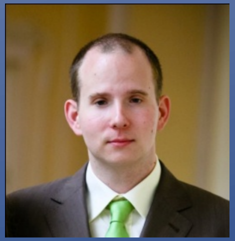
RE-IDENTIFICATION & FINGERPRINTING

2017-01-10

Gábor György Gulyás
Postdoc @ Privatics
http://gulyas.info // @GulyasGG



INTRODUCTION TO RE-IDENTIFICATION

Re-identification

- Definition
  - In a database a set of attributes can be considered as quasi identifiers. The database achieves k-anonymity if for all records there are at least (k-1) other rows with the same quasi identifier.
- Methods: supression or generalization

Explicit ID

Quasi ID

Sensitive attribute

Name	Birth date	City
John	1980-01-31	New York
Emily	1976-06-25	Flint
Bob	1985-09-05	New York
Dave	1973-02-07	South Bend
	…	

Employee database

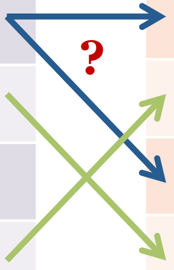
Birth date	City	Diagnosis
1985-09-05	New York	Stroke
1973-02-07	South Bend	-
1980-01-31	New York	Flu
1976-06-25	Flint	AIDS
	…	

Healthcare database

Re-identification & k-anonymity (2)

Employee database

Name	Birth date	City
John	1980-01-31	New York
Emily	1976-06-25	Flint
Bob	1985-09-05	New York
Dave	1973-02-07	South Bend



Healthcare database

Birth date	City	Diagnosis
198*	New York	Stroke
197*	South Bend	-
198*	New York	Flu
197*	Flint	AIDS

Better: P('John has flu')=1 → P('John has flu')= ½

Employee database

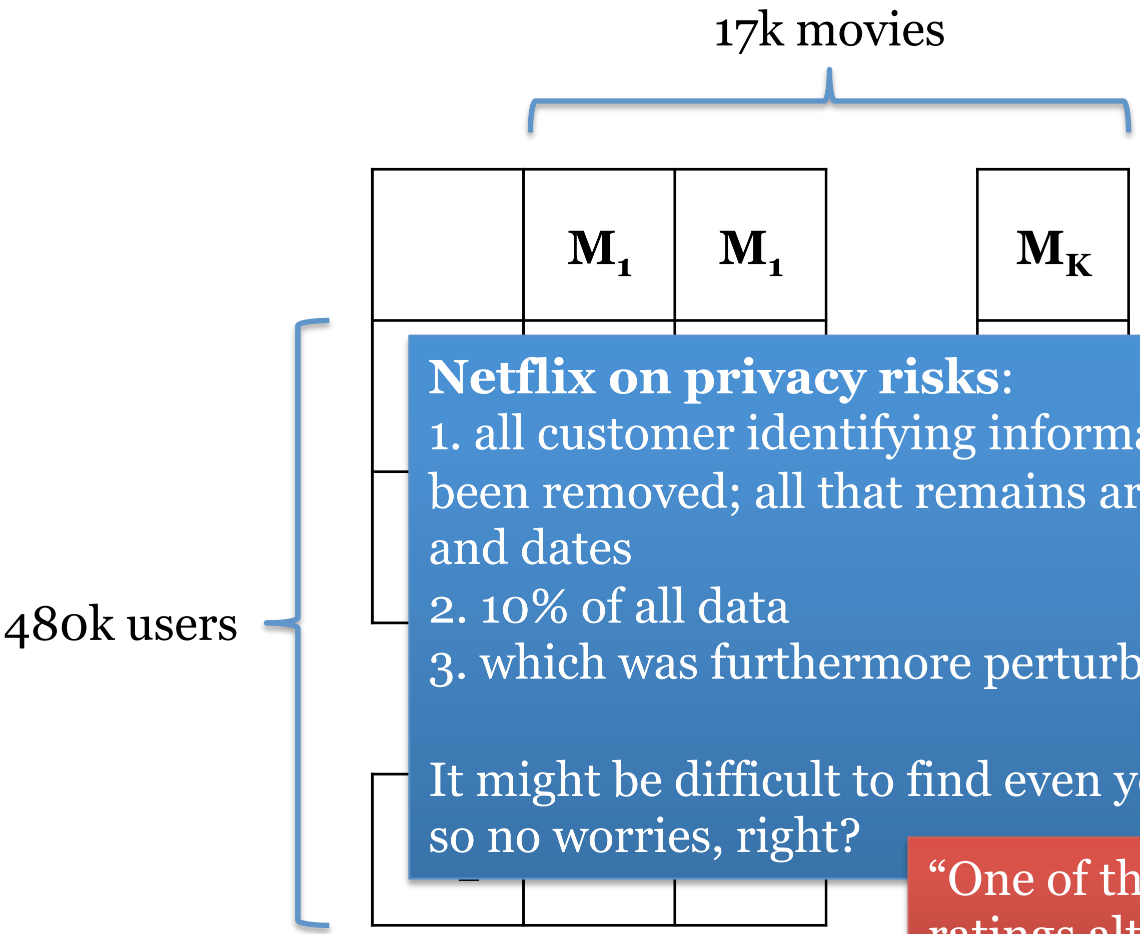
Name	Birth date	City
John	1980-01-31	New York
Emily	1976-06-25	Flint
Bob	1985-09-05	New York
Dave	1973-02-07	South Bend

Healthcare database

Birth date	City	Diagnosis
198*	New York	Stroke
197*	[small city]	-
198*	New York	Flu
197*	[small city]	AIDS

Even better: probs are now ½ for all! (2-anonymity)

The (in)famous Netflix case



100m ratings!
(1-5 stars+date)

Netflix on privacy risks:
1. all customer identifying information has been removed; all that remains are ratings and dates
2. 10% of all data
3. which was furthermore perturbed

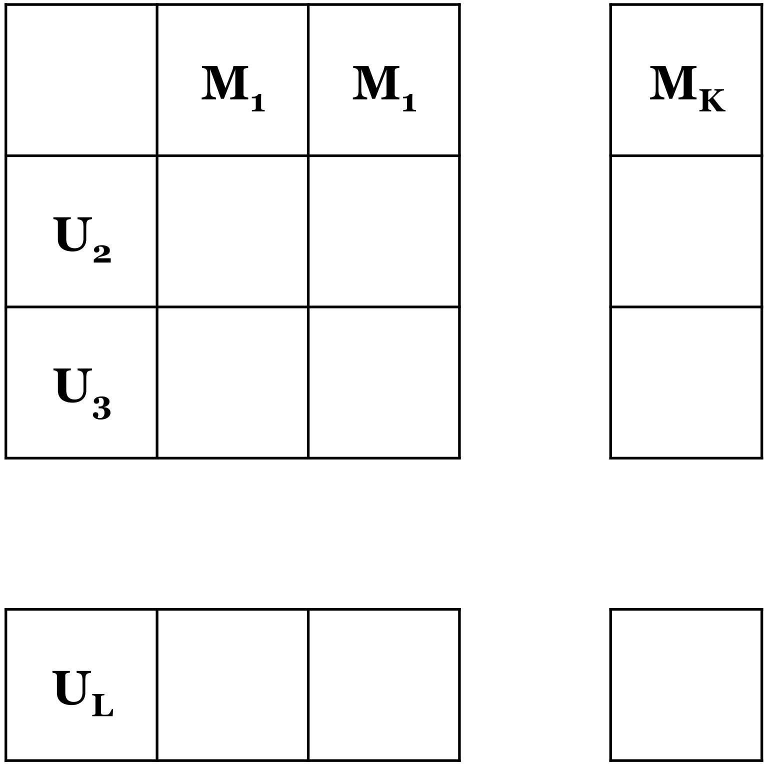
It might be difficult to find even yourself, so no worries, right?

“One of the subscribers had 1 of 306 ratings altered, and the other had 5 of 229 altered.”

Anonymized data r

The (in)famous Netflix case (2)

- Background knowledge?
  - A casual (workplace) conversation
  - Public ratings (IMDb)
  - …
- How to find users by these inaccurate sources?



	M_1	M_1		M_K
U_2				
U_3				
U_L				

Anonymized data release

The (in)famous Netflix case (3)

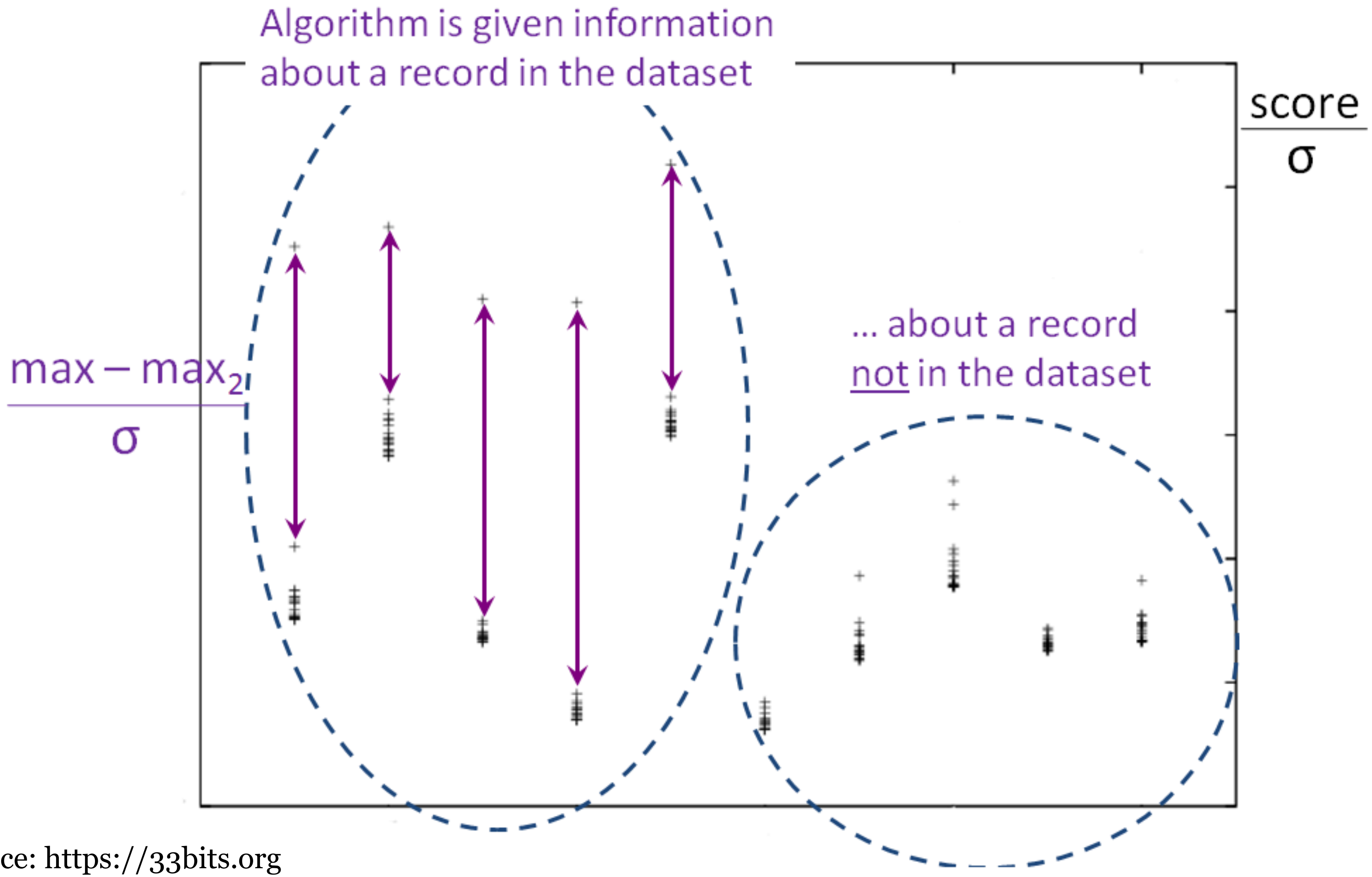
- Attack scheme
  - Obtain a couple (2-8) of ratings
  - Measure similarity against ratings in the dataset
    - Focuing on rarer ratings!
  - Is there a best candidate?
    - Check if it is meaningful!

A teaser from the results

- Exact ratings, dates with ±3/14 days, 5 ratings: de-anonymization with 80%
- Same setting, 7 ratings: above 90%
- Ratings ±1 stars, dates ±14 days
  - 4 ratings: 60% success
  - 8 ratings: 95% success

Anonymized data release

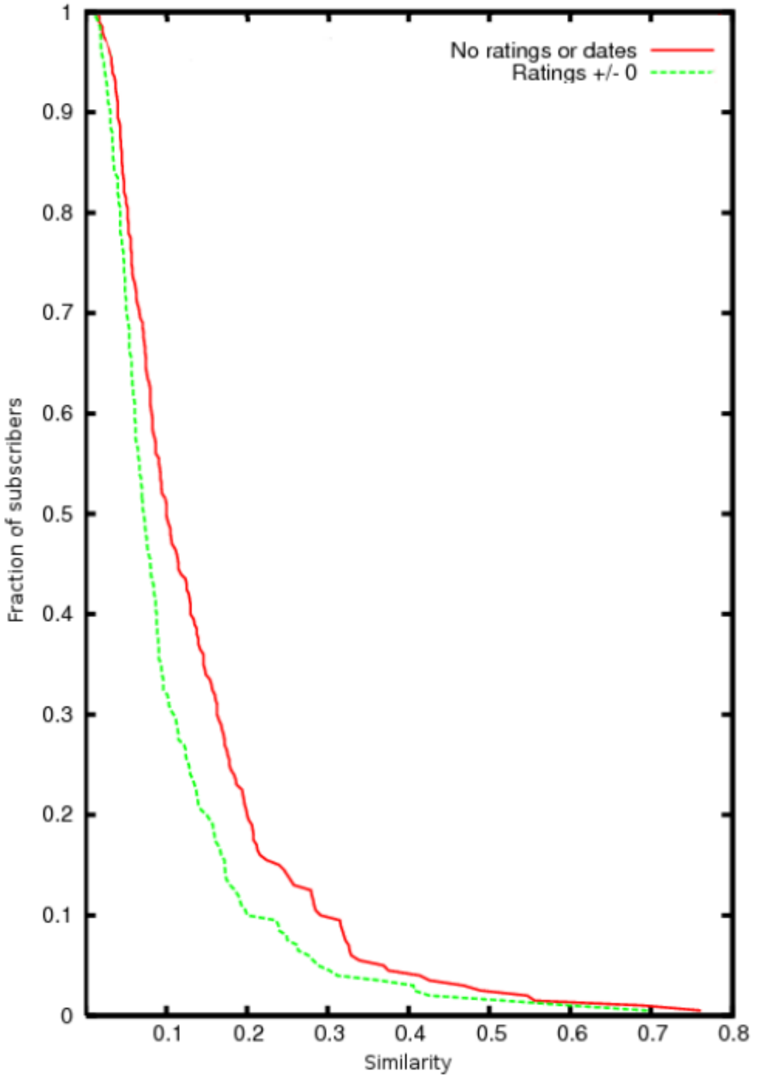
The (in)famous Netflix case (4)



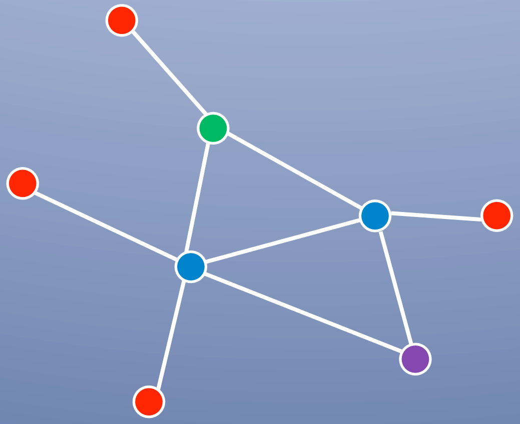
Source: https://33bits.org

Problems summarized

- Little information is enoughfor identification
  - 7 billion → 33 bits of information
- Low similarity of items
  - Large dimensionality of data
  - Heavy tail distribution of used attributes
  - Easy feature selection!
- Std anonymization fails & provability is hard



http://www.cs.cornell.edu/~shmat/shmat_oak08netflix.pdf

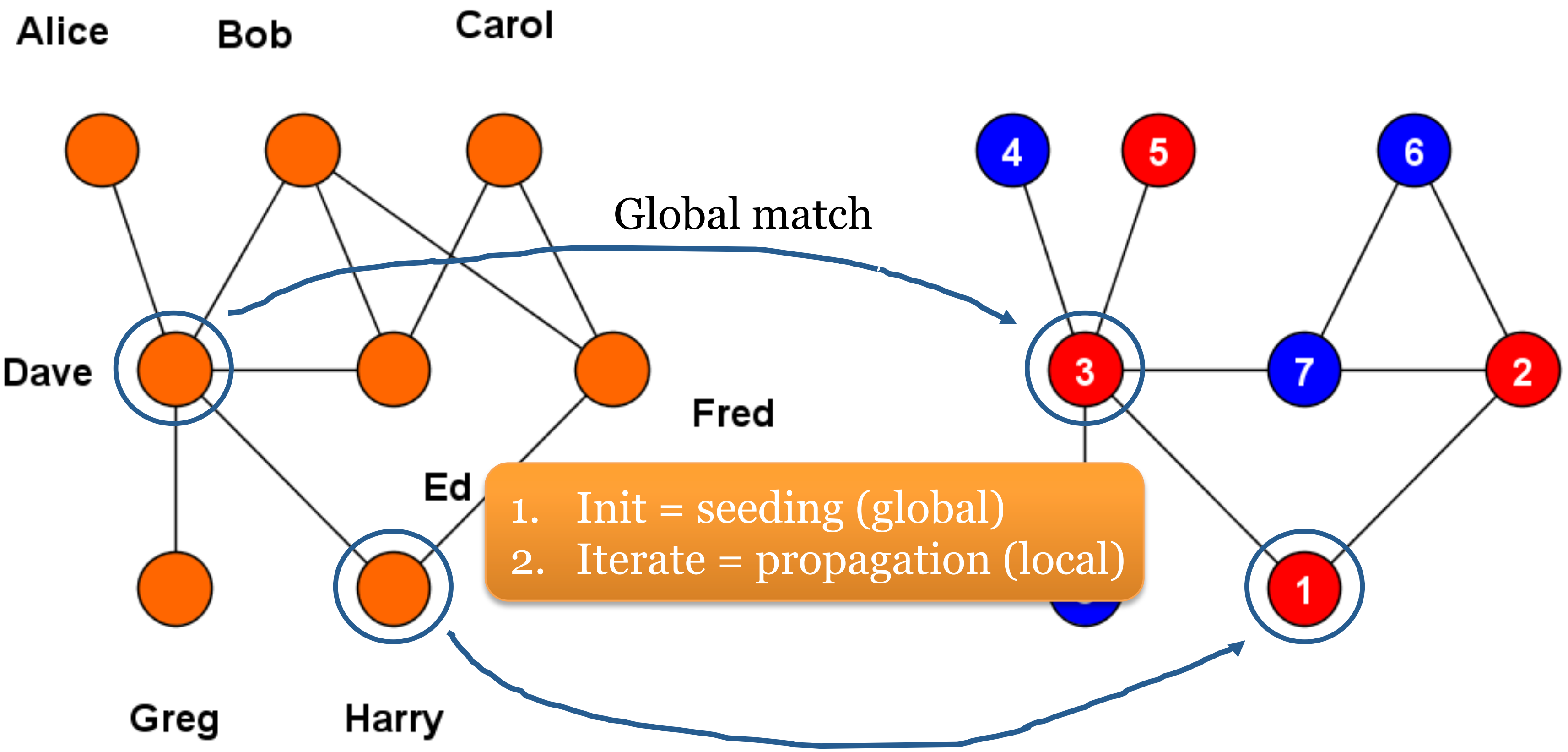


De-anonymizing Social Networks

Re-identification using the structure (2)

Auxiliary information, G_{src}
(a public crawl, e.g., Flickr)

Anonimized graph, G_{tar}
(anonimized export, e.g., Twitter)



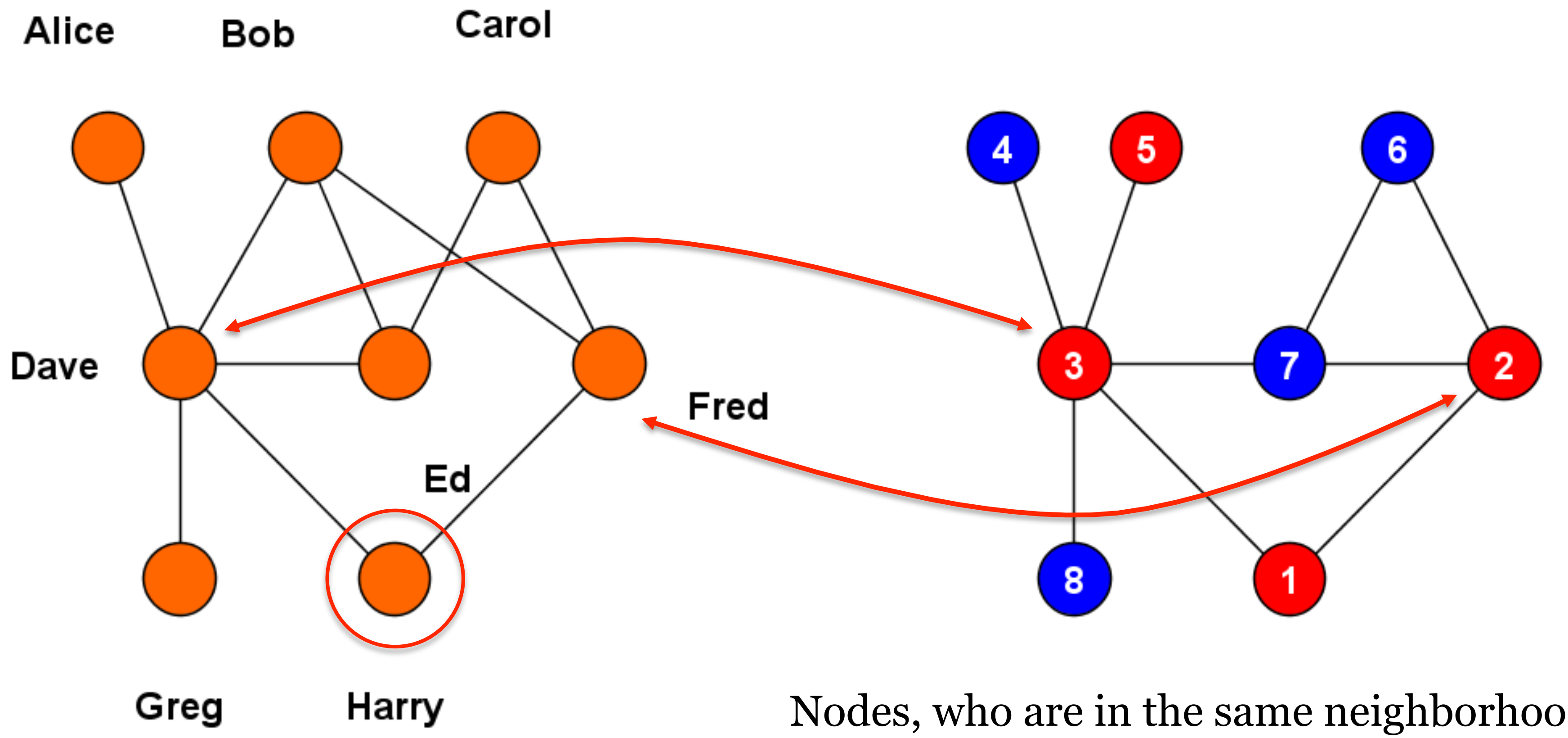
Global match

1. Init = seeding (global)
2. Iterate = propagation (local)

Relative match (local reid.)

