

# Covert & Side Stories: *Threats Evolution in Traditional and Modern Technologies*

Mauro Conti

9 gennaio 2024

NECS – PhD Winter School 2024



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

nick



Password

# Outline



A CRYPTO NERD'S  
IMAGINATION:

HIS LAPTOP'S ENCRYPTED.  
LET'S BUILD A MILLION-DOLLAR  
CLUSTER TO CRACK IT.

BLAST! OUR  
EVIL PLAN  
IS FOILED!

NO GOOD! IT'S  
4096-BIT RSA!



WHAT WOULD  
ACTUALLY HAPPEN:

HIS LAPTOP'S ENCRYPTED.  
DRUG HIM AND HIT HIM WITH  
THIS \$5 WRENCH UNTIL  
HE TELLS US THE PASSWORD.

GOT IT.





SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA







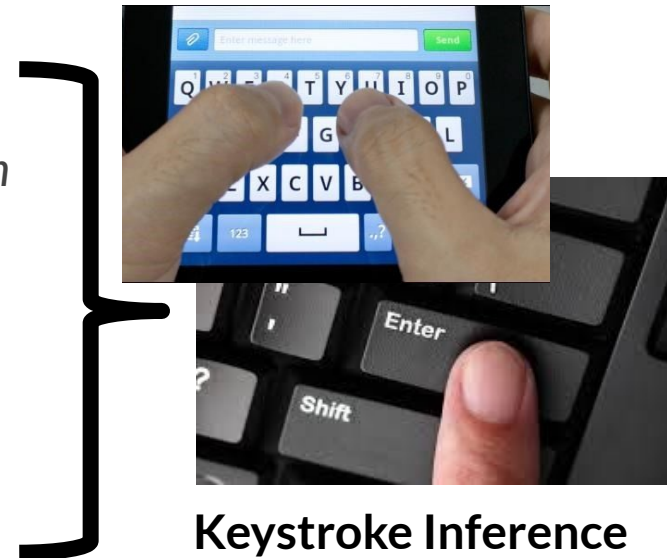
- Covert and Side Channels 101
- Network Traffic Analysis
  - *As a side channel: app and sensitive data inference*
- Energy Consumption
  - *As a side channel: user and app inference*
  - *As a covert channel: data exfiltration*
- Device Movement
  - *As a side channel: smartphone user authentication*
  - *Attacks against biometric authentication*
- Keystroke Timing
  - *As a side channel: text typed on keyboards*
- Acoustic Emanations
  - *As a side channel: text typed on keyboards*

- Covert and Side Channels 101
- Network Traffic Analysis
  - *As a side channel: app and sensitive data inference*
- Energy Consumption
  - *As a side channel: user and app inference*
  - *As a covert channel: data exfiltration*
- Device Movement
  - *As a side channel: smartphone user authentication*
  - *Attacks against biometric authentication*
- Keystroke Timing
  - *As a side channel: text typed on keyboards*
- Acoustic Emanations
  - *As a side channel: text typed on keyboards*



Physical  
Property  
Leveraged

- Covert and Side Channels 101
- Network Traffic Analysis
  - *As a side channel: app and sensitive data inference*
- Energy Consumption
  - *As a side channel: user and app inference*
  - *As a covert channel: data exfiltration*
- Device Movement
  - *As a side channel: smartphone user authentication*
  - *Attacks against biometric authentication*
- Keystroke Timing
  - *As a side channel: text typed on keyboards*
- Acoustic Emanations
  - *As a side channel: text typed on keyboards*



**Keystroke Inference**

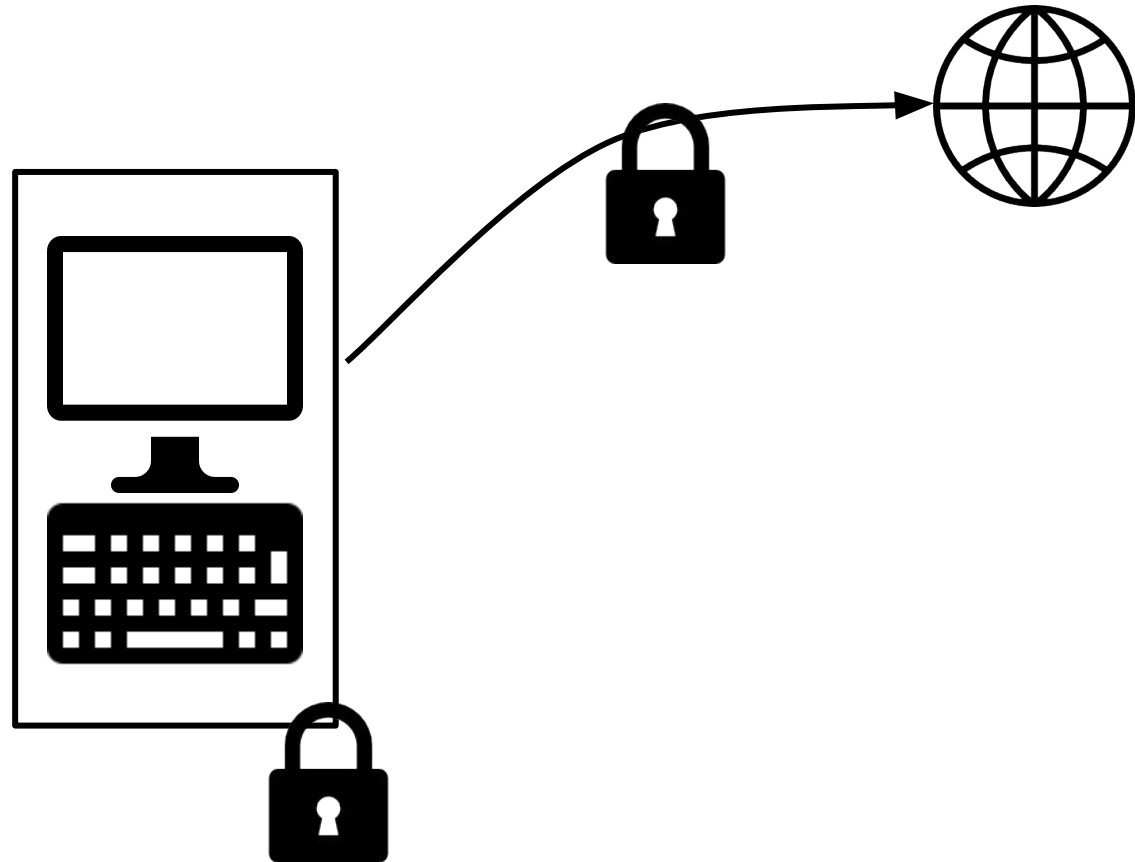


- **Covert and Side Channels 101**
- Network Traffic Analysis
  - *As a side channel: app and sensitive data inference*
- Energy Consumption
  - *As a side channel: user and app inference*
  - *As a covert channel: data exfiltration*
- Device Movement
  - *As a side channel: smartphone user authentication*
  - *Attacks against biometric authentication*
- Keystroke Timing
  - *As a side channel: text typed on keyboards*
- Acoustic Emanations
  - *As a side channel: text typed on keyboards*

