Privacy tangle: HTTPS and third party trackers.

Who is going to protect my daughter?

NNTools 2015

Torino, 19 January 2015

mPlane

mPlane for dummies

Project acronym: mPlane

Project full title: "mPlane – an Intelligent Measurement Plane for Future Network and Application Management"

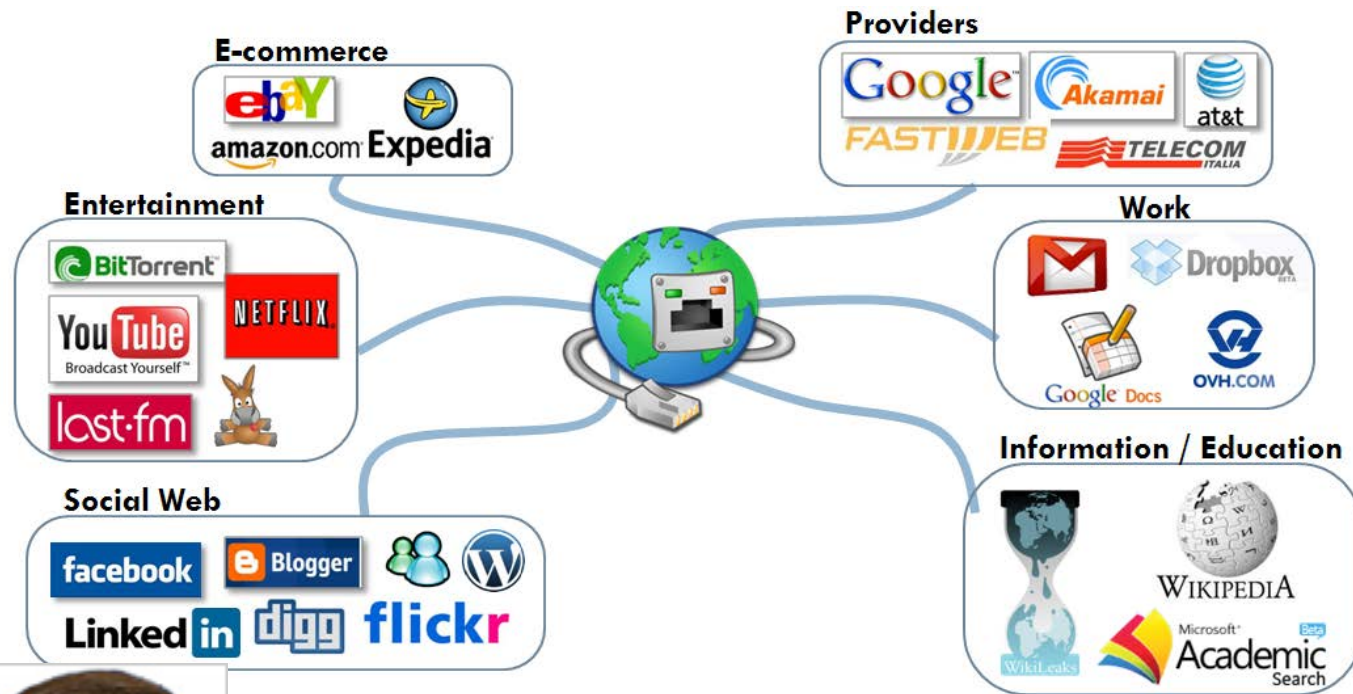Grant agreement no: 318627

Staring Date: November 1st 2012

Total Cost: 11,274,908.00Euros

Duration: 3 years

Partners: 16

Coordinator: **Prof. Marco Mellia** – Politecnico di Torino - IT

The nowadays Internet

The internet is a key infrastructure where different technologies are combined to offer a plethora of services. It's **horribly complicated** and we sorely **miss the technology to understand** what is happening in the network and thus to optimize its performance and utilization.