Nar09 attack: propagation phase (3)

Auxiliary information, G_{src} **Anonimized graph, G_{tar}**



$$\mathrm{CosSim}(v_i, v_j) = \frac{|V_i \cap V_j|}{\sqrt{\cancel{|V_i|} \cdot |V_j|}}$$

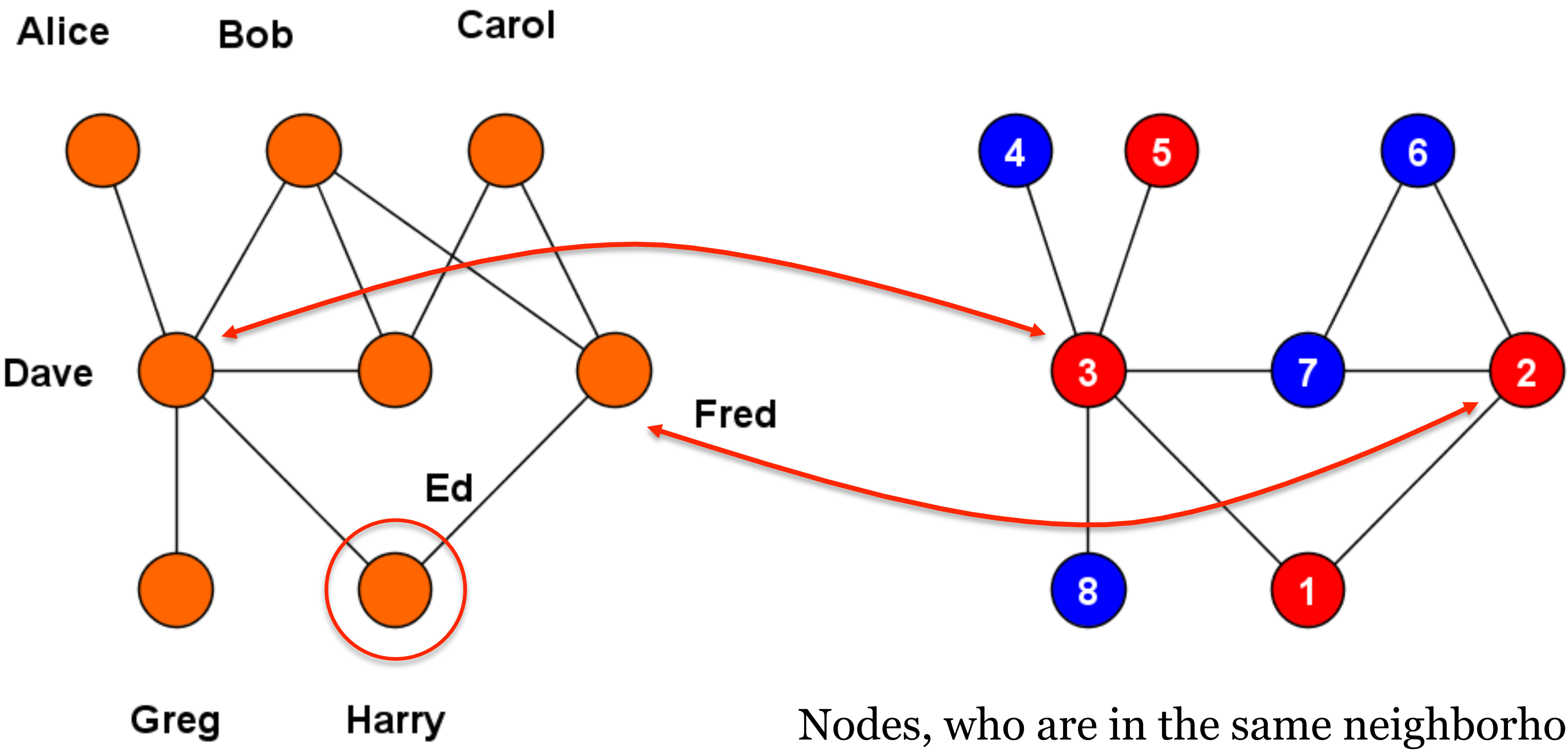
Nodes, who are in the same neighborhood:

$\mathbf{v_1}$	$\mathbf{v_4}$	$\mathbf{v_5}$	$\mathbf{v_6}$	$\mathbf{v_7}$	$\mathbf{v_8}$
1.4	1	1	0.7	1.1	1

↑ Is it good enough?

Nar09 attack: propagation phase (4)

Auxiliary information, G_{src} **Anonimized graph, G_{tar}**



$$\text{Eccentricity}(S) = \frac{\max(S) - \max(\{S \setminus \max(S)\})}{\sigma(S)} \overset{?}{\ge} \Theta$$

$$\text{Eccentricity}(S) = \frac{1.4 - 1.1}{0.22} = 1.36 > 1.0 = \Theta$$

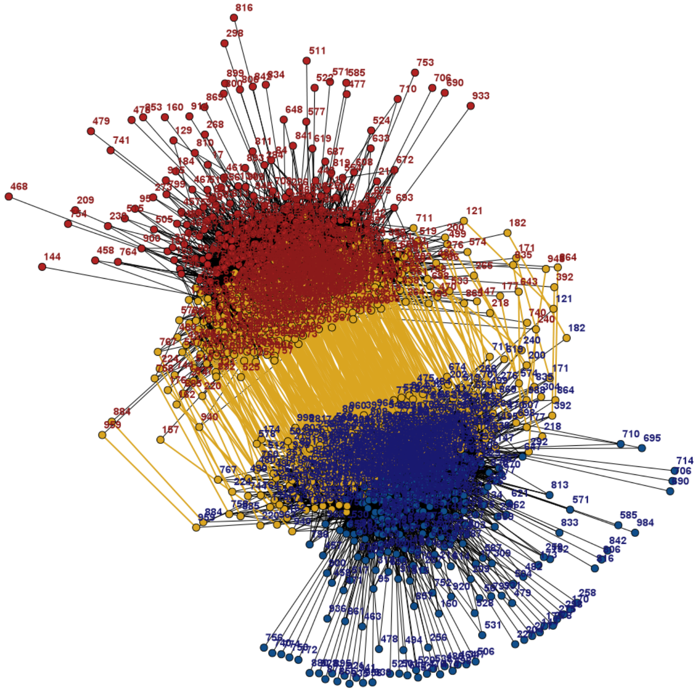
Nodes, who are in the same neighborhood:

v_1	v_4	v_5	v_6	v_7	v_8
1.4	1	1	0.7	1.1	1

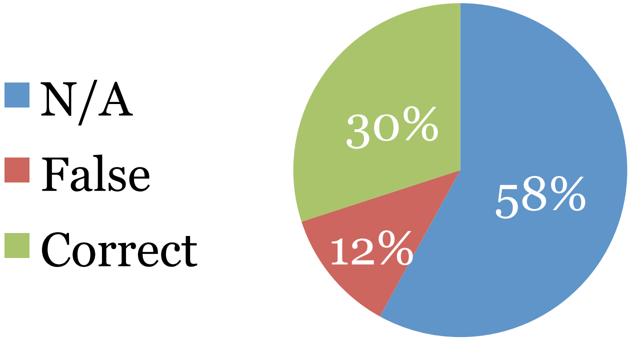
↑ Is it good enough?

Narayanan & Shmatikov results (Nar09)

- Large social networks
  - Background knowledge: Flickr (3,3m ns, 53m es)
  - Anonymous data: Twitter (224k ns, 8,5m es)



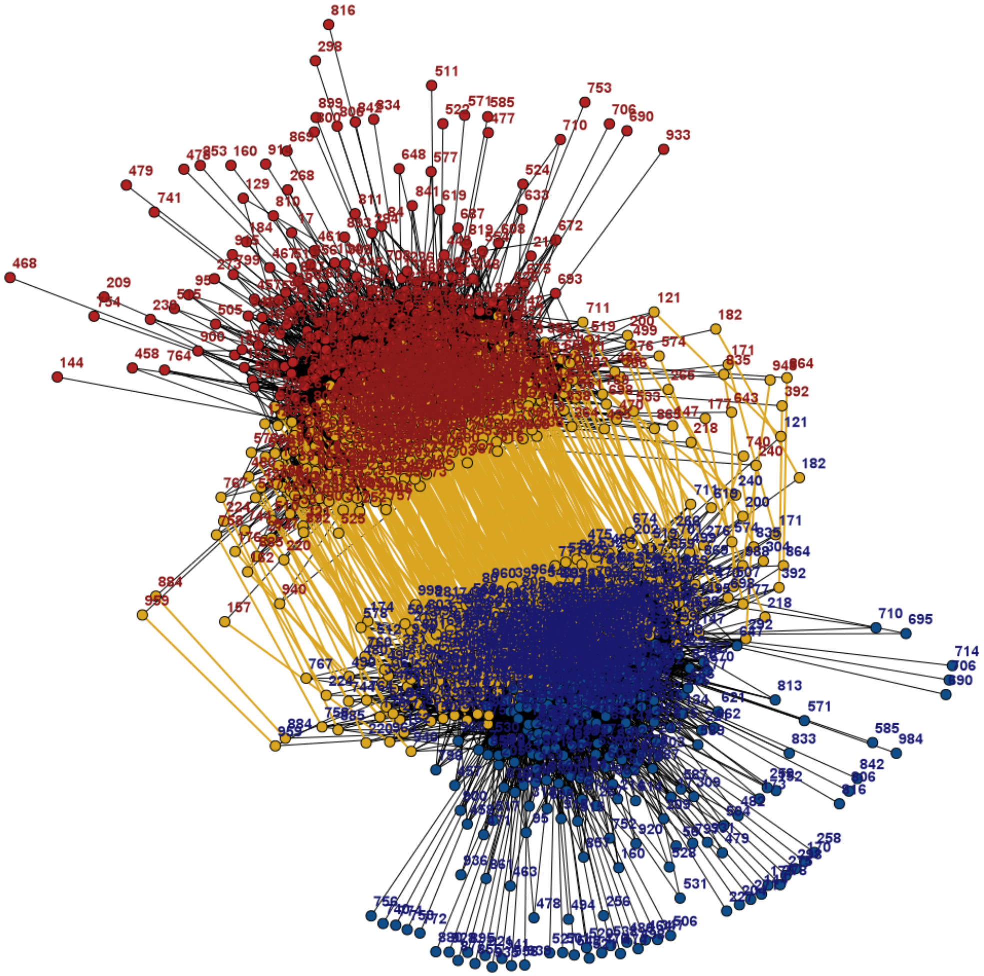
% of nodes



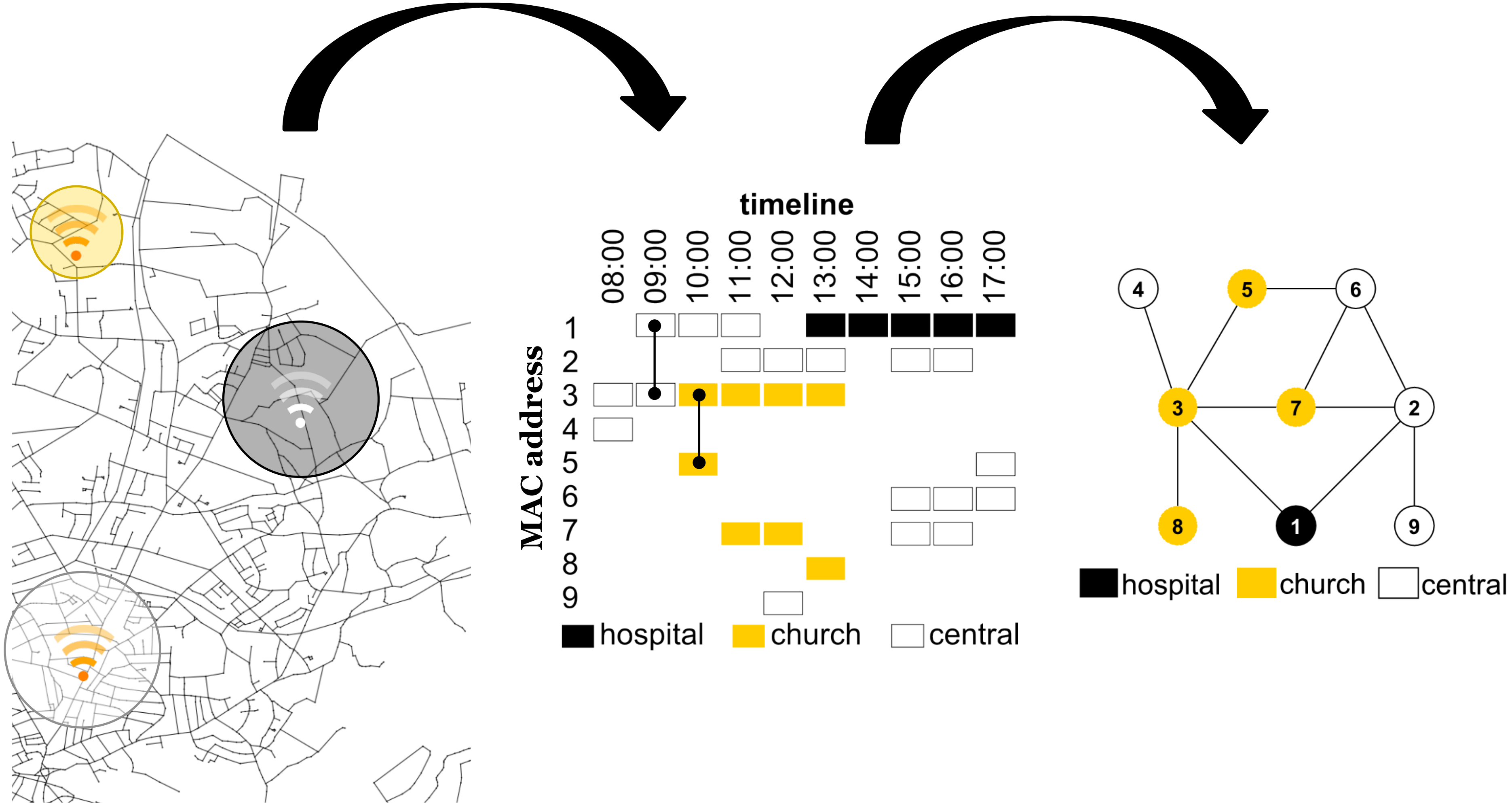
Ground truth of 27k nodes (verified by name/user/loc.)

Implications

- Linking identities in different datasets
  - Email vs. Phone
  - Social networks
  - …
- De-anonymizing anonymously published datasets with public data
  - e.g., other social networks



Implications (2)



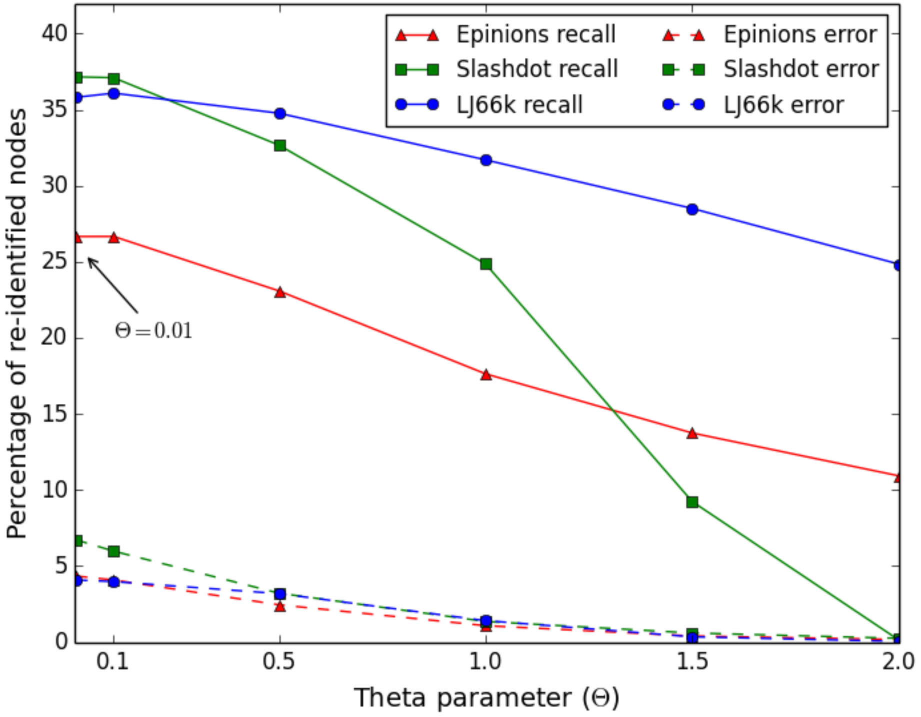
timeline

MAC address	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00
1		□	□	□		■	■	■	■	■
2				□	□	□		□	□	
3	□	□	🟨	🟨	🟨	🟨				
4	□									
5			🟨							□
6								□	□	□
7				🟨	🟨			□	□	
8						🟨				
9					□					

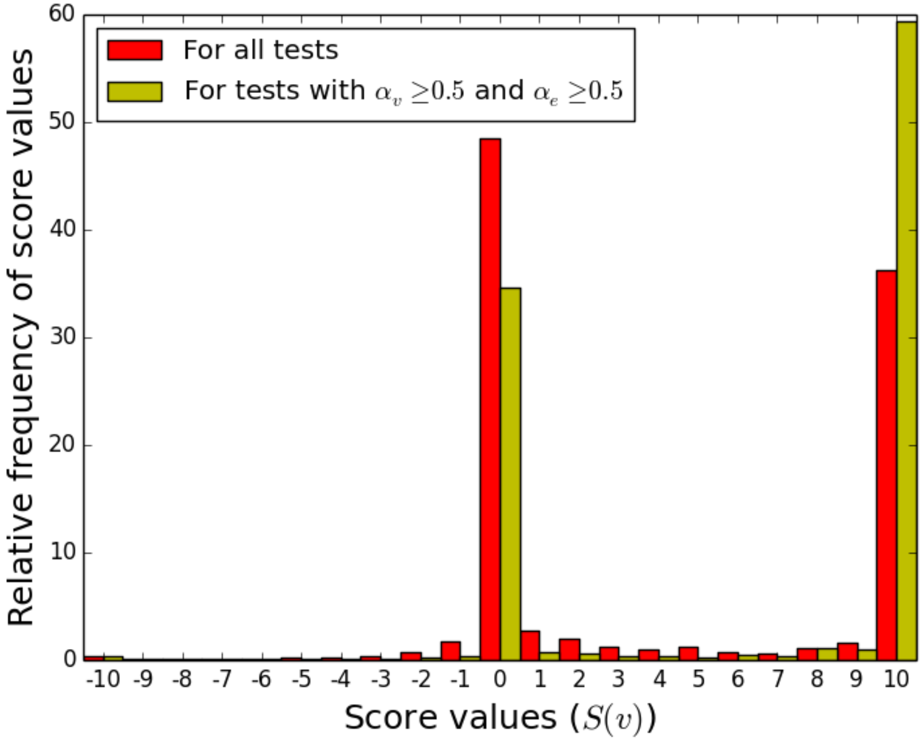
■ hospital 🟨 church □ central

Nar09 attack: properties

- Θ controls yield & error

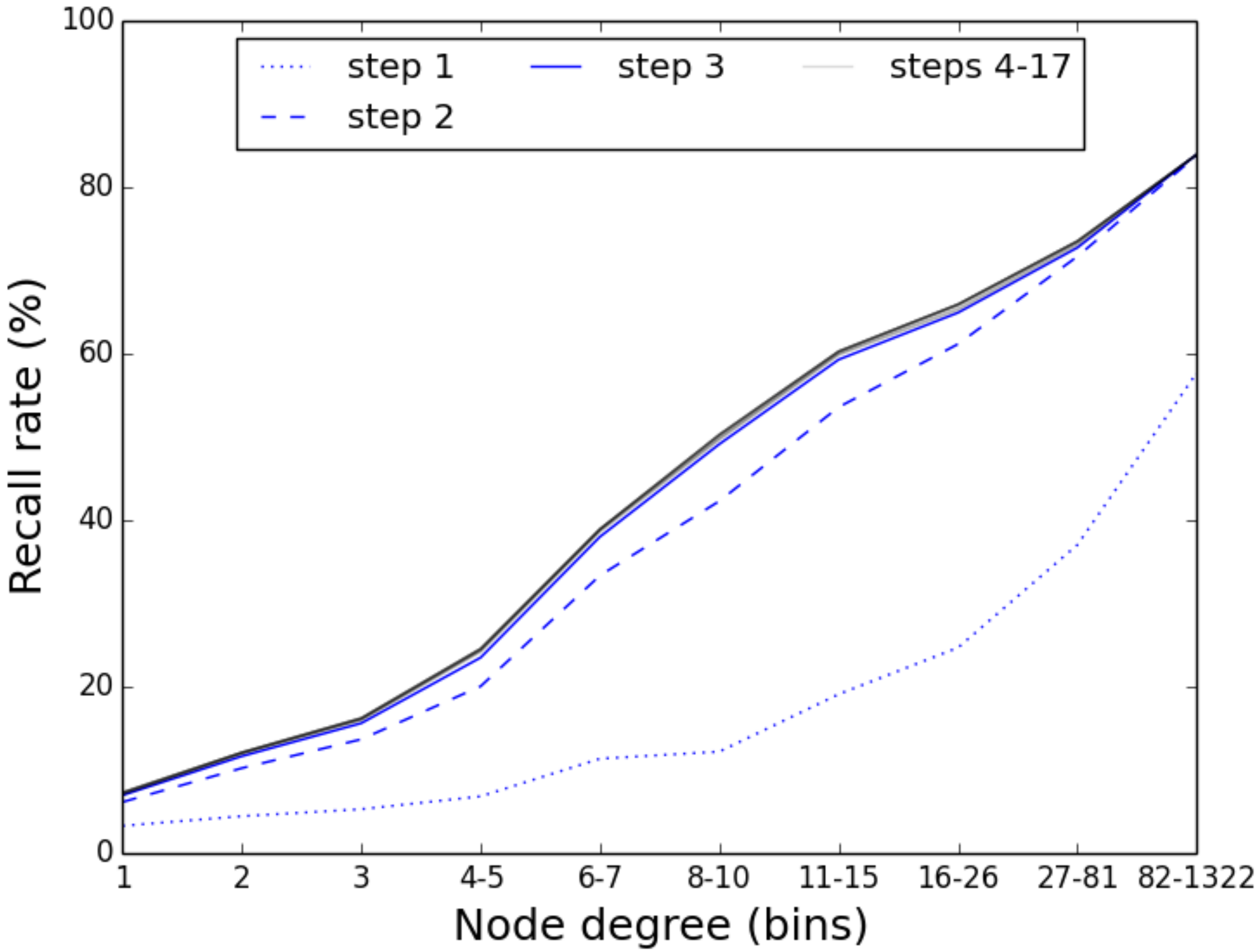


- More-or-less deterministic



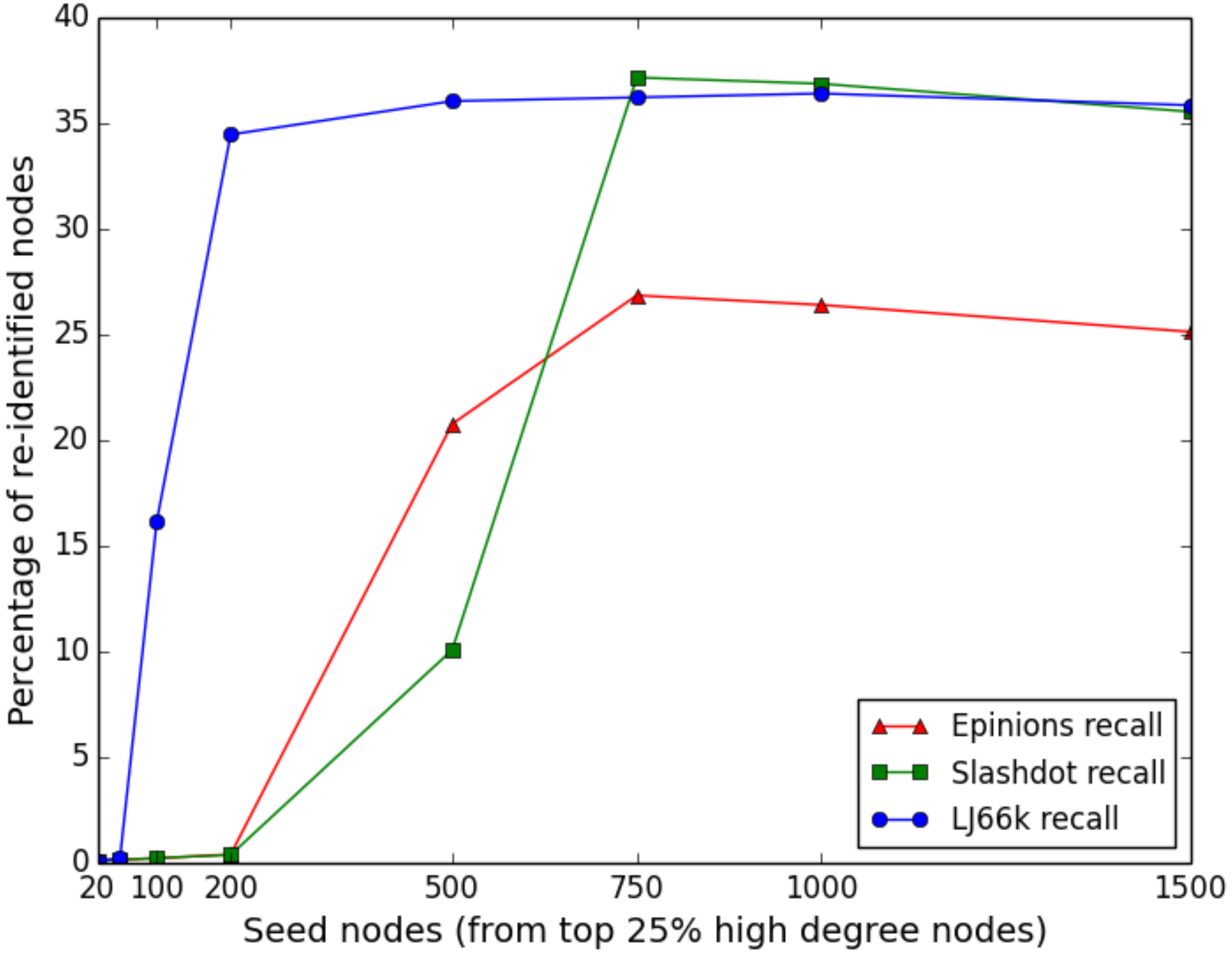
Nar09 attack: properties (2)

