

# Snowden & Internet

Cees de Laat



# The revelations

- A series of exposés beginning June 5, 2013 revealed Internet surveillance programs such as **PRISM**, **XKeyscore** and **Tempora**, as well as the interception of US and European telephone metadata.





**NSA:**

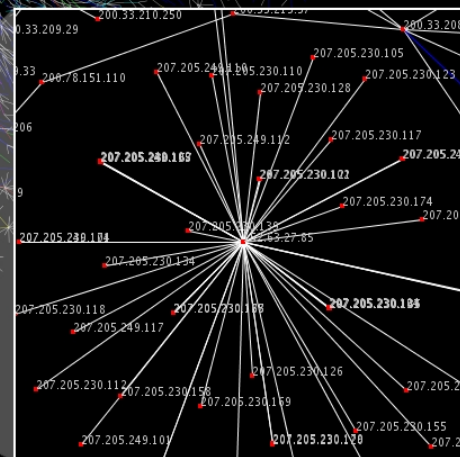
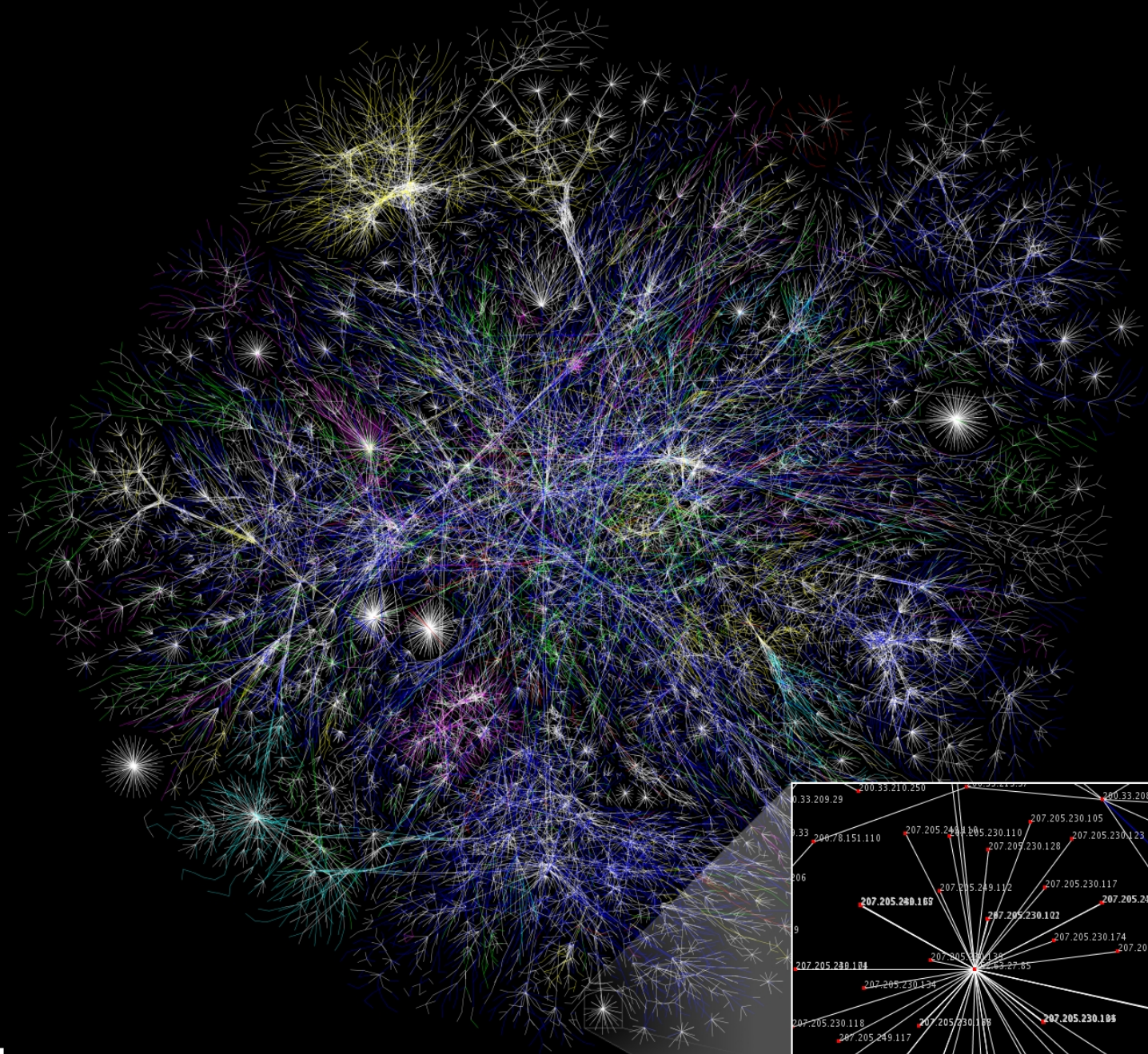
**“Collect it all,  
know it all,  
exploit it all”**

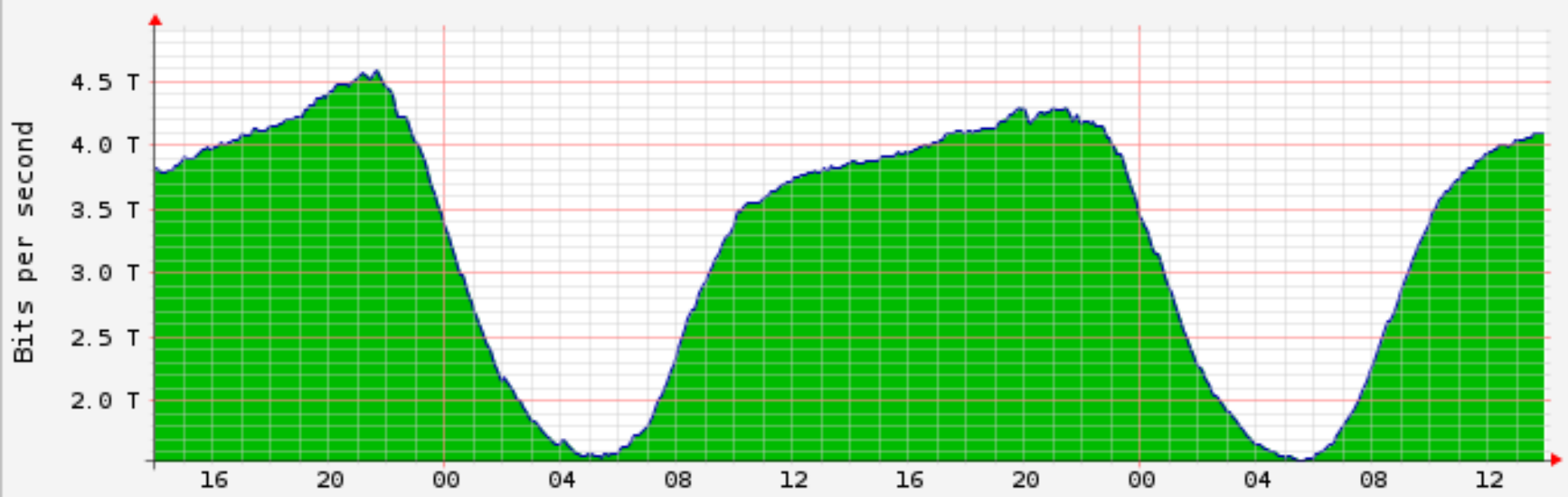


# Snowden personalia

- Edward Joseph Snowden
- Elizabeth City (NC), 21 juni 1983
- Former employee of the CIA
- System manager subcontracted from the company Booz Allen Hamilton by the National Security Agency (NSA)
- In June 2013 Snowden leaked classified information on a number of espionage activities by the NSA on the Internet
- Activities included global surveillance programs run by NSA & Five Eyes Intel Alliance and many other agencies





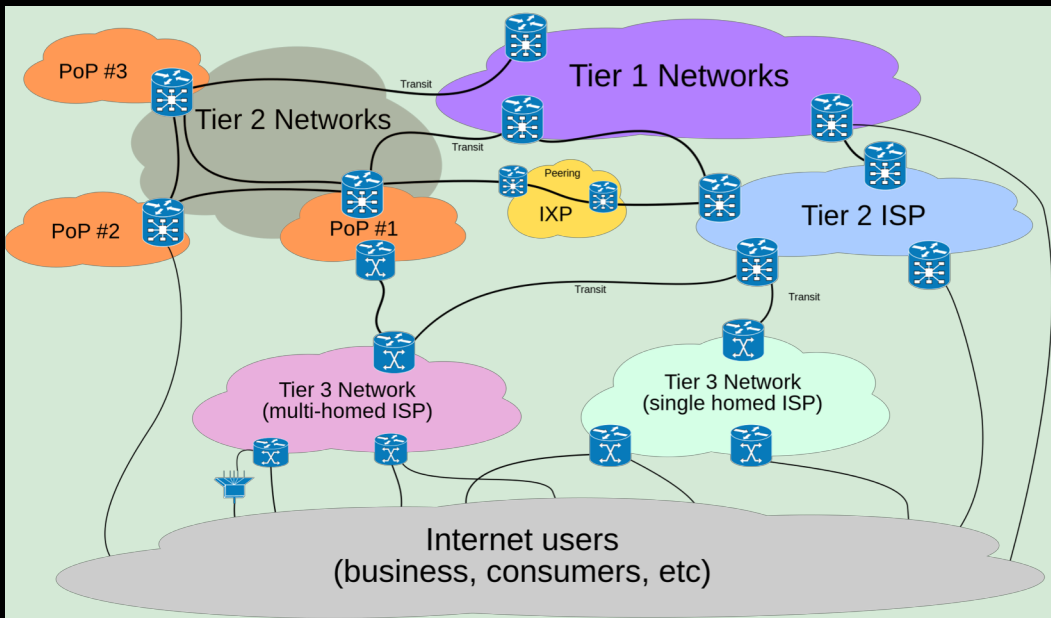
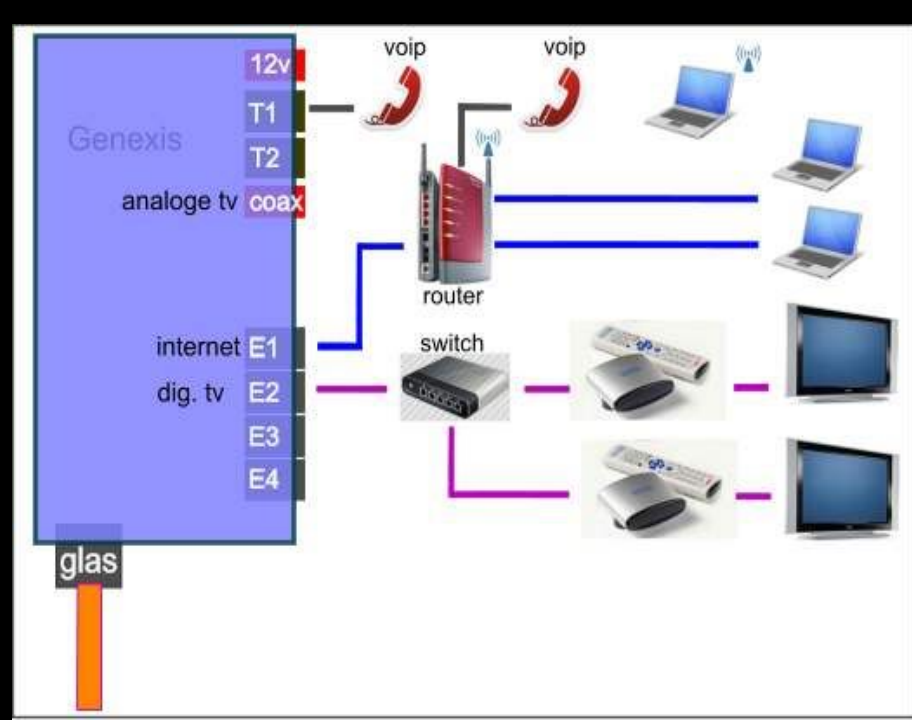
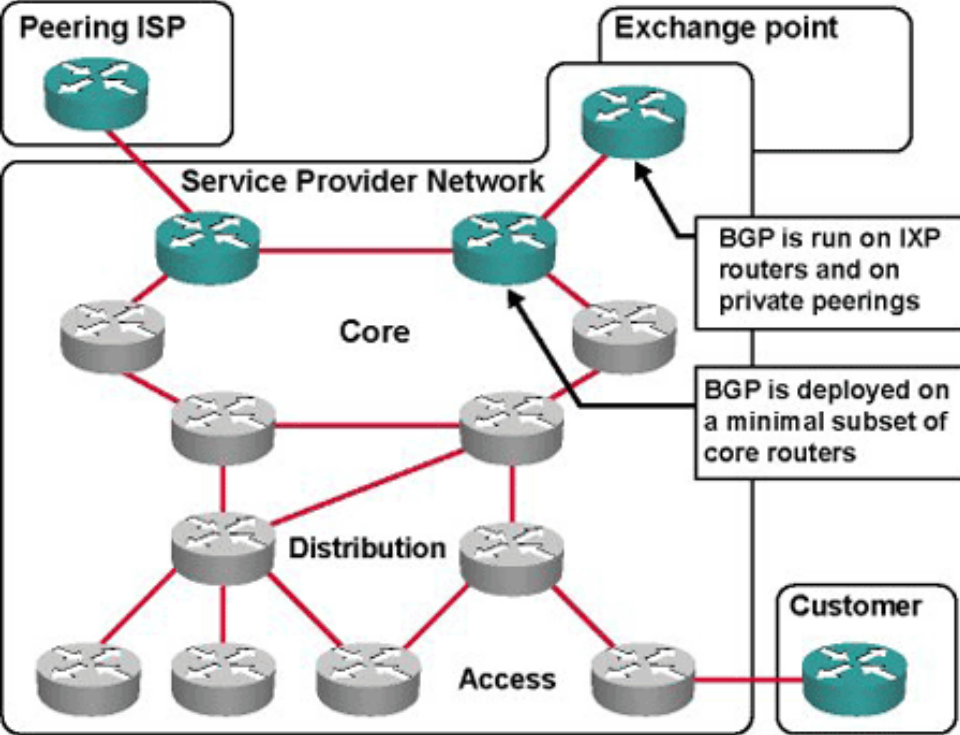


■ Input ■ Output

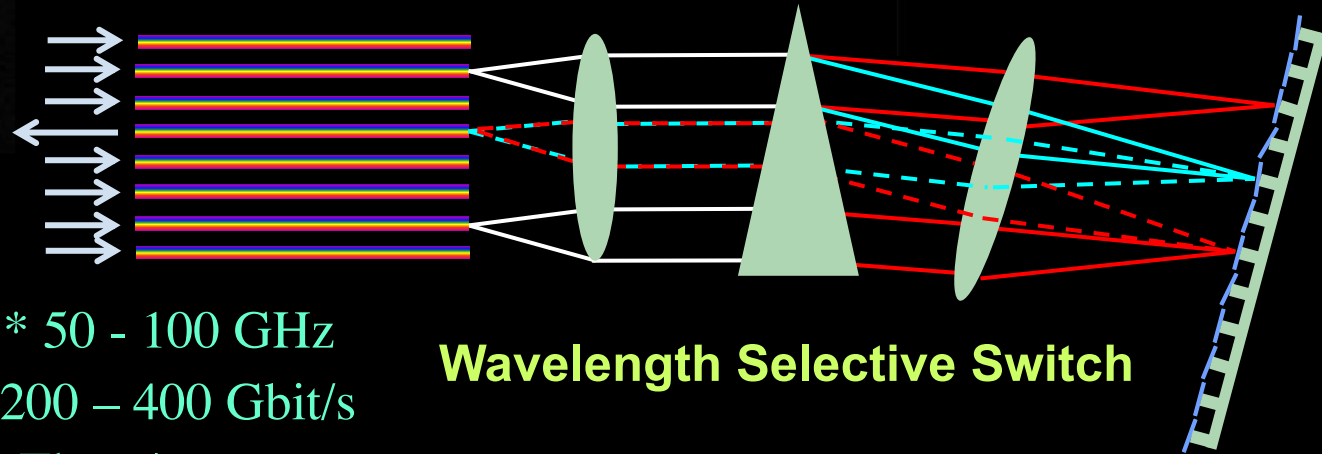
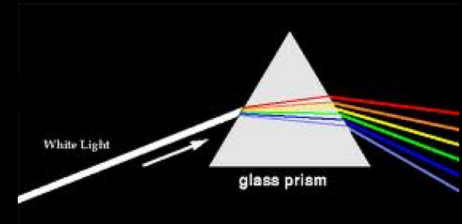
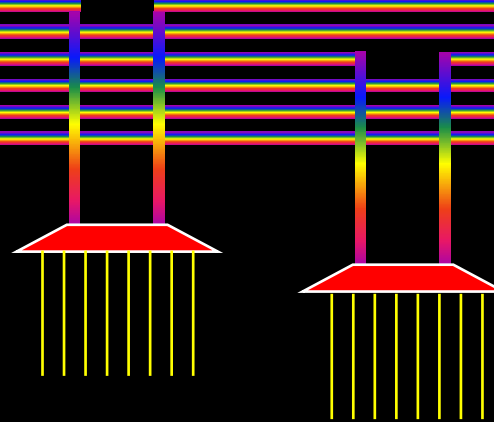
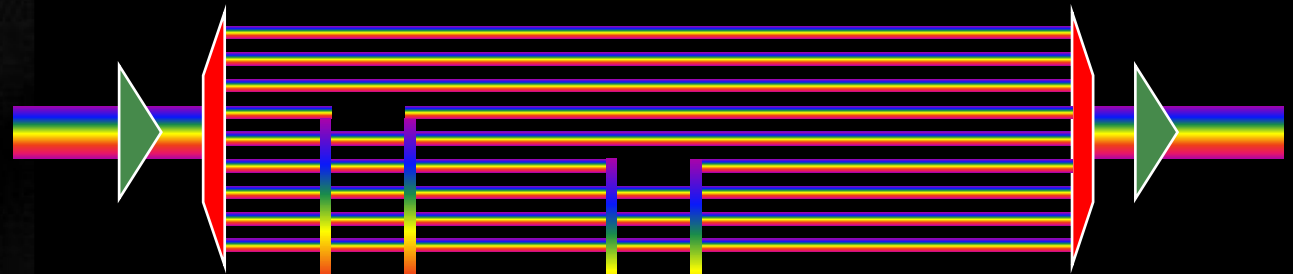
Peak In	: 4.587 Tb/s	Peak Out	: 4.578 Tb/s
Average In	: 3.256 Tb/s	Average Out	: 3.256 Tb/s
Current In	: 4.094 Tb/s	Current Out	: 4.091 Tb/s

