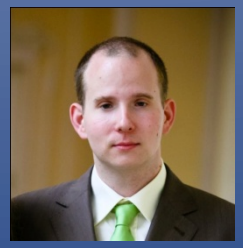
Near-Optimal Fingerprinting with Constraints

Joint work with **G. Ács**, **C. Castelluccia**

PET Symposium, 2016-07-21

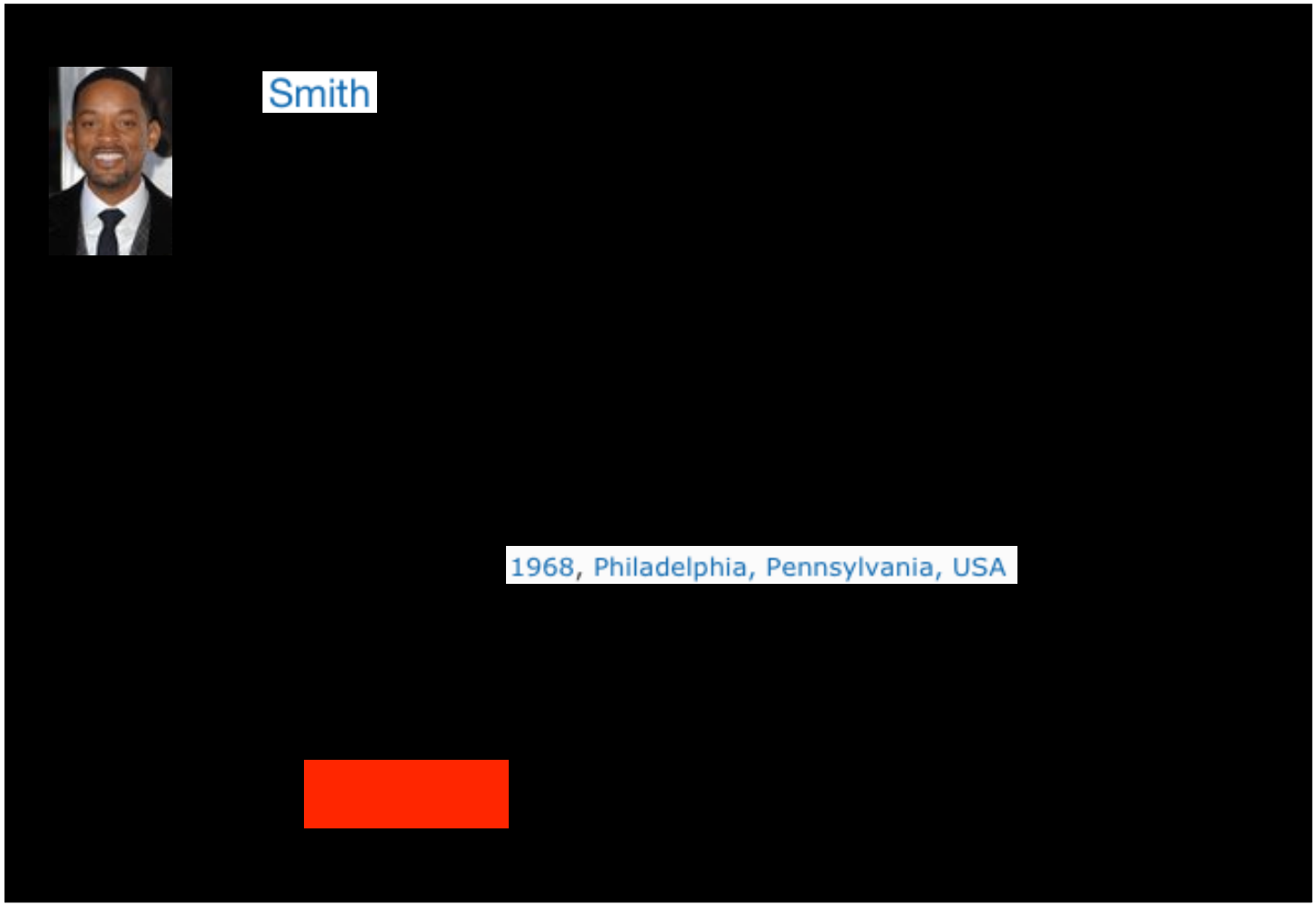
Gábor György Gulyás
Postdoc @ Privatics
http://gulyas.info // @GulyasGG



Inria
INVENTEURS DU MONDE NUMÉRIQUE

Limiting attribute access for protecting privacy?