# Side Channels



Devices, and network communication, are usually **protected and encrypted**



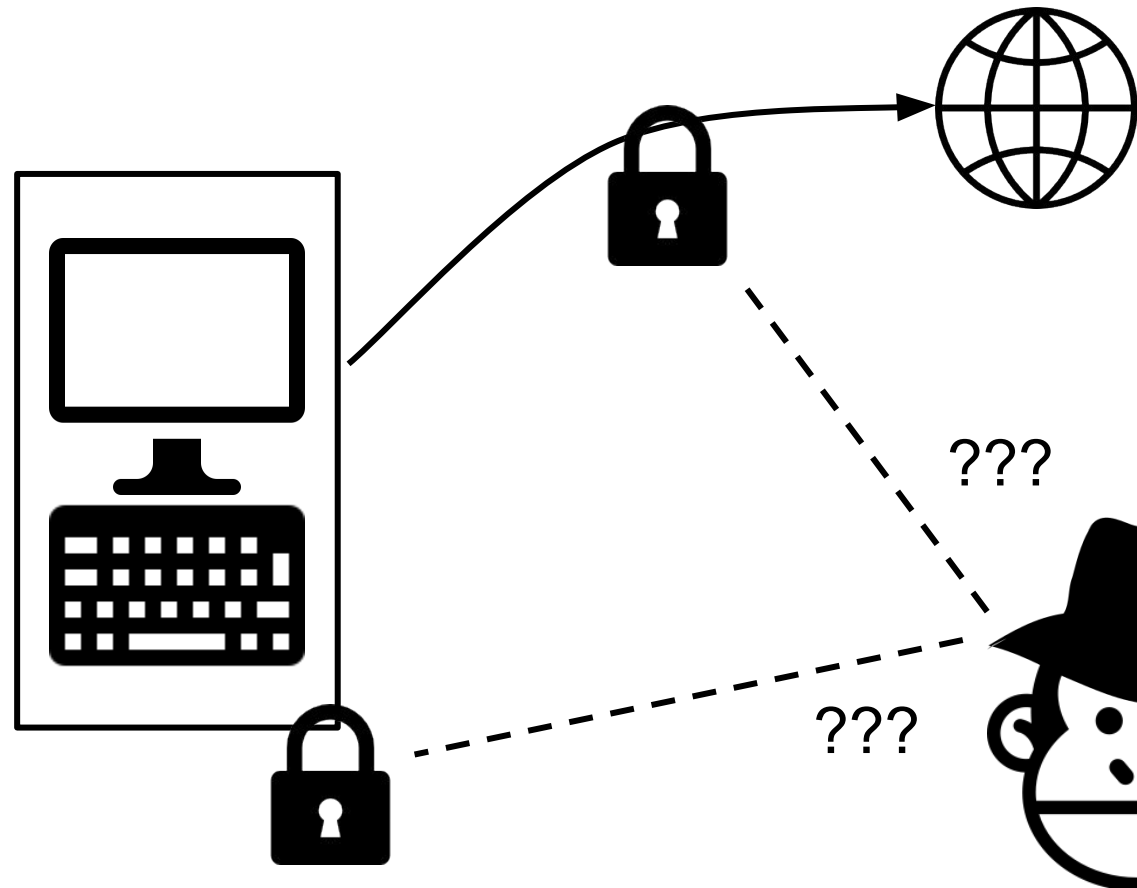


# Side Channels



Devices, and network communication, are usually **protected** and **encrypted**

→ Difficult for **Attackers** to violate such protection

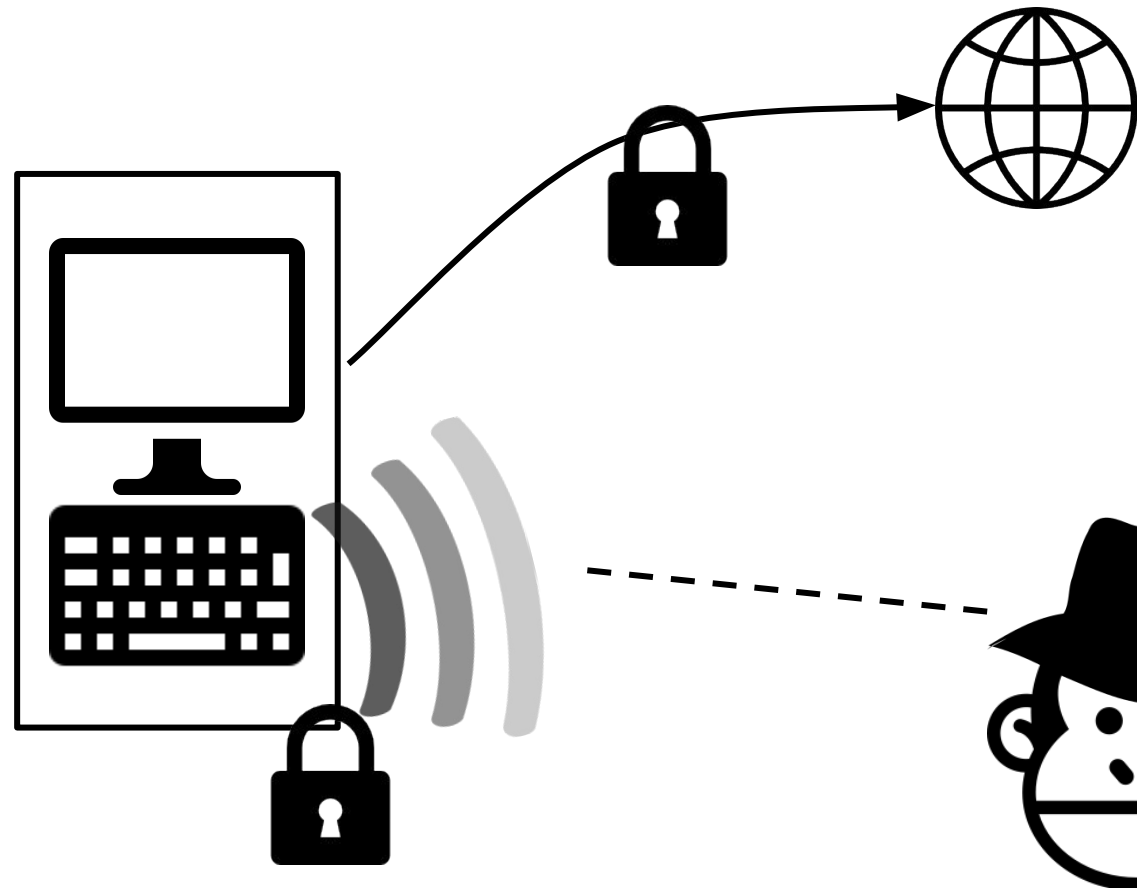


# Side Channels



Observing emanations and  
patterns  
*Can reveal secrets!*

This is called a **side channel**



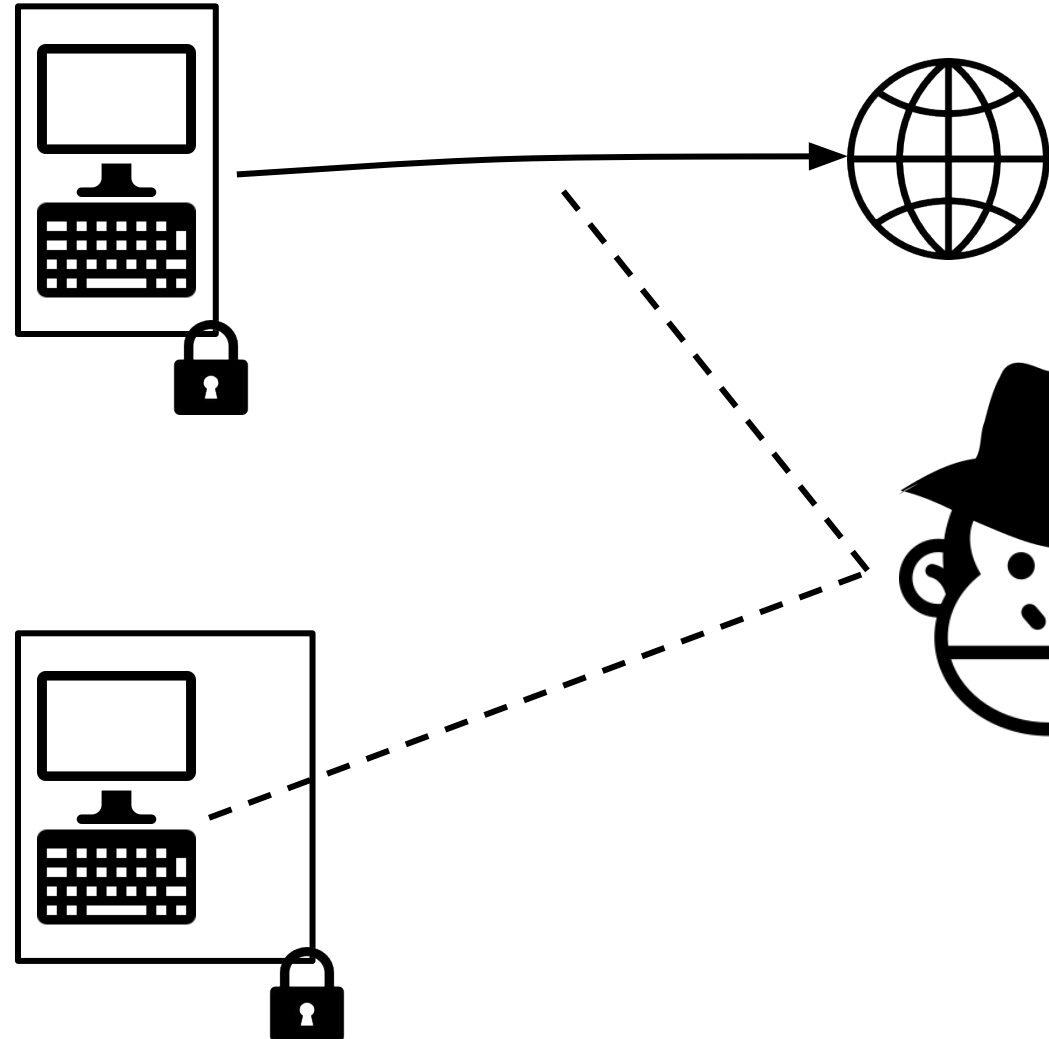
# Covert Channels



**Covert Channels** are used to communicate stealthily.

Either to **avoid listeners in the middle...**

**...or to exfiltrate information.**





- Covert and Side Channels 101
- **Network Traffic Analysis**
  - *As a side channel: app and sensitive data inference*
- Energy Consumption
  - *As a side channel: user and app inference*
  - *As a covert channel: data exfiltration*
- Device Movement
  - *As a side channel: smartphone user authentication*
  - *Attacks against biometric authentication*
- Keystroke Timing
  - *As a side channel: text typed on keyboards*
- Acoustic Emanations