Copyright (c) 2016 AMS-IX B.V. [updated: 09-Oct-2016 13:56:25 +0200]





# Multiple colors / Fiber



Per fiber: ~ 80-100 colors \* 50 - 100 GHz  
Per color: 10 - 40 - 100 - 200 - 400 Gbit/s  
Max total: ~20 Tbit/s = ~2 Tbyte/s

**Wavelength Selective Switch**

**New: Hollow Fiber!**

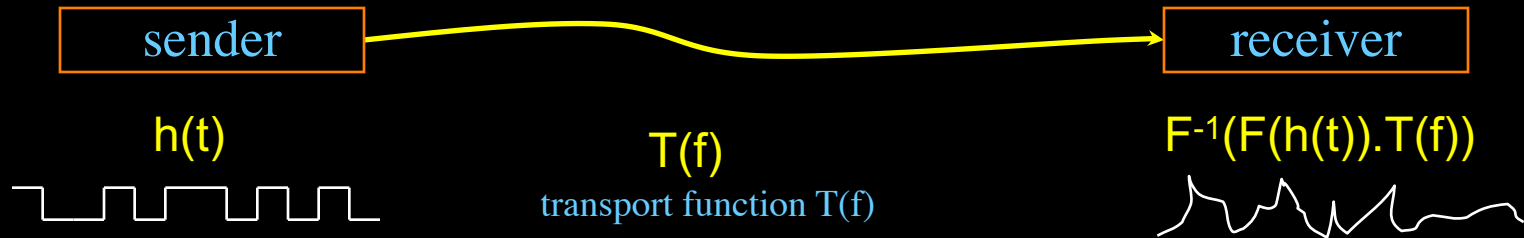
**→ less RTT!**





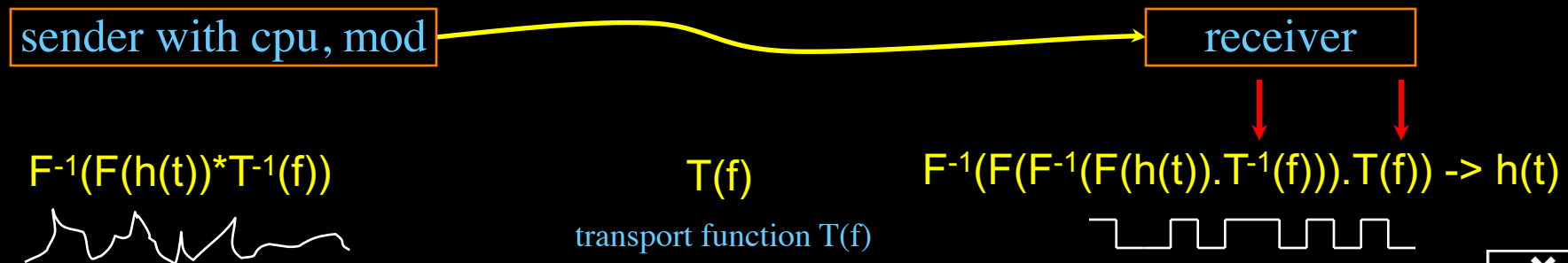
# Dispersion compensating modem: eDCO from NORTEL

(Try to Google eDCO :-)



Solution in 5 easy steps for dummy's :

- try to figure out  $T(f)$  by trial and error
- invert  $T(f) \rightarrow T^{-1}(f)$
- computationally multiply  $T^{-1}(f)$  with Fourier transform of bit pattern to be send
- inverse Fourier transform the result from frequency to time space
- modulate laser with resulting  $h'(t) = F^{-1}(F(h(t)).T^{-1}(f))$



(ps. due to power  $\sim$  square E the signal to send looks like uncompensated received but is not)



# Gridless colors.

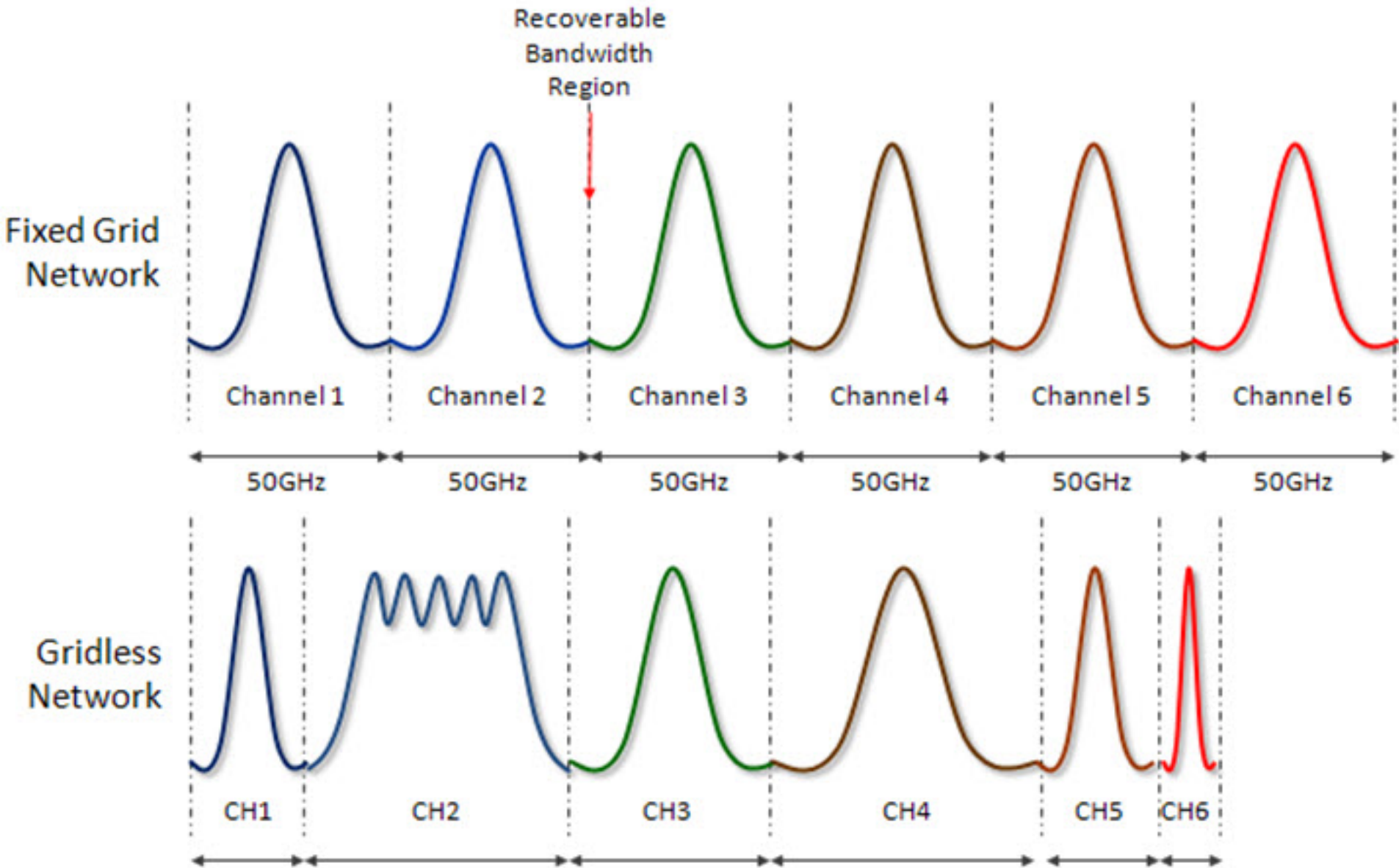
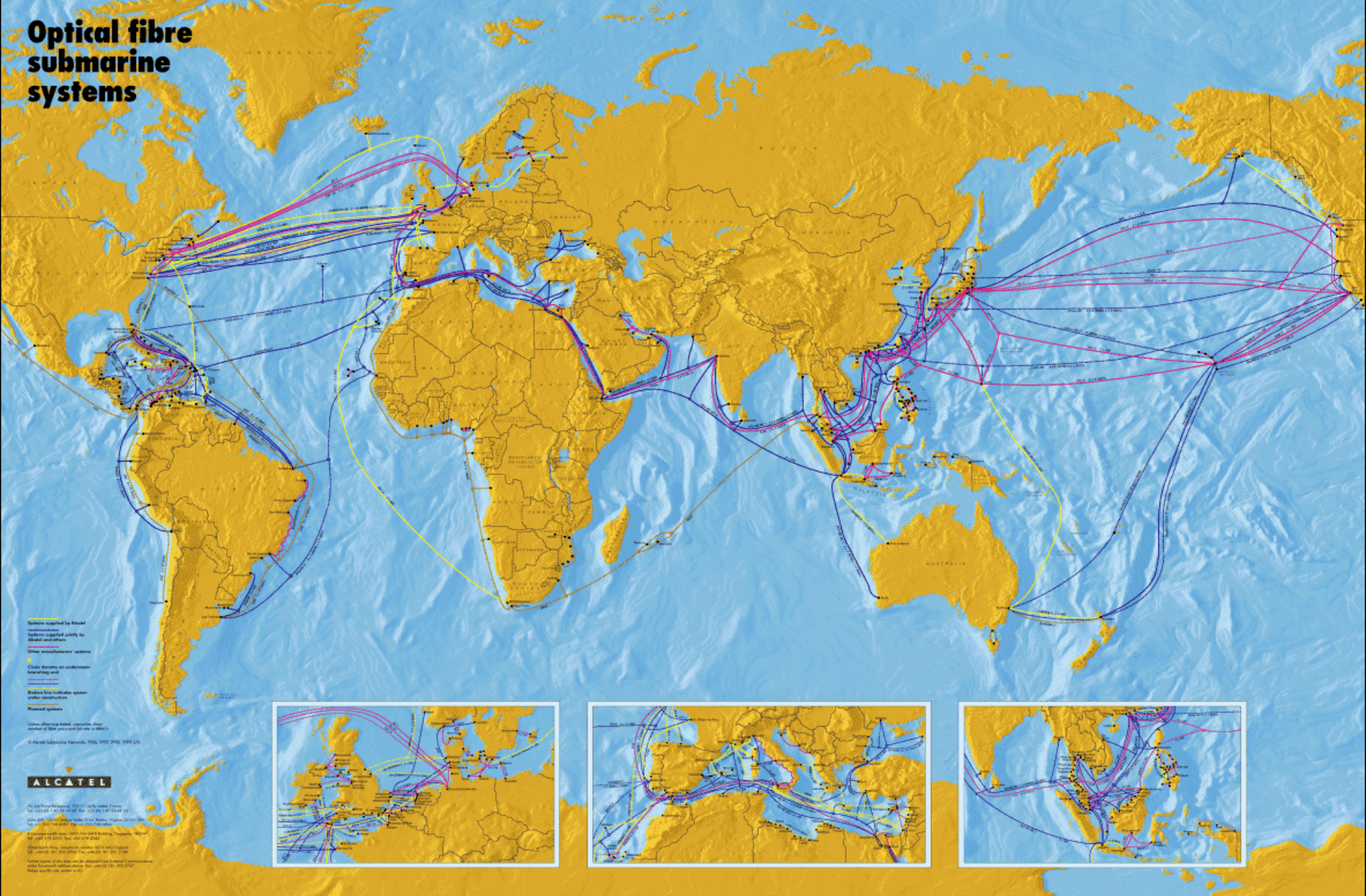


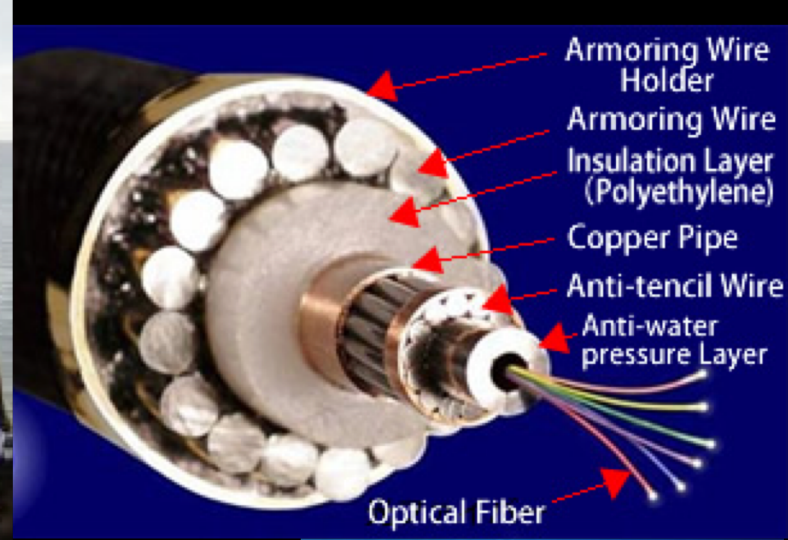
Figure 1 – Fixed Grid vs. Gridless Networking