- Slow convergence + biased towards high degree



Nar09 attack: properties (3)

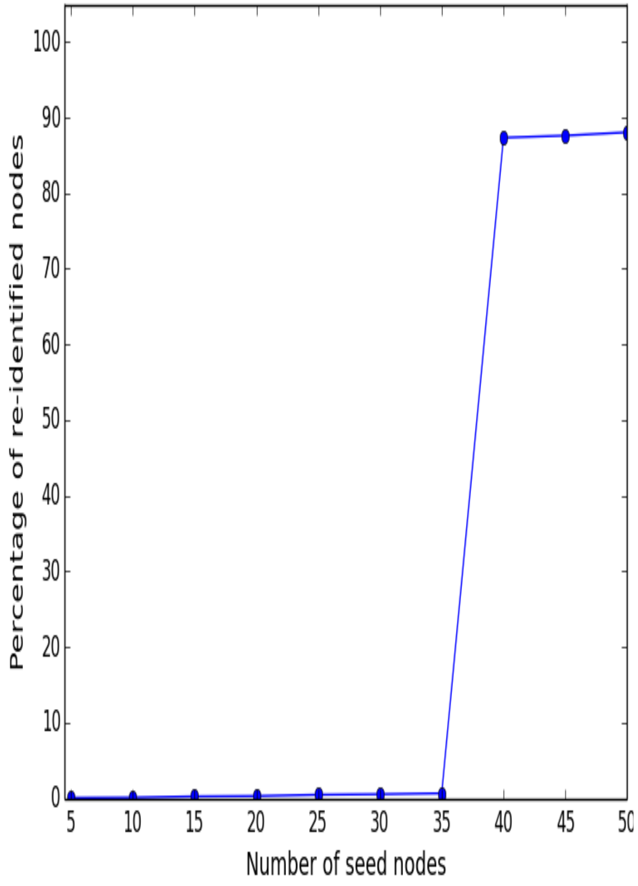
- Phase transition & total yield: depends on network



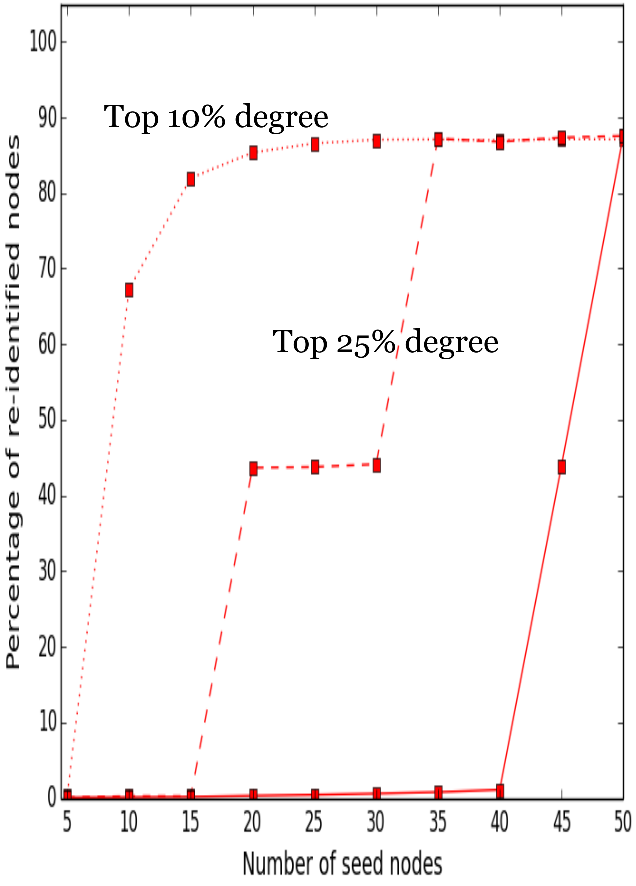
State-of-the-art attack: properties (4)

- Phase transition: also depends on seeding type

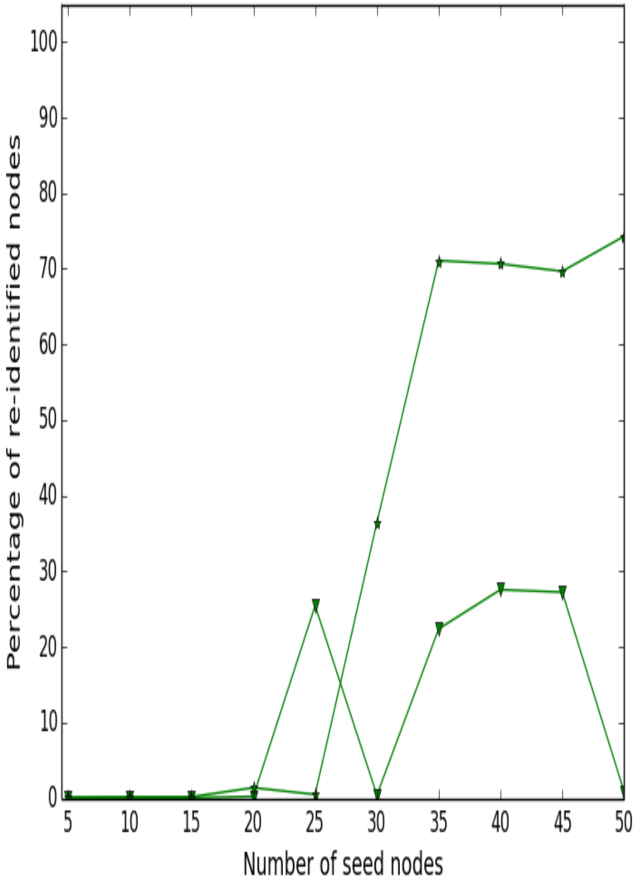
Top degree nodes



High betweenness + degree



Low clustering coeff.

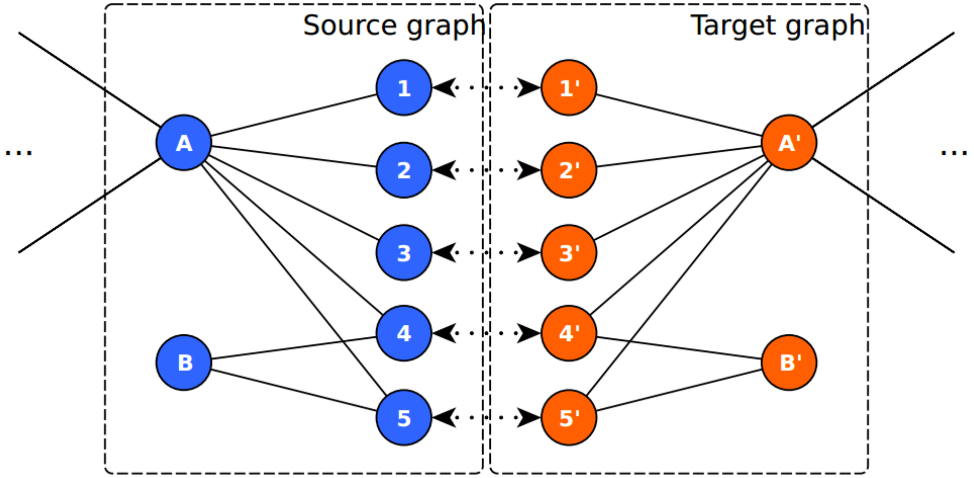


Joint work with **Benedek Simon, Sándor Imre**

[https://gulyas.info/files/publications/GulyasG_WPES16.pdf]

BUMBLEBEE

Motivation for Bumblebee



$$\mathrm{NarSim}(v_i, v_j) = \frac{\#\mathrm{mutual\_nbrs}}{\sqrt{\deg(v_j)}}$$

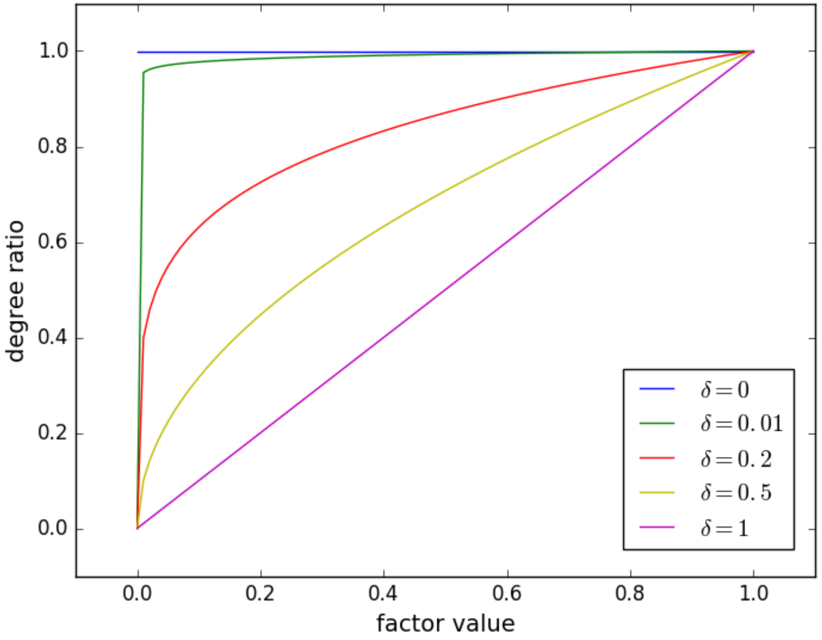
$$\mathrm{BlbSim}(v_i, v_j) = \#\mathrm{mutual\_nbrs} \cdot \left(\min\left(\frac{\deg(v_i)}{\deg(v_j)}, \frac{\deg(v_j)}{\deg(v_i)}\right)\right)^{\delta}$$

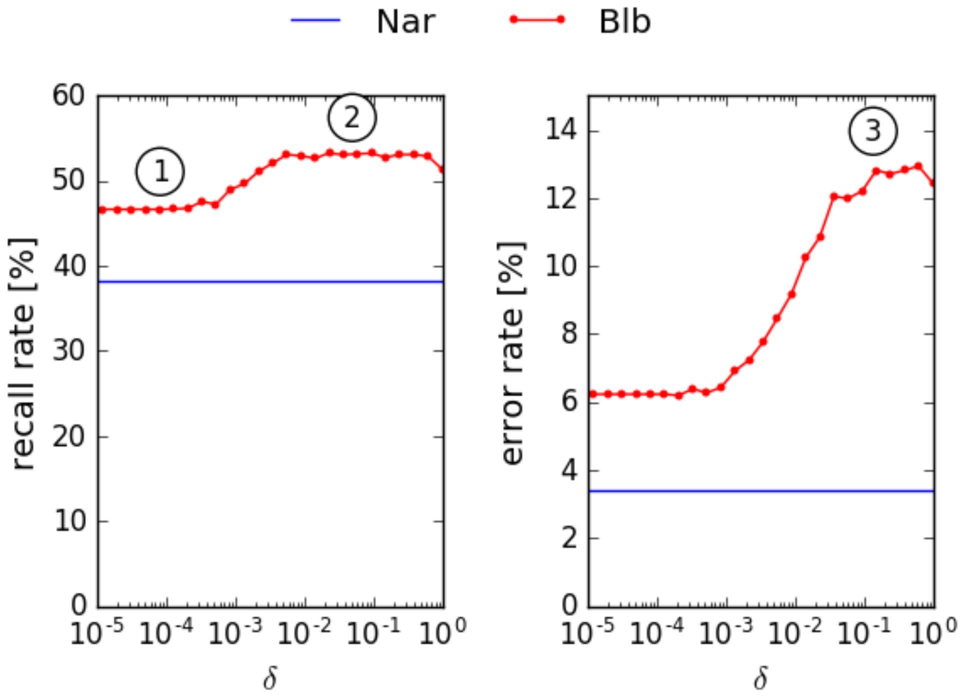
	A’	**B’**	**?**
A	$\frac{5}{\sqrt{100}}$	$\frac{2}{\sqrt{2}}$	B’
B	$\frac{2}{\sqrt{100}}$	$\frac{2}{\sqrt{2}}$	B’

	A’	**B’**	**?**
A	5	0.89	A’
B	0.89	2	B’

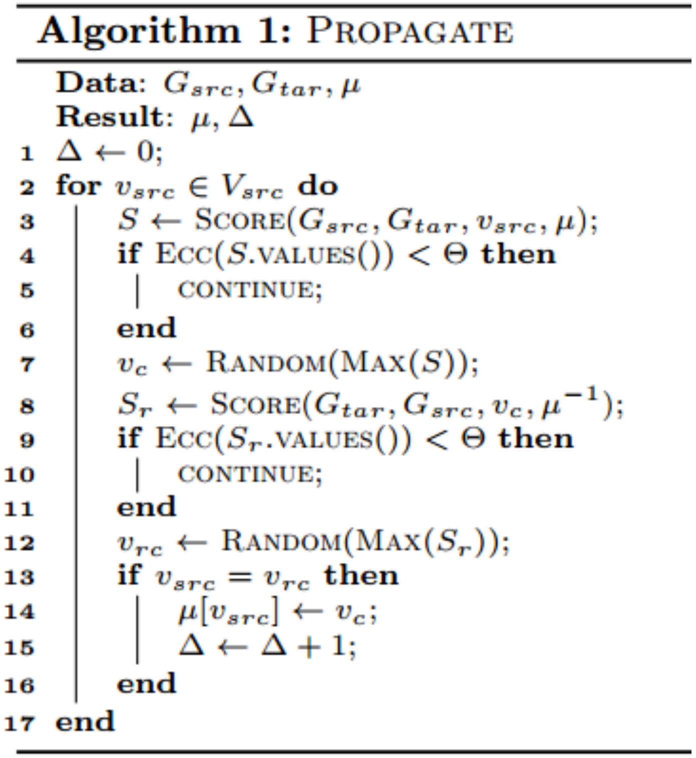
Parameters of the attack – δ

$$\mathrm{BlbSim}(v_i, v_j) = \#\mathrm{mutual\_nbrs} \cdot \left(\min\left(\frac{\deg(v_i)}{\deg(v_j)}, \frac{\deg(v_j)}{\deg(v_i)} \right) \right)^{\delta}$$

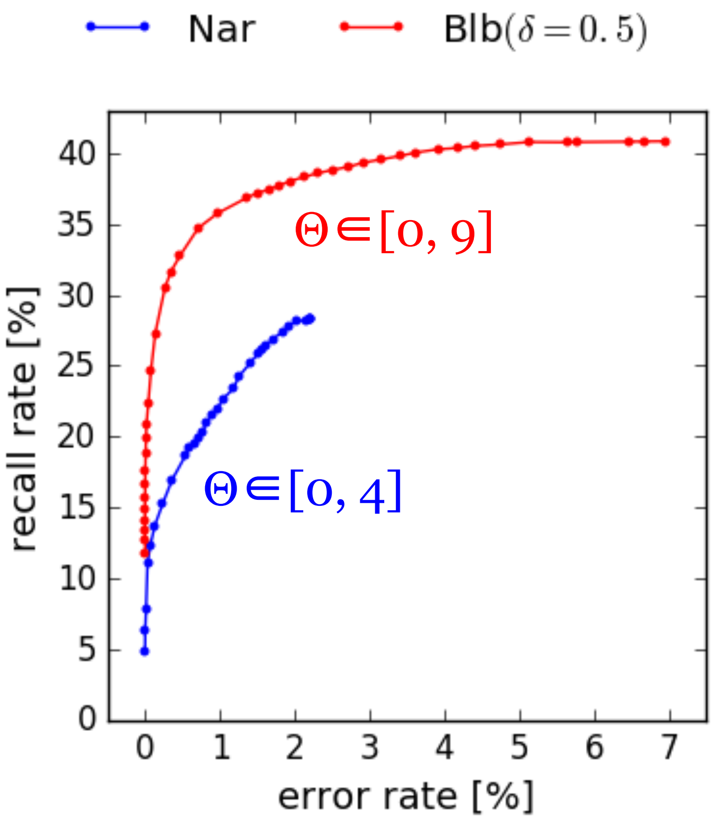




Parameters of the attack (2) – Θ



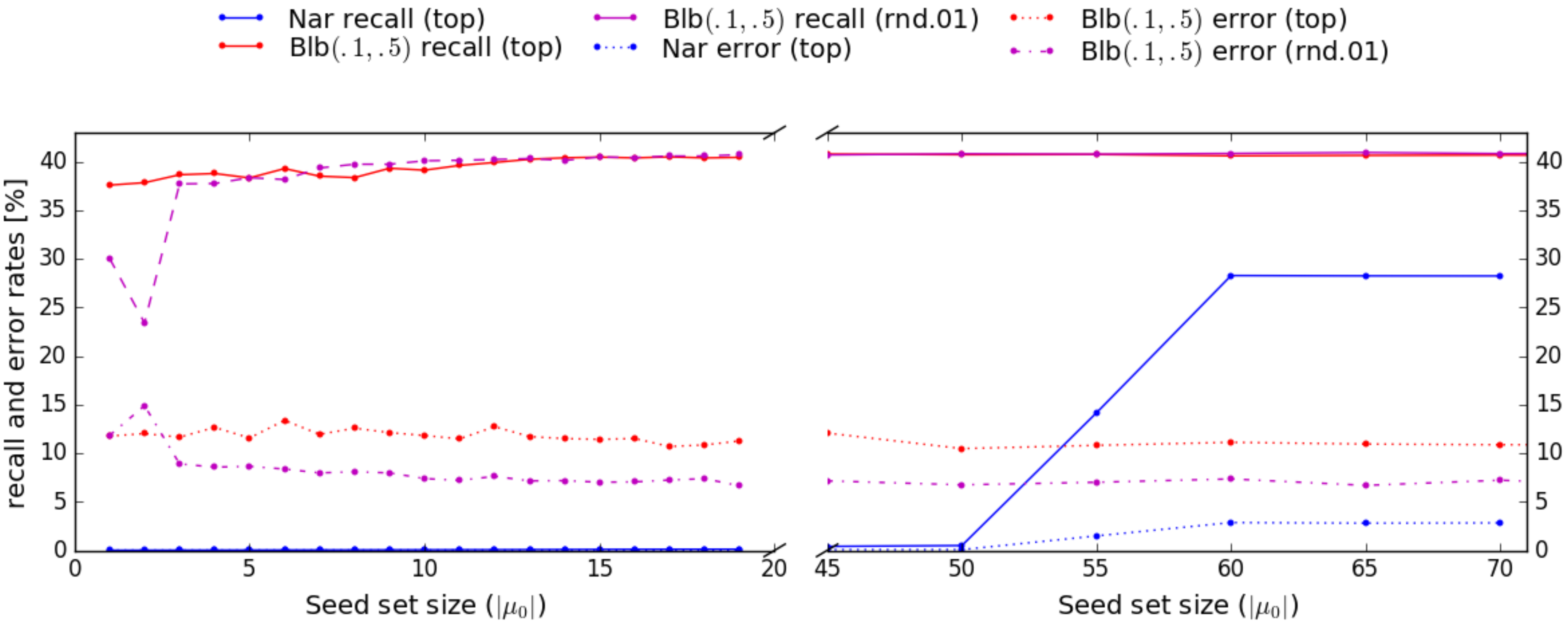
```
Algorithm 1: PROPAGATE
Data: $G_{src}, G_{tar}, \mu$
Result: $\mu, \Delta$
$\Delta \leftarrow 0$;
for $v_{src} \in V_{src}$ do
    $S \leftarrow$ SCORE($G_{src}, G_{tar}, v_{src}, \mu$);
    if ECC($S$.VALUES()) $< \Theta$ then
        CONTINUE;
    end
    $v_c \leftarrow$ RANDOM(MAX($S$));
    $S_r \leftarrow$ SCORE($G_{tar}, G_{src}, v_c, \mu^{-1}$);
    if ECC($S_r$.VALUES()) $< \Theta$ then
        CONTINUE;
    end
    $v_{rc} \leftarrow$ RANDOM(MAX($S_r$));
    if $v_{src} = v_{rc}$ then
        $\mu[v_{src}] \leftarrow v_c$;
        $\Delta \leftarrow \Delta + 1$;
    end
end
```



