Privacy on the Web

Gábor György Gulyás
gulyas.info // @GulyasGG
Laboratory of Cryptography and System Security (CrySyS)
Budapest University of Technology and Economics
www.crysys.hu

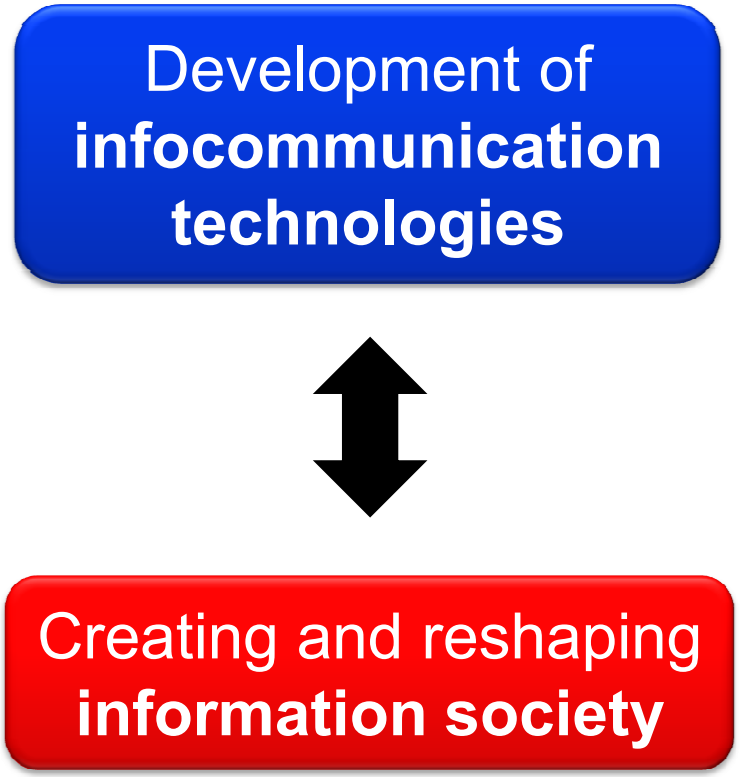
Overview and goal of the talk

Goal: clarify the attacker model concerning online privacy.

- Privacy intro
  - Why it is important
  - How can we define it
  - Web related overview
  - Economic background
- Tracking users on the web
  - Storage based tracking
  - Fingerprinting
- Enhancing privacy
  - Debunking some misbeliefs
  - Guidelines for protecting privacy
- Conclusion

DO WE NEED PRIVACY?

We live in Surveillance? societies



- **Government**: efficiency, ease of communication, greater control, ...
- **Commercial parties**: new technology & data, great opportunities, ...
- **Society itself**: comfort, changes in social interaction, ...

„Anyway, who cares?”

Who would abuse my privacy?

I have nothing to hide! ☺

Think again.

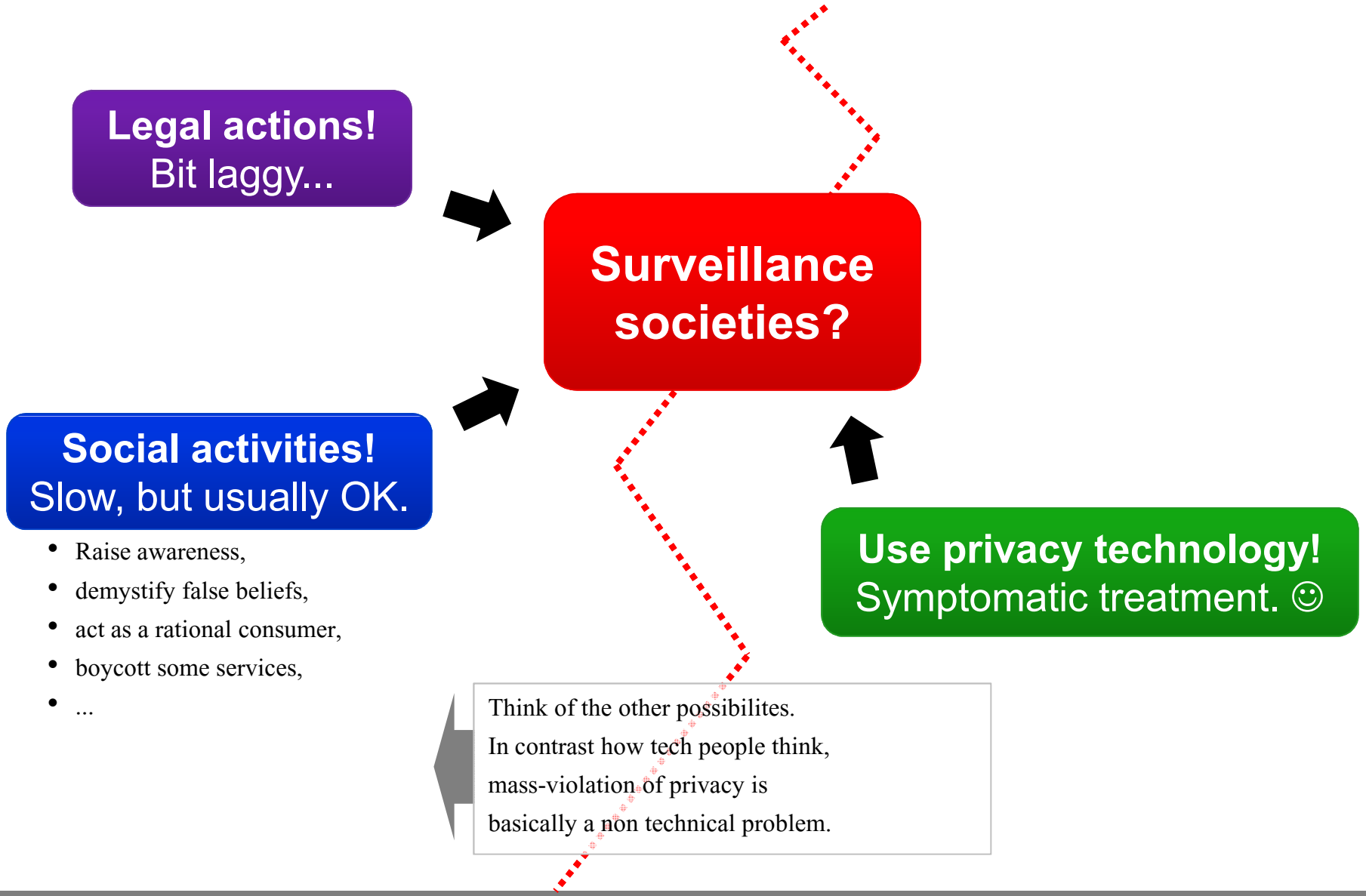
Isn’t it all already lost?

Nope.

Effect on a personal level?

Freedom of speech?

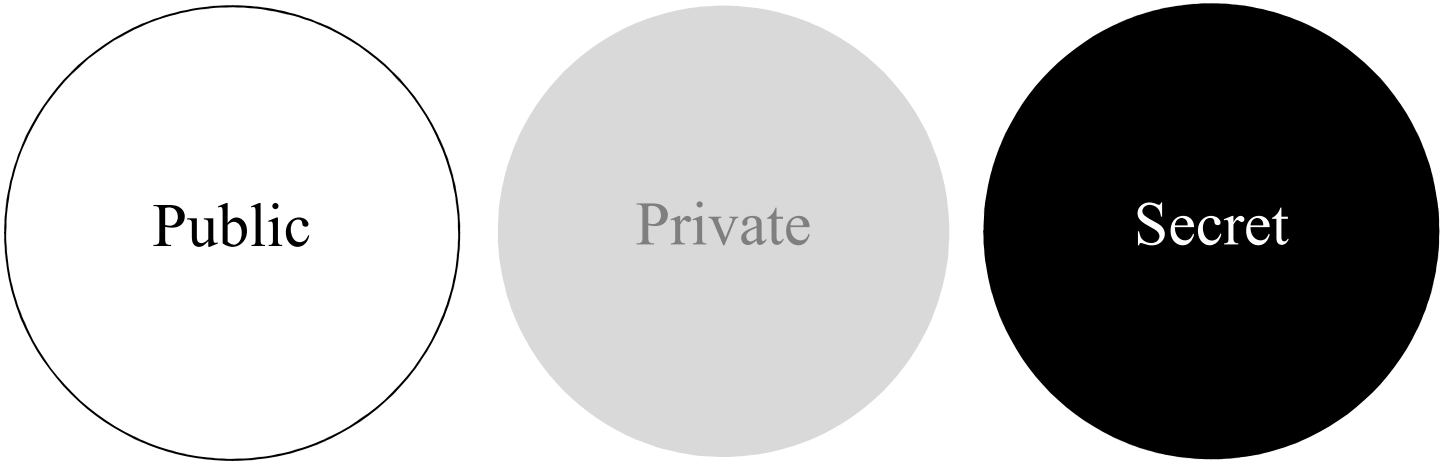
What can we do?



WHAT IS PRIVACY AT ALL?

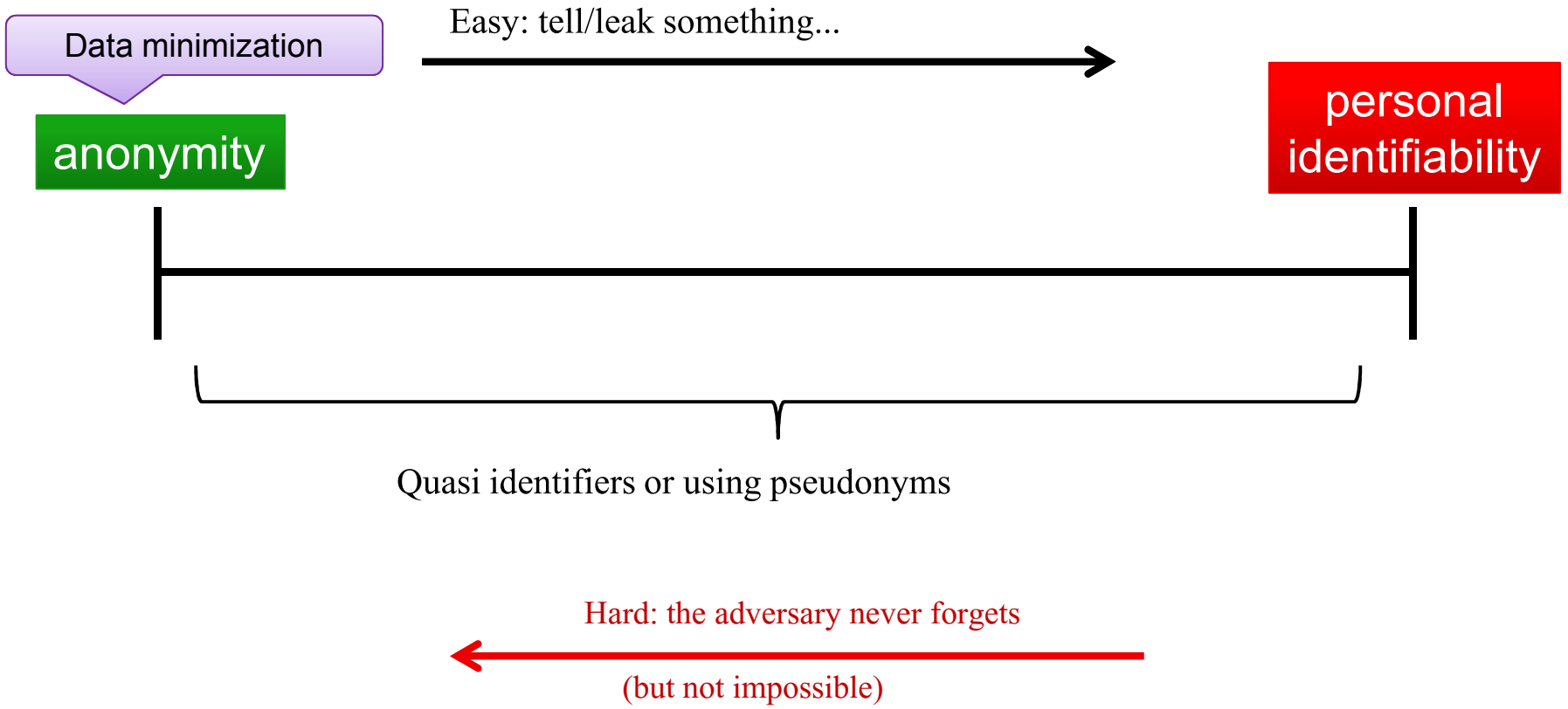
What is privacy?

- 'All human beings have three lives: public, private, and secret.' (Gabriel García Márquez)



Privacy = Freedom

Anonymity vs. Identifiability?



WEB PRIVACY OVERVIEW

Privacy on the web?!

THE WALL STREET JOURNAL. ≡ | TECH

TOP STORIES IN TECH 1 of 12

Apple Weighed Firing Ad Agency

2 of 12

Amazon Teases Home Barcode Scanner for ...

3 Mak Chec

TECHNOLOGY

On Orbitz, Mac Users Steered to Pricier Hotels

Email Print Save 258 Comments A A

By DANA MATTIOLI

Updated Aug. 23, 2012 6:07 p.m. ET

Orbitz has found that Apple users spend as much as 30% more a night on hotels, so the online travel site is starting to show them different, and sometimes costlier, options than Windows visitors see. Dana Mattioli has details on The News Hub. Photo: Bloomberg.

Orbitz Worldwide Inc. OWW -4.57% has found that people who use Apple Inc. AAPL -1.29% 's Mac computers spend as much as 30% more a night on hotels, so the online travel agency is starting to show them different, and sometimes costlier, travel options than Windows visitors see.

The Orbitz effort, which is in its early stages, demonstrates how tracking people's online activities can use even seemingly innocuous information—in this case, the fact that customers are visiting Orbitz.com from a Mac—to start predicting their tastes and spending habits.

Why the Apple Demographic Is So Important ›

Apple is practically creating its own demographic, and researchers are trying to define it. Their goal: to see if higher income levels translate into higher spending. Read more .