# Optical fibre submarine systems



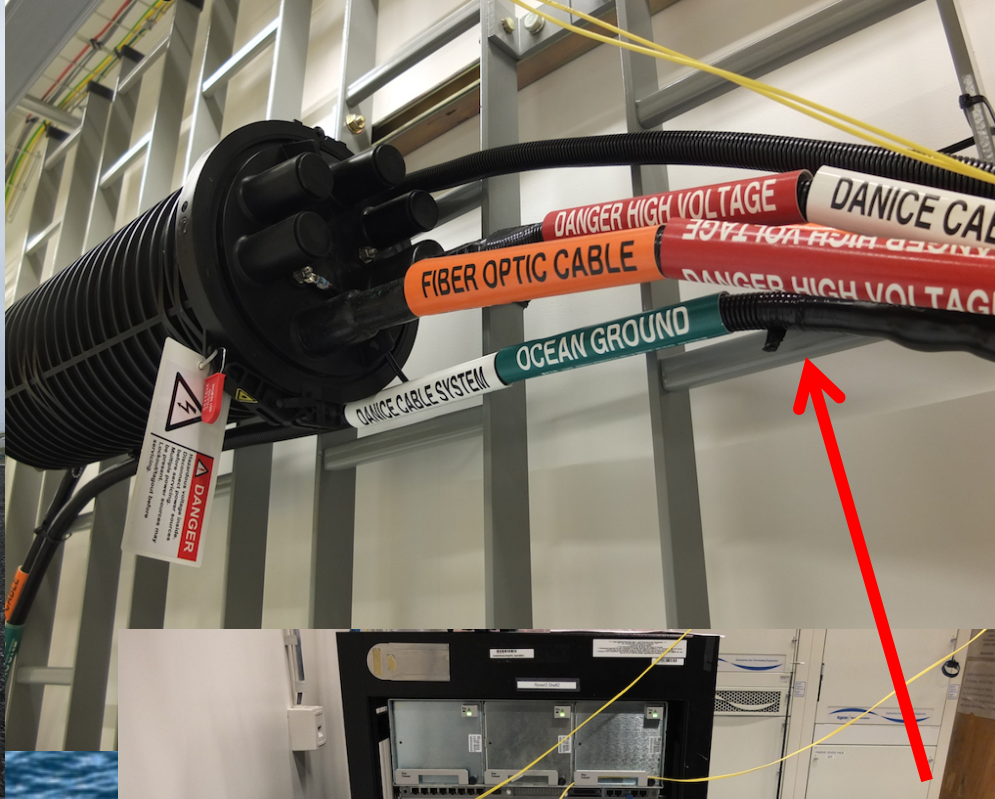
# Undersea Cable System





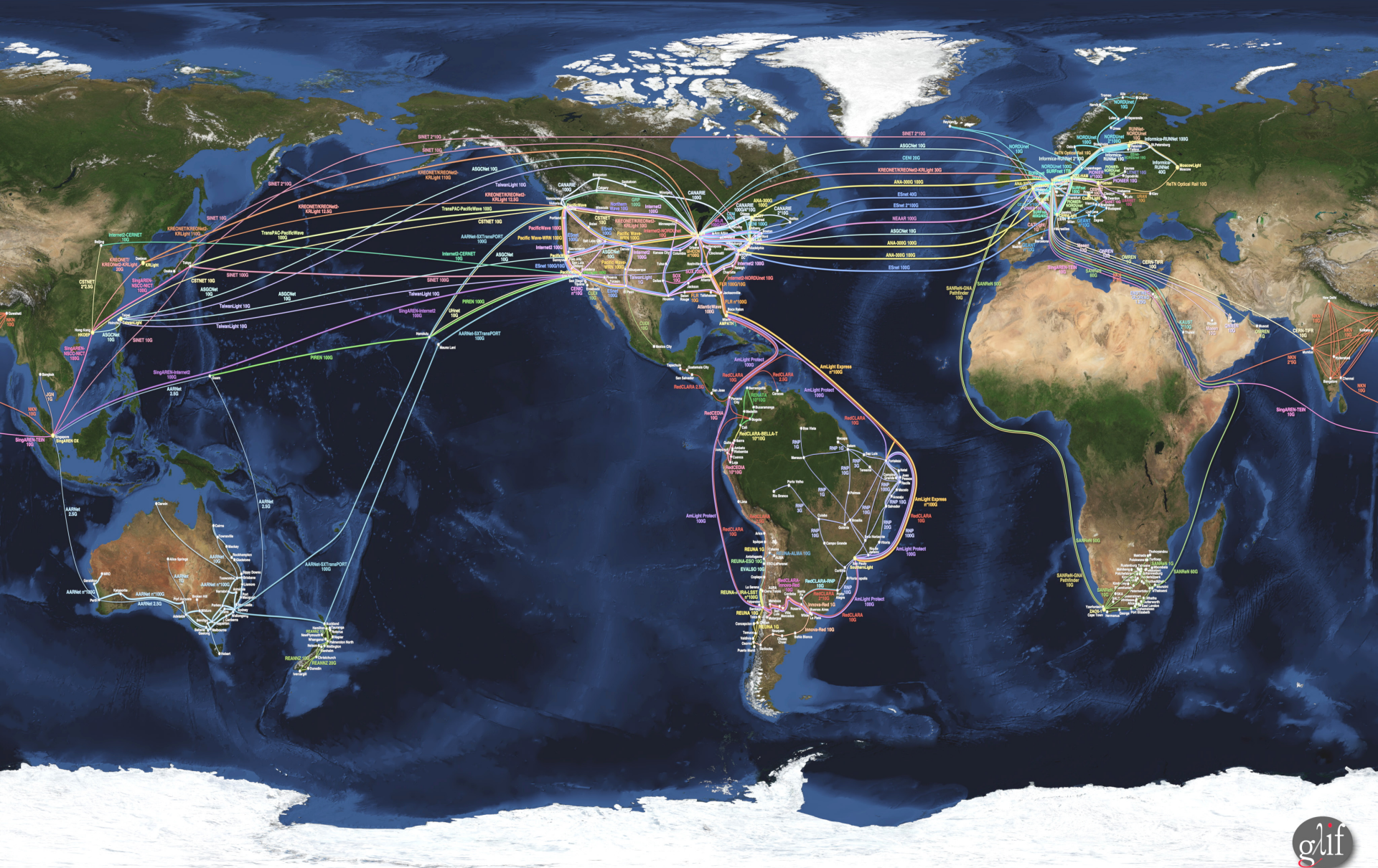
A **cable landing station** may or may not be required, depending on whether, for example, the submarine cable requires power to power submarine repeaters or amplifiers. The voltages applied to the cables can be high **3,000 to 4,000 volts** for a typical trans-Atlantic telecommunications cable system, and 1,000 volts for a cross-channel telecommunications cable system. Submarine power cables can operate at many kilovolts: for example, the [Fenno-Skan power cable operates at 400 kV DC](#).





# Undersea Cable HV





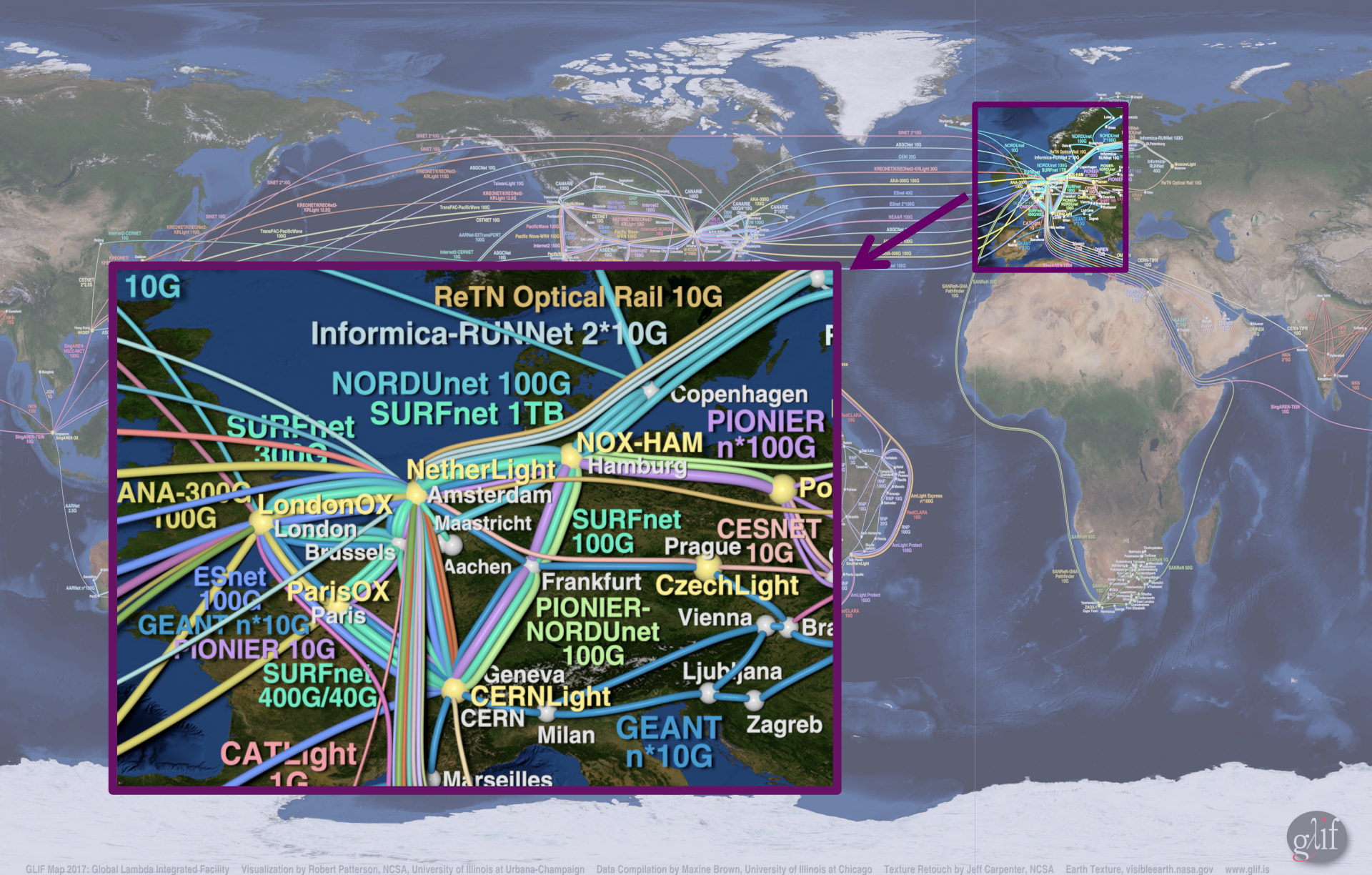
GLIF Map 2017: Global Lambda Integrated Facility Visualization by Robert Patterson, NCSA, University of Illinois at Urbana-Champaign Data Compilation by Maxine Brown, University of Illinois at Chicago Texture Retouch by Jeff Carpenter, NCSA Earth Texture, visibleearth.nasa.gov www.glif.is

F Dijkstra, J van der Ham, P Grosso, C de Laat, "A path finding implementation for multi-layer networks", Future Generation Computer Systems 25 (2), 142-146.



# The GLIF – LightPaths around the World





GLIF Map 2017: Global Lambda Integrated Facility Visualization by Robert Patterson, NCSA, University of Illinois at Urbana-Champaign Data Compilation by Maxine Brown, University of Illinois at Chicago Texture Retouch by Jeff Carpenter, NCSA Earth Texture, visibleearth.nasa.gov www.glif.io

F Dijkstra, J van der Ham, P Grosso, C de Laat, "A path finding implementation for multi-layer networks", Future Generation Computer Systems 25 (2), 142-146.

# The GLIF – LightPaths around the World



# AT&T verzamelt voor veel geld data voor inlichtingendiensten

**Riffy Bol**  
Amsterdam

**De Amerikaanse telecomprovider AT&T biedt Amerikaanse politie en inlichtingendiensten voor miljoenen dollars per jaar toegang tot een dienst om burgers te bespioneren. Voor deze software die klantgegevens analyseert, genaamd 'Hemisphere', hebben overheidsinstanties geen opsporingsbevel nodig. De belofte om Hemisphere niet te noemen in strafrechtelijk onderzoek volstaat, onthulde nieuwssite *The Daily Beast* dinsdag.**

Door zonder formeel arrestatiebevel klantgegevens aan overheidsinstanties te verstrekken, schendt het feitelijk de privacy van zijn ruim honderd miljoen klanten. Analisten van AT&T zoeken met Hemisphere naar verborgen patronen in de zogeheten metadata die het bedrijf van zijn klanten opslaat. Daarmee kunnen zij relaties tussen personen en hun verplaatsingen door de Verenigde Staten nauwkeurig bijhouden.

Telecombedrijven zijn verplicht om data af te geven als opsporingsdiensten daarom vragen. Maar AT&T handelt met de surveillancesoftware Hemisphere vooral uit commercieel oogpunt, zegt beleidsanalist Christopher Soghoian

van burgerrechtenbeweging ACLU tegen *The Daily Beast*. 'AT&T hoeft zijn database niet te datamineren om de politie te helpen aan nieuwe gevallen om te onderzoeken.' Een woordvoerder van AT&T zegt tegen *The Daily Beast* echter dat het bedrijf 'geen speciale database' bijhoudt voor de Amerikaanse overheid.

Politiedepartementen zouden 100 duizend tot één miljoen dollar per jaar betalen voor toegang tot Hemisphere. De bestuurlijke regio waarin de Texaanse stad Houston ligt zou tussen 2007 en 2011 ruim 900 duizend dollar aan de dienst hebben gependend, schrijft *The Daily Beast* op basis van een contract dat het heeft ingezien.

Federale en lokale agenten kunnen niet rechtstreeks bij de data; deze worden op afstand doorgelicht door werknemers van AT&T. Hemisphere zou in zeker 28 inlichtingencentra verspreid door de VS worden gebruikt.

AT&T zit op een immense hoeveelheid gegevens van Amerikaanse burgers. Het bezit meer dan driekwart van de Amerikaanse knooppunten voor vaste telefoonlijnen. En op concurrent Verizon na heeft AT&T het grootste marktaandeel in draadloze infrastructuur en telefoonmasten. Bovendien slaat de provider klantgegevens op die teruggaan tot juli 2008. Verizon bevaart deze naar verluidt slechts een jaar.

Hemisphere werd in 2013 al ontdekt door *The New York Times*, in een PowerPoint-presentatie van de Amerikaanse antinarcoticadienst DEA. Het product van AT&T, dat in 2007 is ontwikkeld, omschreef de krant destijds als een 'essentieel, voorzichtig toegepast gereedschap' in de strijd tegen drugs. *The Daily Beast* laat echter zien dat het telecombedrijf veel verder gaat dan bijklussen voor de Drug Enforcement Agency.

Zo werd de moord op een familie uit Californië opgelost toen gegevens van AT&T vaststelden dat de verdachte op de plaats delict was, twee dagen nadat het gezin van vier vermist was. De telefoon van Charles Merritt maakte contact met een telefoonmast iets ten noordoosten van de vindplaats van de familie McStay.

Een deel van de activiteiten van AT&T's programma blijft in nevelen gehuld. Het gevaar van deze geheimzinnigheid is volgens Adam Schwartz van de Electronic Frontier Foundation, dat data die AT&T levert aan inlichtingendiensten, niet als bewijs opgevoerd kunnen worden in de rechtszaal. Gedagden hebben het recht om te weten waarvan zij worden verdacht en hoe dat bewijs is gewonnen.

Schwartz stelt dat de politie mogelijk eerst het bewijs van AT&T bekijkt, om datzelfde bewijs vervolgens op een andere manier zelf te verzamelen.



# Revelations on:



name	orig	partners	purpose
Xkeyscore	USA	D, S	searching and analyzing global Internet data
PRISM	USA	AU, UK, NL	collect info from Micro\$oft, Google, Facebook, Apple
ECHELON	USA	5Y	global network to eavesdrop on telephones, faxes and computers, bank accounts
Carnivore	USA		Monitor electronic communications using customizable packet sniffer on target user's Internet
DISHFIRE	USA	UK	covert global surveillance collection system and database
Stone Ghost	USA		information sharing and exchange between the United States, the United Kingdom, Canada and Australia
Tempora	UK	USA	Telcos: BT, Interoute, L3, Global Crossing, Verizon, Viatel, Vodafone cable
MUSCULAR	UK	USA	records from internal Yahoo! and Google
Frenchelon	FR		French global network to eavesdrop on telephones, faxes and computers, bank accounts
Fairview	USA	AT&T	collect phone, internet and e-mail data of foreign countries' citizens at major cable landing stations and switching stations inside the United States
MYSTIC	USA		collect the metadata as well as the content of phone calls from several entire countries
DCSN	USA	FBI	surveillance system to perform instant wiretaps on almost any telecommunications device in the US
Boundless Informant	USA		a big data analysis and data visualization tool
BULLRUN	USA		to crack encryption of online communications and data (UK -> Edgehill)
PINWALE	USA		Digital Network Intelligence, including internet e-mail
Stingray	USA		IMSI-catcher, cellular phone surveillance device, manufactured by Harris Corporation
LOVEINT	USA		Spying on colleague's, spouses ☺

- buffer most Internet communications that are extracted from fibre-optic cablesCollate online and telephone traffic
- Data from fibre-optic cable communications.
- Data is preserved for three days, metadata for thirty days.
- By May 2012 300 GCHQ analysts and 250 NSA analysts had been assigned to sort data.[4]
- About 850,000 people have security clearance to access the data.
- Tempora said to include recordings of telephone calls, content of email messages, Facebook entries and personal internet history of users.
- Snowden said of Tempora that "It's not just a U.S. problem. "They [GCHQ] are worse than the U.S."
- Dutch programs (e.g. iColumbo,
- [https://www.flusso.nl/Werk/big\\_data-nationalepolitie/](https://www.flusso.nl/Werk/big_data-nationalepolitie/) )



