mPlane

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 3y

mPlane

- FP7 IP project starting fall 2012
- Goal: build an Intelligent, Standard, Open Measurement Plane for Future Network and Application Management
  - **Probes (WP2)**
    - Build on existing tools/methodologies
    - Offer a flexible, programmable, open platform to run and collect passive, active, hybrid measurement
  - **Repositories (WP3)**
    - Collect in a standard way measurement
    - Offer access to interested parties: ISP, content providers, end-users, regulation agencies, etc.
  - **Intelligent reasoner (WP4)**
    - Mine automatically the data and extract useful information
    - Help in drilling down to the root cause of a problem
- Budget: 11.3ME (14M$)

mPlane

Internet Traffic Monitoring: Discerning Content and Services in a Tangled Web

Marco Mellia

Electronic and Telecommunication Department

Politecnico di Torino

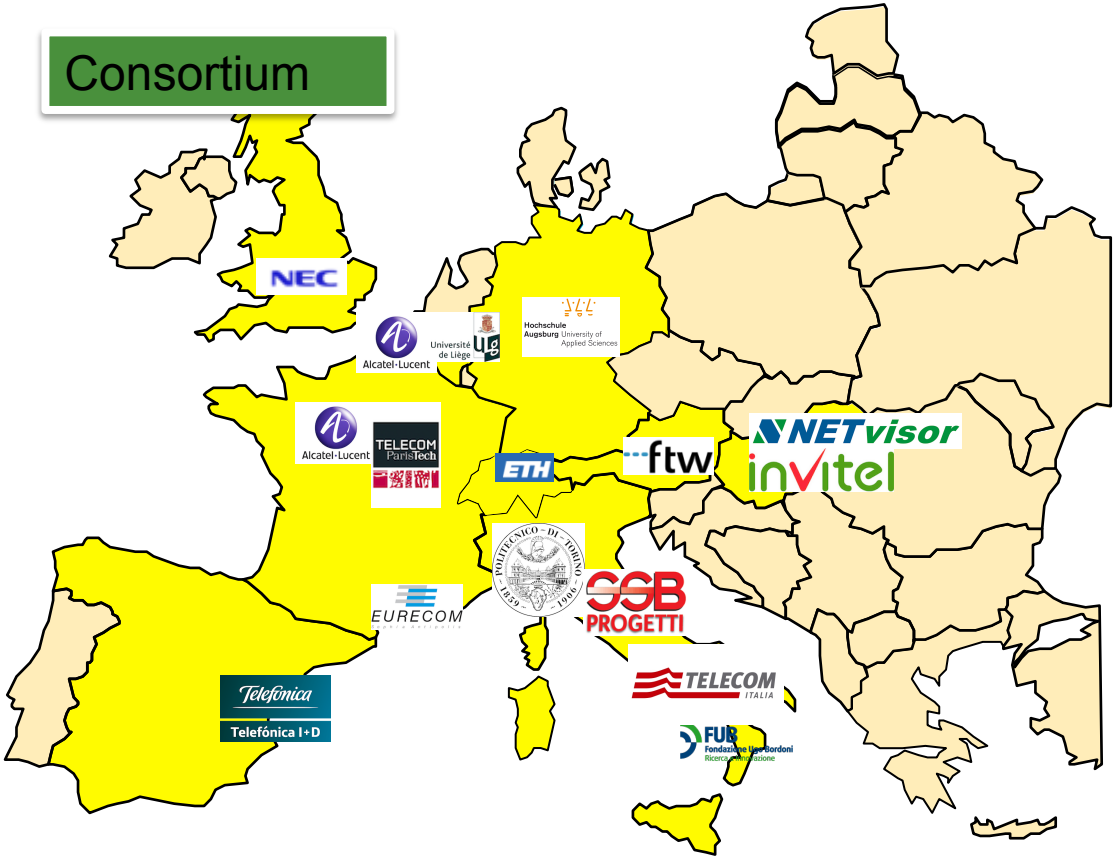
Email:mellia@tlc.polito.it

TELECOMMUNICATION NETWORKS GROUP

NARUS

POLITECNICO DI TORINO 1859 1906

mPlane



- Prof. Marco Mellia (PoliTo)
  Project Coordinator
- Dr Saverio Niccolini (**NEC**)
  Technical Coordinator
- Dr. Dina Papagiannaki (**TID**)
- Dr. Giovanna Carofiglio (**ALU**)
- Dr. Alessandro Capello (**TI**)
- Prof. Fabio Ricciato (FTW)
- Dr. Pedro Casas (FTW)
- Prof. Ernst Biersack (Eurecom)
- Prof. Pietro Michiardi (Eurecom)
- Prof. Dario Rossi (ENST)
- Prof. Laurent Mathy (Univ.Liege)
- Prof. Bennoit Donnet (Univ.Liege)
- Prof. Bernhard Plattner (ETH)
- Dr. Brian Trammel (ETH)
- Dr. Rolf Winter (Hochschule Augsburg)
- …
- …

FOR MORE INFO...

http://www.tlc-networks.polito.it

TNG at-a-glance - Faculty

NETWORKS GROUP

10 strutturati:

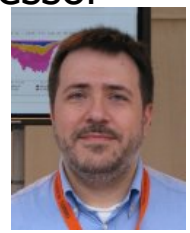
- **2 Full Professors**
- **4 Associate Professors**
- **4 Assistant Professors**

~5 Post-Docs

~25 PhD Stude

Marco Ajmone Marsan – Full Professor
Guido Albertengo – Associate Professor
Andrea Bianco – Full Professor
Claudio Casetti – Assistant Professor
Carla-Fabiana Chiasserini – Associate Professor
Paolo Giaccone – Assistant Professor
Emilio Leonardi – Associate Professor
o Mellia – Assistant Professor
la Meo –Associate Professor
izio Munafò – Assistant Professor

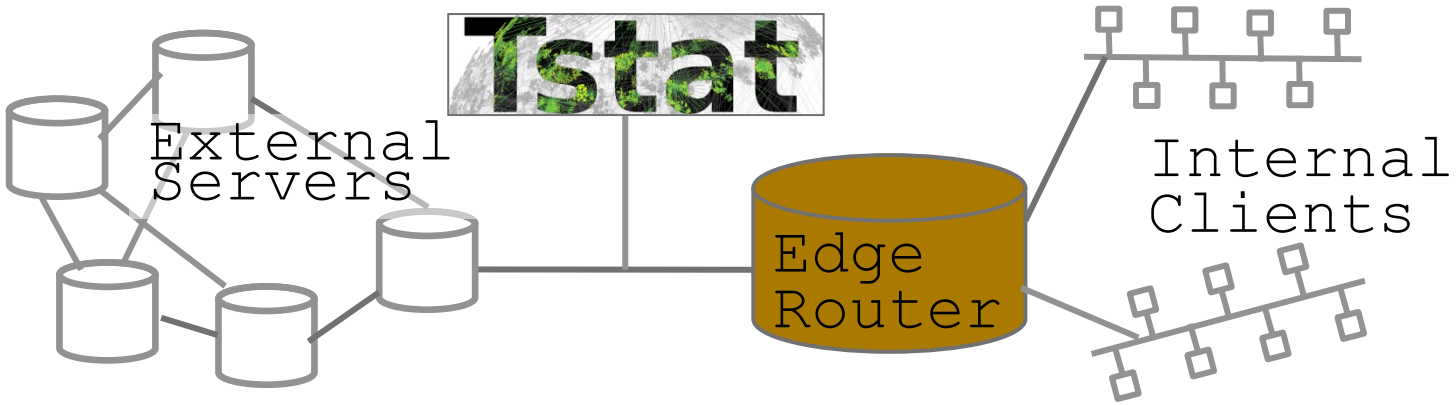
Fabio Neri **- Full Professor**



Traffic Monitoring in modern Internet

- **Why?**
  - Identify normal and anomalous behavior
  - Characterize the network and its users
  - Quality of service
  - Filtering
  - …
  - **Understand today Internet**
- **How?**
  - By means of passive measurement

Scenario



- Traffic classifier
  - Deep packet inspection
  - Statistical methods
- Persistent and scalable monitoring platform
  - Round Robin Database (RRD)
  - Histograms
  - Logs (like narusvectors)

http://tstat.tlc.polito.it

mPlane

Where Tstat Lives

Plane

The past, the present, the future

- The past: we have technology to monitor the Internet
  - Classifiers (DPI, Behavioral)
  - Tools (open software, off-the-shelf hardware, …)
- The present: everything is HTTP, cloud, virtualization, CDN, … everything is becoming a huge tangle:
  - Who owns, serve, control the content?
  - What are the users doing?
  - What is the network doing?
- The future: how to control the tangle?
  - **Today ISPs are struggling**
  - **… and the Internet is controlled by few big players…**

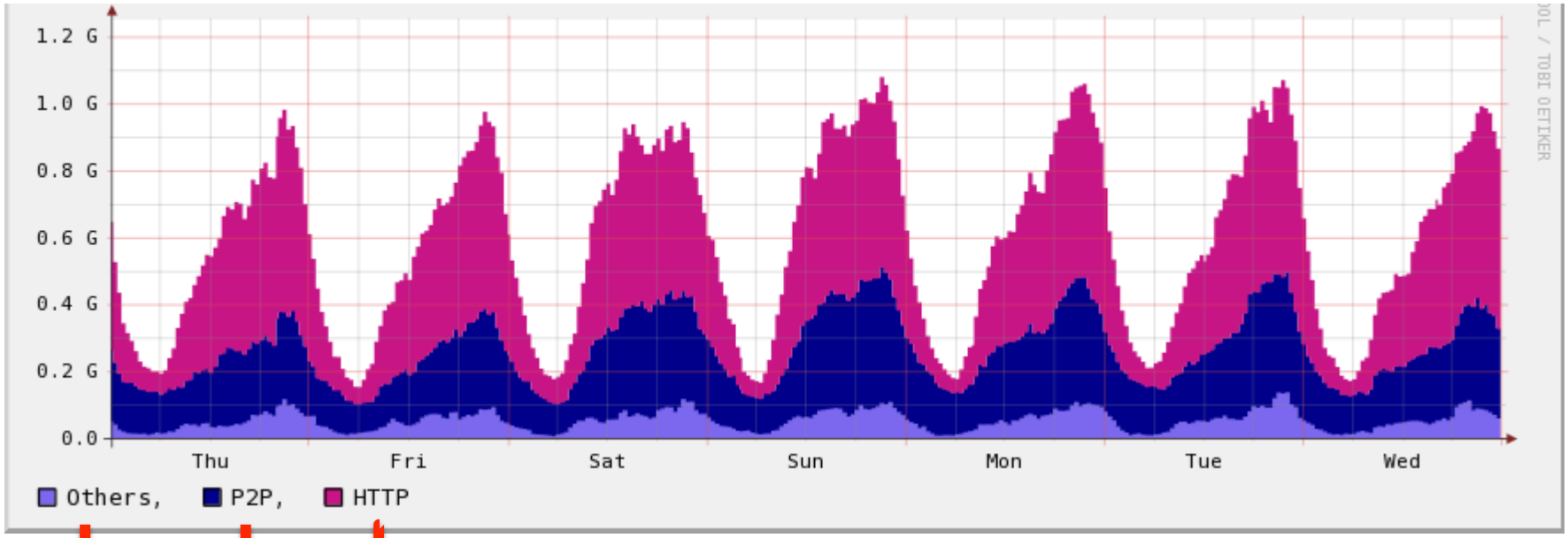
Uncovering the Big Players of the Web

Vinicius Gehlen
Alessandro Finamore
Marco Mellia
Maurizio M. Munafò

Introduction

Nowadays Internet traffic volume is mainly HTTP + P2P

Breakdown of downstream traffic of residential customers



A plethora of services!