Orbitz executives confirmed that the company is experimenting with showing different hotel offers to Mac and PC visitors, but said the company isn't showing the same room to different users at different prices. They also pointed out that users can opt to rank results by price.

Orbitz found Mac users on average spend $20 to $30 more a night on hotels than their PC counterparts, a significant margin given the site's average nightly hotel booking is

Privacy on the web?!

Facebook friends could change your credit score

CNNMoney

By Katie Lobosco @KatieLobosco August 27, 2013: 11:24 AM ET

Some tech startups are using your online social data to determine your creditworthiness.

NEW YORK (CNNMoney)

Choose your Facebook friends wisely; they could help you get approved -- or rejected -- for a loan.

Privacy on the web?!

Is This the Grossest Advertising Strategy of All Time?

A new study claims to identify the times of the week that women are feeling the most insecure about their bodies, and recommends that brands "concentrate media during prime vulnerability moments."

REBECCA J. ROSEN | OCT 3 2013, 1:46 PM ET

More

Reuters

Most of the time, targeted ads are pretty harmless. You searched for a flight to Denver? Here are some hotels in Denver. You looked for new running sneakers? Here are a few options.

But a new "study" from marketing firm PHD recommends a strategy that crosses the line from merely targeted to outright predatory, explicitly advising brands to seize on the times of the day and week when women feel the most insecure about their bodies and overall appearance in order to sell beauty products and other goods.

Is it serious? (2012)

Global recession? Hah!
$36.6 billion just in 2012!

Continouos increase in numbers between 2006-2012.

Third party to first party „conversion”

Social „trackers” around 35%

Alexa top 500:
- 524 trackers,
- 7264 web bugs!

Not only cookies

One web – multiple companies

Some tracker can detect 20%+ of personal online acitivites!

What do they collect?

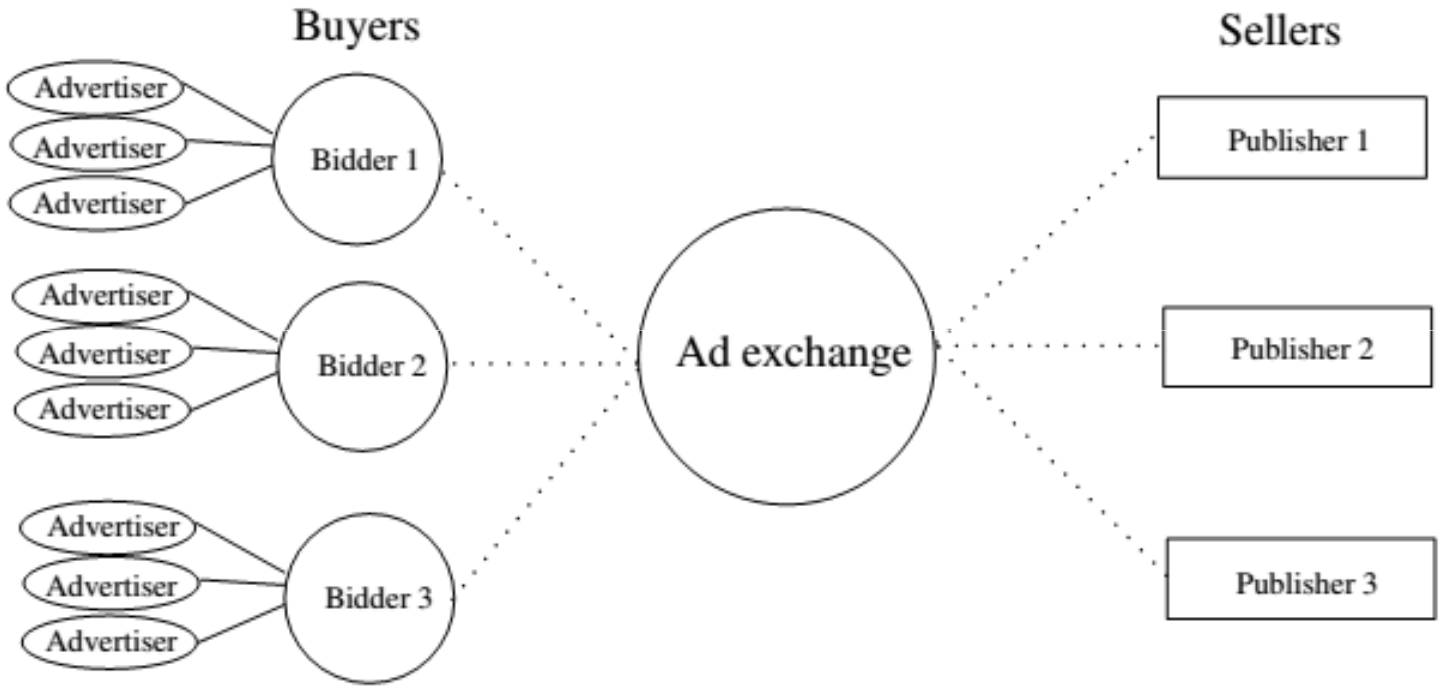
- Visit data
- Clickstream
- Typing, texts, copy-paste
- Content (content, meaning, sentiment analysis)

Superpower trends?

WHAT IS THE PRICE OF OUR PRIVACY?

Auctions, where you are the product

2011: ~10%, 2015: ~25% expected



≤100ms

Olejnik et al., 2013.

What is the price of privacy?

Topic matters

Category	Avg. price
Adult / Mature Content	0.25
Humor	0.25
Sports	0.29
Games	0.32
Blogs / Web Communications	0.33
Entertainment	0.33
Streaming Media / MP3	0.36
Computers / Internet	0.38
News / Media	0.38
Society / Lifestyle	0.38
Vehicles	0.41
Reference	0.48
Restaurants / Food	0.59
Shopping	0.68

Cost-permille impression (CPM)
Prices for 1000 impressions.
Avg. 0.5 CPM → $0.0005
(ca. 0.113 HUF)

Date and location:

Time division	The US	France	Japan
0-8h	0.75 (3246)	0.39 (10621)	0.28 (729)
8-16h	0.68 (2772)	0.36 (11375)	0.22 (732)
16-24h	0.62 (2520)	0.31 (7675)	0.19 (516)

What is the price of privacy? (2)

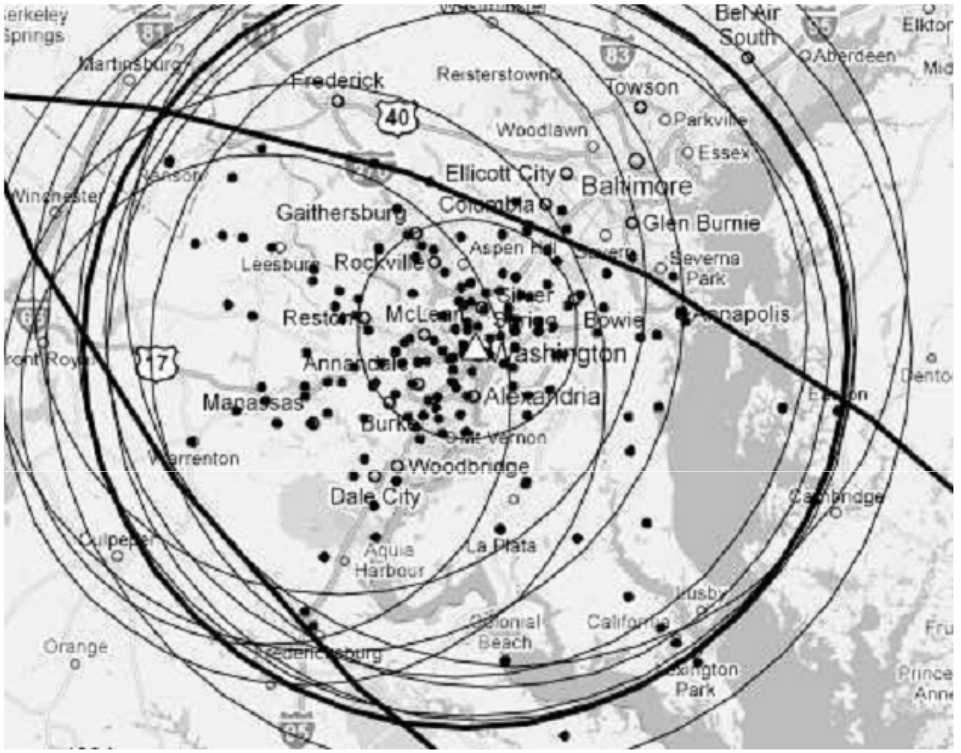
Browsing history

Category	New user		fnac.com	hotels.fr	maty.com
	avg	std	avg	avg	avg
Adult	N/A	N/A	0.56	0.64	1.12
Arts	N/A	N/A	0.52	0.66	1.28
Business	N/A	N/A	0.63	0.61	1.10
Business - Finan. Serv.	N/A	N/A	0.68	0.88	1.31
Computers	N/A	N/A	0.57	0.70	1.18
Games	N/A	N/A	0.74	0.81	1.41
Health	N/A	N/A	0.68	0.81	1.21
Home	N/A	N/A	0.70	0.57	1.00
Kids and Teens	N/A	N/A	0.65	0.74	1.25
News	N/A	N/A	0.72	0.74	1.09
Recreation	N/A	N/A	0.64	0.69	1.12
Science	N/A	N/A	0.60	0.59	1.36
Shopping	N/A	N/A	0.61	0.65	1.21
Sports	N/A	N/A	0.59	0.62	1.17
Average	0.41	0.10	0.64	0.69	1.20

STORAGE-BASED TRACKING

Basic techniques

- IP based tracking
  - Problems: NAT, din. IP
  - Used as auxiliary info
  - Geo-localization
  - Government surveillance
- Geo-localization alternative
  - <1km precise (2011)

