The Washington Post

TUESDAY, OCTOBER 15, 2013

NSA collects millions of e-mail address books globally

Although interception occurs overseas, it sweeps in many Americans' contact lists

BY BARTON GELLMAN AND ASHKAN SOLTANI

The National Security Agency is harvesting hundreds of millions of contact lists from personal e-mail and instant messaging accounts around the world, many of them belonging to Americans, according to senior intelligence officials and top-secret documents provided by former NSA contractor Edward Snowden.

The collection program, which has not been disclosed before, intercepts e-mail address books and "buddy lists" from instant messaging services as they move across global data links. Online services often transmit those contacts when a user logs on, composes a message, or synchronizes a computer or mobile device with information stored on remote servers.

Rather than targeting individual users, the NSA is gathering contact lists in large numbers that amount to a sizable fraction of the world's e-mail and instant messaging accounts. Analysis of that data enables the agency to search for hidden connections and to map relationships within a much smaller universe of foreign intelligence targets.

During a single day last year, the NSA's Special Source Operations branch collected 444,743 e-mail address books from Yahoo, 105,068 from Hotmail, 82,857 from Facebook, 33,697 from Gmail and 22,881 from unspecified other providers, according to an internal NSA PowerPoint presentation. Those figures, described as a typical daily intake in the document, correspond to a rate of more than 250 million a year.

Each day, the presentation said, the NSA collects contacts from an estimated 500,000 buddy lists on live-chat services as well as from the inbox displays of Web-based e-mail accounts.

The collection depends on secret arrangements with foreign telecommunications companies or allied intelligence services in control of facilities that direct traffic along the Internet's main data routes.

Although the collection takes place overseas, two senior U.S. intelligence officials acknowledged that it sweeps in the contacts of many Americans. They declined to offer an estimate but did not dispute that the number is likely to be in the millions or tens of millions.

A spokesman for the Office of the Director of National Intelligence, which oversees the NSA, said the agency "is focused on discovering and developing intelligence about valid foreign intelligence targets like terrorists, human traffickers and drug smugglers. We are not interested in personal information about ordinary Americans."

The spokesman, Shawn Turner, added that rules approved by the attorney general require the NSA to "minimize the acquisition, use and dissemination" of information that identifies a U.S. citizen or permanent resident.

The NSA's collection of nearly all U.S. call records, under a separate program, has generated significant controversy since it was revealed in June. The NSA's director, Gen. Keith B. Alexander, has defended "bulk" collection as an essential counterterrorism and foreign intelligence tool, saying, "You need the haystack to find the needle."

Contact lists stored online provide the NSA with far richer sources of data than call records alone. Address books commonly include not only names and e-mail addresses, but also telephone numbers, street addresses, and business and family information. Inbox listings of e-mail accounts stored in the "cloud" sometimes contain content, such as the first few lines of a message.

Taken together, the data would enable the NSA, if permitted, to draw detailed maps of a person's life, as told by personal, professional, political and religious connections. The picture can also be misleading, creating false "associations" with ex-spouses or people with whom an account holder has had no contact in many years.

The NSA has not been authorized by Congress or the special intelligence court that oversees foreign surveillance to collect contact lists in bulk, and senior intelligence officials said it would be illegal to do so from facilities in the United States. The agency avoids the restrictions in the Foreign Intelligence Surveillance Act by intercepting contact lists from access points "all over the world," one official said, speaking on the condition of anonymity to discuss the classified program. "None of those are on U.S. territory."

Because of the method employed, the agency is not legally required or technically able to restrict its intake to contact lists belonging to specified foreign intelligence targets, he said.

When information passes through "the overseas collection apparatus," the official added, "the assumption is you're not a U.S. person."

In practice, data from Americans is collected in large volumes — in part because they live and work overseas, but also because data crosses international boundaries even when its American owners stay at home. Large technology companies, including Google and Facebook, maintain data centers around the world to balance loads on their servers and work around outages.

A senior U.S. intelligence official said the privacy of Americans is protected, despite mass collection, because "we have checks and balances built into our tools."

NSA analysts, he said, may not search within the contacts database or distribute information from it unless they can "make the case that something in there is a valid foreign intelligence target in and of itself."

In this program, the NSA is obliged to make that case only to itself or others in the executive branch. With few exceptions, intelligence operations overseas fall solely within the president's legal purview. The Foreign Intelligence Surveillance Act, enacted in 1978, imposes restrictions only on electronic surveillance that targets Americans or takes place on U.S. territory.