Eric Schmidt – ex Google Exec. Chairman

Some results

Something to think about…

Questions you may answer

- How much traffic is TCP or UDP?
- How much traffic is HTTP or P2P?
- Is the network neutral?
- …
- …
-
- **How much traffic is HTTPS?**
  - **And what are the "cost of the S"?**

mPlane

SEVENTH FRAMEWORK PROGRAMME

The Cost of the "S" in HTTPS

David Naylor
Peter Steenkiste

Carnegie Mellon University

Alessandro Finamore
Marco Mellia
Maurizio Munafò

Politecnico di Torino

Ilias Leontiadis
Yan Grunenberger
Dina Papagiannaki

Telefónica Research

mPlane

SEVENTH FRAMEWORK PROGRAMME

TLS

Certificates + Validation

Handshake

Cryptography

Authentication

Secrecy

mPlane

SEVENTH FRAMEWORK PROGRAMME

Direct Costs
mechanics of TLS

Indirect Costs
everything is encrypted

COSTS

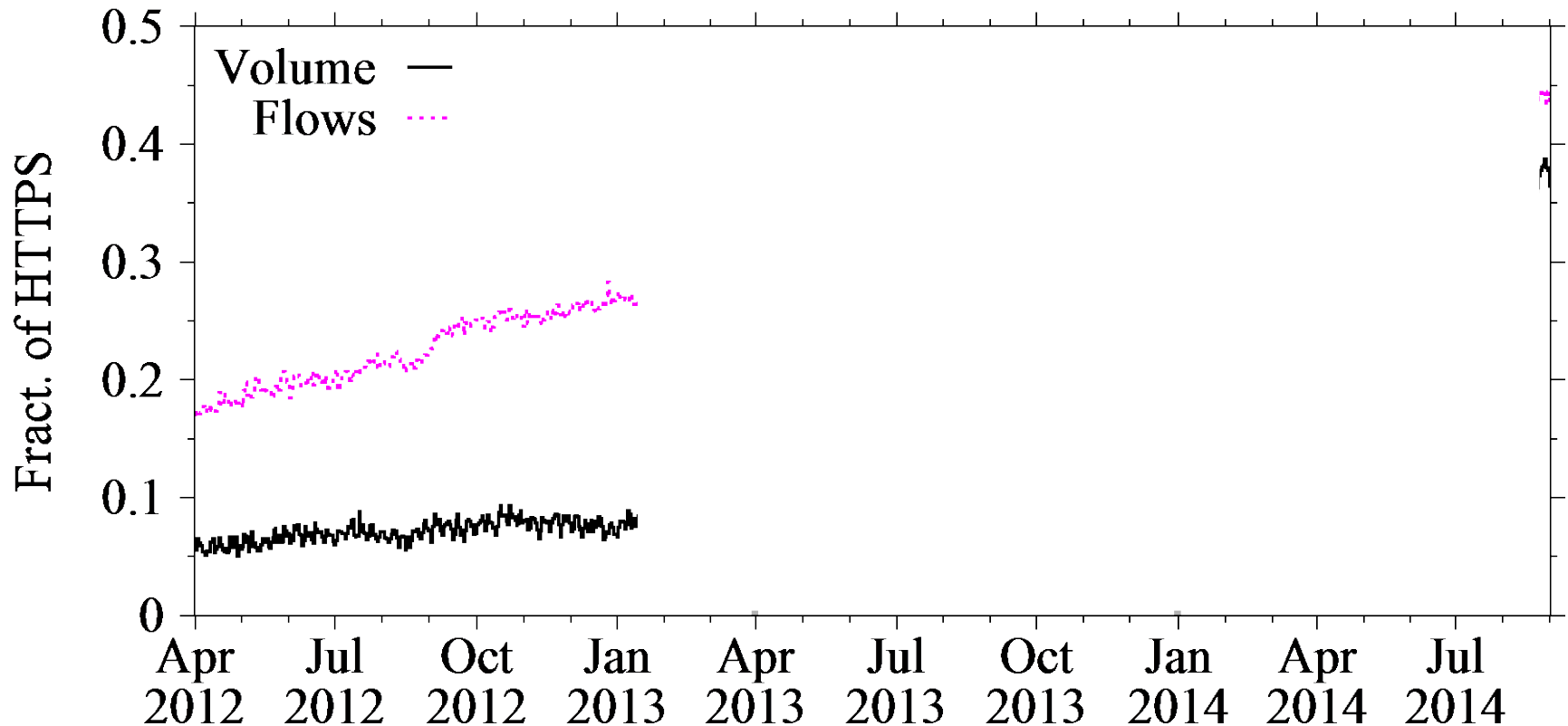|	
site infrastructure	1
load time	2
data usage	3
energy consumption	4
middlebox services	5

