ACLU
AMERICAN CIVIL LIBERTIES UNION

EAVESDROPPING 101: WHAT CAN THE NSA DO?

The recent revelations about illegal eavesdropping on American citizens by the U.S. National Security Agency have raised many questions about just what the agency is doing. Although the facts are just beginning to emerge, information that has come to light about the NSA's activities and capabilities over the years, as well as the recent reporting by the *New York Times* and others, allows us to discern the outlines of what they are likely doing and how they are doing it.

The NSA is not only the world's largest spy agency (far larger than the CIA, for example), but it possesses the most advanced technology for intercepting communications. We know it has long had the ability to focus powerful surveillance capabilities on particular individuals or communications. But the current scandal has indicated two new and significant elements of the agency's eavesdropping:

The NSA has gained direct access to the telecommunications infrastructure through some of America's largest companies
The agency appears to be not only targeting individuals, but also using broad "data mining" systems that allow them to intercept and evaluate the communications of millions of people within the United States.

The ACLU has prepared a map (see page 2) illustrating how all this is believed to work. It shows how the military spying agency has extended its tentacles into much of the U.S. civilian communications infrastructure, including, it appears, the "switches" through which international and some domestic communications are routed, Internet exchange points, individual telephone company central facilities, and Internet Service Providers (ISP). While we cannot be certain about these secretive links, this chart shows a representation of what is, according to recent reports, the most likely picture of what is going on.

CORPORATE BEDFELLOWS

One major new element of the NSA's spying machinery is its ability to tap directly into the major communications switches, routing stations, or access points of the telecommunications system. For example, according to the *New York Times*, the NSA has worked with "the leading companies" in the telecommunications industry to collect communications patterns, and has gained access "to switches that act as gateways" at "some of the main arteries for moving voice and some Internet traffic into and out of the United States."[1]

This new level of direct access apparently includes both some of the gateways through which phone calls are routed, as well as other key nodes through which a large proportion of Internet traffic passes. This new program also recognizes that today's voice and Internet communications systems are increasingly converging, with a rising proportion of even voice phone calls moving to the Internet via VOIP, and parts of the old telephone transmission system being converted to fiber optic cable and used for both data and voice communications. While data and voice sometimes travel together and sometimes do not, and we do not know exactly which "switches" and other access points the NSA has tapped, what appears certain is that the NSA is looking at both.

And most significantly, access to these "switches" and other network hubs give the agency access to a direct feed of all the communications that pass through them, and the ability to filter, sift through, analyze, read, or share those communications as it sees fit.

DATA MINING

The other major novelty in the NSA's activities appears to be the exploitation of a new concept in surveillance that has attracted a lot of attention in the past few years: what is commonly called "data mining." Unlike the agency's longstanding practice of spying on specific individuals and communications based upon some source of suspicion, data mining involves formula-based searches through mountains of data for individuals whose behavior or profile is in some way suspiciously different from the norm.