# TEMPORA





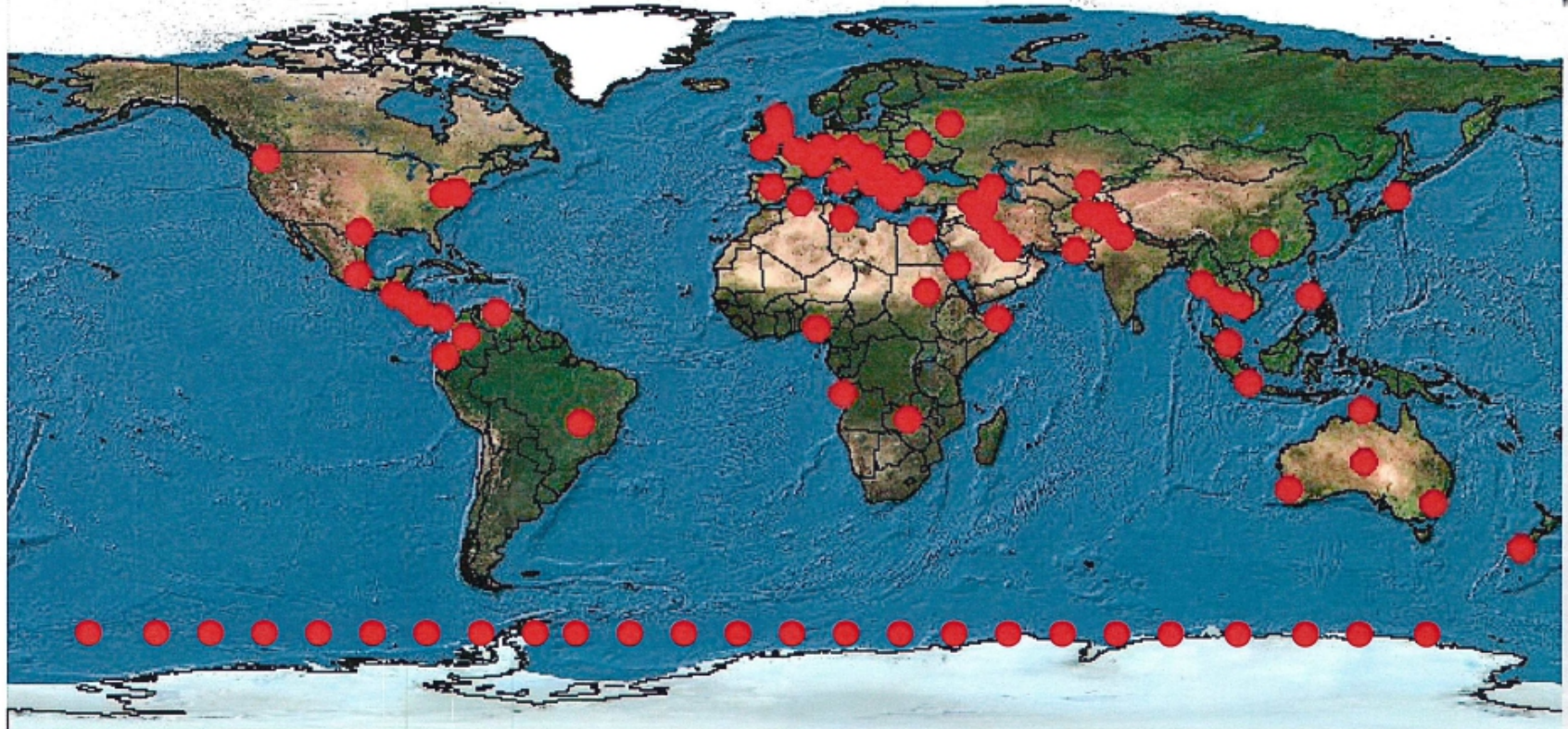
# What is XKEYSCORE?

1. DNI Exploitation System/Analytic Framework
  2. Performs strong (e.g. email) and soft (content) selection
  3. Provides real-time target activity (tipping)
  4. "Rolling Buffer" of ~3 days of ALL unfiltered data seen by XKEYSCORE:
    - Stores full-take data at the collection site – indexed by meta-data
    - Provides a series of viewers for common data types
1. Federated Query system – one query scans all sites
    - Performing full-take allows analysts to find targets that were previously unknown by mining the meta-data





# Where is X-KEYSCORE?



Approximately 150 sites

Over 700 servers

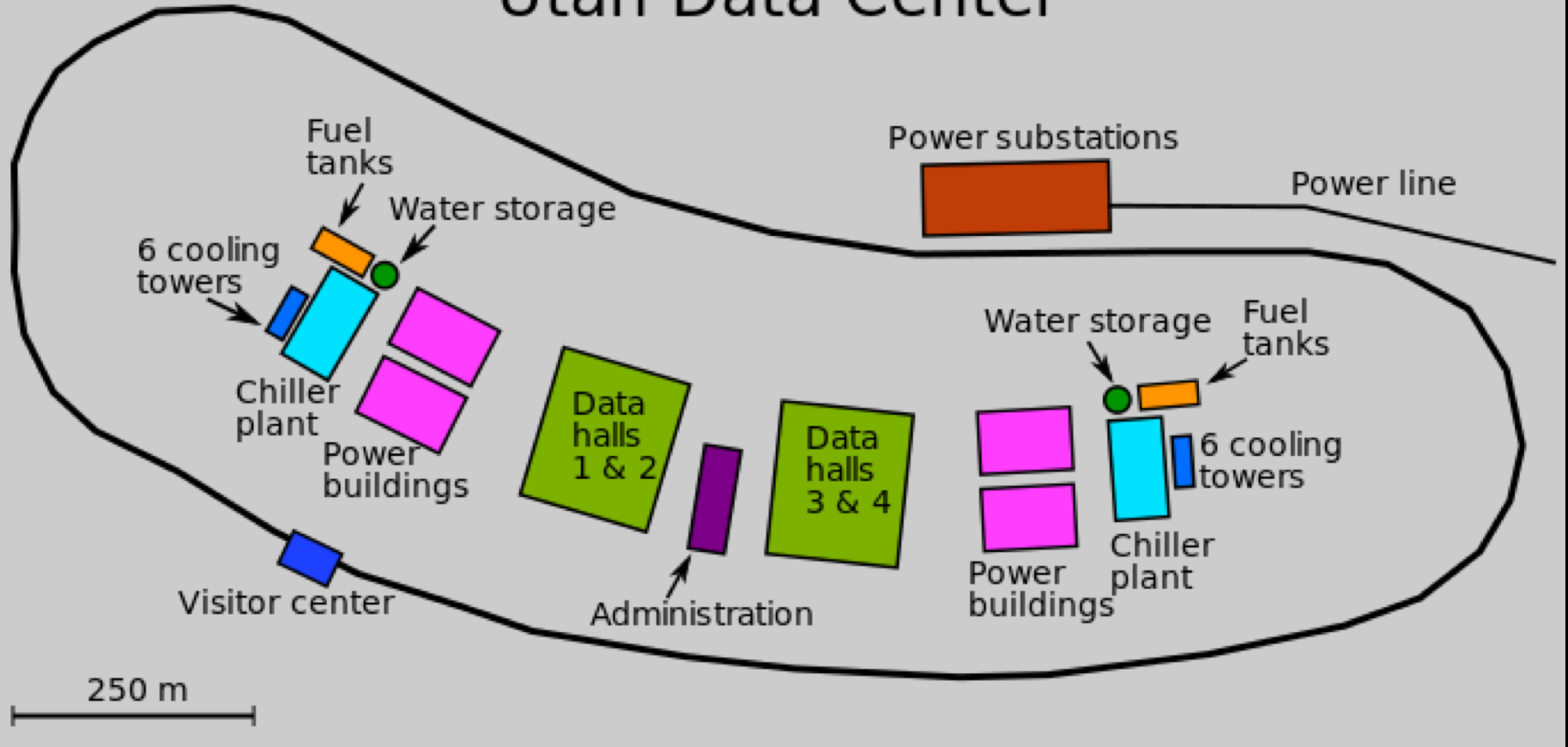


# Utah data center



# Utah data center

## Utah Data Center



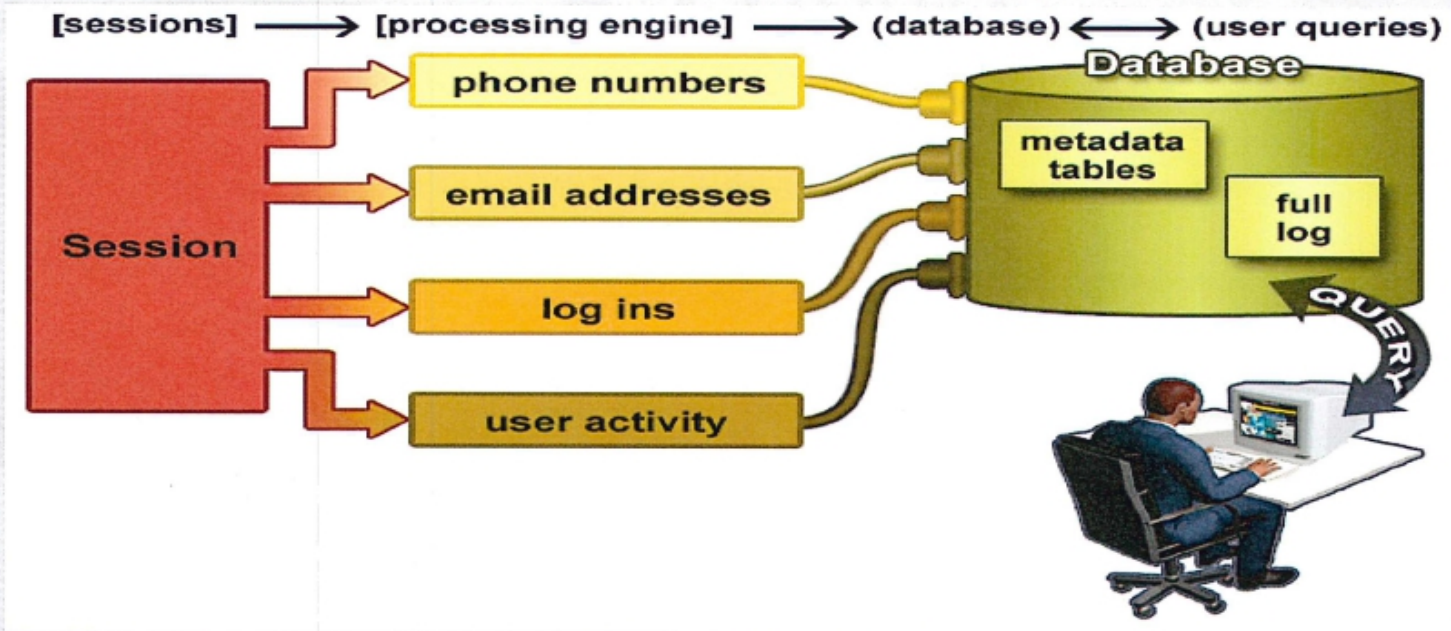
- capable of storing data on the scale of yottabytes (1 yottabyte = 1 trillion terabytes, or 1 quadrillion gigabytes or  $10^{24}$ ).
- Its purpose – as the name implies – is to support the Comprehensive National Cybersecurity Initiative (CNCI), including storing details of people's mobile phone and internet use, though its precise mission is secret
- The planned structure is about 1.25 million square feet and cost \$1.5 billion when it was completed in May 2014.
- One report suggested that it also cost another \$2 billion for hardware, software, and maintenance.
- The facility is estimated to have a power demand of 65 megawatts, costing about \$40 million per year.





# What XKS does with the Sessions

Plug-ins extract and index metadata into tables





# Plug-ins

Plug-in	DESCRIPTION
E-mail Addresses	Indexes every E-mail address seen in a session by both username and domain
Extracted Files	Indexes every file seen in a session by both filename and extension
Full Log	Indexes every DNI session collected. Data is indexed by the standard N-tuple (IP, Port, Casenotation etc.)
HTTP Parser	Indexes the client-side HTTP traffic (examples to follow)
Phone Number	Indexes every phone number seen in a session (e.g. address book entries or signature block)
User Activity	Indexes the Webmail and Chat activity to include username, buddylist, machine specific cookies etc.







# What Can Be Stored?

- Anything you wish to extract
  - Choose your metadata
  - Customizable storage times
  - Ex: HTTP Parser

```
GET /search?hl=en&q=islamabad&meta= HTTP/1.0
Accept: image/gif, image/x-bitmap, image/jpeg, image/pjpeg, application/vnd.ms-
application/msword, application/x-shockwave-flash, */*
Referer: http://www.google.com.pk/
Accept-Language: en-us
User-Agent: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)
Host: www.google.com.pk
```