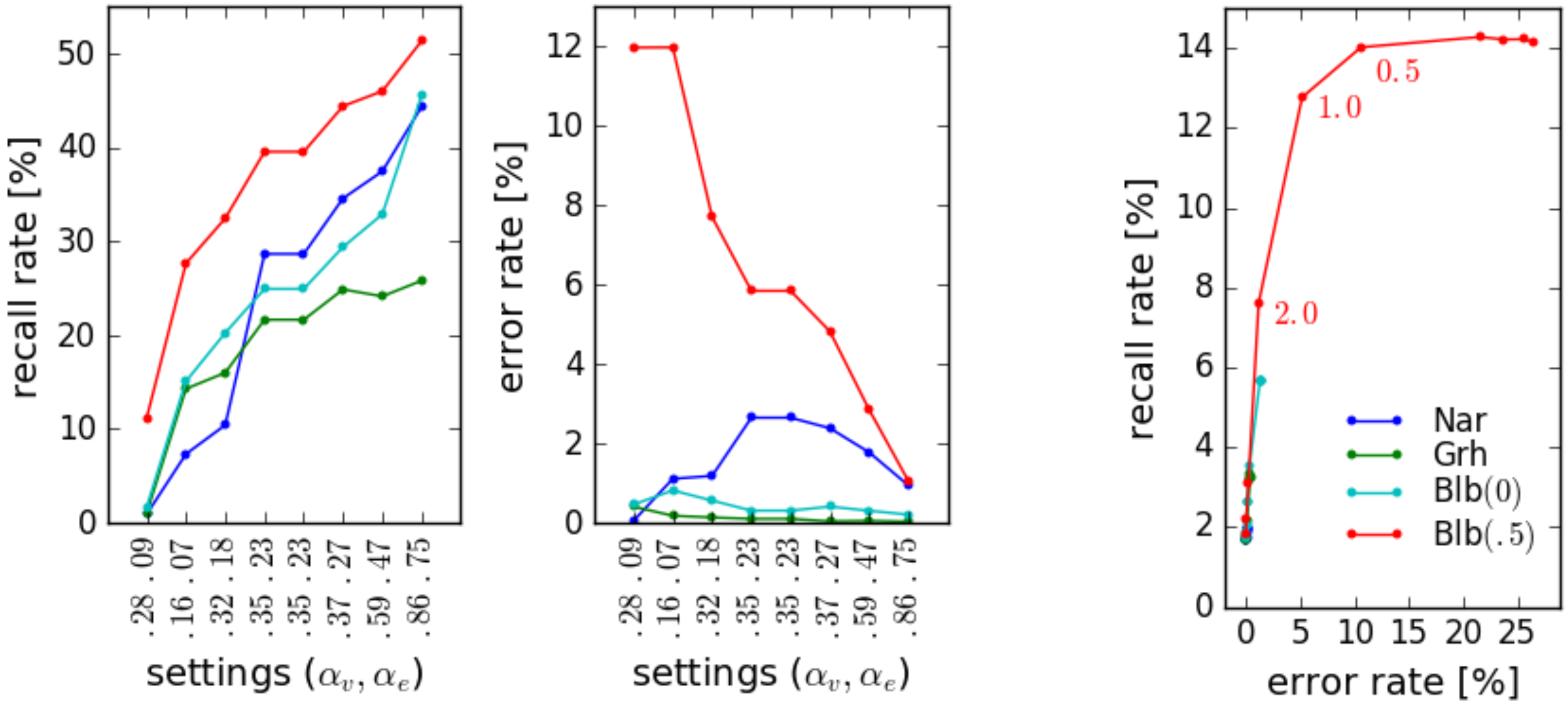
$\Theta \in [0, 9]$

$\Theta \in [0, 4]$

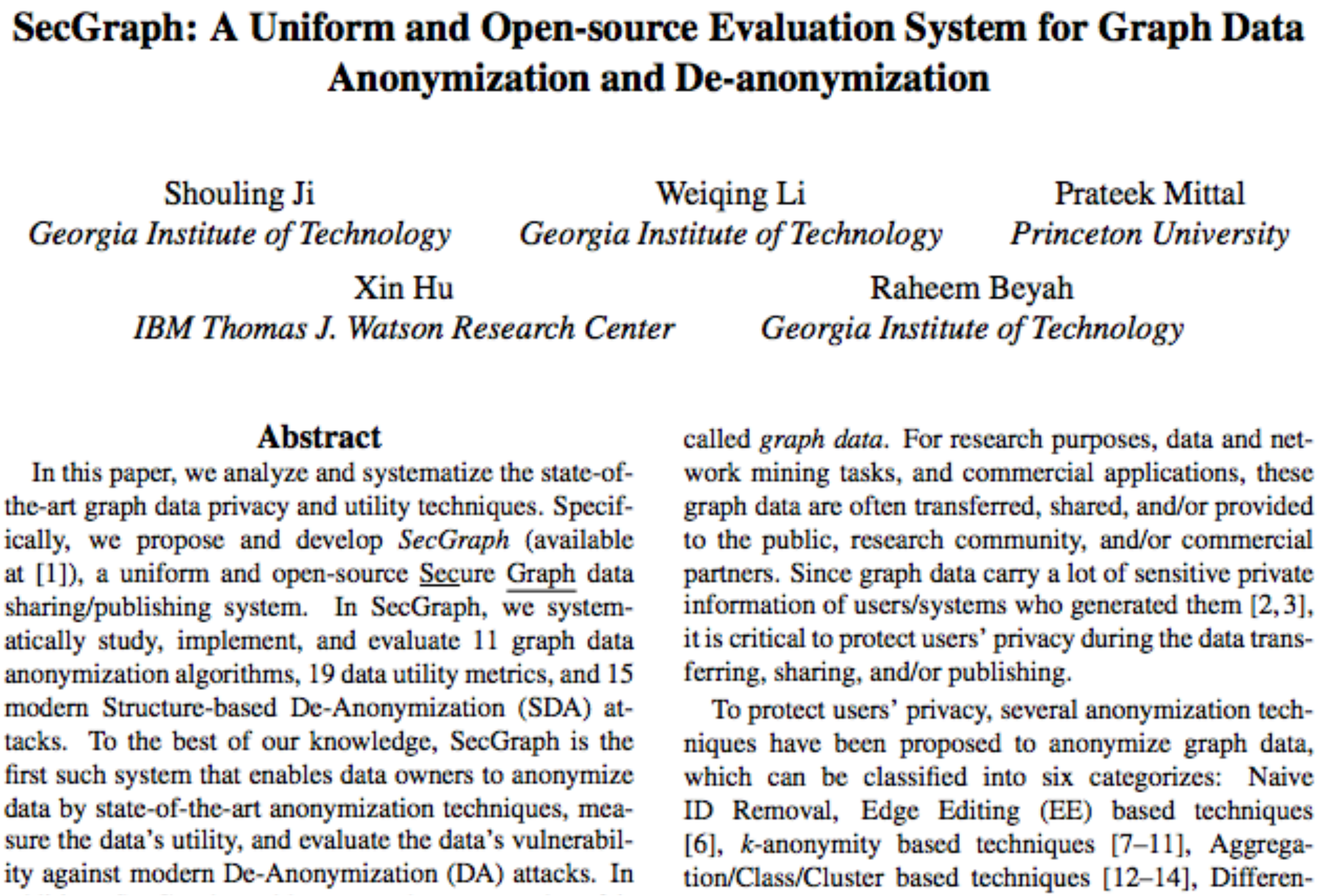
Seeding sensitvity



Robustness to noise



Comparison with other attacks



SecGraph: A Uniform and Open-source Evaluation System for Graph Data Anonymization and De-anonymization

Shouling Ji
Georgia Institute of Technology

Weiqing Li
Georgia Institute of Technology

Prateek Mittal
Princeton University

Xin Hu
IBM Thomas J. Watson Research Center

Raheem Beyah
Georgia Institute of Technology

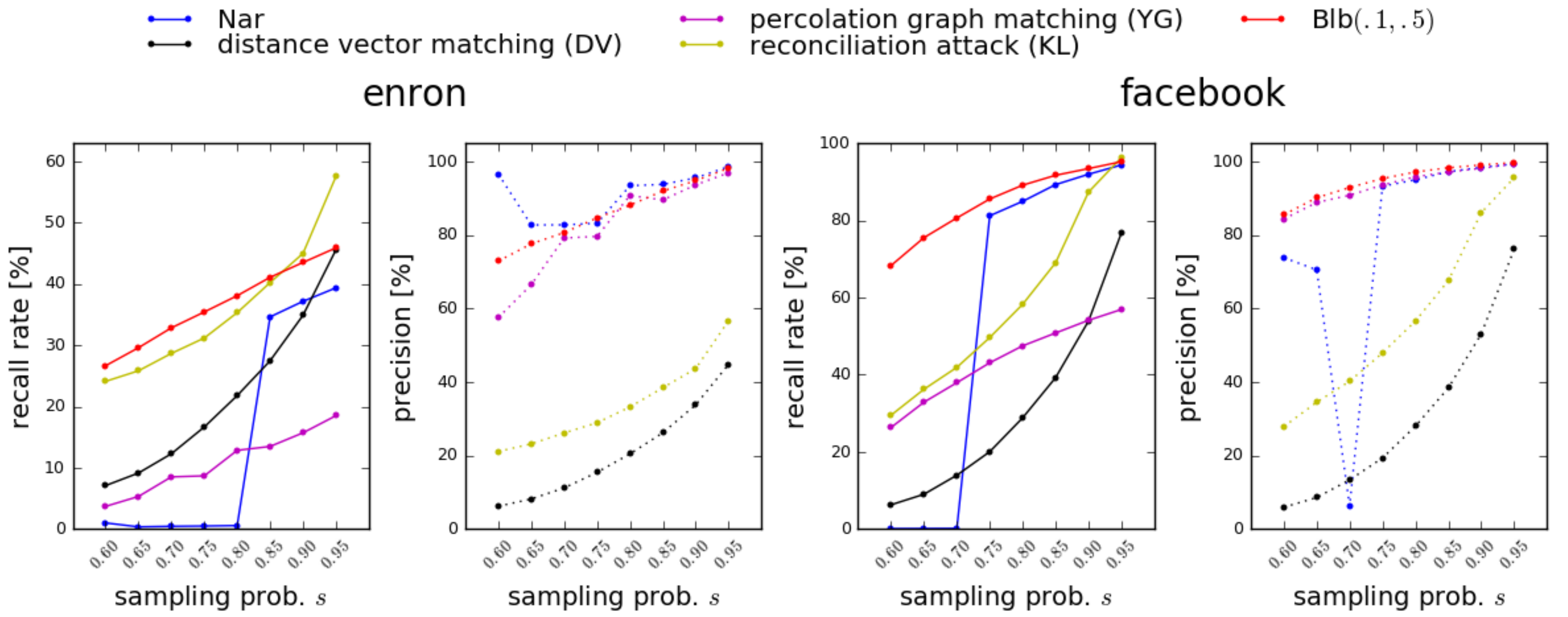
Abstract

In this paper, we analyze and systematize the state-of-the-art graph data privacy and utility techniques. Specifically, we propose and develop *SecGraph* (available at [1]), a uniform and open-source Secure Graph data sharing/publishing system. In SecGraph, we systematically study, implement, and evaluate 11 graph data anonymization algorithms, 19 data utility metrics, and 15 modern Structure-based De-Anonymization (SDA) attacks. To the best of our knowledge, SecGraph is the first such system that enables data owners to anonymize data by state-of-the-art anonymization techniques, measure the data's utility, and evaluate the data's vulnerability against modern De-Anonymization (DA) attacks. In

called *graph data*. For research purposes, data and network mining tasks, and commercial applications, these graph data are often transferred, shared, and/or provided to the public, research community, and/or commercial partners. Since graph data carry a lot of sensitive private information of users/systems who generated them [2, 3], it is critical to protect users' privacy during the data transferring, sharing, and/or publishing.

To protect users' privacy, several anonymization techniques have been proposed to anonymize graph data, which can be classified into six categorizes: Naive ID Removal, Edge Editing (EE) based techniques [6], *k*-anonymity based techniques [7–11], Aggregation/Class/Cluster based techniques [12–14], Differen-

Comparison with other attacks (2)

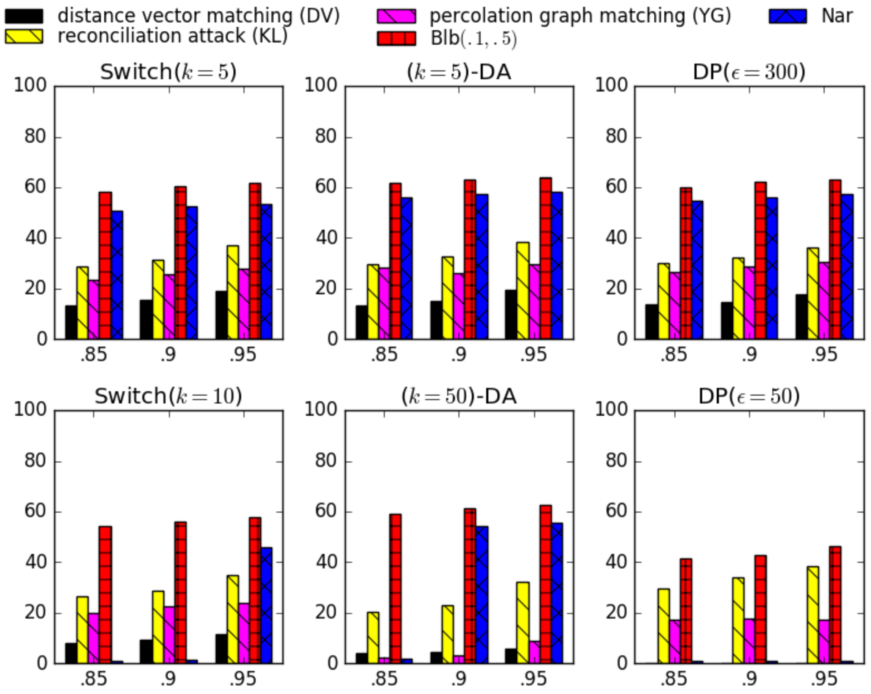


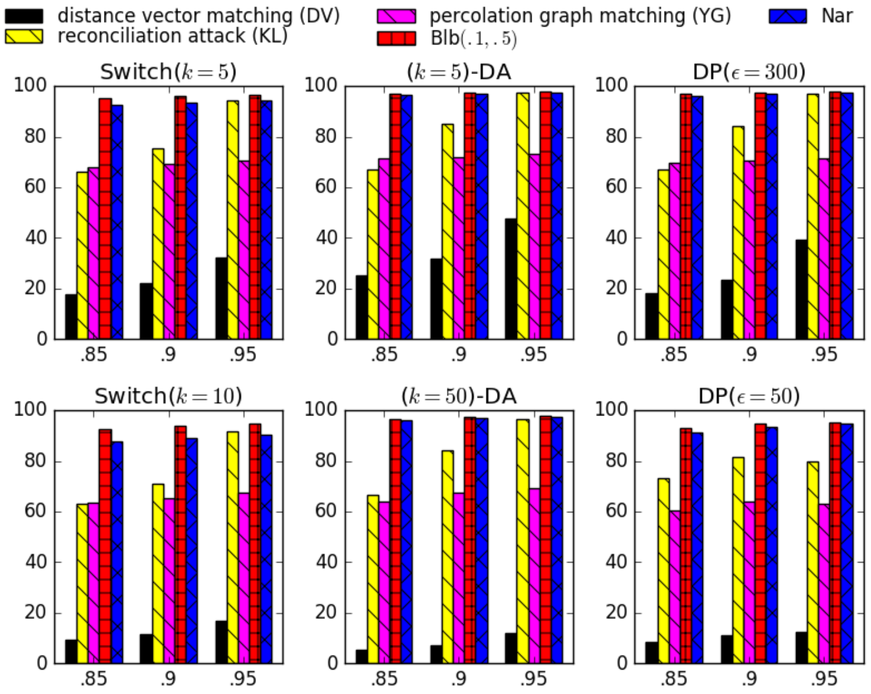
Comparison with other attacks (3)

Enron (36k)

Facebook (63k)

F1-scores

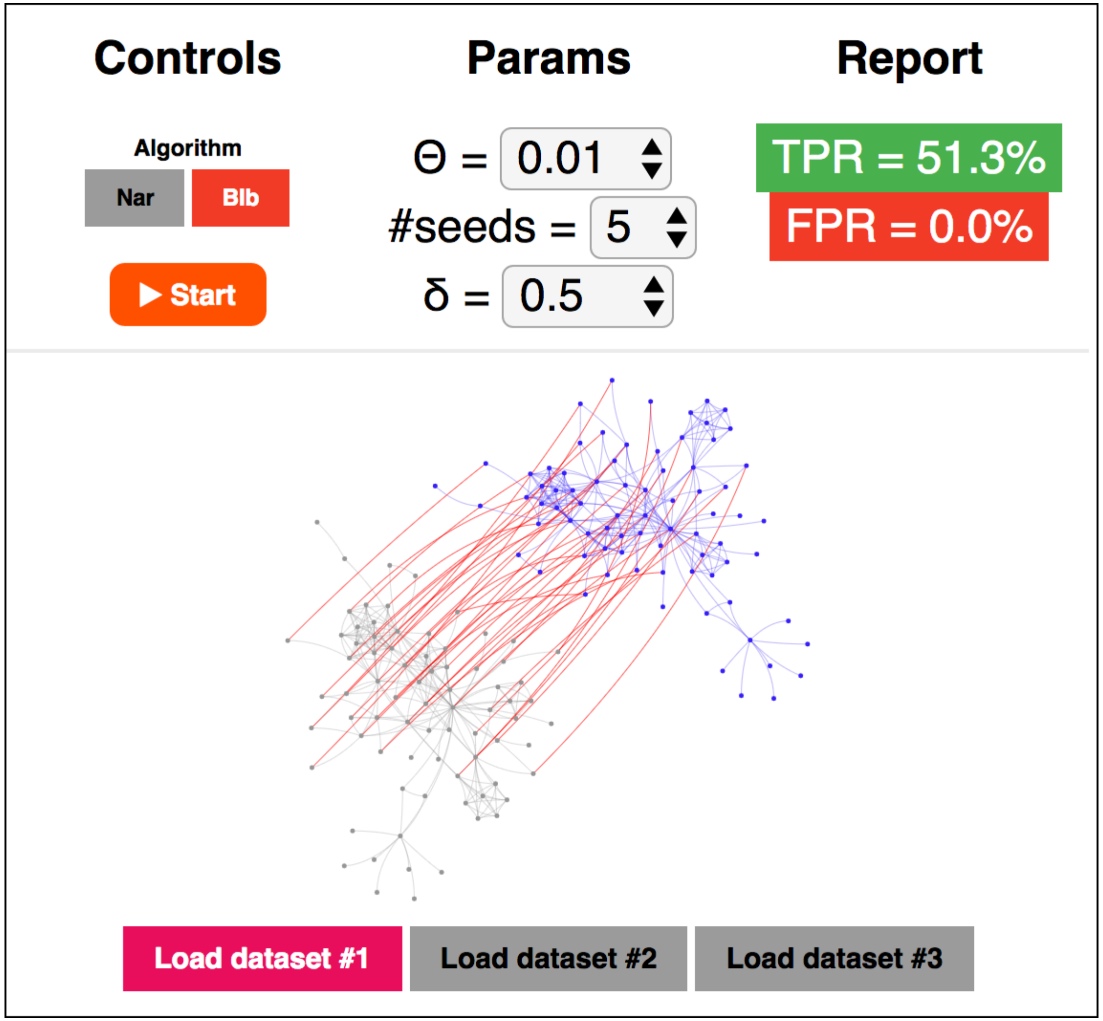




$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Demo time: try them in your browser

https://gulyas.info/snda?tldr



Θ = 0.01

#seeds = 5

δ = 0.5

TPR = 51.3%

FPR = 0.0%

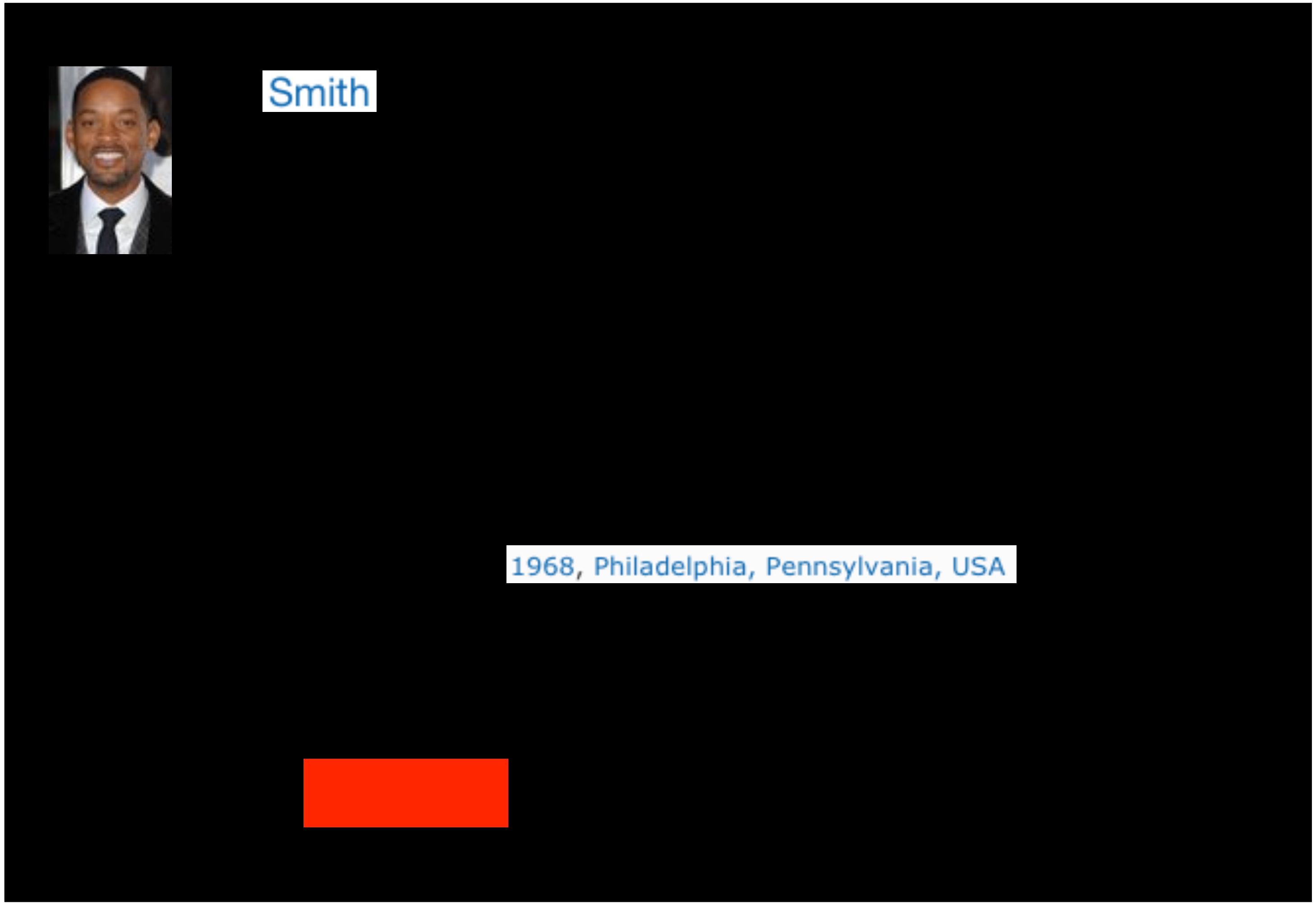
Image: The Photographer / CC-BY-SA-3.0

Joint work with **Gergely Ács, Claude Castelluccia**
[https://gulyas.info/files/publications/GulyasG_PETS16.pdf]

FINGERPRINTING ATTACKS

Limiting attribute access for protecting privacy?

