics/prism-collection-documents/)

6.3 Decryption, or circumventing encryption

One of the NSA's core missions is to decipher and break codes, so it is no surprise that the leaks have shown a primary focus of the agency's surveillance to be decryption. What is interesting, however, is that the majority of their efforts do not involve actually breaking codes, but exploiting human elements and implementation software to circumvent encryption. Further, while mass surveillance programs (see paragraph 6.4.a on XKEYSCORE) aid in identifying machines to exploit, the actual decryption – or circumventing of encryption – is done on a targeted, rather than mass surveillance scale.

6.3.a Defeating encryption

BULLRUN is the NSA's program dealing with defeating encryption. According to the Classification Guide for Project BULLRUN, the project's sources include Computer Network Exploitation (hacking), "interdiction, industry relationships, collaboration with other IC [Intelligence Community] entities, and advanced mathematical techniques."[72] The Guide states that the NSA has "some capabilities" against encryption in HTTPS, VoIP, Secure Sockets Layer (SSL; used for online banking) VPNs, SSH, and Webmail. The

[72] NSA, Cryptanalysis and Exploitation Services, "Classification Guide: PROJECT BULLRUN," June 16, 2010.

Guide also make explicit that "'capabilities against a technology' does not necessarily equate to decryption."

Obtaining keys

One of the NSA's circumvention methods is to obtain the encryption keys. The agency has several ways to do this – court orders requiring companies to turn over keys; hacking into companies to obtain the keys; securing companies' willing collaboration. The agency reportedly maintains of database of the keys it collects.[73]

Backdoors

The NSA has also persuaded or coerced companies into installing backdoors into their security systems. It can then get around the encryption by using these backdoors.

Influencing encryption standards

Another method the NSA purportedly employs is to influence encryption key standards. This would probably be done through the National Institute for Standards and Technology (NIST) "which sets U.S. cryptography standards and is influential worldwide."[74] The MIT Technology Review's IT editor for hardware and software argues that it is unlikely that the NSA compromised many of these most widely used standards because there were developed by open groups outside the U.S. The one standard it did play a significant role in developing is part of a cryptography toolkit – Suite B – used by the U.S. government and its contractors. "Introducing backdoors into that would seem counterproductive to the NSA."[75]

Brute force

The NSA also uses brute computer force to break weak encryption. Therefore, although there is not yet evidence that the NSA has cracked SSL, experts have long warned that the keys typically used with SSL are not long enough. A government agency or a large company with significant resources could break the 1,024 bits long keys that most sites use for SSL. Longer keys are necessary to protect against this kind of attack, but few companies use them. (Google and Facebook are just this year switching to longer keys.)[76]

---

[73] Tom Simonite, "Circumventing Encryption Frees NSA's Hands Online," *MIT Technology Review*, September 6, 2013, http://www.technologyreview.com/news/519131/circumventing-encryption-frees-nsas-hands-online/.
[74] Tom Simonite, "NSA Leak Leaves Crypto-Math Intact but Highlights Known Workarounds," *MIT Technology Review*, September 9, 2013, http://www.technologyreview.com/news/519171/nsa-leak-leaves-crypto-math-intact-but-highlights-known-workarounds/.
[75] Ibid.
[76] Ibid.

6.3.b FOXACID

The NSA's attacks using FOXACID are a good example of how it circumvents encryption and exploits human and software implementation elements. FOXACID is a Computer Network Exploitation system that matches potential targets with prepared attacks. It is a modular system that allows exploits to be changed if discovered and only launches certain attacks against certain targets. FOXACID is used to perform all kinds of attacks, including those against Tor users.[77]

Tor – The Onion Router – is an online anonymity network. It is a problem for law enforcement because criminals use it for communication and it makes identification of the user, hence, the criminal, excessively more difficult.[78] Tor works by routing data packets through multiple nodes, or relays, rather than taking the most direct path. Thus, it removes predictability. Three relays are randomly selected with a preference for those that are further apart geographically. The circuit is switched about every ten minutes.[79] Each relay only knows the relay the data came from and the next relay it is going to. No relay knows the entire path of the data, making it impossible to know the source and destination of the data. The information is encrypted with each hop using different encryption keys. The one unencrypted link is the exit node, where the data packet is decrypted and exits the Tor network to arrive at its destination.[80] To the websites visited, the location of the Tor user appears random.

