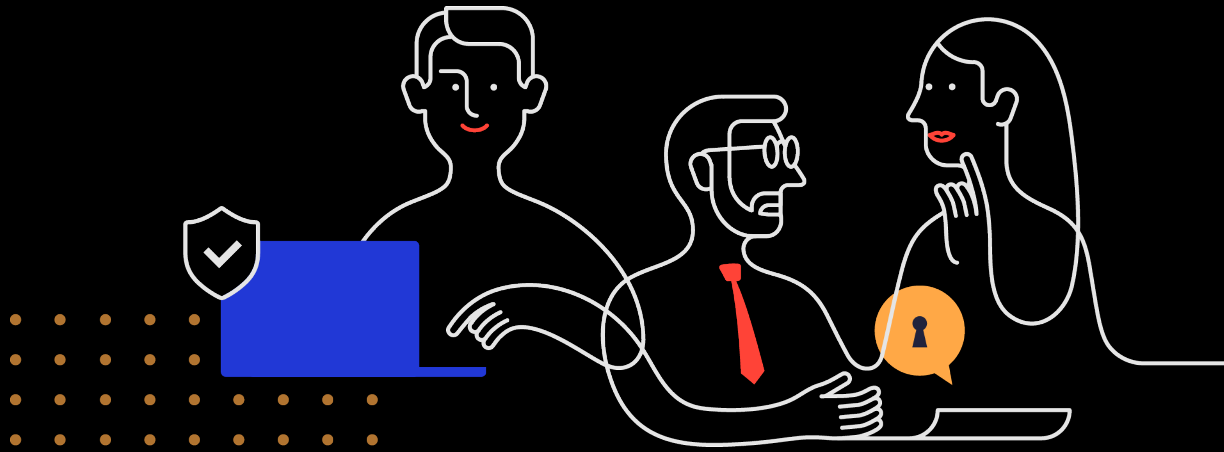
Privacy in the Era of Big Data, Machine Learning, IoT, and 5G



Elisa Bertino

Cs Department, Purdue University

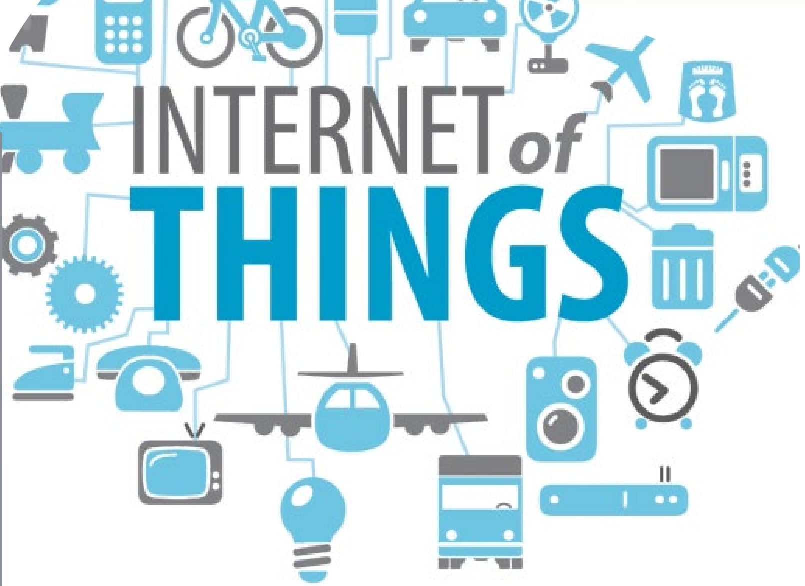
Key Technologies

Increase our capacity for

-collecting and processing data

- obtaining knowledge and recommendation from data

- making devices, control systems, and cyber-physical systems intelligent and autonomous



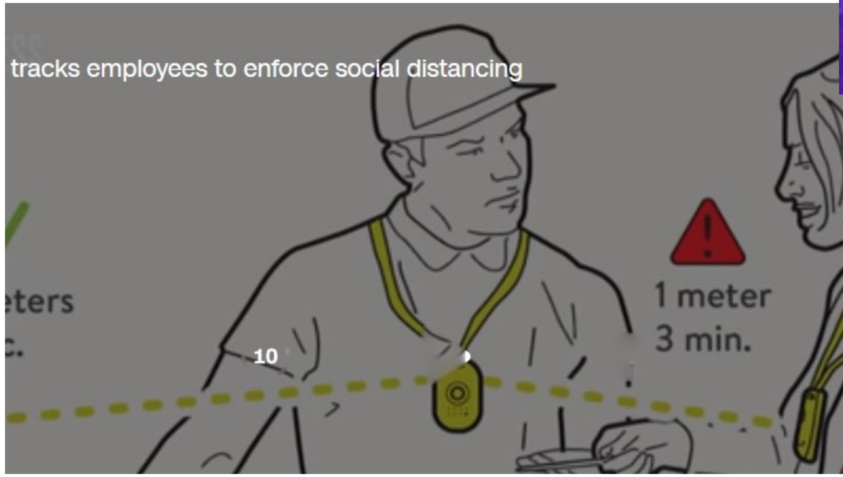
Key Technologies - Improving Security

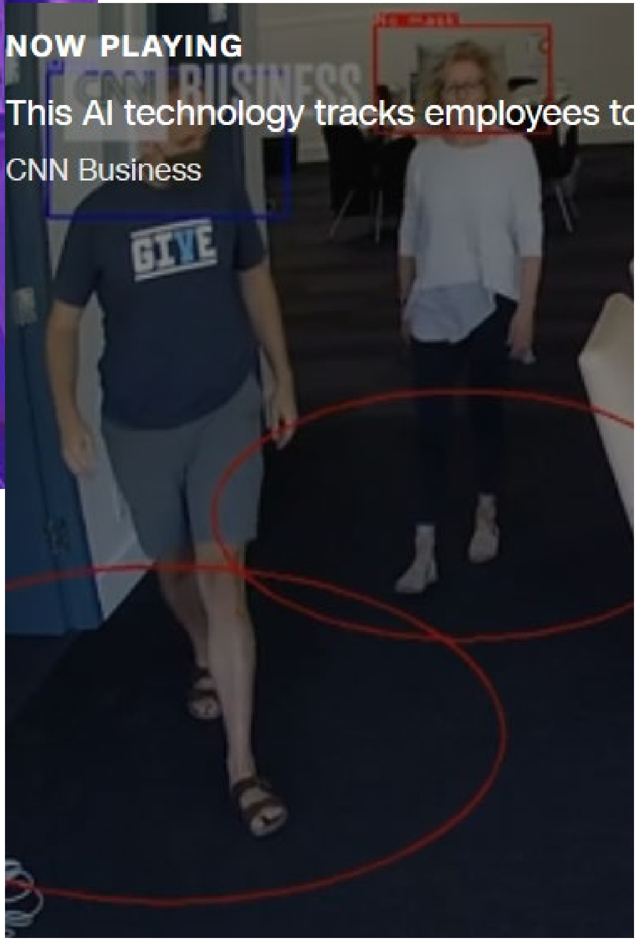
- ***Health Security***
  - Monitoring and prevention of disease spreading
  - Evidence-based healthcare
- *Cyber Security*
  - Security information and event management (SIEM)
  - Authentication (biometrics, continuous user authentication, federated ID management)
  - Access control (e.g. attribute-based, location-based and context-based access control)
  - Insider threat (anomaly detection) and user monitoring
- *Homeland Protection*
  - Identification of links and relationships among individuals in social networks
  - Prediction of attacks
  - Management of emergencies and disasters
- *Food and Water Security*
  - Precision agriculture

