

# The NSA's SKYNET program may be killing thousands of innocent people

---

ars [arstechnica.com/information-technology/2016/02/the-nsas-skynet-program-may-be-killing-thousands-of-innocent-people](http://arstechnica.com/information-technology/2016/02/the-nsas-skynet-program-may-be-killing-thousands-of-innocent-people)

Christian Grothoff & J.M. Porup

In 2014, the former director of both the CIA and NSA proclaimed that "we kill people based on metadata." Now, a new examination of previously published Snowden documents suggests that many of those people may have been innocent.

Last year, The Intercept published documents detailing the NSA's SKYNET programme. According to the documents, SKYNET engages in mass surveillance of Pakistan's mobile phone network, and then uses a machine learning algorithm on the cellular network metadata of 55 million people to try and rate each person's likelihood of being a terrorist.

Patrick Ball—a data scientist and the director of research at the Human Rights Data Analysis Group—who has previously given expert testimony before war crimes tribunals, described the NSA's methods as "ridiculously optimistic" and "completely bullshit." A flaw in how the NSA trains SKYNET's machine learning algorithm to analyse cellular metadata, Ball told Ars, makes the results scientifically unsound.

Somewhere between 2,500 and 4,000 people have been killed by drone strikes in Pakistan since 2004, and most of them were classified by the US government as "extremists," the Bureau of Investigative Journalism reported. Based on the classification date of "20070108" on one of the SKYNET slide decks (which themselves appear to date from 2011 and 2012), the machine learning program may have been in development as early as 2007.

In the years that have followed, thousands of innocent people in Pakistan may have been mislabelled as terrorists by that "scientifically unsound" algorithm, possibly resulting in their untimely demise.

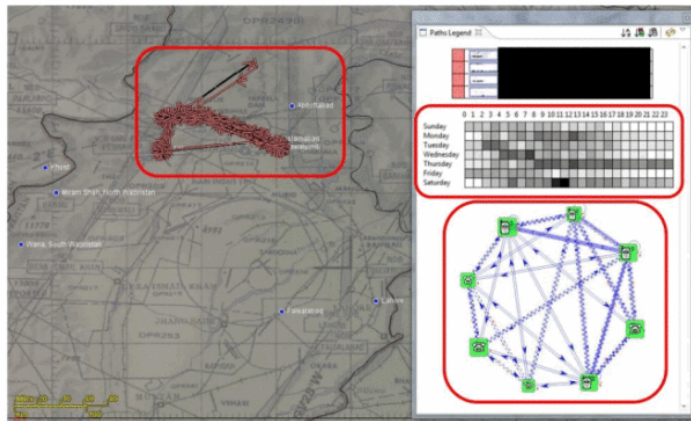
## The siren song of big data

---

SKYNET works like a typical modern Big Data business application. The program collects metadata and stores it on NSA cloud servers, extracts relevant information, and then applies machine learning to identify leads for a targeted campaign. Except instead of trying to sell the targets something, this campaign, given the overall business focus of the US government in Pakistan, likely involves another branch of the US government—the CIA or military—that executes their "Find-Fix-Finish" strategy using Predator drones and on-the-ground death squads.

TOP SECRET//COMINT//REL TO USA, FVEY

From GSM metadata, we can measure aspects of each selector's **pattern-of-life**, **social network**, and **travel behavior**



TOP SECRET//COMINT//REL TO USA, FVEY

Enlarge / From GSM metadata, we can measure aspects of each selector's pattern-of-life, social network, and travel behaviour

In addition to processing logged cellular phone call data (so-called "DNR" or Dialed Number Recognition data, such as time, duration, who called whom, etc.), SKYNET also collects user location, allowing for the creation of detailed travel profiles. Turning off a mobile phone gets flagged as an attempt to evade mass surveillance. Users who swap SIM cards, naively believing this will prevent tracking, also get flagged (the ESN/MEID/IMEI burned into the handset makes the phone trackable across multiple SIM cards).

TOP SECRET//SI//REL TO USA, FVEY

## Cloud Analytic Building Blocks

- Travel Patterns
  - Travel phrases (Locations visited in given timeframe)
  - Regular/repeated visits to locations of interest
- Behavior-Based Analytics
  - Low use, incoming calls only
  - Excessive SIM or Handset swapping
  - Frequent Detach/Power-down
  - Courier machine learning models
- Other Enrichments
  - Travel on particular days of the week
  - Co-travelers
  - Similar travel patterns
  - Common contacts
  - Visits to airports
  - Other countries
  - Overnight trips
  - Permanent move

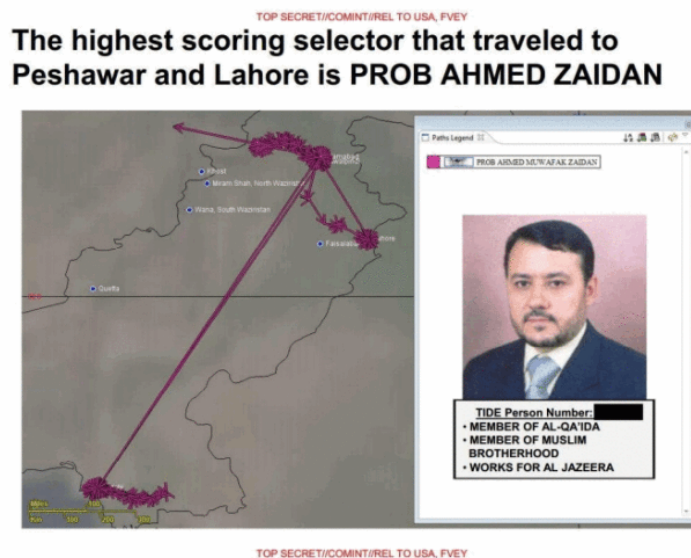
TOP SECRET//SI//REL TO USA, FVEY

Enlarge / Travel patterns, behaviour-based analytics, and other "enrichments" are used to analyse the bulk metadata for terroristiness.

Even handset swapping gets detected and flagged, the slides boast. Such detection, we can only speculate (since the slides do not go into detail on this point), is probably based on the fact that other metadata, such as user location in the real world and social network, remain unchanged.

Given the complete set of metadata, SKYNET pieces together people's typical daily routines—who travels together, have shared contacts, stay overnight with friends, visit other countries, or move permanently. Overall, the slides indicate, the NSA machine learning algorithm uses more than 80 different properties to rate people on their terroristiness.

The program, the slides tell us, is based on the assumption that the behaviour of terrorists differs significantly from that of ordinary citizens with respect to some of these properties. However, as The Intercept's exposé last year made clear, the highest rated target according to this machine learning program was Ahmad Zaidan, Al-Jazeera's long-time bureau chief in Islamabad.



Enlarge / The highest scoring selector who travelled to Peshawar and Lahore is "PROB AHMED ZAIDAN", Al-Jazeera's long-time bureau chief in Islamabad.

As The Intercept reported, Zaidan frequently travels to regions with known terrorist activity in order to interview insurgents and report the news. But rather than questioning the machine learning that produced such a bizarre result, the NSA engineers behind the algorithm instead trumpeted Zaidan as an example of a SKYNET success in their in-house presentation, including a slide that labelled Zaidan as a "MEMBER OF AL-QA'IDA."