

Boundless Informant: the NSA's secret tool to track global surveillance data

9. Juni 2012

James Ball and Spencer Ackerman, The Guardian

The National Security Agency has developed a powerful tool for recording and analysing where its intelligence comes from, raising questions about its repeated assurances to Congress that it cannot keep track of all the surveillance it performs on American communications.

The Guardian has acquired top-secret documents about the NSA datamining tool, called Boundless Informant, that details and even maps by country the voluminous amount of information it collects from computer and telephone networks.

The focus of the internal NSA tool is on counting and categorizing the records of communications, known as metadata, rather than the content of an email or instant message.

The Boundless Informant documents show the agency collecting almost 3 billion pieces of intelligence from US computer networks over a 30-day period ending in March 2013. One document says it is designed to give NSA officials answers to questions like, "What type of coverage do we have on country X" in "near real-time by asking the SIGINT [signals intelligence] infrastructure."

An NSA factsheet about the program, acquired by the Guardian, says: "The tool allows users to select a country on a map and view the metadata volume and select details about the collections against that country."

Under the heading "Sample use cases", the factsheet also states the tool shows information including: "How many records (and what type) are collected against a particular country."

A snapshot of the Boundless Informant data, contained in a top secret NSA "global heat map" seen by the Guardian, shows that in March 2013 the agency collected 97bn pieces of intelligence from computer networks worldwide.

Iran was the country where the largest amount of intelligence was gathered, with more than 14bn reports in that period, followed by 13.5bn from Pakistan. Jordan, one of America's closest Arab allies, came third with 12.7bn, Egypt fourth with 7.6bn and India fifth with 6.3bn.

The heatmap gives each nation a color code based on how extensively it is subjected to NSA surveillance. The color scheme ranges from green (least subjected to surveillance) through yellow and orange to red (most surveillance).

The disclosure of the internal Boundless Informant system comes amid a struggle between the NSA and its overseers in the Senate over whether it can track the intelligence it collects on American communications. The NSA's position is that it is not technologically feasible to do so.

At a hearing of the Senate intelligence committee In March this year, Democratic senator Ron Wyden asked James Clapper, the director of national intelligence: "Does the NSA collect any type of data at all on millions or hundreds of millions of Americans?"

"No sir," replied Clapper.

Judith Emmel, an NSA spokeswoman, told the Guardian in a response to the latest disclosures: "NSA has consistently reported – including to Congress – that we do not have the ability to determine with certainty the identity or location of all communicants within a given communication. That remains the case."

Other documents seen by the Guardian further demonstrate that the NSA does in fact break down its surveillance intercepts which could allow the agency to determine how many of them are from the US. The level of detail includes individual IP addresses.

IP address is not a perfect proxy for someone's physical location but it is rather close, said Chris Soghoian, the principal technologist with the Speech Privacy and Technology Project of the American Civil Liberties Union. "If you don't take steps to hide it, the IP address provided by your internet provider will certainly tell you what country, state and, typically, city you are in," Soghoian said.

That approximation has implications for the ongoing oversight battle between the intelligence agencies and Congress.

On Friday, in his first public response to the Guardian's disclosures this week on NSA surveillance, Barack Obama said that that congressional oversight was the American peoples' best guarantee that they were not being spied on.

"These are the folks you all vote for as your representatives in Congress and they are being fully briefed on these programs," he said. Obama also insisted that any surveillance was "very narrowly circumscribed".

Senators have expressed their frustration at the NSA's refusal to supply statistics. In a letter to NSA director General Keith Alexander in October last year, senator Wyden and his Democratic colleague on the Senate intelligence committee, Mark Udall, noted that "the intelligence community has stated repeatedly that it is not possible to provide even a rough estimate of how many American communications have been collected under the Fisa Amendments Act, and has even declined to estimate the scale of this collection."

At a congressional hearing in March last year, Alexander denied point-blank that the agency had the figures on how many Americans had their electronic communications collected or reviewed. Asked if he had the capability to get them, Alexander said: "No. No. We do not have the technical insights in the United States." He added that "nor do we have the equipment in the United States to actually collect that kind of information".

Soon after, the NSA, through the inspector general of the overall US intelligence community, told the senators that making such a determination would jeopardize US intelligence operations – and might itself violate Americans' privacy.

"All that senator Udall and I are asking for is a ballpark estimate of how many Americans have been monitored under this law, and it is disappointing that the inspectors general cannot provide it," Wyden told Wired magazine at the time.

The documents show that the team responsible for Boundless Informant assured its bosses that the tool is on track for upgrades.

The team will "accept user requests for additional functionality or enhancements," according to the FAQ acquired by the Guardian. "Users are also allowed to vote on which functionality or enhancements are most important to them (as well as add comments). The BOUNDLESSINFORMANT team will periodically review all requests and triage according to level of effort (Easy, Medium, Hard) and mission impact (High, Medium, Low)."

Emmel, the NSA spokeswoman, told the Guardian: "Current technology simply does not permit us to positively identify all of the persons or locations associated with a given communication (for example, it may be possible to say with certainty that a communication traversed a particular path within the internet. It is harder to know the ultimate source or destination, or more particularly the identity of the person represented by the TO:, FROM: or CC: field of an e-mail address or the abstraction of an IP address).

"Thus, we apply rigorous training and technological advancements to combine both our automated and manual (human) processes to characterize communications – ensuring protection of the privacy rights of the American people. This is not just our judgment, but that of the relevant inspectors general, who have also reported this."

She added: "The continued publication of these allegations about highly classified issues, and other information taken out of context, makes it impossible to conduct a reasonable discussion on the merits of these programs."