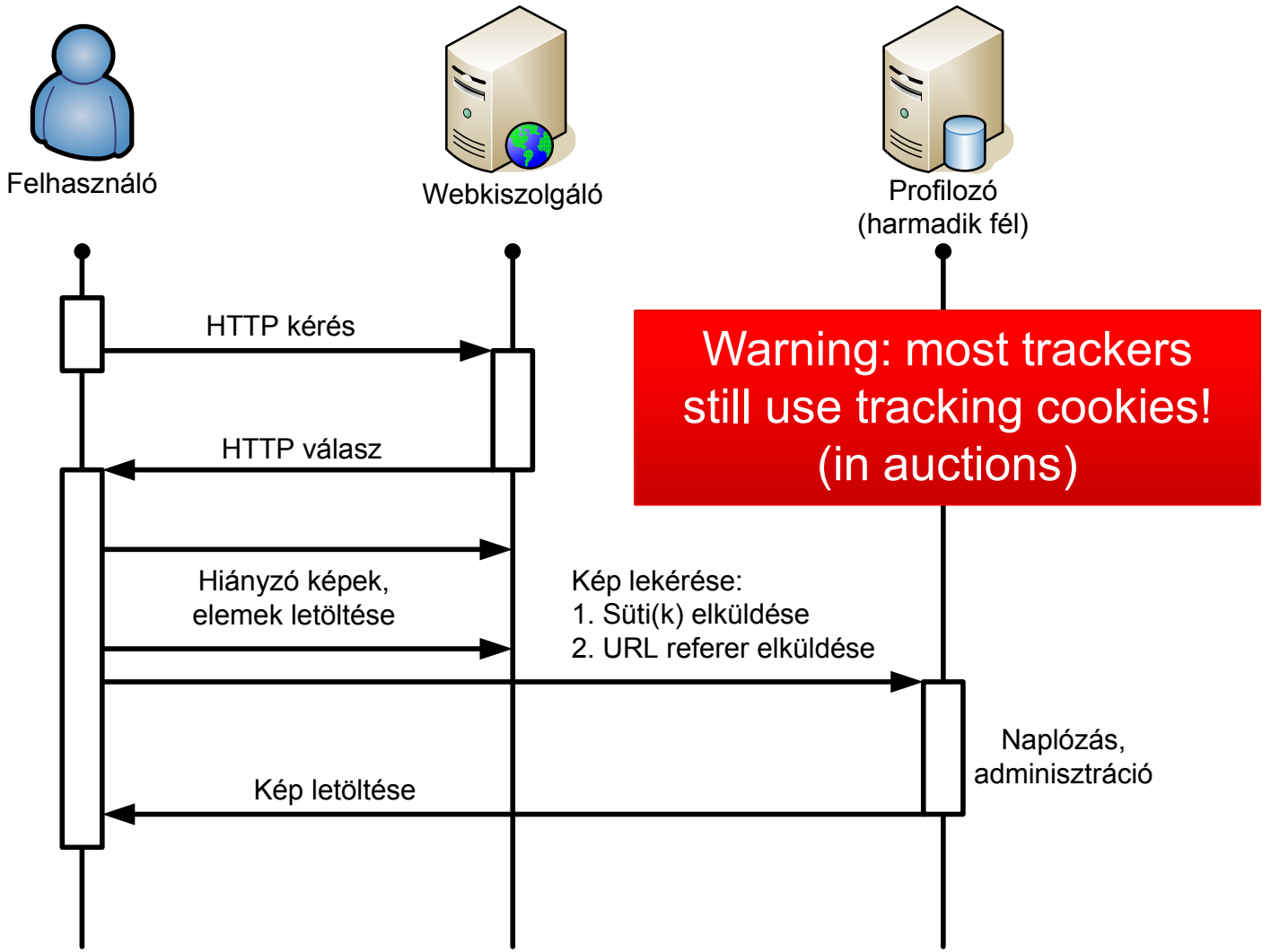


http://static.usenix.org/event/nsdi11/tech/full_papers/Wang_Yong.pdf

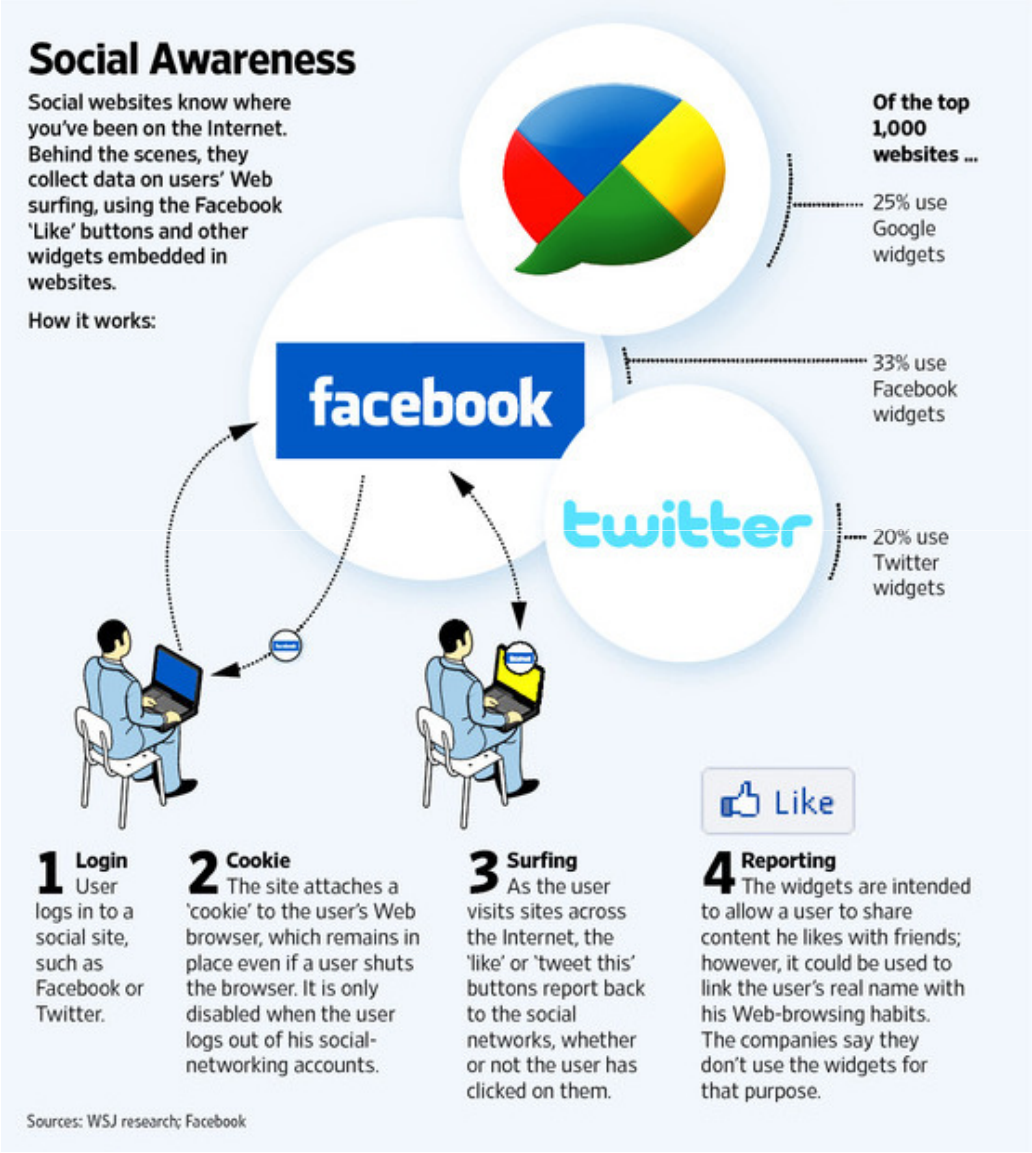
Web bugs

- Web bugs
  - Hidden or camouflaged: 1x1 pixels, transparent GIFs, visible ads, or other resources
  - (concerned techniques discussed later)
  - A.k.a.: web beacon, clear GIF, 1x1 GIF, tracking pixel, pixel tag, pixel
- Other types
  - Email web bugs: have you read it?
  - RSS web bugs:
    - ~25% infection of 40 major RSS feeds (2011)
    - Problem: automatic refresh in browsers, lack of awareness
- URL referrers!

Browser cookies (tracking cookies, third-party cookies)



Social buttons serving your good – or not?

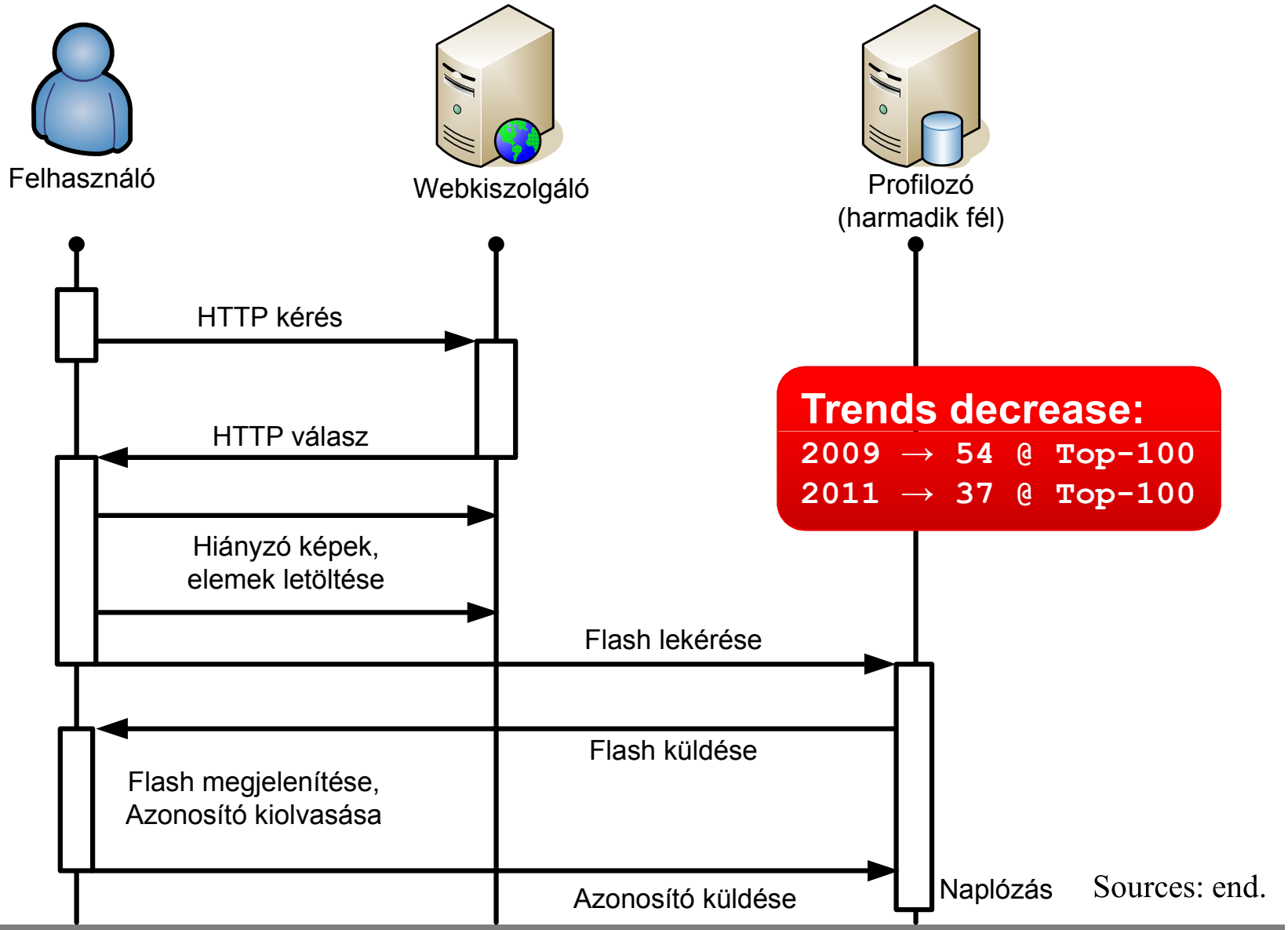


http://online.wsj.com

Flash and SilverLight

- Still high market share (95.63%, 67.98% for PCs in 2011)
- Cookies
  - Flash: LSO = Local Shared Object
  - SilverLight: Isolated Storage
- Flash properties
  - Max 100 KB, never expires
  - Cross-browser storage
- Tracking
  - PIE: Persistent Identification Element (2005)
  - Cookie respawn → backup cookies in PIEs
  - Flash info leak (OS, HW, settings)

Flash tracking



Sources: end.

HTML5

- HTML5 cookie-like storage
  - Client side only, need to be explicitly sent
  - Bigger storage, 5 MB
  - Still never expires
- Real threat?
  - Only in a fraction of all cases (1%-17%)
  - Sometimes used instead of tracking cookies (taboolasyndication.com & krxd.net)
  - Backup for cookies (cookie respawn)

(Sources: Roesner et al., 2012, Ayenson et al., 2011)

Cache exploitation

- Content cache
  - JavaScript
  - CSS
  - Modified pixels in images (stego like)
- Cache controls
  - E-tag
  - Last-mod timestamp
- Operative caches
  - HTTP authentication cache
  - HTTP 301 redirect cache
  - HTTP Strict Transport Security cache

Evercookie: persistent cookies

- Concept:
  - Redundant by using multiple storage possibilities in parallel
- Cons:
  - Initially undeletable, but some browsers now are safe
  - Some storages are „costly” to read
- Evercookie vs. fingerprints
  - Fingerprinting does not work when there is an array of similar computers

http://samy.pl/evercookie/

TRACKING VIA IDENTIFICATION (2)

CSS History Stealing

- Basic concept
  - Browsing history can not be explicitly accessed
  - Inserting links + checking their color → URL list
  - Can be used for various purposes
- Patched since 2010
  - Reason: social sites provided background knowledge for precise personal identification
- There are various alternatives,
  - and new ones pop-up from time-to-time
  - but these are not generic, nor wide-spread

Fingerprinting?

Sweeney, 1990

87% of US population is identifiable by (216 million of 248 million):

`{5 digit ZIP, gender, date of birth}`

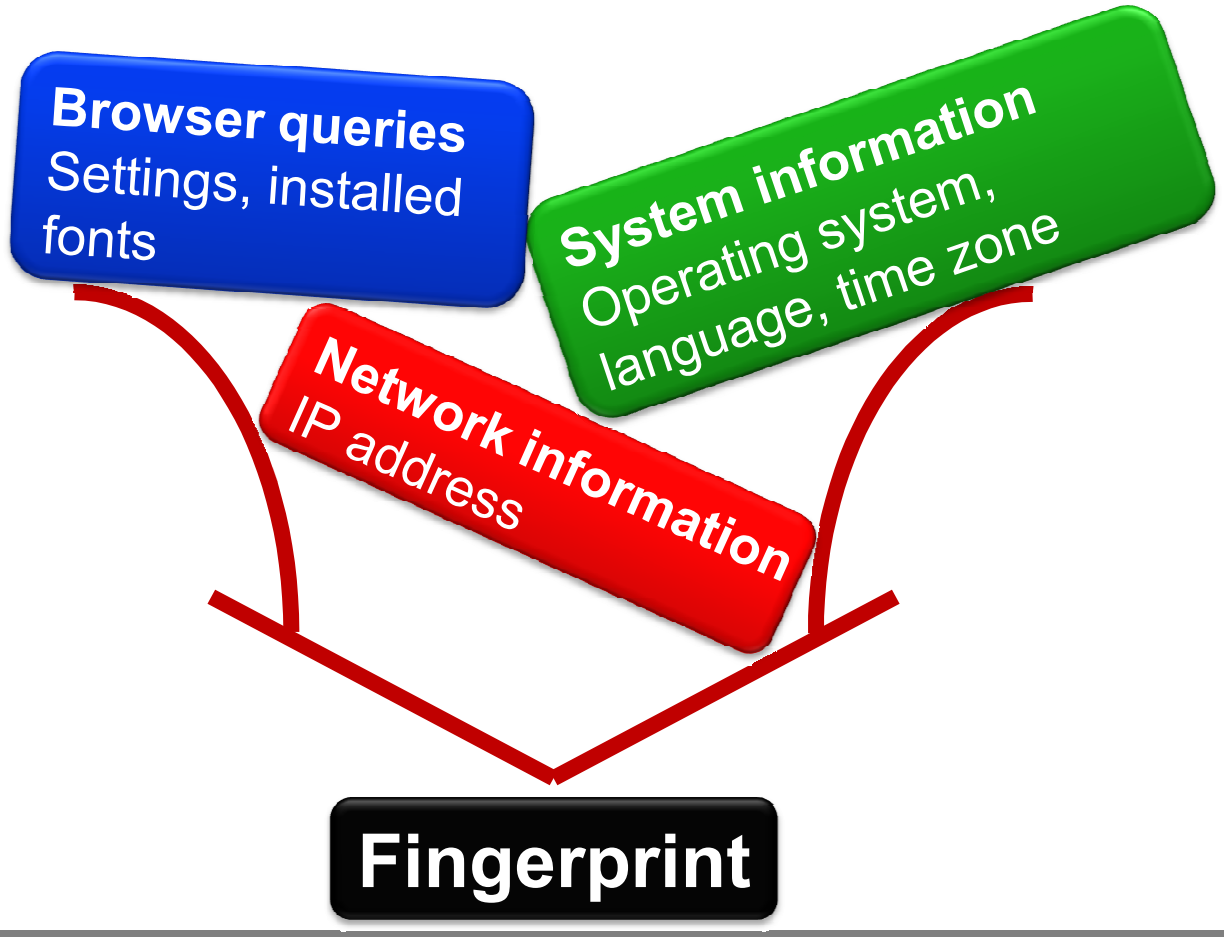
Panopticlick project: browser fingerprinting? (2010)

- Uniqueness based on:
  - User agent string
  - Accept header
  - Plugin list
  - Available fonts
  - Etc.
- Pros:
  - Fast and precise
  - Passive (no data storage)
  - Long term fingerprints

http://panopticlick.eff.org

- Cons:
  - Flash/Java required (for 95% uniqueness)
  - Browser dependent

Cross-browser fingerprinting possible?