+ BitTorrent

Mainly SSH, VoIP, DNS, eMail, etc.

Plane

Methodology

- Focus only on HTTP traffic
- Rely on Tstat to generate flow-level HTTP logs
  - L4: #bytes, #pkts, RTT, etc.
  - L7: service type and “meta-data” (e.g. video)
- Rely on MaxMind organization data base
  - Each server IP is associated to its “owner”
    - **92.122.208.73 →AKAMAI TECHNOLOGIES**

Dataset

- 3 vantage points of an ISP in Italy
  - Residential customers
  - ADSL (VP2, VP3) + Fiber-To-The-Home (VP1)
  - 2 weeks of traffic
    - 20-24 June 2011
    - 1-7 April 2012

Name	Volume [GB]	Flow [M]	# Servers	# Clients
VP1	1745 (35%)	16 (63%)	77,000 (0.14%)	1534 (99%)
VP2	10802 (44%)	84 (53%)	171,000 (0.6%)	11742 (97%)
VP3	13761 (35%)	125 (52%)	215,000 (0.5%)	17168 (98%)

OVERVIEW

Which organizations? Volumes? Popularity?

Top10 (+ 1) organizations (2011)

Rank	Org. Name	% B	% F
1	Google	22.7	12.7
2	Akamai	12.3	16.7
3	Leaseweb	6.3	1.1
4	Megaupload	5.5	0.2
5	Level3	4.7	1.9
6	Limelight	3.9	1.6
7	PSINet	3.2	0.2
8	Webzilla	2.9	0.3
9	Choopa	1.5	0.01
10	OVH	1.0	0.7
11	Facebook	0.9	4.2

- Google handles 2x the Akamai volume
- Besides Google and Akamai, many others
  - known (Level3, Limelight, Leaseweb, Megaupload)
  - less known (PSINet, Webzilla, Choopa)
- >10k organizations but 65% of volume is due to only 11 big players

Top10 (+ 1) organizations (2011)

June 2011

Rank	Organization	% Bytes
1	Google	22.7
2	Akamai	12.3
3	Leaseweb	6.3
4	Megaupload	5.5
5	Level3	4.7
6	Limelight	3.9
7	PSINet	3.2
8	Webzilla	2.9
9	Choopa	1.5
10	OVH	1.0
11	Facebook	0.9
	Total	64.9

April 2012

Rank	Organization	% Bytes	
1	Google	29.8%	+++
2	Akamai	19.2%	+++
3	Level3	5.2%	++
4	Limeligth	4.5%	+
5	Netload	3.1%	New
6	Leaseweb	2.0%	---
7	Edgecast	1.8%	New
8	VideotimeSpa	1.6%	New
9	OVH	1.2%	+
10	Facebook	1.1%	++
11	Amazon	1.1%	New
	Total	**70.6**	

mPlane

Organizations popularity (2011)

% client IP that have contacted the organization at least one time

Organization	% Client	Video Content	SW Update	Adv. & Others
Google		YouTube	-	Google services
Akamai		Vimeo	Microsoft, Apple	Facebook static content, eBay
Leaseweb		Megavideo	-	publicbt.com
Megaupload		Megavideo	-	FileHosting
Level3		YouPorn	-	quantserve, tinypic, Photobucket
Limelight		Pornhub, Veoh	Avast	betclick, wdig, trafficjunky
PSINet		Megavideo	Kaspersky	Imageshack
Webzilla		Adult Video	-	Filesonic, Depositfiles
Choopa		-	-	zShare
OVH		Auditude	-	Telaxo, m2cai
Facebook		Facebook	-	Facebook dynamic content

- 97% of clients contacts Google and Akamai
- 63% of client clients contacts OVH
  - Advertisement
- 90.6% of clients contacts Facebook
  - !??!!?

mPlane

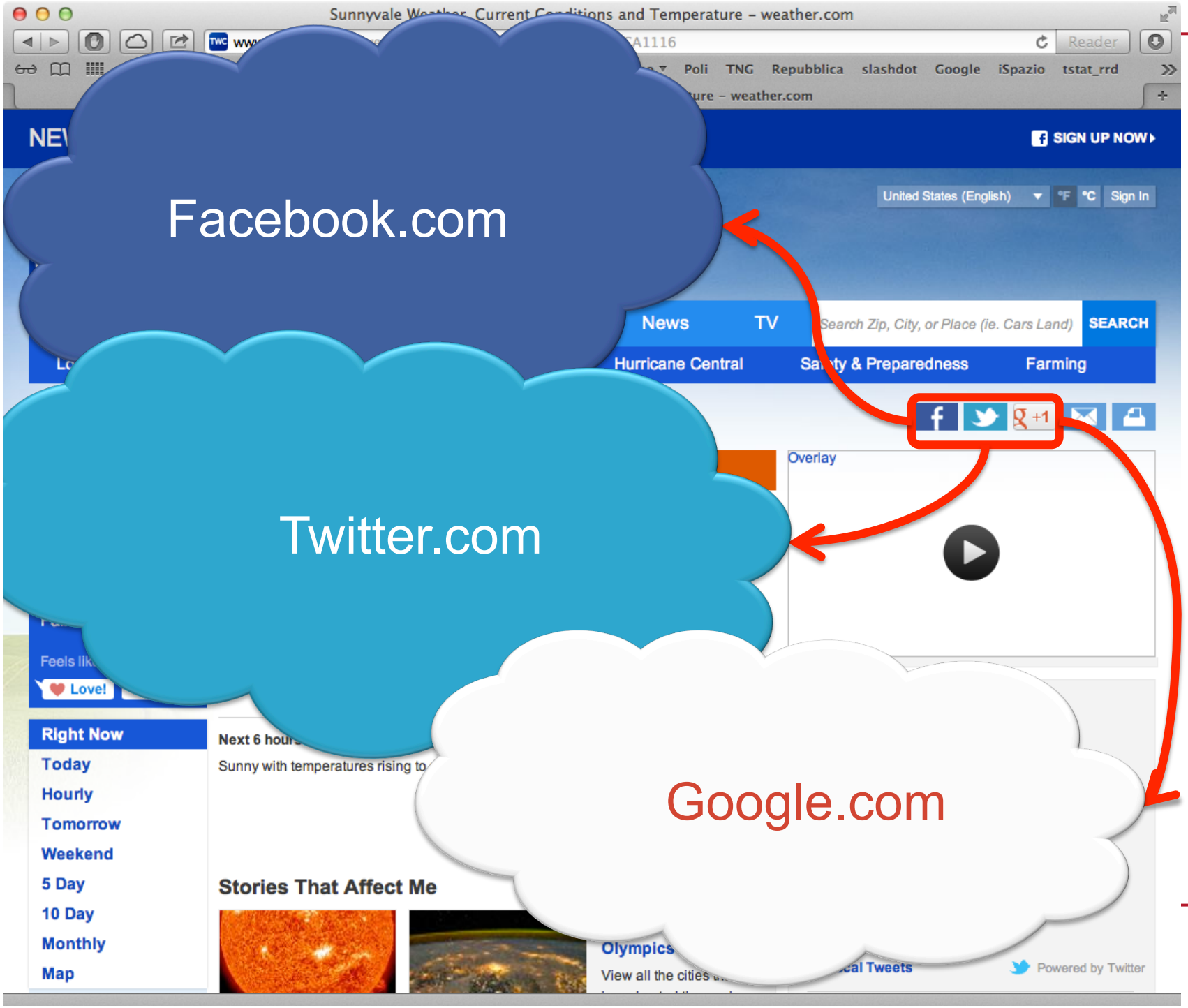
Why FB sees **91%** of clients?

- I visit nutella.com
  - Slurp!
- There is an embedded object pointing to the FB page about Nutella
- This generates a connection to FB
- So **FB knows** that **I like nutella**
  - Privacy anyone?!?!

Plane

Why FB sees 91% of clients?

```
GET /plugins/like.php?href=http%3A%2F%2Fwww.facebook.com
%2FNutella.Italy&layout=box_count&show_faces=false&width=120&action=like&colorscheme
=light&height=65 HTTP/1.1

Host: www.facebook.com

User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_3) AppleWebKit/534.53.11
(KHTML, like Gecko) Version/5.1.3 Safari/534.53.10

Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8

Referer: http://www.nutella.it/

Accept-Language: en-us

Accept-Encoding: gzip, deflate

Cookie:
presence=EM331242829EuserFA21519421867A2EstateFDsb2F0Et2F_5b_5dElm2FnullEuct2F133124
2225BEtrFnullEtwF838980386G331242829049H0EblcF0EsndF1CEchFDsubF_5b0_5dEp_5f151942186
7F2CC; p=6; c_user=1519421867; datr=FOkpTF9NjoDB9oSHmLkP5E_Ong..;
lu=ggoJZj70PxZ6gRARMPcNiRXw;
xs=1%3A889419cc0700aee88497c71b1015fa45%3A0%3A1331242820; locale=en_US
```