By contrast, the NSA draws on authority in the Patriot Act for its bulk collection of domestic phone records, and it gathers online records from U.S. Internet companies, in a program known as PRISM, under powers granted by Congress in the FISA Amendments Act. Those operations are overseen by the Foreign Intelligence Surveillance Court.

Sen. Dianne Feinstein, the California Democrat who chairs the Senate Intelligence Committee, said in August that the committee has less information about, and conducts less oversight of, intelligence gathering that relies solely on presidential authority. She said she planned to ask for more briefings on those programs.

"In general, the committee is far less aware of operations conducted under 12333," said a senior committee staff member, referring to Executive Order 12333, which defines the basic powers and responsibilities of the intelligence agencies. "I believe the NSA would answer questions if we asked them, and if we knew to ask them, but it would not routinely report these things, and, in general, they would not fall within the focus of the committee."

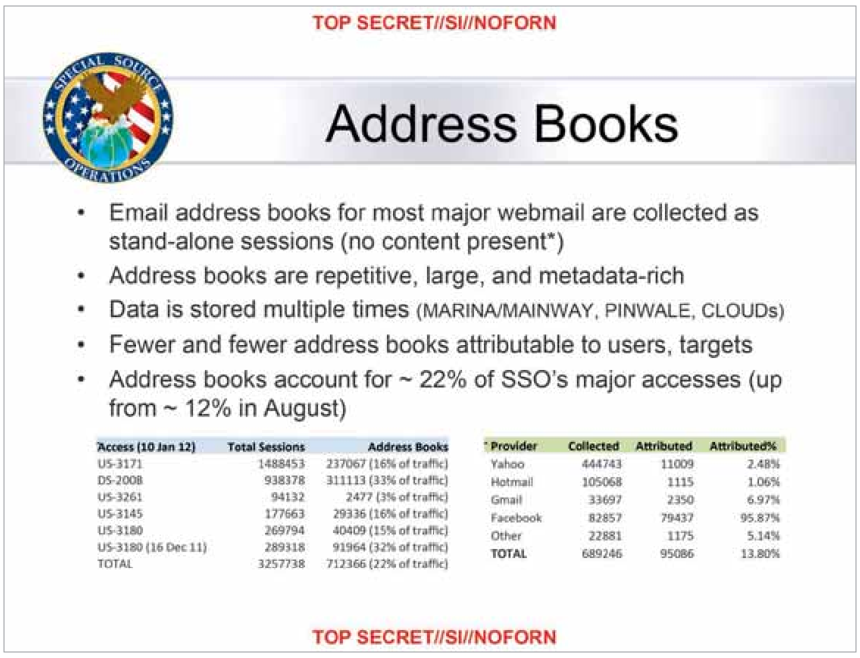
Because the agency captures contact lists "on the fly" as they cross major Internet switches, rather than "at rest" on computer servers, the NSA has no need to notify the U.S. companies that host the information or to ask for help from them.

Problems in the collection of address books

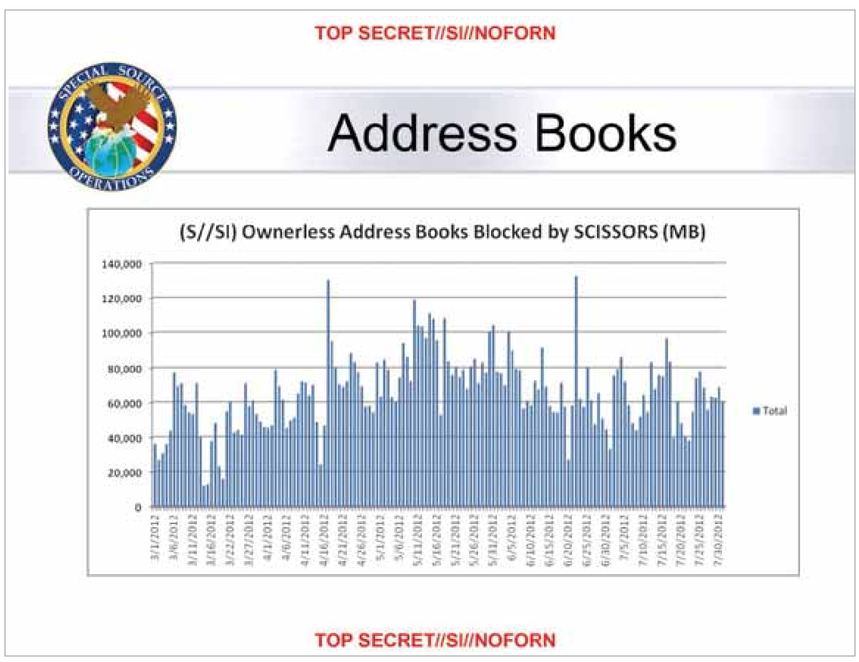
These four slides from a pair of briefings describe problems with overcollection and NSA efforts to filter out what it does not not need. These top-secret documents were provided by former NSA contractor Edward Snowden.

