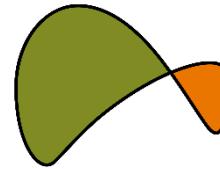


The Online Tracking Horde: a View from Passive Measurements

Hassan Metwally
Stefano Traverso
Marco Mellia
Stanislav Miskovic
Mario Baldi



NARUS



POLITECNICO
DI TORINO

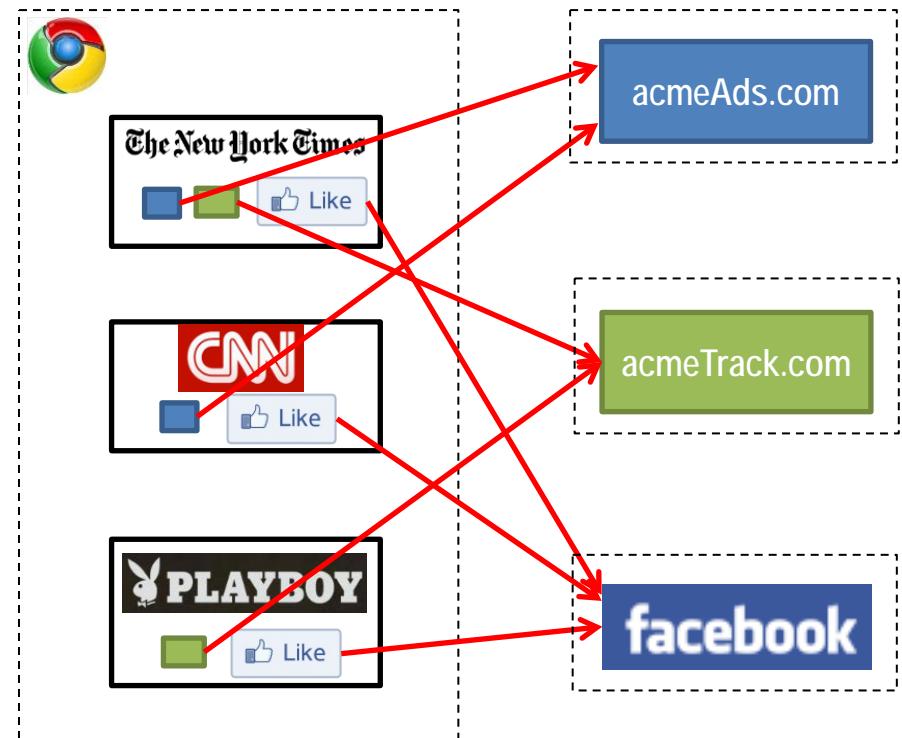
DET

Department of Electronics and Telecommunications

7° International Workshop on Traffic Monitoring and Analysis
Barcelona, Spain

Third-Party Trackers

- Third-party web tracking refers to the practice by which a service records user web activities often for profit
- Many techniques
 - Cookies
 - HTML5 LocalStorage
 - Finger printing (browser/OS/IP)
- Some are visible
 - Like
- Others are not
 - Like



Third-Party Trackers In Real Scenario

4w Marketplace
Acuity Ads
Adform
AppNexus
Audience Science
BidSwitch
BidTheatre
Brandscreen
Connexity
DataXu
Digilant
Dotomi
DoubleClick
Drawbridge
Facebook Connect
Facebook Exchange (FBX)
Facebook Social Plugins
Google Adsense
Google Analytics
Google+ Platform
Infectious Media
Klux Digital
Media Innovation Group
Media Optimizer (Adobe)
MediaMath
NetRatings SiteCensus
NetSeer
OwnerIQ
Quantcast
RadiumOne
Rubicon
ScoreCard Research Beacon
Simpli.fi
Sizmek
taboola
Tapad
TradeDesk
Turn
Twitter Button
Veruta
Videology
Webtrekk
Xaxis