  - *As a side channel: text typed on keyboards*



M. Conti, L. V. Mancini, R. Spolaor, N. V. Verde.

**Can't you hear me knocking: Identification of user actions on Android apps via traffic analysis.**

*In ACM SIGSAC CODASPY 2015*

V. F. Taylor, R. Spolaor, M. Conti, I. Martinovic.

**AppScanner: Automatic Fingerprinting of Smartphone Apps From Encrypted Network Traffic.**

*In IEEE EuroSP 2016*



# Traffic Analysis

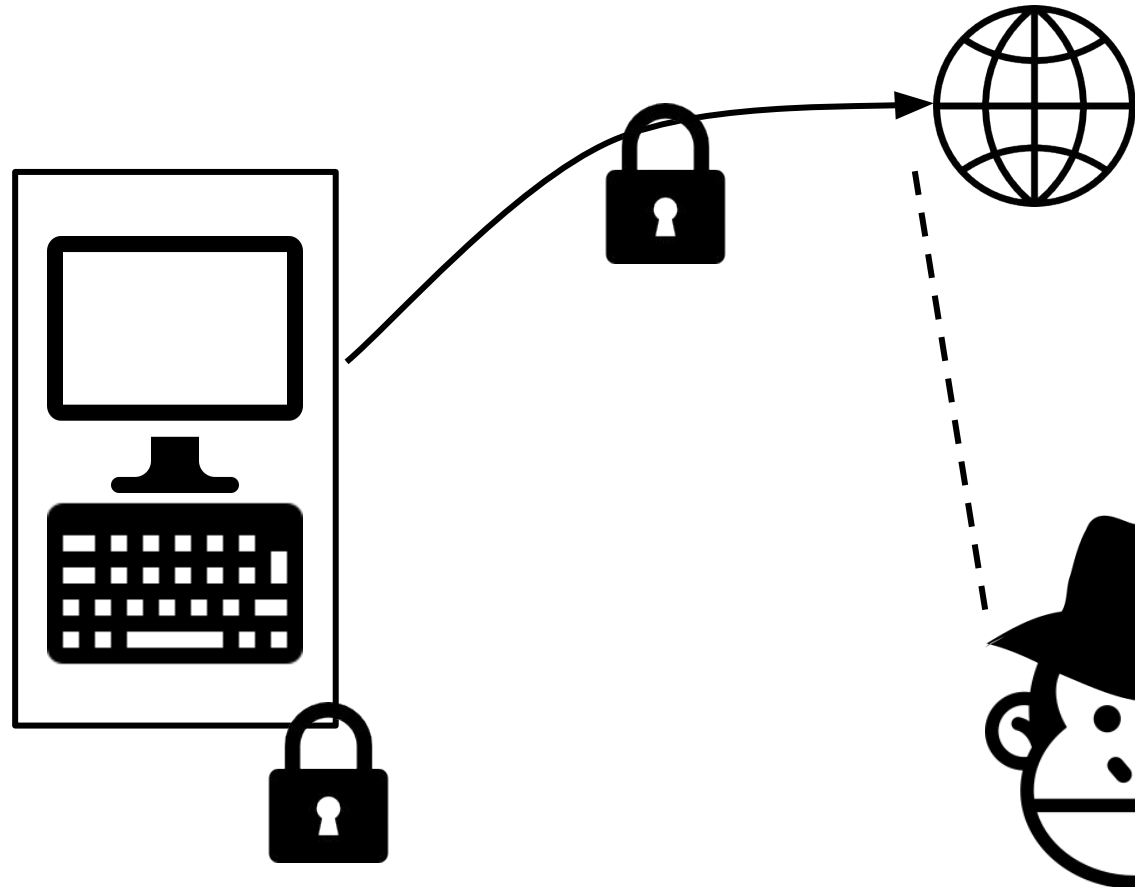


## Traffic patterns

*Can reveal what we are doing!*

Device-platform interaction  
reveals our actions

Called **traffic analysis**



# Can't you hear me knocking (CODASPY '14, TIFS '15)



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



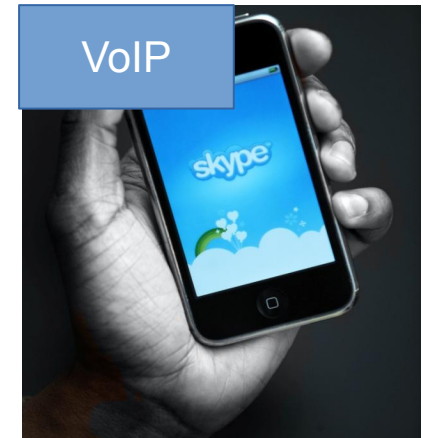
UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

## Motivation

Encryption is not enough!



[Song et al. '11]

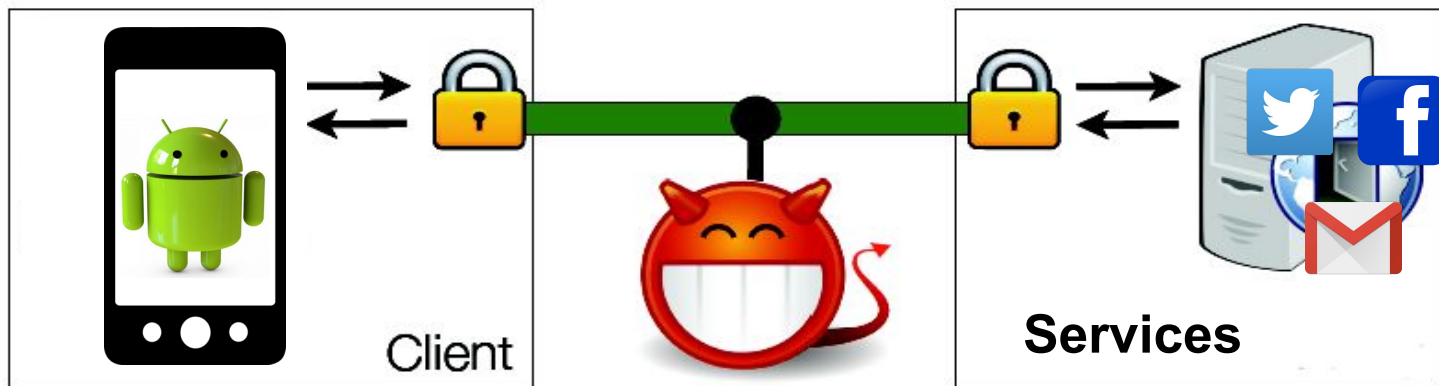


[Wright et al. '08]

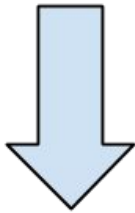
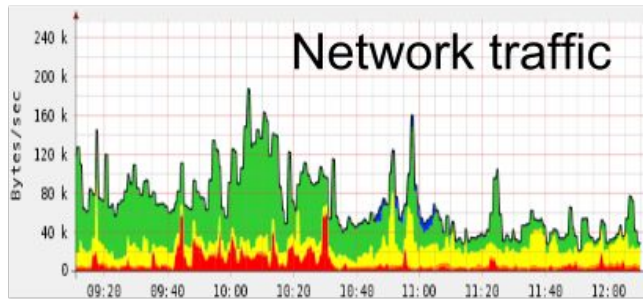
## Attacker's observations

- Coarse features:
  - Packet lengths
  - Packet directions
  - Packet timings
  - ....

