De-anonymization and mass surveillance

Richard Darst

2013-09-13

De-anonymization: Notable instances

- AOL search release
- Hospital discharge database
- Netflix prize data

Basic model

- (see board)
- Privacy breaching is connecting different data
- “Private attributes” vs “quasi-private attributes” vs “public attributes”

Definitions of privacy

- *k*-anonymity
- *l*-diversity
- Differential privacy

Recent works

- Backstrom active attacks on livejournal (2007)
- Netflix prize data (2008)
- Social network mapping (this talk) (2009)
- D4D de-anonymization paper (2013)

“De-anonymizing Social Networks”, Narayanan and Shmatikov, 2009 (focus of this talk)

- (see the file)
- Passive attack
- Seed nodes
- By comparing node degrees, expand this seed to span the entire network.
- Various measures of overlap and success
- Implement attack comparing Twitter to Flickr: 30-70% of “ground truth” mapping recovered.

Implications

- The more data that is de-anonymized or released, the easier it is to get more.
- Privacy laws have not caught up to this, and mostly cover removal of directly identifying information.
- Data release is good for open government, research, etc. But we need a way to do it while preserving privacy properly.

Mass surveillance

Brief history

- At least 100 years
- ECHELON
- 11 September attacks: modern age?
- Many hints since then that this has been going on (and really, it's not that big a surprise).
- Snowden leaks.

Who does it?

- USA: NSA, CIA
- Top-tier partners: UK, Canada, Australia, New Zealand
- Information shared on don't ask basis (and this allows one to get around domestic spying laws).
- Within US, shared to other law enforcement agencies and they are told to cover up the true origin of the data.

Applicable US laws

- Firts amendment (free speech)
- Fourth amendment (unreasonable search and seizure)
- Right to a speedy and public trial
- Warrant requirements

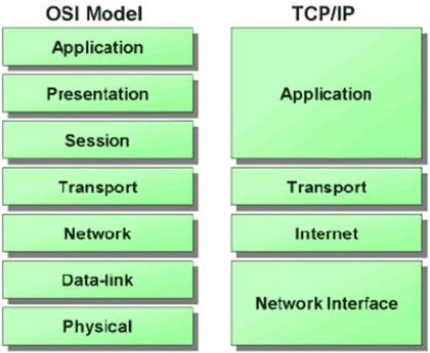
History of Edward Snowden

- Did not graduate high school.
- Worked various NSA/CIA jobs.
- Took a job at Booz Allen Hamilton as a "system administrator" in order to get more documents to leak.
- Contacted journalists and then flew to Hong Kong.
- Barely escaped Hong Kong and flew to Moscow, he got trapped in the international transit zone
- Temporary asylum in Russia under condition that ey doesn't leak anymore.

Mindset of the National Security Agency

- We have the ability to know everything about everyone.
- We have the right to know everything about everyone.
- We have the moral obligation to know everything about everyone.
- We must do this in secret.

