Jonathan Mayer
Stanford Security Laboratory
Department of Computer Science
353 Serra Mall MC 9045
Stanford, CA 94305

October 3, 2013

Review Group on Intelligence and Communications Technologies
Office of the Director of National Intelligence

Re: Internet Surveillance Under Section 702 of the FISA Amendments Act

Dear Review Group:

According to recent disclosures, the National Security Agency has the capability and purported authorization to conduct bulk acquisition of Internet traffic.[1] Coverage has emphasized that this "upstream" surveillance is directed at communications where at least one end is located outside the United States. The NSA's focus has both statutory and constitutional dimensions: Section 702 of the FISA Amendments Act, the alleged authorization, concerns extraterritorial non-United States persons.[2] Programs under Section 702 must also comport with the well-established Fourth Amendment protection for Americans' reasonable expectations of privacy.

---

[1] *See, e.g.*, [Redacted], No. [Redacted] (FISA Ct., Oct. 3, 2011) (describing an upstream collection program), *available at* http://icontherecord.tumblr.com/post/58944252298/dni-declassifies-intelligence-community-documents; Exhibit A, [Redacted], No. [Redacted] (FISA Ct., July 29, 2009) (establishing targeting criteria for Section 702 upstream surveillance), *available at* http://www.theguardian.com/world/interactive/2013/jun/20/exhibit-a-procedures-nsa-document; Glenn Greenwald & James Ball, *The Top Secret Rules That Allow NSA to Use US Data Without a Warrant*, THE GUARDIAN, June 20, 2013, *available at* http://www.theguardian.com/world/2013/jun/20/fisa-court-nsa-without-warrant (explaining Section 702 targeting procedures); Glenn Greenwald & Spencer Ackerman, *NSA Collected US Email Records in Bulk for More than Two Years Under Obama*, THE GUARDIAN, June 27, 2013, *available at* http://www.theguardian.com/world/2013/jun/27/nsa-data-mining-authorised-obama (describing Internet metadata program ended in 2011); Spencer Ackerman, *How the NSA Is Still Harvesting Your Online Data*, THE GUARDIAN, June 27, 2013, *available at* http://www.theguardian.com/world/2013/jun/27/nsa-online-metadata-collection (detailing ongoing bulk Internet metadata program); Charlie Savage, *N.S.A. Said to Search Content of Messages to and from U.S.*, N.Y. TIMES, Aug. 8, 2013, at A1 (describing searches of international Internet traffic based on their contents); Siobhan Gorman & Jennifer Valentino-Devries, *New Details Show Broader NSA Surveillance Reach*, WALL. ST. J., Aug. 20, 2013, at A1 (explaining filtering associated with domestic bulk NSA Internet); Jennifer Valentino-Devries, *What You Need To Know on New Details of NSA Spying*, WALL ST. J., Aug. 20, 2013, *available at* http://online.wsj.com/article/SB10001424127887324108204579025222244858490.html (same in more detail); James Risen & Laura Poitras, *N.S.A. Gathers Data on Social Connections of U.S. Citizens*, N.Y. TIMES, Sept. 28, 2013, at A1 (recounting NSA use of data concerning U.S. persons).

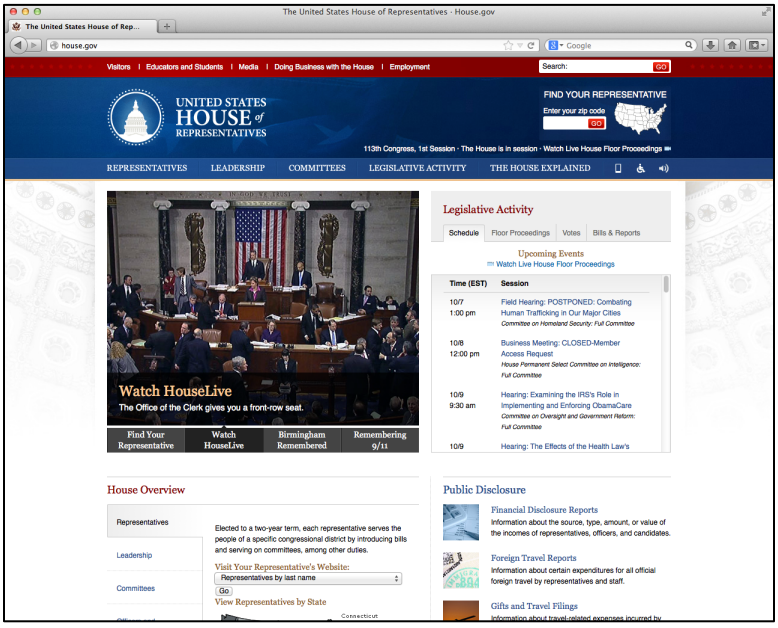
[2] 50 U.S.C. § 1881a(b), (d)(1) (restricting targeting criteria).

I write to explain how, as a technical matter, dragnet acquisition of "one-end foreign" Internet traffic necessarily sweeps in vast amounts of routine activity by ordinary Americans. This recognition raises substantial policy concerns and presses at the legal status of the NSA's Internet surveillance programs.