POLITECNICO
DI TORINO

DE
De

la Repubblica | Mobile | Facebook | Twitter | Google +

R.it | ECONOMIA & Finanza con Bloom

sonale AFFARI & FINANZA Osserva Italia

SCORECARD RESEARCH

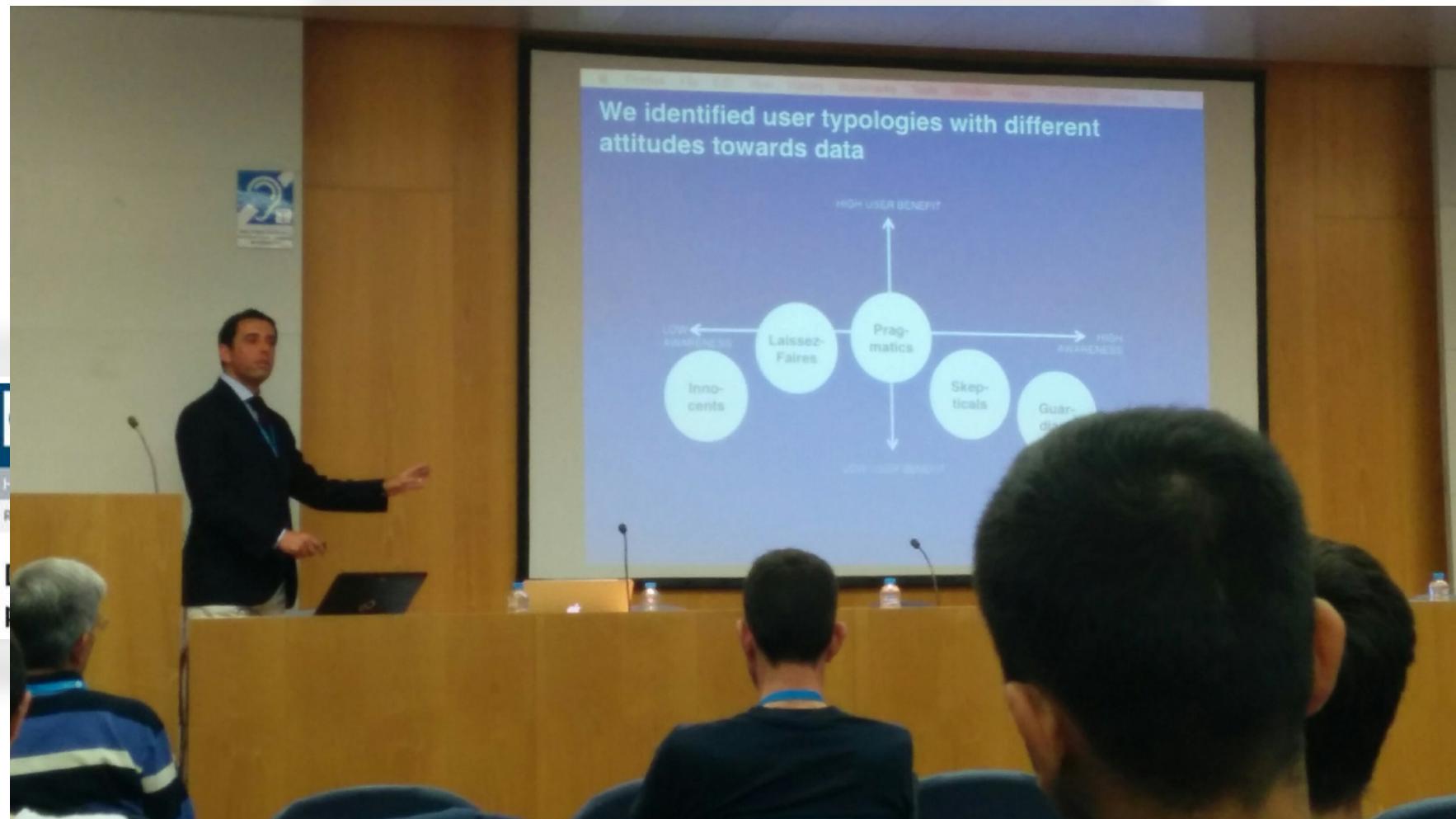
SL 20 MEGA SENZA LIMITI

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

Chi naviga sul sito. «Lo usiamo per sviluppare il prodotto, per migliorare l'esperienza utente». Modificare il carattere di una scritta o il layout di una pagina per far impennare il numero di click. Comprare una pubblicità online, calcolando al centesimo il ritorno atteso. Così importante che nella piramide aziendale Conforti, un passato da consulente strategico, risponde direttamente all'amministratore delegato. «Non puoi gestire ciò che non misuri», sintetizzano due guru dell'economia digitale, Andrew McAfee e Erik Brynjolfsson. Ecco perché i big data e la business intelligence, la possibilità di raccogliere e processare moli di input fino a ieri impossibili da mettere in fila, sono una rivoluzione per il management. Più informazioni uguale migliore conoscenza dell'azienda.

Trackers...why?



POLITECNICO
DI TORINO

DET
Department of Electronics and Telecommunications

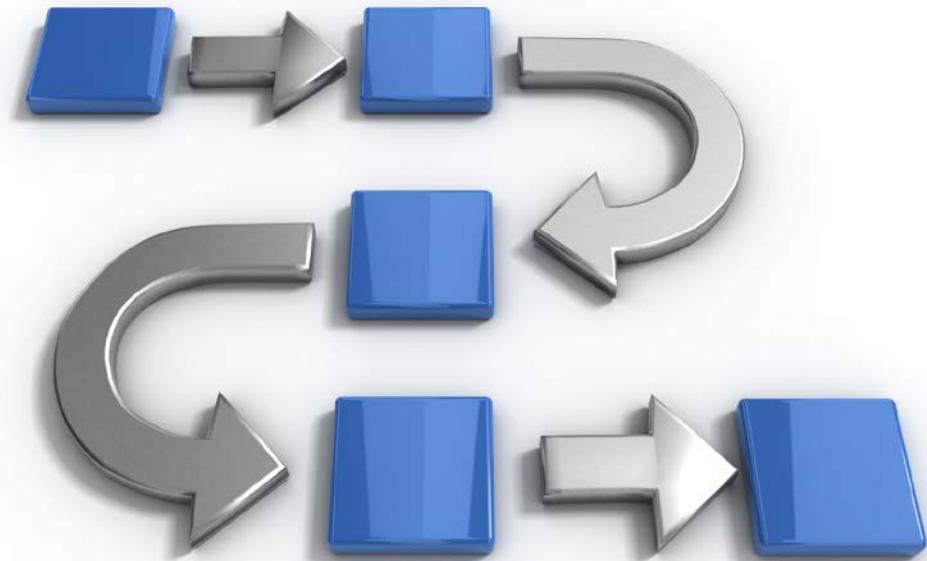
The Online Tracking Horde: a View from Passive Measurements