Profile id: #2adc272d9



Smith

1968, Philadelphia, Pennsylvania, USA

[redacted]

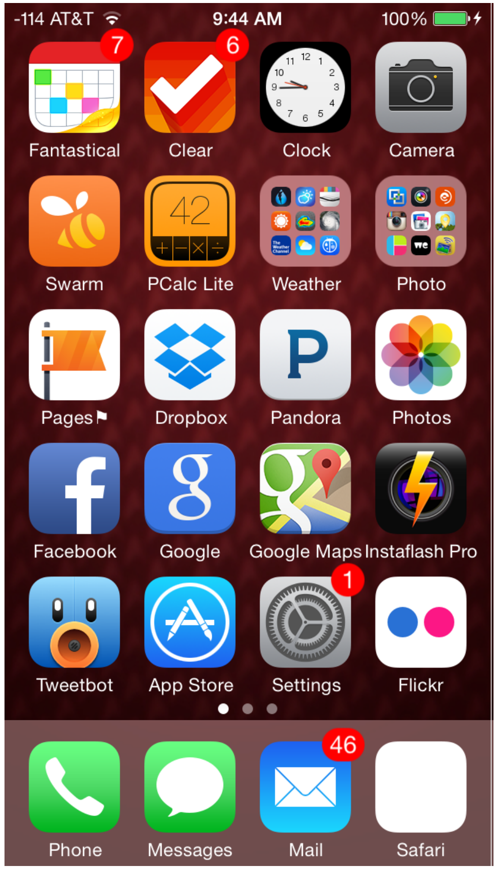
[redacted]

Fingerprinting: privacy in iOS9

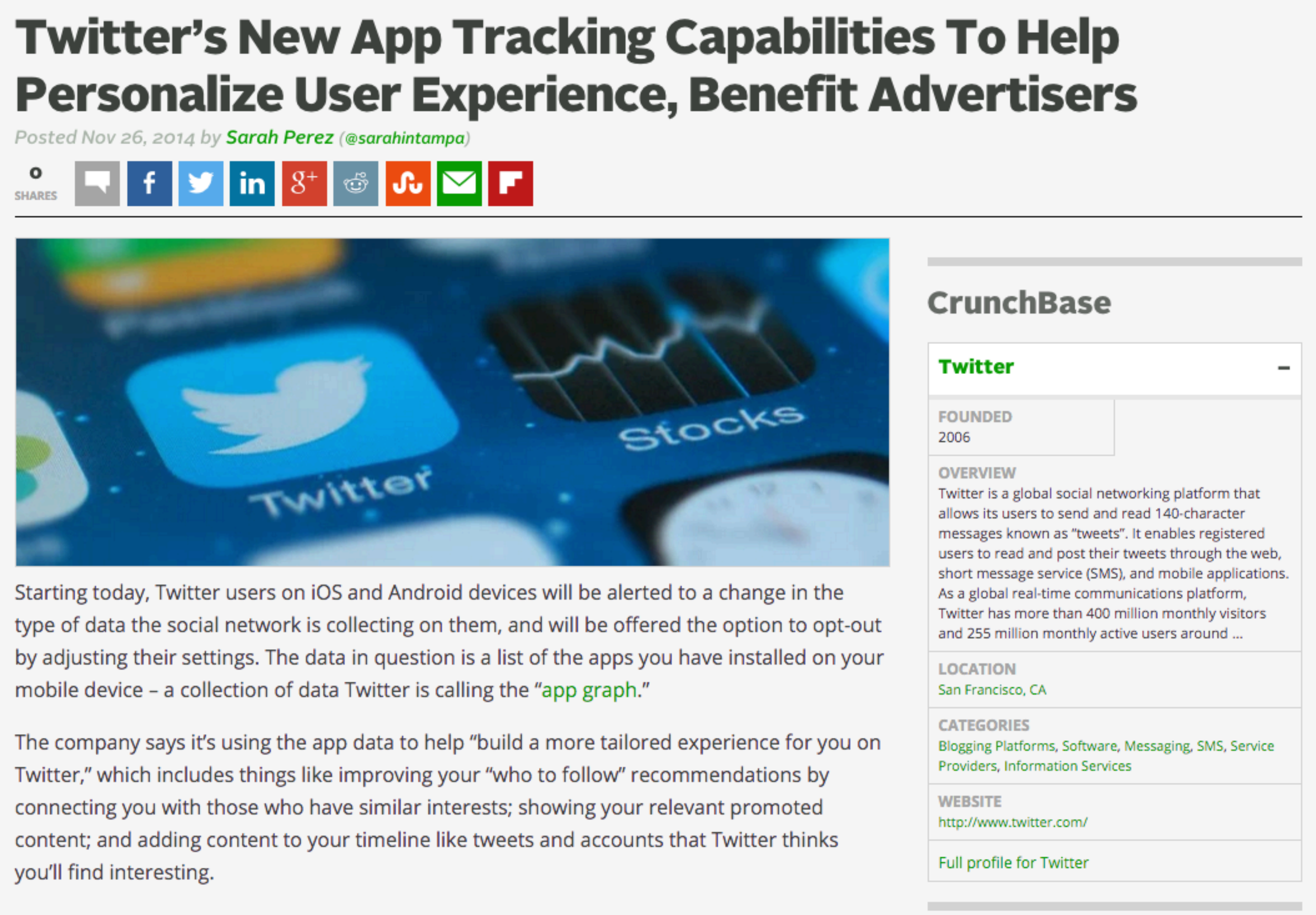
iOS 9

Tor Browser

Location dataset



Original image: Michael Lee (flickr)



Twitter's New App Tracking Capabilities To Help Personalize User Experience, Benefit Advertisers

Posted Nov 26, 2014 by Sarah Perez (@sarahintampa)

0 SHARES

Starting today, Twitter users on iOS and Android devices will be alerted to a change in the type of data the social network is collecting on them, and will be offered the option to opt-out by adjusting their settings. The data in question is a list of the apps you have installed on your mobile device – a collection of data Twitter is calling the "app graph."

The company says it's using the app data to help "build a more tailored experience for you on Twitter," which includes things like improving your "who to follow" recommendations by connecting you with those who have similar interests; showing your relevant promoted content; and adding content to your timeline like tweets and accounts that Twitter thinks you'll find interesting.

CrunchBase

Twitter –

FOUNDED
2006

OVERVIEW
Twitter is a global social networking platform that allows its users to send and read 140-character messages known as "tweets". It enables registered users to read and post their tweets through the web, short message service (SMS), and mobile applications. As a global real-time communications platform, Twitter has more than 400 million monthly visitors and 255 million monthly active users around ...

LOCATION
San Francisco, CA

CATEGORIES
Blogging Platforms, Software, Messaging, SMS, Service Providers, Information Services

WEBSITE
http://www.twitter.com/

Full profile for Twitter

New scheme on iOS 9.0

- Trade-off situation:
  - make apps unable to detect the presence of applications at large scales (e.g., for profiling)
  - but allow legitimate uses (e.g., inter-application collaboration)
- `canOpenURL()` limitations (on e.g., "fb://" or "twitter://")

	Run on iOS 8	Run on iOS 9
Linked to iOS 8	no limits	Max 50 calls (*)
Linked to iOS 9	no limits	Predefined call schemes (unlimited)
Market share (**)	11%	84%

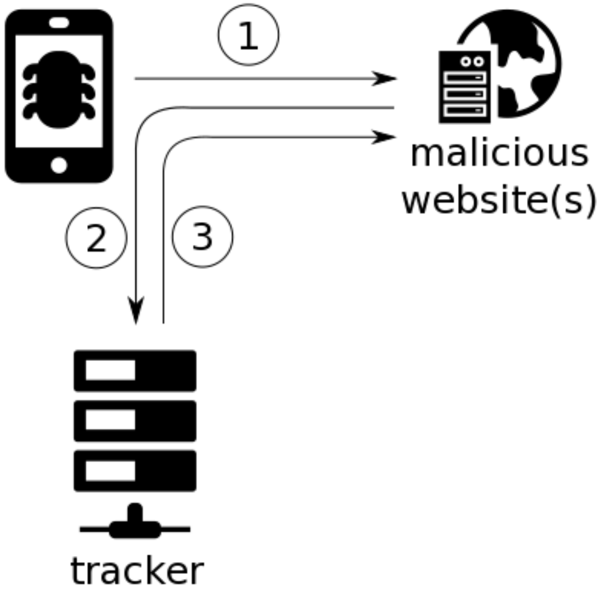
(*) Can be reset with program upgrades and re-installs
(**) As of May 9, 2016, measured by the App Store

Identification may be still possible

- Behavioral identification by applications
  (vs. random identifiers)
  - Works after re-installs
  - Same results for multiple apps
  - Not just for in-app tracking

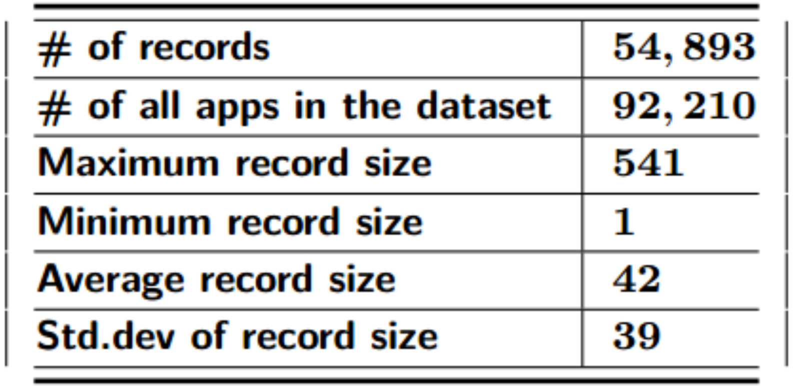
➔ Tracking

➔ Re-identification!

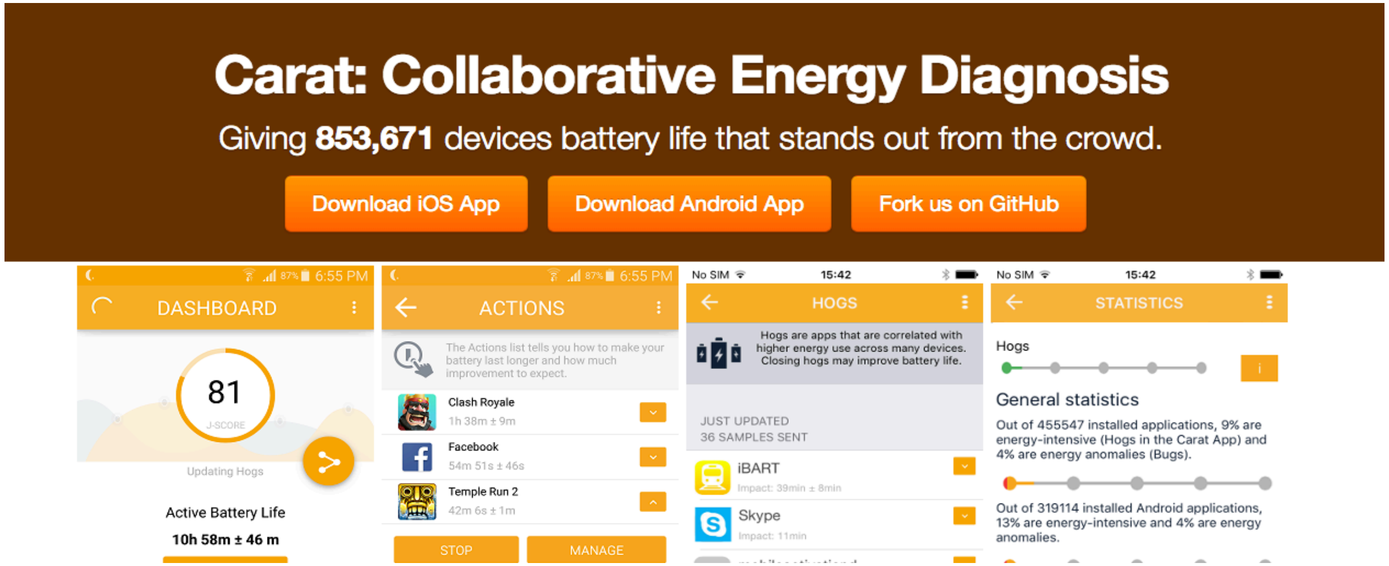


Analysis – data?

- Android apps: Carat project
  - 11/03/2013 & 15/10/2013
  - (without system apps)



# of records	54, 893
# of all apps in the dataset	92, 210
Maximum record size	541
Minimum record size	1
Average record size	42
Std.dev of record size	39



Carat: Collaborative Energy Diagnosis

Giving **853,671** devices battery life that stands out from the crowd.

The Actions list tells you how to make your battery last longer and how much improvement to expect.

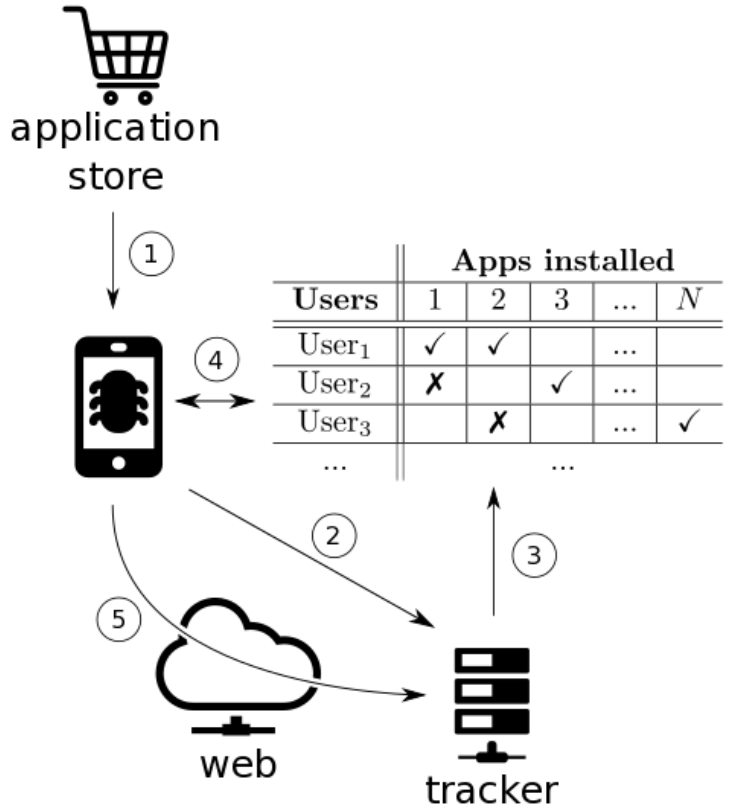
Hogs are apps that are correlated with higher energy use across many devices. Closing hogs may improve battery life.

Out of 455547 installed applications, 9% are energy-intensive (Hogs in the Carat App) and 4% are energy anomalies (Bugs).

Out of 319114 installed Android applications, 13% are energy-intensive and 4% are energy anomalies.

Attack schemes on identification

Targeted fingerprinting (de-anonymization)



	Apps installed				
Users	1	2	3	…	N
User_1	✓	✓		…	
User_2	✗		✓	…	
User_3		✗		…	✓
…			…		

against apps linked to iOS 8

Background knowledge:

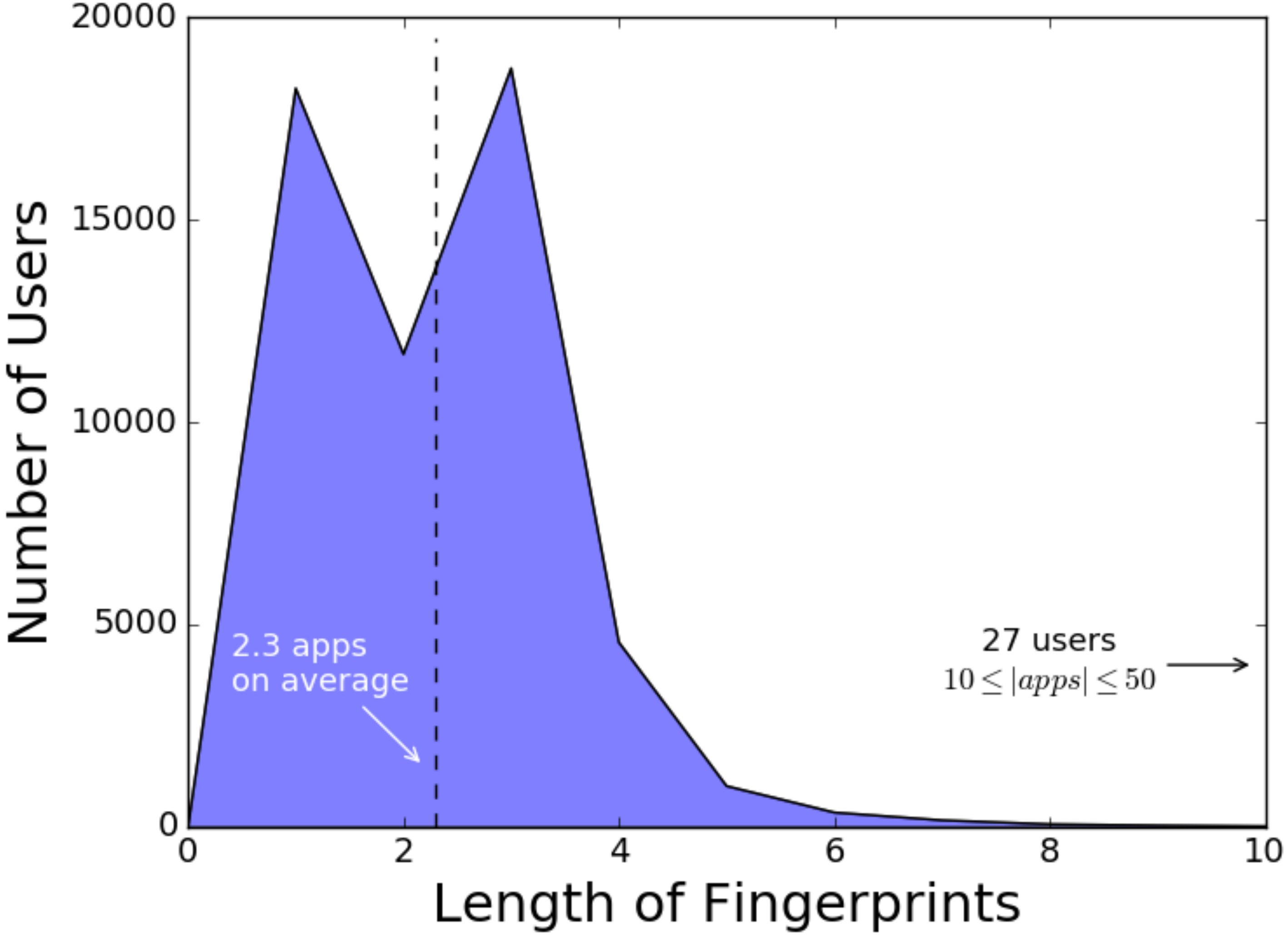
	A_1	A_2	A_3	A_4
U_1	1	0	1	1
U_2	1	1	1	1
U_3	0	1	0	1
U_4	1	0	1	0
U_5	1	1	1	0
U_6	1	1	0	0

#1	4	4	3	2	A_4!
#2	2	1	1	-	A_2!
#3	1	-	0	-	A_3!

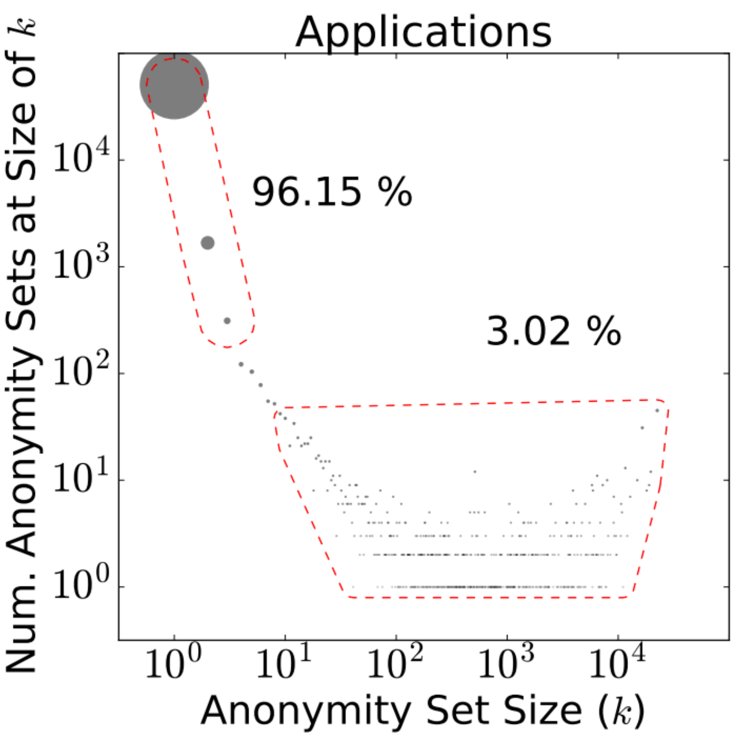
Fingerprint: A_4, A_2, A_3

Targeted fingerprinting

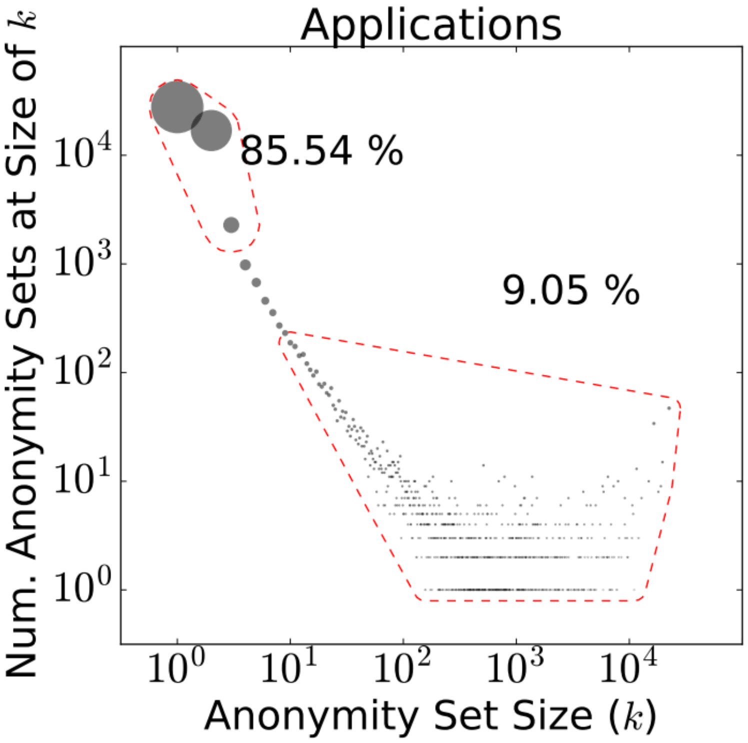
(note: limit of 50 applies)



Targeted fingerprinting (2)



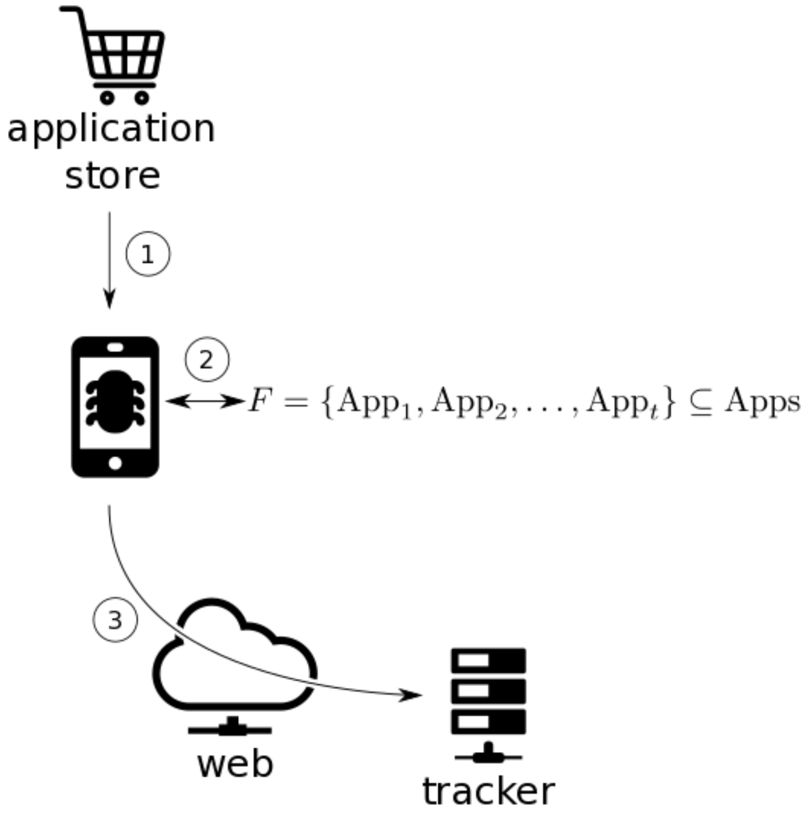
Fingerprint length: 50



Fingerprint length: 2

Attack schemes on identification (2)