No username/strong selector

```
Connection: keep-alive
```



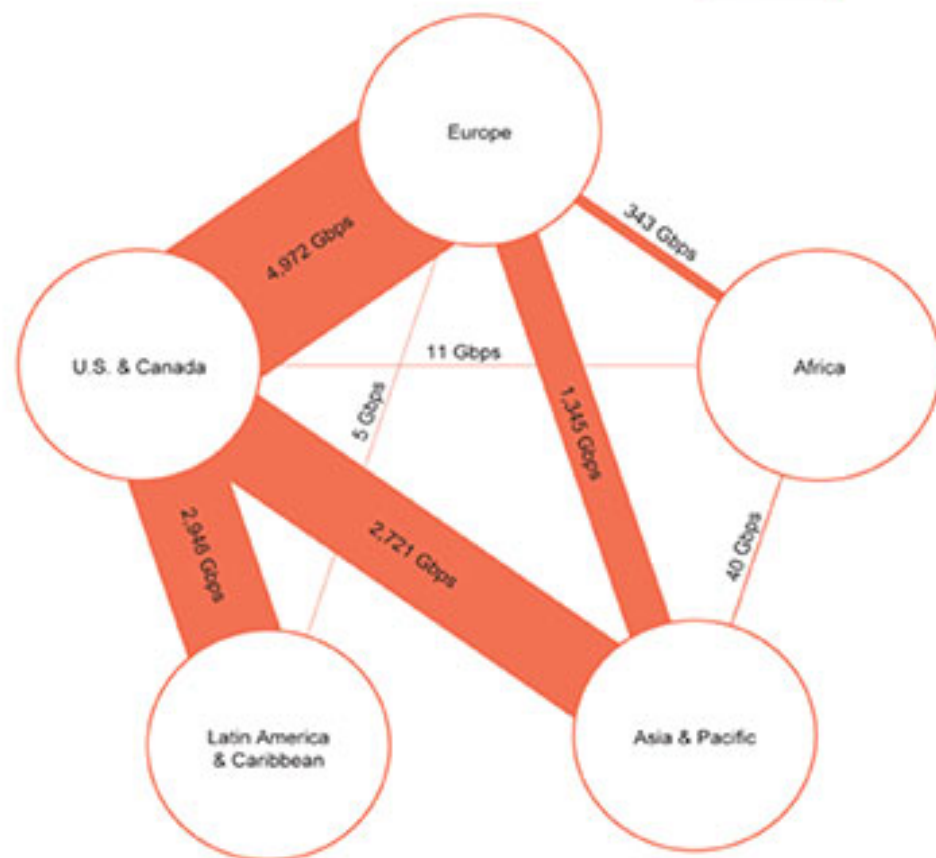


(TS//SI//NF) Introduction

*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



Hotmail

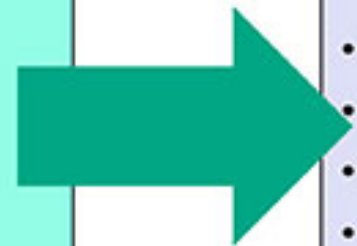


# (TS//SI//NF) PRISM Collection Details



## 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





facebook

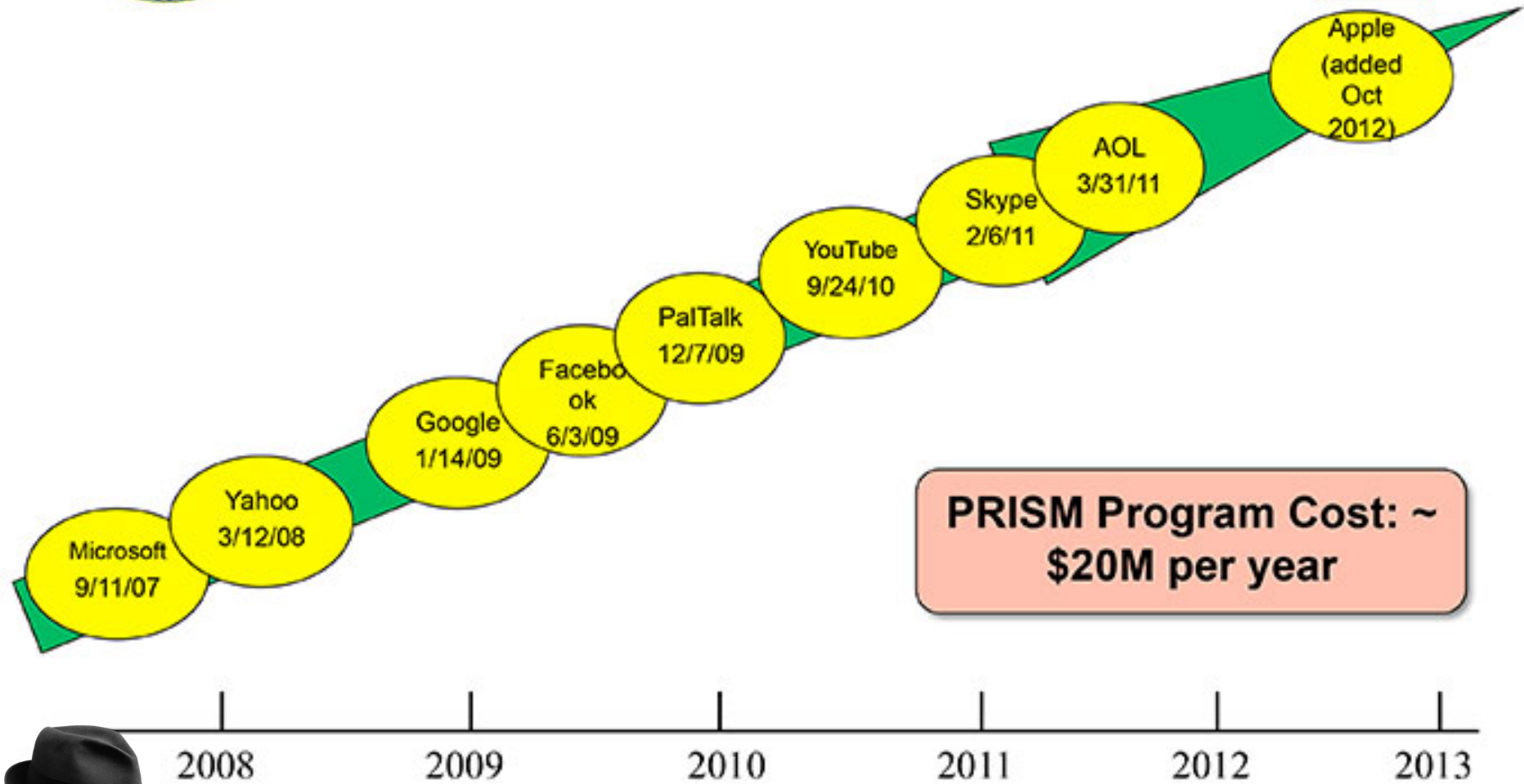


Hotmail

YAHOO!



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



**PRISM Program Cost: ~ \$20M per year**



2008

2009

2010

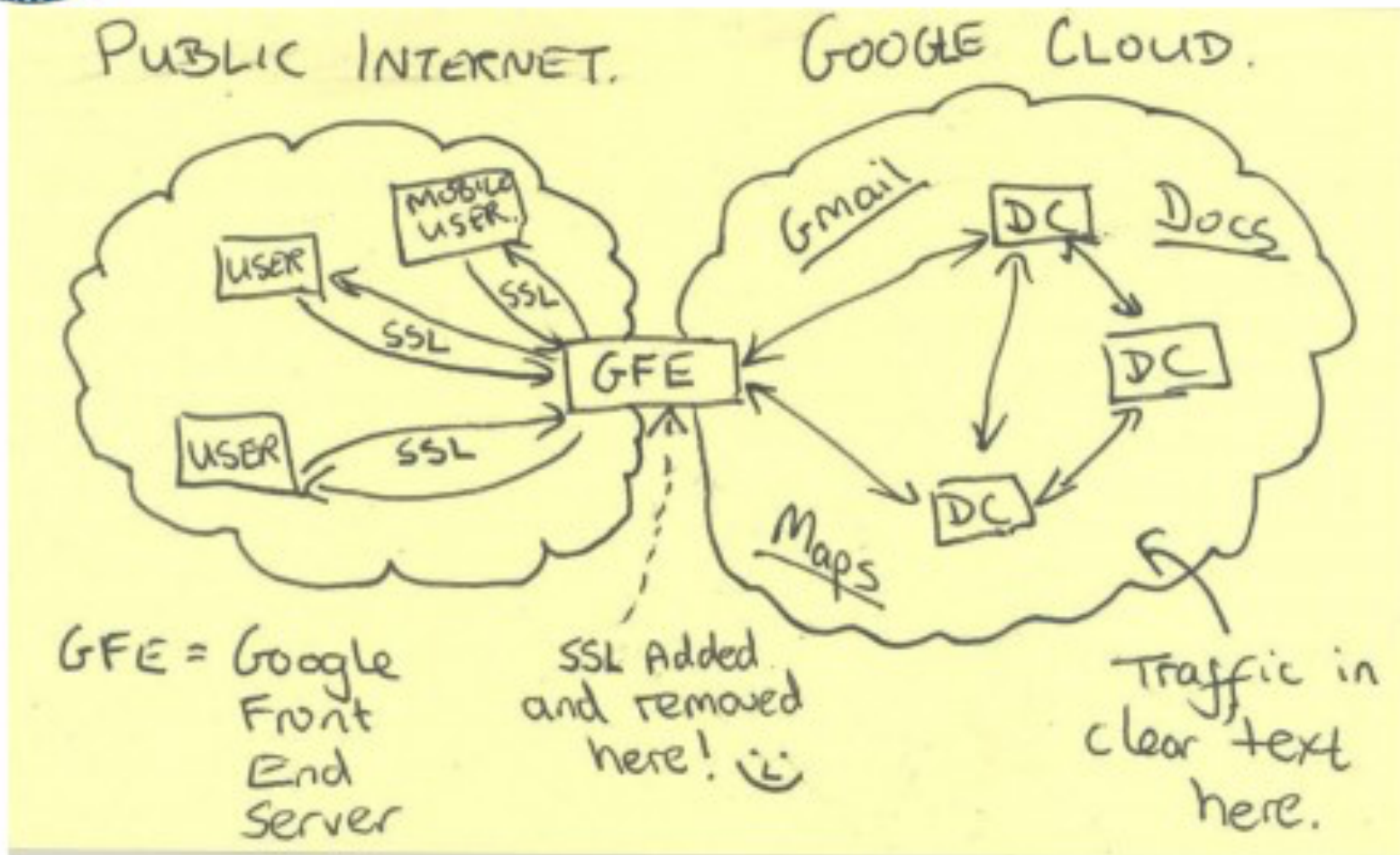
2011

2012

2013

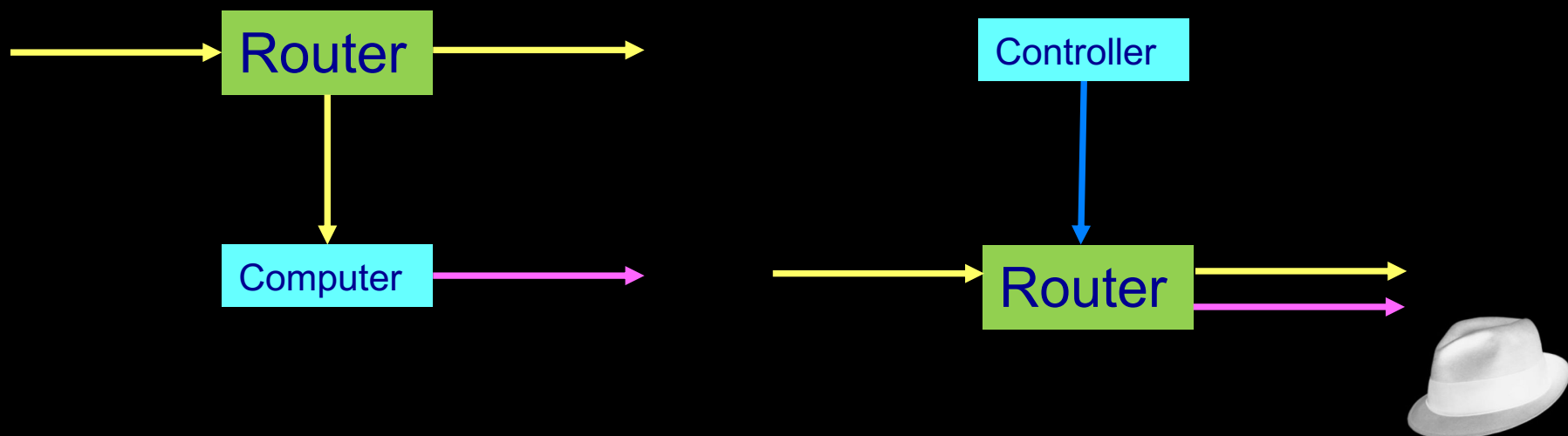


# Current Efforts - Google



# New Internet Technology

- SDN, NFV, OpenFlow
- Decoupling logic from forwarding plane
- Rules that encode in forwarding plane TCAM's
  - Ternary Content Addressable Memory





CHAPPATTE International New York Times



# Encryption?

The NSA follows specific procedures to target non-U.S. persons and to minimize data collection from U.S. persons.

These court-approved policies allow the NSA to:

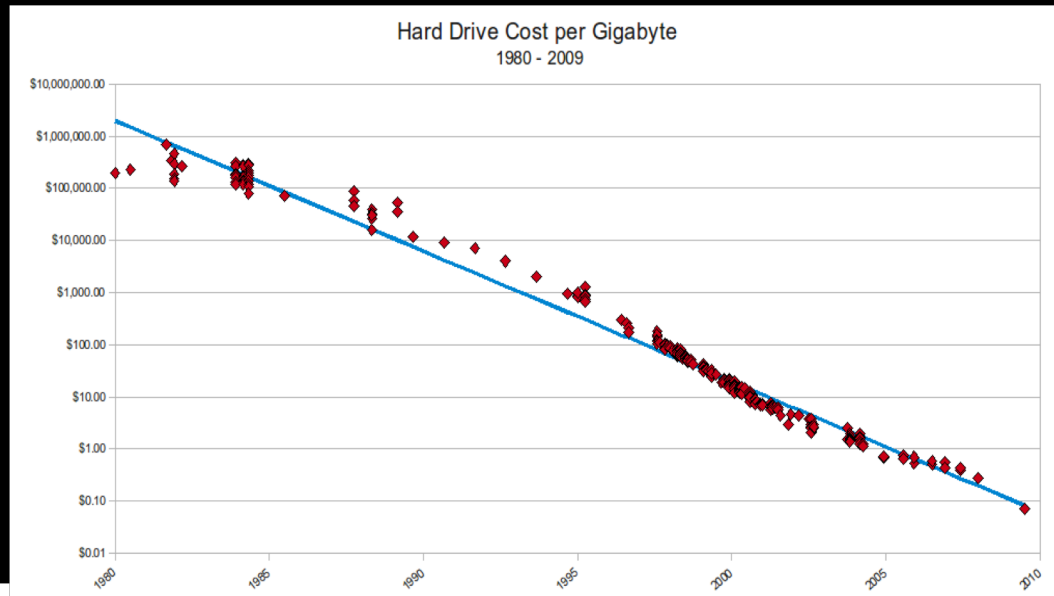
- keep data that could potentially contain details of U.S. persons for up to five years;
- **retain** and make use of "inadvertently acquired" domestic communications if they contain usable intelligence, information on criminal activity, threat of harm to people or property, are **encrypted**, or are believed to contain any information relevant to cybersecurity;
- preserve "foreign intelligence information" contained within [attorney–client communications](#)
- access the content of communications gathered from "U.S. based machine[s]" or phone numbers in order to establish if targets are located in the U.S., for the purposes of ceasing further surveillance.



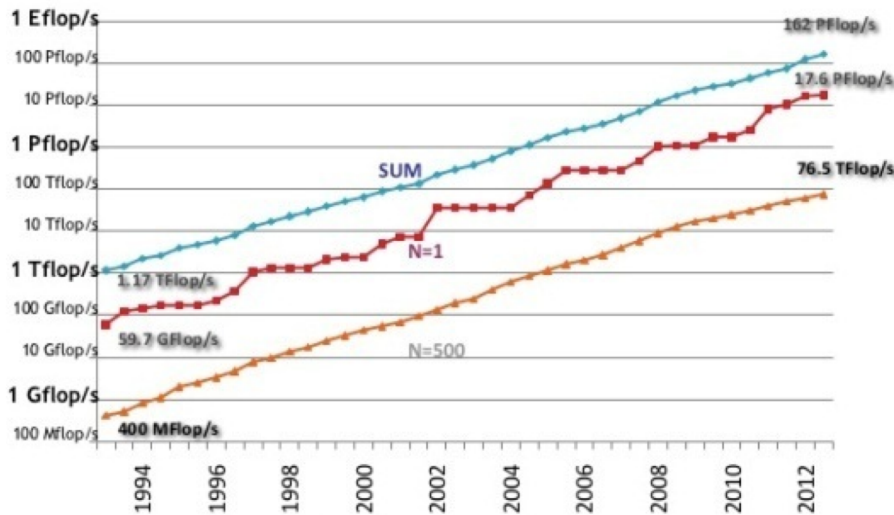


# Computing vs Data

Computing per unit cost has doubled roughly every 18 months (Moore's law).



## Performance Development



Space per unit cost has doubled roughly every 14 months (Kryder's law).

So: data becomes exponentially uncomputable.



# NSA seeks to build quantum computer that could crack most types of encryption



## HOW BAD IS IT?

If you take the development of serious quantum computing power as a given, all of the encryption methods based on factoring primes or doing modular exponentials, most notably RSA, elliptic curve cryptography, and Diffie-Hellman are all in trouble. Specifically, Shor's algorithm, when applied on a quantum computer, will render the previously difficult math problems that underlie these methods trivially easy almost irrespective of chosen key length. That covers most currently used public-key crypto and the key exchange that's used in negotiating an SSL connection.

Post Quantum encryption

- [https://en.wikipedia.org/wiki/Post-quantum\\_cryptography](https://en.wikipedia.org/wiki/Post-quantum_cryptography)
- [https://www.schneier.com/blog/archives/2016/07/googles\\_post-qu.html](https://www.schneier.com/blog/archives/2016/07/googles_post-qu.html)
- <https://eprint.iacr.org/2015/1092.pdf>

<http://hackaday.com/2015/09/29/quantum-computing-kills-encryption/>



Note the difference



# Fact & Fiction

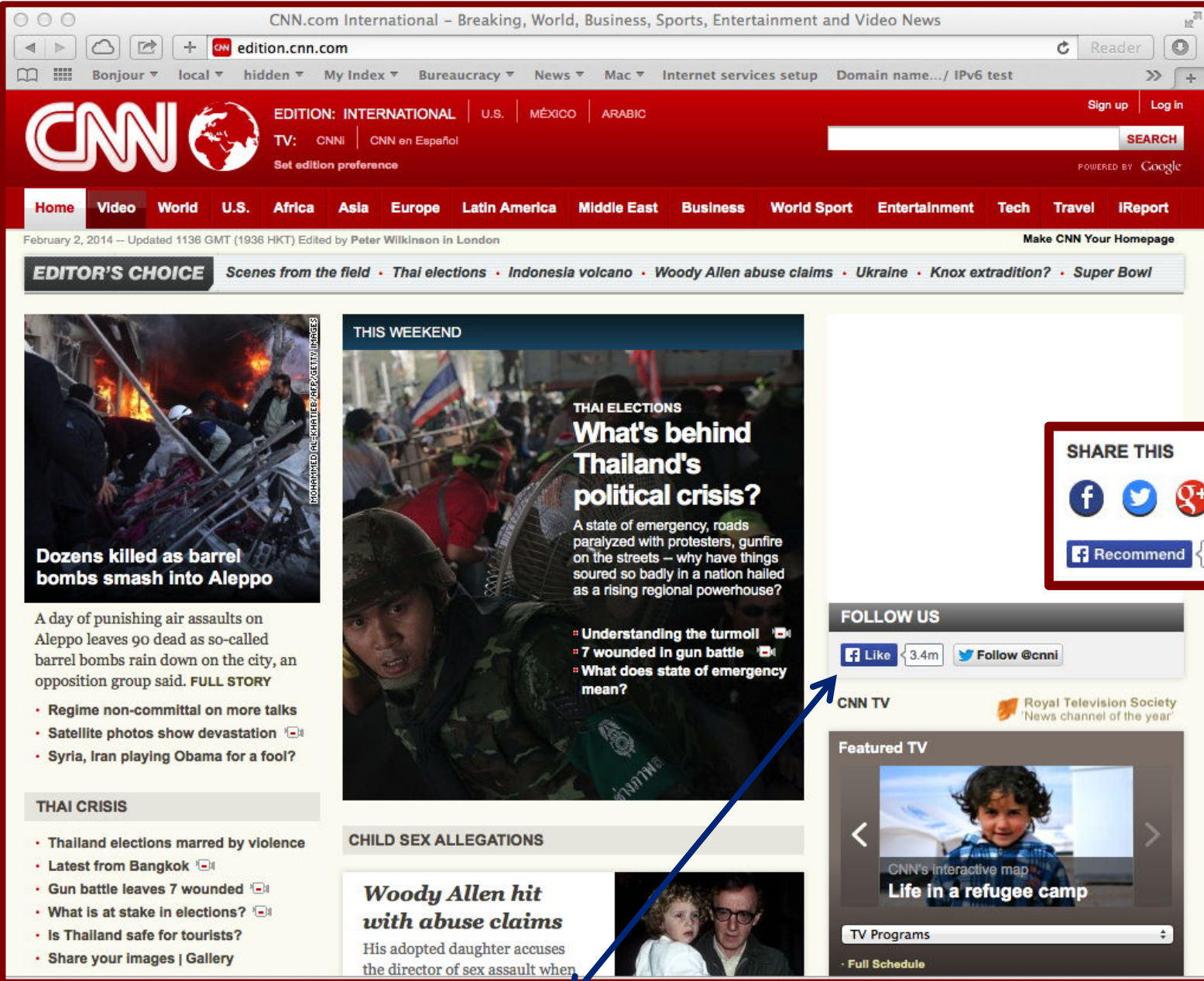
- Claims that BT pre-cooks adsl modems to send information from home networks to NSA and British Intelligence
  - <http://cryptome.org/2013/12/Full-Disclosure.pdf>
  - Modem connects to specific IP addresses at boot time
- Critical responses:
  - DOD uses lots of address space that is not publicly routed
    - <http://blog.erratasec.com/2013/12/dod-address-space-its-not-conspiracy.html>
  - See also the comment:  
"lucent uses 152.148.0.0/16 for 'management' on lots of their old big telco iron as if it was RFC-1918 space. (...)"
  - Also BT-competitor AAISP claims this is FUD:
    - <https://s.aa.net.uk/1871>
  - Claims: "They use DOD space because it's not internet-routable, and it's for the TR-069 ( <http://en.wikipedia.org/wiki/TR-069> ) service. This is \*NOT\* news."
    - <http://www.bit-tech.net/news/hardware/2013/12/17/bt-back-door/1>