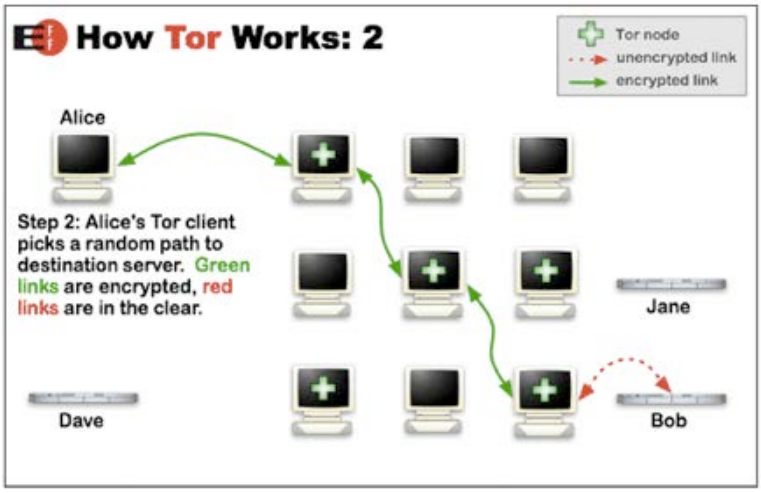

Figure 9: How Tor Works (Source: Tor project https://www.torproject.org/about/overview)

[77] Bruce Schneier, "Attacking Tor: How the NSA Targets Users' Online Anonymity," *The Guardian*, October 4, 2013, sec. World news, http://www.theguardian.com/world/2013/oct/04/tor-attacks-nsa-users-online-anonymity.

[78] A recent (and somewhat amusing) example was that of a Harvard student wishing to avoid an exam, using Tor to send bomb threats via email, resulting in several buildings being evacuated (and his exam cancelled). ("Harvard student tried to dodge exam with bomb hoax, FBI says," NBC news, http://www.nbcnews.com/news/us-news/harvard-student-tried-dodge-exam-bomb-hoax-fbi-says-v21943608)

[79] Karen Reilly, Development Director of The Tor Project in *NSA Surveillance: What We Know; What to Do About It*, Conference, Cato Institute Conference, 2013, http://www.cato.org/events/nsa-surveillance-what-we-know-what-do-about-it.

[80] "Tor," accessed November 26, 2013, https://www.torproject.org/about/overview.html.en.

To execute an attack, the first step is to identify Tor users on the Internet. This is easy for the NSA to do because the characteristics that make Tor anonymous also make all Tor users look the same on the Internet. However, the user's identify and location remain unknown. "The NSA creates 'fingerprints' that detect http requests from the Tor network to particular servers."[81] The fingerprints are then put into a database and data analysis tools are used to sort through all the Internet traffic the NSA monitors to identify Tor connections. (It seems likely that the analysis tool used for this is XKEYSCORE – see paragraph 6.4.a.)

Once a Tor user is identified the NSA uses its secret servers on the Internet backbone, codenamed QUANTUM, to redirect those users to other secret servers, codenamed FOXACID. Because QUANTUM servers are at key locations on the Internet backbone, they can react faster than other websites and thus impersonate the website the user is wanting to access. They can respond to the request before the actual website can and they look the same as the actual website. The target's browser is thereby fooled into contacting the FOXACID server. These kinds of attacks are a sort of race and are hard for anyone besides the NSA to execute because they depend on "a privileged position on the Internet backbone."[82] This is known as a man-in-the-middle attack (see figure 10).