Enable Traffic  
Analysis Attacks



# Attack scenario



## Log actions

- 12.30 Post on wall
- 11.44 Private message
- 11.21 Post on wall
- 10.45 User profile page
- 10.30 Post on wall
- 09.21 Open Facebook

facebook

Search

Facebook Like

Wall Info Resources Stories Facebook Live Press >>

Facebook Don't just watch the U.S. election results, be part of the conversation during a Live Town Hall starting at 7 pm EDT Tuesday from ABC News and Facebook. Ask your own questions, answer surveys and invite your friends to watch with you at <http://apps.facebook.com/twentytenthall/>. Check out U.S. Politics on Facebook and ABC News for more details. 6 hours ago · Comment · Like

54 people like this.

View all 111 comments

Write a comment...

Facebook We're proud to be joining the Alliance To Save Energy and to be working on making the systems that run Facebook even more efficient.

facebook Facebook Friends' the Alliance to Advance the Cause of Saving Energy | Alliance to Save Energy ase.org

In Facebook's explosive six-year history, millions of people around the globe have shared stories, made new connections and strengthened old friendships on the social networking site. But what many users don't know is that Facebook, which boasts more than 500 million users, also is a pioneer in ener

Saturday at 7:29am · Comment · Like · Share

11,158 people like this.

View all 1,922 comments

Write a comment...

facebook No one wants spam on their favorite Pages, so we've launched new filters for Page admins to help improve the quality of posts you see. If you run a Page, be sure to like the Facebook Pages page for more updates.

Improving Page Content on Your Wall

Facebook Pages are intended to help people engage and interact with high quality content from their favorite brands and celebrities....

By: Facebook Pages

Saturday at 3:12am · Comment · Like · Share

Lidor Beck and 13,397 others like this.

Information

Founded: February 4, 2004

82 Friends Like This

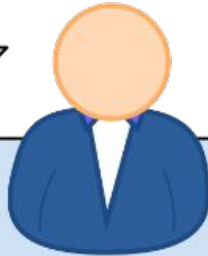
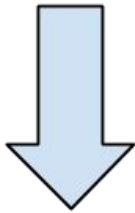
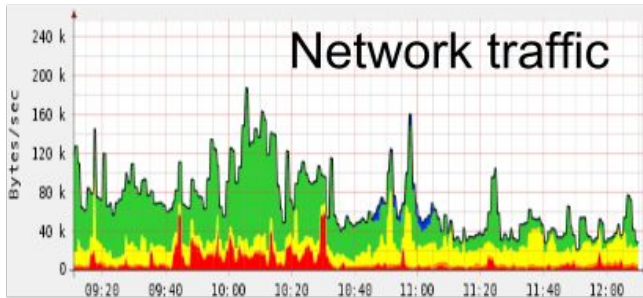
6 of 82 Friends See All

Michael Kempton-Jones Rodney Bethune Crystal Merritt Benjamin Patch John Bobry Siddartha Thota

24,369,086 People Like This

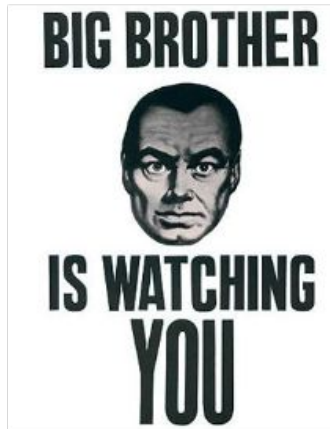


# Attack scenario



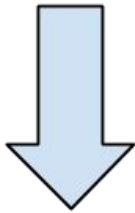
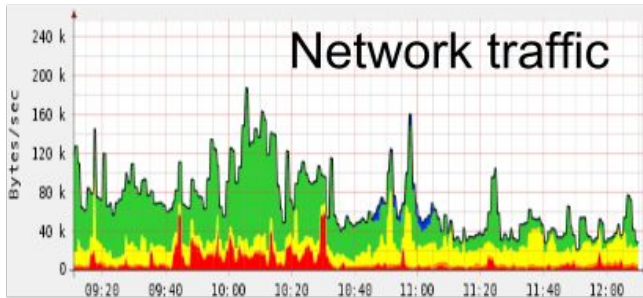
## Log actions

- 12.30 Post on wall
- 11.44 Private message
- 11.21 Post on wall
- 10.45 User profile page
- 10.30 Post on wall
- 09.21 Open Facebook



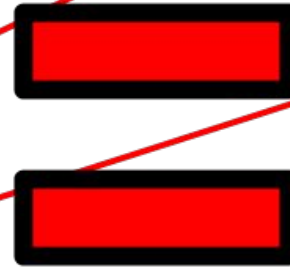
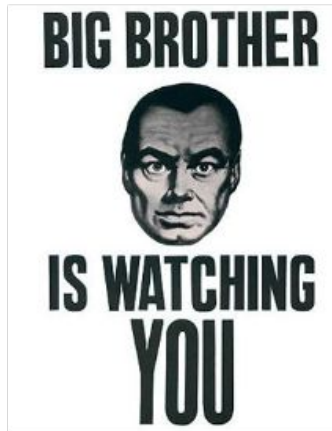
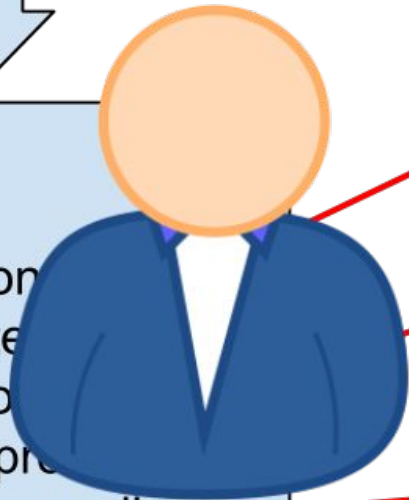


# Attack scenario



**Log actions**

- 12.30 Post on
- 11.44 Private
- 11.21 Post on
- 10.45 User pr
- 10.30 Post on wall
- 09.21 Open Facebook



facebook

Facebook Don't just watch the U.S. election results, be part of the conversation during a Live Town Hall starting at 7 pm EDT Tuesday from ABC News and Facebook. Ask your own questions, answer surveys and invite your friends to watch with you at <http://apps.facebook.com/twentytownhall/>. Check out U.S. Politics on Facebook and ABC News for more details. 6 hours ago · Comment · Like

54 people like this.

View all 111 comments

Write a comment...

facebook We're proud to be joining the Alliance To Save Energy and to be working on making the systems that run Facebook even more efficient.

facebook Facebook 'Friends' the Alliance to Advance the Cause of Saving Energy | Alliance to Save Energy ase.org

In Facebook's explosive six-year history, millions of people around the globe have shared stories, made new connections and strengthened old friendships on the social networking site. But what many users don't know is that Facebook, which boasts more than 900 million users, is also a pioneer in ener

Saturday at 7:29am · Comment · Like · Share

11,158 people like this.

View all 1,922 comments

Write a comment...

facebook No one wants spam on their favorite Pages, so we've launched a new tool for Page admins to help improve the quality of posts you see. If you're a Page admin, be sure to like the Facebook Pages page for more updates.

Improving Page Content on Your Wall

Facebook Pages are intended to help people engage and interact with high quality content from their favorite brands and celebrities...

By: Facebook Pages

Patch

24,369,086 People Like This

Saturday at 3:12am · Comment · Like · Share

Lidor Beck and 13,397 others like this.



- **To identify communicating parties**
  - **from sending/receiving pattern**
- **Behavioural profiling**
  - **to improve fingerprintings**
  - **for marketing reasons**
  - **...**



## The goal

Can an attacker recognize actions that a user performs on some android app by analyzing the **encrypted network traffic**?