# What has this to do with the National Science quiz 2013?

- Q13: For an illness that 1 out of 1000 people suffer, a 99% accurate test is developed. You are tested with that method and found bearer of the illness. What is the probability that you really have the specific illness?
- Choose: [ A: 99%, B: 50%, C: 9% ]
- Answer C: because you are in the set of true and false positives!
- Suppose the accuracy of PRISM, Tempora, Xkeyscore, etc. is 99% and 1 out of 100000 of the subjects are indeed terrorists
- False positives among 100k ... ~1000 !
- Send in the drones: <http://www.businessinsider.com/nsa-cia-drone-program-2013-10?international=true&r=US&IR=T>





I will follow you!



```
<iframe
src="//www.facebook.com/plugins/like.php?href=http%3A%2F%2Fwww.facebook.com%2Fcnni
nternational&send=false&layout=button_count&show_faces=false&action=like&colorscheme=light&font=arial&height=21" ...></iframe>
```

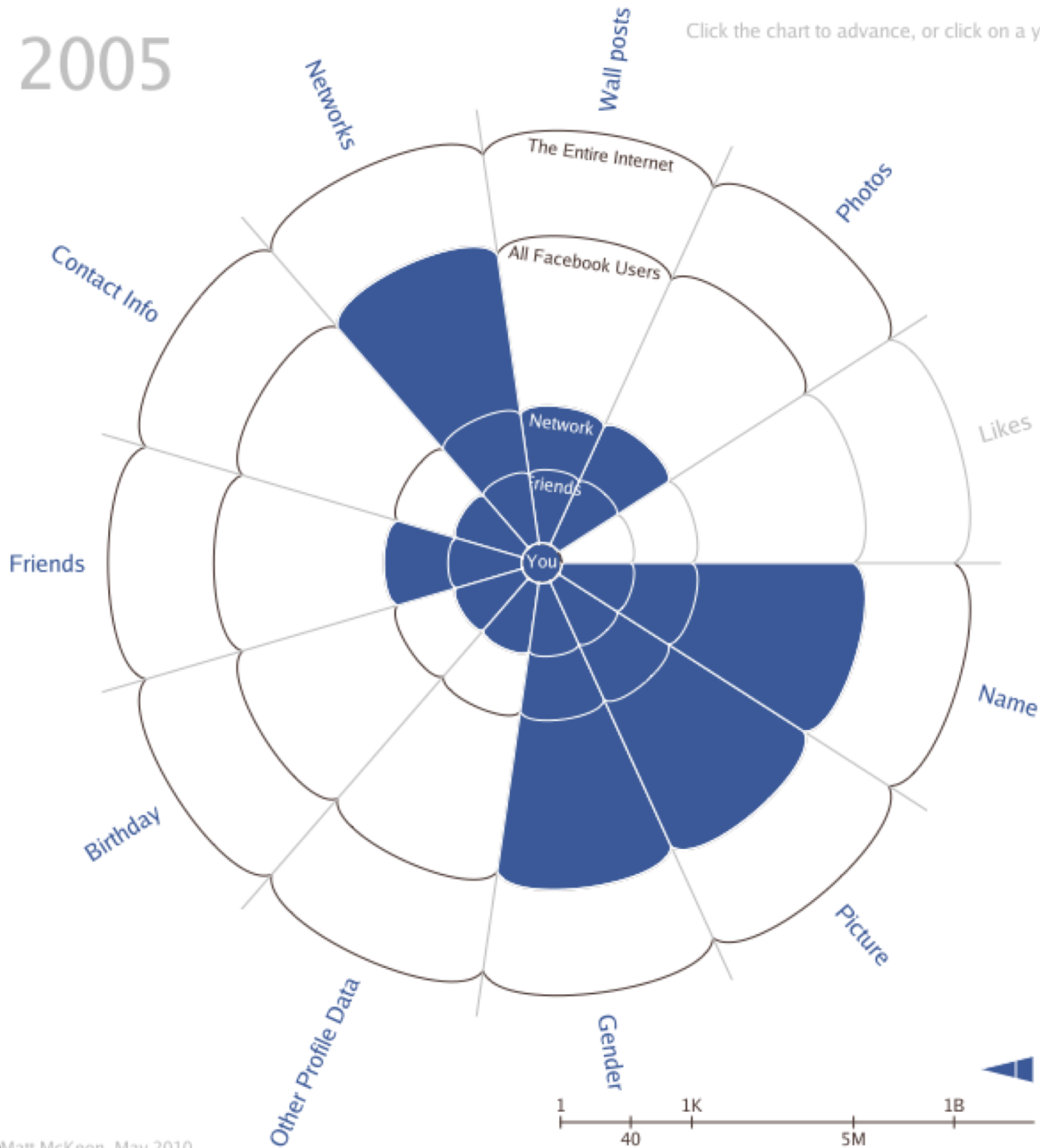


2005

Click the chart to advance, or click on a year

2005

- 2006
- 2007
- 2009 (Nov)
- 2009 (Dec)
- 2010 (Apr)



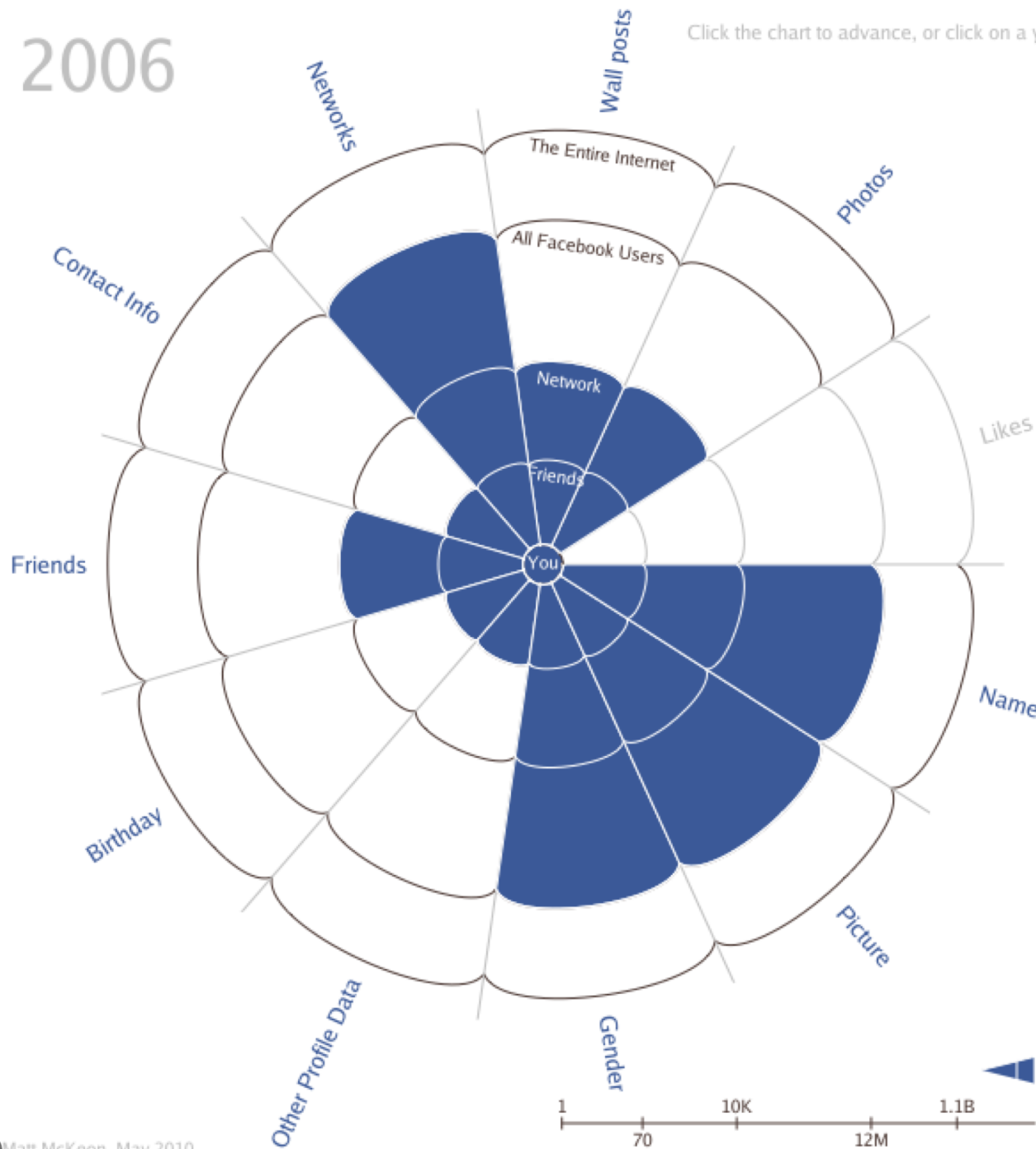
▲ Availability of your personal data on Facebook (default settings)  
Number of People



2006

Click the chart to advance, or click on a year

- 2005
- 2006**
- 2007
- 2009 (Nov)
- 2009 (Dec)
- 2010 (Apr)



▲ Availability of your personal data on Facebook (default settings)  
Number of People

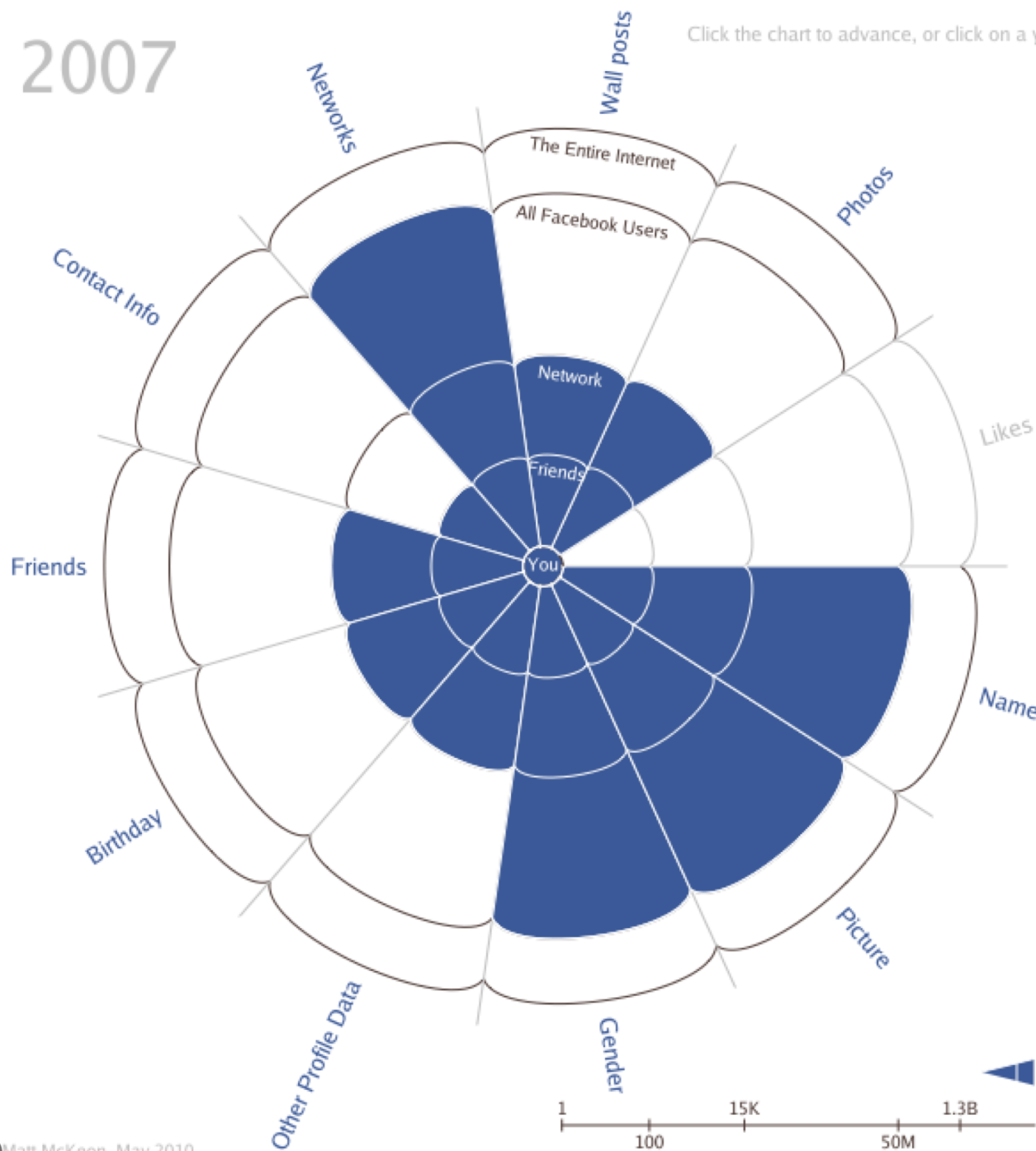
1 70 10K 12M 1.1B



# 2007

Click the chart to advance, or click on a year

- 2005
- 2006
- 2007**
- 2009 (Nov)
- 2009 (Dec)
- 2010 (Apr)



▲ Availability of your personal data on Facebook (default settings)  
Number of People