General fingerprinting (linking attacks)

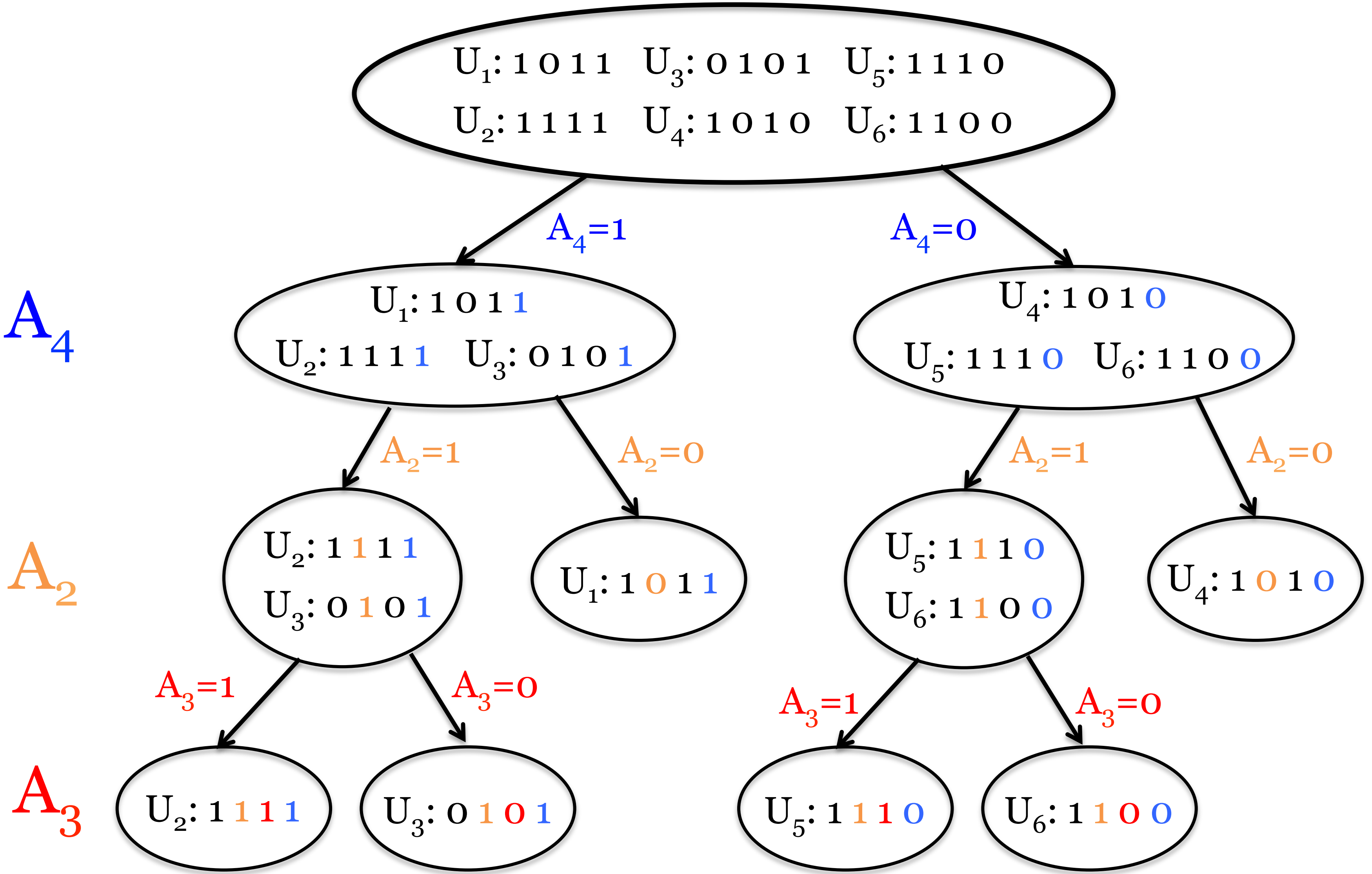

application store
1
2
$F = \{\mathrm{App}_1, \mathrm{App}_2, \ldots, \mathrm{App}_t\} \subseteq \mathrm{Apps}$
3
web
tracker

Background knowledge:

	A_1	A_2	A_3	A_4
U_1	1	0	1	1
U_2	1	1	1	1
U_3	0	1	0	1
U_4	1	0	1	0
U_5	1	1	1	0
U_6	1	1	0	0

against apps linked to iOS 9

Attack schemes on identification (4)



U_1: 1 0 1 1 U_3: 0 1 0 1 U_5: 1 1 1 0

U_2: 1 1 1 1 U_4: 1 0 1 0 U_6: 1 1 0 0

A_4=1

A_4=0

U_1: 1 0 1 1

U_2: 1 1 1 1 U_3: 0 1 0 1

U_4: 1 0 1 0

U_5: 1 1 1 0 U_6: 1 1 0 0

A_4

A_2=1

A_2=0

A_2=1

A_2=0

U_2: 1 1 1 1

U_3: 0 1 0 1

U_1: 1 0 1 1

U_5: 1 1 1 0

U_6: 1 1 0 0

U_4: 1 0 1 0

A_2

A_3=1

A_3=0

A_3=1

A_3=0

A_3

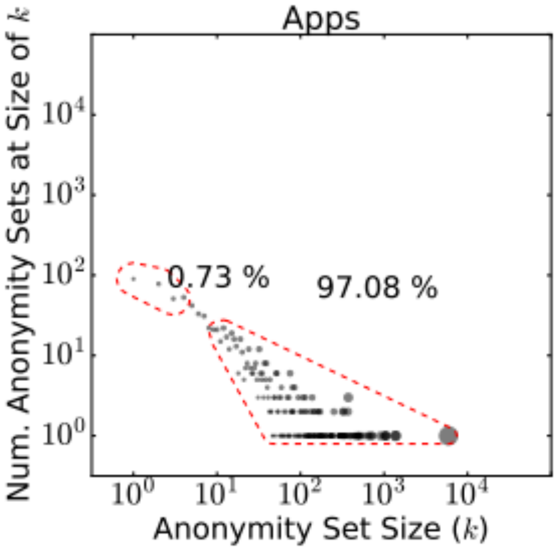
U_2: 1 1 1 1

U_3: 0 1 0 1

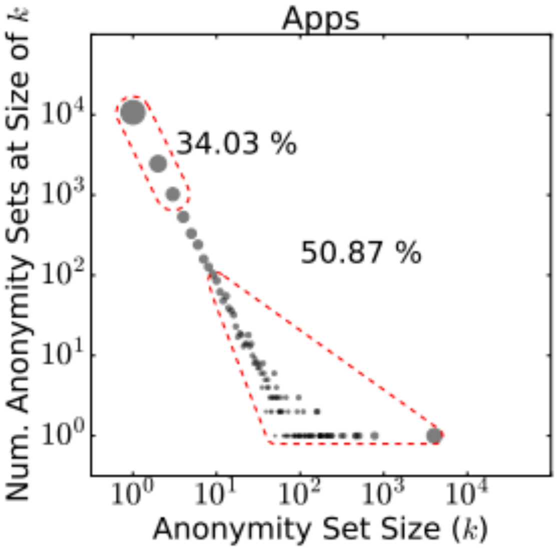
U_5: 1 1 1 0

U_6: 1 1 0 0

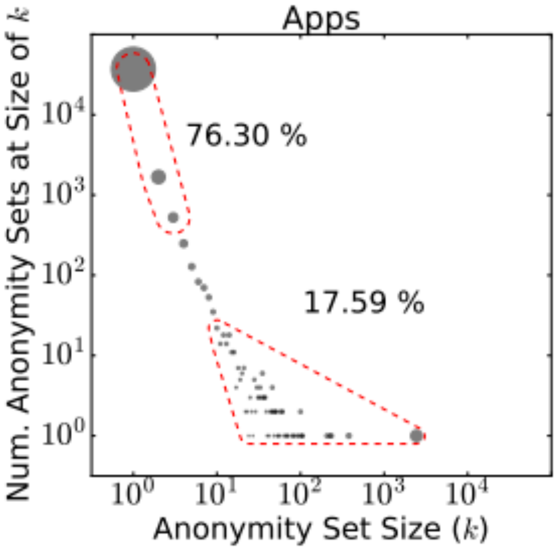
General fingerprinting



Fingerprint length: 10



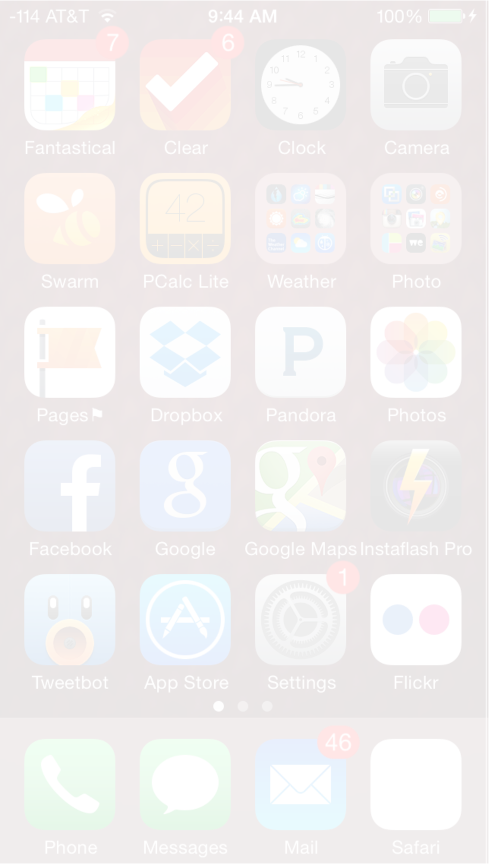
Fingerprint length: 20



Fingerprint length: 50

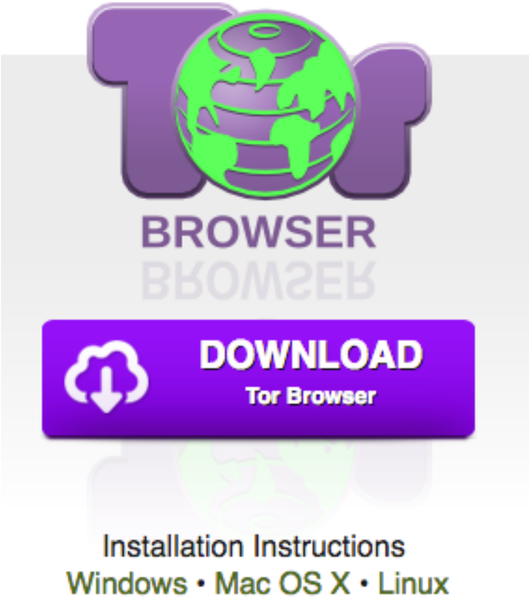
Fingerprinting: the Tor Browser

iOS 9



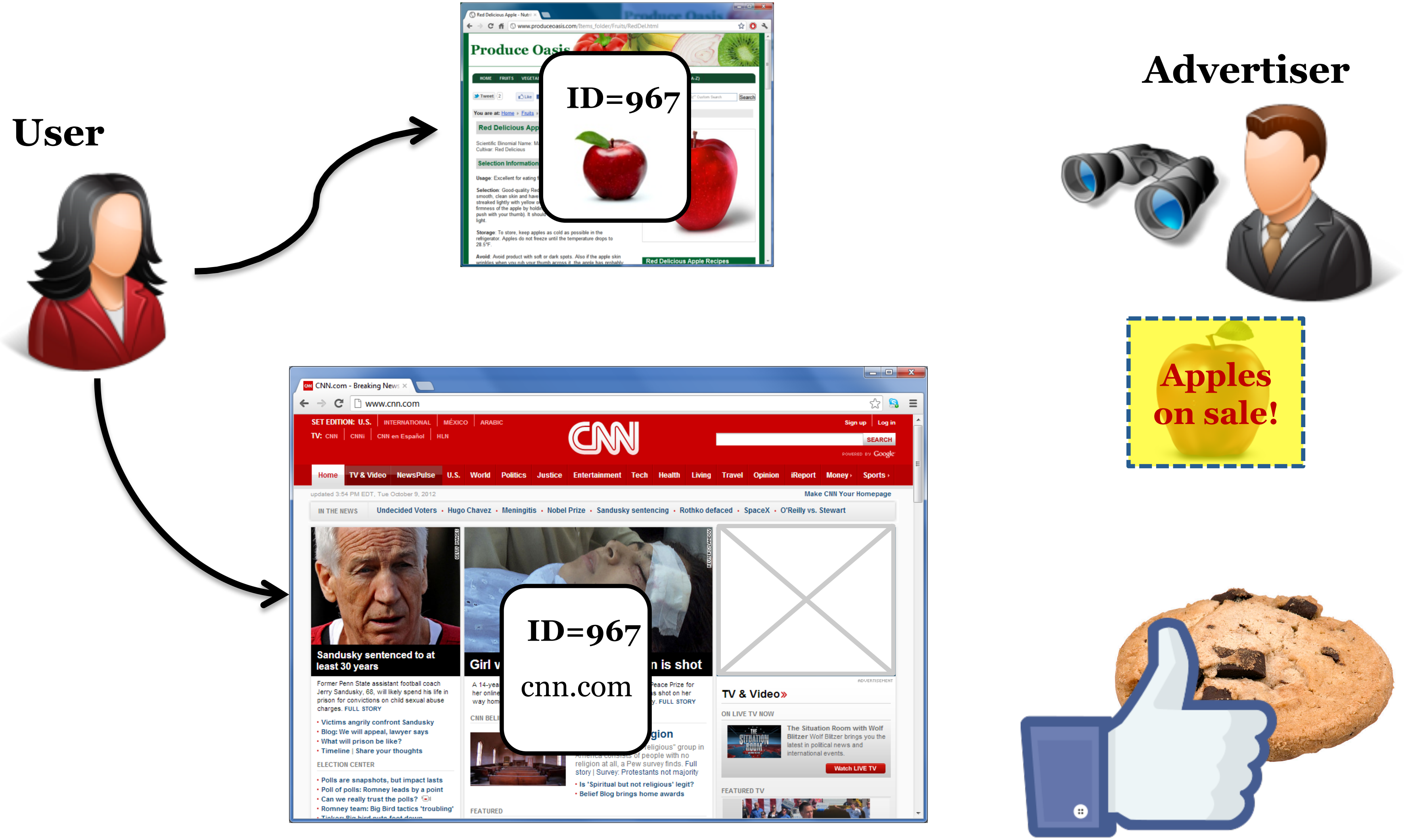
Original image: Michael Lee (flickr)

Tor Browser



Location dataset

The business model of the web



Red Delicious Apple - Nutri

www.produceoasis.com/Items_folder/Fruits/RedDel.html

Produce Oasis

HOME FRUITS VEGETA A-Z)

Tweet 2 Like Custom Search Search

You are at: Home ▸ Fruits ▸

Red Delicious App

Scientific Binomial Name: M
Cultivar: Red Delicious

Selection Information

Usage: Excellent for eating f

Selection: Good-quality Red smooth, clean skin and have streaked lightly with yellow o firmness of the apple by holdin push with your thumb). It should light.

Storage: To store, keep apples as cold as possible in the refrigerator. Apples do not freeze until the temperature drops to 28.5°F.

Avoid: Avoid product with soft or dark spots. Also if the apple skin wrinkles when you rub your thumb across it, the apple has probably

Red Delicious Apple Recipes

CNN.com - Breaking News

www.cnn.com

SET EDITION: U.S. | INTERNATIONAL | MÉXICO | ARABIC

TV: CNN | CNNi | CNN en Español | HLN

Sign up | Log in

SEARCH

POWERED BY Google

Home | TV & Video | NewsPulse | U.S. | World | Politics | Justice | Entertainment | Tech | Health | Living | Travel | Opinion | iReport | Money › | Sports ›

updated 3:54 PM EDT, Tue October 9, 2012

Make CNN Your Homepage

IN THE NEWS Undecided Voters • Hugo Chavez • Meningitis • Nobel Prize • Sandusky sentencing • Rothko defaced • SpaceX • O'Reilly vs. Stewart

Sandusky sentenced to at least 30 years

Former Penn State assistant football coach Jerry Sandusky, 68, will likely spend his life in prison for convictions on child sexual abuse charges. FULL STORY

- Victims angrily confront Sandusky
- Blog: We will appeal, lawyer says
- What will prison be like?
- Timeline | Share your thoughts

ELECTION CENTER

- Polls are snapshots, but impact lasts
- Poll of polls: Romney leads by a point
- Can we really trust the polls?
- Romney team: Big Bird tactics 'troubling'

Girl v n is shot

A 14-yea Peace Prize for her online s shot on her way hom y. FULL STORY

CNN BELI

gion

religious" group in America consists of people with no religion at all, a Pew survey finds. Full story | Survey: Protestants not majority

- Is 'Spiritual but not religious' legit?
- Belief Blog brings home awards

FEATURED

ADVERTISEMENT

TV & Video»

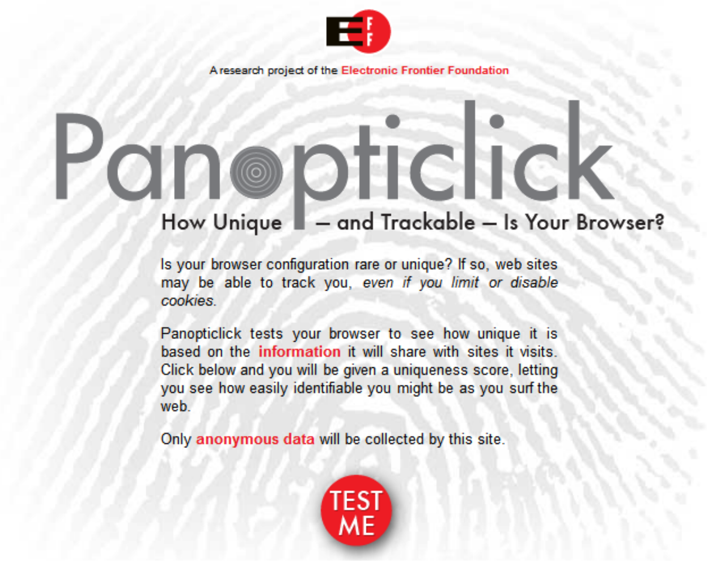
ON LIVE TV NOW

The Situation Room with Wolf Blitzer Wolf Blitzer brings you the latest in political news and international events.

Watch LIVE TV

FEATURED TV

Browser fingerprinting appears (2010-2012)



A research project of the Electronic Frontier Foundation

Panopticlick

How Unique — and Trackable — Is Your Browser?

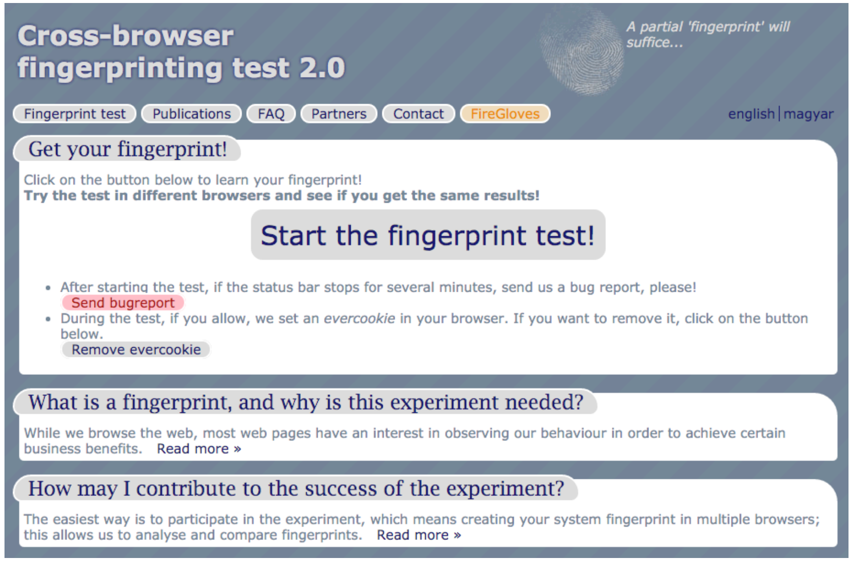
Is your browser configuration rare or unique? If so, web sites may be able to track you, *even if you limit or disable cookies.*

Panopticlick tests your browser to see how unique it is based on the information it will share with sites it visits. Click below and you will be given a uniqueness score, letting you see how easily identifiable you might be as you surf the web.

Only anonymous data will be collected by this site.

TEST ME

http://panopticlick.eff.org



Cross-browser fingerprinting test 2.0

A partial 'fingerprint' will suffice...

Get your fingerprint!

Click on the button below to learn your fingerprint!
Try the test in different browsers and see if you get the same results!

Start the fingerprint test!

- After starting the test, if the status bar stops for several minutes, send us a bug report, please!
  Send bugreport
- During the test, if you allow, we set an *evercookie* in your browser. If you want to remove it, click on the button below.
  Remove evercookie

What is a fingerprint, and why is this experiment needed?