### Contribution

- We prove that it is possible, with an accuracy  $> 95\%$
- Traffic analysis using **machine learning** techniques

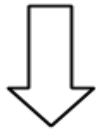
## Key Concepts

**Interactions**



Input on a device

E.g., tap, swipe,  
key press



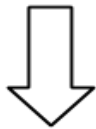
used to achieve

**User actions**



Operation on apps

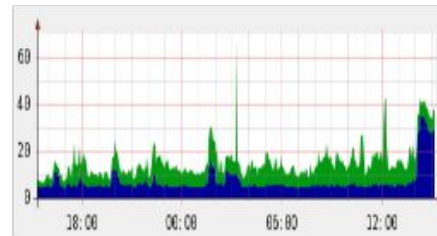
E.g., send an email,  
open a page



produce

**Network flows**

**tumblr.**



Sequence of packets

Couple of IP addresses  
and ports

# Can't you hear me knocking (CODASPY '14, TIFS '15)

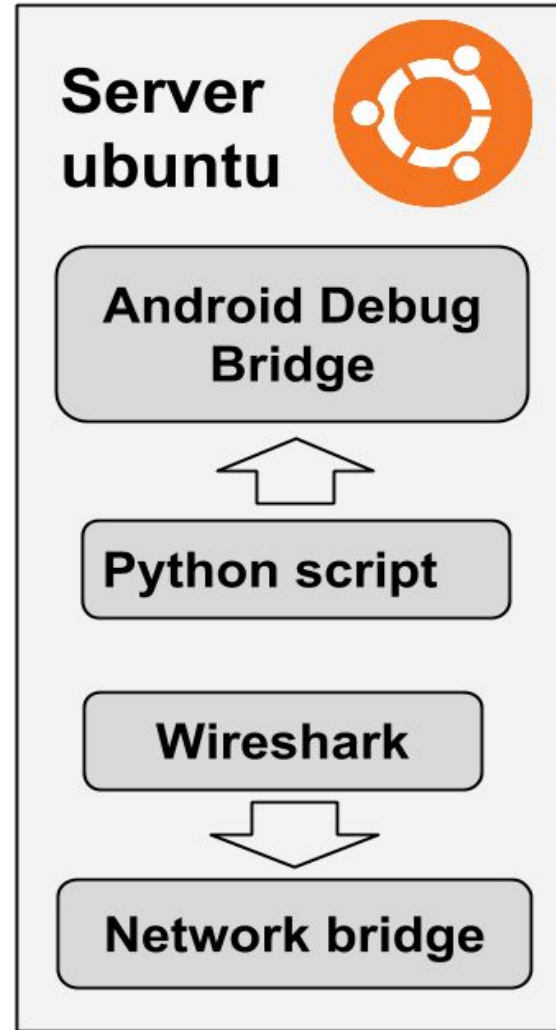


SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP

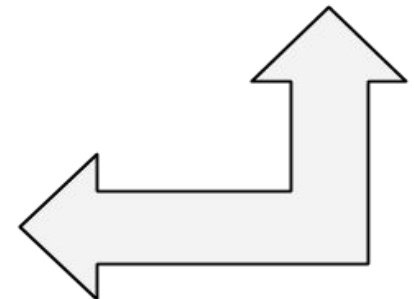


UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

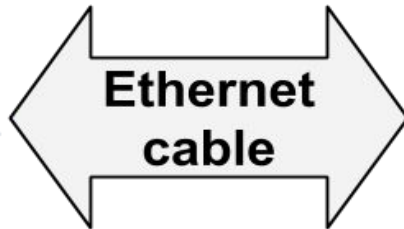
## Dataset collection



**The Internet**

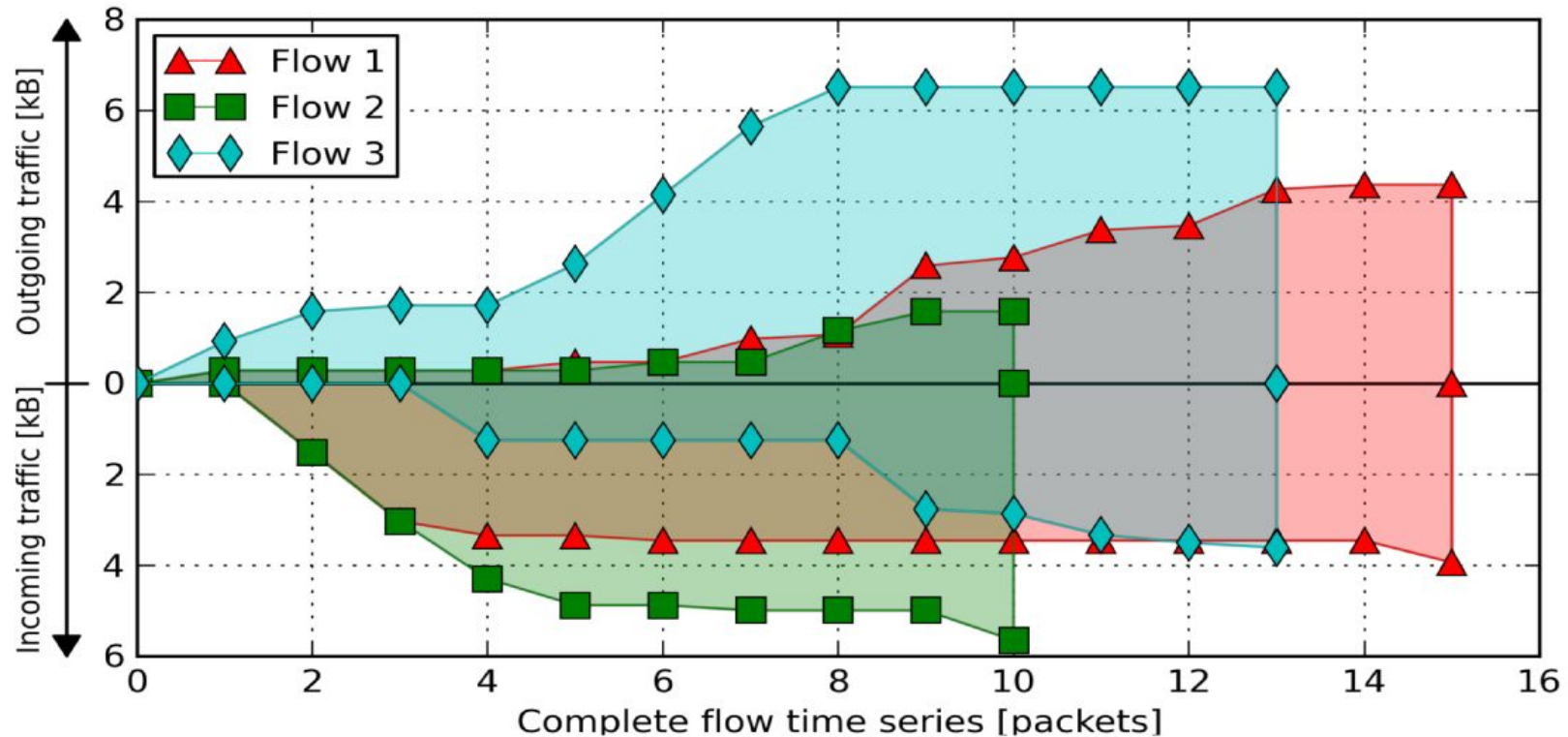


**Access Point**





## Network Traffic Flows Representation



Flow ID	Flow time series
Flow 1	[282, -1514, -1514, -315, 188, -113, 514, 96, 1514, 179, 603, 98, 801, 98, -477]
Flow 2	[282, -1514, -1514, -1266, -582, 188, -113, 692, 423, -661]
Flow 3	[926, 655, 136, -1245, 913, 1514, 1514, 863, -1514, -107, -465, -172, -111]

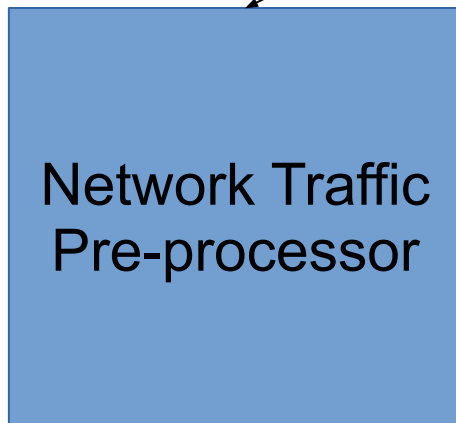
## The framework

**Labeled Dataset**



**Phase 1.  
Training**

**Phase 2.  
Testing**

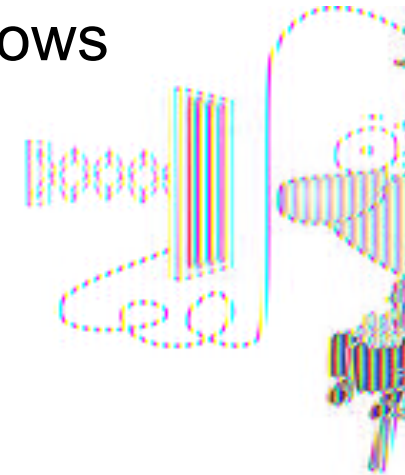


**Predictions**

- Tweet sent
- email answered
- tweeter contact opened...

