THE TRIAGE DOCUMENT

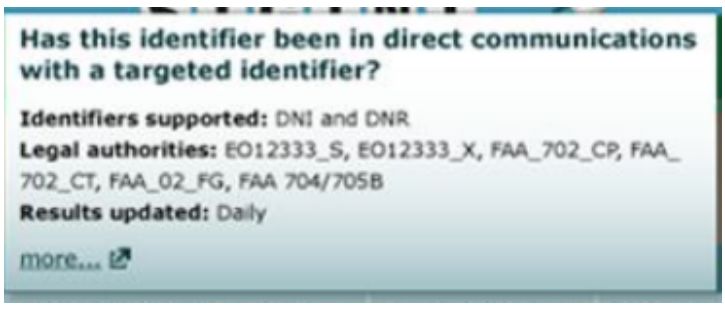
Accompanying a new story on GCHQ/NSA cooperation yesterday, the Intercept released one of the most revealing documents about NSA spying yet. It describes efforts to use Identifier Scoreboard to triage leads such that analysts spend manual time only with the most promising leads. Basically, the NSA aims to use this process to differentiate the 75% of metadata they collect that is interesting but not of high interest into different categories for further analysis.

It does so by checking the leads — which are identifiers like email addresses and phone numbers — against collected data (and this extends beyond just stuff collected on the wires; it includes captured media) to see what kind of contacts with existing targets there have been. Not only does the system pull up what prior contacts of interest exist, but also what time frame those occurred and in what number. From there, the analyst can link directly to either the collected knowledge about a target **or the content**.

Before I get into the significance, a few details.

First, the system works with both phone and Internet metadata. That's not surprising, and it does not yet prove they're chaining across platforms. But it is another piece of evidence supporting that conclusion.

More importantly, look at the authorities in question:



First, FAA. The CP and CT are almost certainly

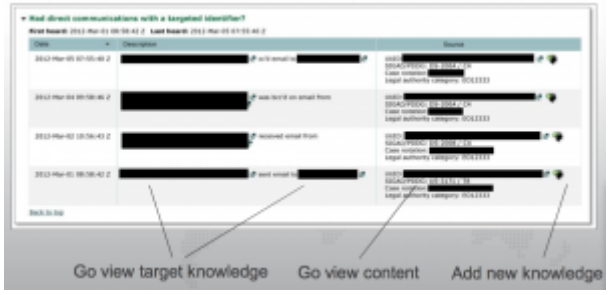
certificates, the authority to collect on counterproliferation and counterterrorism targets. But note what's not there? Cybersecurity, the third known certificate (there was a third certificate reapproved in 2011, so it was active at this time). Which says they may be using that certificate differently (which might make sense, given that you'd be more interested in forensic flows, but this triage system is used with things like TAO which presumably include cyber targets).

There is, however, a second kind of FAA, "FG." That may be upstream or it may be something else (FG could certainly stand for "Foreign Government, which would be consistent with a great deal of other data). If it's something else, it supports the notion that there's some quirk to how the government is using FAA that differs from what they've told PCLOB and the Presidential Review Group, which have both said there are just those 3 certificates.

Then there's FAA 704/705B. This is collection on US person overseas. Note that FAA 703 (collection on US person who is located overseas but the collection on whom is in the US) is not included. Again, this shows something about how they use these authorities.

Finally, there are two EO12333s. In other slides, we've seen an EO12333 and an EO123333 SPCMA (which means you can collect and chain through Americans), and that may be what this is. Update: One other possibility is that this distinguishes between EO12333 data collected by the US and by second parties (the Five Eyes).

Now go to what happens when an identifier has had contact with a target — and remember, these identifiers are just random IDs at this point.



The triage program automatically pulls up prior contacts with targets. Realize what this is? It's a backdoor search, conducted off an identifier about which the NSA has little knowledge.

And the triage provides a link directly from that the metadata describing when the contact occurred and who initiated it **to the content**.

When James Clapper and Theresa Shea describe the metadata serving as a kind of index that helps prioritize what content they read, this is part of what they're referring to. That — for communications involving people who have already been targeted under whatever legal regime — the metadata leads directly to the content. (Note, this triage does not apparently include BR FISA or PRTT data — that is, metadata collected in the US — which says there are interim steps before such data will lead directly to content, though if that data can be replicated under EO 12333, as analysts are trained to do, it could more directly lead to this content.)

So they find the identifiers, search on prior contact with targets, then pull up that data, at least in the case of EO12333 data. (Another caution, these screens date from a period when NSA was just rolling out its back door search authorities for US persons, and there's nothing here that indicates these were US persons, though it does make clear why — as last year's audit shows — NSA has had numerous instances where they've done back door searches on US person identifiers they didn't know were US person identifiers.)

Finally, look at the sources. The communications identified here all came off EO12333

communications (interestingly, this screen doesn't ID whether we're looking at EO12333_X or _S data). As was noted to me this morning, the SIGADS that are known here are offshore. But significantly, they include MUSCULAR, where NSA steals from Google overseas.

That is, this screen shows NSA matching metadata with metadata and content that they otherwise might get under FAA, legally, within the US. They're identifying that as EO12333 data. EO12333 data, of course, gets little of the oversight that FAA does.

At the very least, this shows the NSA engaging in such tracking, including back door searches, off a bunch of US providers, yet identifying it as EO12333 collection.

Update: Two more things on this. Remember NSA has been trying, unsuccessfully, to replace its phone dragnet "alert" function since 2009 when the function was a big part of its violations (a process got approved in 2012, but the NSA has not been able to meet the terms of it technically, as of the last 215 order). This triage process is similar — a process to use with fairly nondescript identifiers to determine whether they're worth more analysis. So we should assume that, while BR FISA (US collected phone dragnet) information is not yet involved in this, the NSA aspires to do so. There are a number of reasons to believe that moving to having the providers do the initial sort (as both the RuppRoge plan offered by the House Intelligence Committee and Obama's plan do) would bring us closer to that point.

Finally, consider what this says about probable cause (especially if I'm correct that EO12333_S is the SPMCA that includes US persons). Underlying all this triage is a theory of what constitutes risk. It measures risk in terms of conversations —how often, how long, how many times — with "dangerous" people. While that may well be a fair measure in some cases, it may not be (I've suggested, for example, that people who don't know they may be at risk are more likely

to speak openly and at length, and those conversations then serve as a kind of camouflage for the truly interesting, rare by operational security conversations). But this theory (though not this particular tool) likely lies behind a lot of the young men who’ve been targeted by FBI.