NSA PRIVACY OFFICER REBECCA RICHARDS EXPLAINS WHAT CONNECTION CHAINING IS!

Update: I checked with the FBI on whether they were going to do a similar privacy report. After checking around, a spokesperson said, "We are not aware of our folks preparing any such similar public report."

You'll recall that for the year and a half that Congress was percolating over USA Freedom Act, I was trying to figure out what "connection chaining" was, but no one knew or would say?

The description of phone dragnet hops as "connections" rather than calls showed up in early versions of the bill and in dragnet orders since 2014. Ultimately, the final bill used language to describe hops that was even less explanatory, as all it requires is a session identifier connection (which could include things like cookies), without any call or text exchanged.

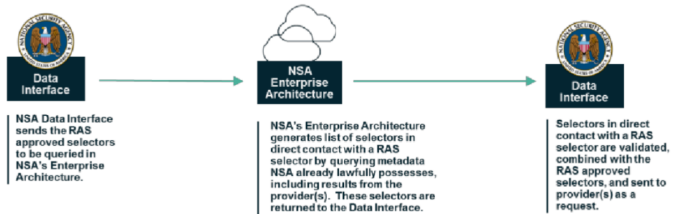
> (iii) provide that the Government may require the prompt production of a first set of call detail records using the specific selection term that satisfies the standard required under subsection (b)(2)(C)(ii);
>
> (iv) provide that the Government may require the prompt production of a second set of call detail records using session-identifying information or a telephone calling card number identified by the specific selection term used to produce call detail records under clause (iii);

In documents released yesterday, NSA's Privacy Officer Rebecca Richards has offered the first explanation of what that chaining process looks like. NSA's Civil Liberties and Privacy Office released a privacy report and minimization procedures on USAF.

Curiously, the privacy report doesn't describe two hops of provider data, though that's meaningless, as the queries will automatically repeat "periodically" (described as daily in the bill), so the government would obtain a second hop from providers by the second day at the latest. Rather, it describes a first hop as occurring within NSA's Enterprise Architecture, and the results of that query to be sent to providers for a second hop.

> Collection: The FISC-approved specific selection term, along with any one-hop results generated from metadata NSA already lawfully possesses from previous results returned from the provider(s) and other authorities, will be submitted to the authorized provider(s). The provider(s) will return CDRs that are responsive to the request, meaning the results will consist of CDRs that are within one or two hops of a FISC-approved specific selection term. This step will be repeated periodically for the duration of the order to capture any new, responsive CDRs but in no case will the procedures generate third or further hops from a FISC-approved specific selection term.

Here's the key part of the picture included to describe the NSA hop that precedes the provider hop.



NSA Data Interface sends the RAS approved selectors to be queried in NSA's Enterprise Architecture.

NSA's Enterprise Architecture generates list of selectors in direct contact with a RAS selector by querying metadata NSA already lawfully possesses, including results from the provider(s). These selectors are returned to the Data Interface.

Selectors in direct contact with a RAS selector are validated, combined with the RAS approved selectors, and sent to provider(s) as a request.

The report is laudable for its very existence (I'm pestering FBI to see if we'll get one from them) and for its willingness to use real NSA terms like "Enterprise Architecture." It is coy in other ways, such as the full role of the FBI, the type of records queried, and — especially — the type of providers included; for the latter, the report cites page 17 of the House report, which only describes providers in this paragraph, using terms — phone company and telecommunications carrier — that are ambiguous and undefined (though someone like Apple could launch a nice lawsuit on the latter term, especially given that they are refusing to provide a back door in a case in EDNY based on the claim they're not a carrier).

> The government may require the production of up to two ''hops''—i.e., the call detail records associated with the initial seed telephone number and call detail records (CDRs) associated with the CDRs identified in an initial ''hop.'' Subparagraph (F)(iii) provides that the government can obtain the first set of CDRs using the specific selection term approved by the FISC. In addition, the government can use the FISC-approved specific selection term to identify CDRs from metadata it already lawfully possesses. Together, the CDRs produced by the phone companies and those identified independently by the government constitute the first ''hop.'' Under subparagraph (F)(iv), the government can then present session identifying information or calling card numbers (which are components of a CDR, as defined in section 107) identified in the first ''hop'' CDRs to *phone companies* to serve as the basis for companies to return the second ''hop'' of CDRs. As with the first ''hop,'' a second ''hop'' cannot be based on, nor return, cell site or GPS location information. It also does not include an individual listed in a telephone contact

> list, or on a personal device that uses the same wireless router as the seed, or that has similar calling patterns as the seed. Nor does it exist merely because a personal device has been in the proximity of another personal device. These types of information are not maintained by *telecommunications carriers* in the normal course of business and, regardless, are prohibited under the definition of ''call detail records.'' [my emphasis]

That said, we know the term provider must be understood fairly broadly given the expanded number of providers who will be included in this program.