## Training phase

1. Unsupervised learning → **Clusters** of similar flows
  - **Dynamic Time Warping** (DTW) [Müller 2007] as metric
  - The **number of clusters** is a parameter to tune
2. Training set building
  - User actions → Classes
  - Cluster labels → Features



IDs	user actions	cluster 0	cluster 1	...	cluster k	...	cluster N-1	cluster N
001	send mail	0	1	...	1	...	0	0
002	send mail	0	1	...	1	...	0	0
003	send reply	1	0	...	2	...	1	0
....	...	...	...	...	...	...	...	...

3. Supervised learning → Random Forest **classifier**

## Evaluation phase

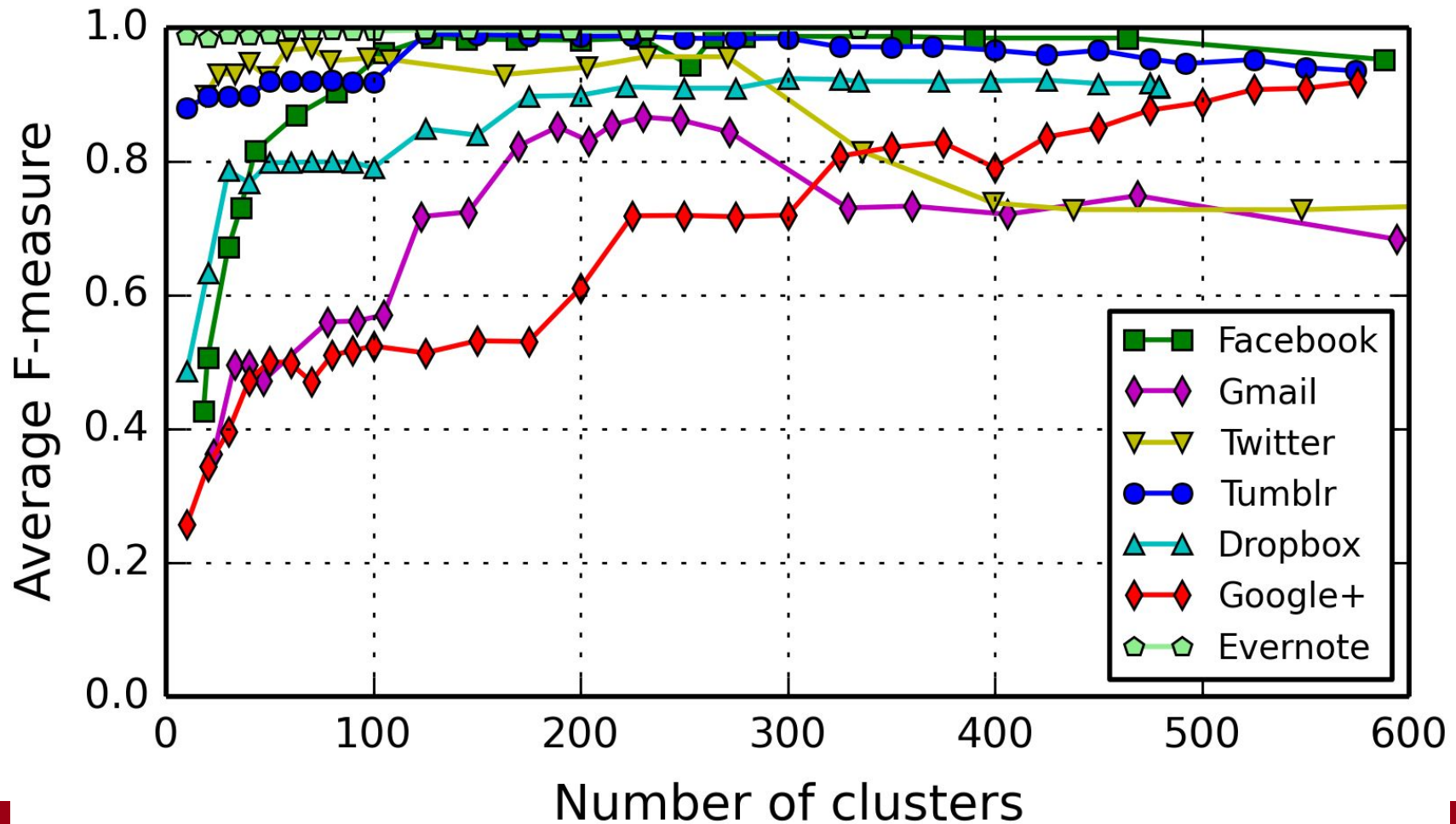
1. User actions produce **unseen flows**
2. Assign each **unseen flow** to a **cluster**
  - clusters used in **training** phase and **DTW** as metric
3. Test set building
  - (similarly to training set)
  - User actions → **unknown classes**
  - Cluster labels → Features
4. User action **recognition**



© Ron Leishman \* [www.ClipartOf.com/439797](http://www.ClipartOf.com/439797)



## Accuracy vs. number of clusters

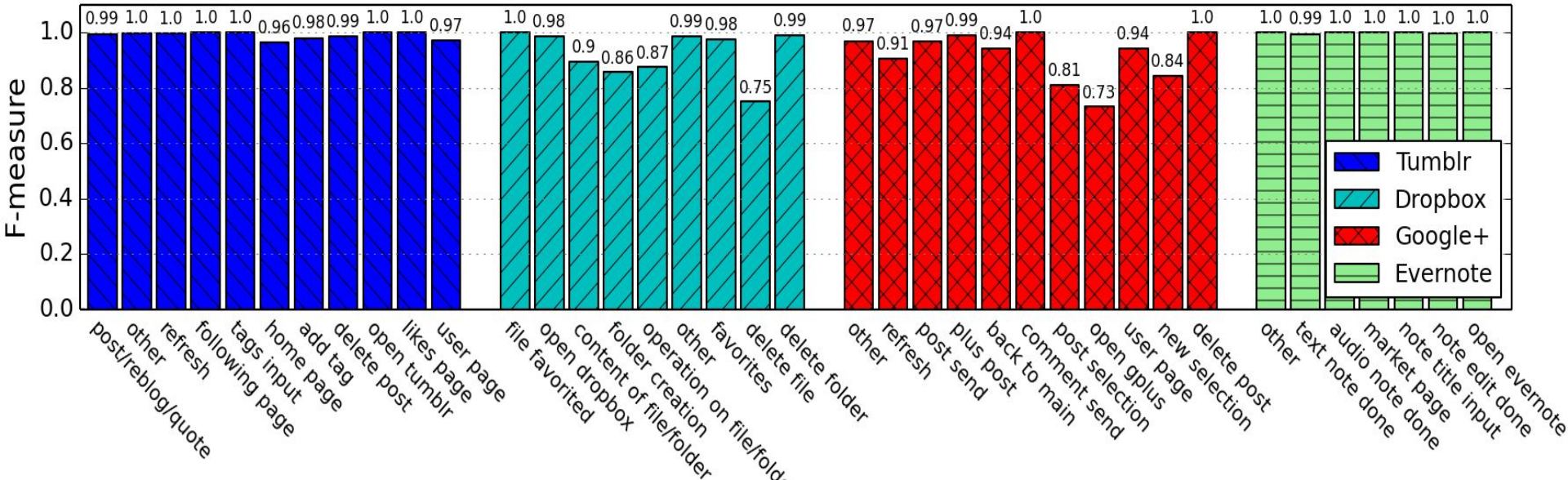
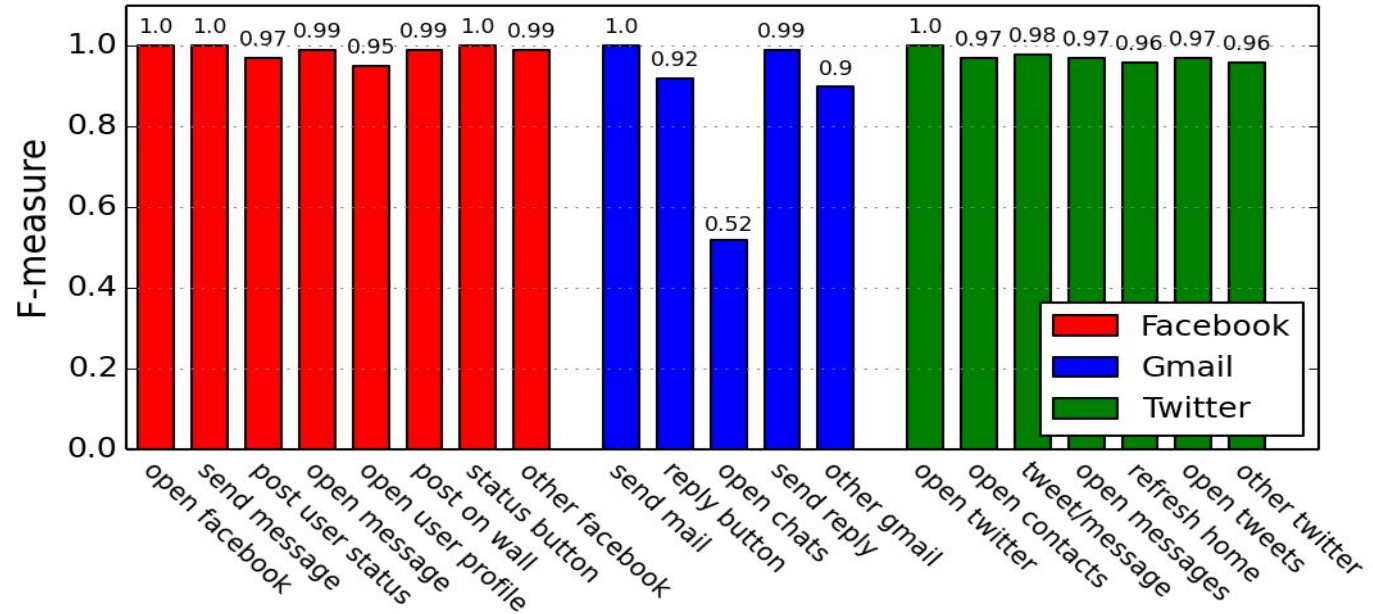