Profile id: #2adc272d9



Discussed topics (in the paper)

iOS 9

Original image: Michael Lee (flickr)

Tor Browser

Location dataset

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%

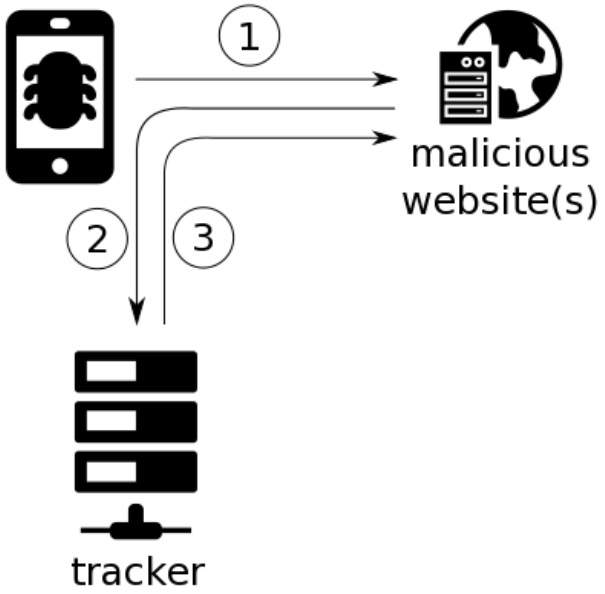
(*) Can be reset with program upgrades and re-installs
(**) As of May 9, 2016, measured by the App Store

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

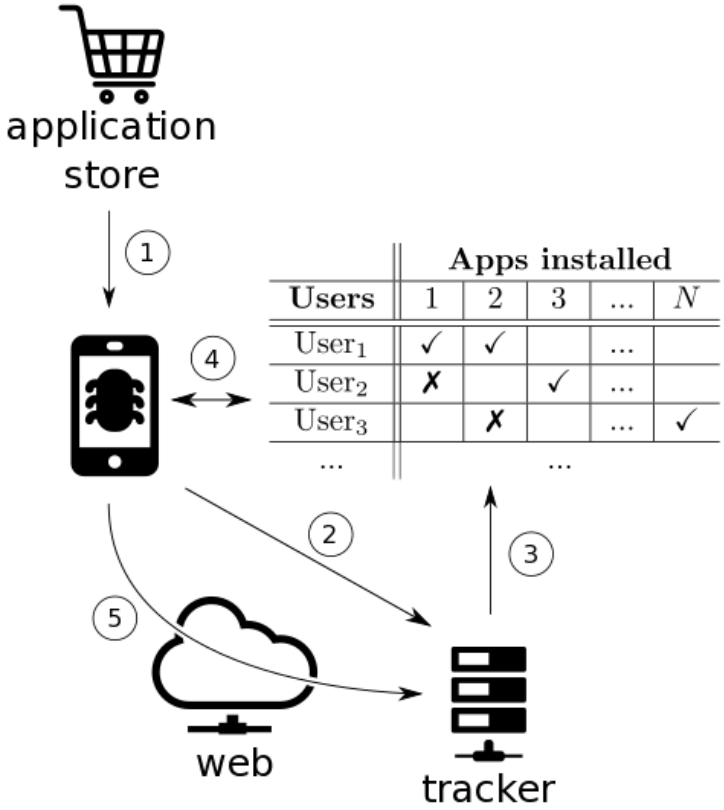
➔ Re-identification!



Have a try:
https://webbug.eu/ios_device_id/

Attack schemes on identification

Targeted fingerprinting (de-anonymization)

