

THE NSA'S PURGE OBFUSCATIONS

One thing that the 2011 702 documents Charlie Savage liberated make clear is that the government is (and was) obtaining more domestic communications – but purging them – than it wants to let on (and the numbers have surely gotten worse since 2011).

In a hearing on September 7, 2011, the first question that John Bates asked (starting at PDF 35) about the sampling the NSA had done is how many communications had been purged before the agency started counting its sample, a sample that included both PRISM and upstream collection. As Bates noted, it would be one thing if the NSA were purging half its collection and then counting than if it only had to purge a small amount.

So that 13.25 million sample that you saw is reflective of what was in our systems as of that particular date. As you march forward in time from the time we took that sample, you would expect that other items that were collected during that time could also be similarly identified through our compliance process and purged.

THE COURT: Well, let me -- this is not intended to be reflective of the truth and is simply for demonstrative purposes, but if you purged 13 million from during that time period, then all the numbers that you're presenting would really only be half of the picture in terms of what's collected.

See, I'm interested ultimately in what's being collected, not what happens to be sitting in your data files at a particular point in time. If the purging that takes place as a normal course of business is half of the material, then it changes everything just from that perspective even without knowing what the purged material is, whether it's of a different nature and richer in terms of wholly domestic communications or not. Just in terms of raw numbers, it would really alter

During this exchange, the government was careful to limit their discussion of purged communications to upstream MCT related collection.

██████: In our previous filings, we have indicated that prior to our statistical analysis we had not identified any communications of the MCT type that were wholly domestic, which would require purge.

When the government responded (starting at PDF 117), it provided numbers for just what had gotten purged from upstream collection.

was NSA's intent to capture for further manual review a truly representative sample of Internet communications acquired through NSA's upstream collection. Nevertheless, in order to ensure that the Government drew an appropriately representative sample of Internet communications with which to conduct its manual review, the Court requested to know the total number of Internet communications acquired by NSA during this six month period and the total number of Internet communications purged during this six month period. NSA reports that from January 1, 2011, to June 30, 2011, approximately 18,446 upstream transactions were acquired and thereafter purged from ████████ during that same time period.² The 18,446 transactions were purged for various reasons, such as a target traveling to the United States and other matters more specifically reported pursuant to Rule 13(b) of the Rules of the Foreign Intelligence Surveillance Court, including the Quarterly Reports Concerning Compliance Matters under Section 702 of FISA. For example, many related to two over-collection incidents previously reported to the Court on February 9, 2011 and June 24, 2011. Aside from the possibility of a target traveling to the United States, as otherwise reported to the Court, none of the transactions which were purged related to NSA's discovery of a wholly domestic communication acquired through its upstream techniques. ~~(TS//SI//NF)~~

I'm not entirely sure their claim that none of this purged information was "upstream" collection – as opposed to MCT collection – is correct (as a post on the violations will explore). But they make it clear: the 18,446 purged communications were just Internet upstream. For every upstream record purged because the target had roamed into the US, there might be correlated telephony collection that would get purged – some of the most commonly discussed purged communications. It might also include PRISM production that would have to get purged (if, for example, the target continued to use Gmail while in the US). In addition, there might be targets discovered to be (perhaps by reading that PRISM production) Americans. So the 18,446 is just a portion of what got purged – but the government pointedly avoided telling Bates how much of the other kind there was.

Of the upstream Internet collection in 2011, .1% was getting purged.

The purge numbers for telephony and PRISM would not be the same as for upstream. The telephony numbers might be far far higher, given public reporting from the period. The NSA was working off some overcollection that was limited to upstream during this period, which would lead to

more upstream communications being purged. But the rules on domestic collection of PRISM communications are different than they are for upstream.

In any case, the government's careful dodge of providing Bates the full purge number suggests the telephony and PRISM purge numbers might be substantial, too. But we don't get that number.