"By the time the NSA tricks a target into visiting one of those servers, it already knows exactly who that target is, who wants him eavesdropped on, and the expected value of the data it hopes to receive."[83] Based on this, FOXACID automatically determines which exploits are best to serve against the particular target. It performs a risk-benefit analysis, considering factors such as the technical sophistication of the target, the value gained from a successful attack, the risk of discovery, the value and rarity of the exploit, etc.[84] In the case of Tor, FOXACID attacks through the Tor browser bundle (a group of programs designed to make the installation and use of Tor software easier), exploiting vulnerabilities in the Firefox web browser. Once an attack has been successfully executed, the infected computer calls back to the FOXACID server, which then further infects the computer, compromising it long-term and providing the NSA with ongoing information.[85]

Note that although the NSA can execute these attacks against Tor it cannot do so on a large scale. Nor can it do so on demand. These are targeted attacks against individual users, not blanket attacks against all Tor users. According to 2012 NSA PowerPoint slides, the agency "will never be able to de-anonymize all Tor users all the time." It has had success de-anonymizing "a **very small fraction** of Tor users" with manual analysis,

[81] Schneier, "Attacking Tor."
[82] Ibid.
[83] Bruce Schneier, "Crypto-Gram Newsletter, October 15, 2013," October 15, 2013, https://www.schneier.com/crypto-gram-1310.html.
[84] Ibid.
[85] Schneier, "Attacking Tor."

but has had no success in de-anonymizing a Tor user on demand (bold in original document).[86]

This is apparently the same kind of attack that was used against Belgacom. When targeted employees visited their LinkedIn profiles a secret server responded with a fake page that infected their computer with malware.[87]

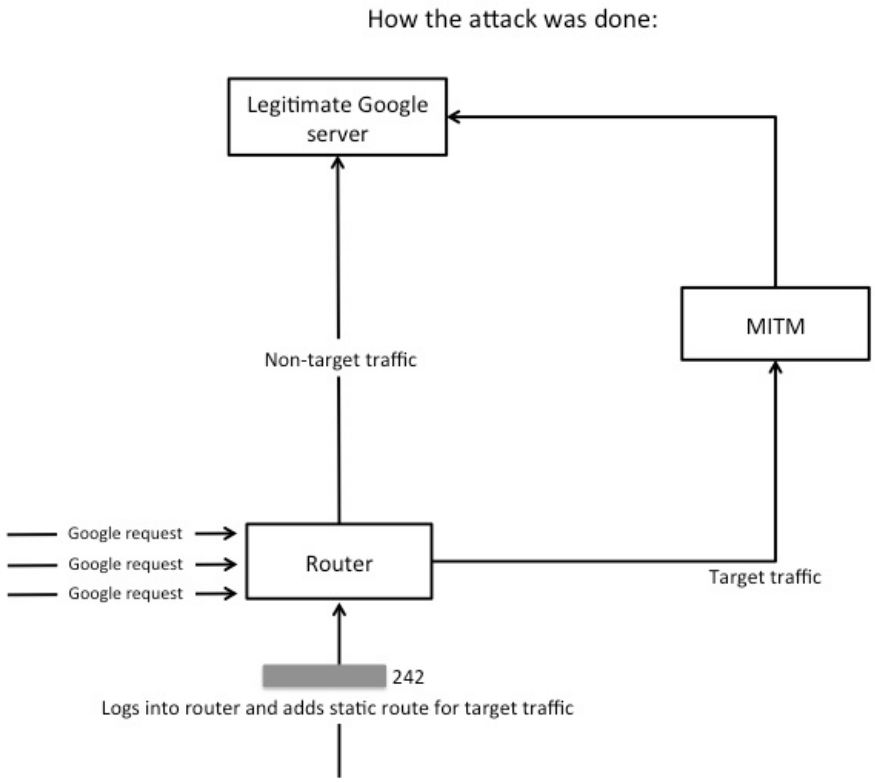

Figure 10: Illustration of a man-in-the-middle attack, with a QUANTUM server impersonating Google (After source: NSA diagram published by *The Guardian*, "Attacking Tor: how the NSA targets users' online anonymity," 4 Oct. 2013, http://www.theguardian.com/world/2013/oct/04/tor-attacks-nsa-users-online-anonymity)

The NSA capabilities under BULLRUN are interesting because they show the limits of its abilities. The agency cannot break encryption. Its attacks against encryption use must be done on a case-by-case basis. However, it also illustrates its ability (and potential) to bring more data under its surveillance by bypassing encryption. Obtaining each individual key or backdoor access is targeted action; what this yields could be targeted or mass surveillance (e.g. it could read all the traffic of a company once it has the encryption key).