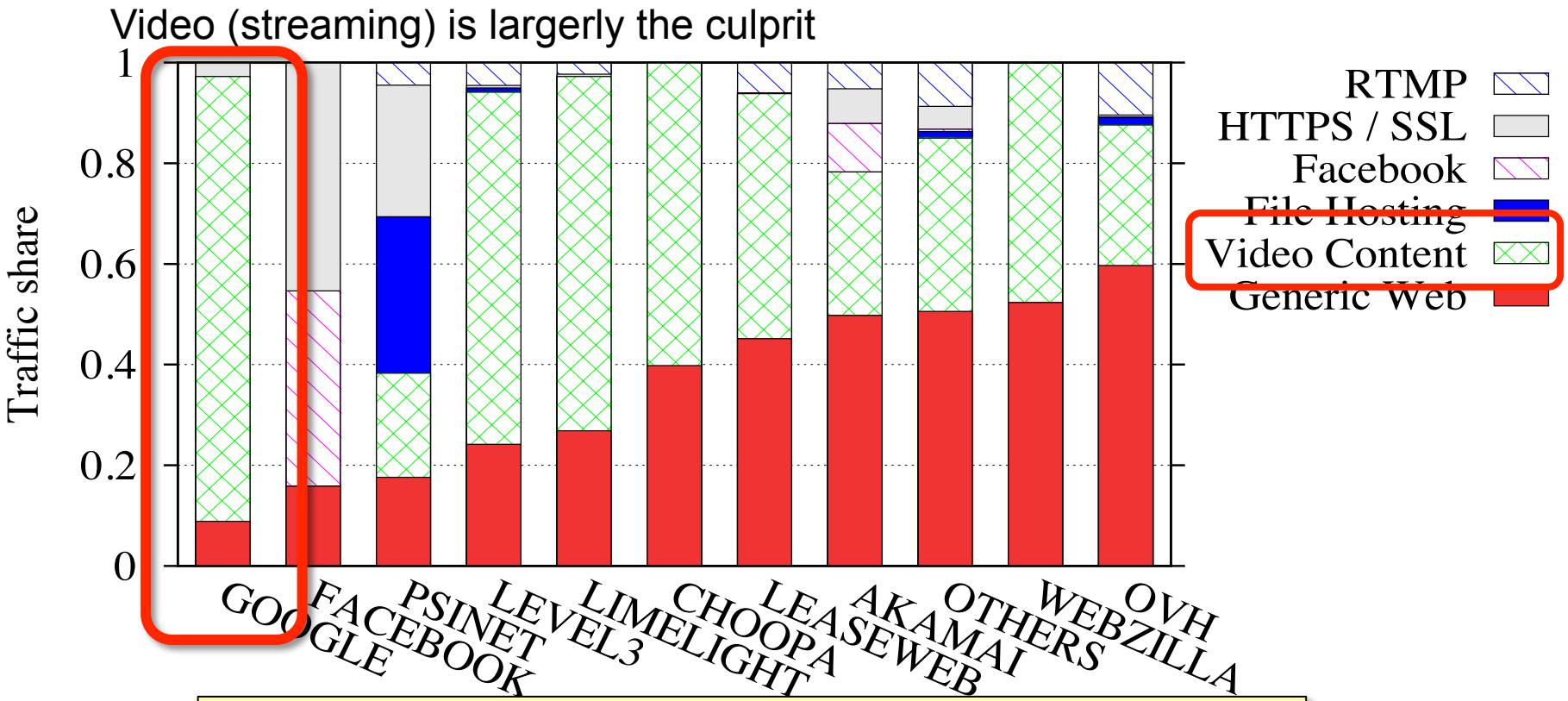

Facebook.com
Twitter.com
Google.com
Sunnyvale Weather, Current Conditions and Temperature – weather.com
SIGN UP NOW
United States (English)
Sign In
News
TV
Search Zip, City, or Place (ie. Cars Land)
SEARCH
Hurricane Central
Farming
Overlay
Right Now
Today
Hourly
Tomorrow
Weekend
5 Day
10 Day
Monthly
Map
Stories That Affect Me
Olympics
Powered by Twitter

CONTENT served

How much Video?

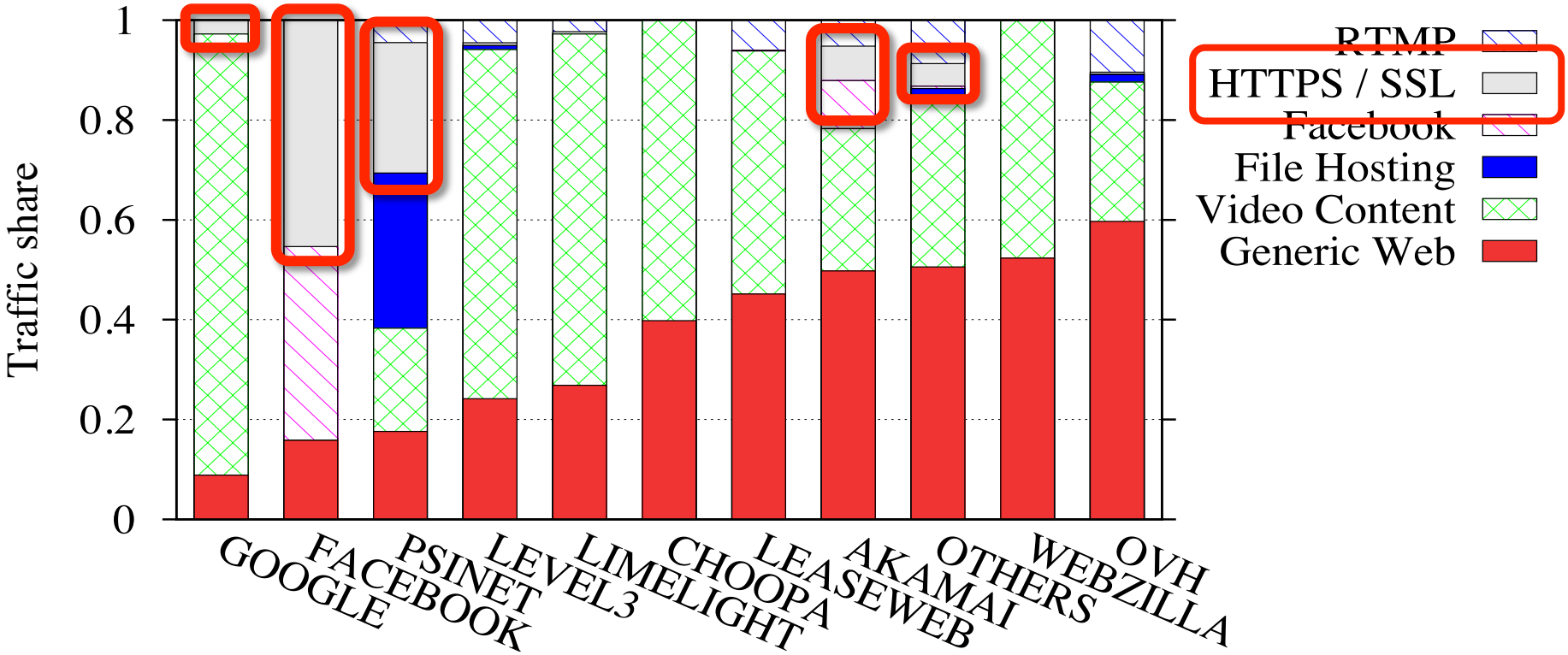
How much encryption ?

What type of traffic they serve (2012)



- ❑ 90% of Google traffic is YouTube
- ❑ Video is a large fraction of the share in any CDN

What type of traffic they serve (2012)

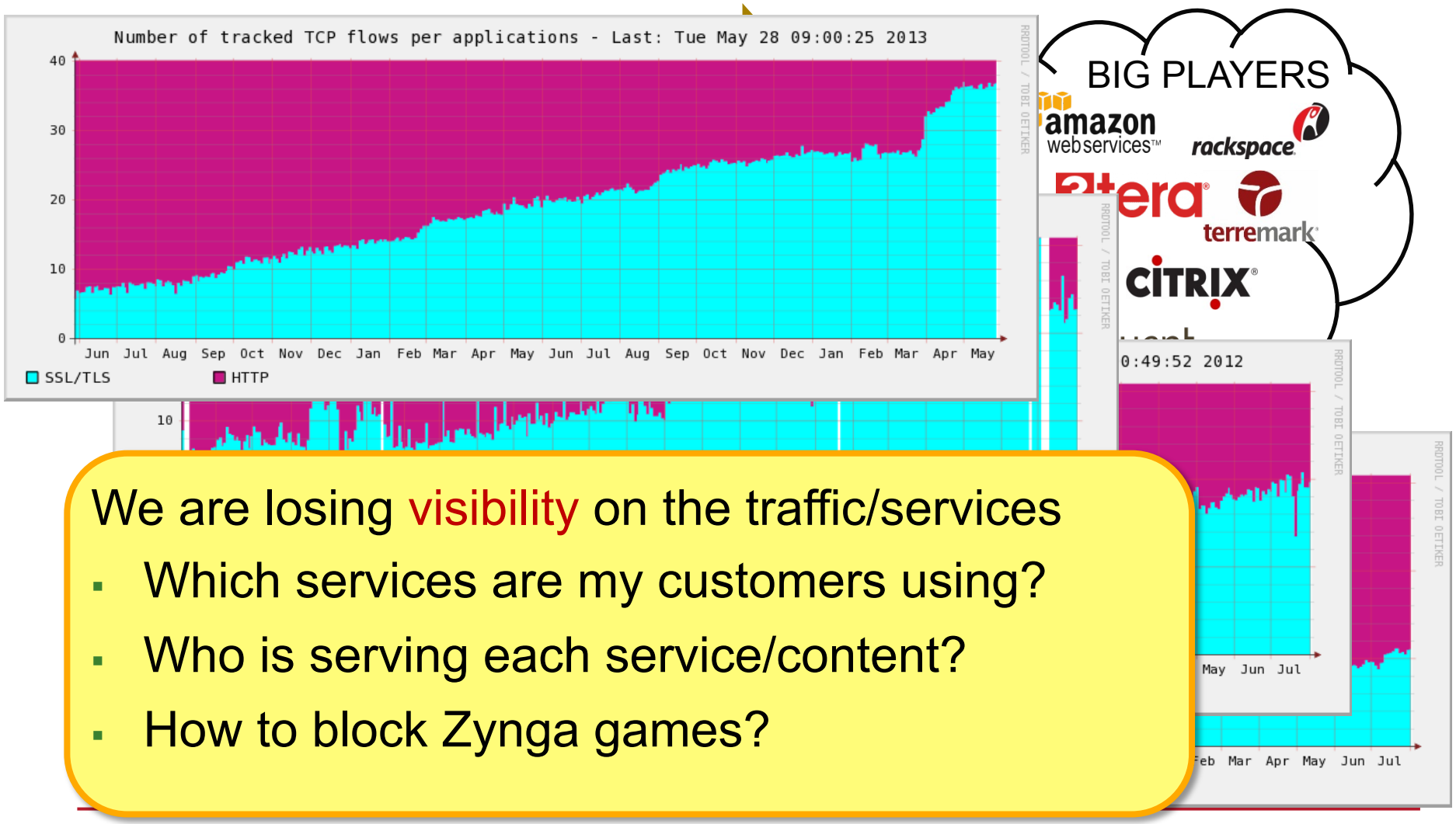


- Google, Akamai, Psinet and Facebook have a large share of HTTPS

This is going to impair traffic visibility **a lot!!**

Plane

A tangled and obfuscated WEB



We are losing visibility on the traffic/services

- Which services are my customers using?
- Who is serving each service/content?
- How to block Zynga games?