Health Security – How IoT and AI can help

- touchless entry
- thermal temperature scanning
- managing and tracking physical interactions among individuals
- enforcing safe distancing

Images from Forbes and CNN


tracks employees to enforce social distancing


NOW PLAYING

This AI technology tracks employees to

CNN Business

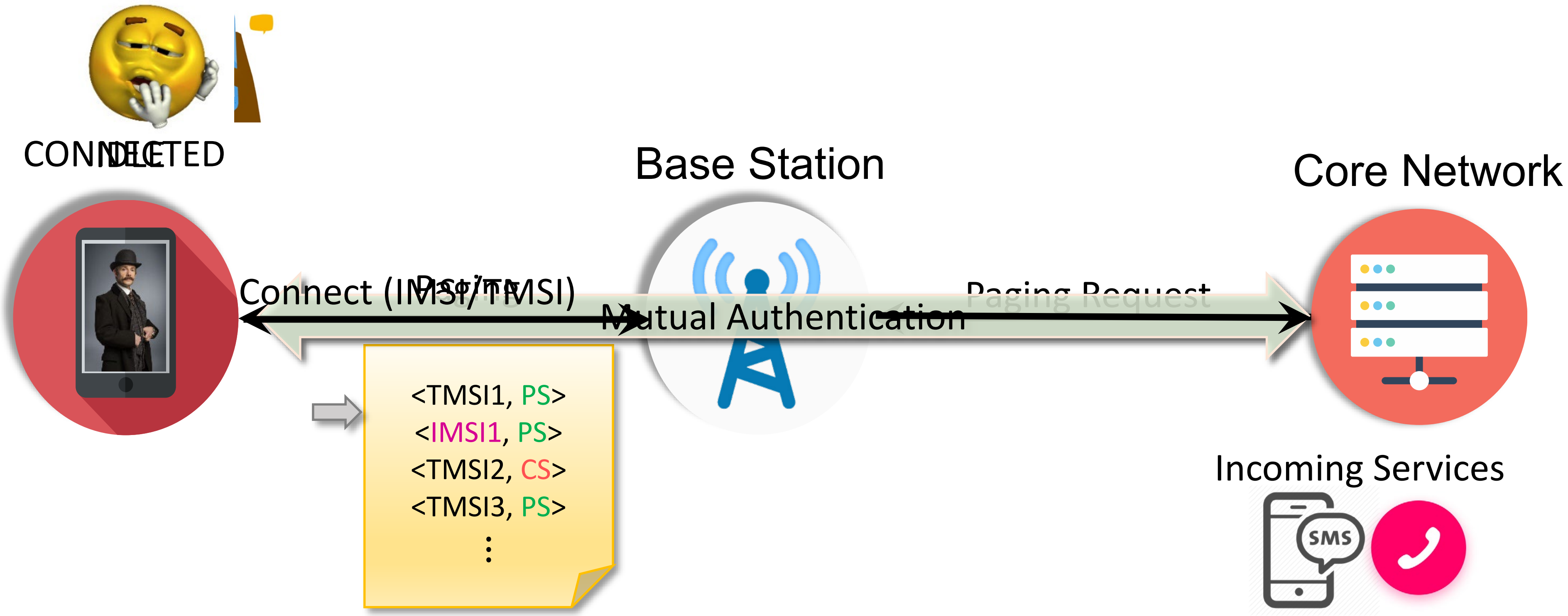
Privacy Threats

- ***Cellular Networks***
  - Matching of mobile users to access points at the physical layer
  - Traceability attacks via IMSI catching (addressed by TMSI, GUTI in 5G)
  - Exploitation of paging occasions (ToRPEDO attack)
- ***Data***
  - Data linkage
  - Lack of data security
  - Unproper use of data
- ***Mobile Applications***
  - Vulnerable mobile applications
  - “Curious” mobile applications
- ***AI and Machine Learning***
  - Inversion attacks
  - Uneven data privacy for specific subsets of users
- ***Wearable devices and continuous data streaming***

TORPEDO ATTACK – Paging Procedure

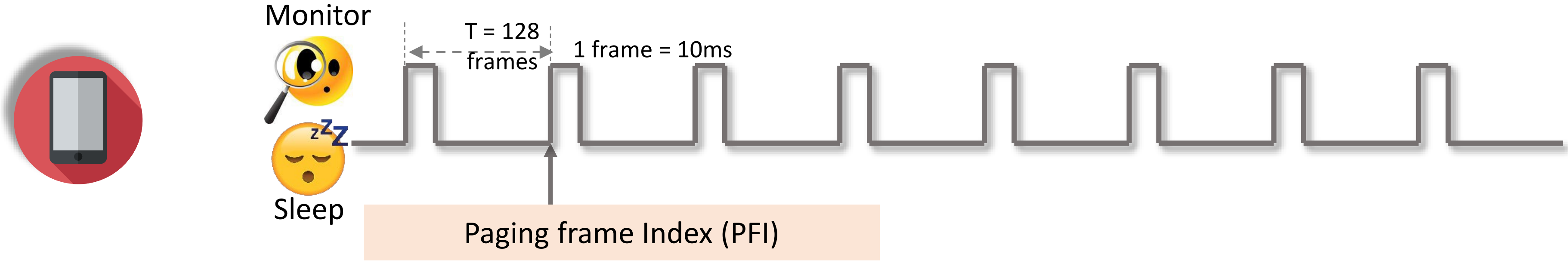
IMSI: International Mobile Subscriber Identity

TMSI: Temporary Mobile Subscriber Identity



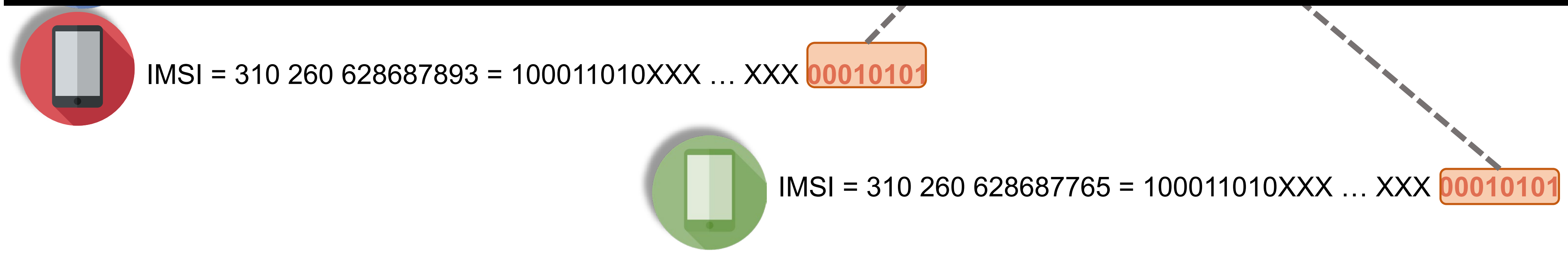
<TMSI1, PS>
<IMSI1, PS>
<TMSI2, CS>
<TMSI3, PS>
⋮

TORPEDO ATTACK – Paging Occasion



?