While we browse the web, most web pages have an interest in observing our behaviour in order to achieve certain business benefits. Read more »

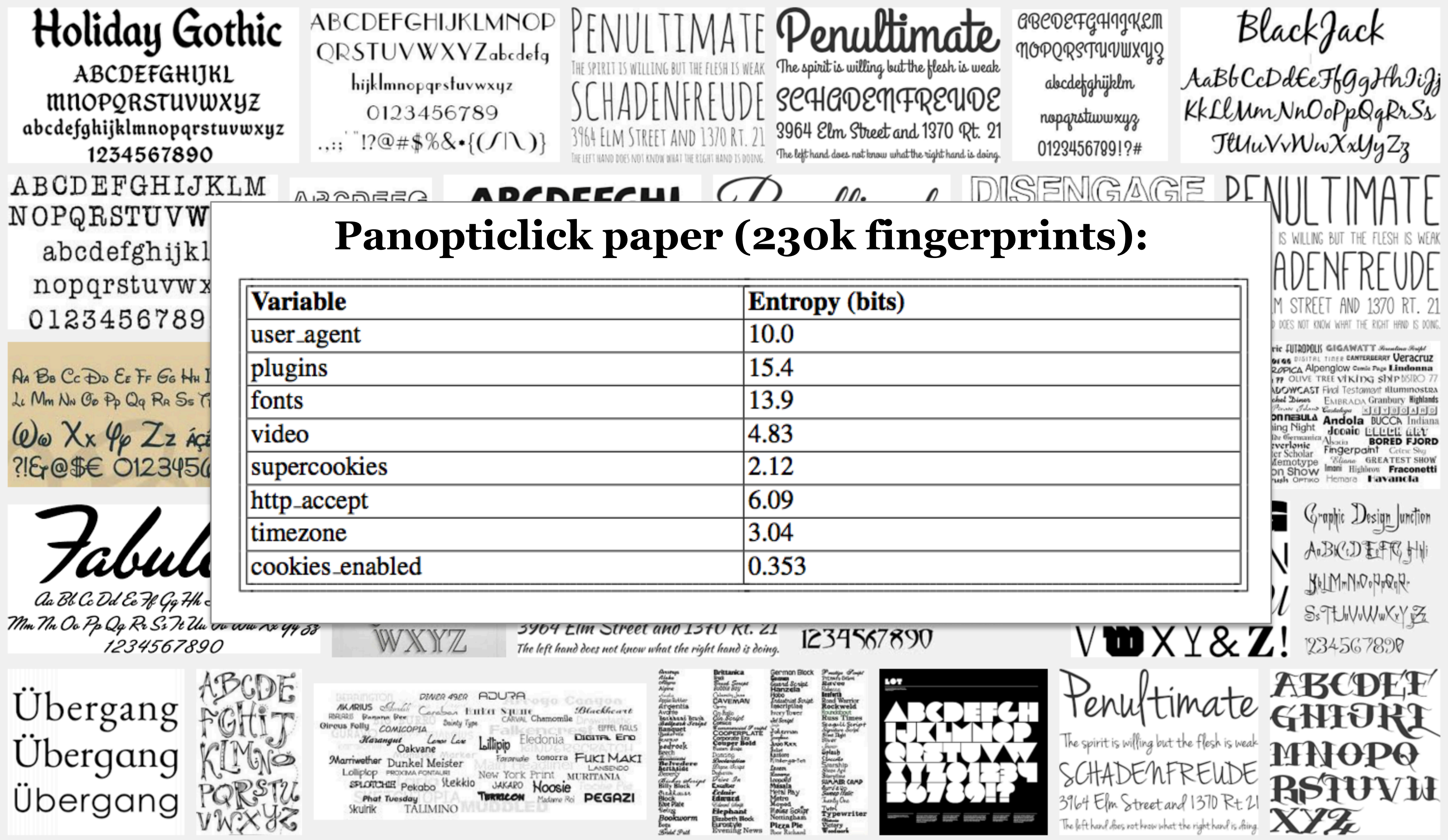
How may I contribute to the success of the experiment?

The easiest way is to participate in the experiment, which means creating your system fingerprint in multiple browsers; this allows us to analyse and compare fingerprints. Read more »

https://fingerprint.pet-portal.eu

- Browser fingerprint
  - Flash/Java required (for 95% uniqueness)
  - Browser dependent
- Cross-browser fingp.
  - Device fingerprint
  - No plugins, just JS
  - Concept appears later in the wild

Browser fingerprinting – a crucial ingredient



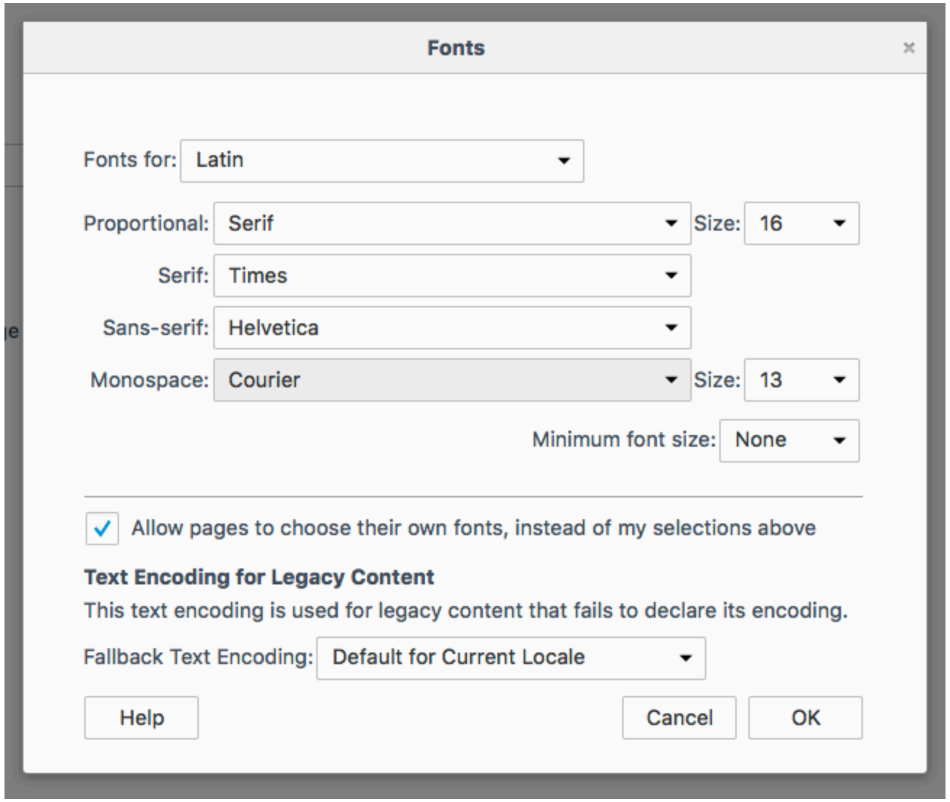
Panopticlick paper (230k fingerprints):

Variable	Entropy (bits)
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

TOR Browser: blocks font querying…

<v5.5

- Firefox: binary protection system
- TOR Browser with custom limits on
  - number of avail. fonts
  - load attempts
- `about:config`



Fonts

Fonts for: Latin

Proportional: Serif Size: 16

Serif: Times

Sans-serif: Helvetica

Monospace: Courier Size: 13

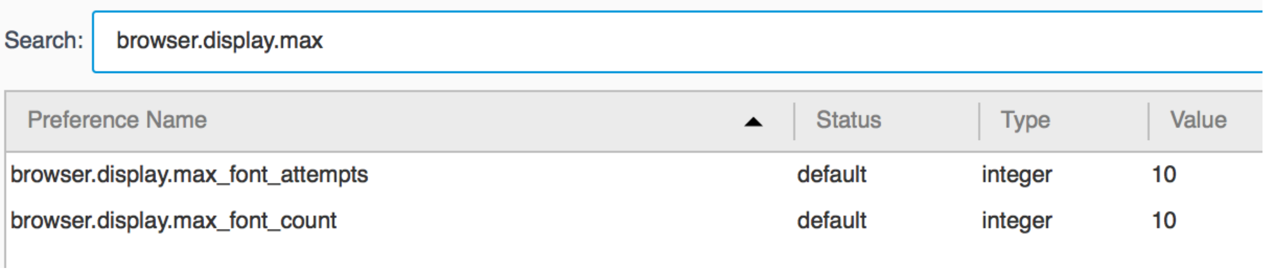
Minimum font size: None

☑ Allow pages to choose their own fonts, instead of my selections above

Text Encoding for Legacy Content

This text encoding is used for legacy content that fails to declare its encoding.

Fallback Text Encoding: Default for Current Locale



Search: browser.display.max

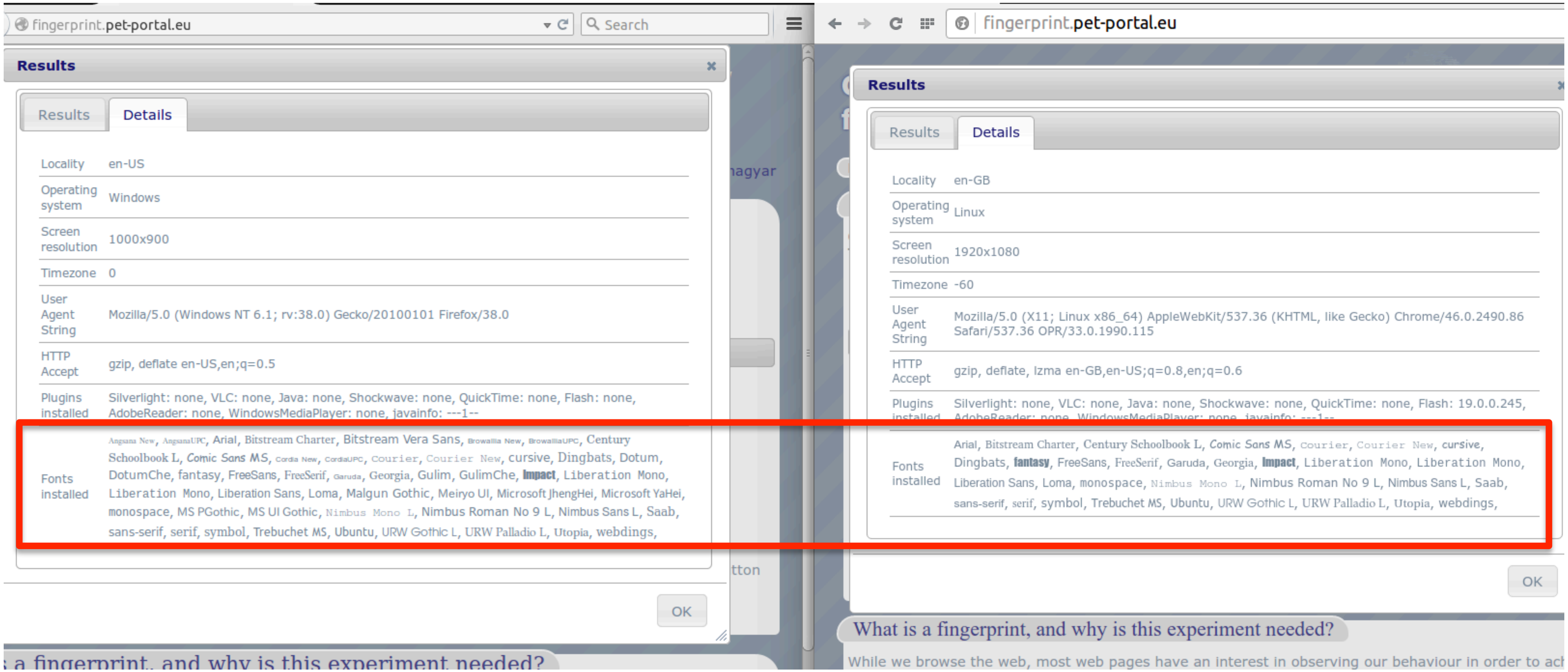
Preference Name	Status	Type	Value
browser.display.max_font_attempts	default	integer	10
browser.display.max_font_count	default	integer	10

… when it works ☺

<v5.5

TOR Browser

regular browser



Locality	en-US
Operating system	Windows
Screen resolution	1000x900
Timezone	0
User Agent String	Mozilla/5.0 (Windows NT 6.1; rv:38.0) Gecko/20100101 Firefox/38.0
HTTP Accept	gzip, deflate en-US,en;q=0.5
Plugins installed	Silverlight: none, VLC: none, Java: none, Shockwave: none, QuickTime: none, Flash: none, AdobeReader: none, WindowsMediaPlayer: none, javainfo: ---1--
Fonts installed	Angsana New, AngsanaUPC, Arial, Bitstream Charter, Bitstream Vera Sans, Browallia New, BrowalliaUPC, Century Schoolbook L, Comic Sans MS, Cordia New, CordiaUPC, Courier, Courier New, cursive, Dingbats, Dotum, DotumChe, fantasy, FreeSans, FreeSerif, Garuda, Georgia, Gulim, GulimChe, Impact, Liberation Mono, Liberation Mono, Liberation Sans, Loma, Malgun Gothic, Meiryo UI, Microsoft JhengHei, Microsoft YaHei, monospace, MS PGothic, MS UI Gothic, Nimbus Mono L, Nimbus Roman No 9 L, Nimbus Sans L, Saab, sans-serif, serif, symbol, Trebuchet MS, Ubuntu, URW Gothic L, URW Palladio L, Utopia, webdings,

a fingerprint, and why is this experiment needed?

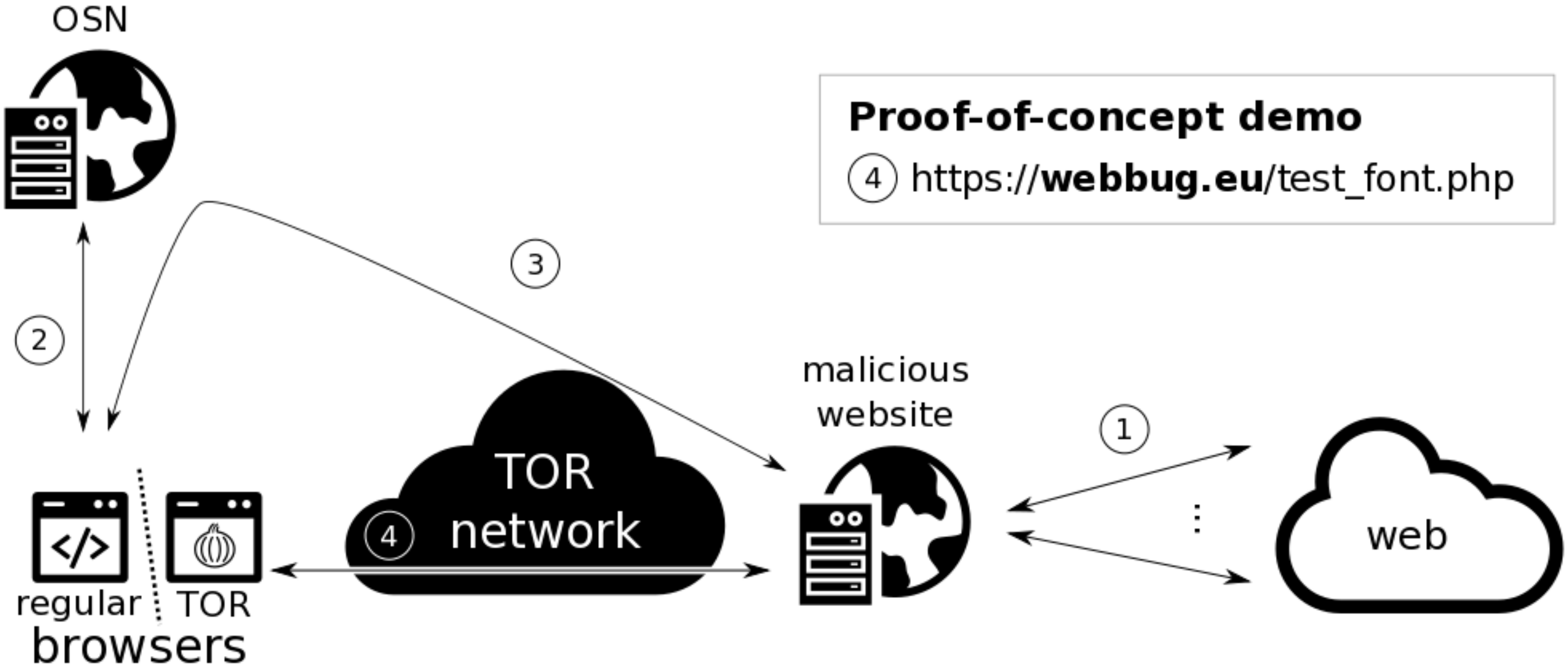
Locality	en-GB
Operating system	Linux
Screen resolution	1920x1080
Timezone	-60
User Agent String	Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/46.0.2490.86 Safari/537.36 OPR/33.0.1990.115
HTTP Accept	gzip, deflate, lzma en-GB,en-US;q=0.8,en;q=0.6
Plugins installed	Silverlight: none, VLC: none, Java: none, Shockwave: none, QuickTime: none, Flash: 19.0.0.245, AdobeReader: none, WindowsMediaPlayer: none, javainfo: ---1--
Fonts installed	Arial, Bitstream Charter, Century Schoolbook L, Comic Sans MS, Courier, Courier New, cursive, Dingbats, fantasy, FreeSans, FreeSerif, Garuda, Georgia, Impact, Liberation Mono, Liberation Mono, Liberation Sans, Loma, monospace, Nimbus Mono L, Nimbus Roman No 9 L, Nimbus Sans L, Saab, sans-serif, serif, symbol, Trebuchet MS, Ubuntu, URW Gothic L, URW Palladio L, Utopia, webdings,

What is a fingerprint, and why is this experiment needed?

While we browse the web, most web pages have an interest in observing our behaviour in order to act

➔ We found the issue: November 2015

Our attack on TOR's scheme



Proof-of-concept demo

4 https://**webbug.eu**/test_font.php

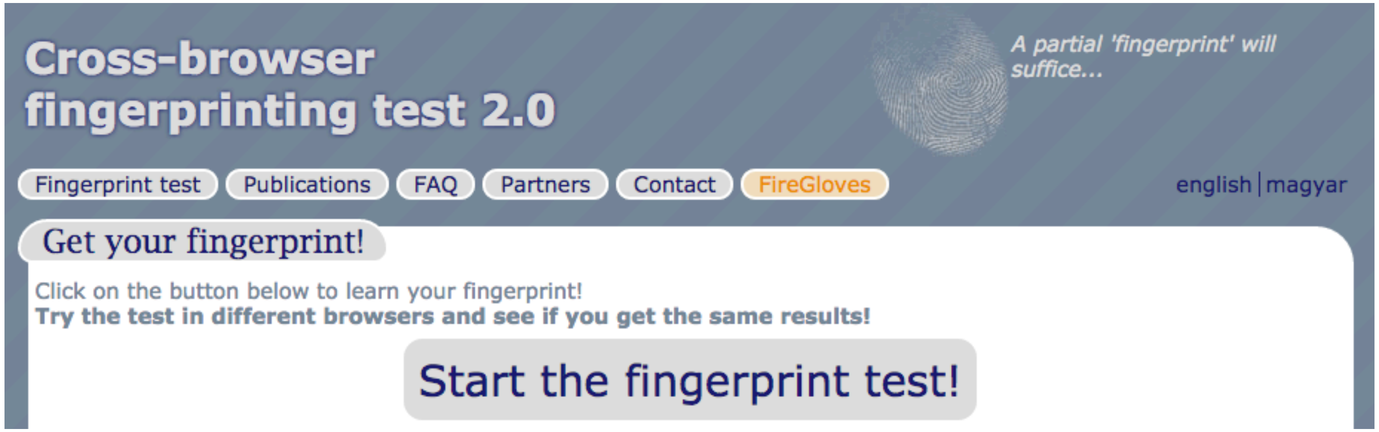
De-anonymization
(targeted fingerprinting)

U_1 fingerprint: [f93 (+), f12 (-), f67 (+)]
U_2 fingerprint: [f11 (+), f12 (+)]
…

Tracking
(general fingerprinting)

Fingerprint: [f1, f2, ..., f10]

Cleaned dataset from the cross-browser test

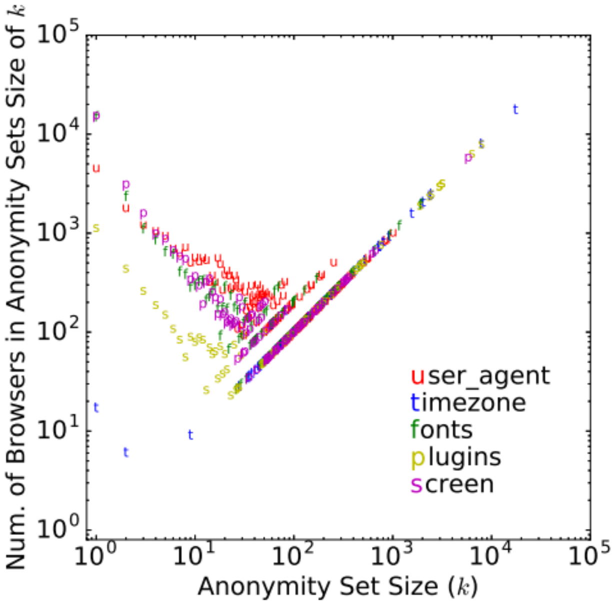


Cross-browser fingerprinting test 2.0
A partial 'fingerprint' will suffice...
Fingerprint test
Publications
FAQ
Partners
Contact
FireGloves
english | magyar

Get your fingerprint!

Click on the button below to learn your fingerprint!
Try the test in different browsers and see if you get the same results!

Start the fingerprint test!

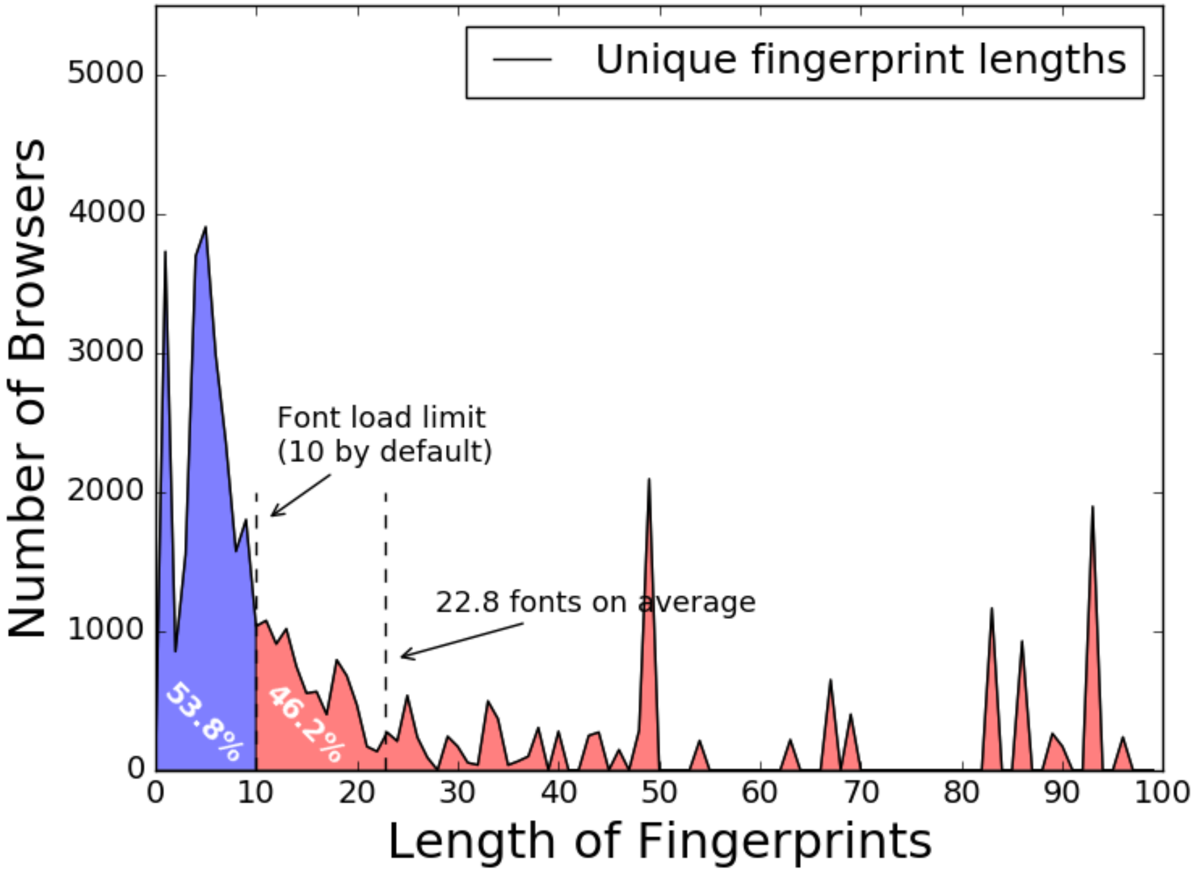


43k user fingerprints in total

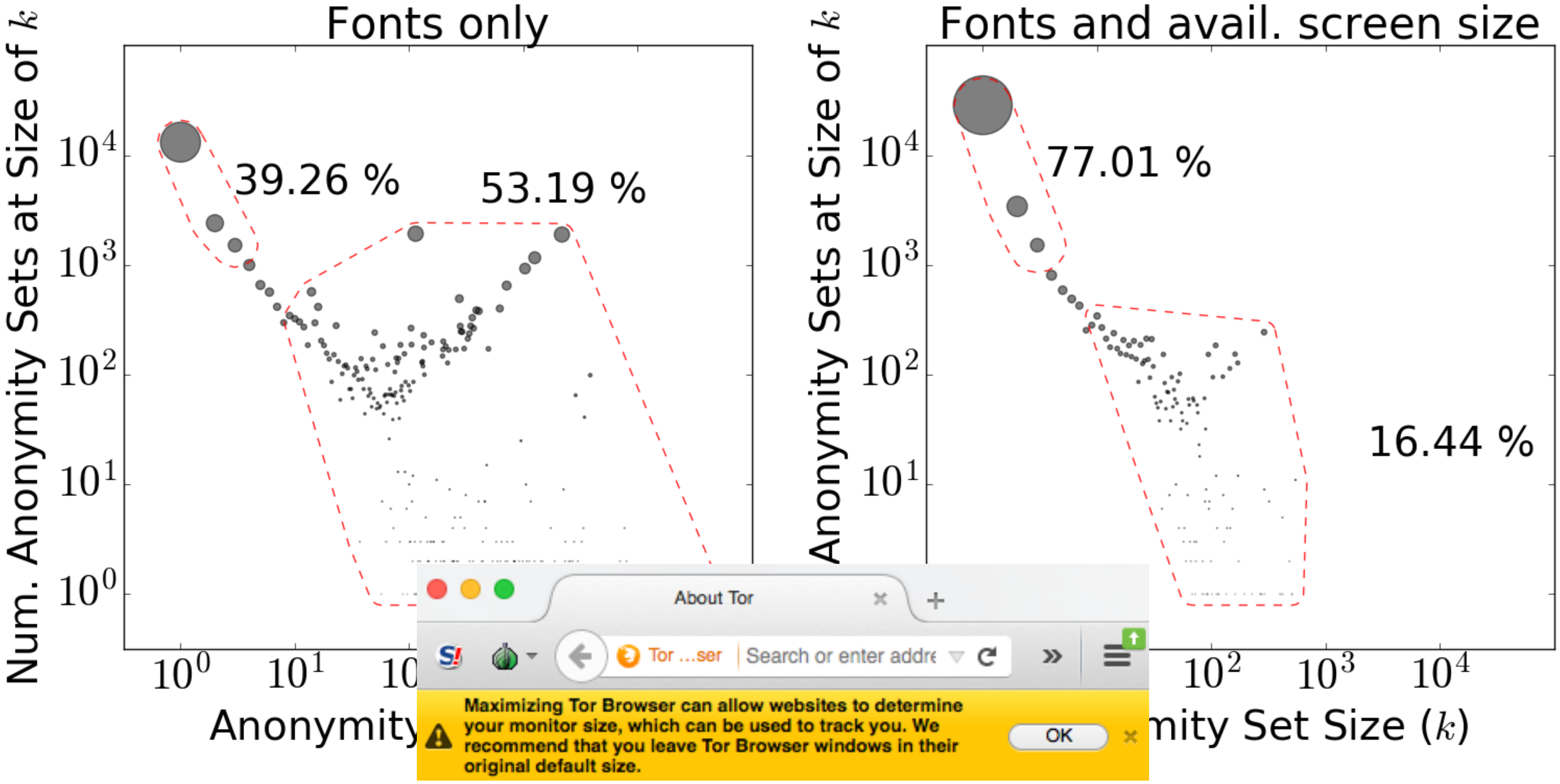
	Panopticlick	current
User agent string	10.0	7.18
Timezone	3.04	2.23
All fonts	13.9	7.79
Plugins	15.4	7.91
Screen	4.83	3.34