Plane

Privacy

- facebook -> HTTPS
- twitter -> HTTPS
- Google France -> HTTPS
- YourFavouriteSite -> HTTPS
- …
- This is to protect your **privacy**…

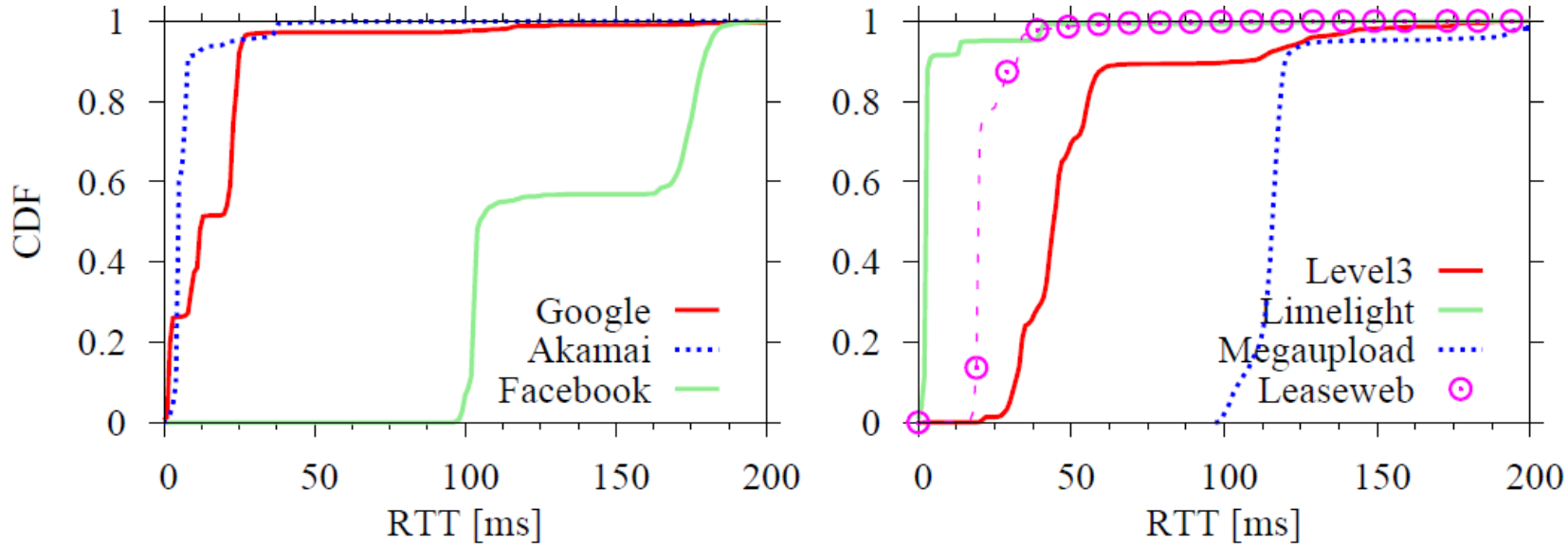
- … but then why the facebook app on iOS uses HTTP?!?!?

Organizations' Infrastructure

Behind the scene

RTT – Latency towards the Internet

CDF of minimum RTT, measured on per-flow base (2011)

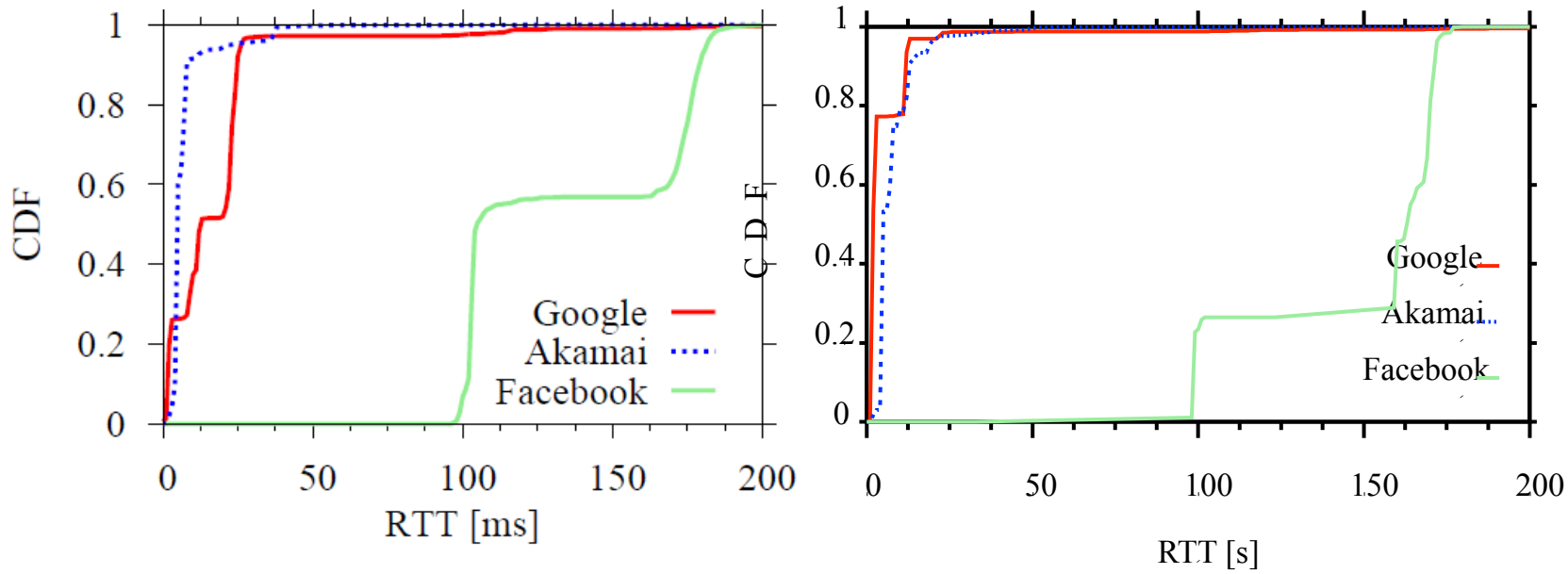


- Facebook has 2 locations (100ms and 170ms)
- 3 Google datacenters are preferred
  - Only <30% of request are served by the closest one (12ms)

Plane

RTT – Latency towards the Internet

CDF of minimum RTT, measured on per-flow base

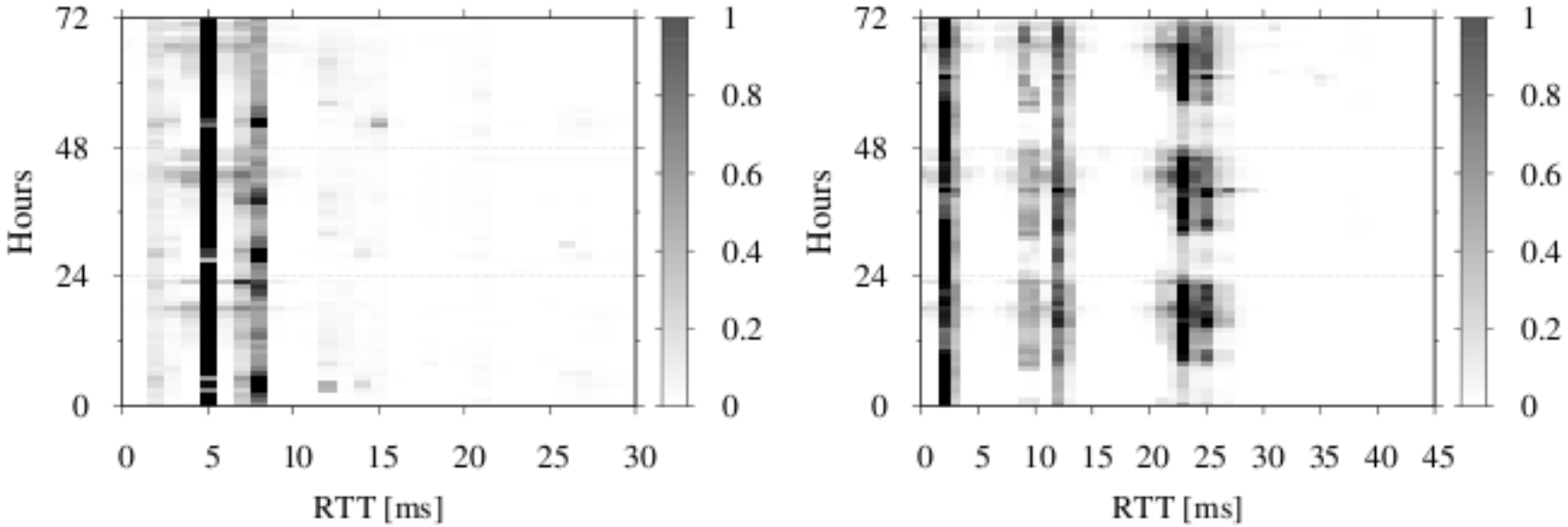


June 2011

April 2012

- Some **months** later …

RTT variation during the day



(a) Akamai (b) Google

- Some **hours** later …

Who controls the network?

mPlane

Something we observe

Something we observe

SUBNET
173.194.18
173.194.19
173.194.2
173.194.20
173.194.208
173.194.5
173.194.6
173.194.62
173.194.9
208.117.236
208.117.248
208.117.250
208.117.252
208.117.254
74.125.105
74.125.13
74.125.14
74.125.216
74.125.218
74.125.4
74.125.99

Something we observe

SUBNET	NAME with AIRPORT code
173.194.18	fra02s08.c.youtube.com
173.194.19	fra02s15.c.youtube.com
173.194.2	mil01s12.c.youtube.com
173.194.20	par08s06.c.youtube.com
173.194.208	par08s06.c.youtube.com
173.194.5	lhr14s08.c.youtube.com
173.194.6	fra07s13.c.youtube.com
173.194.62	fra07s19.c.youtube.com
173.194.9	par03s06.c.youtube.com
208.117.236	par03x04.c.youtube.com
208.117.248	mia02s11.c.youtube.com
208.117.250	ams09x06.c.youtube.com
208.117.252	dfw06x02.c.youtube.com
208.117.254	fra07x03.c.youtube.com
74.125.105	lhr22s16.c.youtube.com
74.125.13	zrh04s03.c.youtube.com
74.125.14	mil02s01.c.youtube.com
74.125.216	bru02t11.c.youtube.com
74.125.218	fra07t13.c.youtube.com
74.125.4	lhr22s11.c.youtube.com
74.125.99	fra07s03.c.youtube.com

Something we observe

SUBNET	NAME with AIRPORT code	5-May #flow	5-May Tru avg
173.194.18	fra02s08.c.youtube.com	-1	-1
173.194.19	fra02s15.c.youtube.com	-1	-1
173.194.2	mil01s12.c.youtube.com	17054	1333.46
173.194.20	par08s06.c.youtube.com	-1	-1
173.194.208	par08s06.c.youtube.com	-1	-1
173.194.5	lhr14s08.c.youtube.com	449	1819.57
173.194.6	fra07s13.c.youtube.com	-1	-1
173.194.62	fra07s19.c.youtube.com	-1	-1
173.194.9	par03s06.c.youtube.com	-1	-1
208.117.236	par03x04.c.youtube.com	179	164.18
208.117.248	mia02s11.c.youtube.com	-1	-1
208.117.250	ams09x06.c.youtube.com	41430	679
208.117.252	dfw06x02.c.youtube.com	-1	-1
208.117.254	fra07x03.c.youtube.com	838	667.29
74.125.105	lhr22s16.c.youtube.com	1829	1551.78
74.125.13	zrh04s03.c.youtube.com	719	1074.15
74.125.14	mil02s01.c.youtube.com	48366	1234.82
74.125.216	bru02t11.c.youtube.com	-1	-1
74.125.218	fra07t13.c.youtube.com	8697	1355.33
74.125.4	lhr22s11.c.youtube.com	1496	1846.25
74.125.99	fra07s03.c.youtube.com	-1	-1