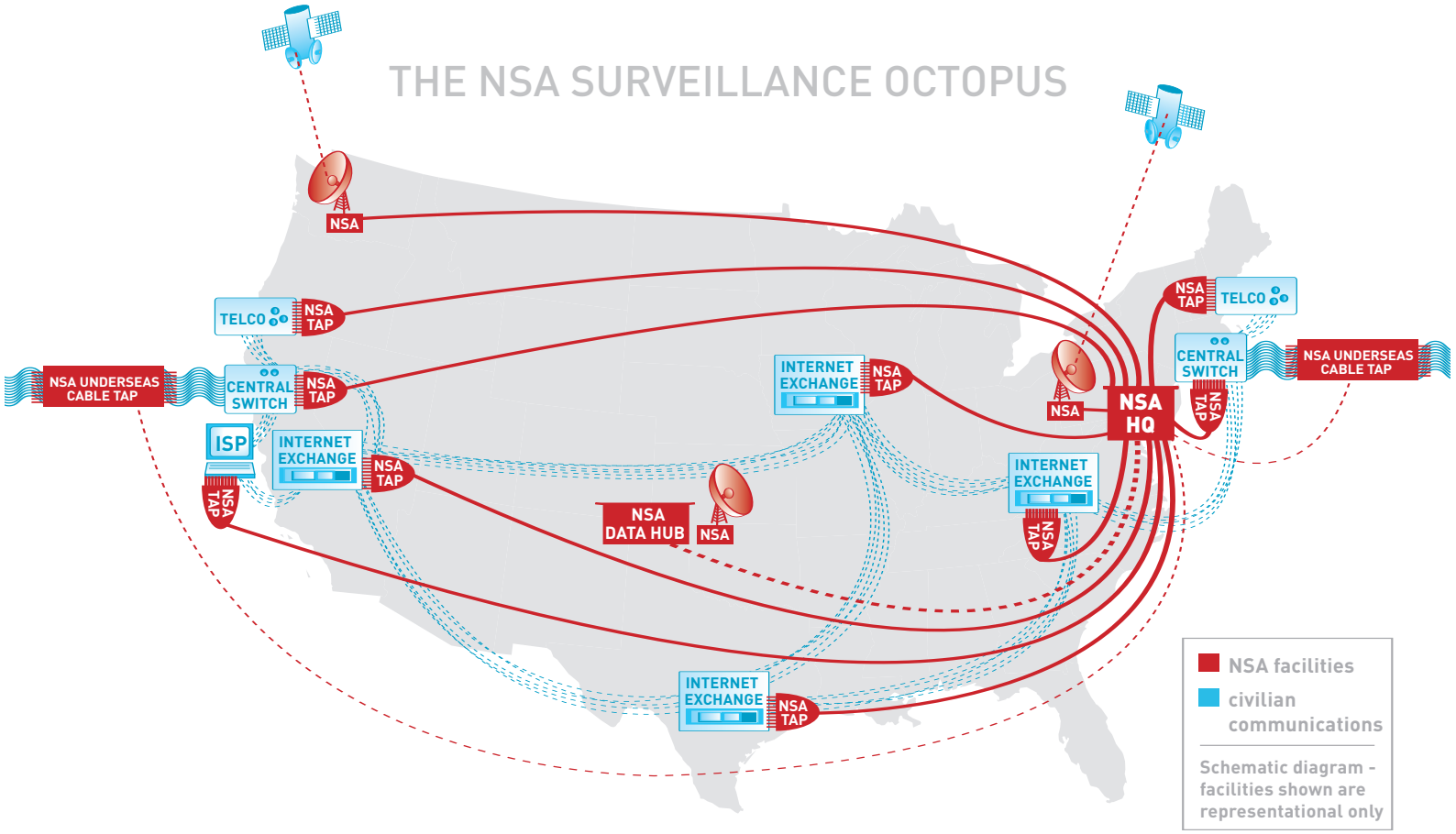
Data mining is a broad dragnet. Instead of targeting you because you once received a telephone call from a person who received a telephone call from a person who is a suspected terrorist, you might be targeted because the NSA's computers have analyzed your communications and have determined that they contain certain words or word combinations, addressing information, or other factors with a frequency that deviates from the average, and which they have decided might be an indication of suspiciousness. The NSA has no prior reason to suspect you, and you are in no way tied to any other suspicious individuals – you have just been plucked out of the crowd by a computer algorithm's analysis of your behavior.

Use of these statistical fishing expeditions has been made possible by the access to communications streams granted by key corporations. The NSA may also be engaging in "geographic targeting," in which they listen in on communications between the United States and a particular foreign country or region. More broadly, data mining has been greatly facilitated by underlying changes in technology that have taken place in the past few years (see page 3).

This dragnet approach is not only bad for civil liberties – it is also a bad use of our scarce security and law enforcement resources. In fact, the creation of large numbers of wasteful and distracting leads is one of the primary reasons that many security experts say data mining and other dragnet strategies are a poor way of preventing crime and terrorism. The *New York Times* confirmed that point, with its report that the NSA has sent the FBI a "flood" of tips generated by mass domestic eavesdropping and data mining, virtually all of which led to dead ends that wasted the FBI's resources. "We'd chase a number, find it's a schoolteacher with no indication they've ever been involved in international terrorism," one former FBI agent told the *Times*. "After you get a thousand numbers and not one is turning up anything, you get some frustration."[2]

COMBINING TELECOMMUNICATIONS AND OTHER PRIVATE DATA?