The system fingerprint project (2011)

http://pet-portal.eu/fingerprint

Az Ön számítógépének ujjlenyomata:

Ujjlenyomat:	8bf2bcc936e098497691f3782c3c6fa62705f51e
A böngésző adatait tartalmazó karakterlánc:	Mozilla/5.0 (Windows; U; Windows NT 6.1; hu; rv:1.9.2.9) Gecko/20100824 Firefox/3.6.9 (.NET CLR 3.5.30729)
Operációs rendszer:	Windows
Képernyő felbontás:	1680x1050
Időzóna:	-120
Telepített betűkészletek:	

Aharoni, **ALGERIAN**, Andalus, Angsana New, AngsanaUPC, Aparajita, Arabic Transparent, Arabic Typesetting, Arial, Arial Baltic, **Arial Black**, Arial CE, Arial CYR, Arial Greek, Arial Narrow, Arial TUR, Arial Unicode MS, Baskerville Old Face, Batang, BatangChe, **Bauhaus 93**, Bell MT, Berlin Sans FB, **Berlin Sans FB Demi**, **Bernard MT Condensed**, Bodoni MT Poster Compressed, Book Antiqua, Bookman Old Style, Bookshelf Symbol 7, **Britannic bold**, **Broadway**, Browallia New, BrowalliaUPC, Brush Script MT, Calibri, Californian FB, Cambria, Cambria Math, Candara, Centaur, Century, Century Gothic, Chiller, Consolas, Constantia, **Cooper Black**, Corbel, Cordia New, CordiaUPC, Courier New CE, Courier New CYR, Courier New Greek, DejaVu Sans, DejaVu Sans Condensed, DejaVu Sans Light, DejaVu Sans Mono, DejaVu Serif, DejaVu Serif Condensed, DFKai-SB, DilleniaUPC, DokChampa, Dotum, DotumChe, Ebrima, Ezra SIL, Ezra SIL SR, FangSong, fantasy, footlight mt light, Franklin Gothic Book, Franklin Gothic Medium, Franklin Gothic Medium Cond, FrankRuehl, Gabriola, Garamond, Gautami, Gentium Basic, Gentium Book Basic, Georgia, Gisha,

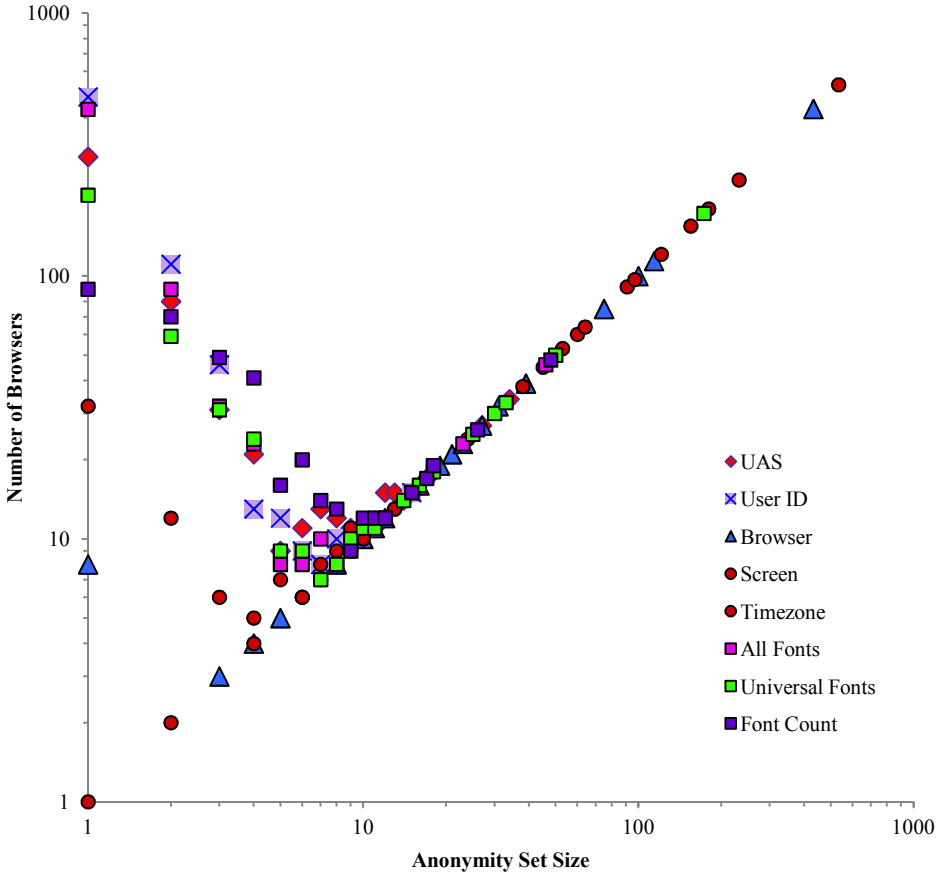
Goals:
- Cross-browser
- Plugin independent
- (IP-range indep.)

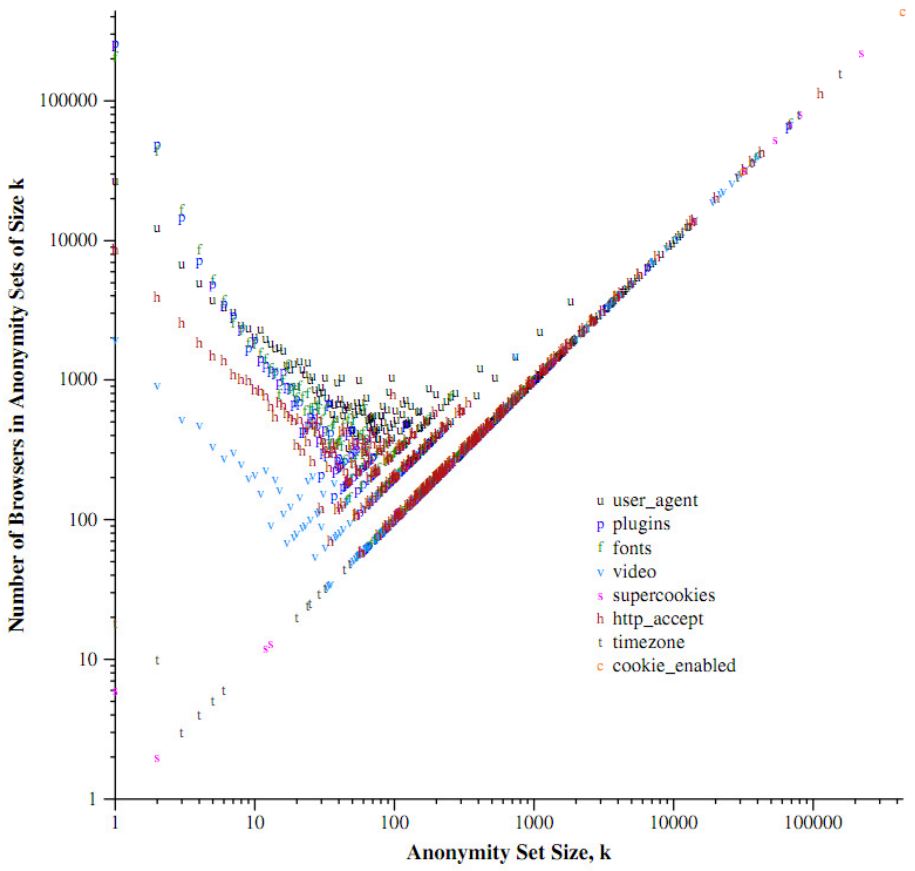
Dataset insight:
- 2010/09 – 2011/03
- ~1k fingerprints
- Locality: 649+340

Validating our dataset

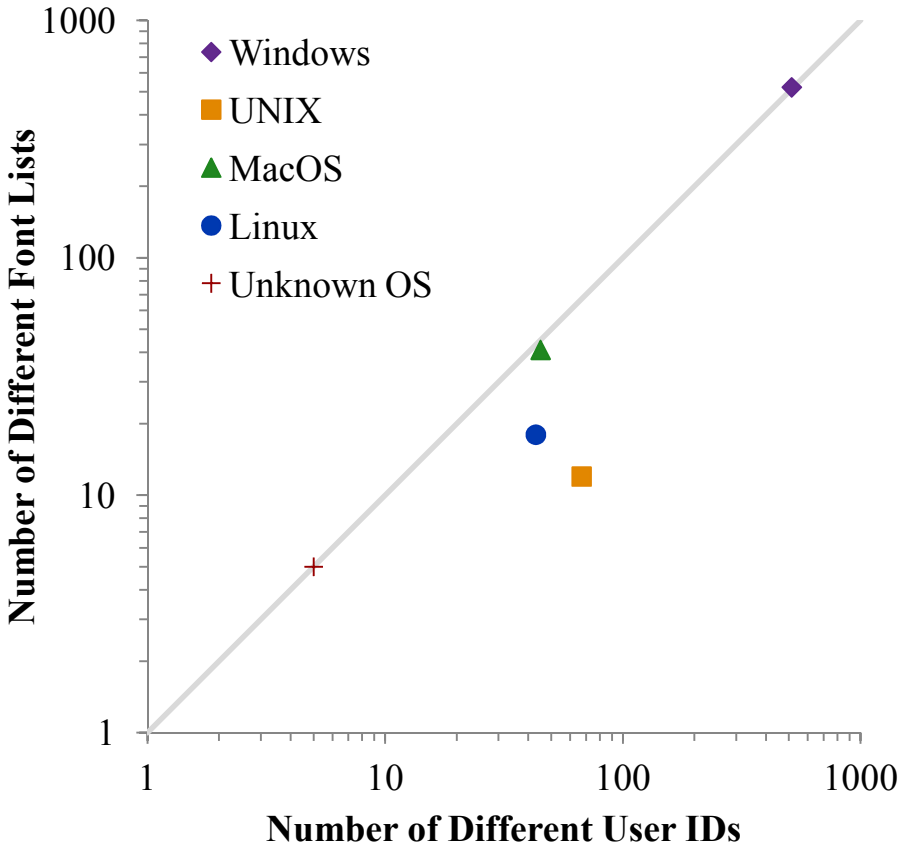
- Our dataset:
- Panopticlick:

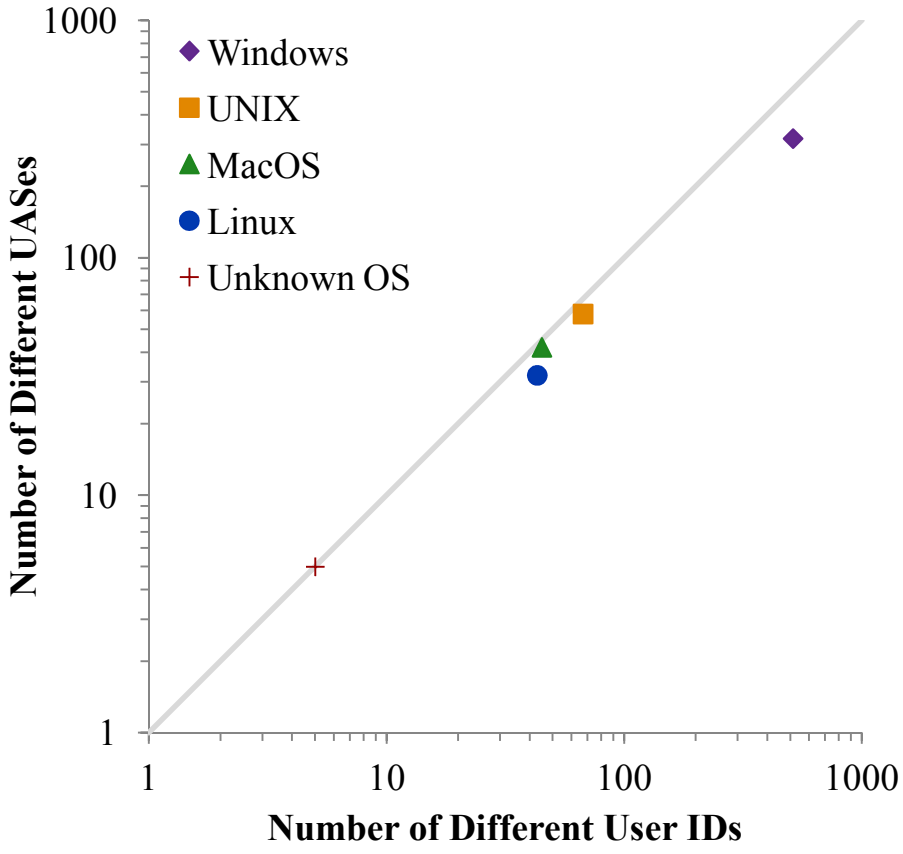
Entropy values are also similar.