Plane

Something we observe

		5-May		6-May	
SUBNET	NAME with AIRPORT code	#flow	Tru avg	#flow	Tru avg
173.194.18	fra02s08.c.youtube.com	-1	-1	-1	-1
173.194.19	fra02s15.c.youtube.com	-1	-1	-1	-1
173.194.2	mil01s12.c.youtube.com	17054	1333.46	15470	1276.31
173.194.20	par08s06.c.youtube.com	-1	-1	-1	-1
173.194.208	par08s06.c.youtube.com	-1	-1	-1	-1
173.194.5	lhr14s08.c.youtube.com	449	1819.57	283	1658.45
173.194.6	fra07s13.c.youtube.com	-1	-1	-1	-1
173.194.62	fra07s19.c.youtube.com	-1	-1	-1	-1
173.194.9	par03s06.c.youtube.com	-1	-1	-1	-1
208.117.236	par03x04.c.youtube.com	179	164.18	4250	540.16
208.117.248	mia02s11.c.youtube.com	-1	-1	77	552
208.117.250	ams09x06.c.youtube.com	41430	679	49437	656.39
208.117.252	dfw06x02.c.youtube.com	-1	-1	51	285.63
208.117.254	fra07x03.c.youtube.com	838	667.29	2130	852.53
74.125.105	lhr22s16.c.youtube.com	1829	1551.78	1655	1185.94
74.125.13	zrh04s03.c.youtube.com	719	1074.15	499	2264.09
74.125.14	mil02s01.c.youtube.com	48366	1234.82	37968	1253.01
74.125.216	bru02t11.c.youtube.com	-1	-1	-1	-1
74.125.218	fra07t13.c.youtube.com	8697	1355.33	12579	1338.71
74.125.4	lhr22s11.c.youtube.com	1496	1846.25	2488	1034.78
74.125.99	fra07s03.c.youtube.com	-1	-1	-1	-1

Something we observe

		5-May		6-May		7-May	
SUBNET	NAME with AIRPORT code	#flow	Tru avg	#flow	Tru avg	#flow	Tru avg
173.194.18	fra02s08.c.youtube.com	-1	-1	-1	-1	-1	-1
173.194.19	fra02s15.c.youtube.com	-1	-1	-1	-1	-1	-1
173.194.2	mil01s12.c.youtube.com	17054	1333.46	15470	1276.31	13655	**1259.63**
173.194.20	par08s06.c.youtube.com	-1	-1	-1	-1	-1	-1
173.194.208	par08s06.c.youtube.com	-1	-1	-1	-1	-1	-1
173.194.5	lhr14s08.c.youtube.com	449	1819.57	283	1658.45	-1	-1
173.194.6	fra07s13.c.youtube.com	-1	-1	-1	-1	-1	-1
173.194.62	fra07s19.c.youtube.com	-1	-1	-1	-1	-1	-1
173.194.9	par03s06.c.youtube.com	-1	-1	-1	-1	-1	-1
208.117.236	par03x04.c.youtube.com	179	164.18	4250	540.16	957	496.91
208.117.248	mia02s11.c.youtube.com	-1	-1	77	552	-1	-1
208.117.250	ams09x06.c.youtube.com	41430	679	49437	656.39	57675	**653.81**
208.117.252	dfw06x02.c.youtube.com	-1	-1	51	285.63	-1	-1
208.117.254	fra07x03.c.youtube.com	838	667.29	2130	852.53	-1	-1
74.125.105	lhr22s16.c.youtube.com	1829	1551.78	1655	1185.94	3957	942.47
74.125.13	zrh04s03.c.youtube.com	719	1074.15	499	2264.09	82	1302.03
74.125.14	mil02s01.c.youtube.com	48366	1234.82	37968	1253.01	37182	**1162.85**
74.125.216	bru02t11.c.youtube.com	-1	-1	-1	-1	-1	-1
74.125.218	fra07t13.c.youtube.com	8697	1355.33	12579	1338.71	8560	1239
74.125.4	lhr22s11.c.youtube.com	1496	1846.25	2488	1034.78	4146	**1363.63**
74.125.99	fra07s03.c.youtube.com	-1	-1	-1	-1	-1	-1

Plane

Something we observe

		5-May		6-May		7-May		8-May	
SUBNET	NAME with AIRPORT code	#flow	Tru avg	#flow	Tru avg	#flow	Tru avg	#flow	Tru avg
173.194.18	fra02s08.c.youtube.com	-1	-1	-1	-1	-1	-1	6139	**368.93**
173.194.19	fra02s15.c.youtube.com	-1	-1	-1	-1	-1	-1	9940	**258.18**
173.194.2	mil01s12.c.youtube.com	17054	1333.46	15470	1276.31	13655	**1259.63**	14186	1296.07
173.194.20	par08s06.c.youtube.com	-1	-1	-1	-1	-1	-1	-1	-1
173.194.208	par08s06.c.youtube.com	-1	-1	-1	-1	-1	-1	-1	-1
173.194.5	lhr14s08.c.youtube.com	449	1819.57	283	1658.45	-1	-1	3470	937.18
173.194.6	fra07s13.c.youtube.com	-1	-1	-1	-1	-1	-1	4924	**412.17**
173.194.62	fra07s19.c.youtube.com	-1	-1	-1	-1	-1	-1	6160	**325.82**
173.194.9	par03s06.c.youtube.com	-1	-1	-1	-1	-1	-1	-1	-1
208.117.236	par03x04.c.youtube.com	179	164.18	4250	540.16	957	496.91	-1	-1
208.117.248	mia02s11.c.youtube.com	-1	-1	77	552	-1	-1	-1	-1
208.117.250	ams09x06.c.youtube.com	41430	679	49437	656.39	57675	**653.81**	567	906.65
208.117.252	dfw06x02.c.youtube.com	-1	-1	51	285.63	-1	-1	-1	-1
208.117.254	fra07x03.c.youtube.com	838	667.29	2130	852.53	-1	-1	465	606.1
74.125.105	lhr22s16.c.youtube.com	1829	1551.78	1655	1185.94	3957	942.47	3454	990.64
74.125.13	zrh04s03.c.youtube.com	719	1074.15	499	2264.09	82	1302.03	-1	-1
74.125.14	mil02s01.c.youtube.com	48366	1234.82	37968	1253.01	37182	**1162.85**	47844	1298.45
74.125.216	bru02t11.c.youtube.com	-1	-1	-1	-1	-1	-1	-1	-1
74.125.218	fra07t13.c.youtube.com	8697	1355.33	12579	1338.71	8560	1239	11469	1256.32
74.125.4	lhr22s11.c.youtube.com	1496	1846.25	2488	1034.78	4146	**1363.63**	-1	-1
74.125.99	fra07s03.c.youtube.com	-1	-1	-1	-1	-1	-1	4221	**187.84**

mPlane

Something we observe

		5-May		6-May		7-May		8-May		9-May	
SUBNET	NAME with AIRPORT code	#flow	Tru avg	#flow	Tru avg	#flow	Tru avg	#flow	Tru avg	#flow	Tru avg
173.194.18	fra02s08.c.youtube.com	-1	-1	-1	-1	-1	-1	6139	**368.93**	6266	298.4
173.194.19	fra02s15.c.youtube.com	-1	-1	-1	-1	-1	-1	9940	**258.18**	12893	196.06
173.194.2	mil01s12.c.youtube.com	17054	1333.46	15470	1276.31	13655	**1259.63**	14186	1296.07	12616	1197.64
173.194.20	par08s06.c.youtube.com	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
173.194.208	par08s06.c.youtube.com	-1	-1	-1	-1	-1	-1	-1	-1	487	414.63
173.194.5	lhr14s08.c.youtube.com	449	1819.57	283	1658.45	-1	-1	3470	937.18	4222	1025.49
173.194.6	fra07s13.c.youtube.com	-1	-1	-1	-1	-1	-1	4924	**412.17**	8749	331.14
173.194.62	fra07s19.c.youtube.com	-1	-1	-1	-1	-1	-1	6160	**325.82**	6877	248.39
173.194.9	par03s06.c.youtube.com	-1	-1	-1	-1	-1	-1	-1	-1	87	355.1
208.117.236	par03x04.c.youtube.com	179	164.18	4250	540.16	957	496.91	-1	-1	-1	-1
208.117.248	mia02s11.c.youtube.com	-1	-1	77	552	-1	-1	-1	-1	-1	-1
208.117.250	ams09x06.c.youtube.com	41430	679	49437	656.39	57675	**653.81**	567	906.65	-1	-1
208.117.252	dfw06x02.c.youtube.com	-1	-1	51	285.63	-1	-1	-1	-1	-1	-1
208.117.254	fra07x03.c.youtube.com	838	667.29	2130	852.53	-1	-1	465	606.1	126	1146.87
74.125.105	lhr22s16.c.youtube.com	1829	1551.78	1655	1185.94	3957	942.47	3454	990.64	4116	1061.72
74.125.13	zrh04s03.c.youtube.com	719	1074.15	499	2264.09	82	1302.03	-1	-1	-1	-1
74.125.14	mil02s01.c.youtube.com	48366	1234.82	37968	1253.01	37182	**1162.85**	47844	1298.45	52594	1226.85
74.125.216	bru02t11.c.youtube.com	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
74.125.218	fra07t13.c.youtube.com	8697	1355.33	12579	1338.71	8560	1239	11469	1256.32	11633	1292.58
74.125.4	lhr22s11.c.youtube.com	1496	1846.25	2488	1034.78	4146	**1363.63**	-1	-1	-1	-1
74.125.99	fra07s03.c.youtube.com	-1	-1	-1	-1	-1	-1	4221	**187.84**	4913	189.64

Y

Who controls the traffic??

mPlane

Conclusions

- Today the Internet is a very complicated tangle
- Few big players control
  - In terms of content
  - In terms of information/ knowledge about the customers
- ISPs are being cut out of the picture
  - And they are struggling to control their network
  - … revenues are vanishing

Who will give you 100Mb/s?

mPlane

DNS to the rescue: seeing some light at the end of the tunnel

Ignacio Bermudez
Ram Keralapura
Marco Mellia
Maurizio Munafò
Antonio Nucci

Observe is the first step toward control

- Today picture:
  - Encryption ++ →
  - CDN ++ →
  - Cloud ++ →
- E.g.
  - Lives on facebook
  - Run on amazon web services
  - Relies on Akamai

- Tools?
  - No DPI
  - No IP server info
  - No spatial correlation
    - It changes over time
    - It changes over space

Is there a way to understand…

- What the users are doing?
  - Which service is popular?
  - Which impact it has?
  - Is there any QoS/ QoE problem?
- What is the network doing to serve the content?
  - Which CDN is being used?
  - Which datacenter is being used?
  - Is there a better way to do it?