"S" usage trends

COSTS

site infrastructure / 1

load time / 2

data usage / 3

energy consumption / 4

middlebox services / 5

5

VALUE-ADDED SERVICES

LOTS OF MIDDLEBOXES

(that operate on packet contents)

compression

forensics

virus scanning

ad blocking

parental filtering

intrusion prevention

app-aware load balancing

transcoding

caching

app analytics

EXAMPLE: PARENTAL FILTERING
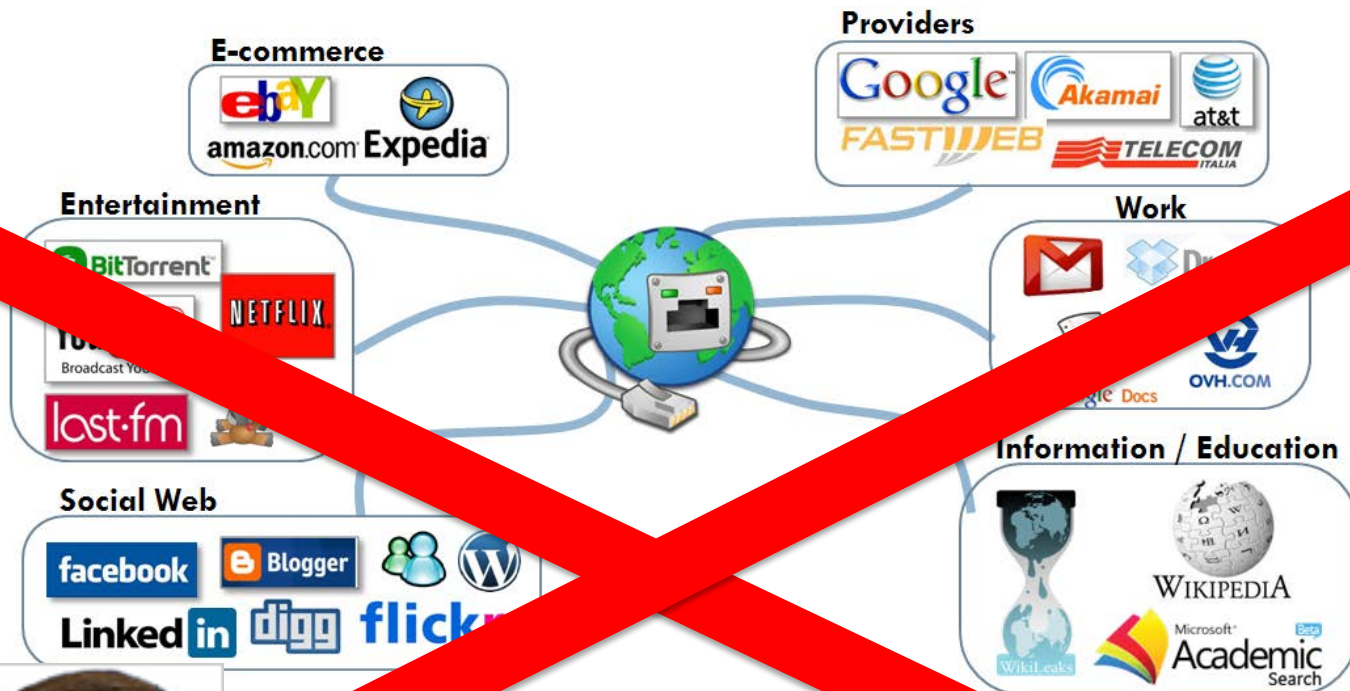
Internet Watch Foundation Blacklist
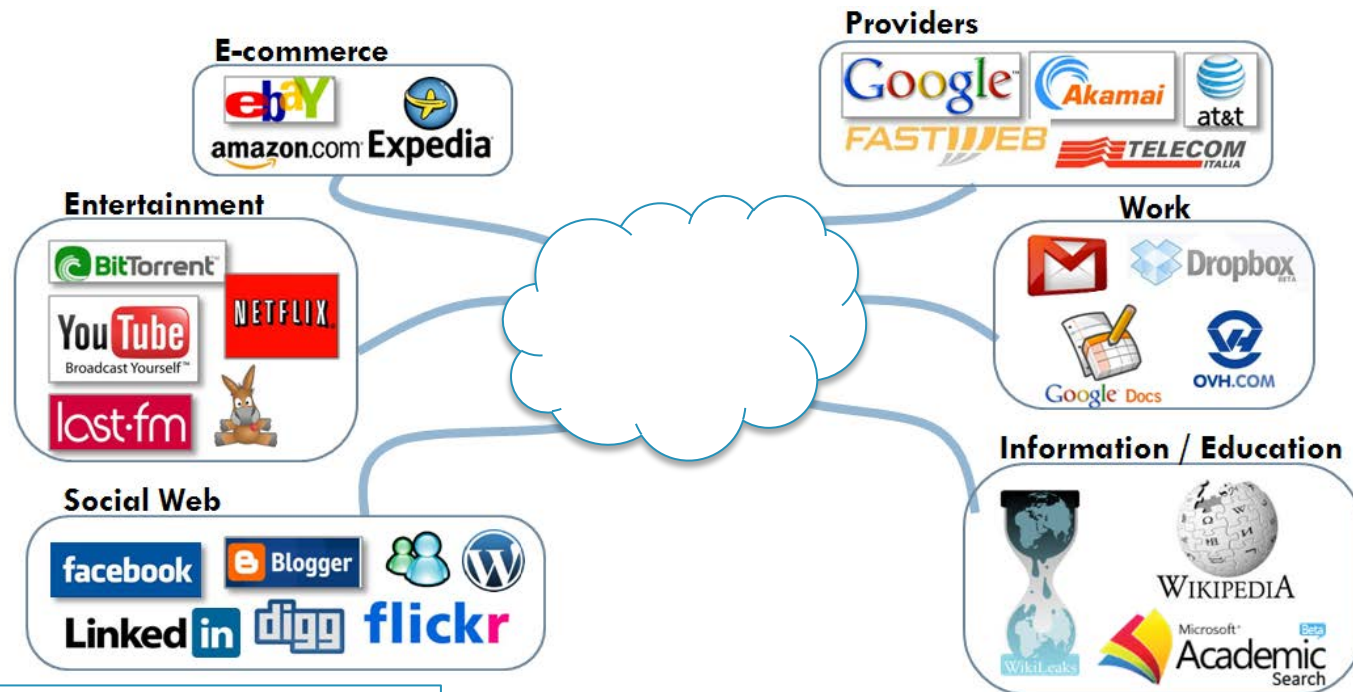
5% pure domain or subdomain

…Rewind…

The nowadays Internet

"The Internet is the first thing that humanity has built that humanity doesn't **understand**, the largest experiment in **anarchy** that we have ever had."

Eric Schmidt – ex Google Exec. Chairman

mPlane

The nowadays Internet

If you ask my ...er "What is the Internet?"…

What is the Internet?

Internet Usage
A look at the behavior or Internet users

96% ACCESS THE INTERNET DAILY

Internet users nearly universally ...cessed the ...nce a day.

90% INDICATE THEY USE SOCIAL MEDIA

A majority of users access social media sites, 60% access them on a daily basis.