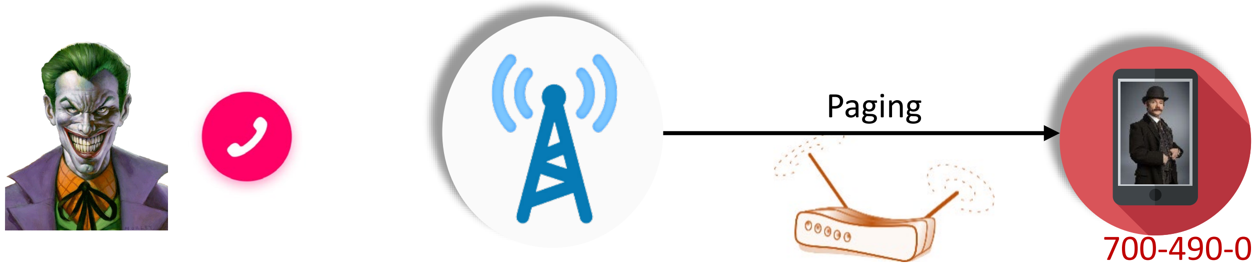
Can a passive adversary only knowing victim’s phone number/Twitter handle Identify/track the victim’s presence in a target area?



IMSI = 310 260 628687893 = 100011010XXX … XXX 00010101

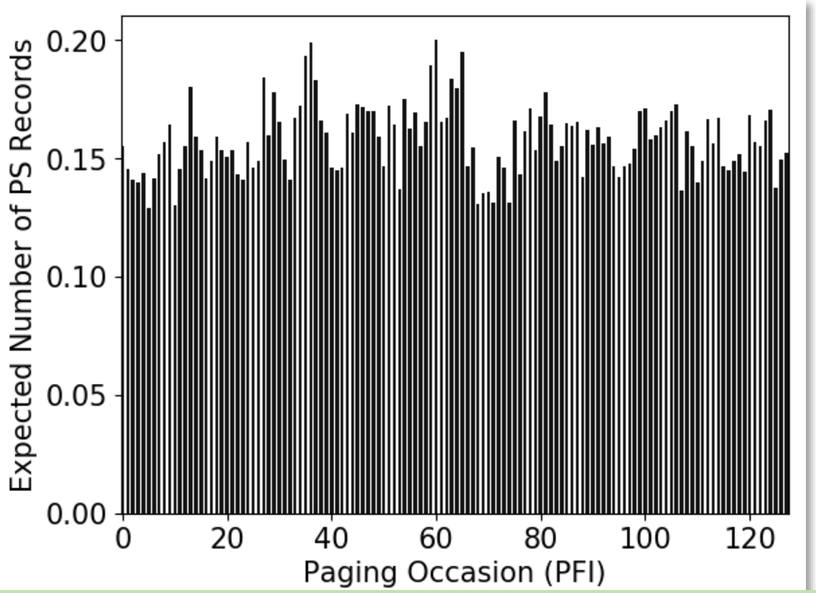
IMSI = 310 260 628687765 = 100011010XXX … XXX 00010101

ToRPEDO **TR**acking via **P**aging m**E**ssage **D**istributi**O**n

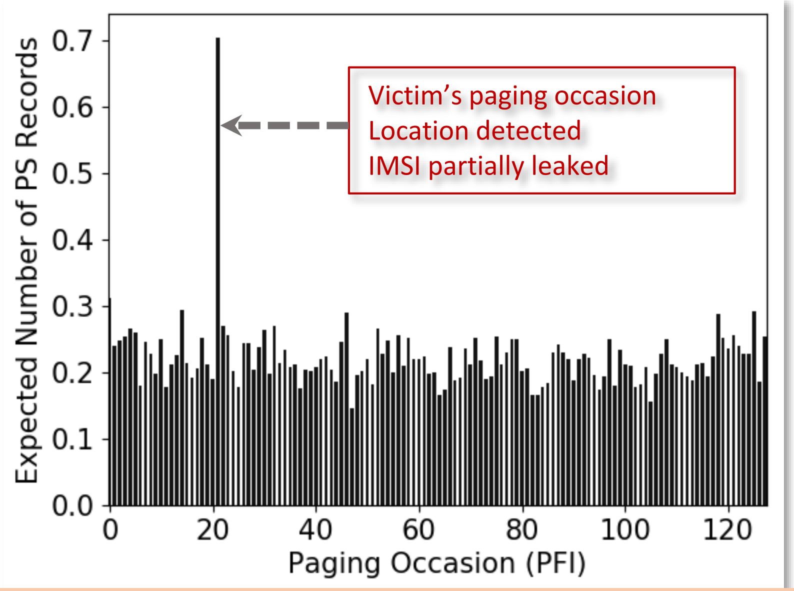


700-490-0301

https://twitter.com/victim



Distribution of paging messages (PS records) when attacker makes no phone call



Victim’s paging occasion
Location detected
IMSI partially leaked

Distribution of paging messages (PS records) when attacker makes silent phone calls