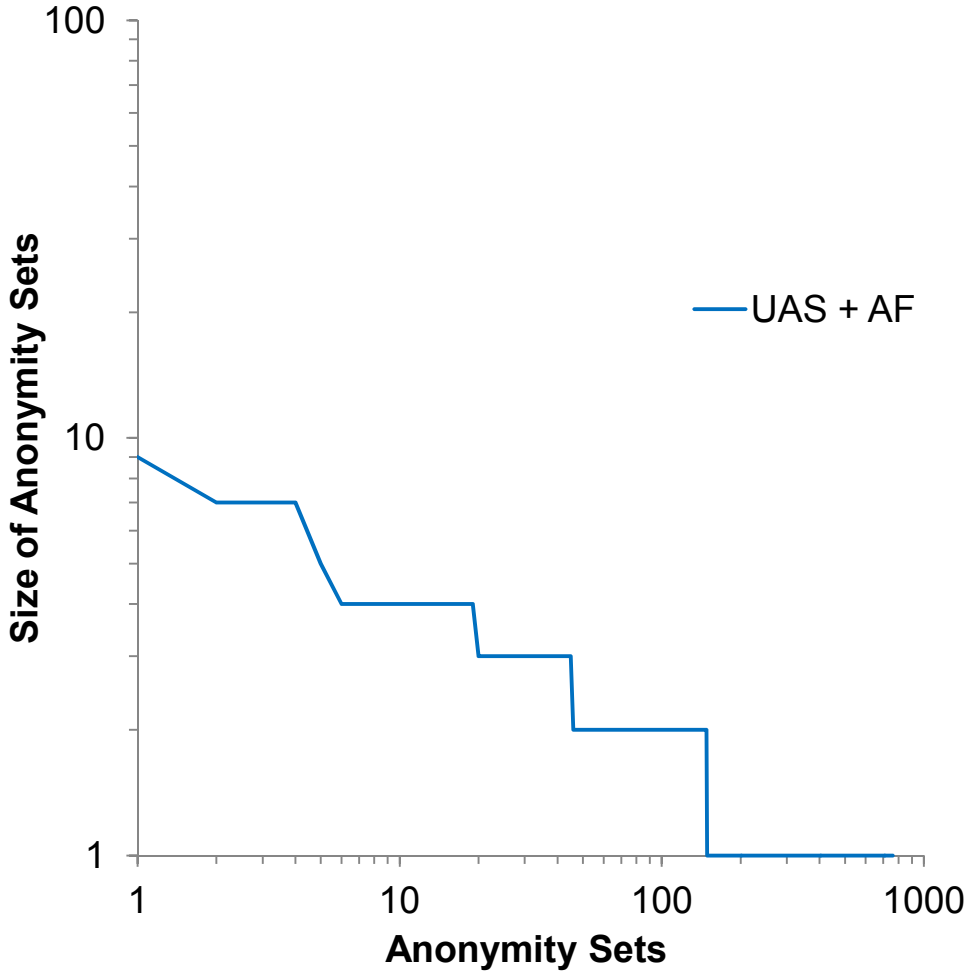
Key ingredients: font lists & UASes





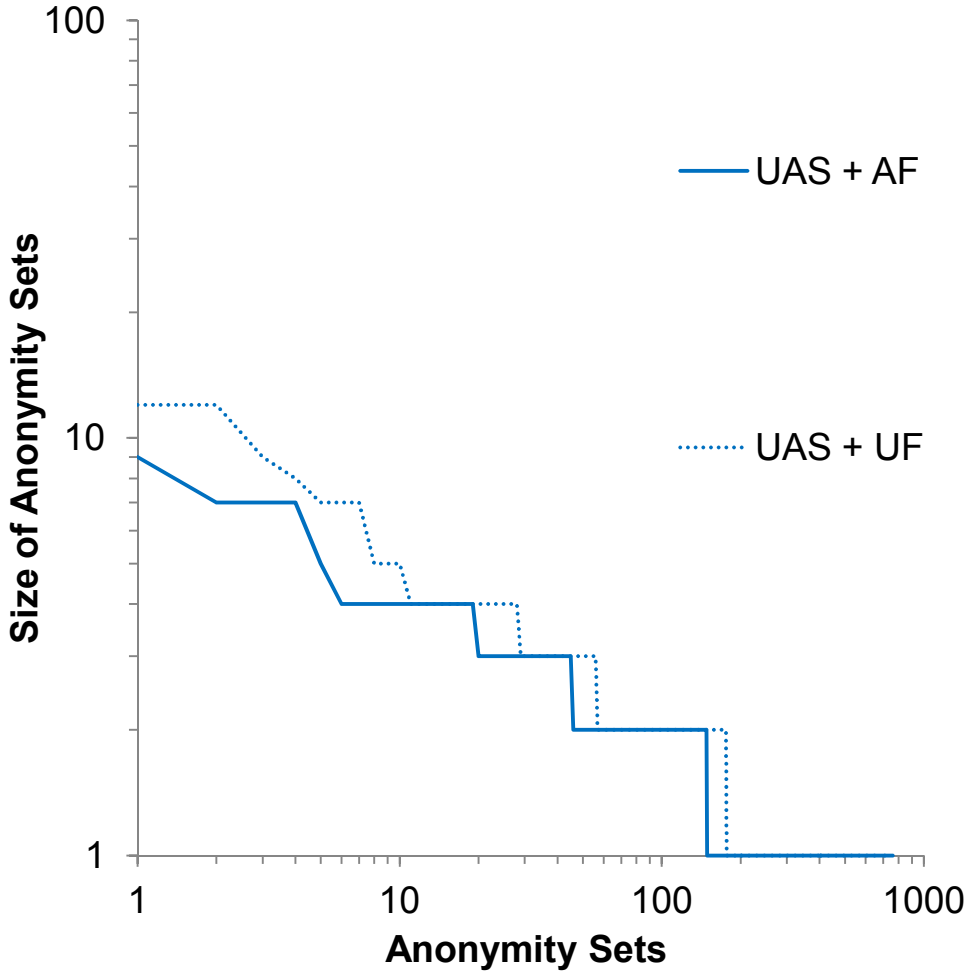
Checking the anonymity sets

- AF: all fonts available

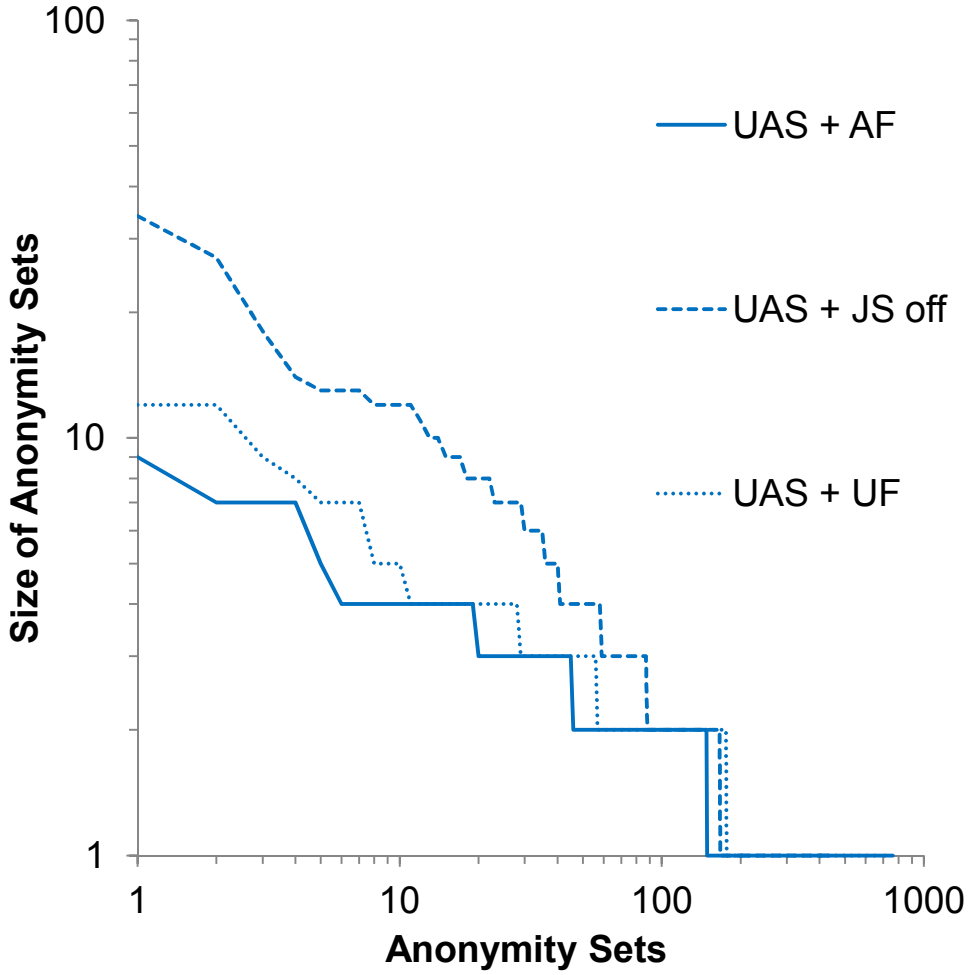


Checking the anonymity sets (2)

- AF: all fonts available
- UF: unified fontset

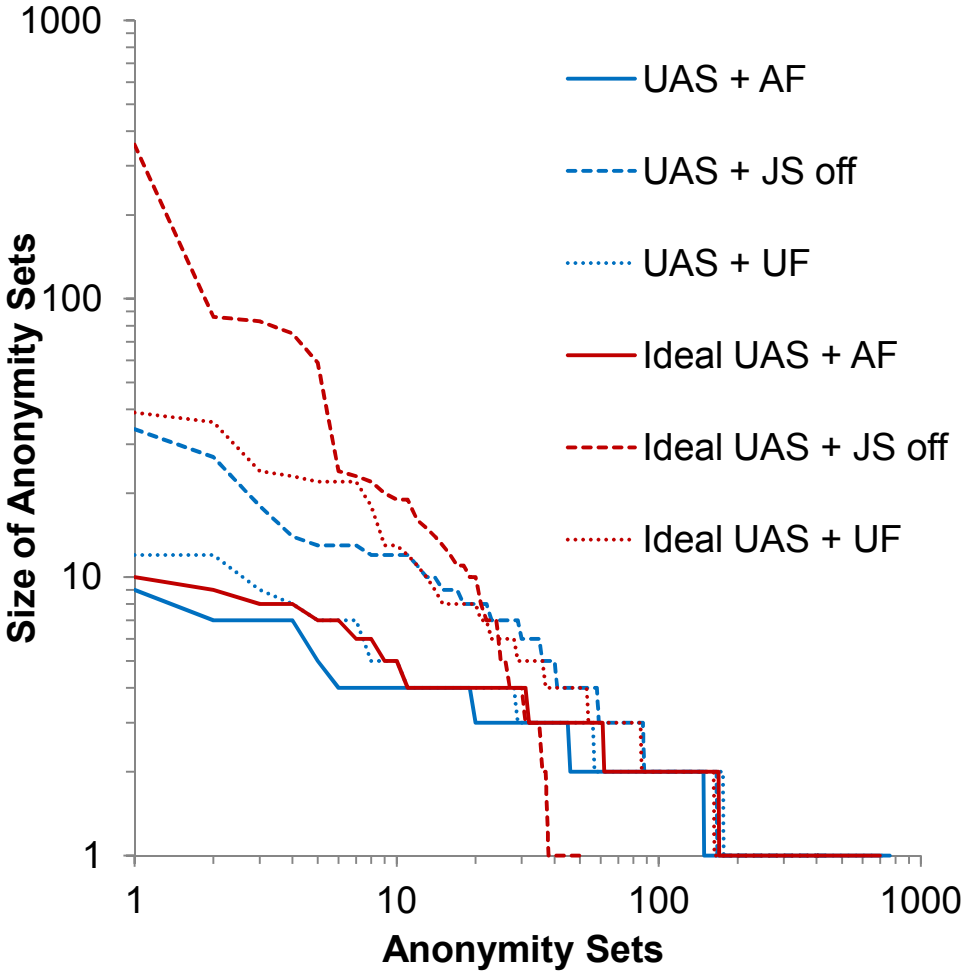


Checking the anonymity sets (3)



- AF: all fonts available
- UF: unified fontset
- JS off: no fonts can be detected

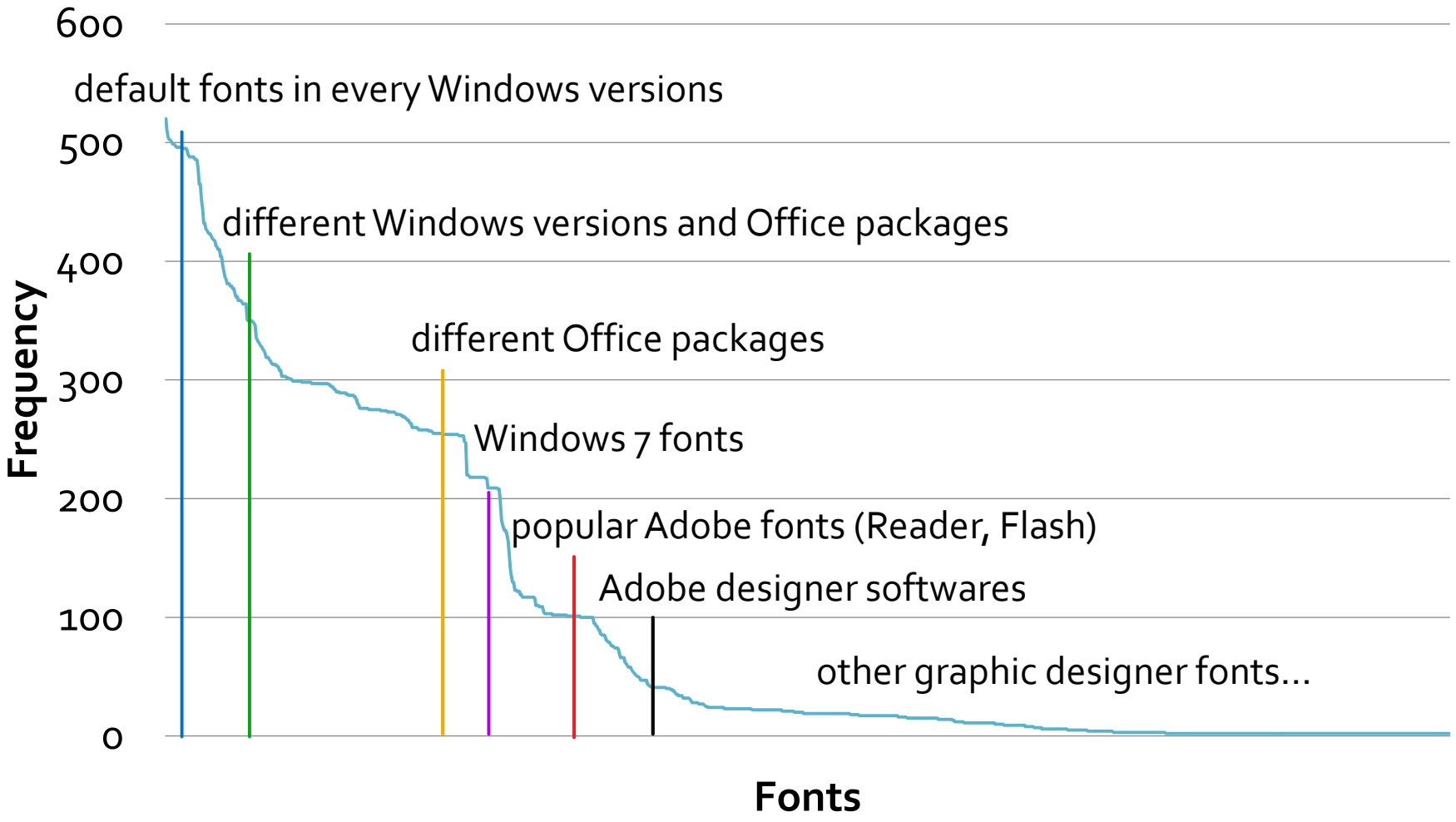
Checking the anonymity sets (4)