What this means, in effect, is that NSA and FBI (the latter does the actual application) will get a specific identifier — which could be a phone number, a SIM card number, a handset identifier, or a credit card [correction: this should be "calling card"], among other things — approved at the FISC, then go back to at least NSA's data (and quite possibly FBI's), and find all the contacts with something deemed to "be" that identifier that would be meaningful for a "phone company" to query their own records with, up to and including a cookie (which is, by definition, a session identifier).

Even in the report's description of this process, there's some slippage in the NSA query step, from an initial RAS approved phone number (202) 555-1234 to an NSA identified number from the (202) area code not provided, making an additional call.

> To illustrate the process, assume an NSA intelligence analyst identifies or learns that phone number (202) 555-1234 is being used by a suspected international terrorist. This is the "specific selection term" or "selector" that will be submitted to the FISC (or the Attorney General in an emergency)

> for approval using the RAS standard. Also assume that, through NSA's examination of metadata produced by the provider(s) or in NSA's possession as a result of the Agency's otherwise lawfully permitted signals intelligence activities (e.g., activities conducted pursuant to Section 1.7(c)(1) of Executive Order 12333, as amended), NSA determines that the suspected terrorist has used a 202 area code phone number to call (301) 555-4321. The phone number with the 301 area code is a "first-hop" result. In turn, assume that further analysis or production from the provider(s) reveals (301) 555-4321 was used to call (410) 555-5678. The number with the 410 area code is a "second-hop" result.

And in this part of the report, the provider query will return any session identifier that includes the selection terms (though elsewhere the report implies only contacts will be returned).

> Once the one-hop results are retrieved from the NSA's internal holdings, the list of FISC-approved specific selection terms, along with NSA's internal one-hop results, are submitted to the provider(s). The provider(s) respond to the request based on the data within their holdings with CDRs that contain FISC-approved specific selection terms or the one-hop selection term. One-hop returns from providers are placed in NSA's holdings and become part of subsequent query requests, which are executed on a periodic basis.

Described in this way, the query process sounds a lot more like what the version of the bill I dubbed USA Freedumber authorized than what the language of USA F-ReDux authorized: two steps of provider queries based off the connected selectors identified at NSA.

> (iii) provide that the Government may require the prompt production of call detail records—
>
> (I) using the specific selection term that satisfies the standard required under subsection (b)(2)(C)(ii) as the basis for production; and
>
> (II) using call detail records with a direct connection to such specific selection term as the basis for production of a second set of call detail records;

Given the breathtaking variety of selector types the NSA uses, this could represent a great deal of queries on the provider side, many tracking user activity rather than user communications. And, at least given how the privacy report describes the transparency reporting, neither those interim NSA selectors nor cookies showing user activity but not communication of information would get counted in transparency reports.

> The number of targets under each order: Defined as the person using the selector. For example, if a target has a set of four selectors that have been approved, NSA will count one target, not four. Alternatively, if two targets are using one selector that has been approved, NSA will count two targets.
>
> The number of unique identifiers used to communicate information collected pursuant to an order: Defined as each unique record sent back from the provider(s).

This approach seems to solve a problem the NSA appears to have been having since 2009, how to query entirely domestic records with identifiers that have been algorithmically determined to be used by the same person. Here, the NSA will be able to match connected selectors to an approved one, and then send all of them to providers to obtain entirely domestic records.

But if I’m right in my reading of this, it leaves one hole in the privacy analysis of the this report.

Richards measures USAF, as she has other programs, against the Fair Information Practice Principles, which include a measure of Data Quality and Integrity. But the report’s analysis of that in this program completely ignores how central NSA’s own data is in the process.

> Each CDR is a business record generated by a provider for the provider□’s own business use. NSA plays no role in ensuring that the provider-generated CDRs accurately reflect the calling events that occurred over the provider’s infrastructure, but the provider(s) have their own policies, practices, and incentives for ensuring the accuracy of their records˙. NSA’s requirements for ensuring accurate, relevant, timely, and complete CDRs begin when NSA submits query requests to the provider(s), and the provider(s), in response, produce CDRs to the Agency.

At least given the description laid out throughout this report, that’s entirely wrong! NSA is centrally involved in getting from the initial selector to the selectors submitted to the providers for query. So if the NSA’s analysis, which as described may include algorithmic matching of records, is inaccurate (say, by matching burner phones inaccurately), than the provider query will return the phone and other records of completely unassociated individuals. I can’t see any way that the NSA’s own query can be exempted from accuracy review here, but it has been.

I absolutely assume NSA is confident in its analysis, but to just dismiss it as uninvolved when it precedes the provider query ignores the implementation architecture laid out in this report.

In any case, I’m grateful we’ve got this report (I may have more to say on the minimization procedures, but they, like the report, are far clearer than the ones included in the old dragnet and for Section 702, perhaps because of the involvement of a Privacy Officer). I’m still thinking through the privacy implications of this. But really, this querying process should have been revealed from the start.