Add mobile devices into this nightmare…

mPlane

What ISPs want

- The boss asks Netadmin to setup fine-grained SLA
  - ~~… answer is "NO"~~
  - Low-latency for Gmail and Dropbox
  - YouTube must not exceed 10Mbps
  - Facebook permitted, but not Zynga games
- Above services can use HTTPS, so classical tools fail

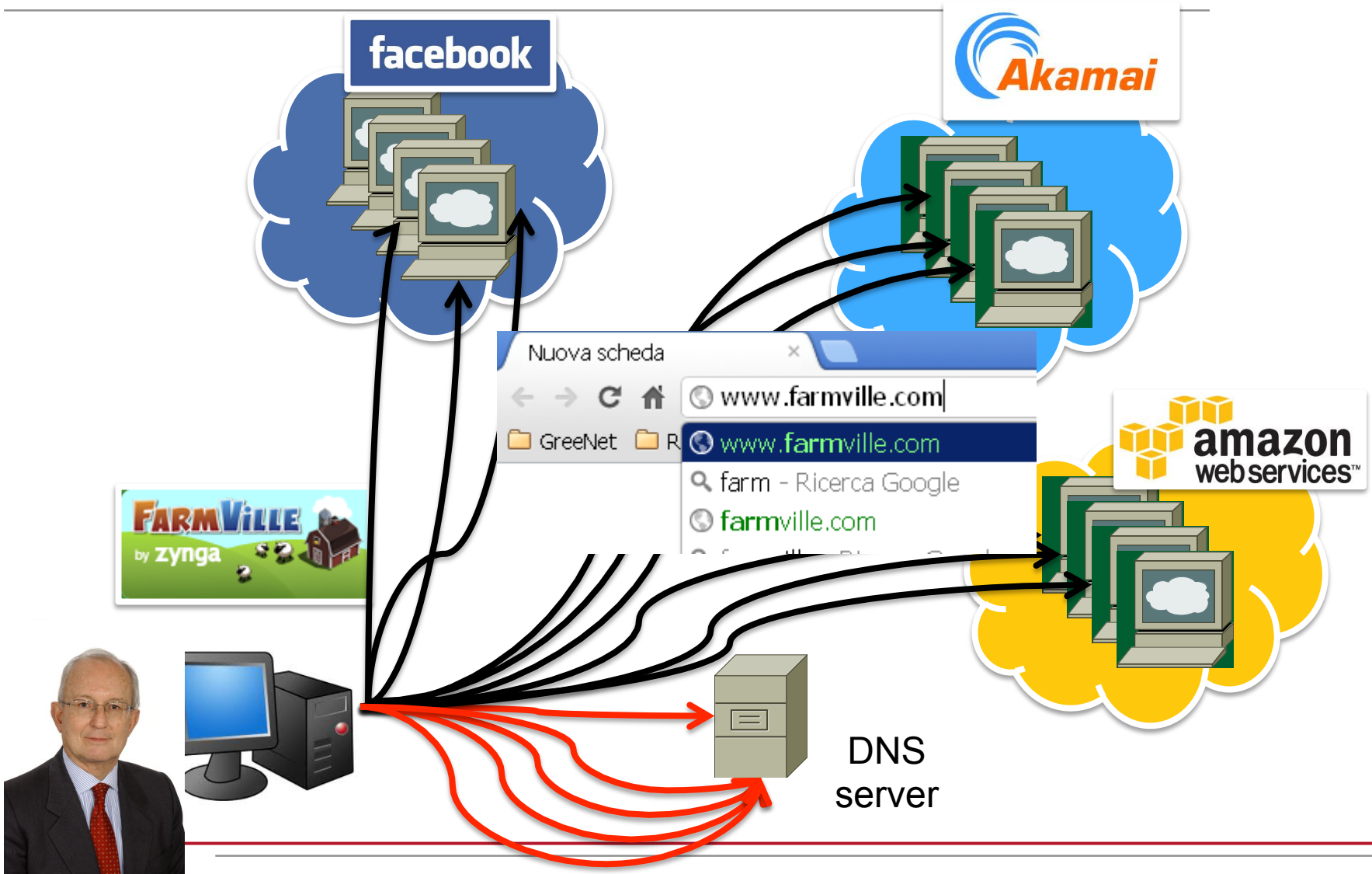
SLA in terms of CONTENT

- ISP-A has a peering-link with Google
  - It is used to access the preferred cache
- Google redirects requests to another cache
  - ISP-A traffic goes through another expensive link
- How to reveal this and help ISP-A to optimize paths to route traffic back through the original link?

What network administrator want

- Netadmin sees lot of requests going to IP 173.194.78.141
  - `wg-in-f141.1e100.net`
  - owned by Google
- Protocol is unknown
  - Some binary protocol
- Should Maurizio block it?
  - That could be `www.google.com`
- Boss is asking to block requests to Farmville
  - it runs on **Amazon**
- And to improve performance of Dropbox
  - It runs on **Amazon**
- Netadmin's firewall would either block both, or let everyone enjoy Farmville!

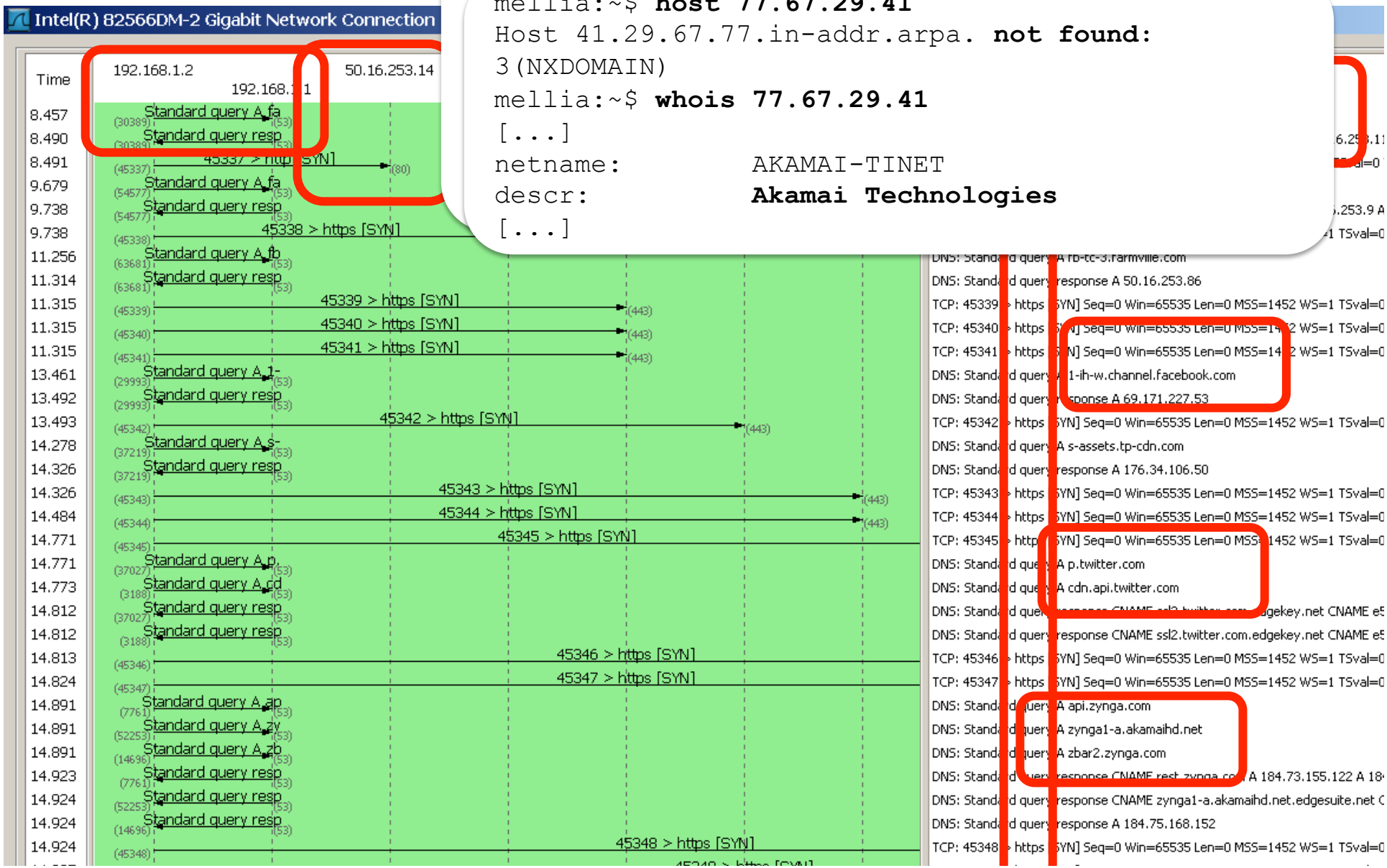
The intuition



mPlane

The intuition

```
mellia:~$ host 50.16.253.14
mellia:~$ host 77.67.29.41
Host 41.29.67.77.in-addr.arpa. not found:
3(NXDOMAIN)
mellia:~$ whois 77.67.29.41
[...]
netname:        AKAMAI-TINET
descr:          Akamai Technologies
[...]
```



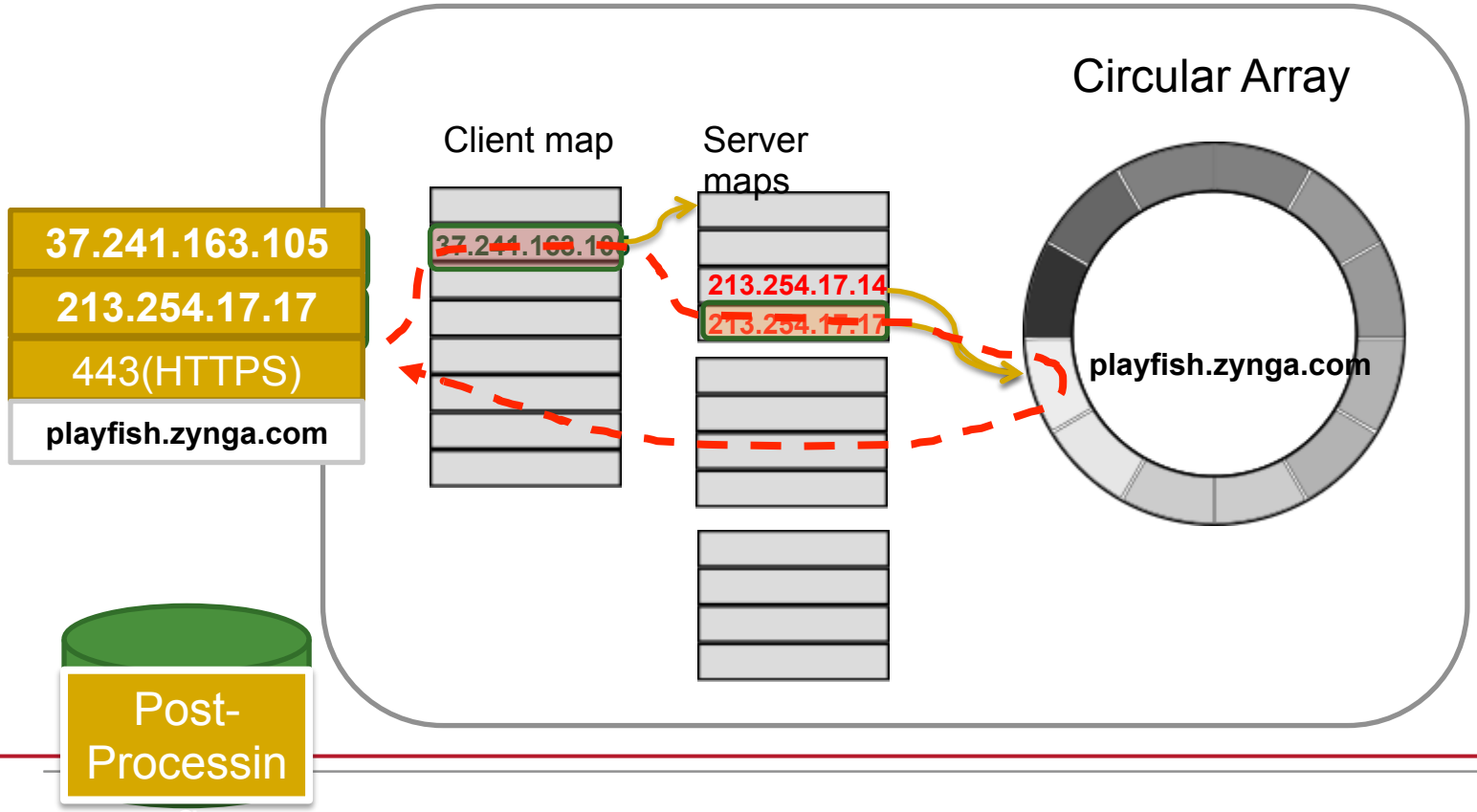
- Observation: Most **client-server applications need DNS** to retrieve the server IP-address