- AF: all fonts available
- UF: unified fontset
- JS off: no fonts can be detected

- Plus: simplifying the UAS

Unique software set → unique font list



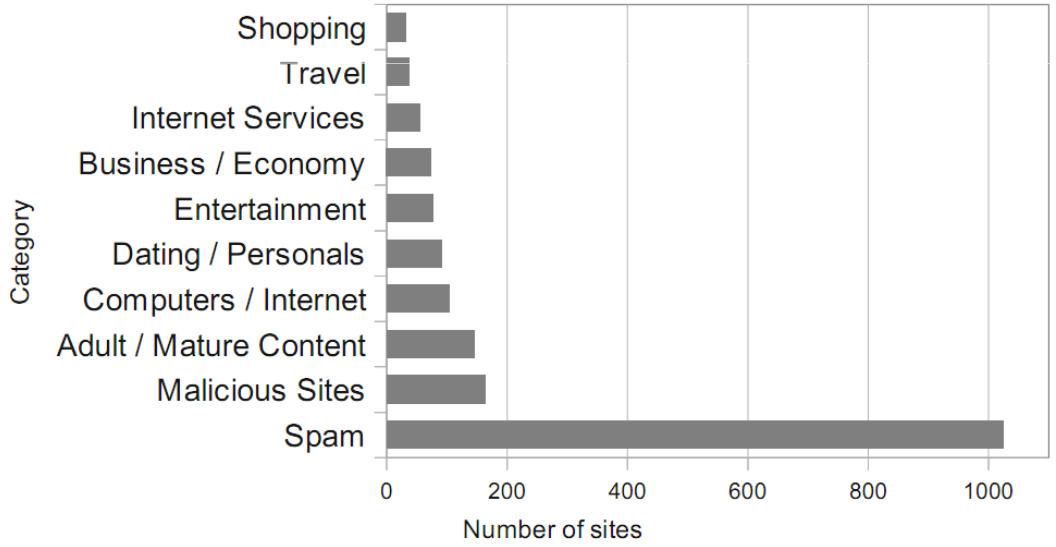
Cross-browser fingerprinting test 2.0 (2012)

Is fingerprinting widely adopted?

(Src: Nikiforakis et al., 2013)

- Alexa top 10 000, 20 pages deep
  - Only 0,4% adoption (40 sites)
  - Skype.com, porn and dating sites
- 3 804 less popular sites have it

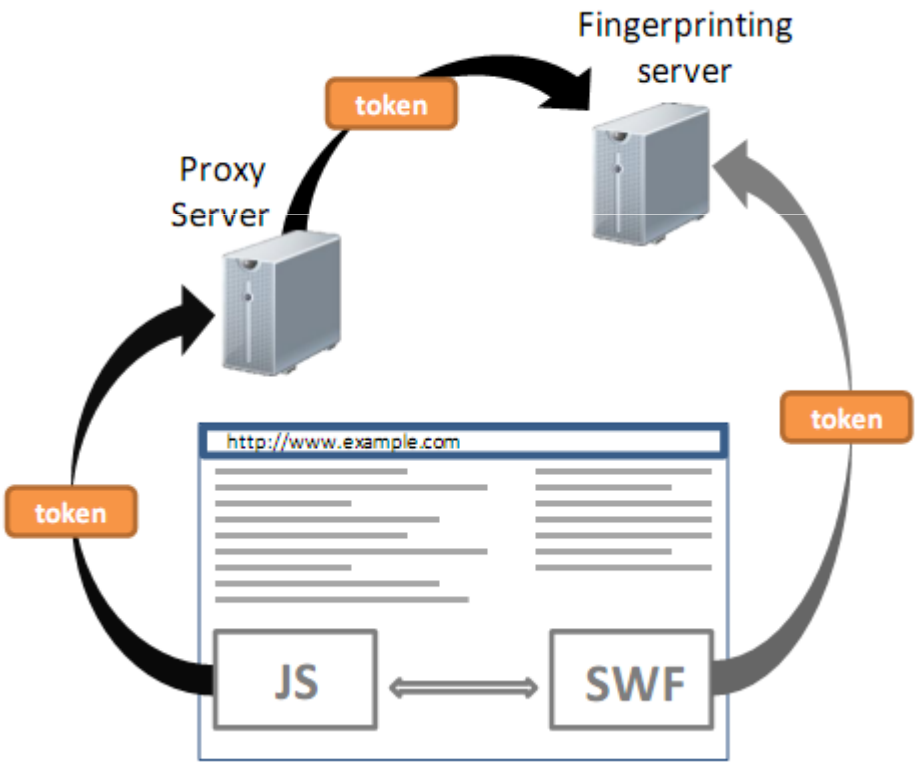
ThreatMetrix



How do they work?

(Src: Nikiforakis et al., 2013)

- No Java, but use Flash



- Firefox: Linux x86 64
- Flash: Linux 3.2.0-26-generic

- Firefox: 1280x720
- Flash: 2560x720

- Firefox: JS font list
- Flash: unique list

How do they work? (2)

(Src: Nikiforakis et al., 2013)

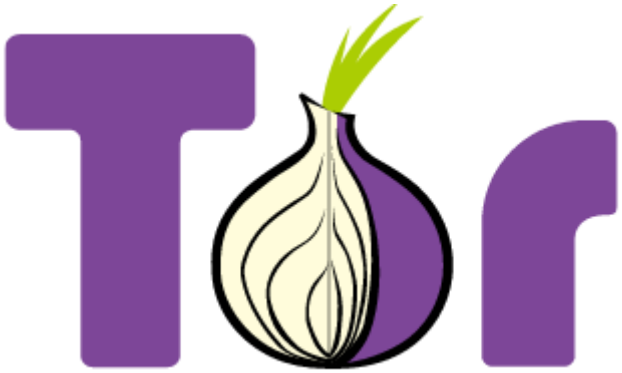
- Browser dependent fingerprinting techniques
- Native fingerprinting libraries with Internet Explorer (deployment?)
  - the hard disk's identifier, TCP/IP parameters, the computer's name, Internet Explorer's product identifier, the installation date of Windows, the Windows Digital Product Id and the installed system drivers
- Fingerprint DES encoded, public key wrapped → the tracker gets inescapable!

Fingerprinting Category	Panopticlick	BlueCava	Iovation ReputationManager	ThreatMetrix
Browser customizations	Plugin enumeration$_{(JS)}$ Mime-type enumeration$_{(JS)}$ ActiveX + 8 CLSIDs$_{(JS)}$	Plugin enumeration$_{(JS)}$ ActiveX + 53 CLSIDs$_{(JS)}$ Google Gears Detection$_{(JS)}$		Plugin enumeration$_{(JS)}$ Mime-type enumeration$_{(JS)}$ ActiveX + 6 CLSIDs$_{(JS)}$ Flash Manufacturer$_{(FLASH)}$
Browser-level user configurations	Cookies enabled$_{(HTTP)}$ Timezone$_{(JS)}$ Flash enabled$_{(JS)}$	System/Browser/User Language$_{(JS)}$ Timezone$_{(JS)}$ Flash enabled$_{(JS)}$ Do-Not-Track User Choice$_{(JS)}$ MSIE Security Policy$_{(JS)}$	Browser Language$_{(HTTP, JS)}$ Timezone$_{(JS)}$ Flash enabled$_{(JS)}$ Date & time$_{(JS)}$ Proxy Detection$_{(FLASH)}$	Browser Language$_{(FLASH)}$ Timezone$_{(JS, FLASH)}$ Flash enabled$_{(JS)}$ Proxy Detection$_{(FLASH)}$
Browser family & version	User-agent$_{(HTTP)}$ ACCEPT-Header$_{(HTTP)}$ Partial S.Cookie test$_{(JS)}$	User-agent$_{(JS)}$ Math constants$_{(JS)}$ AJAX Implementation$_{(JS)}$	User-agent$_{(HTTP, JS)}$	User-agent$_{(JS)}$
Operating System & Applications	User-agent$_{(HTTP)}$ Font Detection$_{(FLASH, JAVA)}$	User-agent$_{(JS)}$ Font Detection$_{(JS, FLASH)}$ Windows Registry$_{(SFP)}$	User-agent$_{(HTTP, JS)}$ Windows Registry$_{(SFP)}$ MSIE Product key$_{(SFP)}$	User-agent$_{(JS)}$ Font Detection$_{(FLASH)}$ OS+Kernel version$_{(FLASH)}$
Hardware & Network	Screen Resolution$_{(JS)}$	Screen Resolution$_{(JS)}$ Driver Enumeration$_{(SFP)}$ IP Address$_{(HTTP)}$ TCP/IP Parameters$_{(SFP)}$	Screen Resolution$_{(JS)}$ Device Identifiers$_{(SFP)}$ TCP/IP Parameters$_{(SFP)}$	Screen Resolution$_{(JS, FLASH)}$

MISBELIEFS ON PROTECTION

Using TOR alone provides privacy

- Hides some meta-data, e.g., IP, timings.
- Often praised as anti-gov hiding tool.
- But:
  - Identification & tracking on higher levels (e.g., fingerprinting)
  - Personal identification with strong background knowledge (e.g., gov.)



Anonabox

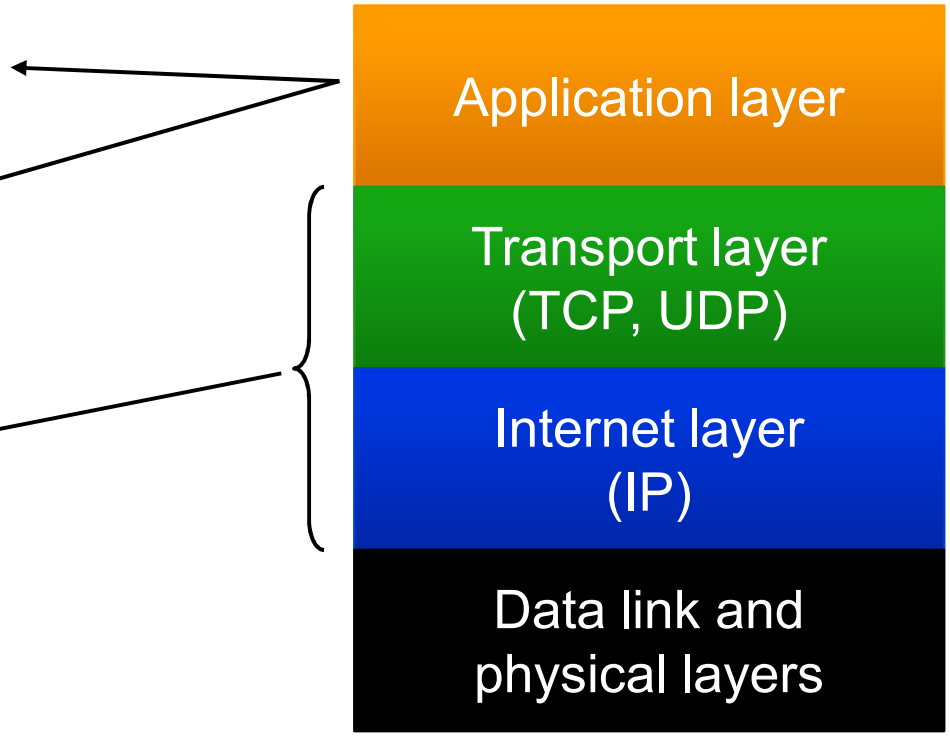
Using TOR alone provides privacy (2)

- Not totally pointless
  - Makes surveillance harder for smaller gov's
  - Circumvents malicious ISP activites

HOW TO PROTECT OURSELVES

Protection mechanisms

- **Partial solutions**
  - Use the correct settings of the browser
  - Filtering content (ads, web bugs, etc.)
  - Spyware/malware protection
  - TOR/JondoNym
- **Complex solutions**
  - Anonymous browsers
  - Tails



TCP/IP stack

Privacy/security vs. usability

Regular use

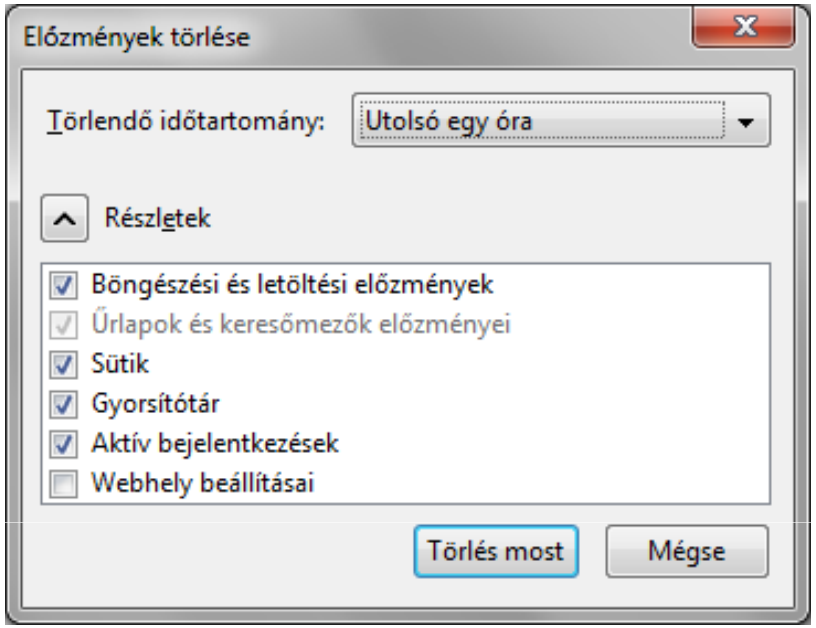
- Majority of problems can be covered with simple solutions:
  - Setting correctly
  - Adding simple filters: Ghostery/AdBlock/NoScript
  - Or anonymous browsers without anonymized network