Paging Procedure – Design Vulnerabilities

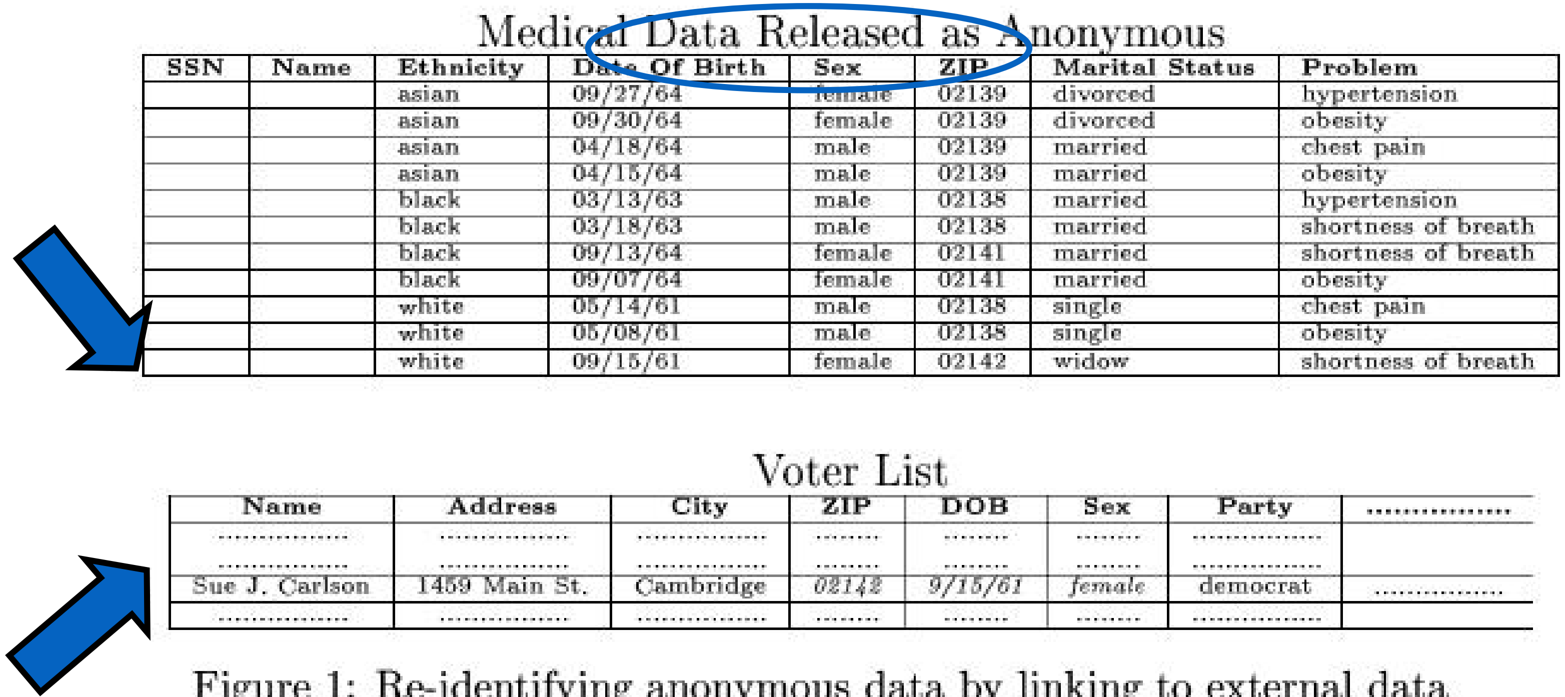
TMSI sent in plaintext and not updated frequently

Fixed paging occasion

Lack of authentication on paging messages

Data Linkage L. Sweeney’Attack (1997)

Massachusetts hospital discharge dataset



Medical Data Released as Anonymous

SSN	Name	Ethnicity	Date Of Birth	Sex	ZIP	Marital Status	Problem
		asian	09/27/64	female	02139	divorced	hypertension
		asian	09/30/64	female	02139	divorced	obesity
		asian	04/18/64	male	02139	married	chest pain
		asian	04/15/64	male	02139	married	obesity
		black	03/13/63	male	02138	married	hypertension
		black	03/18/63	male	02138	married	shortness of breath
		black	09/13/64	female	02141	married	shortness of breath
		black	09/07/64	female	02141	married	obesity
		white	05/14/61	male	02138	single	chest pain
		white	05/08/61	male	02138	single	obesity
		white	09/15/61	female	02142	widow	shortness of breath

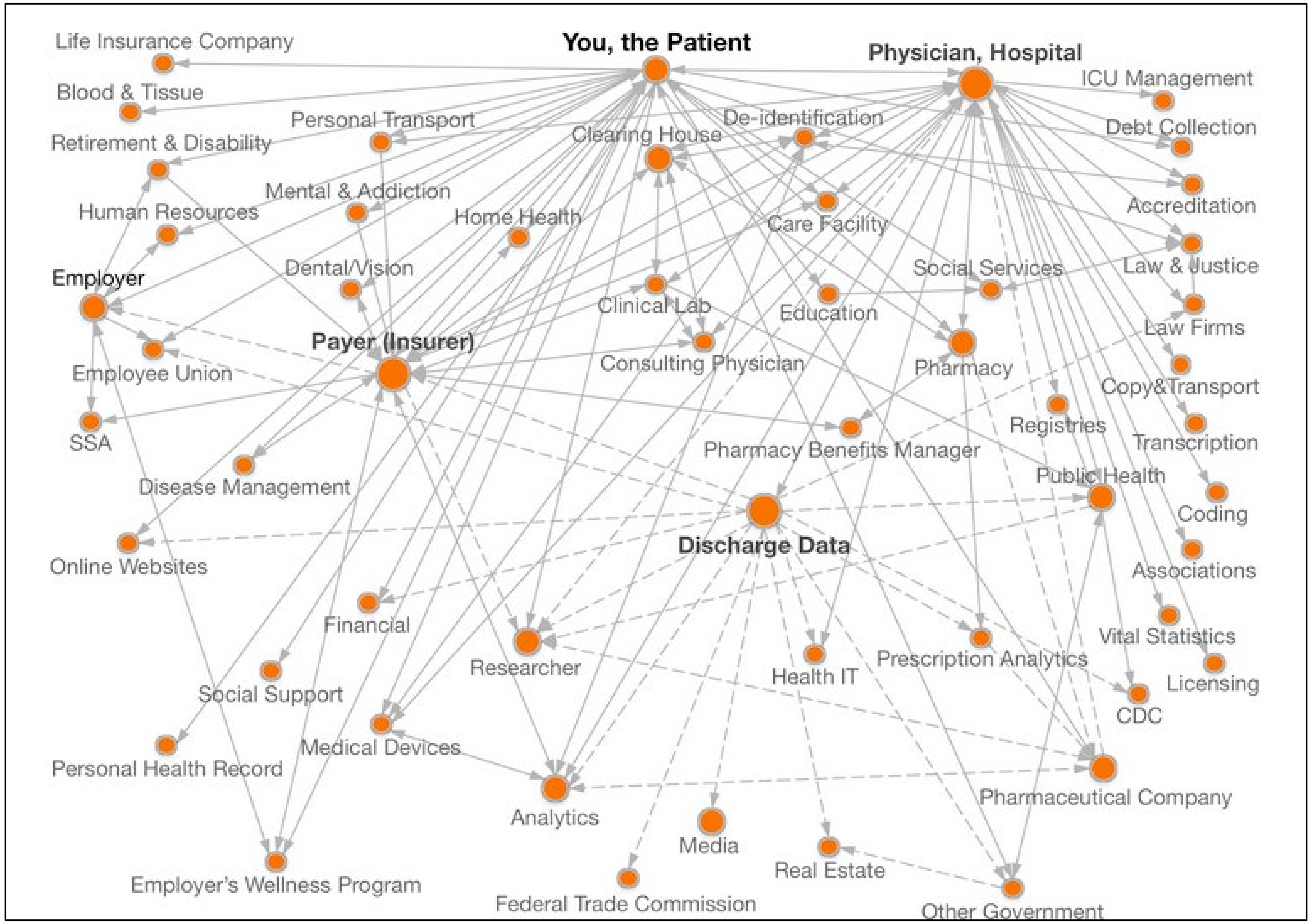
Voter List

Name	Address	City	ZIP	DOB	Sex	Party	....
....	....	....	....	....	....	....	
....	....	....	....	....	....	....	
Sue J. Carlson	1459 Main St.	Cambridge	*02142*	*9/15/61*	*female*	democrat	....
....	....	....	....	....	....	....	