[86] NSA, "Tor Stinks," Powerpoint presentation, June 2012.
[87] Staff, "GCHQ Targets Engineers with Fake LinkedIn Pages," *Spiegel Online*, November 11, 2013, http://www.spiegel.de/international/world/ghcq-targets-engineers-with-fake-linkedin-pages-a-932821.html.

6.4 Analysis tools and Databases

The NSA has a myriad of data analysis tools and databases to process the data it collects. Different databases store different kinds of data. For example, MARINA stores telephone metadata, while PINWALE stores recorded signals intercepts.[88] Marc Ambinder has an ongoing Mind Map of the NSA, in which he counts 20 SIGINT analytical and processing tools and 12 databases.[89] There are no doubt still dozens more. These systems do not collect information, but allow NSA analysts to sort through, locate, and analyze the troves of data that the NSA collects. XKEYSCORE is both an analytical tool and database. We have chosen to focus on this system as an example of NSA databases, as it is the one that we hear the most about via the leaked documents, and is at the same time, perhaps, the most complex and difficult to understand.

6.4.a XKEYSCORE

After the attacks of 9/11 the NSA needed a quick way to increase its Internet surveillance. It did this by purchasing "off-the-shelf" systems such as Narus. Narus is a two-part deep packet inspection system. As discussed in paragraph 6.1.b, part one of the system is the Narus Semantic Traffic Analyzer 6400 and part two is the Narus Logic Server. Part one monitors data packets for metadata that matches "key pairs" and part two processes the matching packets, mining, indexing, and finally depositing them into a database.

The amount of Internet traffic that Narus can handle monitoring is directly related to how many rules have been loaded into the machine watching the metadata flow past (part 1 of the system). The more rules, or pre-configured filters, turned on, "the more compute power burned and memory consumed per packet, and the fewer packets that can be handled simultaneously."[90] In the Narus system, if all the pre-configured filters are turned on then the system can monitor 12 gigabits out of 20 (on a two-way 10 gigabit Ethernet connection). To make the system more efficient and able to monitor more of the 20 gigabits flowing past, some of these filters have to be turned off. "In other words, to handle really big volumes of data and not miss anything with a traffic analyzer, you have to widen the scope of what you collect."[91]

This meant that a lot of data was collected. The problem for the NSA was where to store it and how to get it there. "Even when you store just the cream skimmed off the top of the 129.6 terabytes per day that can be collected from a 10-gigabit network tap, you're still faced with at least tens of terabytes of data per tap that need to be written to a database."[92] It was physically impossible for the NSA to get all that information back to its central database. So it created XKEYSCORE. XKEYSCORE solves the problem by storing

[88] Ambinder, "Solving the Mystery of PRISM."
[89] Marc Ambinder, "Mind Map - National Security Agency / Central Security Servi...," *MindMeister*, accessed December 10, 2013, http://www.mindmeister.com/326632176/national-security-agency-central-security-service-ts-comint-eci-orcon-noforn.
[90] Gallagher, "Building a Panopticon."
[91] Ibid.
[92] Ibid.

the data packets in local caches rather than sending it back to a central database as in the AT&T case. Since the advent of XKEYSCORE the agency can now store 3 days worth of raw packet data and 30 days worth of metadata in the local caches.

XKEYSCORE *is not* the system that actually captures the data via the fiber-optic cable tap. Rather it is the analyzing system and the database. The system that captures data packets from NSA wire taps is code-named TURMOIL. XKEYSCORE processes the data that TURMOIL brings in.