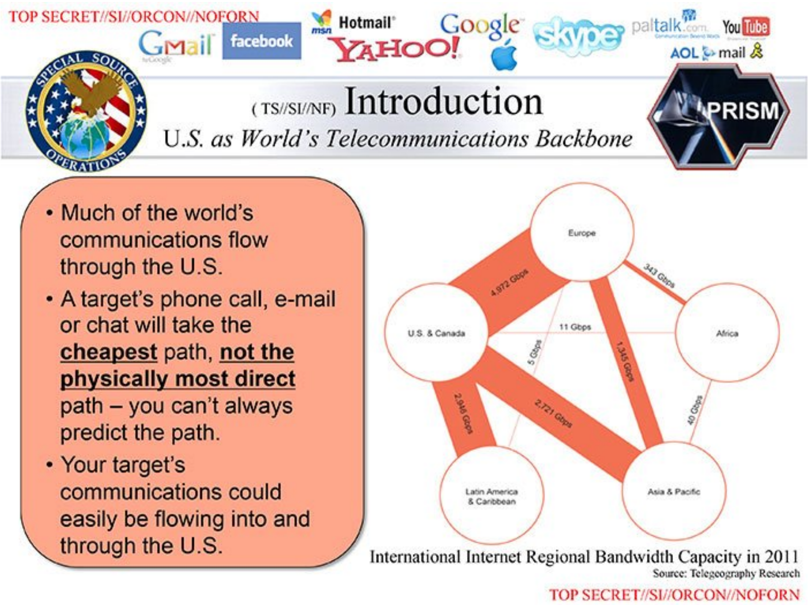
Summary of NSA activities



TCP/IP and the OSI model

Physical/network layers

- Cable tapping
- Access at telecom endpoints (Room 641A)
- See: map of undersea communication cables



TOP SECRET//SI//ORCON//NOFORN

(TS//SI//NF) Introduction

U.S. as World's Telecommunications Backbone

- Much of the world's communications flow through the U.S.
- A target's phone call, e-mail or chat will take the **cheapest** path, **not the physically most direct** path – you can't always predict the path.
- Your target's communications could easily be flowing into and through the U.S.

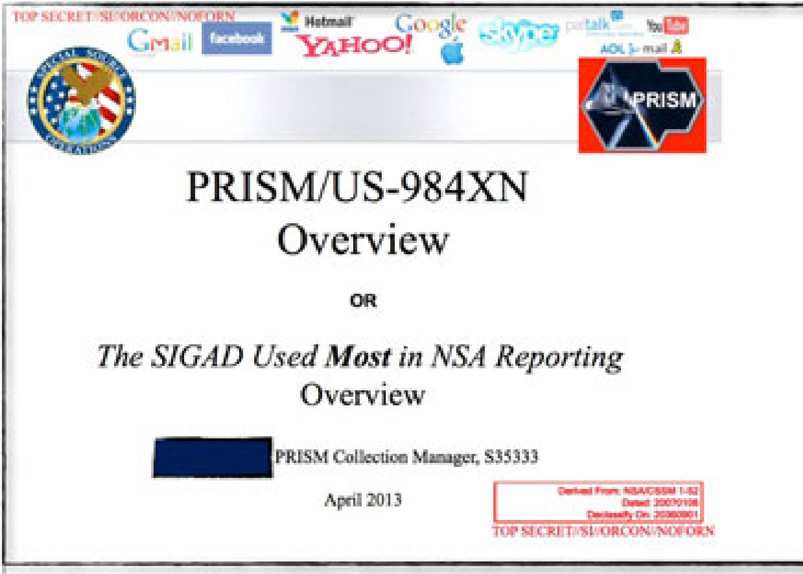
International Internet Regional Bandwidth Capacity in 2011

Source: Telegeography Research

TOP SECRET//SI//ORCON//NOFORN

Cloud attacks

- PRISM: direct access to corporations' servers
- Companies deny it, but they wouldn't even know.



Gmail
facebook
Hotmail
YAHOO!
Google
skype
paltalk
YouTube
AOL mail
PRISM

TOP SECRET//SI//ORCON//NOFORN

PRISM/US-984XN
Overview

OR

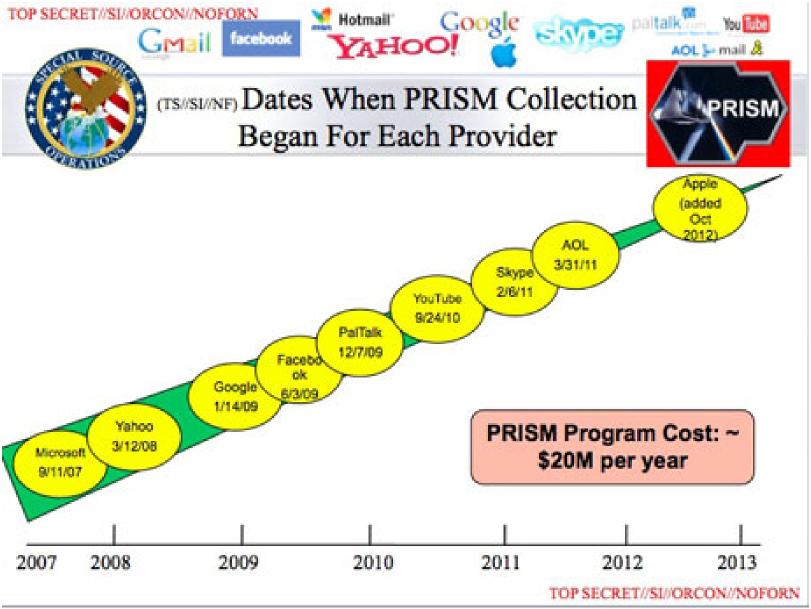
*The SIGAD Used **Most** in NSA Reporting*
Overview