Figure 1: Re-identifying anonymous data by linking to external data

Public voter dataset

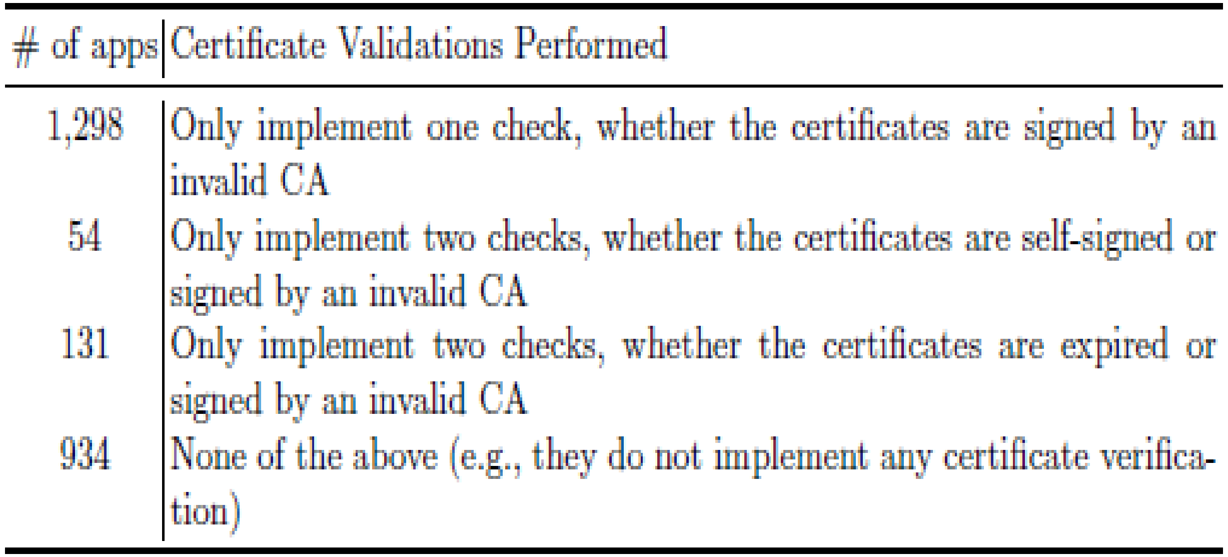
Where does your data go?



Map from the DataMap Project, https://thedatamap.org/

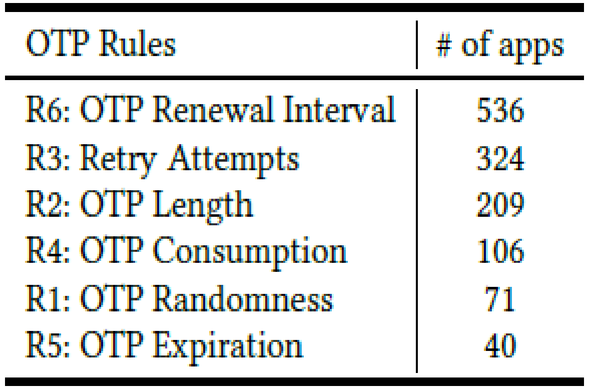
UNSECURE AND “Curious” Mobile Applications

Analysis of 13,687 apps done in 2019

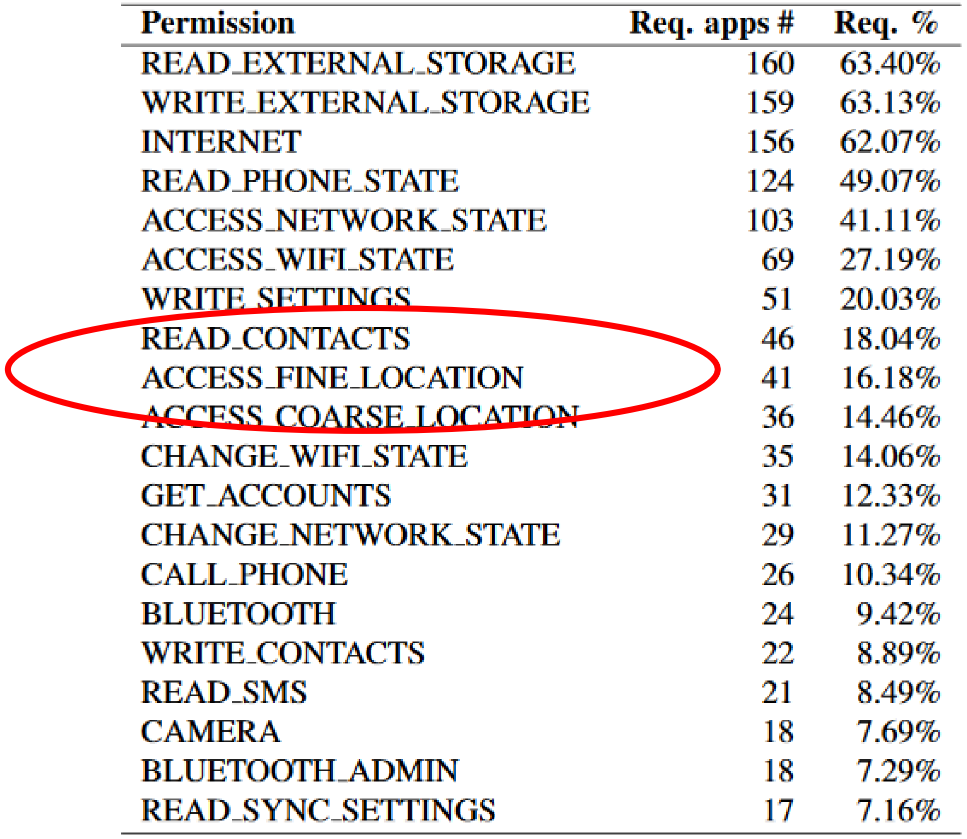


# of apps	Certificate Validations Performed
1,298	Only implement one check, whether the certificates are signed by an invalid CA
54	Only implement two checks, whether the certificates are self-signed or signed by an invalid CA
131	Only implement two checks, whether the certificates are expired or signed by an invalid CA
934	None of the above (e.g., they do not implement any certificate verification)

Analysis of 3,303 apps using OTP in 2019



OTP Rules	# of apps
R6: OTP Renewal Interval	536
R3: Retry Attempts	324
R2: OTP Length	209
R4: OTP Consumption	106
R1: OTP Randomness	71
R5: OTP Expiration	40