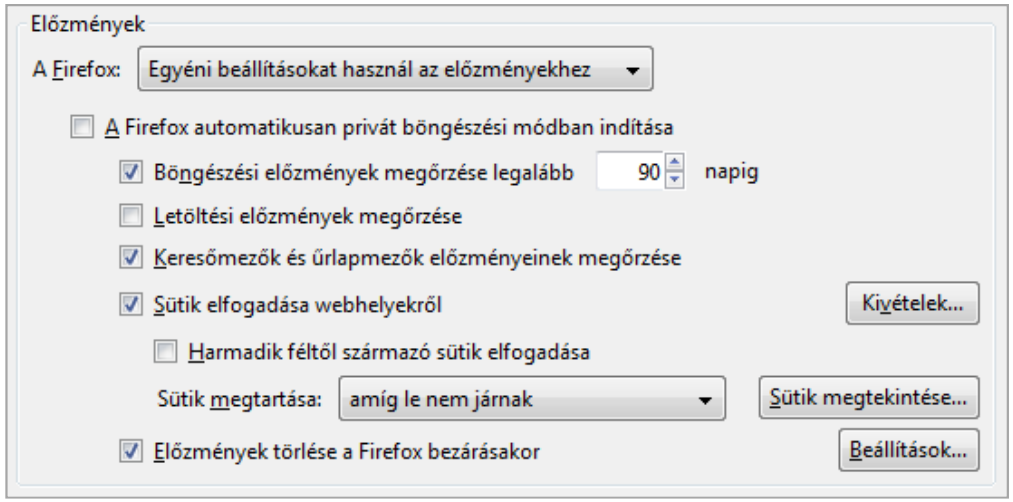
Hard-core use

- Anonymous browsers
  - JondoFox + JondoNym
  - Tor Browser Bundle
- Tails

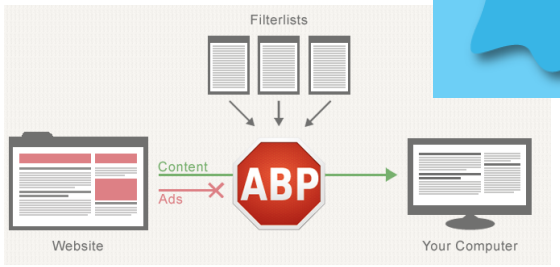
Setting the browser correctly + filtering

- Wipe private data when closed
- No third-party cookies
- Leave DNT unchanged
  - Adds tracking info
  - Who really cares?
- SSL 3.0



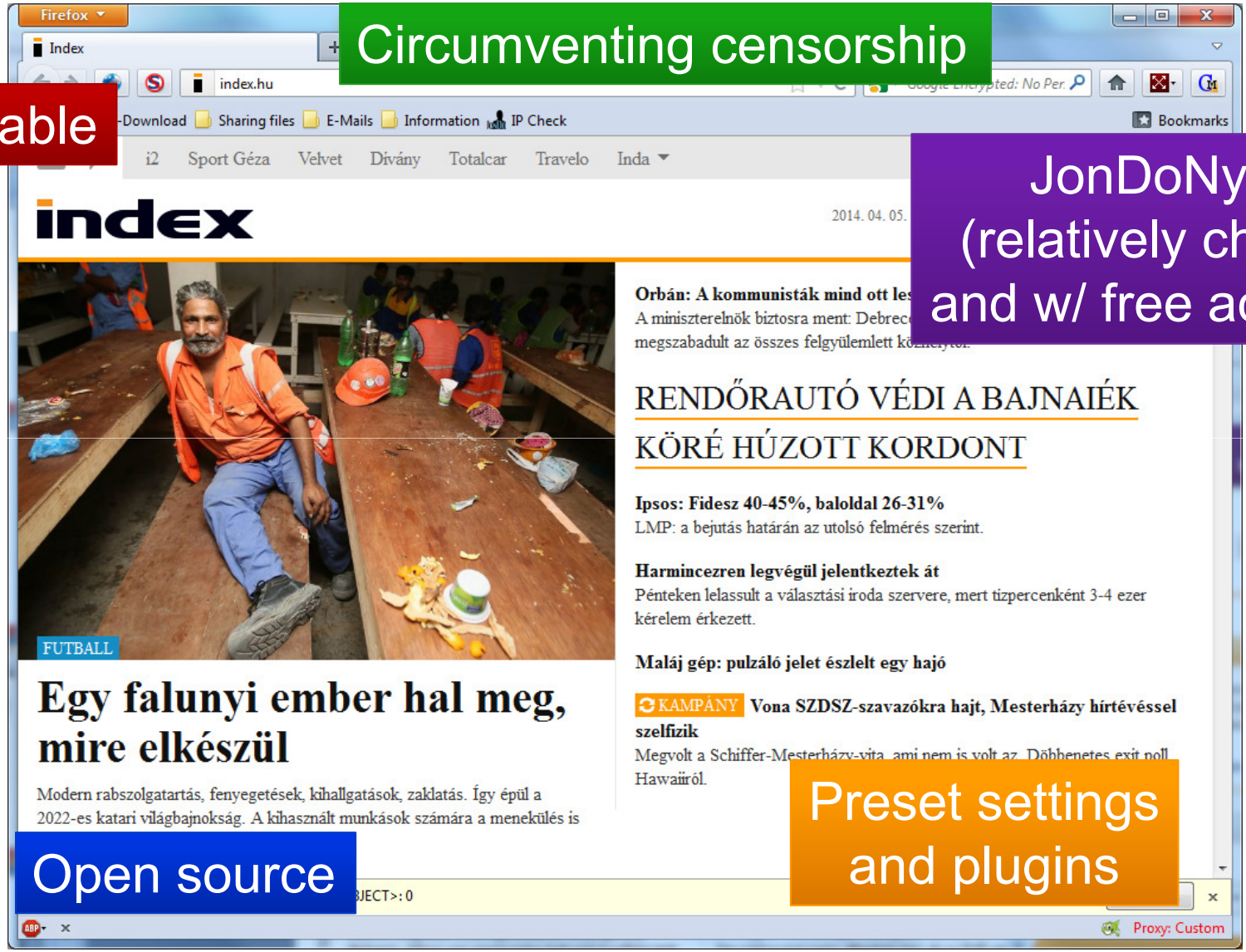


Ghostery



Adblock

JondoFox: german precision & respect for privacy



Portable

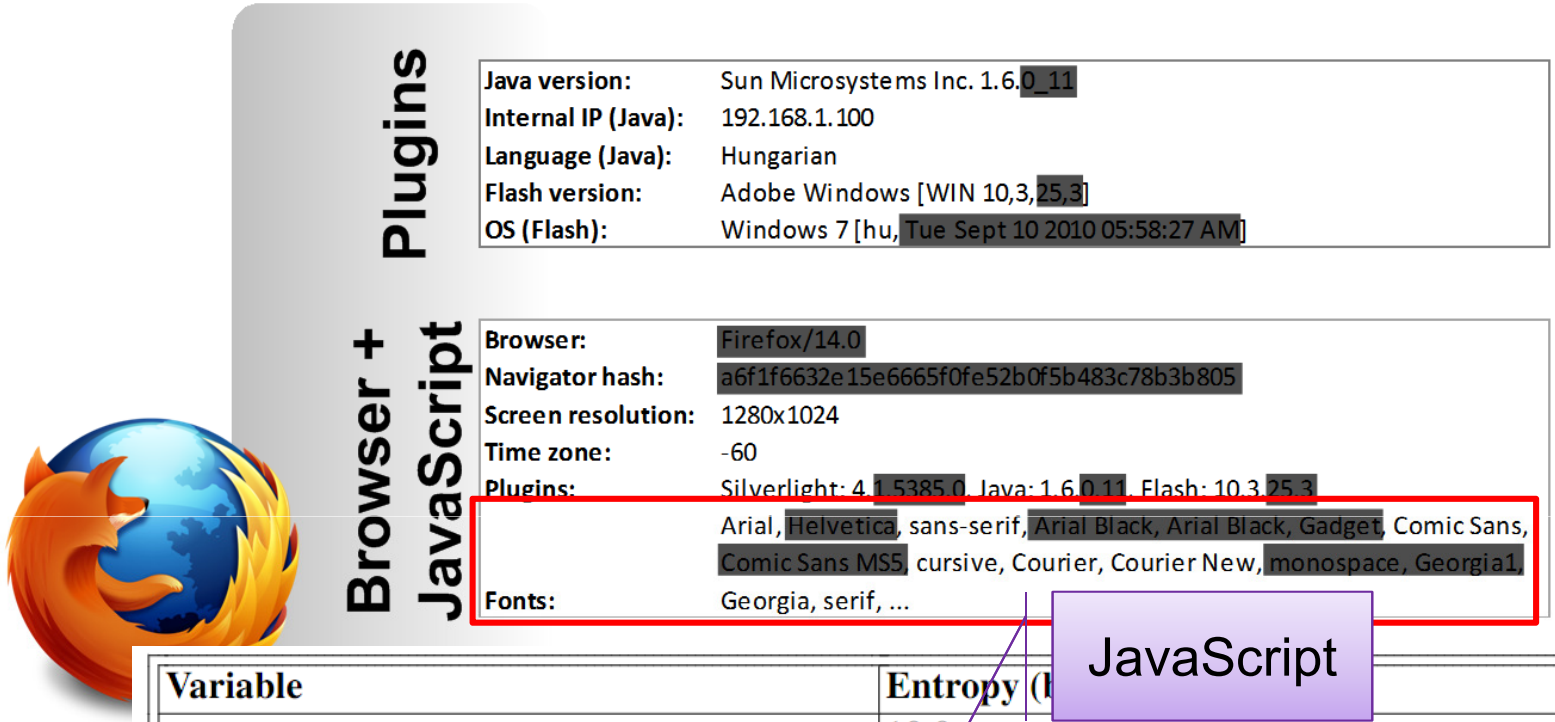
Circumventing censorship

JonDoNym (relatively cheap, and w/ free access)

Open source

Preset settings and plugins

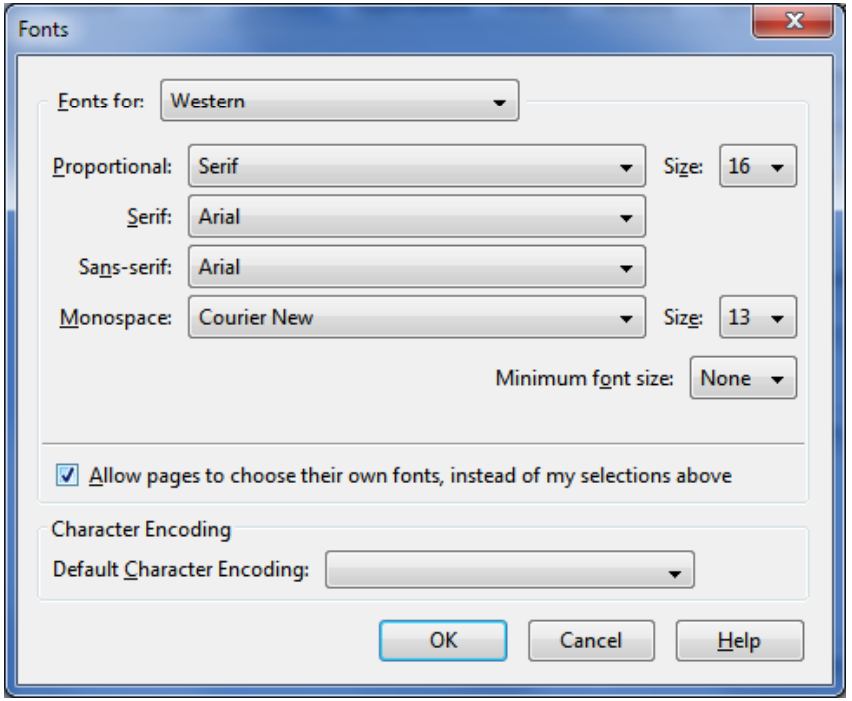
What about fingerprinting?



Variable	Entropy (b
user_agent	10.0
plugins	15.4
fonts	13.9
video	4.83
supercookies	2.12
http_accept	6.09
timezone	3.04
cookies_enabled	0.353

What about fingerprinting? (2)

- None of storage based techniques should work
  - e.g., JondoFox
- Home-brew solutions ⇒ customized Firefox portable
  - User experience?
  - Requires waste amount of time
  - You should be a professional



Beware: every setting, plugin, etc. could eventually lead get you an easy prey for fingerprinting!