1 100 15K 50M 1.3B

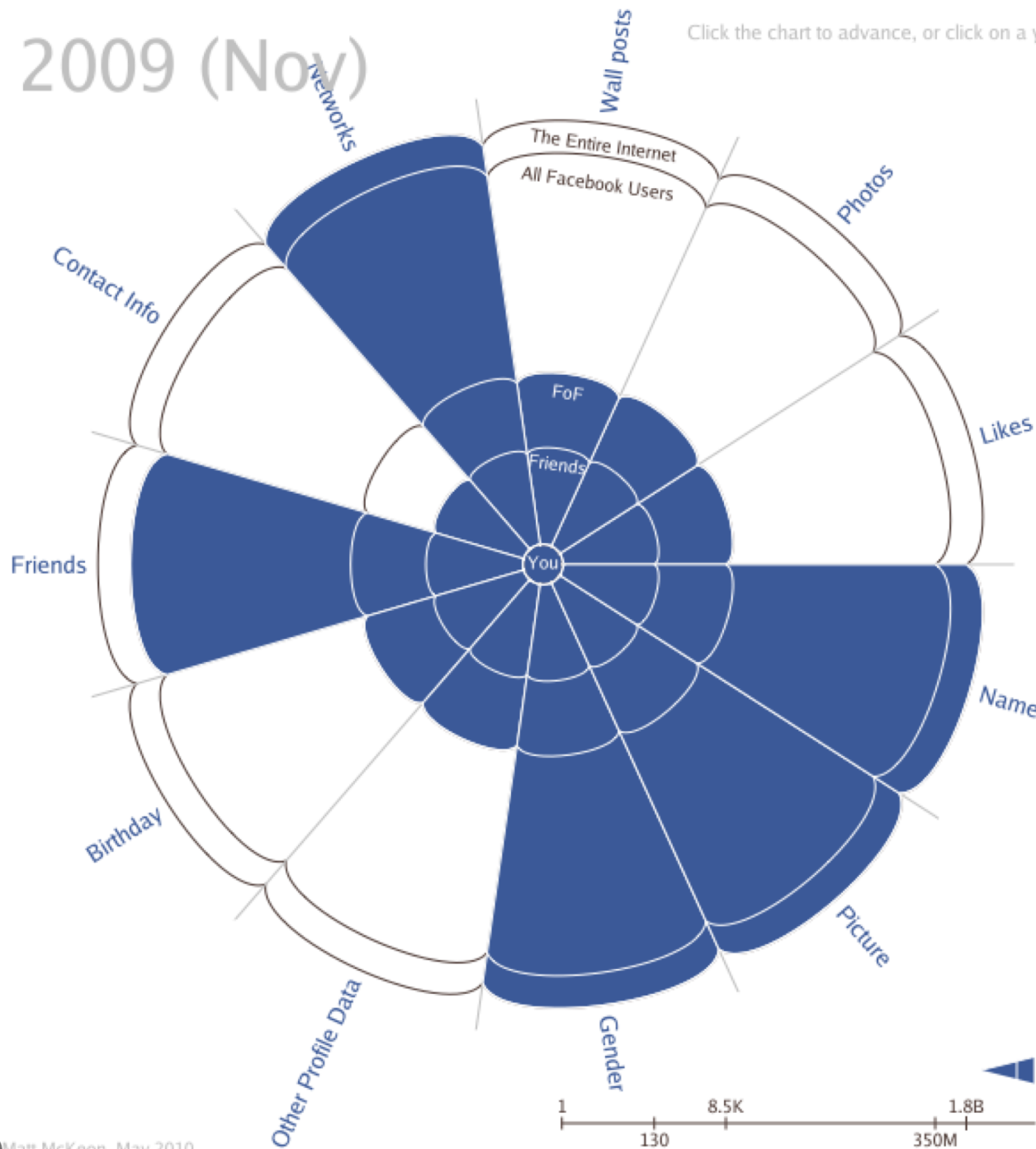




# 2009 (Nov)

Click the chart to advance, or click on a year

- 2005
- 2006
- 2007
- 2009 (Nov)**
- 2009 (Dec)
- 2010 (Apr)



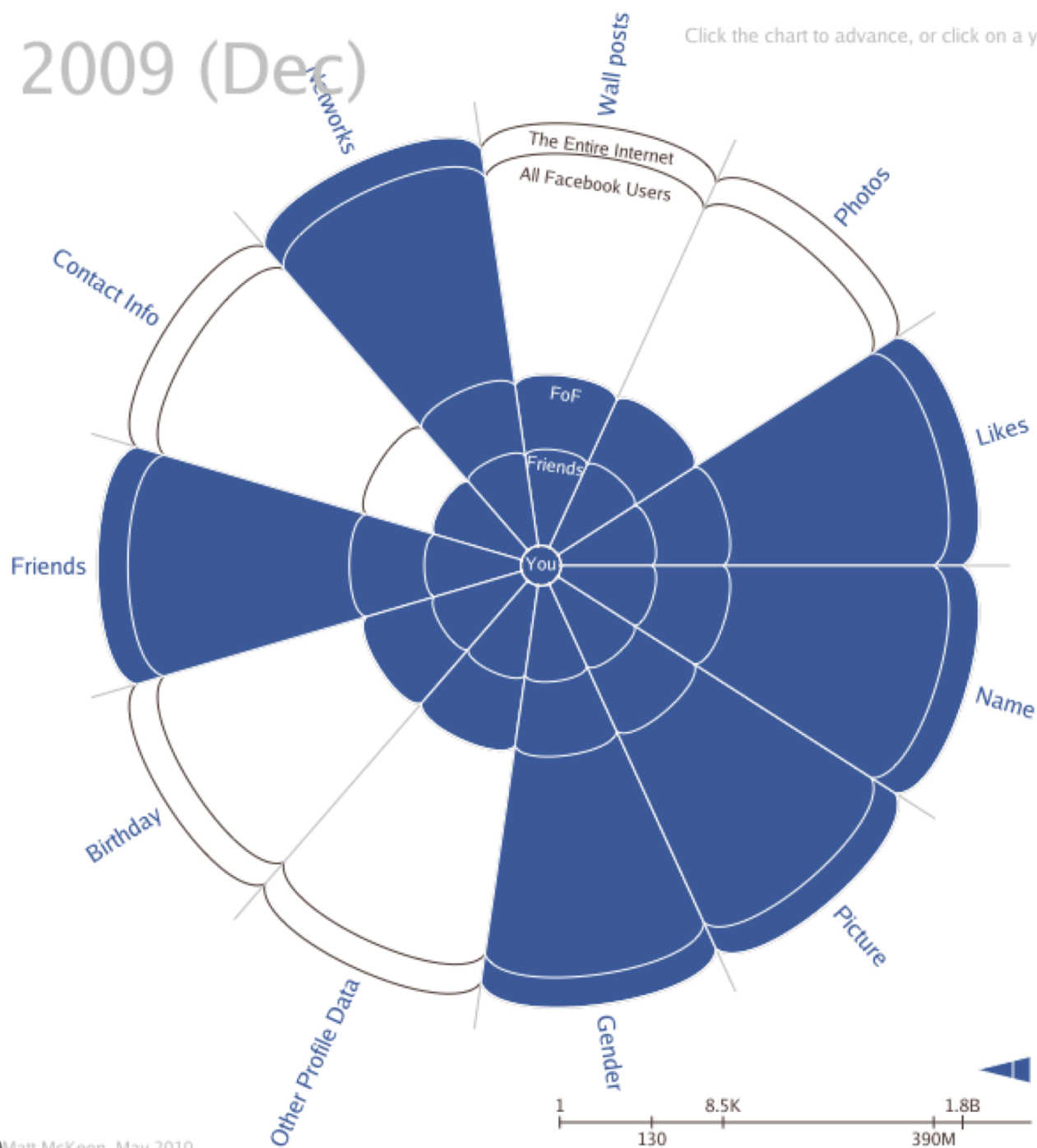
▲ Availability of your personal data on Facebook (default settings)  
Number of People



# 2009 (Dec)

Click the chart to advance, or click on a year

- 2005
- 2006
- 2007
- 2009 (Nov)
- 2009 (Dec)**
- 2010 (Apr)



▲ Availability of your personal data on Facebook (default settings)  
Number of People

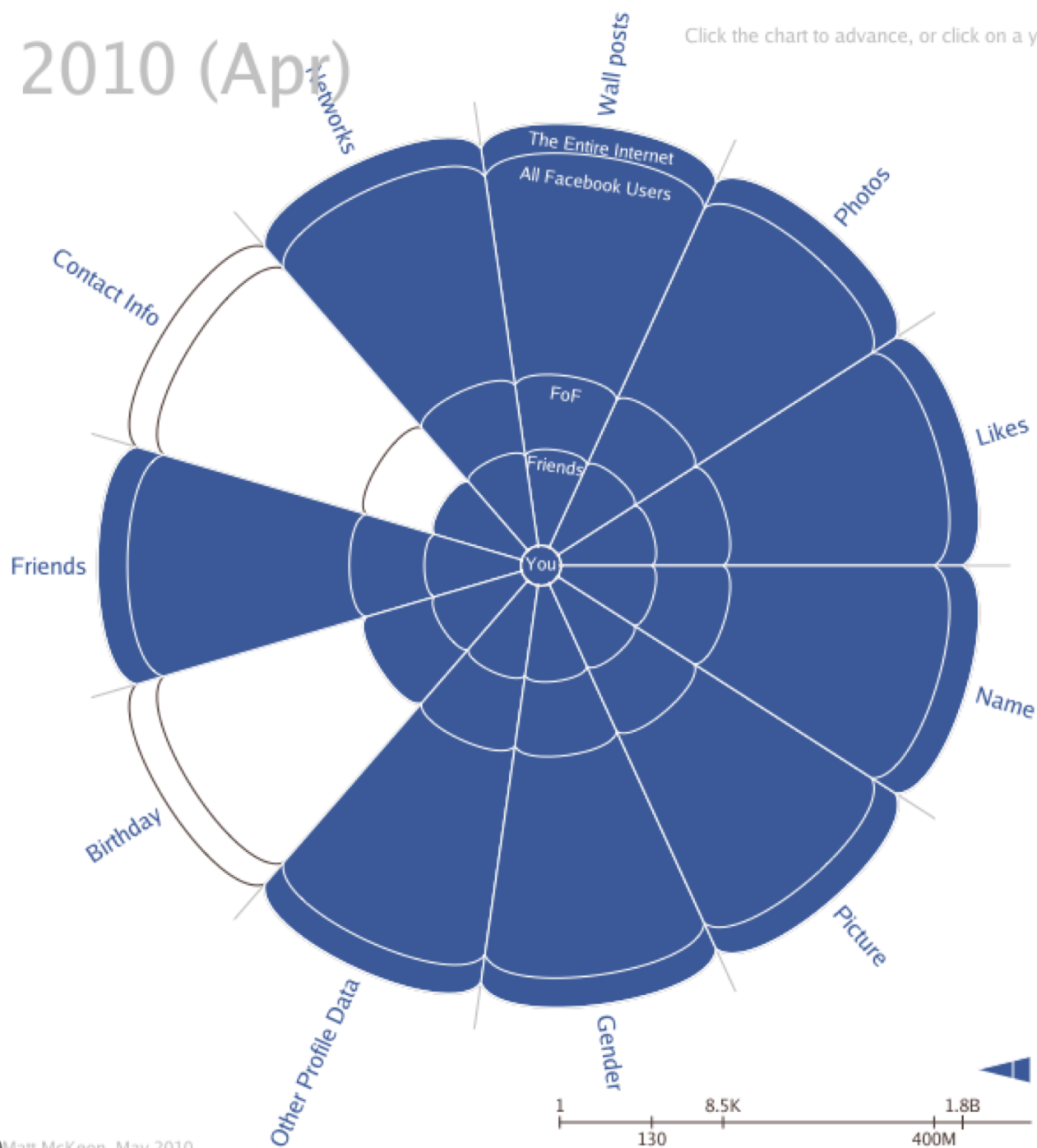
1 130 8.5K 390M 1.8B



# 2010 (Apr)

Click the chart to advance, or click on a year

- 2005
- 2006
- 2007
- 2009 (Nov)
- 2009 (Dec)
- 2010 (Apr)**



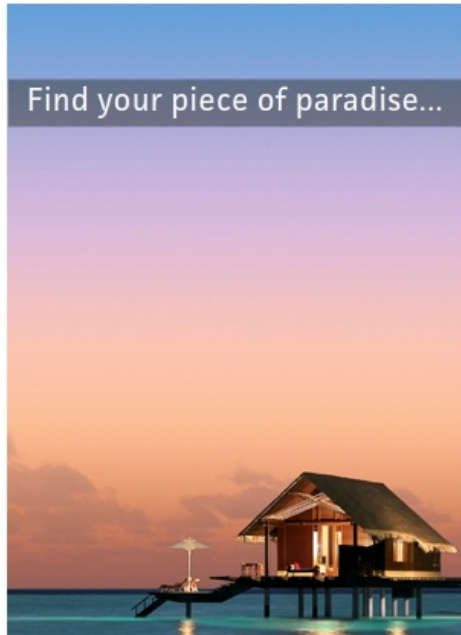
Availability of your personal data on Facebook (default settings)  
Number of People





# You are Facebook's product, not customer

TECHNOLOGY / 21 SEPTEMBER 11 / by OLIVIA SOLON ↗



People need to understand that they are the product of Facebook and not the customer, according to media theorist and writer Douglas Rushkoff.

Speaking at the inaugural Hello Etsy conference in Berlin, the author of *Program or Be Programmed* said: "Ask a kid what Facebook is for and they'll answer 'it's there to help me make friends'. Facebook's boardroom isn't talking about how to make Johnny more friends. It's talking about how to monetise Johnny's social graph."



[Flickr.com/designbyfront](https://www.flickr.com/photos/designbyfront/)





# US Senate Interviews



# http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/

## World's Biggest Data Breaches

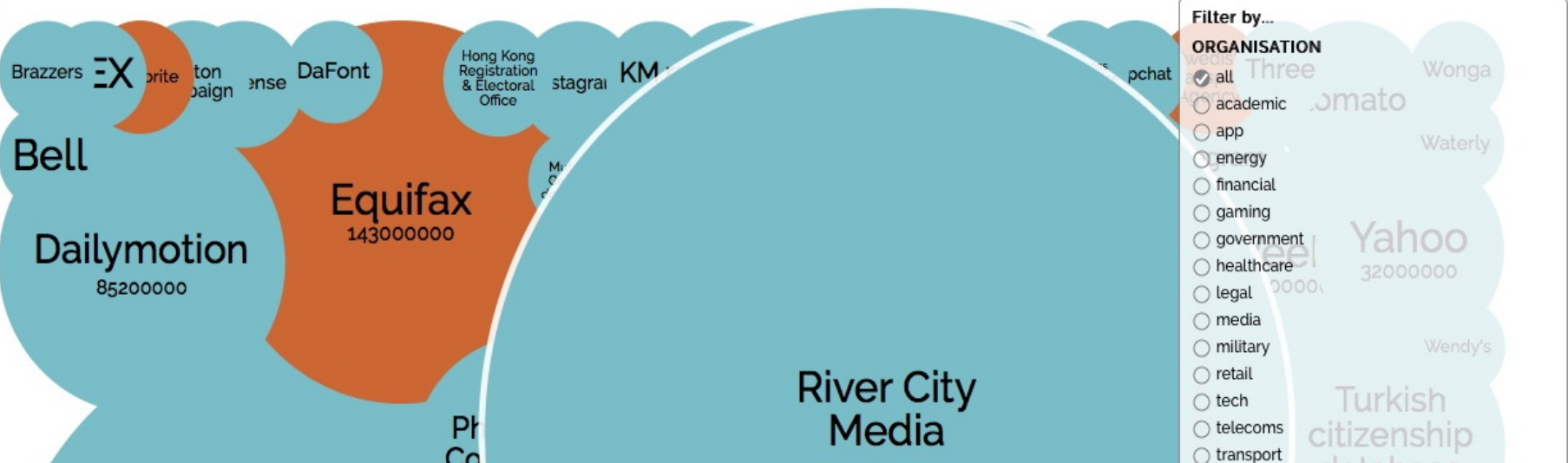
Selected losses greater than 30,000 records

(updated 10th Sep 2017)

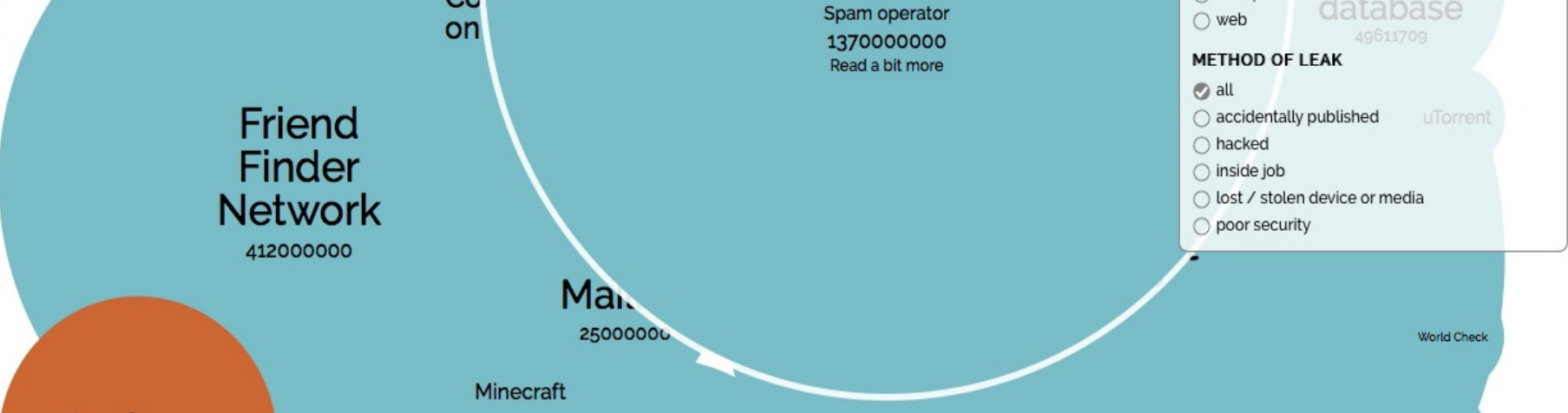
interesting story

YEAR BUBBLE COLOUR YEAR METHOD OF LEAK BUBBLE SIZE NO OF RECORDS STOLEN DATA SENSITIVITY HIDE FILTER

2017



2016



Filter by...