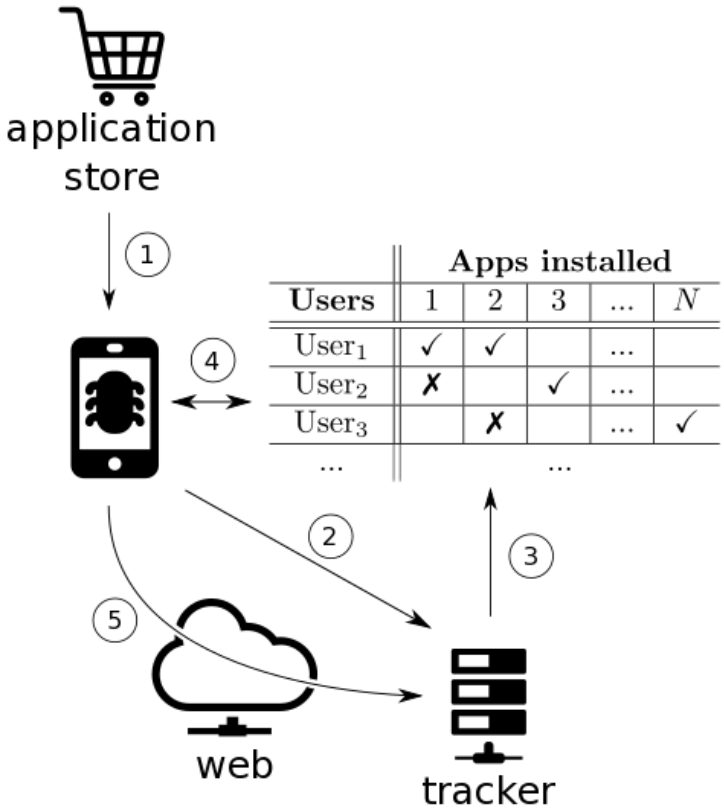


Discussed problems are NP-hard
➔ Greedy heuristics

against apps linked to iOS 8

Attack schemes on identification (2)

Targeted fingerprinting (de-anonymization)



against apps linked to iOS 8

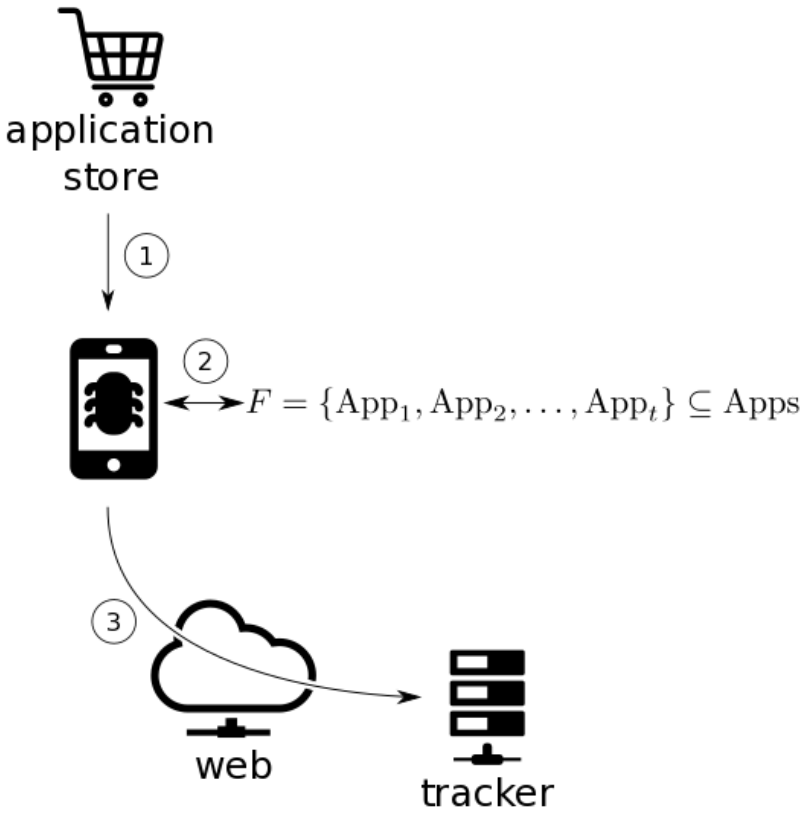
Background knowledge:

	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

Attack schemes on identification (3)

General fingerprinting (linking attacks)

