WHY DID GOOGLE MISS A LOT OF USERS AFFECTED BY FISA?

There's been some bad news in the transparency reports issued by America's tech companies thus far. First, Apple revealed a huge spike in FISA requests.

> the number of national security orders, including secret rulings from the Foreign Intelligence Surveillance Court, spiked during the period.
>
> The company received between 13,250 and 13,499 national security orders, affecting between 9,000 and 9,249 accounts.
>
> That's a threefold increase compared to the year earlier, which saw up to 2,999 orders for the period.
>
> It's the largest number of national security orders that Apple has ever reported in five years of publishing transparency reports.

My guess is this reflects increasing reliance on requests to Apple to obtain information that would otherwise be encrypted (it might even suggest Apple was forced to put a back door into their phones, though there has been no declassified FISC opinion that would reflect that, so I doubt that's it). I'm wondering, because of the change Apple just made in iOS 11 that requires passwords before a phone trusts a computer, whether Apple has been asked to turn over backups of iPhones shared to iTunes, but that's admittedly a wildarseguess.

Then, in addition to an new high in standard government information requests, Google also revised its previously issued national security request numbers to reflect (on the most part) significantly more users and/or accounts

affected (CNet reported this here).

Reporting period	Number of requests	Users/Accounts (ORIGINAL)	Users/Accounts (REVISED)
Jan 2017 - Jun 2017	Data subject to six month reporting delay	Data subject to six month reporting delay	Data subject to six month reporting delay
Jul 2016 - Dec 2016	500 - 999	35000 - 35499	35000 - 35499
Jan 2016 - Jun 2016	500 - 999	18500 - 18999	25000 - 25499
Jul 2015 - Dec 2015	500 - 999	21000 - 21499	22500 - 22999
Jan 2015 - Jun 2015	500 - 999	16000 - 16499	19000 - 19499
Jul 2014 - Dec 2014	500 - 999	17500 - 17999	18500 - 18999
Jan 2014 - Jun 2014	500 - 999	15000 - 15499	17000 - 17499
Jul 2013 - Dec 2013	500 - 999	15500 -15999	15500 - 15999
Jan 2013 - Jun 2013	500 - 999	9500 - 9999	14000 - 14499
Jul 2012 - Dec 2012	500 - 999	12500 - 12999	12500 - 12999
Jan 2012 - Jun 2012	500 - 999	8000 - 8499	10500 - 10999
Jul 2011 - Dec 2011	500 - 999	9500 - 9999	10000 - 10499
Jan 2011 - Jun 2011	500 - 999	7000 - 7499	7000 - 7499
Jul 2010 - Dec 2010	0 - 499	5000 - 5499	4000 - 4499
Jan 2010 - Jun 2010	0 - 499	3500 - 3999	3500 - 3999
Jul 2009 - Dec 2009	0 - 499	3500 - 3999	3500 - 3999
Jan 2009 - Jun 2009	0 - 499	2000 - 2499	2000 - 2499

At first I thought this might reflect either the two-year delayed reporting on new services being requested or delayed collection off an original target (which might happen if someone commented, four years later, on a YouTube video posted by an account being tasked). And while some combination of those might be involved, Google claims this was an inadvertent undercounting

> We've also posted updated figures for the number of users/accounts impacted by Foreign Intelligence Surveillance Act (FISA) requests for content in previous reporting periods. While the total number of FISA content requests was reported accurately, we inadvertently under-reported the user/account figures in some reporting periods and over-reported the user/account figures in the second half of 2010. The corrected figures are in the latest report and reflected on our visible changes page. [my emphasis]

Which suggests it may instead pertain to uncertainty — on the part of the government, especially — of which selectors relate to a natural person.

As I have noted, in the government's own

transparency reporting, they provide estimated numbers of targets for both 702 and traditional FISA. The reason they can only provide estimates is almost certainly because for both authorities (and for much of NSA's 12333 targeting) they're targeting selectors of interest, only some of which they've tied to a known person's identity. And it's likely they have selectors that are interesting because of their contacts and other behaviors that belong to already known targets using other selectors.

I provided some background on why this is the case in this post on changes in the reporting provisions the 2015 version of USA Freedom Act.