The NSA has historically been in the business of intercepting and analyzing communications data. One question is whether or not this communications data is being combined with other intimate details about our lives. A few years ago, the Pentagon began work on an breathtaking data mining program called Total Information Awareness, which envisioned programming computers to trawl through an extensive list of information on Americans (including, according to the program's own materials, "Financial, Education, Travel, Medical, Veterinary, Country Entry, Place/Event Entry, Transportation, Housing, Critical Resources, Government, Communications") in the hunt for "suspicious" patterns of activity. Congress decisively



Yakima listening post One way that telephone calls and other communications are sent from the United States to Asia and other destinations is via satellite and microwave transmissions. This NSA satellite facility on a restricted Army firing range in Yakima, Washington sweeps in millions of communications an hour from international communications satellites.

Sugar Grove listening post One way that telephone calls and other communications are sent from the United States to Europe and other destinations is via satellite and microwave transmissions. This NSA satellite facility, located in an isolated valley in Sugar Grove, West Virginia, sweeps in millions of communications an hour from international communications satellites.

ISP

Internet Service Provider (ISP) The NSA may be forcing ISPs to provide it with information in the form of a computer tap (similar to a controversial FBI device dubbed "Carnivore") that scans all the communications that reach that ISP.

CENTRAL SWITCH

Central switch These facilities, one in New York and one in Northern California, are operated by major telecommunications companies. They are a primary means by which a mix of voice and data communications, including those that travel over transoceanic undersea fiber optic cables, are routed ("switched") toward their proper destination. Because they serve as central switching points, they offer the NSA access to a large volume of communications.

INTERNET EXCHANGE

Internet exchange These publicly or privately owned "Internet exchanges" are where Internet traffic is exchanged between the sub-networks that make up the Internet. These public or privately owned facilities are divided into Tier 1, Tier 2, and Tier 3 exchanges. The Tier 1 exchanges, typically located in big cities, are the ones that have national and global reach and are likely to be of most interest to the NSA.

NSA UNDERSEAS CABLE TAP

Underseas cable tap According to published reports, American divers were able to install surveillance devices onto the transoceanic cables that carry phone calls and data across the seas. One of these taps was discovered in 1982, but other devices apparently continued to function undetected. The advent of fiber-optic cables posed challenges for the NSA, but there is no reason to believe that that problem remained unsolved by the agency.

NSA HQ

The NSA's headquarters Tens of thousands of people, including intelligence analysts, linguists and computer professionals, work at this complex in Fort Meade, Maryland outside of Washington, DC. NSA headquarters is where the millions of intercepted communications are processed and analyzed.

TELCO

Telco: Domestic telephone company The NSA is apparently hooking in to U.S. telephone companies, which have not only networks that can be tapped into, but also records of customer communications.

NSA DATA HUB

NSA Data Hub: Domestic Warning Hub and Data Warehouse, Aurora, CO The NSA is reportedly building a massive data storage facility in this Denver suburb, and also operates a reconaissance satellite dish here. This may be where the agency's data mining operations take place. A CIA facility and the military's Northern Command (NORTHCOM) are also located here.

rejected this approach, voting to shut down the program, at least for domestic use – but we know Congress allowed elements of the program to be moved undercover, into the bowels of the Pentagon, while supposedly being restricted to non-Americans. We also know that the NSA is sharing its information with other security services. What we do not know is whether any of information from TIA-like enterprises is being combined with the NSA's communications intercepts.

HOW THE NSA SEARCHES FOR TARGETS

There are a range of techniques that are probably used by the NSA to sift through the sea of communications it steals from the world's cables and airwaves:

Keywords. In this longstanding technique, the agency maintains a watch list or "dictionary" of key words, individuals, telephone numbers and presumably now computer IP addresses. It uses that list to pick out potentially relevant communications from all the data that it gathers. These keywords are often provided to the NSA by other security agencies, and the NSA passes the resulting intelligence "take" back to the other agencies or officials. According to the law, the NSA must strip out the names and other identifying information of Americans captured inadvertently, a process called "minimization." (According to published reports, those minimization procedures are not being properly observed.) In the 1990s, it was revealed that the NSA had used the word "Greenpeace" and "Amnesty" (as in the human rights group Amnesty International) as keywords as part of its "Echelon" program (see below).