Our Goal

- Quantifying the pervasiveness and extensiveness of online tracking
 - 1. Monitoring the activity of trackers
 - 2. Via Passive Measurements
 - Factor the natural interactions among users and trackers
 - Past works used active crawling to get a first picture of the problem
- To Increase awareness of
 - Users
 - Regulators
 - Practitioners



Methodology and Dataset



**POLITECNICO
DI TORINO**

DET
Department of Electronics and Telecommunications

The Online Tracking Horde: a View from Passive Measurements

Methodology

1. List of possible third party tracking services

- Firefox + Ghostery
- Abine
- Manual identification



List of Trackers



Ghostery Identification



Abine List



Manual Identification

443 distinct tracking services



POLITECNICO
DI TORINO

DET
Department of Electronics and Telecommunications

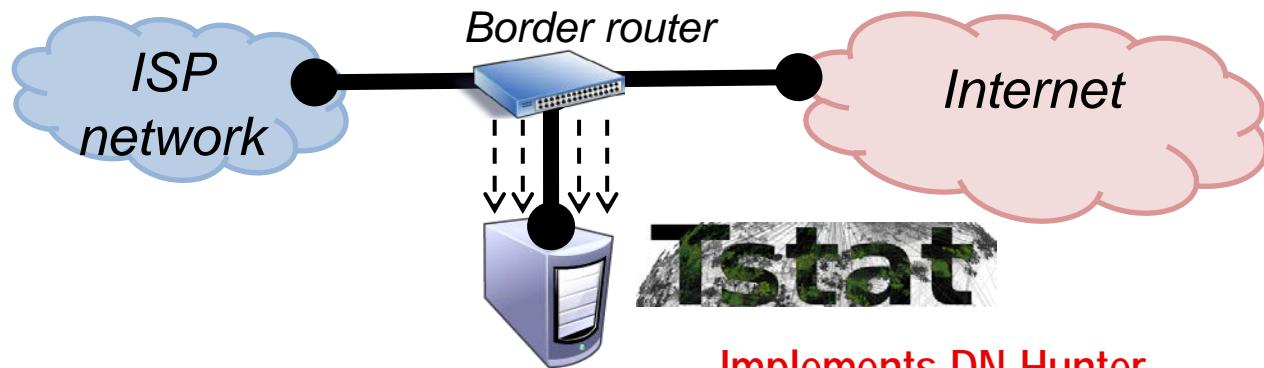
The Online Tracking Horde: a View from Passive Measurements

Methodology

1. List of possible third party tracking services
 - Firefox + Ghostery
 - Abine
 - Manual identification
2. Passive Monitoring of Internet Traffic
 - Tstat + DN-Hunter



Tstat + DN-Hunter



Tstat

- Captures traffic on the network interface and processes it **real-time**
- Statistics computed at the end of the flow

Implements DN-Hunter

- Track DNS conversations needed to retrieve the server IP-address
- Associate **network flows** with their corresponding **domain name**

c_ip	c_port	c_pkts	c_bytes	s_ip	s_port	s_pkts	s_bytes	FQDN
12.132.54.94	1197	1	18938	87.250.137.92	443	1	992221	Acme.com

Tstat

DN-Hunter

c_ip	//	time_abs	method	//	FQDN	referer	User_agent
12.132.54.94	//	1427293122663	GET	//	Scorecardresearch.com	Acme.com	Mozilla/5.0....



Methodology

1. List of possible third party tracking services
 - Firefox + Ghostery
 - Abine
 - Manual identification
2. Passive Monitoring of Internet Traffic
 - Tstat + DN-Hunter
3. Dataset