> First, the reporting provisions as a whole move from tracking "individuals whose communications were collected" to "unique identifiers used to communicate information." They probably did that because they don't really have a handle on which of the identifiers all represent the same natural person (and some aren't natural persons), and don't plan on ever getting a handle on that number. Under last year's bill, ONDI could certify to Congress that he couldn't count that number (and then as an interim measure I understand they were going to let them do that, but require a deadline on when they would be able to count it). Now, they've eliminated such certification for all but 702 metadata back door searches (that certification will apply exclusively to CIA, since FBI is exempted). In other words, part of this is just an admission that ODNI does not know and does not planning on knowing how many of the identifiers they target actually fit together to individual targets.
>
> But since they're breaking things out into identifiers now, I suspect they're unwilling to give that number because

> for each of the 93,000 targets they're currently collecting on, they're probably collecting on at least 10 unique identifiers and probably usually far, far more.
>
> Just as an example (this is an inapt case because Hassanshahi, as a US person, could not be a PRISM target, but it does show the bare minimum of what a PRISM target would get), the two reports Google provided in response to administrative subpoenas for information on Shantia Hassanshahi, the guy caught using the DEA phone dragnet (these were subpoenas almost certainly used to parallel construct data obtained from the DEA phone dragnet and PRISM targeted at the Iranian, "Sheikhi," they found him through), included:
>
> - *a primary gmail account*
> - *two secondary gmail accounts*
> - *a second name tied to one of those gmail accounts*
> - *a backup email (Yahoo) address*
> - *a backup phone (unknown provider) account*
> - *Google phone number*
> - *Google SMS number*
> - *a primary login IP*
> - *4 other IP logins they were tracking*
> - *3 credit card accounts*
> - *Respectively 40, 5, and 11 Google services tied to the primary and two secondary Google*

> *accounts, much of which would be treated as separate, correlated identifiers*
>
> So just for this person who might be targeted under the new phone dragnet (though they'd have to play the same game of treating Iran as a terrorist organization that they currently do, but I assume they will), you'd have upwards of 15 unique identifiers obtained *just from Google*. And that doesn't include a single cookie, which I've seen other subpoenas to Google return.
>
> In other words, one likely reason the IC has decided, now that they're going to report in terms of unique identifiers, they can't report the number of identifiers targeted under PRISM is because it would make it clear that those 93,000 targets represent, very conservatively, over a million identifiers — and once you add in cookies, maybe a billion identifiers — targeted. And reporting *that* would make it clear what kind of identifier soup the IC is swimming in.

Here's another list of the kinds of identifiers the government seeks with just a 2703(d) order (remember, under PRISM, the government would get both this list of the identifiers, as well as the content or other activity, including location data, tied to the identifiers).

> A. The following information about the customers or subscribers of the Account:
> 1. Names (including subscriber names, user names, and screen names);
> 2. Addresses (including mailing addresses, residential addresses, business addresses, and e-mail addresses);
> 3. Local and long distance telephone

connection records;
4. Records of session times and durations, and the temporarily assigned network addresses (such as Internet Protocol (“IP”) addresses) associated with those sessions;
5. Length of service (including start date) and types of service utilized;
6. Telephone or instrument numbers (including MAC addresses);
7. Other subscriber numbers or identities (including temporarily assigned network addresses and registration Internet Protocol (“IP”) addresses (including carrier grade natting addresses or ports)); and
8. Means and source of payment for such service (including any credit card or bank account number) and billing records.

B. All records and other information (not including the contents of communications) relating to the Account, including:
1. Records of user activity for each connection made to or from the Account, including log files; messaging logs; the date, time, length, and method of connections; data transfer volume; user names; and source and destination Internet Protocol addresses;
2. Information about each communication sent or received by the Account, including the date and time of the communication, the method of communication, and the source and destination of the communication (such as source and destination email addresses, IP addresses, and telephone numbers);
3. Records of any accounts registered with the same email address, phone number(s), method(s) of payment, or IP address as either of the accounts listed in Part 1; and Records of any accounts that are linked to either of the

> accounts listed in Part 1 by machine cookies (meaning all Google user IDs that logged into any Google account by the same machine as either of the accounts in Part A).

But for PRISM requests (as opposed to the new phone dragnet implemented in 2006), this works in reverse, with the government providing long lists of identifiers it wants to task, which may or may not reflect groupings using NSA's own correlation process into identifiable targets. While the government surely asks for all Google content knowingly tied to all accounts of a known identifier (so, for example, if the government tasked "emptywheel" they also might get random Google accounts I set up under different names years ago, as well as accounts they connect by common use of the same cookie), it's possible the government submits selectors believing they belong to the same person when in fact they are separate individuals.

Particularly once you're tying collection to an IP address, it's likely you'll get multiple people off the same selector. And it may take Google some time to sort all that out. So that's my guess of what's going on: the change in numbers reflects the degree of uncertainty — even for Google! — regarding how many people are actually being targeted here.

That said, given the obviously different methodologies in counting these numbers, it may also work the other way. That is, Google may at first believe it has just turned over the data for, say, 10 of a user's Google services, only to later realize it has also provided content or ad profile or Google map location data or Google pay.

Whatever it is, it is telling that even Google (!!!) can't track how many targets FISA collection involves in real time.