Permission	Req. apps #	Req. %
READ_EXTERNAL_STORAGE	160	63.40%
WRITE_EXTERNAL_STORAGE	159	63.13%
INTERNET	156	62.07%
READ_PHONE_STATE	124	49.07%
ACCESS_NETWORK_STATE	103	41.11%
ACCESS_WIFI_STATE	69	27.19%
WRITE_SETTINGS	51	20.03%
READ_CONTACTS	46	18.04%
ACCESS_FINE_LOCATION	41	16.18%
ACCESS_COARSE_LOCATION	36	14.46%
CHANGE_WIFI_STATE	35	14.06%
GET_ACCOUNTS	31	12.33%
CHANGE_NETWORK_STATE	29	11.27%
CALL_PHONE	26	10.34%
BLUETOOTH	24	9.42%
WRITE_CONTACTS	22	8.89%
READ_SMS	21	8.49%
CAMERA	18	7.69%
BLUETOOTH_ADMIN	18	7.29%
READ_SYNC_SETTINGS	17	7.16%

Top 20 requested sensitive permissions from the top 250 applications on Google Play – survey done in 2014

It Seems that Privacy is Long Gone

1993

2015

On the Internet, nobody knowns you're a dog

Peter Steiner's cartoon, as published in *The New Yorker*

"Remember when, on the Internet, nobody knew who you were?"

Kaamran Hafeez' cartoon, New Yorker, Feb.2015

What can we do?

We have a lot of privacy preserving technologies!!

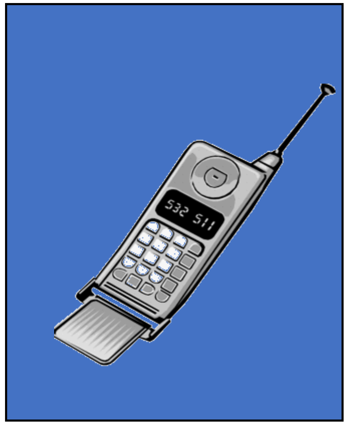
- Privacy-preserving data linkage techniques, protection against AI model inversion attacks, and privacy preserving AI
- Network anonymizers
- SMC, practical homomorphic encryption (see IBM recently released toolkit, June 4, 2020)
- Privacy-preserving digital identity management, including pseudonym systems
- <u>Access control (AC) punctuations for streaming data</u>
- <u>Anonymous “mode” for mobile applications</u>

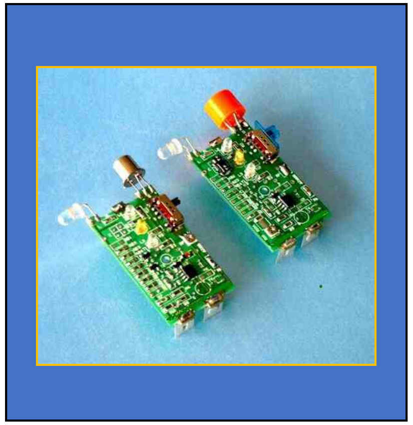
However privacy is always very personal and

different individuals often have different privacy preferences

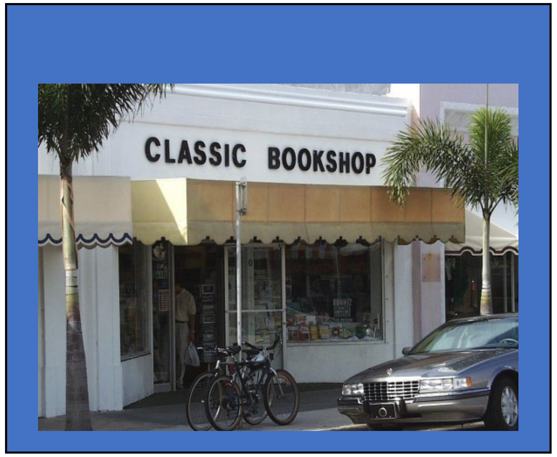
AC for Streaming data – Data Providers and Query Specifiers

Data Providers – send streaming data (objects)