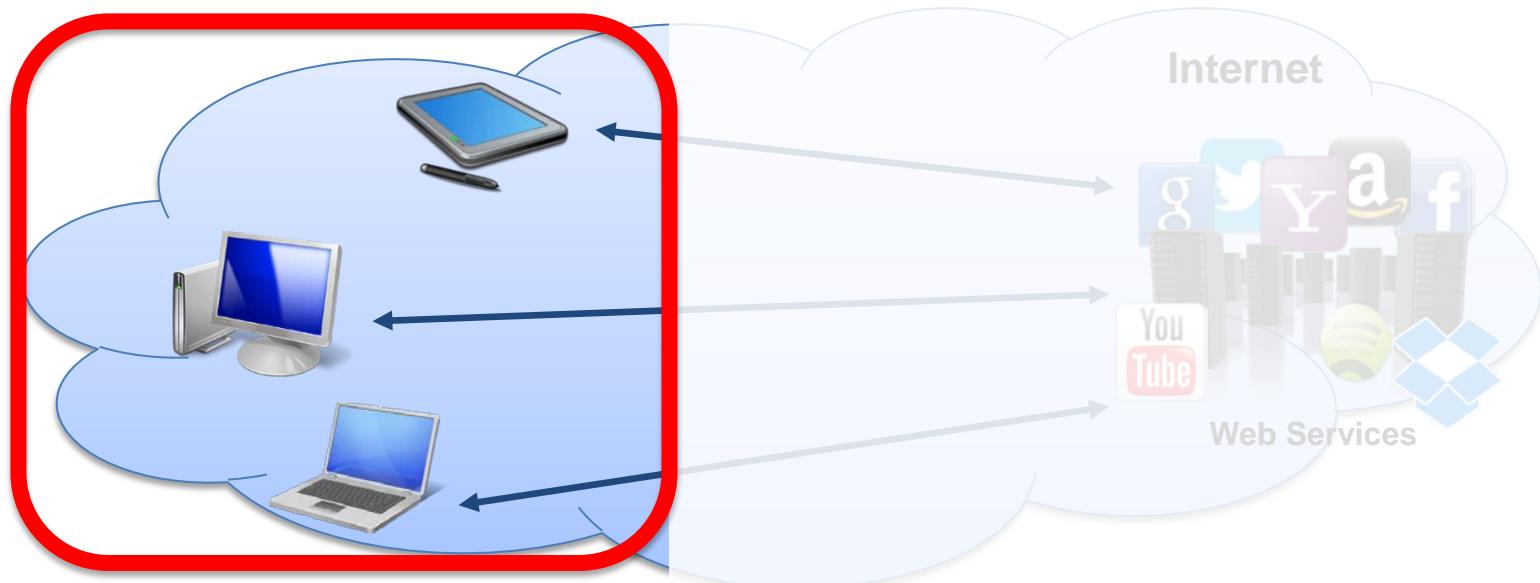
Dataset

- Tstat installed at three PoPs of an Italian ISP
- 2 weeks of data collection
- Focus on 13,000 households
- IP addresses are anonymized, URL are shortened at the FQDN
- Consider only aggregated data

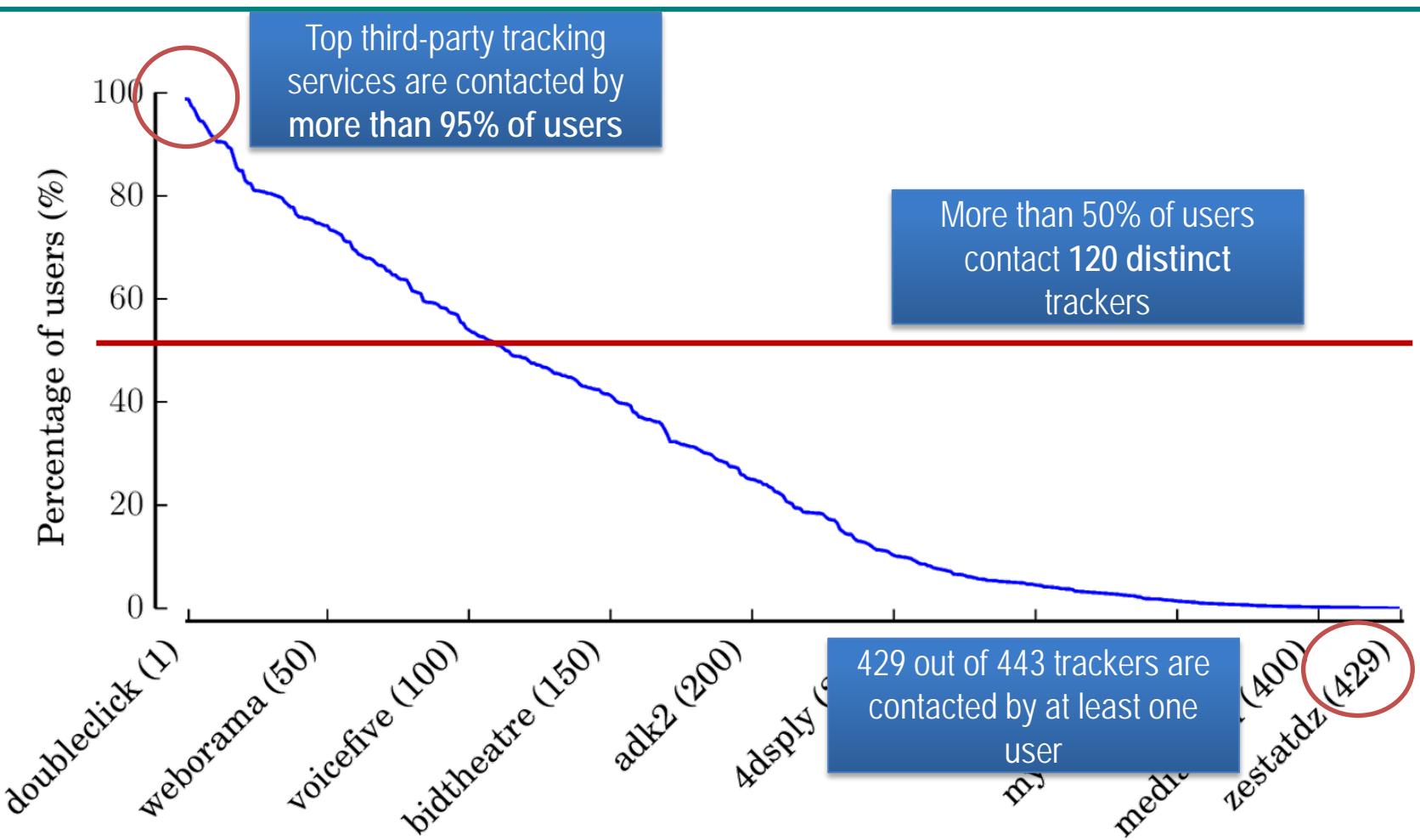
Trace	Households	Connections	Period
Vp1-10d	13408	1M	13-23/10/2014
Vp1-1d-05/14	10458	238K	07/05/2014
Vp1-1d-05/13	12218	240K	08/05/2013
Vp1-1d-05/12	11660	200K	09/05/2012