[redacted] PRISM Collection Manager, S35333

April 2013

Derived From: NSA/CSSM 1-52
Dated: 20070108
Declassify On: 20360901

TOP SECRET//SI//ORCON//NOFORN

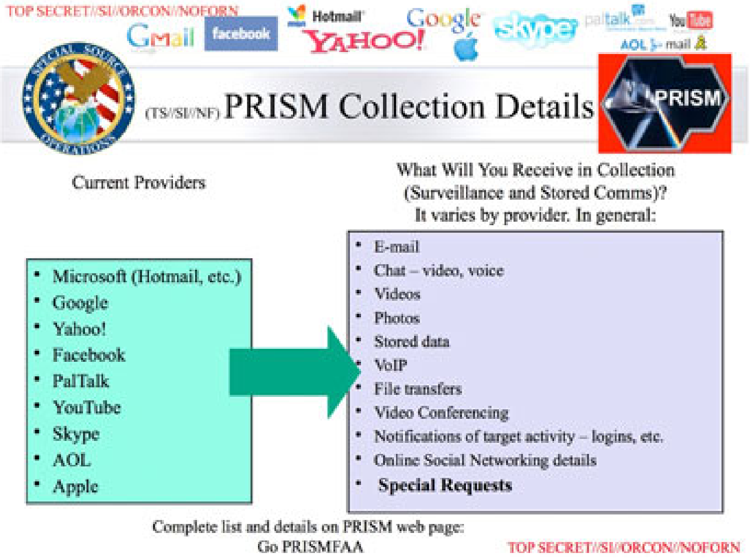


TOP SECRET//SI//ORCON//NOFORN

(TS//SI//NF) Dates When PRISM Collection Began For Each Provider

PRISM Program Cost: ~ $20M per year

TOP SECRET//SI//ORCON//NOFORN


Gmail
facebook
Hotmail
YAHOO!
Google
skype
paltalk
YouTube
AOL mail
PRISM

TOP SECRET//SI//ORCON//NOFORN

(TS//SI//NF) PRISM Collection Details

Current Providers

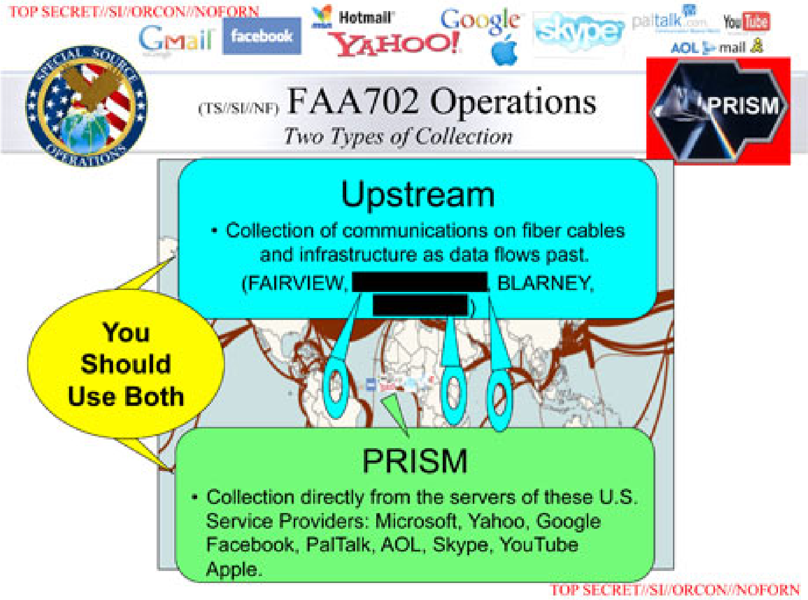
- Microsoft (Hotmail, etc.)
- Google
- Yahoo!
- Facebook
- PalTalk
- YouTube
- Skype
- AOL
- Apple