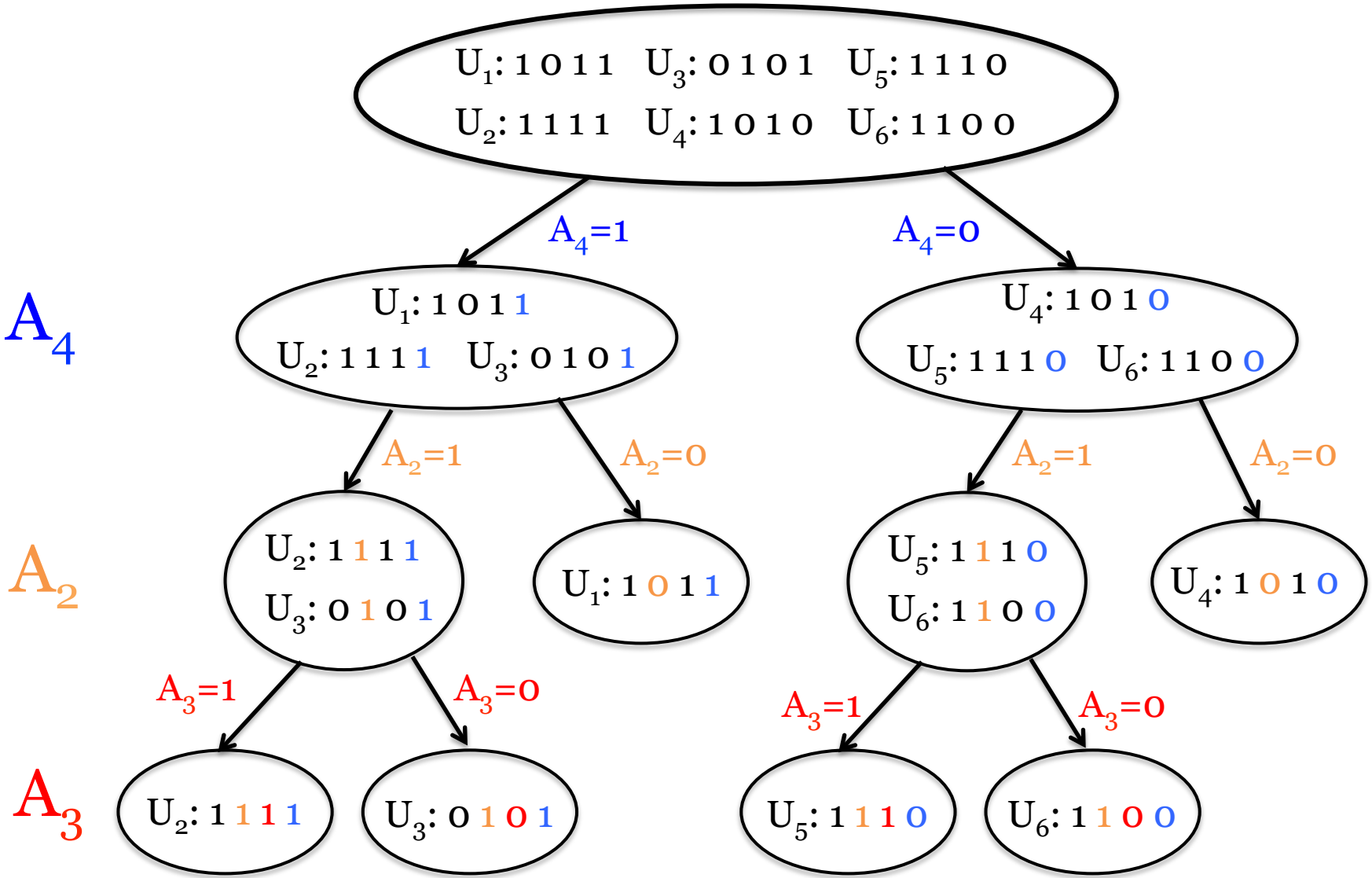


Background knowledge:

	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)