Targeted fingerprinting

- Fingerprint:
  - shortest (greedy) list of most distinguishing fonts
  - either a font installed, either another which is not

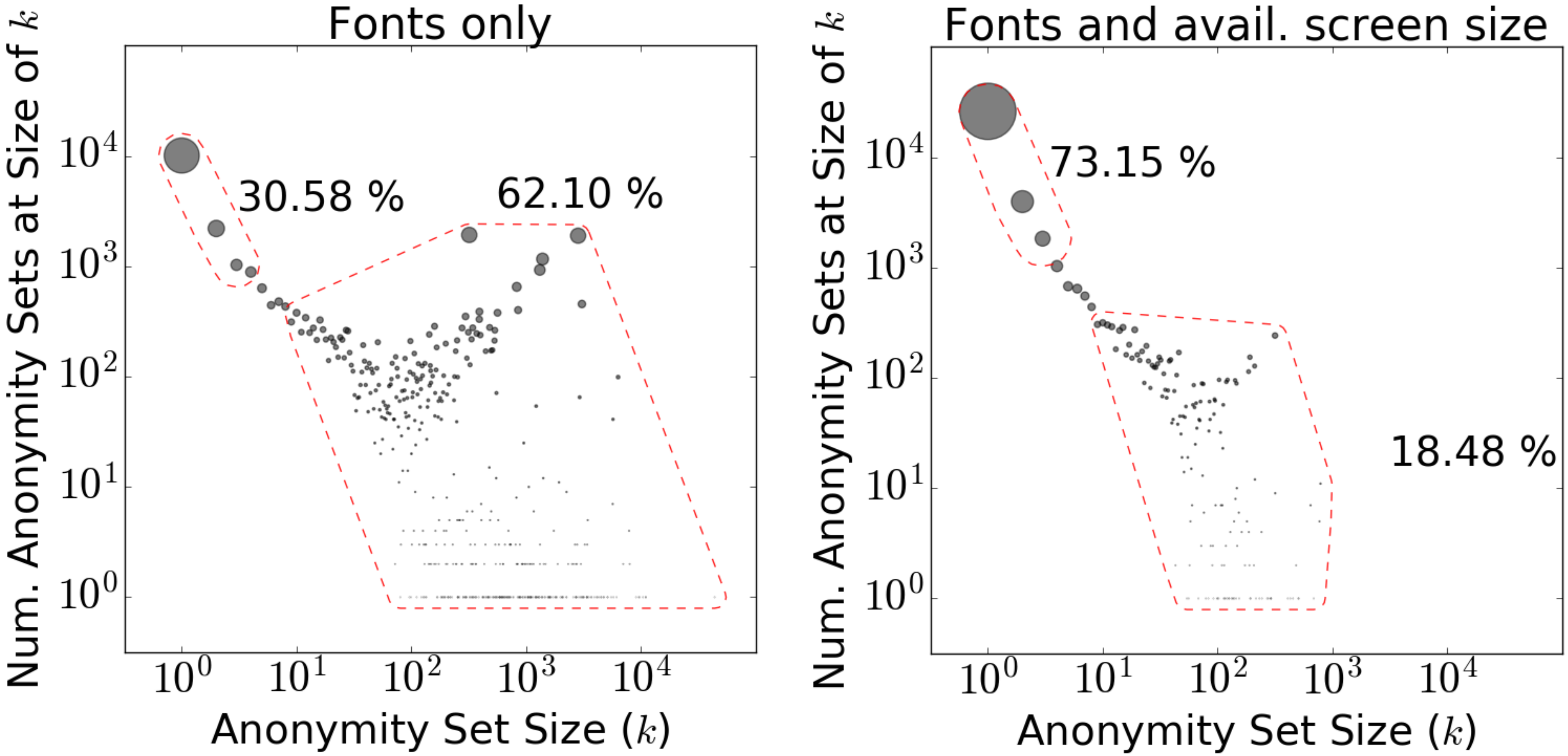


Targeted fingerprinting (2)

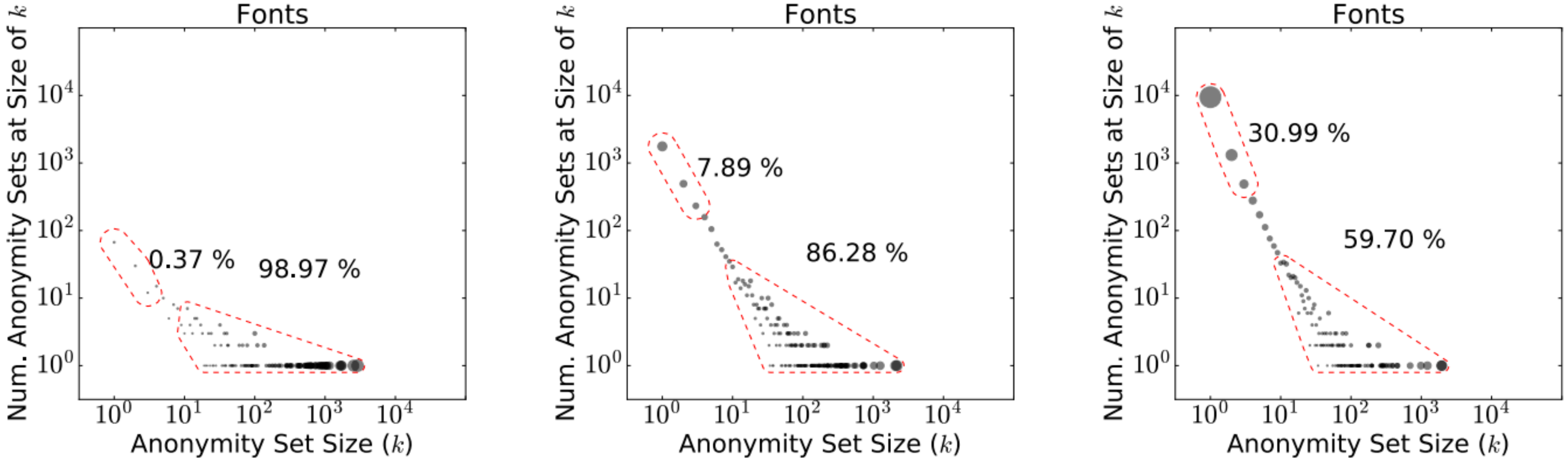


Maximizing Tor Browser can allow websites to determine your monitor size, which can be used to track you. We recommend that you leave Tor Browser windows in their original default size.

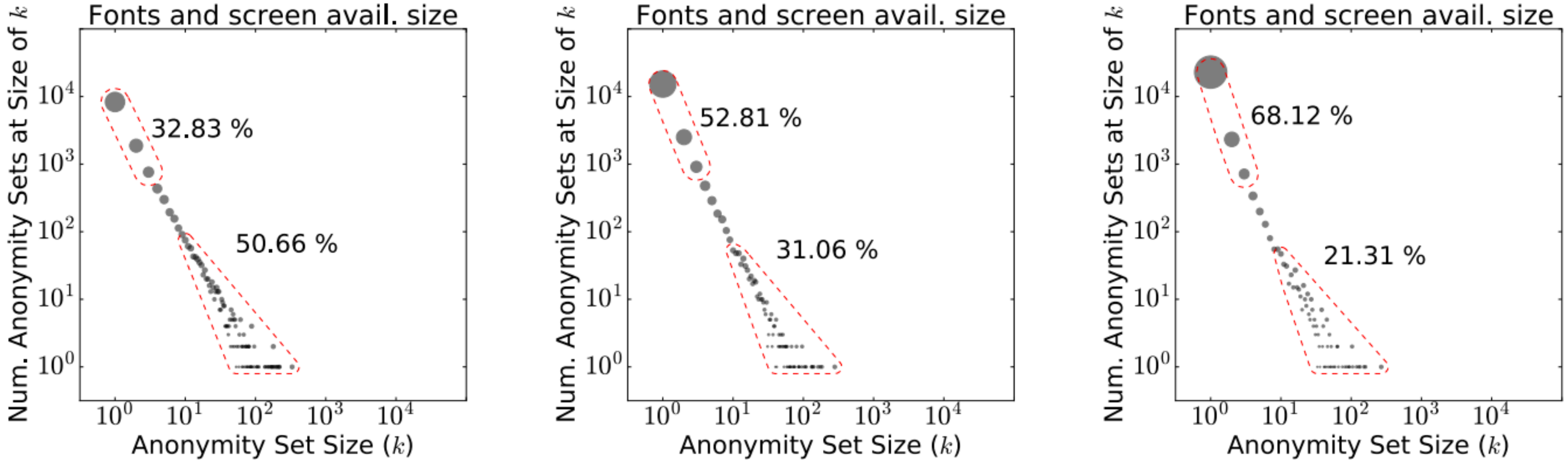
Targeted fingerprinting – 5 fonts max! (3)



General fingerprinting

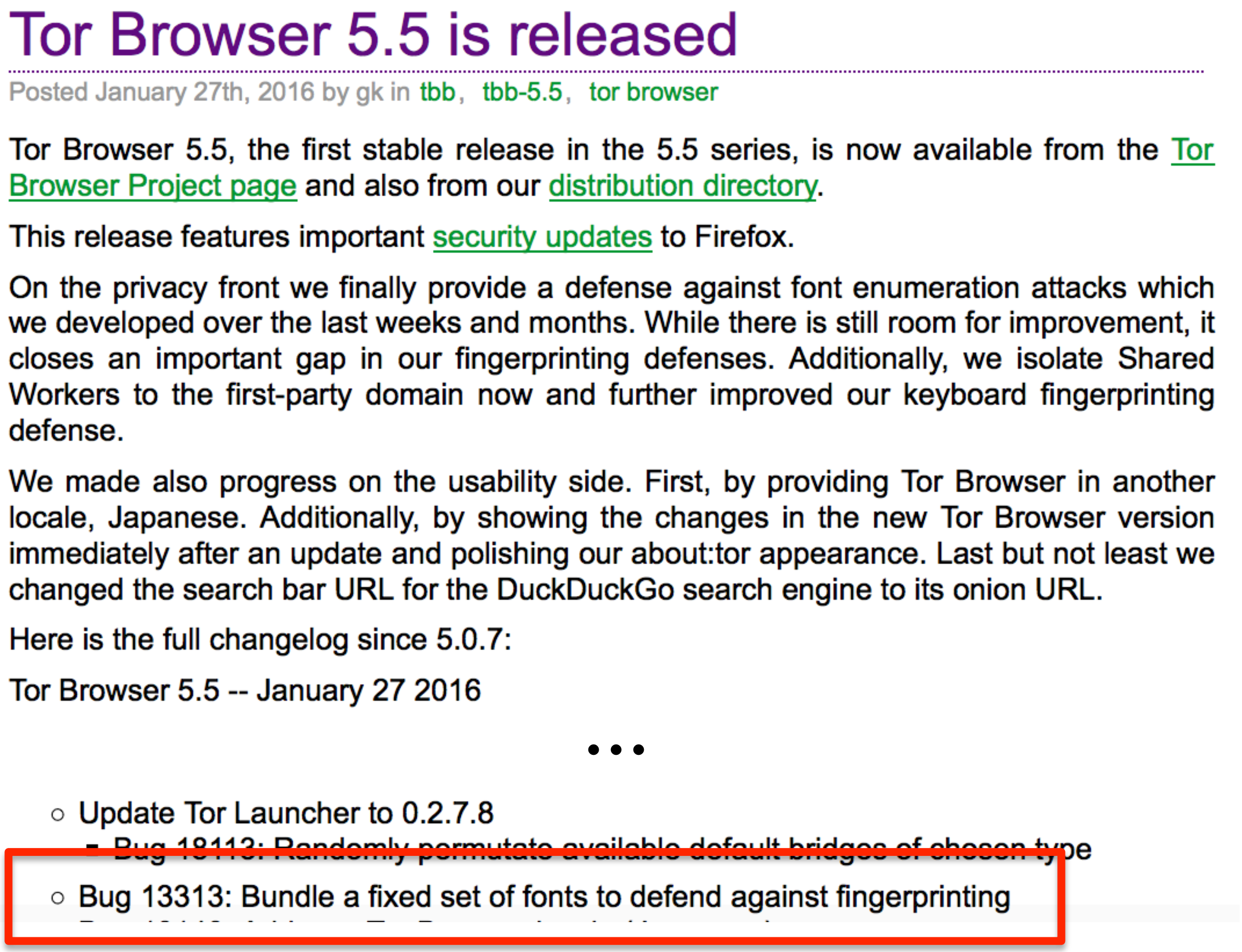


(a) $s = 10$ **(b)** $s = 25$ **(c)** $s = 100$



(a) $s = 10$ **(b)** $s = 25$ **(c)** $s = 100$

Current stats of TOR: patched



Tor Browser 5.5 is released

Posted January 27th, 2016 by gk in tbb, tbb-5.5, tor browser

Tor Browser 5.5, the first stable release in the 5.5 series, is now available from the Tor Browser Project page and also from our distribution directory.

This release features important security updates to Firefox.

On the privacy front we finally provide a defense against font enumeration attacks which we developed over the last weeks and months. While there is still room for improvement, it closes an important gap in our fingerprinting defenses. Additionally, we isolate Shared Workers to the first-party domain now and further improved our keyboard fingerprinting defense.

We made also progress on the usability side. First, by providing Tor Browser in another locale, Japanese. Additionally, by showing the changes in the new Tor Browser version immediately after an update and polishing our about:tor appearance. Last but not least we changed the search bar URL for the DuckDuckGo search engine to its onion URL.

Here is the full changelog since 5.0.7:

Tor Browser 5.5 -- January 27 2016

• • •

- ○ Update Tor Launcher to 0.2.7.8
  - Bug 18113: Randomly permutate available default bridges of chosen type
- ○ Bug 13313: Bundle a fixed set of fonts to defend against fingerprinting

➔ January 27, 2016

Conclusion

- Limiting the number of queries is a risky idea
  - As there are conceptual problems: even with low limits user privacy can be still at stake
  - Should be applied with precaution; e.g., it is better where the number of expected users is high
    - these attacks are not against the whole community (just against the sub-community visiting a site or installing an app)

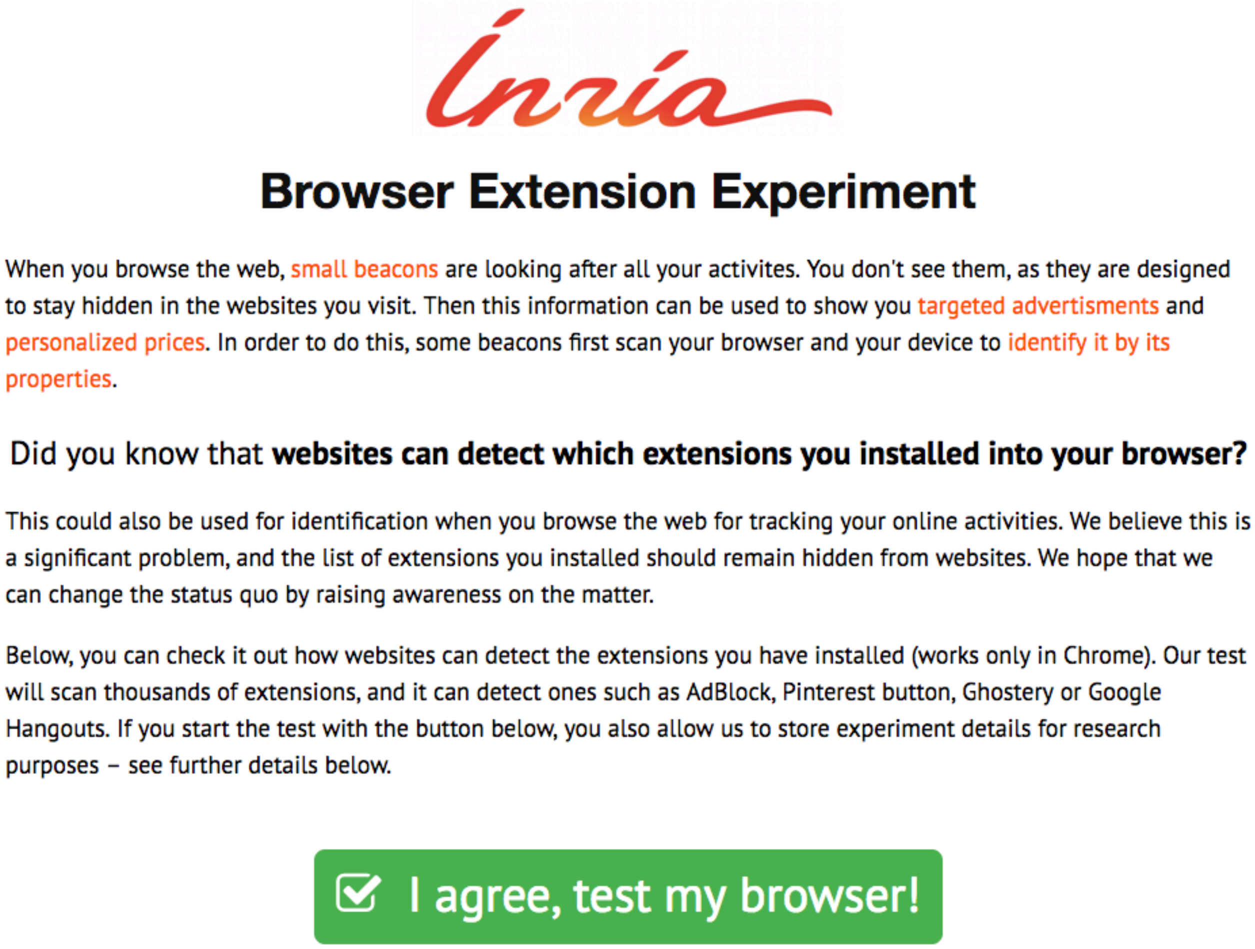
- See the paper for details and other results!

- Code:

https://github.com/gaborgulyas/constrainted_fingerprinting

Demo time: how unique are you?

https://extensions.inrialpes.fr



Browser Extension Experiment

When you browse the web, small beacons are looking after all your activites. You don't see them, as they are designed to stay hidden in the websites you visit. Then this information can be used to show you targeted advertisments and personalized prices. In order to do this, some beacons first scan your browser and your device to identify it by its properties.

Did you know that **websites can detect which extensions you installed into your browser?**

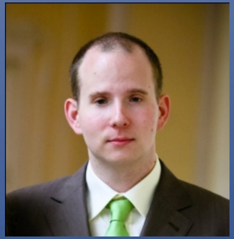
This could also be used for identification when you browse the web for tracking your online activities. We believe this is a significant problem, and the list of extensions you installed should remain hidden from websites. We hope that we can change the status quo by raising awareness on the matter.

Below, you can check it out how websites can detect the extensions you have installed (works only in Chrome). Our test will scan thousands of extensions, and it can detect ones such as AdBlock, Pinterest button, Ghostery or Google Hangouts. If you start the test with the button below, you also allow us to store experiment details for research purposes – see further details below.

Thank you for your attention!

ANY QUESTIONS?

Gábor György Gulyás
Postdoc @ Privatics
http://gulyas.info // @GulyasGG



References (in order as appeared)

- Latanya Sweeney. 2002. *k*-anonymity: a model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst*. 10, 5 (October 2002), 557-570.
- Narayanan, Arvind, and Vitaly Shmatikov. "Robust de-anonymization of large sparse datasets." *2008 IEEE Symposium on Security and Privacy (sp 2008)*. IEEE, 2008.
- Narayanan, Arvind, and Vitaly Shmatikov. "De-anonymizing social networks." *2009 30th IEEE symposium on security and privacy*. IEEE, 2009.
- G. Gy. Gulyás, S. Imre: Measuring Importance of Seeding for Structural De-anonymization Attacks in Social Networks. In Proc. of 6th IEEE International Workshop on SEcurity and SOCial Networking (SESOC) held with PerCom. 28 March. 2014.
- G. Gy. Gulyás, B. Simon, S. Imre: An Efficient and Robust Social Network De-anonymization Attack. In Proc. of the Workshop on Privacy in the Electronic Society (WPES'16), held in conjunction with the ACM CCS'16.
- G. Gy. Gulyás, G. Ács, C. Castelluccia: Near-Optimal Fingerprinting with Constraints. Proceedings on Privacy Enhancing Technologies. Volume 2016, Issue 4, July 2016