It processes it by running plug-ins, analysis engines that look for specific content in the captured data packets. XKEYSCORE has plug-ins for email addresses, phone numbers, webmail and chat activity, and extracted files, among others. "For selected traffic, XKEYSCORE can also generate a full replay of a network session between two Internet addresses."[93]

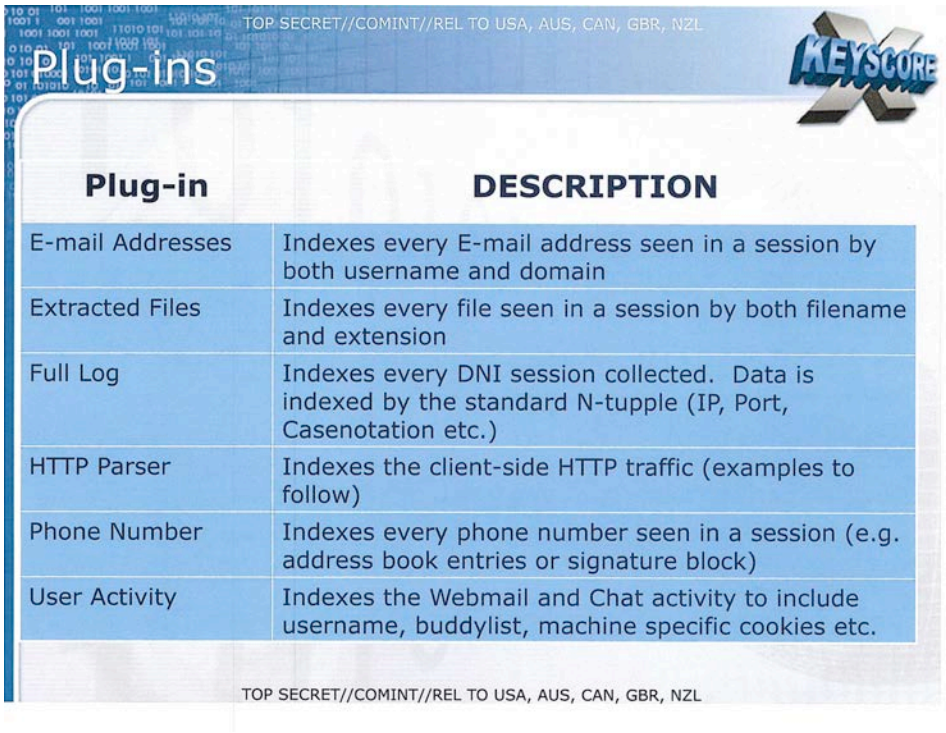

Figure 11: Slide from XKEYSCORE presentation showing examples of plug-ins

The plug-ins extract the metadata from each Internet session and index it into tables. XKEYSCORE can track, cross-index, and search any kind of metadata that can be extracted from an Internet session – log ins, email addresses, the use of encryption, language use, IP address geolocation, etc.[94]

There are approximately 150 XKEYSCORE sites around the world. These sites include wiretaps at telecommunications companies' peering sites (such as the AT&T case), systems connected to friendly foreign intelligence agencies' collection sites, and mid-

[93] Ibid.
[94] Ibid.

ocean fiber-optic cable taps (executed by F6, the joint CIA-NSA Special Collections Service).[95] According to XKEYSCORE slides, other NSA collection sites select and forward to PINWALE less than 5% of the data they process. The rest of the data is lost. By comparison, XKEYSCORE can store three days worth of raw packet data and 30 days worth of metadata in local caches. Only information that is related to specific cases is sent back to the NSA's central database. The data in the local caches is available to analysts through federated search while it is being stored.[96]