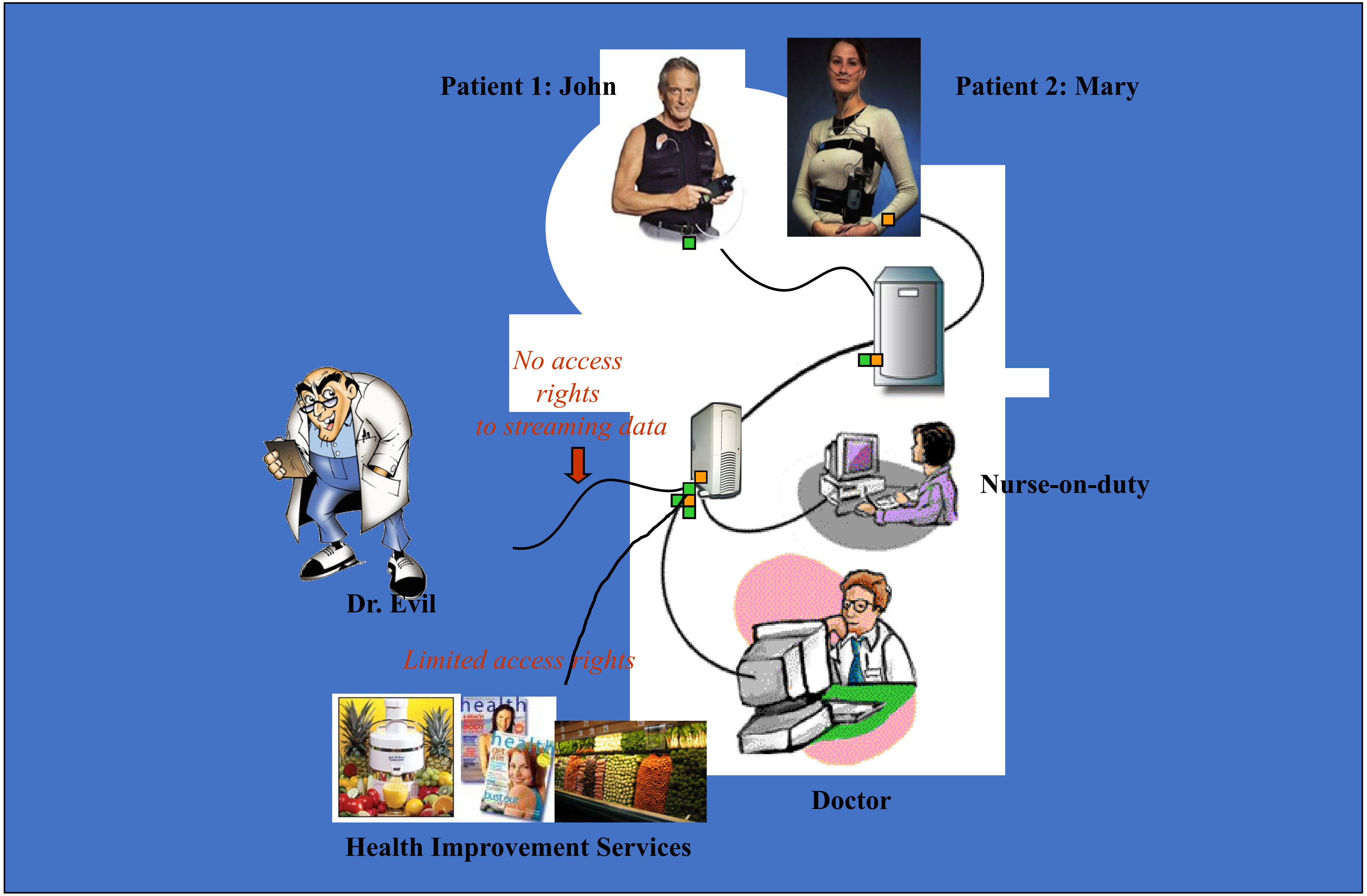




Query Specifiers (subjects) – query streaming data

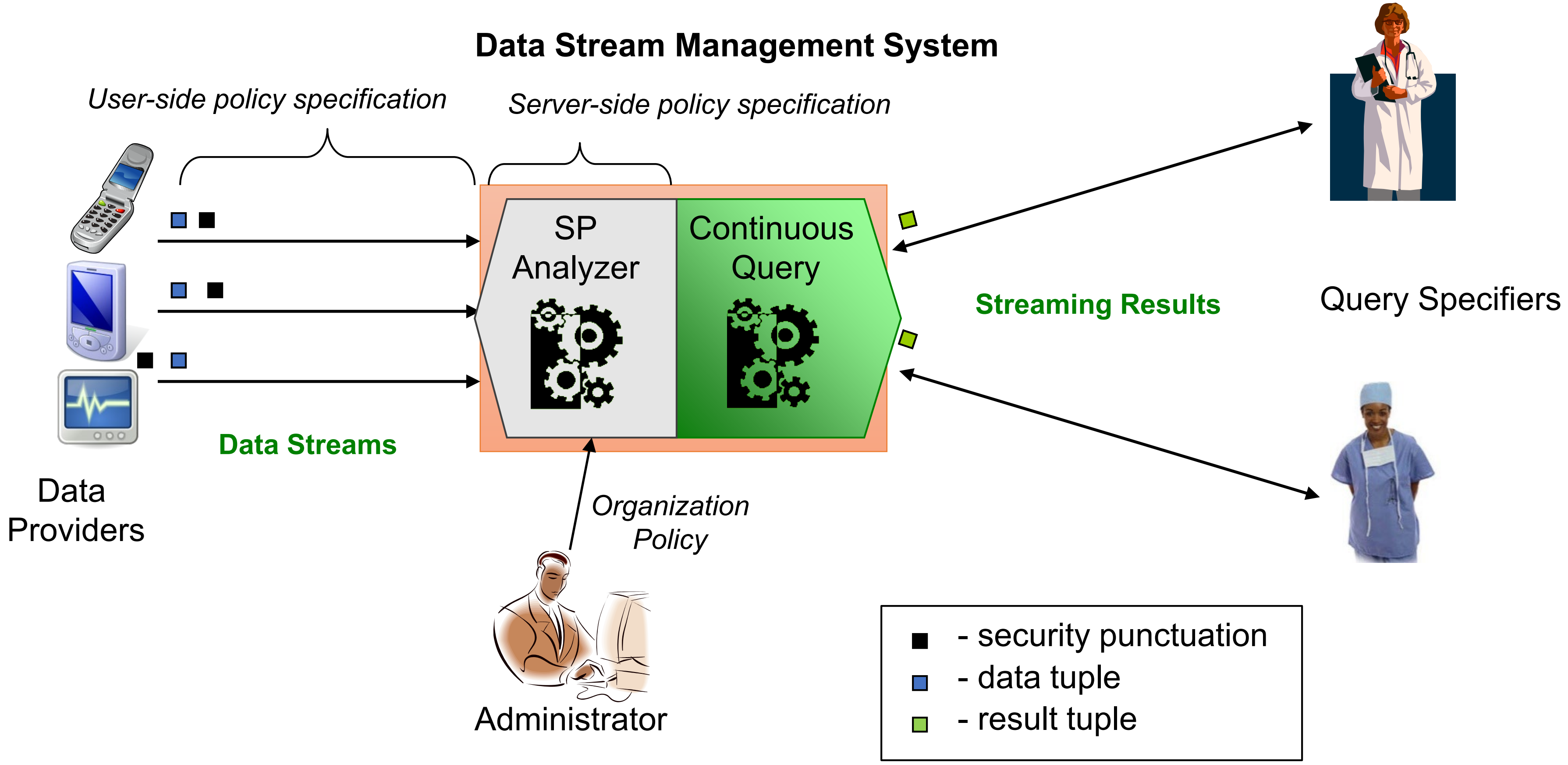


Motivating Example: Patient Monitoring



Security Punctuations (SPs) Conceptual View

- **Security Punctuations**:
  - Metadata with security semantics
  - Embedded inside data streams



Data Stream Management System

User-side policy specification

Server-side policy specification

SP Analyzer

Continuous Query

Streaming Results

Query Specifiers

Data Streams

Data Providers

Organization Policy

Administrator

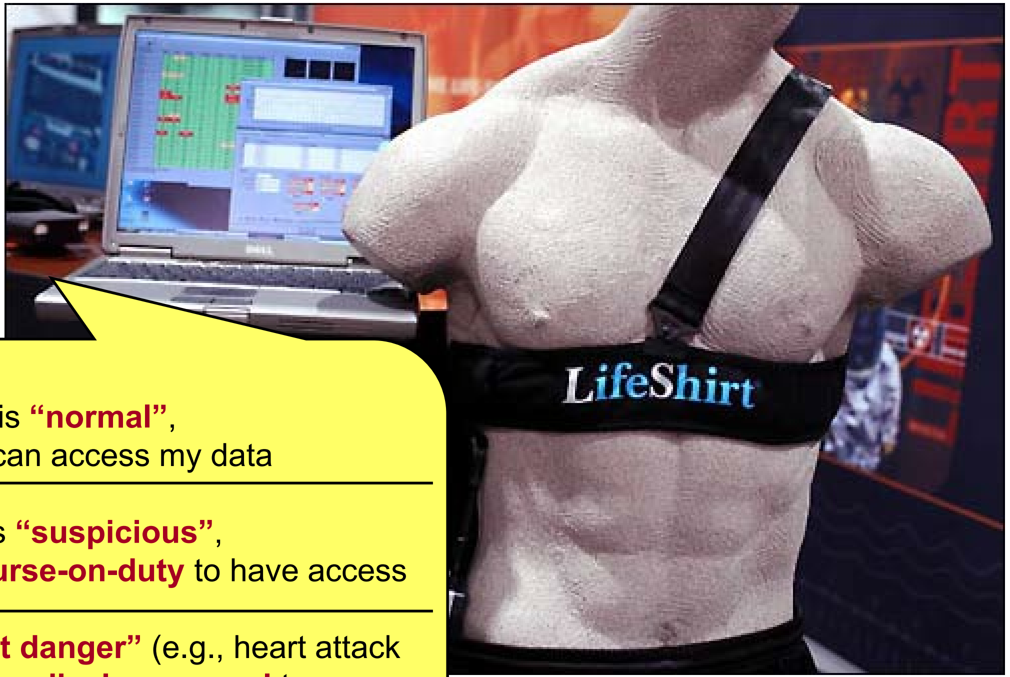
■ - security punctuation

🟦 - data tuple

🟩 - result tuple

How do security punctuations get into streams?