What Will You Receive in Collection (Surveillance and Stored Comms)?
It varies by provider. In general:

- E-mail
- Chat – video, voice
- Videos
- Photos
- Stored data
- VoIP
- File transfers
- Video Conferencing
- Notifications of target activity – logins, etc.
- Online Social Networking details
- **Special Requests**

Complete list and details on PRISM web page:
Go PRISMFAA

TOP SECRET//SI//ORCON//NOFORN



Gmail
facebook
Hotmail
YAHOO!
Google
skype
paltalk.com
YouTube
AOL mail
PRISM
SPECIAL SOURCE OPERATIONS

TOP SECRET//SI//ORCON//NOFORN

(TS//SI//NF) FAA702 Operations

Two Types of Collection

Upstream

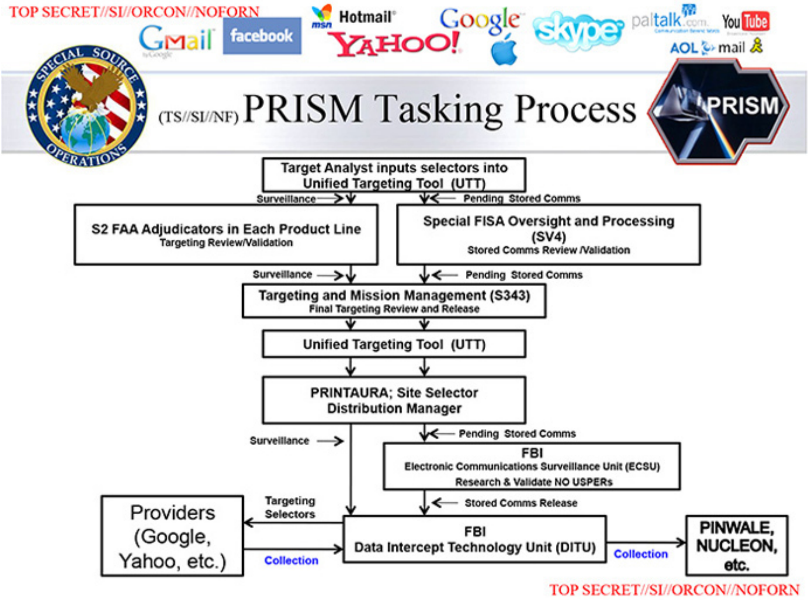
- Collection of communications on fiber cables and infrastructure as data flows past. (FAIRVIEW, [redacted], BLARNEY, [redacted])

You Should Use Both

PRISM

- Collection directly from the servers of these U.S. Service Providers: Microsoft, Yahoo, Google Facebook, PalTalk, AOL, Skype, YouTube Apple.

TOP SECRET//SI//ORCON//NOFORN



TOP SECRET//SI//ORCON//NOFORN

(TS//SI//NF) PRISM Tasking Process

Target Analyst inputs selectors into Unified Targeting Tool (UTT)

Surveillance → ← Pending Stored Comms

S2 FAA Adjudicators in Each Product Line
Targeting Review/Validation

Special FISA Oversight and Processing (SV4)
Stored Comms Review /Validation

Surveillance → ← Pending Stored Comms

Targeting and Mission Management (S343)
Final Targeting Review and Release

Unified Targeting Tool (UTT)

PRINTAURA; Site Selector Distribution Manager

Surveillance → ← Pending Stored Comms

FBI
Electronic Communications Surveillance Unit (ECSU)
Research & Validate NO USPERs

← Stored Comms Release

Providers (Google, Yahoo, etc.)