In order to illustrate the issue, I conducted an analysis of websites that are popular among U.S. users.[3] The research results reflect three problematic phenomena.

First, many domestic websites embed content from international websites.[4] In the design of the web, embedded content is generally notified of its context.[5] Consequently, in instances where an American reasonably expects to interact with a domestic website—and is, in fact, interacting with a domestic website—his or her browsing activity may nevertheless flow across international boundaries.[6]

For an in-depth example, consider the website for the United States House of Representatives, http://house.gov/.



[3] Specifically, I conducted a crawl of the Quantcast U.S. top 2,500 websites using the FourthParty measurement platform. The crawler followed five links from each homepage and spent fifteen seconds on each page to allow dynamic content to load. Once the crawl was complete, I identified leakage of browsing activity by searching HTTP `Request-URI` and `Referer` headers for the URLs of pages visited during the crawl. I then used the MaxMind GeoLite Country database to identify leak recipients potentially located outside the United States. Finally, I confirmed the location of servers by manually inspecting the output of the `traceroute` utility.

[4] For an overview of third-party web content, see Jonathan R. Mayer & John C. Mitchell, *Third-Party Web Tracking: Policy and Technology*, PROC. 2012 IEEE SYMP. ON SECURITY 413 (2012).

[5] Most commonly, embedded content receives an HTTP referrer header that specifies the page it is embedded in. Some sites use alternative approaches, such as passing context in a URL parameter.

[6] The study in this comment centers on one-end foreign Internet traffic. Similar issues may arise for purely domestic Internet traffic that happens to travel an international route.

Based on the website's federal provenance, .gov domain, and governmental content, a visitor might understandably conclude that the site is hosted within the United States. In fact, the `traceroute` utility indicates that the site is served from the Washington, D.C. area.

But the site is not entirely domestic. In order to assist visitors who have difficulty reading, the House website embeds a read-aloud widget from Texthelp, a business that is incorporated in the United Kingdom. In the course of loading a page on the main House of Representatives website, an American user's browser will contact a webserver at `babm.texthelp.com`, which appears to be located in London. The HTTP request will take a form analogous to:

```
GET /Detect.ashx HTTP/1.1
Host: babm.texthelp.com
. . .
Origin: http://house.gov
. . .
Referer: http://house.gov/legislative/date/2013-10-4
. . .
```

In this example, the HTTP referrer header conveys the context for the read-aloud widget: the user is reading the House of Representatives legislative calendar for October 4, 2013. Thus, merely by acquiring one-end foreign traffic, the NSA would possess information about what Americans are reading on the House of Representatives website.

International leakage of domestic browsing activity is remarkably common. The analysis I conducted indicated international leaks across many categories of domestic websites, including political commentary (e.g. National Review and Talking Points Memo), popular culture (e.g. Buzzfeed and Parade), sports (e.g. Major League Baseball and the PGA Tour), travel (e.g. Lonely Planet), consumer products (e.g. Nike), retail (e.g. Overstock), and health (e.g. Medicare.gov). I do not mean to single out these sites for opprobrium—in my view, there is no inherent error in embedding an international advertisement, analytics service, functional widget, or any other sort of content. My point is solely that, owing to the web's architecture, seemingly domestic browsing activity will often transit national boundaries. The NSA's reported acquisition of one-end foreign Internet traffic, therefore, necessarily includes the browsing activity of millions of Americans on domestic websites. Collection on that scale may run afoul of Section 702's targeting requirements as well as transgress the Fourth Amendment's protection of reasonable privacy expectations.

The second problem reflected in the study is that many websites that are popular among U.S. users are themselves hosted abroad. Some sites are unambiguously international, such as the widely read news outlets The Guardian and Der Spiegel. Other sites may be less overtly foreign, such as the music streaming service Spotify, the dating website Plenty of Fish, the scheduling service Doodle, the Reuters news blogs, or the link shortener is.gd. For policy and legal reasons, domestic users may merit protection in their frequent interactions with these foreign websites.

The third problem apparent in the course of the study is that geolocation based on Internet Protocol addresses can be quite imprecise. The database that I used consistently erred in locating,

for example, Facebook in Ireland and Amazon in the Netherlands. Computer science researchers have catalogued shortcomings in geolocation in great depth.[7] If the NSA were to mistakenly classify domestic servers as outside the United States, even at low rates, it would acquire a substantial amount of purely domestic Internet traffic.

My hope is that the Review Group will inquire into the three problems encountered in the study. What are the NSA's procedures for American activity online that is one-end foreign owing to inadvertent international leakage? Where a domestic user has inadvertently used an international site? Where a domestic user has intentionally used an international site? Has the Foreign Intelligence Surveillance Court reviewed and authorized these sorts of acquisitions? Finally, has the NSA carefully audited its geolocation information to ensure it is not acquiring domestic communications?

I would be pleased to respond to any legal or technical questions that the Review Group may have.

Sincerely,

Jonathan Mayer

[7] *See, e.g.*, Ingmar Poese et al., *IP Geolocation Databases: Unreliable?*, ACM SIGCOMM COMPUTER COMM. REV., Apr. 2011, at 53, 56 (reporting 2-4% error rates in country geolocation with commercial databases).