- Users can either ***manually*** inject security punctuations at run-time
- Devices come pre-set with a set of rules (customizable) that ***dynamically*** adjust security settings based on user preferences
- ***Machine learning*** can be used to learn security punctations and/or customize rules

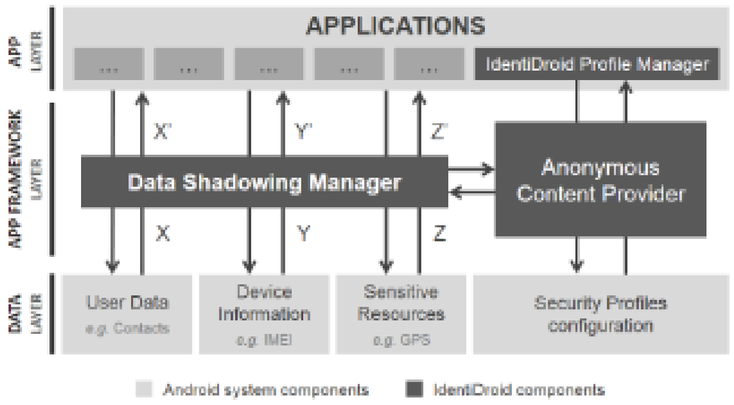


Rule 1: When everything is **"normal"**,
only my doctor can access my data

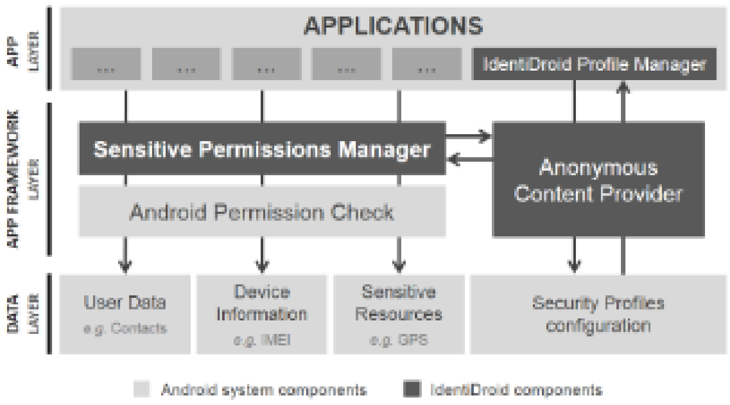
Rule 2: If something looks **"suspicious"**,
allow a **current nurse-on-duty** to have access

Rule 3: If I am in **"iminent danger"** (e.g., heart attack signs), allow **any medical personnel** to access my data

IdentiDroid – Anonymous “mode” for Apps



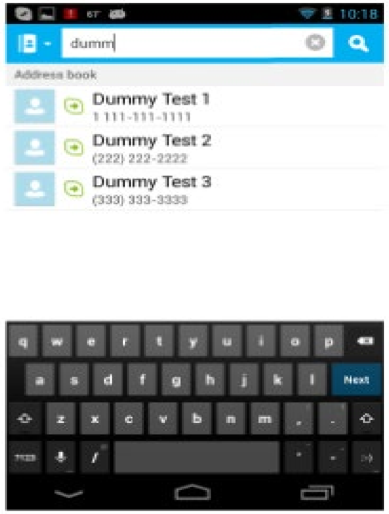
(a) Data Shadowing Manager.



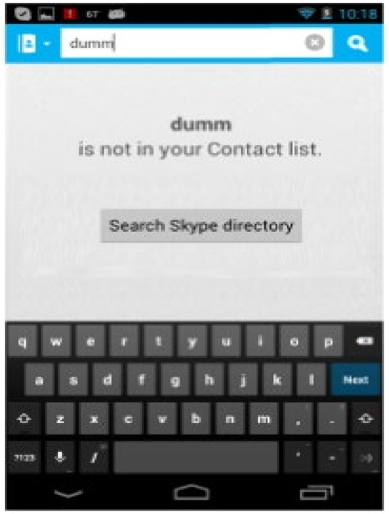
(b) Sensitive Permission Manager.

Main features

- Data shadowing
- Dynamic permission revocation
- Fresh start feature for apps



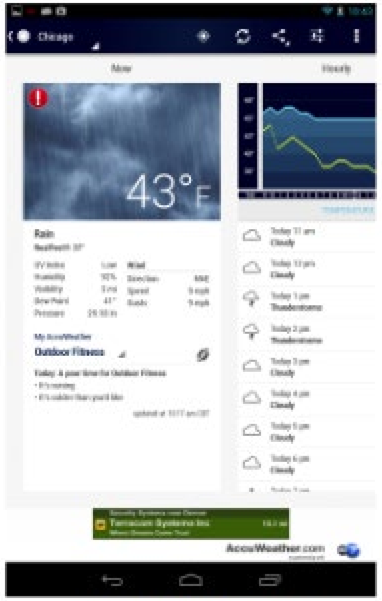
(a) Skype with “read contacts” permission granted



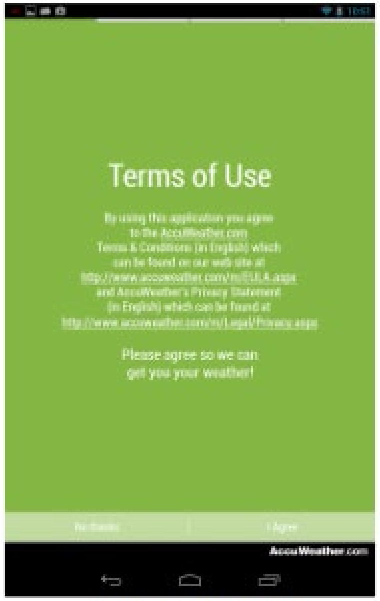
dumm
is not in your Contact list.

Search Skype directory

(b) Skype with “read contacts” permission revoked



(a) *Fresh Start* not activated



Terms of Use

By using this application you agree to the AccuWeather.com Terms & Conditions (in English) which can be found on our web site at http://www.accuweather.com/m/EULA.aspx and AccuWeather's Privacy Statement (in English) which can be found at http://www.accuweather.com/m/Legal/Privacy.aspx

Please agree so we can get you your weather!

(b) *Fresh Start* activated

So what is needed?

Combine those approaches for “**privacy protection in depth**” by developing holistic privacy-preserving environments

However a key question is “personal privacy versus collective safety”.

How can we make possible for people to make their choices about this question?

How can we make possible to reconcile those two seemingly opposing goals?

I believe that data transparency and policy-based use of data are two key elements relevant to these issues

Questions?
Thank You