Newsweek

Why Did Yahoo Allow the Government to Scan Emails?

Julian Sanchez

October 8, 2016

Reuters dropped a bombshell story on Tuesday, reporting that in 2015 Yahoo agreed to scan all its users' incoming emails on behalf of a U.S. intelligence agency, hunting for a particular "character string" and turning over messages where it found a match to the government.

Yet the vagueness of the story—which appears to be based on sources with limited access to the details of the surveillance—leaves a maddening number of unanswered questions.

Yahoo did not greatly help matters with a meticulously worded non-denial, calling the story "misleading" without calling it substantively false, and asserting that the "scanning described in the article does not exist on our systems." (Obvious follow-up questions: *Did* it exist in 2015? *Does* it now exist on some other systems?)

Then, on Wednesday, Charlie Savage and Nicole Perlroth of *The New York Times* published a follow-up article fleshing out some of the details: The bulk scan was conducted pursuant to an order from the secretive Foreign Intelligence Surveillance Court (FISA) and hunted for a "digital signature" associated with a foreign "state-sponsored terror organization."

What's troubling here is that it suggests Yahoo was asked to scan the*contents* of all messages for a string of characters indicating that the message was produced using a particular software tool—such as, for instance, the "Mujahideen Secrets" encryption software used by Al-Qaeda.

There is, of course, nothing inherently wrong with targeting tools associated with known adversaries, but this does represent a dramatic inversion of the traditional way surveillance is conducted.

Normally, we expect that the government will identify a "communications facility" being used by a particular target, then proceed to scrutinize its communications. Here, the government has scrutinized *an entire stream of communications* in bulk, searching for something in the contents that would allow them to identify the target!

It is not hard to see why intelligence agencies would find such scans useful, but it would be a serious mistake to normalize the bulk scanning of communications content—an indiscriminate "search" of people not known to be foreign intelligence targets—even if one is not overly dismayed by a particular application of that approach.

Surveillance architectures create their own institutional momentum, and a software tool designed to scan for digital fingerprints can just as easily scan for words or phrases in messages written by humans, or for cryptographic tools used by many innocent people seeking to protect their privacy, as well as a few bad actors. This is a possibility that becomes far more tempting once the necessary technical infrastructure is in place.

That the government would employ this approach, however, should not exactly come as a great surprise. The National Security Agency's (NSA) targeting procedures for §702 of the FISA Amendments Act of 2008, disclosed three years ago by Edward Snowden, seem to contemplate keying surveillance to such signatures. One of the criteria mentioned for verifying the "foreignness" of a surveillance target is:

> *Information indicates that Internet Protocol ranges and/or specific electronic identifiers or signatures (e.g., specific types of cryptology or steganography) are used almost exclusively by individuals associated with a foreign power or foreign territory, or are extensively used by individuals associated with a foreign power or foreign territory.*

Note that "exclusively"—which may sound reassuring—is quickly followed by "extensively," which would no doubt encompass a great deal of privacy-protecting technology used by many law-abiding Americans as well as foreign criminals or violent extremists.

It also seems at least plausible that searches of this sort have been carried out domestically for far longer than a year. Under the FISA Amendments Act's §702, the government can designate foreign targets for intelligence collection—including electronic surveillance carried out domestically—under broad targeting procedures approved by the FISA Court, without any specific judicial approval of individual targets.

At last count, there were 94,368 such "targets" being monitored under a single blanket authorization. As we know, thanks to the Snowden disclosures, one way they conduct §702 surveillance is known asPrism collection and is carried out with the cooperation of communications providers like Yahoo or Google.

The other main mechanism is known as "Upstream" collection and involves scanning of traffic on the internet backbone—including not just message headers but also the contents of messages—for "selectors" associated with approved targets. This has sometimes been dubbed "about collection," because it means that the NSA would intercept not only messages *to* or *from* the email address used as a selector but also messages mentioning or "about" that selector.

It is at least possible that the government has been routinely using digital fingerprints—the text that says, in effect, "The following message is encrypted with a certain type of software"—along with more conventional selectors like email or IP addresses to scan internet traffic in bulk.

Why, then, would the government ask Yahoo to start doing such scanning for it in 2015? One possibility is that ever since the Snowden revelations began, more and more companies have been encrypting their traffic by default, using a protocol known as Transport Layer Security, or TLS (the email-specific version of which is called STARTTLS).

The wider adoption of such encryption means data that *would* have been visible to an NSA sniffer sitting on the internet backbone is now scrambled and unintelligible, making Upstream increasingly useless.

Even for the NSA, breaking the encryption on traffic wholesale is likely infeasible, but aside from any message content separately encrypted by individual users, that traffic would be readable once it had arrived at Yahoo and been decrypted with the company's private keys. Yahoobegan making such encryption the default in 2014.

One obvious question is whether Yahoo is the only company to be served with such an order or whether it reflects a more widespread practice.Sam Biddle of The Intercept queried some major providers and got relatively straightforward denials from Google, Facebook, Twitter and Apple, though it remains possible that this is a byproduct of the Reuters story having gotten some details of the story wrong.

Microsoft said it had "never engaged in the secret scanning of email traffic like what has been reported" but "would not comment on the record as to whether the company has ever received such a request," which could reflect simple legal caution. Intelligence surveillance requests are invariably covered by broad gag orders, and it gets awkward quickly if you deny getting some types but "no comment" others. Or it could be an indication that the company received a similar demand but successfully fought it.

A second, perhaps less obvious question, is: Why would the scanning be limited to incoming messages (rather than messages either sent or received by Yahoo users) and only to real-time scanning (rather than encompassing older messages stored in the company's servers)?

One possibility has to do with affecting the "facility" at which the surveillance was "directed." In NSA jargon, there is the "target" of surveillance (the person or entity about or from whom information is sought); the "selector" (the specific term used to filter out the information to be collected); and the "facility" at which surveillance is directed (the physical or virtual communications channel from which the information is obtained).

In the simplest type of case, these could all be the same. There is a target known only as the user of a particular email address, which serves as both the "selector" and—when communications are obtained from the provider that hosts that account—the "facility" at which surveillance is directed. But they might also all be different.

An individual target might have several associated "selectors" (different email accounts or other digital identifiers), and as the case of Upstream "about collection" shows, the "facility" might be an internet routing switch rather than a particular repository of stored messages associated with that account.

My (possibly incomplete) understanding from discussions with intelligence officials is that a scan of the *content* of a message sitting in a particular user's inbox would be considered surveillance "directed at the facility" of the individual user's account, even if the scan was based on some different selector. Such a scan would likely require that the person whose inbox it was be considered a "target."

However, intelligence community lawyers have conceivably decided the situation is different if the scans are conducted before messages are routed to specific inboxes—at which stage they're treated like Upstream traffic.

In other words, before messages arrive in the recipient's inbox, there might be no "particular, known U.S. person" who could be considered a "target" of the scan, triggering a laxer set of rules constraining searches.

Whatever the reality, the government should now release an appropriately redacted version of the FISA Court opinion authorizing this bulk email scanning—which appears to have come just months before the passage of the USA Freedom Act, under which it would be *obligated* to prepare an unclassified legal opinion for public release.

If the government is going to be compelling companies to scan everyone's communications in its hunt for extremist attackers, the public is entitled to understand the legal framework within which it plans to do so—and to modify or reject that framework if it fails to meet Fourth Amendment standards.

Julian Sanchez is a senior fellow at the Cato Institute and for Reason *magazine.*