

PCLOB ESTIMATES 120 MILLION PHONE NUMBERS IN CORPORATE STORE

PCLOB's report confirms something ACLU's Patrick Toomey and I have been harping on. One of the biggest risks of the phone dragnet stems not from the initial queries themselves, but from NSA's storage of query results in the "corporate store," permanently, where they can be accessed without the restrictions required for access to the full database, and exposed to all the rest of NSA's neat toys.

According to the FISA court's orders, records that have been moved into the corporate store may be searched by authorized personnel "for valid foreign intelligence purposes, without the requirement that those searches use only RAS-approved selection terms."⁷¹ Analysts therefore can query the records in the corporate store with terms that are not reasonably suspected of association with terrorism. They also are permitted to analyze records in the corporate store through means other than individual contact-chaining queries that begin with a single selection term: because the records in the corporate store all stem from RAS-approved queries, the agency is allowed to apply other analytic methods and techniques to the query results.⁷² For instance, such calling records may be integrated with data acquired under other authorities for further analysis. The FISA court's orders expressly state that the NSA may apply "the full range" of signals intelligence analytic tradecraft to the calling records that are responsive to a query, which includes every record in the corporate store.⁷³

PCLoB doesn't say it, but NSA's SID Director Theresa Shea has: those other authorities include content collection, which means coming up in a query can lead directly to someone reading your content.

Section 215 bulk telephony metadata complements other counterterrorist-related collection sources by serving as a significant enabler for NSA intelligence analysis. It assists the NSA in applying limited linguistic resources available to the counterterrorism mission against links that have the highest probability of connection to terrorist targets. Put another way, while Section 215 does not contain content, analysis of the Section 215 metadata can help the NSA prioritize for content analysis communications of non-U.S. persons which it acquires under other authorities. Such persons are of heightened interest if they are in a communication network with persons located in the U.S. Thus, Section 215 metadata can provide the means for steering and applying content analysis so that the U.S. Government gains the best possible understanding of terrorist target actions and intentions. [my emphasis]

Plus, those authorities will include datamining, including with other data collected by NSA, like a user's Internet habits and financial records.

Then, PCLoB does some math to estimate how many numbers might be in the corporate store.

If a seed number has seventy-five direct contacts, for instance, and each of these first-hop contact has seventy-five new contacts of its own, then each query would provide the government with the complete calling records of 5,625 telephone numbers. And if each of those second-hop numbers has seventy-five new

contacts of its own, a single query would result in a batch of calling records involving over 420,000 telephone numbers.

[snip]

If the NSA queries around 300 seed numbers a year, as it did in 2012, then based on the estimates provided earlier about the number of records produced in response to a single query, the corporate store would contain records involving over 120 million telephone numbers.⁷⁴

⁷⁴ While fewer than 300 identifiers were used to query the call detail records in 2012, that number “has varied over the years.” Shea Decl. ¶ 24.

Some might quibble with these numbers: other estimates use 40 contacts per person (though remember, there’s 5 years of data), and the estimate doesn’t seem to account for mutual contacts. Plus, remember this is unique phone numbers: we should expect it to include fewer people, because people – especially people trying to hide – change phones regularly. Further, remember a whole lot of foreign numbers will be in there.

But other things suggest it might be conservative. As a recent Stanford study showed, if the NSA isn’t really diligent about removing high volume numbers, then queries could quickly include everyone; certainly, NSA could have deliberately populated the corporate store by leaving such identifiers in. We know there were 27,000 people cleared for RAS in 2008 and 17,000 on an alert list in 2009, meaning the query numbers for earlier years are effectively much much higher (which seems to be the point of footnote 74).

Plus, remember that PCLOB gave their descriptive sections to the NSA to review for accuracy. So I assume NSA did not object to the estimate.

So 120 million phone numbers might be a reasonable estimate.

That's a lot of Americans exposed to the level of data analysis permissible in the corporate store.