Quantify tracking activity



Penetration – 10 days



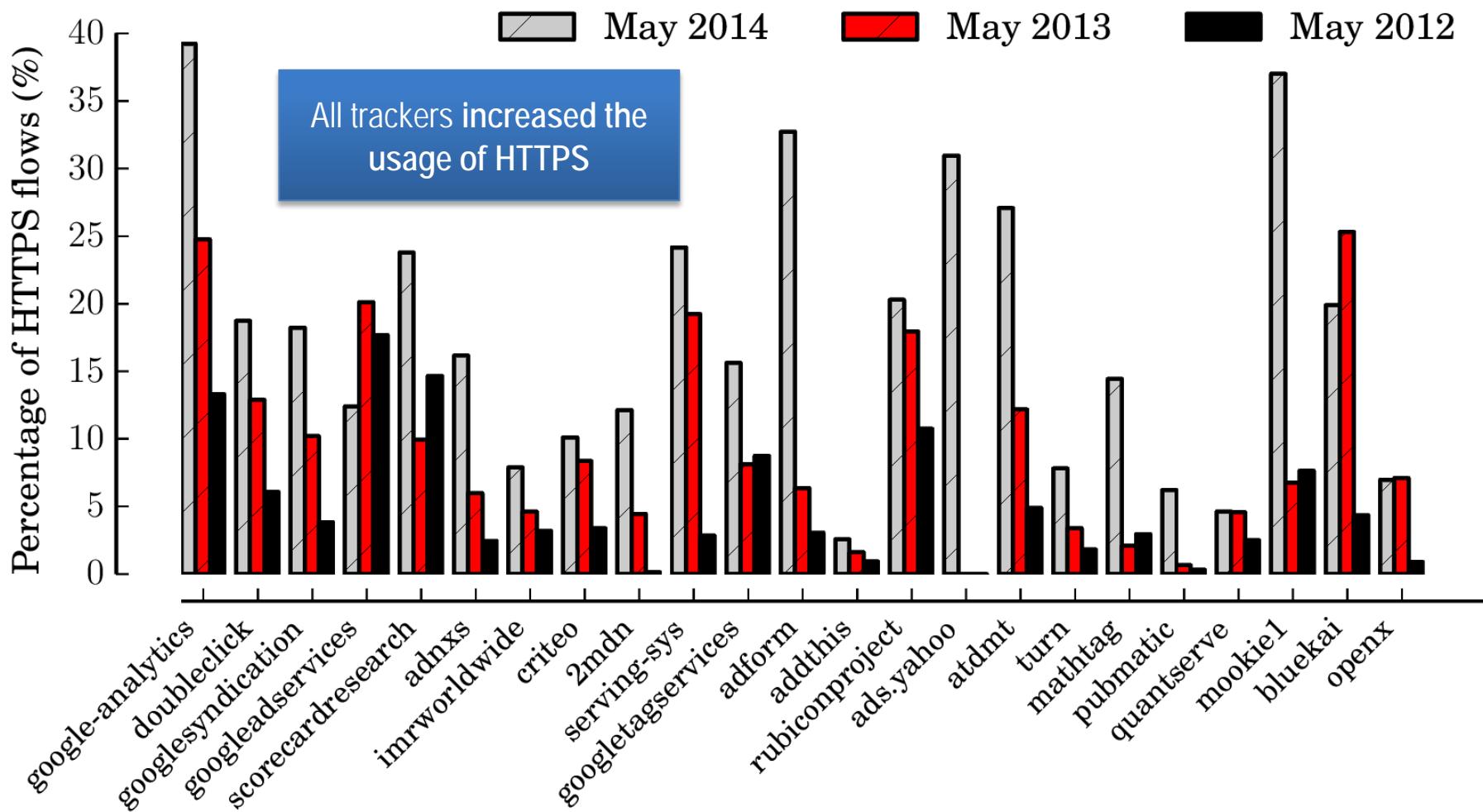
Privacy and tracking: HTTP or HTTPS?

- HTTPS usage is increasing [1]
 - This clearly improves people privacy from the network point of view
 - But it makes it harder to verify tracking services from a neutral third party observing traffic...
- How many third party tracking services are using HTTPS?

[1] Naylor, D. et al. The Cost of the "S" in HTTPS. In: ACM CoNEXT. (2014)



Privacy and tracking: HTTP or HTTPS?



Time to Be Tracked

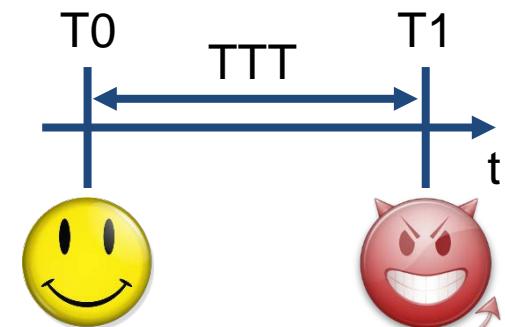
- How much time a user spends online before encountering the first tracker?