Figure 12: Location of XKEYSCORE sites

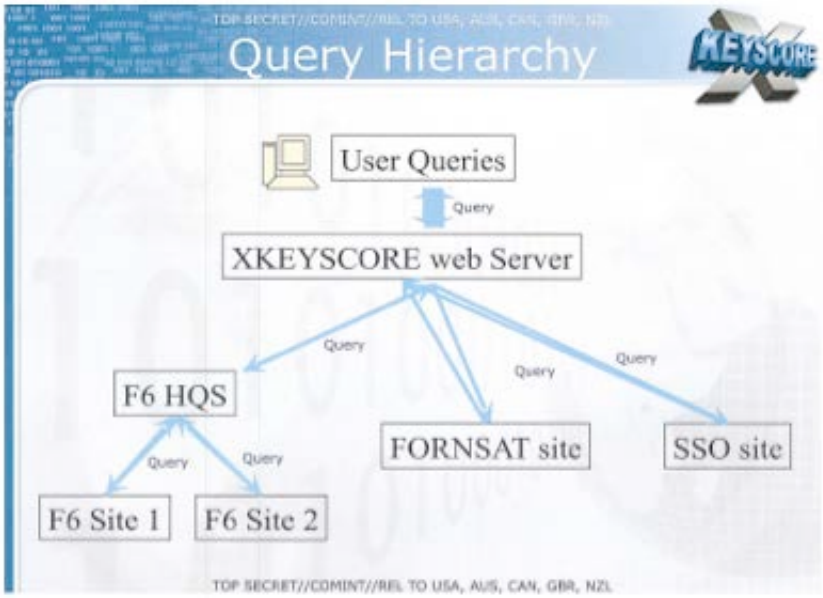

Figure 13: Slide indicating the types of XKEYSCORE sites and the user query being sent to all of them.

[95] Ibid.
[96] Ibid.

To perform a search request an NSA analyst creates a query.[97] This query is sent to all the XKEYSCORE sites. Analysts can search by hard selectors (e.g. email addresses) or soft selectors (e.g. language). So if an analyst does not have a hard selector, such as an email address or phone number for a known target, they can do a search for a category of information, such as all encrypted Word documents or all VPN (virtual private network) startups in a given country.[98] Any kind of query can be created as long as the plugin exists. XKEYSCORE combines and returns all the responses to the query.

Using XKEYSCORE is apparently how the NSA can identify Tor users – this is a category that can be queried. Another category that can be searched is exploitable machines – "Show me all exploitable machines in country X."[99] When the NSA unit, Tailored Access Operations (TAO) identifies a computer as a target it loads its fingerprints, or unique identifiers, into XKEYSCORE's application/ fingerprint ID engine. Whether the search for all exploitable machines refers to all machines that TAO has targeted for exploitation, or those that are vulnerable at that moment to be exploited among those it has targeted, is unclear. In any case, from an NSA presentation, it appears that in the case of Microsoft system crashes XKEYSCORE identifies these error reports and sends an automatic notification, enabling TAO to exploit the machine.[100] For any other exploits that require a certain type of Internet activity or activity that can be identified via Internet traffic, it is likely that XKEYSCORE plays a role.

XKEYSCORE is a powerful tool. It is even possible to search in near real time if the data flow is low past a given tap point. The only factors that limit this analysis tool and database are that the data packets have to pass through one of the NSA's intercept points and that most of the data is only retained for 3 days.[101]

Like Narus, XKEYSCORE performs best when it 'goes shallow,' that is, fewer filters are applied to determine which data packets are captured. This means that a lot of information is collected, including, undoubtedly, information unrelated to NSA targets. There have, however, "been steady improvements to the filter hardware that does the collection for XKEYSCORE."[102] But it is unclear whether these improvements mean that the filters are better targeted to collect only relevant data, or that they are improved to capture more traffic as a whole.[103]

---

[97] The query is done using "an internal Web front-end on the Joint Worldwide Intelligence Communications System (JWICS), the top-secret/ sensitive compartmented information (TS/SCI) network shared by the intelligence community and the Department of Defense." (Gallagher, "Building a Panopticon")

[98] NSA, XKEYSCORE slides, February 25, 2008.

[99] Ibid.

[100] SPIEGEL Staff, "Inside TAO: Documents Reveal Top NSA Hacking Unit," *Spiegel Online*, December 29, 2013, http://www.spiegel.de/international/world/the-nsa-uses-powerful-toolbox-in-effort-to-spy-on-global-networks-a-940969.html.