Link analysis. It is believed that another manner in which individuals are now being added to the watch lists is through a process often called "link analysis." Link analysis can work like this: the CIA captures a terrorist's computer on the battlefield and finds a list of phone numbers, including some U.S. numbers. The NSA puts those numbers on their watch list. They add the people that are called from those numbers to their list. They could then in turn add the people called from *those* numbers to their list. How far they carry that process and what standards if any govern the process is unknown.

Other screening techniques. There may be other techniques that the NSA could be using to pluck out potential targets. One example is voice pattern analysis, in which computers listen for the sound of, say, Osama Bin Laden's voice. No one knows how accurate the NSA's computers may be at such tasks, but if commercial attempts at analogous activities such as face recognition are any guide, they would also be likely to generate enormous numbers of false hits.

A THREE-STAGE PROCESS

So how are all these new techniques and capabilities being put into practice? Presumably, "The Program" (as insiders reportedly refer to the illegal practices) continues to employ watch lists and dictionaries. We do not know how the newer and more sophisticated link analysis and statistical data mining techniques are being used.

But, a good guess is that the NSA is following a three-stage process for the broadest portion of its sweep through the communications infrastructure:

1. The Dragnet: a search for targets. In this stage, the NSA sifts through the data coursing through the arteries of our telecom systems, making use of such factors as keyword searches, telephone number and IP address targeting, and techniques such as link analysis, and "data mining." At this stage, the communications of millions of people may be scrutinized.

2. Human review: making the target list. Communications and individuals that are flagged by the system for one reason or another are presumably then subject to human review. An analyst looks at the origin, destination and content of the communication and makes a determination as to whether further eavesdropping or investigation is desired. We have absolutely no idea what kind of numbers are involved at this stage.

3. The Microscope: targeting listed individuals. Finally, individuals determined to be suspicious in phase two are presumably placed on a target list so that they are placed under the full scrutiny of the NSA's giant surveillance microscope, with all their communications captured and analyzed.

EXPANDING SURVEILLANCE AS TECHNOLOGY CHANGES

Today's NSA spying is a response to, and has been made possible by, some of the fundamental technological changes that have taken place in recent years. Around the end of 1990s, the NSA began to complain privately – and occasionally publicly – that they were being overrun by technology as communications increasingly went digital. One change in particular was especially significant: electronic communications ranging from email to voice conversations were increasingly using the new and different protocols of the Internet.

The consequence of this change was that the NSA felt it was forced to change the points in the communications infrastructure that it targeted – but having done that, it gained the ability to analyze vastly more and richer communications.

The Internet and technologies that rely upon it (such as electronic mail, web surfing and Internet-based telephones known as Voice over IP or VOIP) works by breaking information into small "packets." Each packet is then routed across the network of computers that make up the Internet according to the most efficient path at that moment, like a driver trying to avoid traffic jams as he makes his way across a city. Once all the packets – which are labeled with their origin, destination and other "header" information – have arrived, they are then reassembled.

An important result of this technology is that on the Internet, there is no longer a meaningful distinction between "domestic" and "international" routes of a communication. It was once relatively easy for the NSA, which by law is limited to "foreign intelligence," to aim its interception technologies at purely "foreign" communications. But now, an e-mail sent from London to Paris, for example, might well be routed through the west coast of the United States (when, for example, it is a busy mid-morning in Europe but the middle of the night in California) along the same path traveled by mail between Los Angeles and San Francisco.

That system makes the NSA all the more eager to get access to centralized Internet exchange points operated by a few telecommunications giants. But because of the way this technology works, eavesdropping on an IP communication is a completely different ballgame from using an old-fashioned "wiretap" on a single line. The packets of interest to the eavesdropper are mixed in with all the other traffic that crosses through that pathway – domestic and international.