Source: Internet Society Survey 2012

A TOOL FOR EXPRESSION

89% AGREE

agreed or agreed strongly that Internet access allows freedom of expression on all subjects.

A HUMAN RIGHT

83% AGREE

agreed or agreed strongly that access to the Internet should be considered a basic human right.

Plane

SEVENTH FRAMEWORK PROGRAMME

But "who" is the Internet?

RANK	April 2012 CDN	Byte	March 2013 CDN	Byte	July 2013 CDN	Byte	October 2013 CDN	Byte
1	Google	29.80%	Google	29.97%	Google	31.11%	Google	30.04%
2	Akamai	19.20%	Akamai	24.97%	Akamai	20.58%	Akamai	19.67%
3	Level3	5.20%	Limelight	5.52%	Limelight	6.19%	Limelight	5.49%
4	Limelight	4.50%	Level.3	3.58%	Level.3	4.50%	Level.3	3.95%
5	Netload	3.10%	Leaseweb	2.56%	Vkontakte.ltd	3.35%	Cogent	2.59%
6	Leaseweb	2.00%	Vkontakte.ltd	2.37%	Leaseweb	2.31%	Vkontakte.ltd	2.42%
7	Edgecast	1.80%	Ovh	1.74%	Edgecast	1.50%	Eweka	1.88%
8	VideotimeSpa	1.60%	Edgecast	1.74%	Cogent	1.42%	Videotime.spa	1.85%
9	OVH	1.20%	Videotime.spa	1.39%	Telecom.italia	1.25%	OVH	1.82%
10	Facebook	1.10%	Telecom.italia	1.29%	M247.ltd	1.19%	Leaseweb	1.82%
11	Amazon	1.10%	Amazon	1.12%	Haldex.ltd	1.01%	Deutsche.telekom	1.67%
12			M247.ltd	1.08%	Ovh	1.00%	Edgecast	1.37%
13			Haldex.ltd	0.91%	Deutsche.telekom	0.97%	Amazon	1.27%
14			Aruba.s.p.a	0.88%	Amazon	0.92%	Telecom.italia	0.99%
1			Deutsche.telekom	0.88%	Tripartz.b.v	0.87%	Fdcservers.net	0.93%

What services does my daughter use?

Google

Third-Party sites

SCORECARDRESEARCH

Welcome
*ScorecardResearch, […] a leading global market research effort that **studies and reports on Internet trends and behavior**.*
*ScorecardResearch conducts research by **collecting Internet web browsing data** and then uses that data to help show how people use the Internet, **what they like** about it, and what they don't.*

ScorecardResearch collects data through […] web tagging.

Take 100 users, and observe which services they contact while on the Internet…

*… **most of those are so called "third-party sites" - aka tracking services…***

Plane

SEVENTH FRAMEWORK PROGRAMME

My daughter and the Internet

	September 2013	September 2014
Average **number** of contacted tracking services per user per day		
Percentage of **HTTPS** flows directed to tracking servers		

Tracking services

- **233** out of the top-500 hostnames contacted by users are tracking service
- **196** of this 233 are contacted also using **HTTPS**
  - **To protect my daughter's privacy?**
  - **Or to protect their business?**
- First tracker contacted within **1 second** for 78% of users

- **What is the Internet??**

We can define policies ….

We can define policies…

… But how to verify them?

- Define mechanisms and protocols to identify violations
    - To allow third party verification of services
    - Using active and passive monitoring systems
    - Where end-terminals, networks and **servers** can be instrumented to cooperate

- First step: illuminating the obscure dynamics of the Internet **privacy tangle** via **measurements**

- The Internet is not anymore about packets…

                          … it's about humans and data

Plane

SEVENTH FRAMEWORK PROGRAMME

Perguntas

Fragen Domande Galdera

Otázky

Questions

Spørgsmål Pertanyaan kysymykset

Spørsmål

Frågor Cwestiynau

вопросы Preguntes Sorular

Въпроси

Vragen

Pytania

mPlane