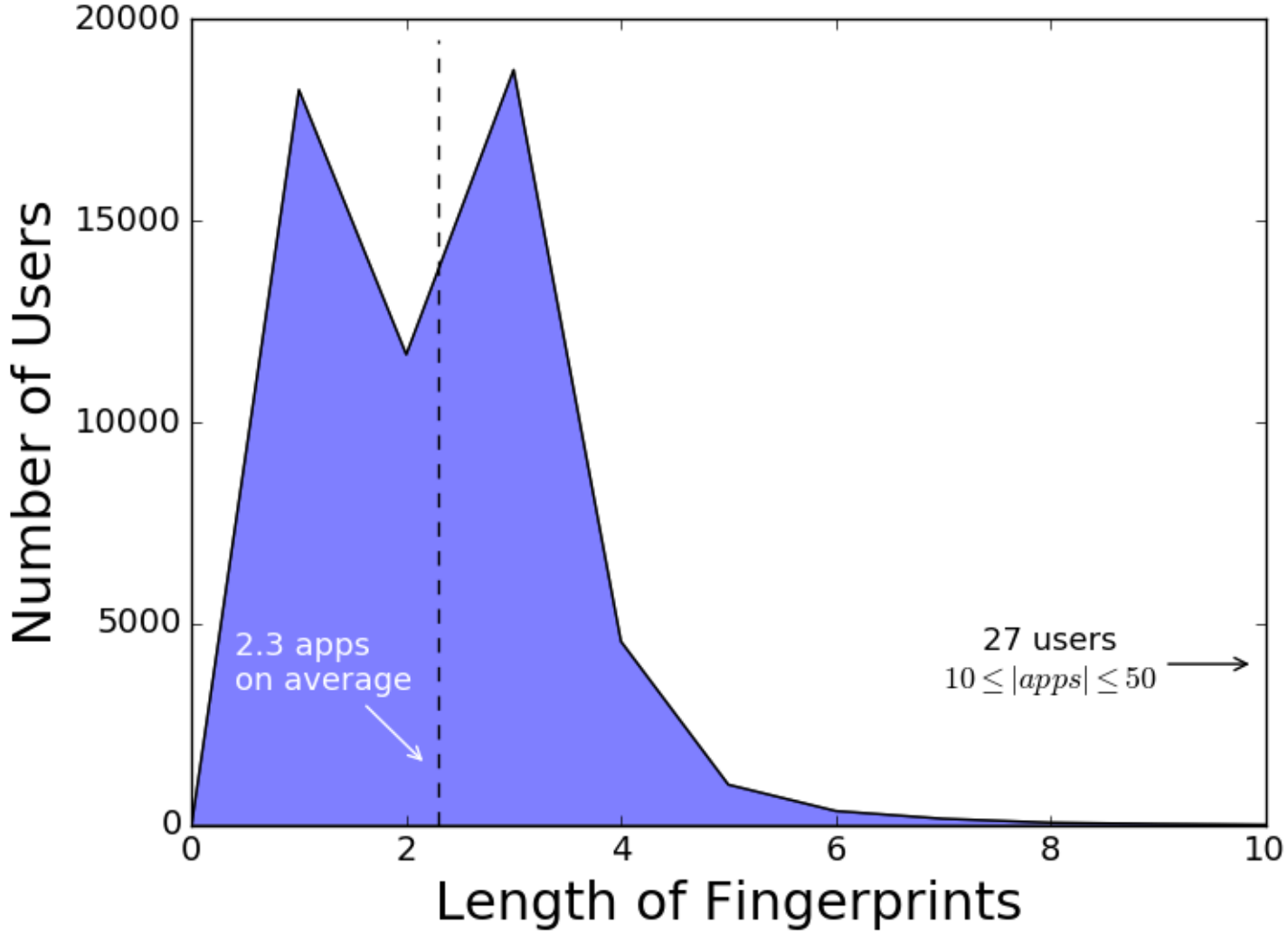
Analysis – data?

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

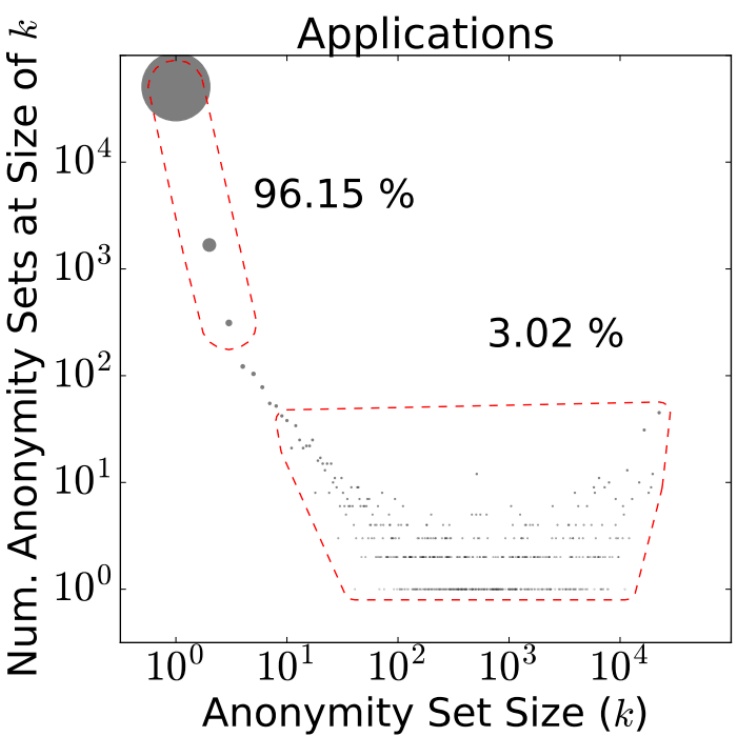
# of records	54, 893
# of all apps in the dataset	92, 210
Maximum record size	541
Minimum record size	1
Average record size	42
Std.dev of record size	39