T_0 = time of first flow generated by user

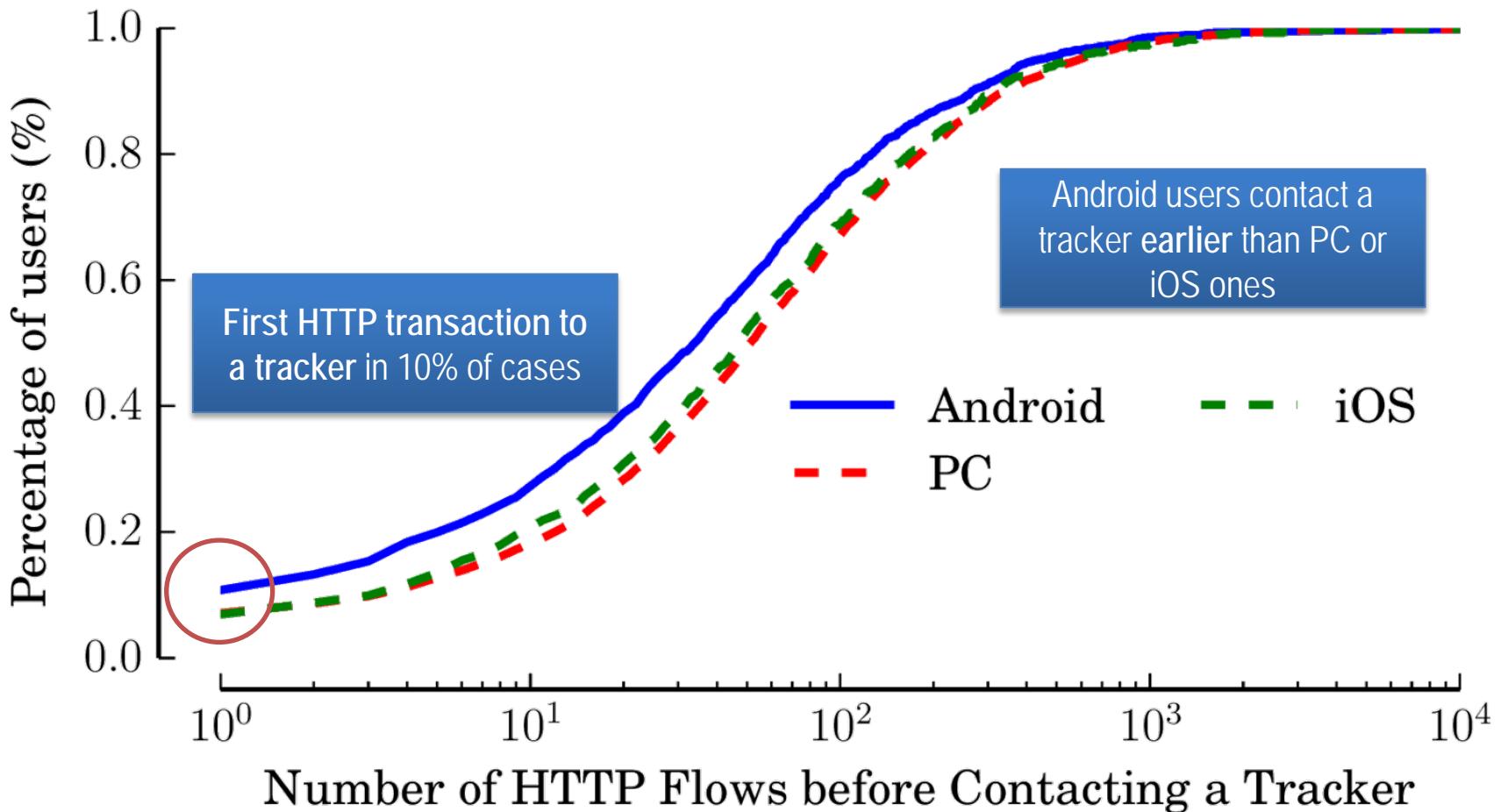
T_1 = time of first flow to any tracker

TTT (Time-To-Tracker) = $T_1 - T_0$

- TTT smaller than 1 s in 77% of PC users
- TTT smaller than 100 s for all PC users
- Let's zoom into this
 - Request to first tracker instead of TTT