Targeting Selectors

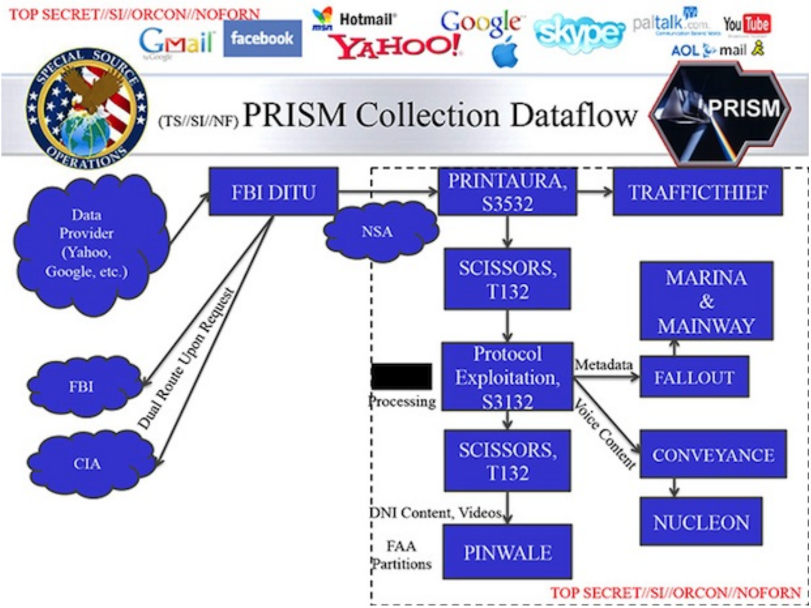
Collection

FBI
Data Intercept Technology Unit (DITU)

Collection

PINWALE, NUCLEON, etc.

TOP SECRET//SI//ORCON//NOFORN



TOP SECRET//SI//ORCON//NOFORN

(TS//SI//NF) PRISM Collection Dataflow

Data Provider (Yahoo, Google, etc.)

FBI DITU

NSA

PRINTAURA, S3532

TRAFFICTHIEF

SCISSORS, T132

MARINA & MAINWAY

[redacted]

Processing

Protocol Exploitation, S3132

Metadata

FALLOUT

Voice Content

CONVEYANCE

NUCLEON

SCISSORS, T132

DNI Content, Videos,

FAA Partitions

PINWALE

FBI

CIA

Dual Route Upon Request

TOP SECRET//SI//ORCON//NOFORN

Social networks

- They consider surveillance of anyone three hops away from a target to be acceptable.
- I'm not sure of extent of mining, but they certainly have access to all data.

Legal aspects

- National security letters
- FISA Court
- “Public-private partnerships”

Who is spied on??

- Legal to spy on anyone who's not a US citizen or on US soil
- "Three hops" away from a target.
- "Secret" list of terrorists that includes people like protesters or journalists.

Attacks on encryption

- Properly-done encryption is probably secure.
- NSA applies pressure to companies to weaken it client-side.
- Certificate authority compromise?

Active hacking

The NSA engages in active hacking against different organizations:

- UN, diplomatic embassies of allies, EU offices
- G20 summit (fake internet cafes)
- Aeroflot, Petrobras, Chinese universities,

Practical issues

- NSA Utah data center
- Teams dedicated to every level, solving every scaling problem
- Largest data management problem ever (?)

Conclusions

Good news

- Properly-implemented strong encryption is probably safe
- Open source and open protocols are probably secure
- Mostly passive attacks, active attacks only on high-value targets.
- We know about it now

Bad news

- Any popular commercial product is probably compromised.
- No rights to due process or public oversight.
- From an engineering point of view, the internet needs fixing.
- "The cloud" is broken from a privacy standpoint.