This slide laments a Yahoo Messenger problem that forced the NSA to stop collecting important information about Greece and Libya. A session is another term for a data interchange between two computers, such as when you log into a service or mail is transferred. Selectors are the NSA's term for what it is searching for — such as an e-mail address or phone number. Detasking means that the agency stops collection.

Buddy lists sometimes include the text of messages waiting to be delivered, which count as content. Web mail inboxes, which list new messages, often include a line or two of the text. When the NSA searches for a specific target, it usually finds only a listing in someone else's address book. More-valuable finds — the target's address book, a person communicating with the target or a message mentioning the target — are rarer.



This slide lays out the number of **contact lists collected** on a single day, Jan. 10, 2012, from the six top overseas access points. The "US" prefix denotes an NSA access point, and "DS" refers to the NSA's Australian counterpart. Address books make up an unexpectedly large share of information pulled in by the NSA. Many of them are less useful because they are "unattributed," with the owners unknown.



SCISSORS is an NSA system that helps parse electronic communications. For "ownerless" address books, which the NSA cannot attribute to a specific account holder, SCISSORS tries to block collection of content. This chart shows how often that happened in mid-summer 2012.

"We have neither knowledge of nor participation in this mass collection of web-mail addresses or chat lists by the government," said Google spokeswoman Niki Fenwick.

At Microsoft, spokeswoman Nicole Miller said the company "does not provide any government with direct or unfettered access to our customers' data," adding that "we would have significant concerns if these allegations about government actions are true."

Facebook spokeswoman Jodi Seth said that "we did not know and did not assist" in the NSA's interception of contact lists.

It is unclear why the NSA collects more than twice as many address books from Yahoo than the other big services combined. One possibility is that Yahoo, unlike other service providers, has left connections to its users unencrypted by default.

Suzanne Philion, a Yahoo spokeswoman, said Monday in response to an inquiry from The Washington Post that, beginning in January, Yahoo would begin encrypting all its e-mail connections.

Google was the first to secure all its e-mail connections, turning on "SSL encryption" globally in 2010. People with inside knowledge said the move was intended in part to thwart large-scale collection of its users' information by the NSA and other intelligence agencies.

The volume of NSA contacts collection is so high that it has occasionally threatened to overwhelm storage repositories, forcing the agency to halt its intake with "emergency detasking" orders. Three NSA documents describe short-term efforts to build an "across-the-board technology throttle for truly heinous data" and longer-term efforts to filter out information that the NSA does not need.

Spam has proven to be a significant problem for the NSA — clogging databases with information that holds no foreign intelligence value. The majority of all e-mails, one NSA document says, "are SPAM from 'fake' addresses and never 'delivered' to targets."

In fall 2011, according to an NSA presentation, the Yahoo account of an Iranian target was "hacked by an unknown actor," who used it to send spam. The Iranian had "a number of Yahoo groups in his/her contact list, some with many hundreds or thousands of members."

The cascading effects of repeated spam messages, compounded by the automatic addition of the Iranian's contacts to other people's address books, led to a massive spike in the volume of traffic collected by the Australian intelligence service on the NSA's behalf.

After nine days of data-bombing, the Iranian's contact book and contact books for several people within it were "emergency detasked."

In a briefing from the NSA's Large Access Exploitation working group, that example was used to illustrate the need to narrow the criteria for data interception. It called for a "shifting collection philosophy": "Memorialize what you need" vs. "Order one of everything off the menu and eat what you want."

bart.gellman@washpost.com

Julie Tate contributed to this report.
Soltani is an independent security researcher and consultant.