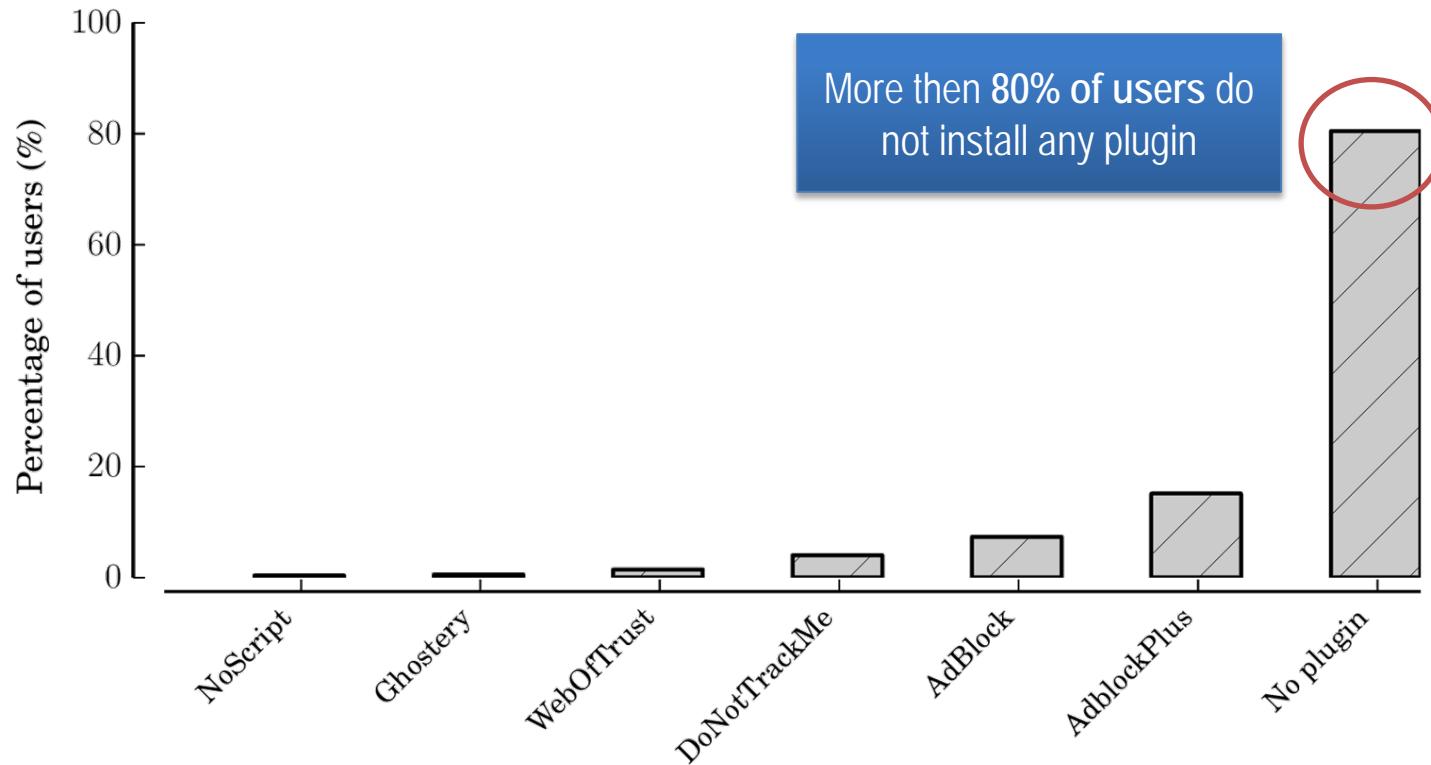


Requests to first tracker

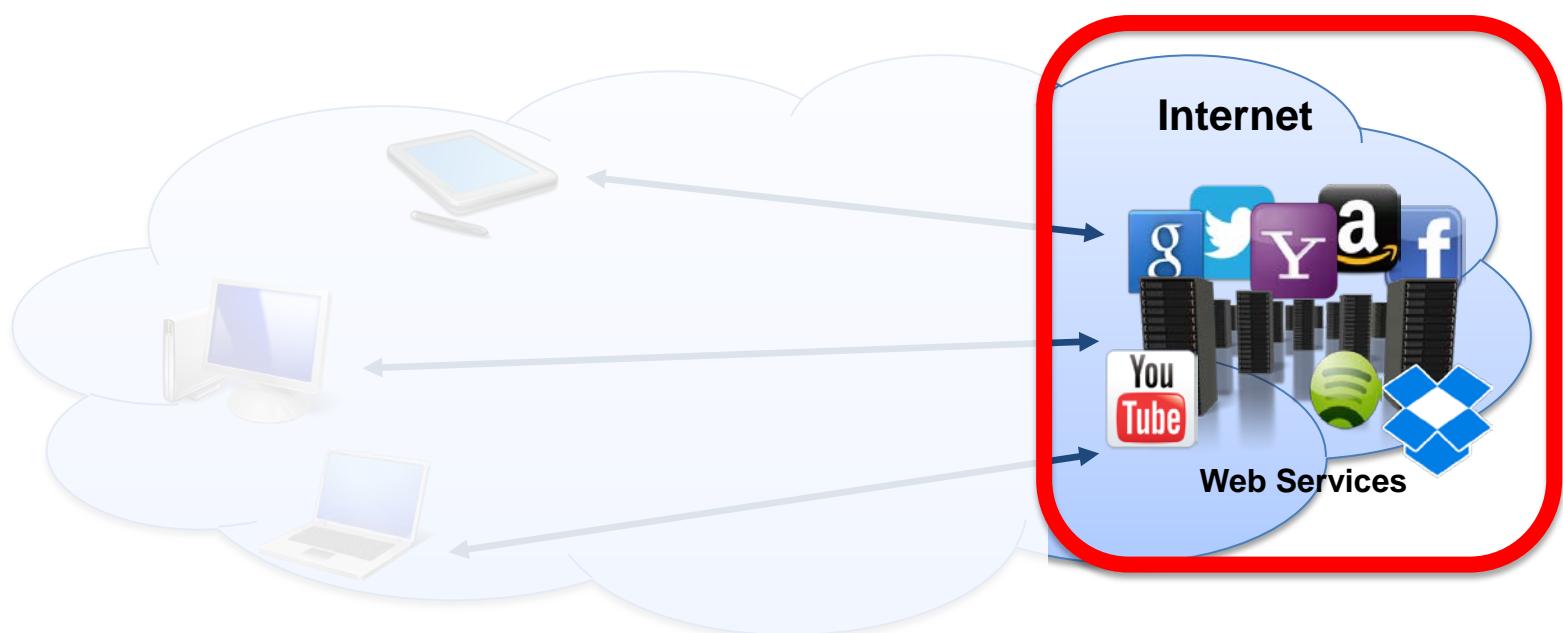


Countermeasures?