# Can't you hear me knocking (CODASPY '14, TIFS '15)



## Accuracy per user action





## Conclusions

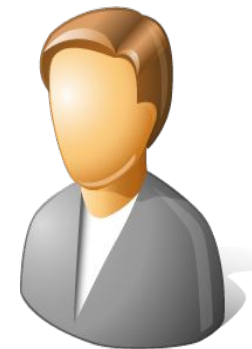
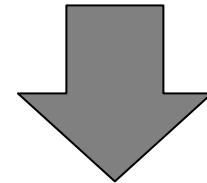
- Encryption does not hide communication patterns
  - We shown that user actions performed on Android apps can be detected by analyzing the encrypted network traffic
- Attackers can leverage our framework to undermine user privacy:
  - Learn user habits
  - Gain commercial or intelligence advantage against some competitor
  - Attribution of social network pseudonyms
- Countermeasures to this type of attacks are needed...



## Motivation (1)

From the set of **apps installed** on a device can be inferred private information about her **owner**:

- Age
- Sex
- Religion
- Relationship status
- Spoken languages
- Countries of interest

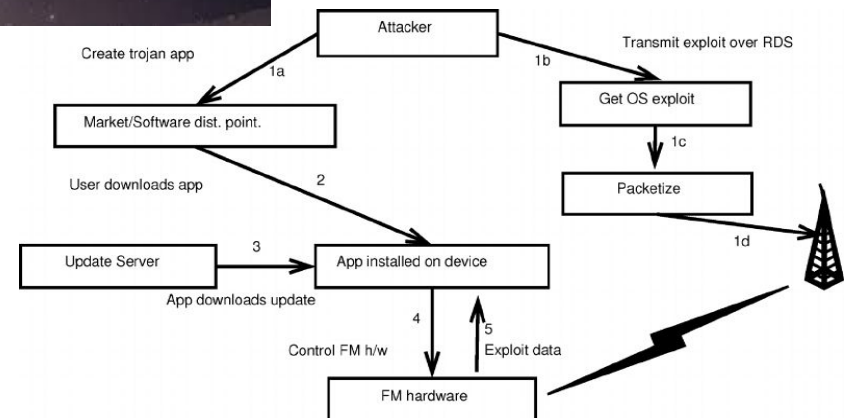
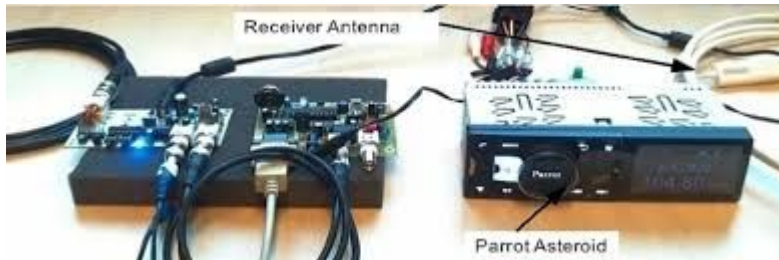
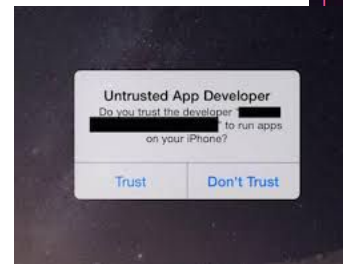


*S. Seneviratne, A. Seneviratne, P. Mohapatra, A. Mahanti. "Predicting User Traits From a Snapshot of Apps Installed on a Smartphone" in ACM SIGMOBILE Mobile Computing and Communications Review 2014.*

## Motivation (2)

Knowing a presence of a specific app  
Hence specific vulnerabilities

Possible ad-hoc attacks  
E.g., zero day exploits





## Motivation

- Given a target app X
- Identify the presence of X in a mobile device
- Using network traffic analysis



## Motivation

- Given a target app X
- Identify the presence of X in a mobile device
- Using network traffic analysis

**It isn't so easy!**



## Motivation

- Given a target app X
- Identify the presence of X in a mobile device
- Using network traffic analysis

**It isn't so easy!**

- Encryption → Payload inspection is not feasible



## Motivation

- Given a target app X
- Identify the presence of X in a mobile device
- Using network traffic analysis

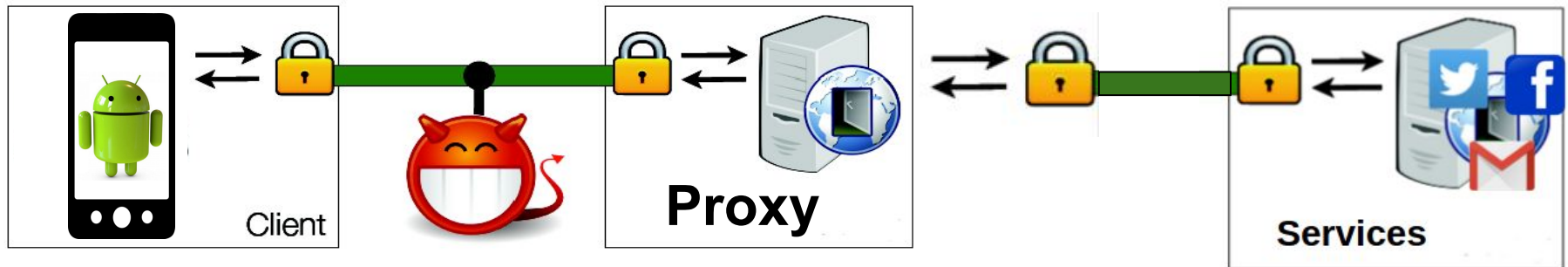
**It isn't so easy!**

- Encryption → Payload inspection is not feasible
- Owner of Destination IP  $\neq$  App
  - Content Delivery Network (CDN)
  - Proxy