Targeted fingerprinting

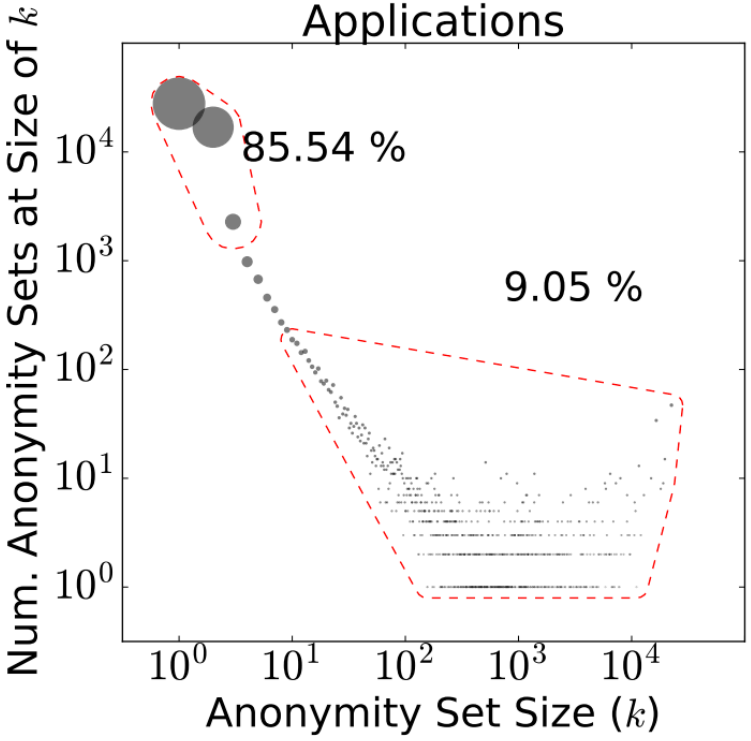
(note: limit of 50 applies)



Targeted fingerprinting (2)