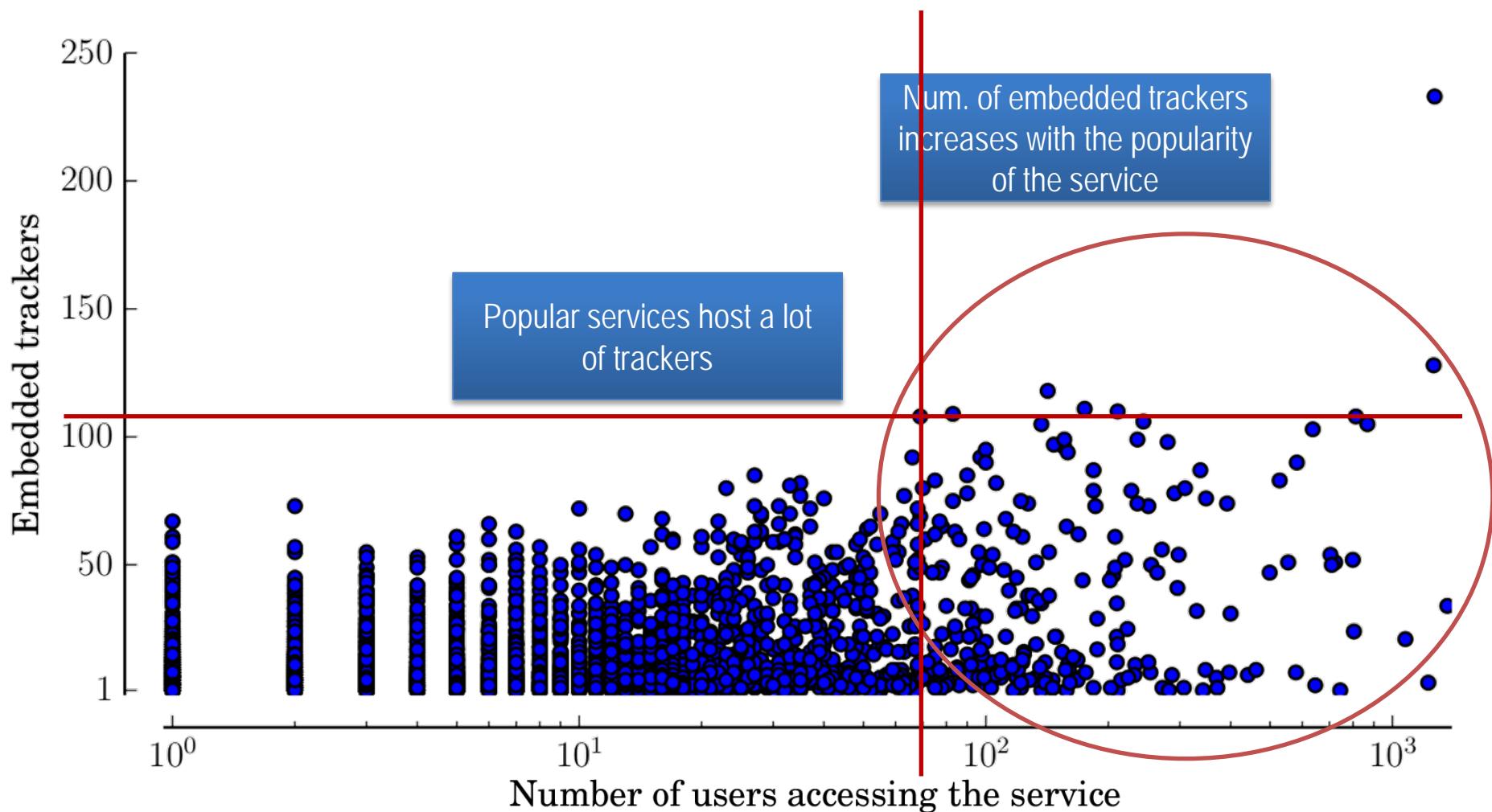
- Some countermeasures are available as **browser extensions**
 - Disable cookie sending, disable javascript, blacklisting,...
-but few users use them



Quantify tracking activity



Trackers per Service



Conclusions

Are you tracked??

Top third-party tracking services are contacted by **more than 95% of users**

77% of users contact the first tracker in less than 1 second

All trackers **increased the usage of HTTPS**

71% of services embed at least one tracking service

Yes, you are!!



THANKS FOR YOUR ATTENTION!



Hassan Metwally – Politecnico di Torino
<hassan.metwally@polito.it>



POLITECNICO
DI TORINO

DET
Department of Electronics and Telecommunications

The Online Tracking Horde: a View from Passive Measurements

Trackers per Service