- Key idea: Associate **network flows** with their corresponding **domain name**

```
A name indicates what we seek.
An address indicates where it is.
A route indicates how we get there.

        -- Jon Postel (1943-1998), RFC 791, "Internet
Protocol", 1981
```

37.241.163.105	SRC_IP
213.254.17.17	SERVER IP
443(HTTPS)	PORT

TCP/UDP Flows



Why this is useful?

- By exploring the FQDN information we can
  - Unveil which content the user is requesting
    - Even if HTTPS is used
    - Even when served by a generic CDN
  - We can track how the CDN replies
    - Observe if association changes over time
    - Unveil how many possible servers/CDNs can serve the same content
    - Optimize traffic in the ISP network
  - Discover additional information by mining the FQDN
    - Which service runs on port 1080, 1337 or 5223?
    - Which service IP 173.194.35.50 handles?

Service Tag extraction

```
TAG EXTRACTION(dPort, limit)
Input: targeted dPort, limit of tags to return
Output: The ranked list of tags
DomainNameSet ← FlowDB.query(dPort)
for all FQDN in DomainNameSet do
    TokenSet ←
    DomainName.split(NoTLD|No2ndDomain)
end for
for all Token in TokenSet do
    Token.score.update()
end for
Return(Tokens.sort(limit))
```

Algorithm 2: L4-Analyzer service tag extraction pseudo-code

- Aim: group all flows that are destined to the same server port
- Analyze FQDN terms to extract most frequent one
- Can they provide hints on the service running on such port?

Service Tag extraction

- Maurizio sees lot of flows going to port 6969
  - DPI cannot identify it
- What is that?

- Group all flows going to port 6969
- Extract FQDNs
- Analyze frequencies of terms
- Get the answer

Content discovery

```
CONTENT DISCOVERY(ServerIPSet)
Input: The list of targeted serverIP
Output: The list of handled FQDNs
DomainNameSet ← FlowDB.query(ServerIPSet)
for all FQDN in DomainNameSet do
    TokenSet ← DomainName.split(FQDN)
end for
for all Token in TokenSet do
    Token.score.update()
end for
Return(Tokens.sort())
```

Algorithm 3: L4-Analyzer Content Discovery pseudo-code

- **Which services are being used by customers?**

Some HTTPS services running on Google

Plane

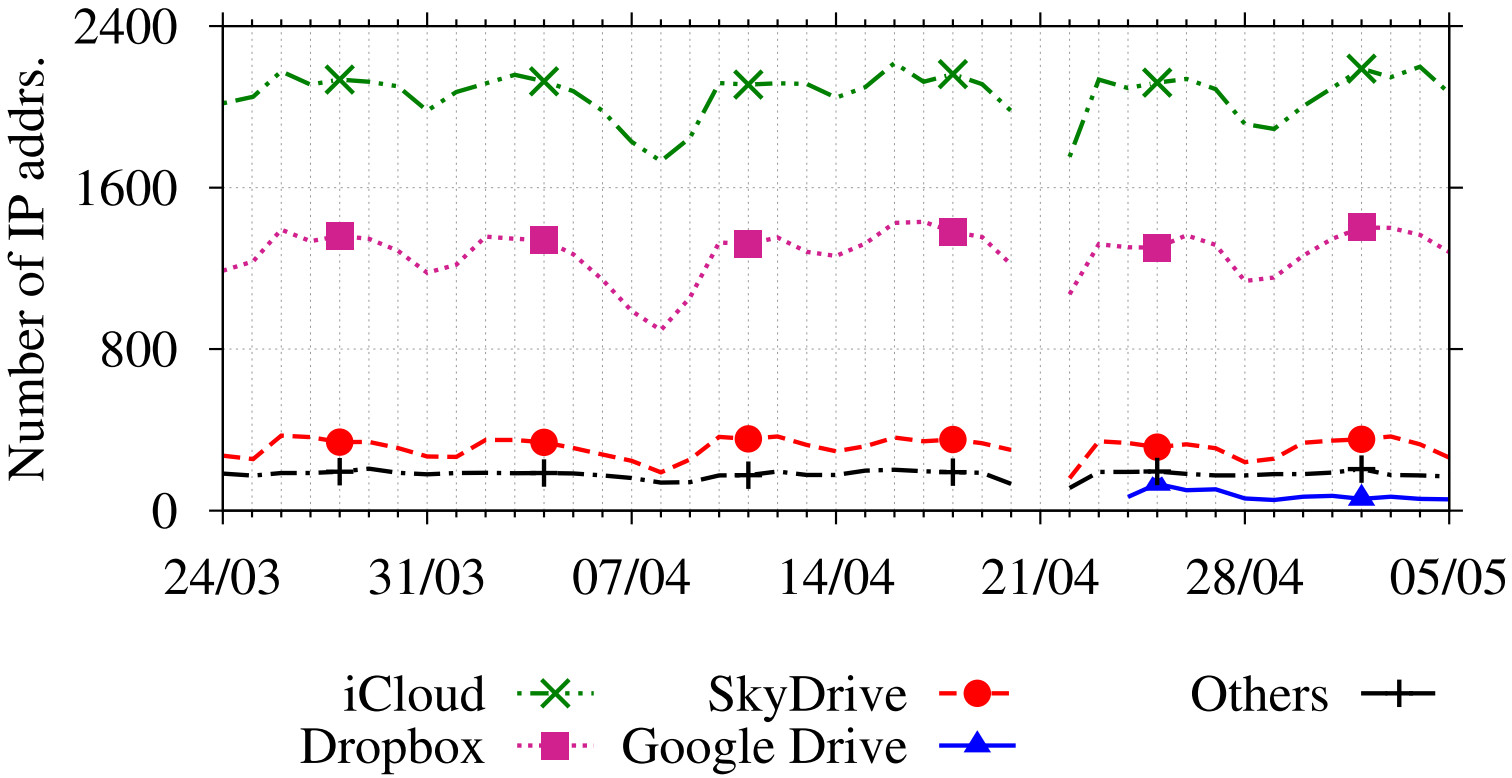
The most popular cloud storage?

- Someone wants to know which is the most popular cloud storage service

- Given a definition of popular
  - Number of hosts generating traffic to ...
- Look at FQDN
  - *dropbox*
  - *drive.google*
  - *icloud*
  - *skydrive*
  - …

The most popular cloud storage?

- And now you can write a paper about Dropbox ☺



Drago, I., M. Mellia, M. Munafo', A. Sperotto, R. Sadre, and A. Pras, "*Inside Dropbox: Understanding Personal Cloud Storage Services*", Internet Measurement Conference - IMC, Boston, MA, ACM, 11/2012

Spatial discovery

```
SPATIAL DISCOVERY(FQDN)
Input: The targeted FQDN
Output: ranked list of serverIP addresses
2ndDomain ← FQDN.split()
ServerSet ←
FlowDB.queryByDomainName(2ndDomain)
FQDNset ← 2ndDomain.query()
for all FQDN in FQDNSet do
    FQDN.ServerSet ←
    FlowDB.queryByDomainName(FQDN)
end for
Return(FQDN.ServerSet.sort(), ServerSet.sort())
```

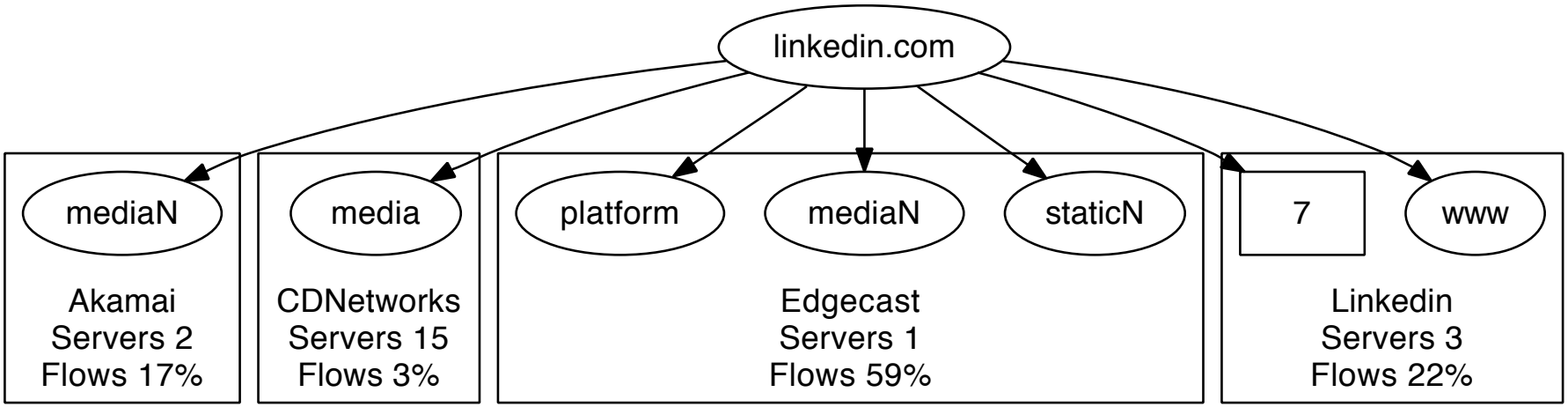
Algorithm 4: FQDN-Analyzer Spatial Discovery pseudo-code

- **Who serves a given content?**

mPlane

LinkedIn "cloud"

- How many servers are used by linkedin.com?
- Which CDNs are involved?