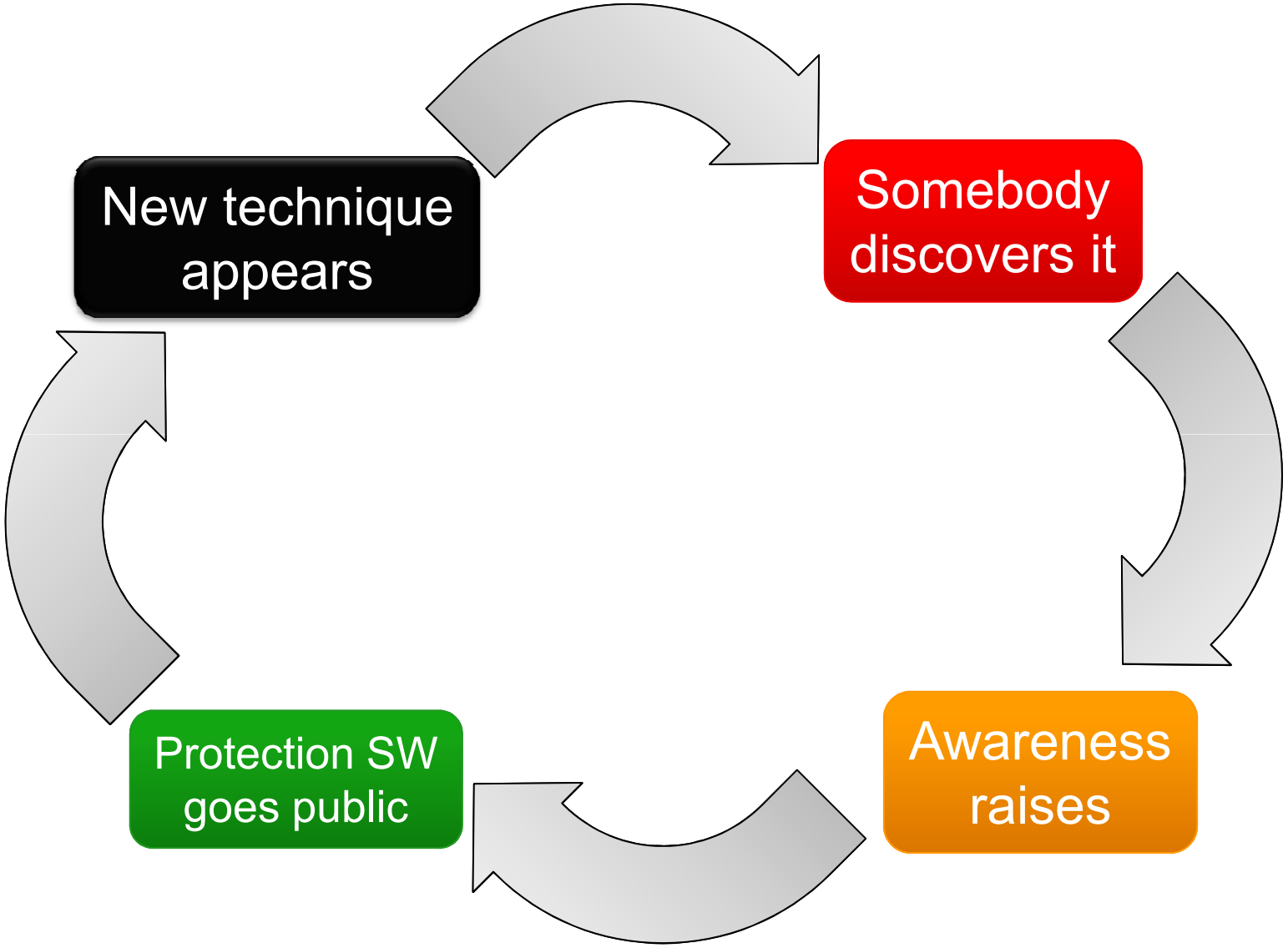
Anonymity paradox!

What about fingerprinting? (3)

- Tor/JondoFox was designed to limit font loads (these are custom builds). Check *about:config*:
  - *browser.display.max_font_count* = 5
  - *browser.display.max_font_attempts* = 10
- Not available in mainstream Firefox versions (yet?)
- Still a bit problematic: font listing can be done within a single session.
- Our proposed, but unheard solution:
  - Cache loaded fonts,
  - For every single domain!

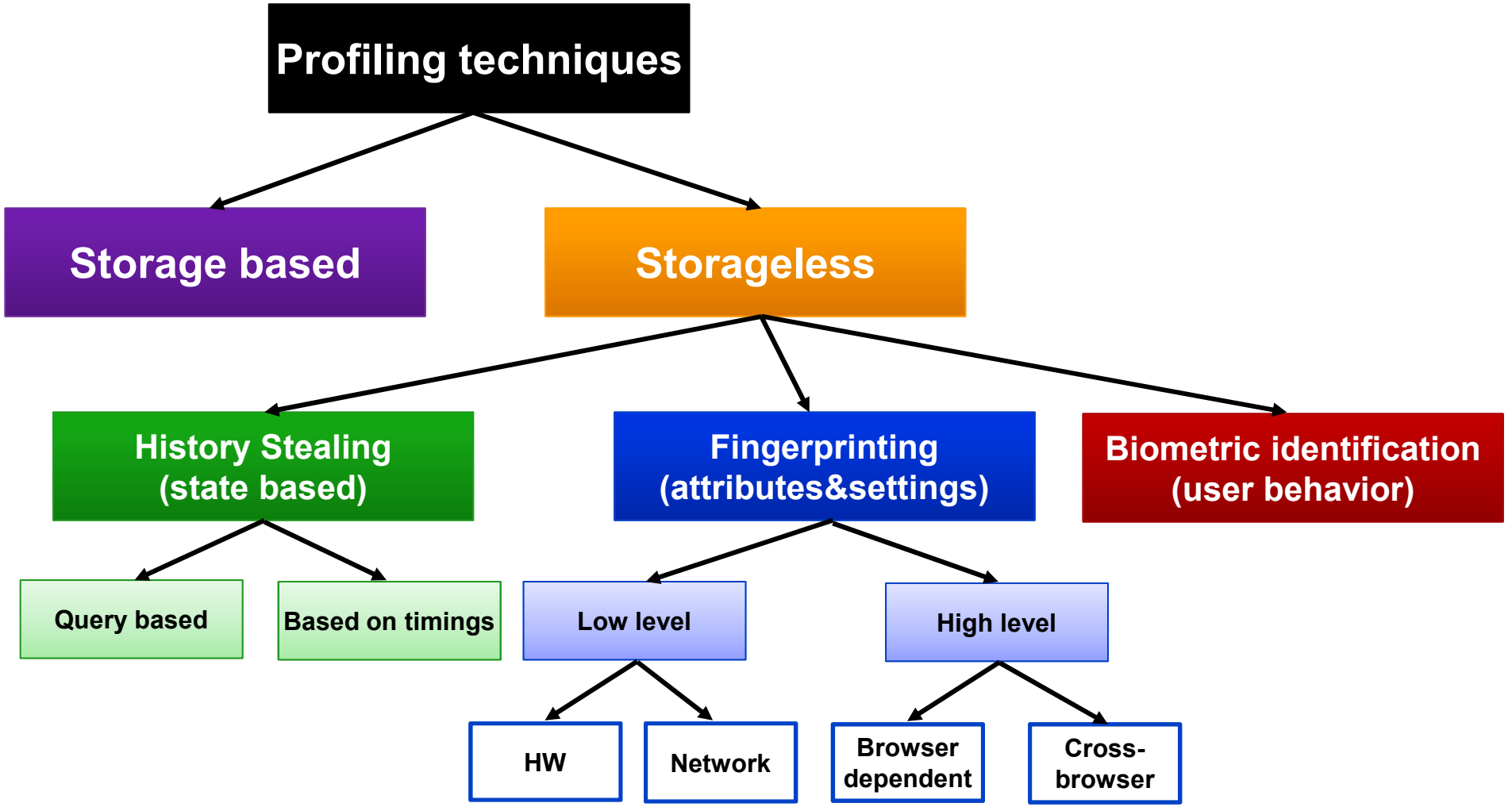
CONCLUSION

„Privacy-circle”

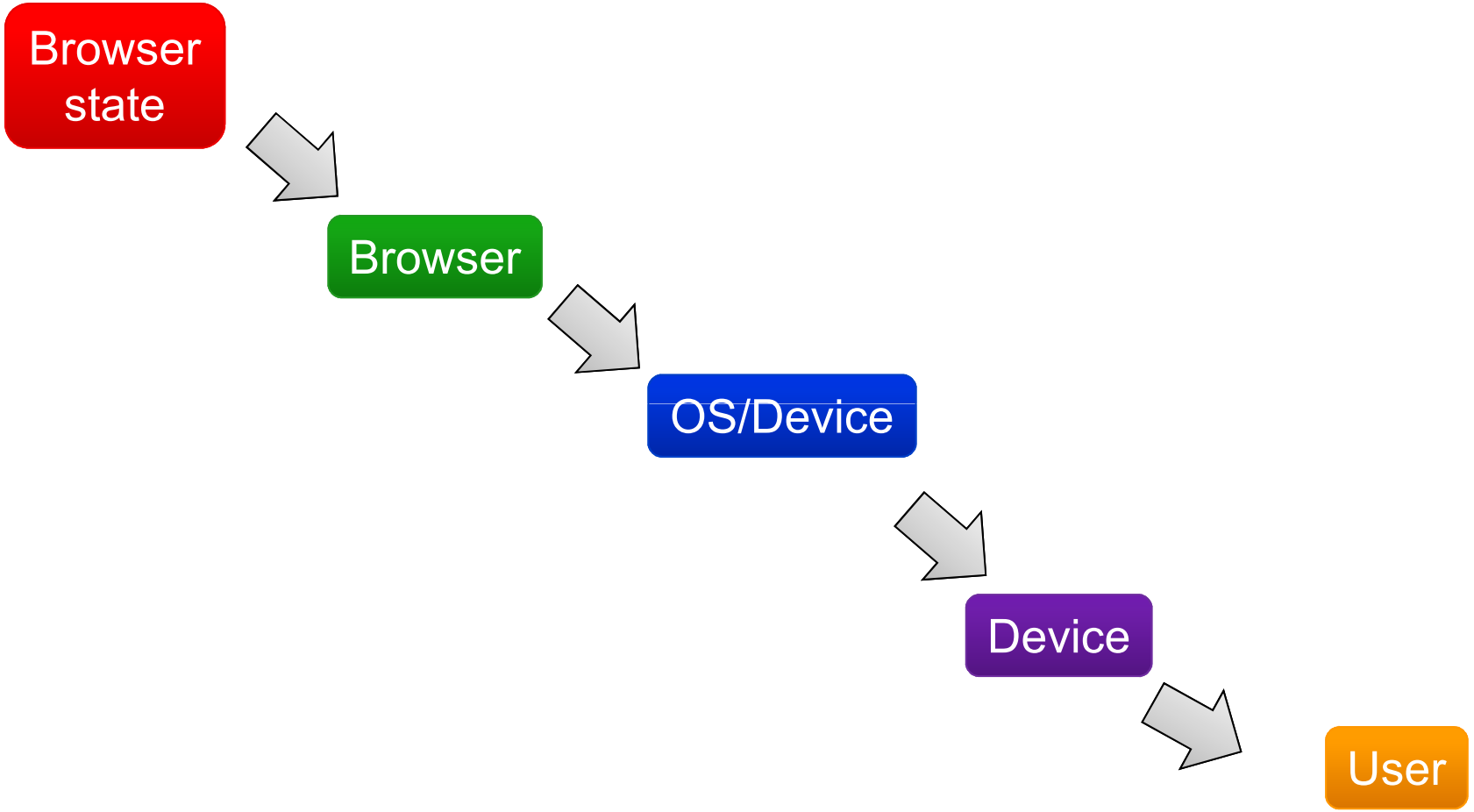


Taxonomy of attacks

Source: http://gulyas.info/upload/GulyasG_Fingerprinting2013.pdf



Where are we today?



What's next?

- Long term effects on society, economy, etc.?
  - Safe-point-of-return?
- Google AdID, Apple, Facebook?
- Mobile web → mobile apps
- Wearable tech
- Clear trends, no real solutions favoring privacy

Thank you for your attention!
Any questions?

Gábor György Gulyás
gulyas.info // @GulyasGG
Laboratory of Cryptography and System Security (CrySyS)
Budapest University of Technology and Economics
www.crysys.hu

References

- Interactive Advertising Bureau (IAB). (June 11, 2012). Internet Advertising Revenues Set First Quarter Record at $8.4 Billion, Interactive Advertising Bureau.
- Krishnamurthy, B., & Wills, C.E. (2006). Generating a privacy footprint on the Internet. In Proc. of the 6th ACM Conference on Internet Measurement.
- Krishnamurthy, B., & Wills, C.E. (2009). Privacy diffusion on the web: A longitudinal perspective. In Proc. of the 18th Conference on the World Wide Web.
- Krishnamurthy, B. (2010). Privacy leakage on the Internet. Presented at IETF 77, March 2010.
- Mayer, J.R., & Mitchell, J.C. (2012). Third-Party Web Tracking: Policy and Technology. In Proc. of the IEEE Symposium on Security and Privacy 2012.
- Latanya Sweeney: Uniqueness of simple demographics in the US population. *LIDAP-WP4. Carnegie Mellon University, Laboratory for International Data Privacy, Pittsburgh, PA* (2000).
- Philippe Golle: Revisiting the uniqueness of simple demographics in the US population. *Proceedings of the 5th ACM workshop on Privacy in electronic society*. ACM, 2006.

References (2)

- Roesner, F., Kohno, T., & Wetherall, D. (2012). Detecting and Defending Against Third-Party Tracking on the Web. In Proc. of 9th USENIX Symposium on Networked Systems Design and Implementation. San Jose, CA, USA
- Kontaxis, G., Polychronakis, M., Keromytis, A.D., Markatos, E.P. (2012). Privacy-Preserving Social Plugins. In Proc. of 12th USENIX Security Symposium. Bellevue, WA, USA
- Ayenson, M., Wambach, D.J., Soltani, A., Good, N., & Hoofnagle, C.J. (July 29, 2011). Flash Cookies and Privacy II: Now with HTML5 and ETag Respawning, SSRN. Retrieved from http://ssrn.com/abstract=1898390
- Benninger, C. (2006). AJAX Storage: A Look at Flash Cookies and Internet Explorer Persistence. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.128.2523

References (3)

- Davidov, M. (2011). The Double-Edged Sword of HSTS Persistence and Privacy. *Leviathan Security Group*. Retrieved from http://www.leviathansecurity.com/blog/archives/12-The-Double-Edged-Sword-of-HSTS-Persistence-and-Privacy.html
- Bursztein, E. (2011). Tracking users that block cookies with a HTTP redirect. Retrieved from http://elie.im/blog/security/tracking-users-that-block-cookies-with-a-http-redirect/
- Grossmann, J. (2007). Tracking users with Basic Auth. Retrieved from http://jeremiahgrossman.blogspot.hu/2007/04/tracking-users-without-cookies.html
- Olejnik, L., Minh-Dung, T., Castelluccia, C. (2013). Selling Off Privacy at Auction.
- N. Nikiforakis, et al. Cookieless monster: Exploring the ecosystem of web-based device fingerprinting. *Security and Privacy (SP), 2013 IEEE Symposium on*. IEEE, 2013.