**ORGANISATION**

- all
- academic
- app
- energy
- financial
- gaming
- government
- healthcare
- legal
- media
- military
- retail
- tech
- telecoms
- transport
- web

**METHOD OF LEAK**

- all
- accidentally published
- hacked
- inside job
- lost / stolen device or media
- poor security

# <http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/>

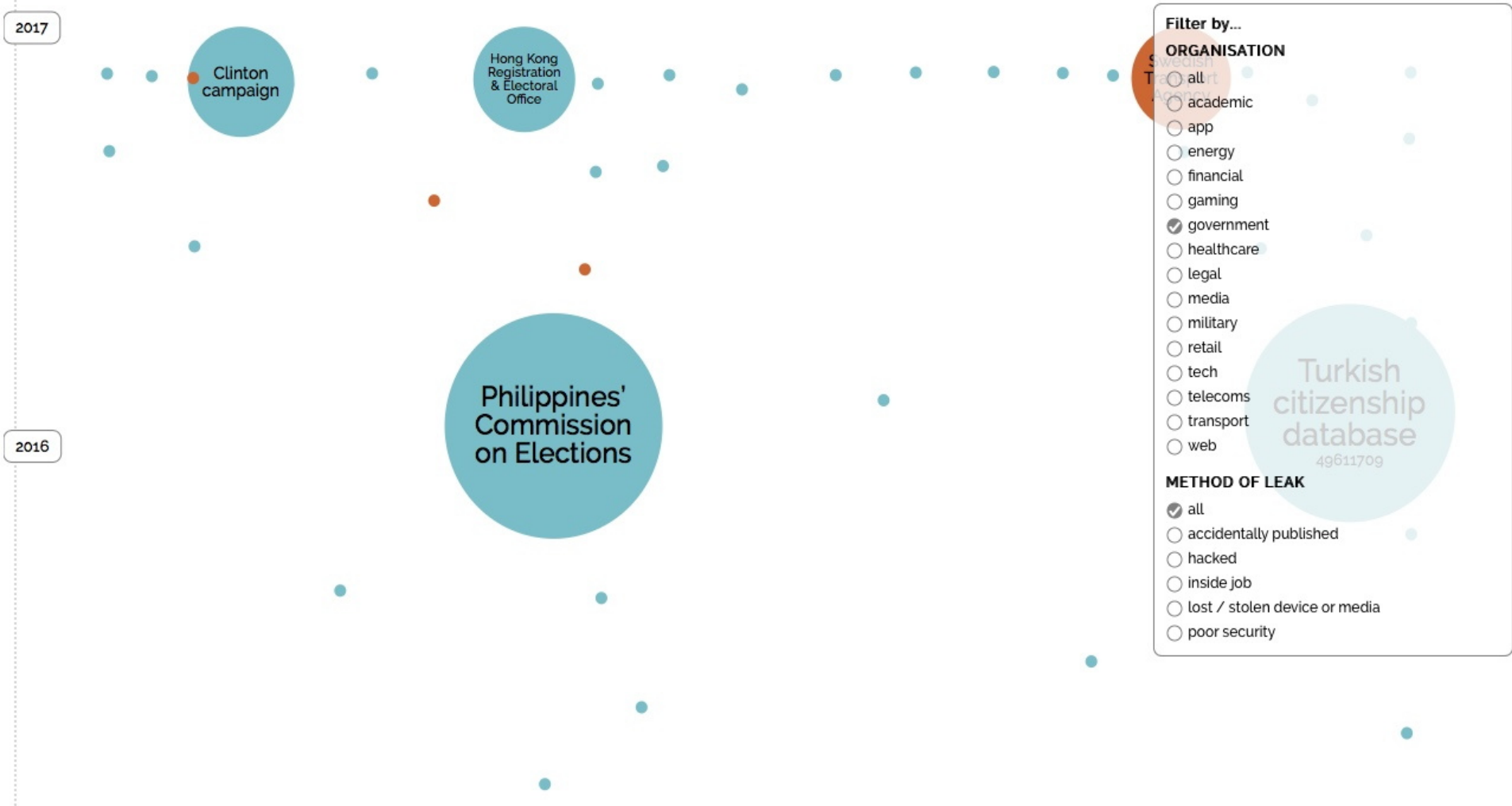
## World's Biggest Data Breaches

Selected losses greater than 30,000 records

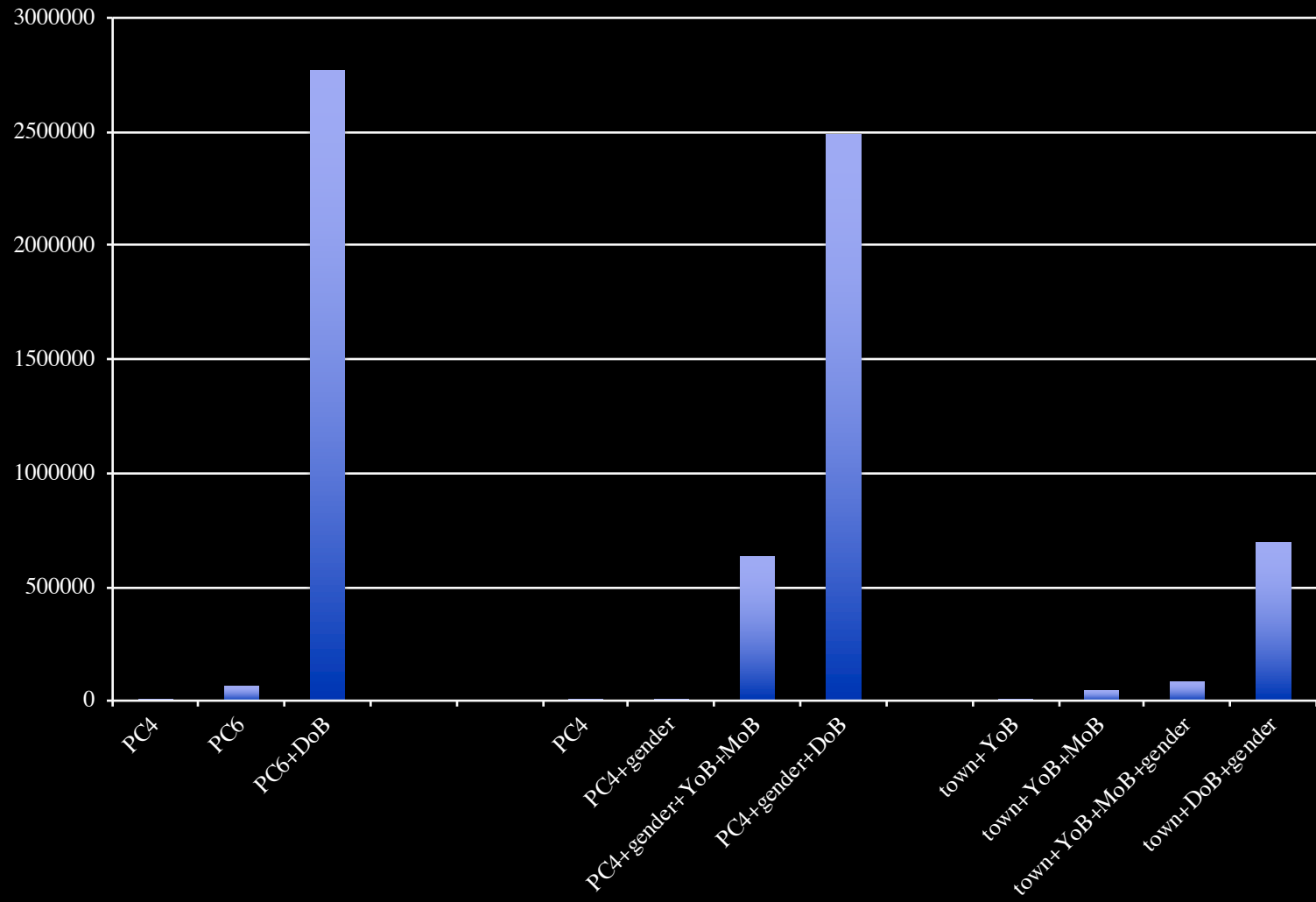
(updated 10th Sep 2017)

interesting story

YEAR BUBBLE COLOUR METHOD OF LEAK BUBBLE SIZE NO OF RECORDS STOLEN DATA SENSITIVITY HIDE FILTER



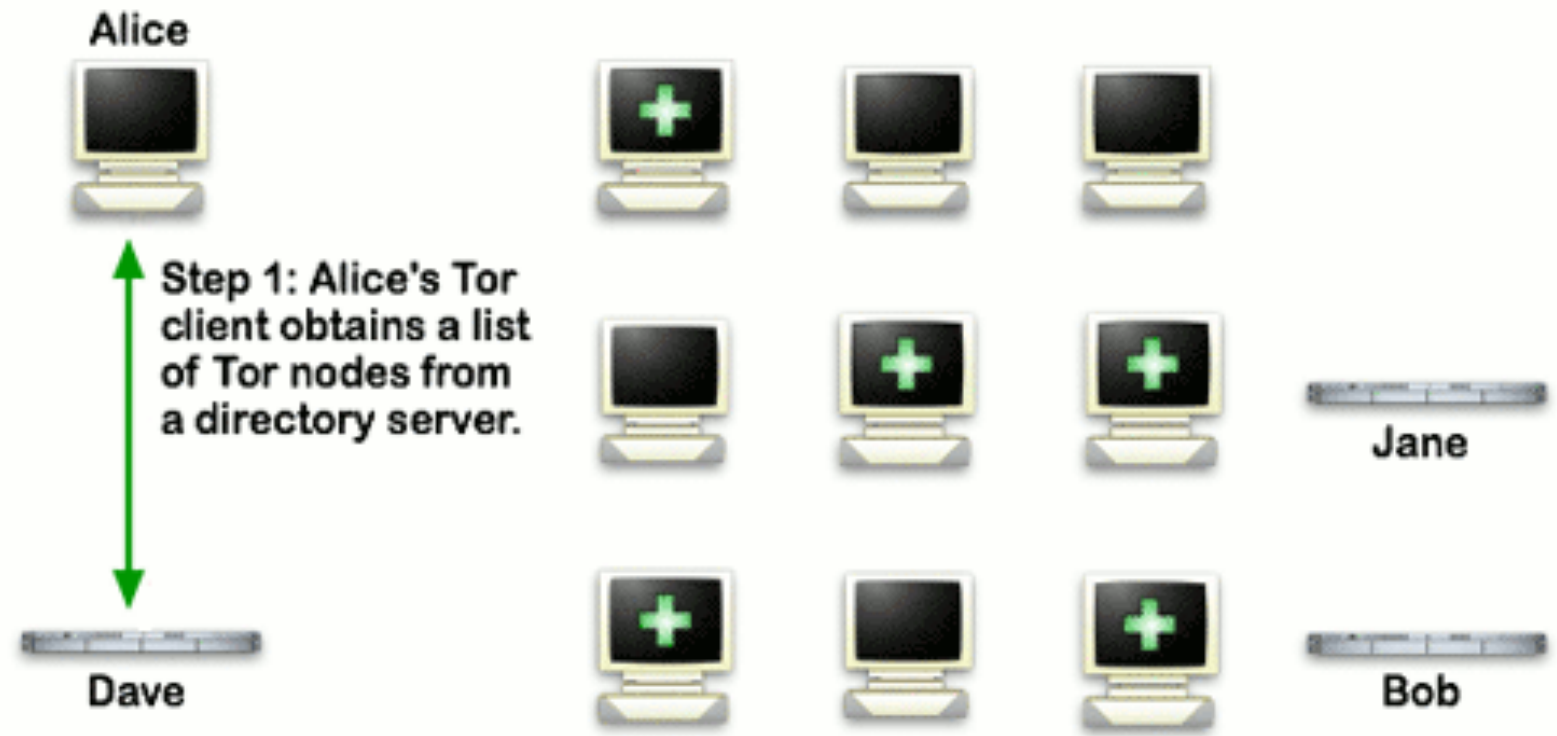
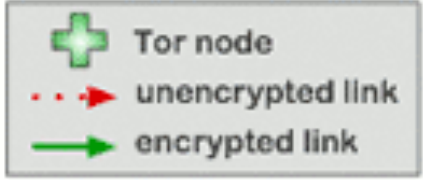
# Thesis Matthijs Koot





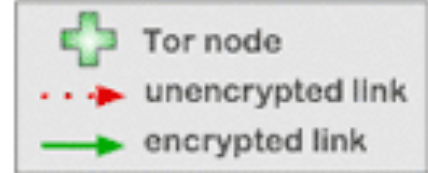
# TOR: third-generation onion routing project of the U.S. Naval Research Laboratory.

## How Tor Works: 1



# TOR: third-generation onion routing project of the U.S. Naval Research Laboratory.

## How Tor Works: 2



Alice



Step 2: Alice's Tor client picks a random path to destination server. **Green links** are encrypted, **red links** are in the clear.



Jane



Dave

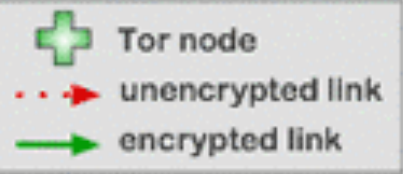


Bob

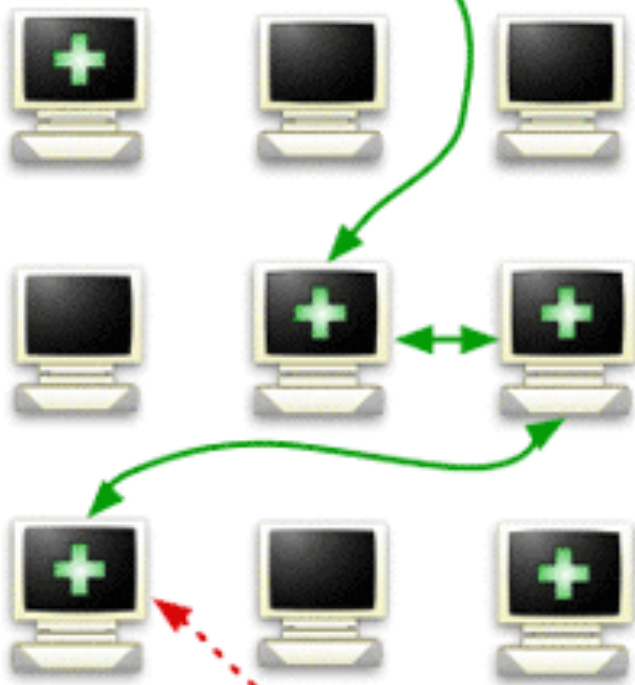


# TOR: third-generation onion routing project of the U.S. Naval Research Laboratory.

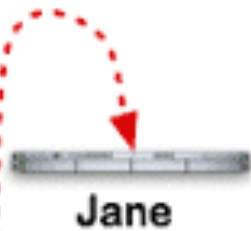
## How Tor Works: 3



Alice



Step 3: If at a later time, the user visits another site, Alice's tor client selects a second random path. Again, **green links** are encrypted, **red links** are in the clear.



# Some remarks

- Not everyone is interesting
- False positives disastrous
- The Internet does not forget
- Asymtotic loss of privacy
- Trying to hide can also trigger!
- Governments may be spooky, don't forget Industry!
- NSA candy store:
  - [http://en.wikipedia.org/wiki/NSA\\_ANT\\_catalog](http://en.wikipedia.org/wiki/NSA_ANT_catalog)



# IETF

RFC 7258, Pervasive Monitoring Is an Attack.

Author: S. Farrell

This document states that "Pervasive monitoring is a technical attack that should be mitigated in the design of IETF protocols, where possible."



# ECIS

## *Ethics Committee for Information Sciences*

- Until very recently, ethical discussions were only relevant to fields of research in which research is conducted on humans, such as medicine and some social sciences. However, due to the increased involvement of humans as in (in)direct research objects in the Information Sciences (IS), these ethical discussions are also becoming important in our field.
- <http://delaat.net/ecis>



# Q & A