About cloud and FQDNs…

- Do you know where are the Amazon datacenters?
  - Look at the FQDNs
  - Hint: Check the IATA airport code
- Pretty popular way to assign FQDNs
  - Same for YouTube

```
ams1 r.cloudfront.net.
ams5 .r.cloudfront.net.
arn1 r.cloudfront.net.
cdg3 r.cloudfront.net.
cdg5 .r.cloudfront.net.
dfw3 r.cloudfront.net.
dfw5 .r.cloudfront.net.
dub2 r.cloudfront.net.
ewr2 r.cloudfront.net.
fra2 r.cloudfront.net.
fra6 r.cloudfront.net.
gru1 r.cloudfront.net.
hkg1 r.cloudfront.net.
iad1 .r.cloudfront.net.
iad2 r.cloudfront.net.
ind6 r.cloudfront.net.
jax1 r.cloudfront.net.
jfk1 r.cloudfront.net.
jfk5 r.cloudfront.net.
lax1 r.cloudfront.net.
lax3 r.cloudfront.net.
lhr3 r.cloudfront.net.
lhr5 r.cloudfront.net.
mia3 r.cloudfront.net.
mxp4 r.cloudfront.net.
nrt5 .r.cloudfront.net.
nrt5 .r.cloudfront.net.
sea4 r.cloudfront.net.
```

What is known about Amazon cloud?

- And now you can write a paper about Amazon ☺

	ID	#IPs		Exchanged Data (%)		Avg. RTT [ms]	β^{RTT} [ms]		β^{AS}	β^{km} [km]
		EC2	S3	EC2	S3		EC2	S3		
Datacenters	IAD	6429	121	85.31%	64.22%	132.13	113.97	116.18	3	6709
	DUB	1167	24	12.65%	35.14%	45.10	48.73	43.77	3	1365
	SJC	632	12	1.71%	–	203.06	182.14	174.81	4	9556
	NAR	18	0	–	–	298.67	–	–	4	9843
	SIN	71	0	0.03%	–	235.60	228.10	–	3	10390
	SEA	0	32	–	0.02%	196.04	–	214.79	4	8617
				97.26GB	37.13GB					

	ID	#IPs	Exchanged Data (%)	Avg. RTT [ms]	β^{RTT} [ms]	β^{AS}	$\beta^{km}[km]$
Caches	IAD	2	–	132.13	102.75	3	6709
	DUB	222	0.05%	45.10	49.76	3	1365
	SJC	–	–	–	–	4	9556
	NAR	115	–	298.67	–	4	9843
	SIN	51	–	235.60	–	3	10390
	SEA	64	–	196.04	–	4	8617
	SFO	253	0.83%	172.80	175.21	4	9537
	CDG	246	0.13%	32.09	38.43	3	584
	FRA	245	0.17%	19.56	21.87	2	566
	MXP	232	98.03%	18.34	21.26	3	124
	EWR	208	–	105.28	109.53	3	6380
	AMS	205	0.04%	23.94	29.88	3	837
	LHR	182	0.17%	31.84	31.60	3	920
	ANR	151	0.56%	41.02	41.48	3	1734
			104.19GB				

I. Bermudez, S. Traverso, M. Mellia, M. Munafo,
"*Exploring the Cloud from Passive Measurements: the Amazon AWS Case*"
IEEE INFOCOM, Turin, Italy, April 14-19, 2013

Comparing performance of Open Resolver

- Is it worth to use GoogleDNS or OpenDNS instead of the ISP DNS resolver?

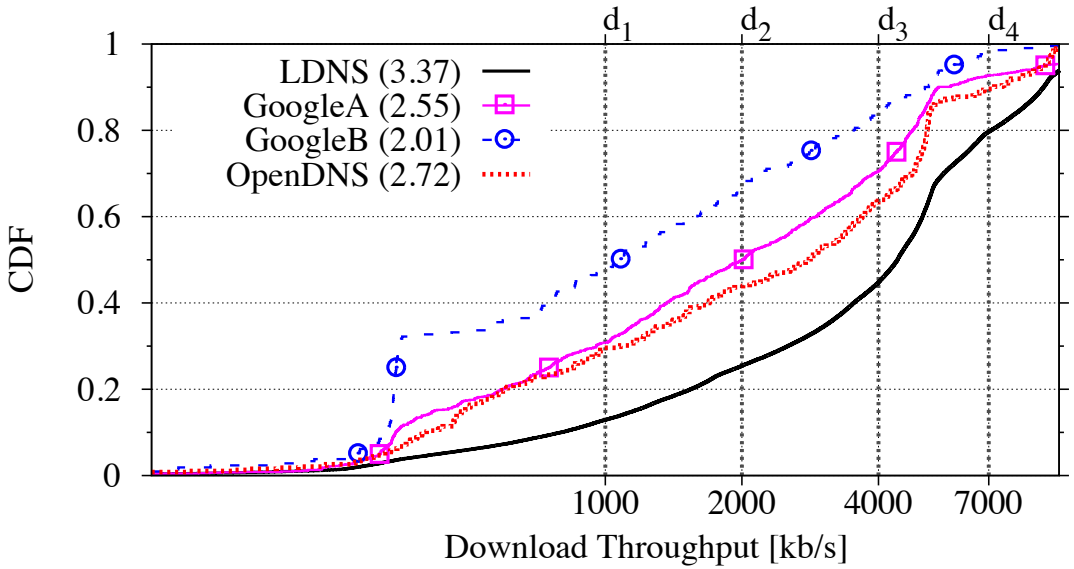
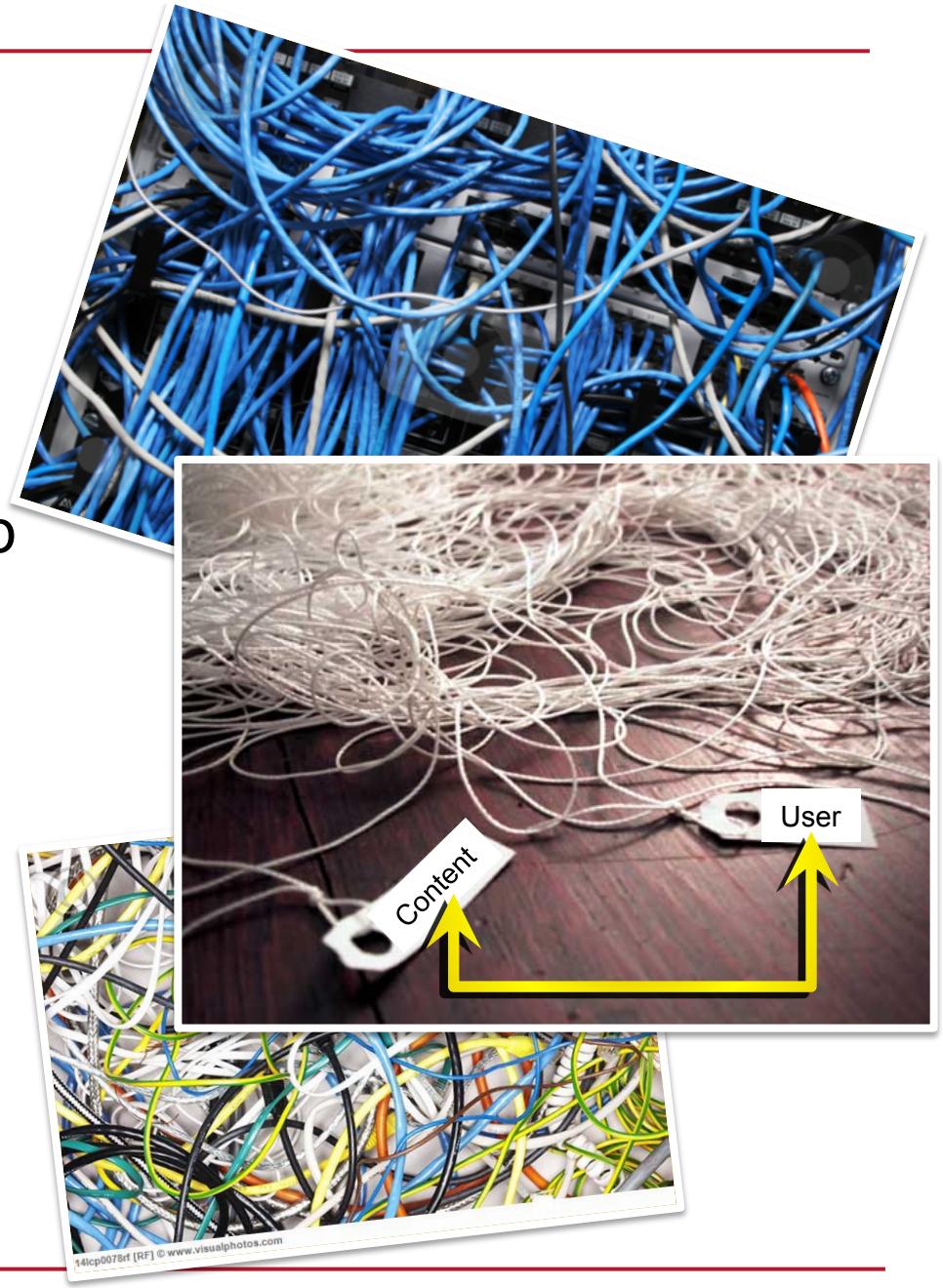


Figure 7: Average download throughput for Apple iTunes/App. Scores are reported in brackets.

I. Bermudez, A. Finamore, M. Mellia,
"Public DNS Resolvers: Friends or Foes? Implications of Resolver Choice for ISP Customers", *Submitted to IMC'13*

Conclusions

- The DNS is the link to discern services and content in the tangled web
- DNS naturally exposes information about
  - The content and the service being accessed
  - The way this content is served by CDNs/clouds
- It gives back network visibility



mPlane

Some references

1. A.Finamore, M.Mellia, M.Meo, M.Munafò, D.Rossi “Experiences of Internet Traffic Monitoring with Tstat IEEE Network”, March/April 2011", Vol. 25, No. 3, March 2011.
2. R.Torres, A.Finamore, J.R.Kim, M.Mellia, M.Munafò, S.Rao, “Dissecting Video Server Selection Strategies in the YouTube CDN,” ICDCS, Minneapolis, MN, June 2011.
3. V.Gehlen, A.Finamore, M.Mellia, M.Munafò “Uncovering the Big Players of the Web”, Traffic Monitoring and Analysis - 4th International Workshop, TMA 2012 Vienna, March 2012.
4. A.Finamore, V.Gehlen, M. Mellia, M.Munafò, “**The Need for an Intelligent Measurement Plane: the Example of Time-Variant CDN Policies**”, Networks, Rome, IT, October 2012.
5. I.Bermudez, M.Mellia, M.Munafò, R.Keralapura, A.Nucci, “**DNS to the rescue: Discerning Content and Services in a Tangled Web**”, ACM IMC, Boston, Nov. 2012.
6. I.Drago, M.Mellia, M. Munafò, R.Sadre, A.Sperotto, A.Pras, “Inside Dropbox: Understanding Personal Cloud Storage Services”, ACM IMC, Boston, Nov. 2012.
7. I.Bermudez, S.Traverso, M.Mellia, M.Munafò, “**Exploring the Cloud from Passive Measurements: the Amazon AWS case**”, IEEE Infocom ’13.

More from http://tstat.tlc.polito.it/publications.php


Perguntas
Domande
Galdera
Fragen
Otázky
Questions
Spørgsmål
Pertanyaan
kysymykset
Frågor
Spørsmål
Cwestiynau
вопросы
Preguntes
Sorular
Въпроси
Vragen
Pytania