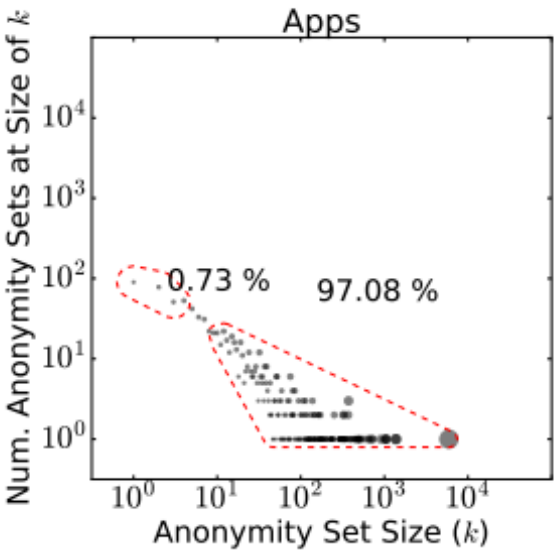
Fingerprint length: 50



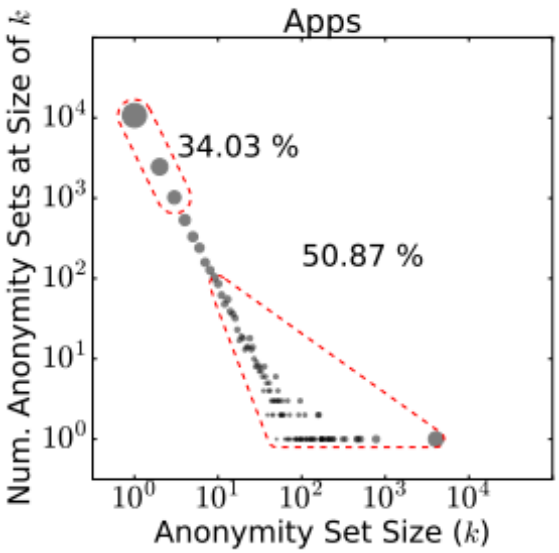
Fingerprint length: 2

Do you like these figures? Have a try! ☺
https://github.com/gaborgulyas/kmap

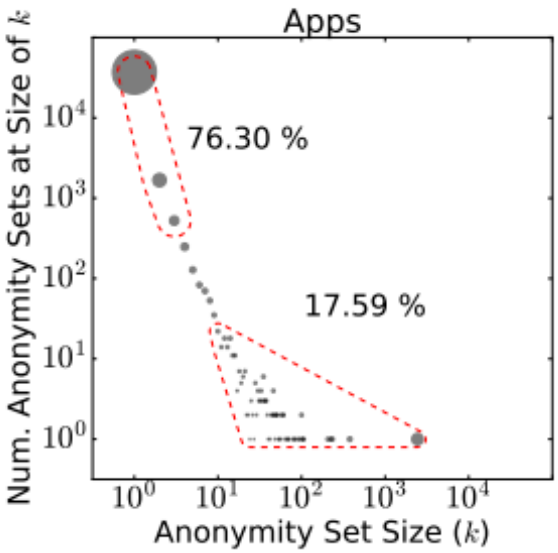
General fingerprinting



Fingerprint length: 10



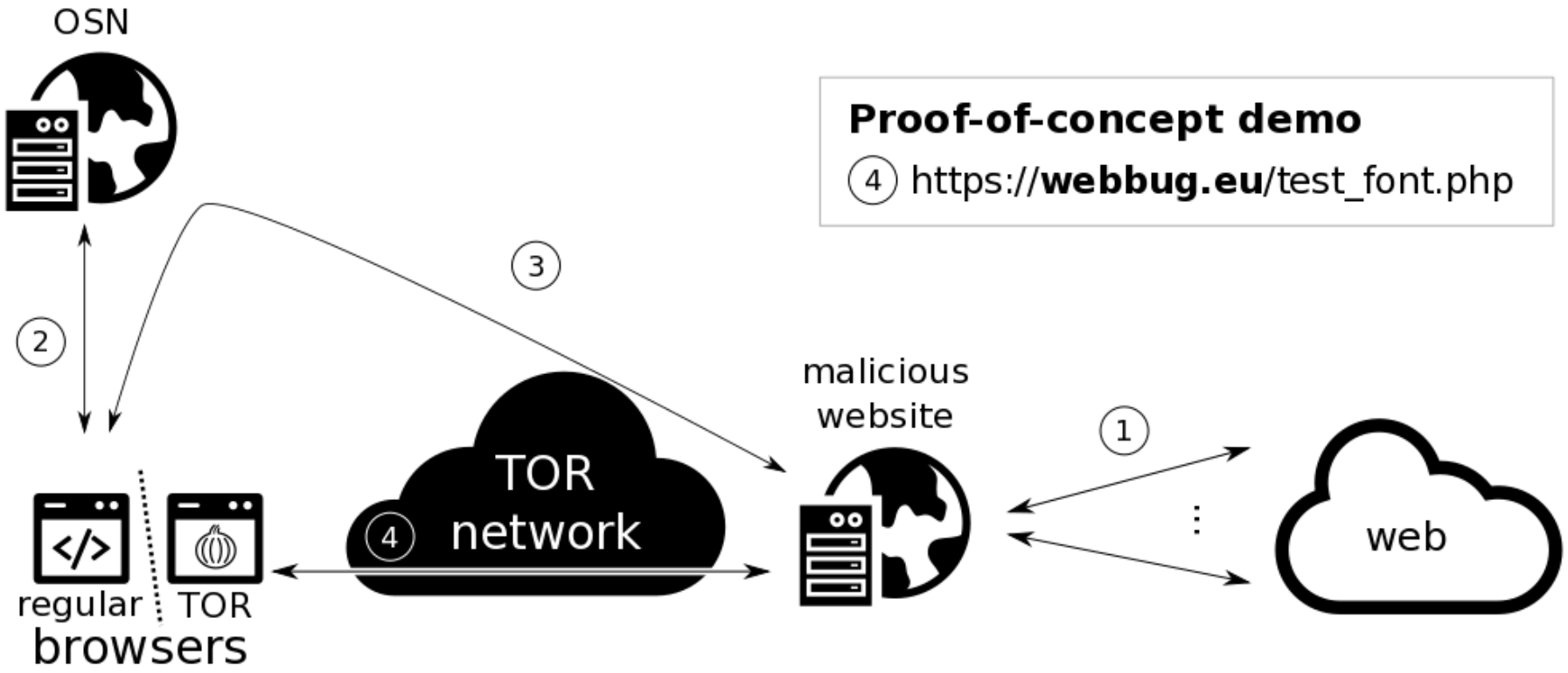
Fingerprint length: 20



Fingerprint length: 50

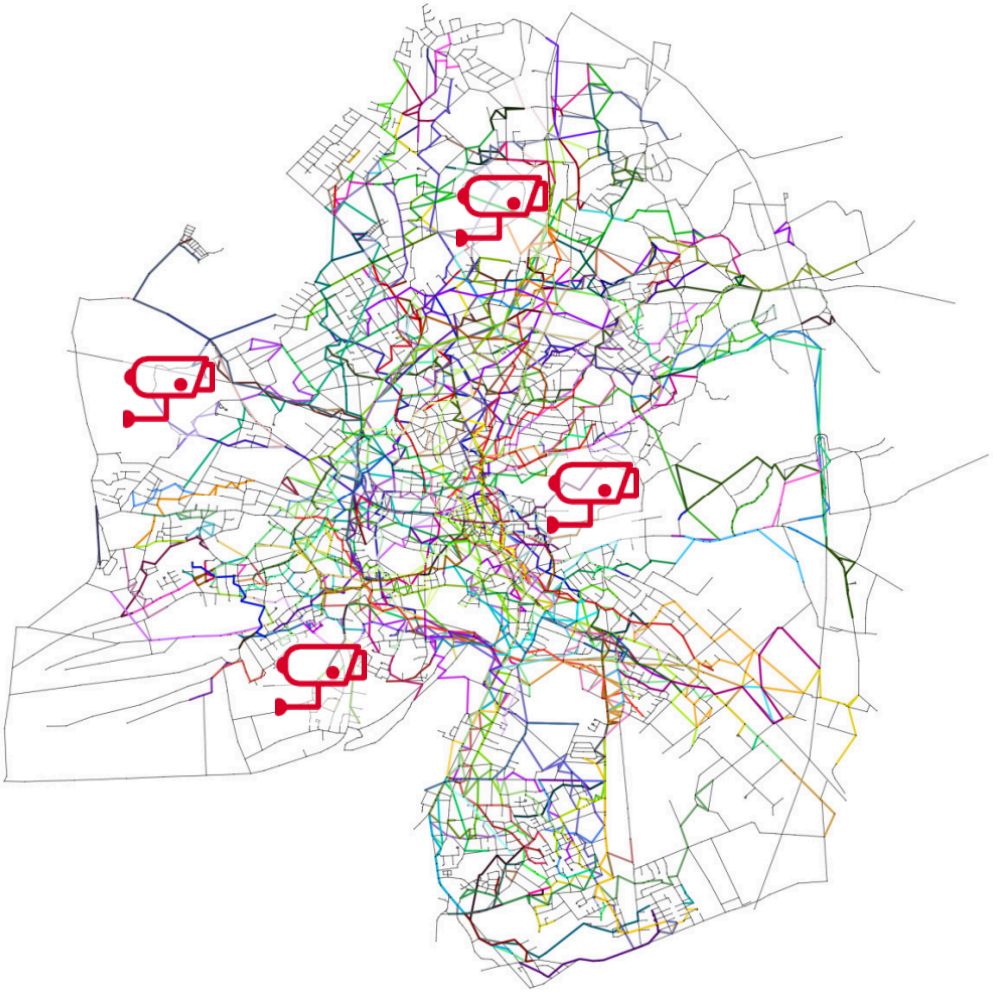
Do you like these figures? Have a try! ☺
https://github.com/gaborgulyas/kmap

Other results – checking the Tor Browser



- Attacks:
  - De-anonymization
  - Tracking
- They could work with high probability

Other results (2) – location privacy



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

Thank you for your attention!

ANY QUESTIONS?

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