[101] Gallagher, "Building a Panopticon."

[102] Ibid.

[103] "[T]he SSO deployed a new system in 2012 that it said allows 'more than 75 percent of the traffic to pass through the filter,' according to information from *The Guardian*. That means that the large majority of traffic passing through US telecommunications peering points can be screened based on the rule sets

7 Conclusion

The NSA's surveillance program covers the full range of surveillance from aiming at precise, specific targets to sweeping up Internet flows en masse. In this deliverable we have focused on mass surveillance. Mass surveillance has been greatly enabled by modern technology. This technology – the Internet, laptops, handheld devices, GPS, VoIP – has made it easy for companies providing these services to access the information communicated by their users. The NSA, in turn, has piggybacked on these companies and found ways to obtain copies of this data.

The number of NSA documents nabbed by Edward Snowden – at least 50,000[104] – ensures that the media coverage and stories on the NSA will continue. We believe, however, that any new technology revealed will largely fall under the categories outlined here. That is, although the leak might reveal a new operation or make a sensational headline, the underlying technology will fall under the broad categories of NSA surveillance discussed here – tapping fiber-optic cables, circumventing encryption, launching cyber attacks, gathering phone metadata, and utilizing traditional spying methods such as bugging embassies and tapping political leaders' phones. The article published by Der Spiegel in December 2013 is a case in point. This sensational article unveiled a catalog-of-sorts of spy technology. The hardware and software described is impressive. RAGEMASTER, for example, is a hardware implant that allows the NSA to see whatever appears on the monitor of the targeted computer. SOMBERKNAVE, a software implant, makes the targeted device controllable remotely. These impressive capabilities, however, are targeted surveillance. They are attacks executed against chosen targets.[105] The astounding ability and reach of the NSA via these devices could, and perhaps did, overshadow the fact that this surveillance is targeted and is exactly the kind of activity spy agencies should be engaged in.

In examining NSA surveillance technology as it pertains to mass surveillance we have discovered that some of the technology discussed in the recent leaks in fact is targeted surveillance and other technology falls into somewhat of a grey area between targeted and mass surveillance. Wire-tapping certainly falls under mass surveillance, as does gathering all phone metadata. Cyber attacks and software and hardware implants are deployed against chosen targets. PRISM, also, is to gain information on specific user

---

used for packet capture. Depending on how wide the aperture of those rules are, that could either mean that the NSA is able to "go deep" on 75 percent of traffic and capture just the information they're looking for (with 25 percent of traffic slipping by untouched), or that 75 percent of traffic is getting dumped to cache to be processed—and is searchable with XKEYSCORE while it's sitting there." (Gallagher, "Building a Panopticon")

[104] Dan Murphy, "Listening to Edward Snowden at SXSW," *The Christian Science Monitor*, March 10, 2014, http://www.csmonitor.com/World/Security-Watch/Backchannels/2014/0310/Listening-to-Edward-Snowden-at-SXSW-video.

[105] According to some sources at least one of these spyware systems – NIGHTSTAND – has been used from drones to target certain geographical areas. In this case, the surveillance would obviously be moving from targeted to mass. (*30c3: To Protect And Infect, Part 2*, 2013, http://www.youtube.com/watch?v=b0w36GAyZIA&feature=youtube_gdata_player)

accounts. Defeating or circumventing encryption is more of a grey area. On the one hand, this category is targeted because the NSA is not capable of defeating encryption on a mass scale. However, some of its methods for circumventing encryption, such as implanting backdoors or obtaining keys can give the agency access to large amounts of untargeted information.

It is pretty clear that the NSA has the capability to spy on virtually anyone, anywhere and that it is collecting massive amounts of information. We can expect that the publications revealing whom it has been spying on will continue. It is also possible that there will be further clarification that will show that what was previously leaked was not accurate, as in the case of PRISM. This is less likely to make as big of a splash and therefore will require more digging on the part of the reader to find.