ECHELON

Much of what we know about the NSA's spying prior to the recent revelations comes from the late 1990s, when a fair amount of information emerged about a system popularly referred to by the name "Echelon" – a codename the NSA had used at least at one time (although their continued use of the term, if at all, is unknown). Echelon was a system for mass eavesdropping on communications around the world by the NSA and its allies

among the intelligence agencies of other nations. The best source of information on Echelon was two reports commissioned by the European Parliament (in part due to suspicions among Europeans that the NSA was carrying out economic espionage on behalf of American corporations). Other bits of information were gleaned from documents obtained through the U.S. Freedom of Information Act, as well as statements by foreign governments that were partners in the program (the UK, Australia, Canada, and New Zealand).

As of the late 1990s/early 2000s, Echelon swept up global communications using two primary methods:

1. The interception of satellite and microwave signals. One way that telephone calls and other communications are sent from the United States to Europe and other destinations is via satellite and microwave transmissions. ECHELON was known to use numerous satellite receivers ("dishes") – located on the east and west coasts of the United States, in England, Australia, Germany, and elsewhere around the globe – to vacuum up the "spillover" broadcasts from these satellite transmissions.

2.Transoceanic cable tapping. ECHELON's other primary eavesdropping method was to tap into the transoceanic cables that also carry phone calls across the seas. According to published reports, American divers were able to install surveillance devices onto these cables. One of these taps was discovered in 1982, but other devices apparently continued to function undetected. It is more difficult to tap into fiber-optic cables (which unlike other cables do not "leak" radio signals that can be picked up by a device attached to the outside of the cable), but there is no reason to believe that that problem remained unsolved by the agency.

We do not know the extent to which these sources of data continue to be significant for the NSA, or the extent to which they have been superseded by the agency's new direct access to the infrastructure, including the Internet itself, over which both voice and data communications travel.

UNANSWERED QUESTIONS

The bottom line is that the NSA appears to be capable not only of intercepting the international communications of a relatively small number of targeted Americans, but also of intercepting a sweeping amount of U.S. communications (through corporate-granted access to communications "pipes" and "boxes"), and of performing mass analysis on those communications (through data mining and other techniques).

Despite the fuzzy picture of "The Program" that we now possess, the current spying scandal has highlighted many unanswered questions about the NSA's current activities. They include:

- Just what kinds of communications arteries has the NSA tapped into?
- What kinds of filters or analysis is the NSA applying to the data that flows through those arteries? How are data mining and other new techniques are being used?
- Which telecom providers are cooperating with the NSA?
- How are subjects selected for targeted intercepts?
- What kinds of information exchange are taking place between the NSA and other security agencies? We know they probably turn over to other agencies any data turned up by watch list entries submitted by those other agencies, and they are also apparently passing along data mining-generated "cold hits" to the FBI and perhaps other security agencies for further investigation. Does information flow the other way as well – are other agencies giving data to the NSA for help in that second phase of deciding who gets put under the microscope?
- Is data that NSA collects, under whatever rubric, being merged with other data, either by NSA or another agency? Is communications data being merged with other transactional information, such as credit card, travel, and financial data, in the fashion of the infamous "Total Information Awareness" data mining program? (TIA, while prohibited by Congress from engaging in "domestic" activities, still exists within the Pentagon – and can be used for "foreign intelligence purposes.) Just how many schoolteachers and other innocent Americans have been investigated as a result of "The Program?" And just how much privacy invasion are they subject to before the FBI can conclude they are not "involved in international terrorism"?

Rarely if ever in American history has a government agency possessed so much power subject to so little oversight. Given that situation, abuses were inevitable – and any limits to those abuses a matter of mere good fortune. If our generation of leaders and citizens does not rise to the occasion, we will prove ourselves to be unworthy of the heritage that we have been so fortunate to inherit from our Founders.

ENDNOTES

[1] Eric Lichtblau and James Risen, "Spy Agency Mined Vast Data Trove, Officials Report," *New York Times*, December 24, 2005; http://select.nytimes.com/search/restricted/article?res=FA0714F63E540C778EDDAB0994DD404482

[2] Lowell Bergman, Eric Lichtblau, Scott Shane and Don Van Natta Jr., "Spy Agency Data After Sept. 11 Led F.B.I. to Dead Ends," *New York Times*, January 17, 2006; http://www.nytimes.com/2006/01/17/politics/17spy.html.