## Attacker's observations (similarly to the previous work)

- Packet length
- Packet directions
- Packet timings

Enable Traffic  
Analysis Attacks





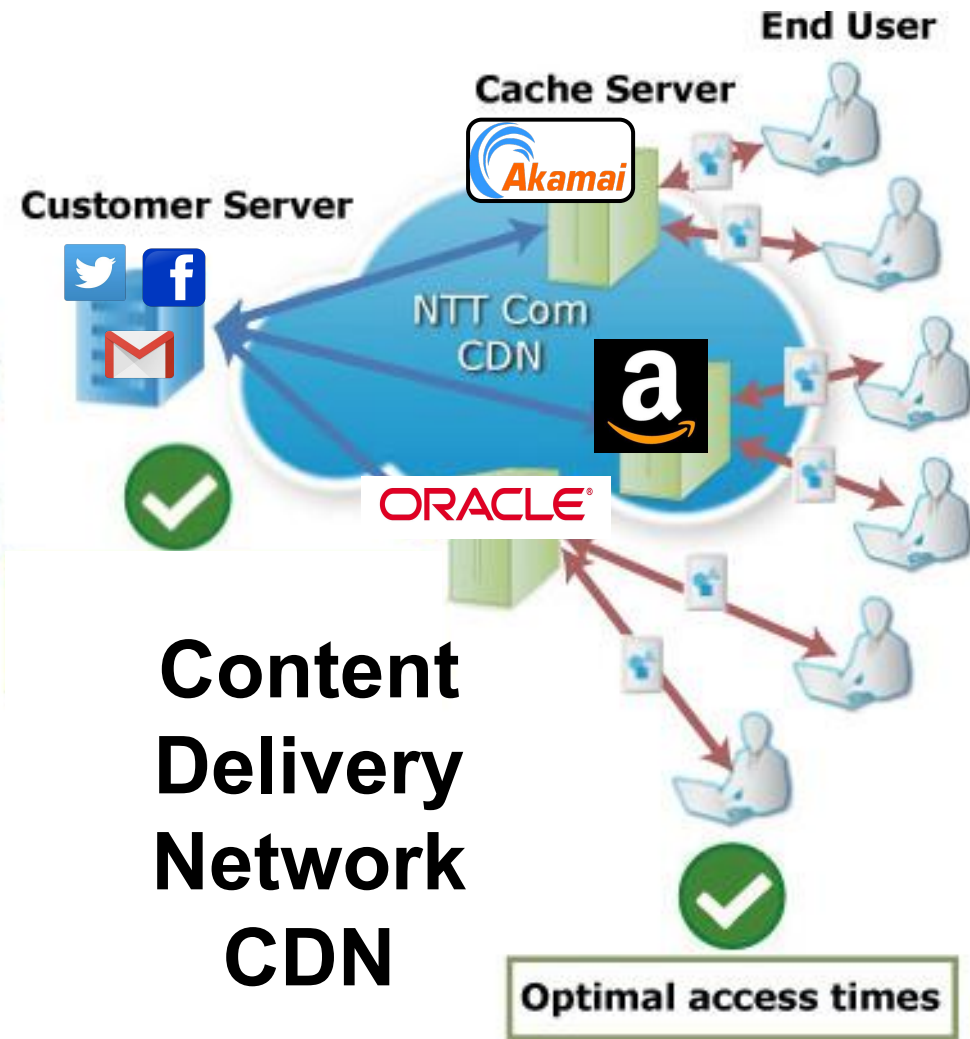
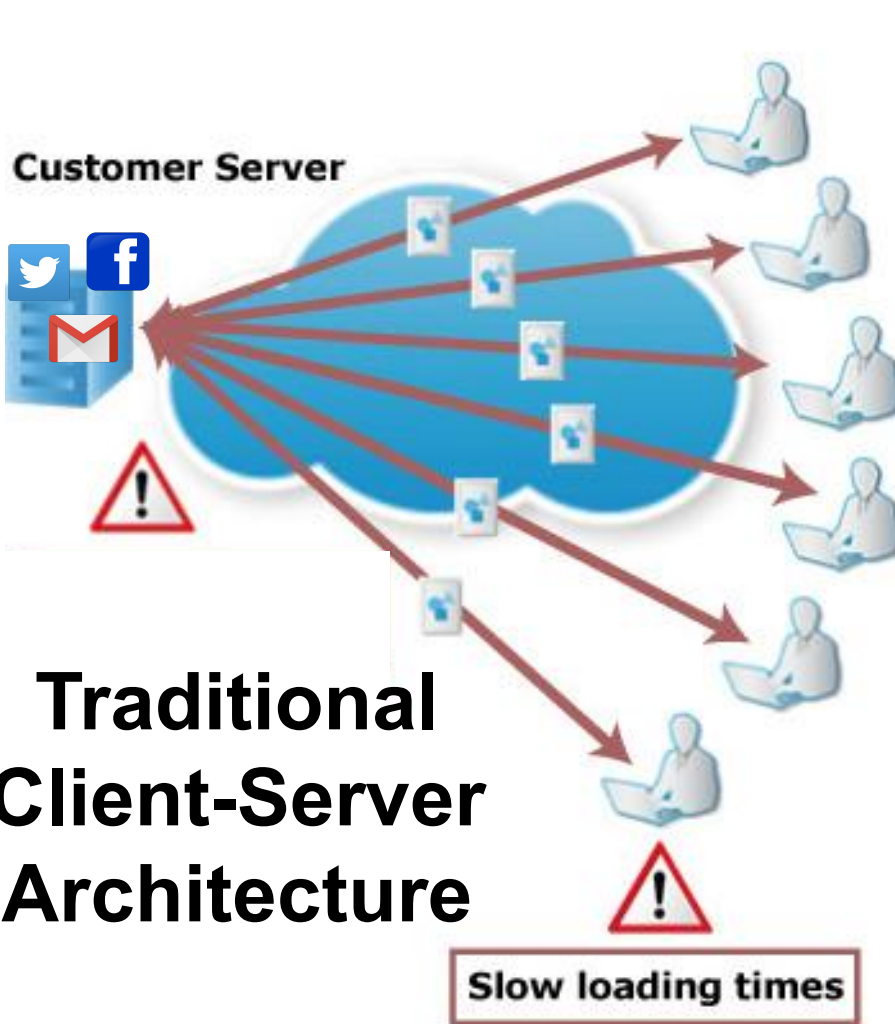
# AppScanner (IEEE EuroS&P '16)



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



**Three** different approaches proposed:

## 1. **Per flow Multi-class** length classification

- A classifier for each length
- No out-of-order packets resiliency, but fast



**Three** different approaches proposed:



**Three** different approaches proposed:

## 1. **Per flow Multi-class** length classification

- A classifier for each length
- No out-of-order packets resiliency, but fast

## 2. **Large Multi-class** classification

- Uses statistics on network flows
- It works on a **set of apps**
- **High Accuracy** and out-of-order packets resiliency, but slow



**Three** different approaches proposed:

## 1. **Per flow Multi-class** length classification

- A classifier for each length
- No out-of-order packets resiliency, but fast

## 2. **Large Multi-class** classification

- Uses statistics on network flows
- It works on a **set of apps**
- **High Accuracy** and out-of-order packets resiliency, but slow

## 3. **Per App** classification

- Uses statistics on network flows
- It focuses on a **specific app**
- Binary classification (app is present or not)





## Building the dataset

### TCP Packets captured

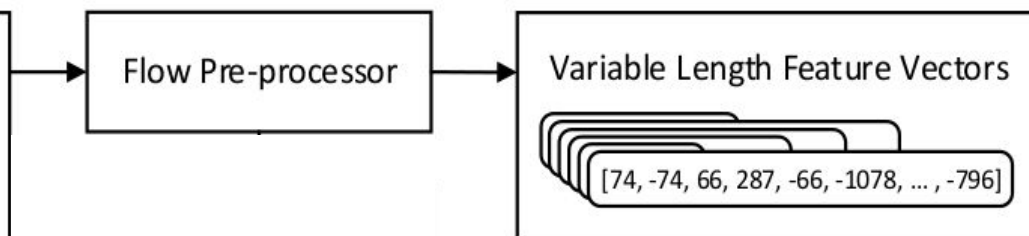
SOURCE_IP	DEST_IP	PROTO	LEN
192.168.137.2	23.23.162.140	TCP	74
23.23.162.140	192.168.137.2	TCP	74
192.168.137.2	23.23.162.140	TCP	66
192.168.137.2	23.23.162.140	TLSv1	287
23.23.162.140	192.168.137.2	TCP	66
23.23.162.140	192.168.137.2	TLSv1	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	114
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TLSv1	796



## Building the dataset

### TCP Packets captured

SOURCE_IP	DEST_IP	PROTO	LEN
192.168.137.2	23.23.162.140	TCP	74
23.23.162.140	192.168.137.2	TCP	74
192.168.137.2	23.23.162.140	TCP	66
192.168.137.2	23.23.162.140	TLSv1	287
23.23.162.140	192.168.137.2	TCP	66
23.23.162.140	192.168.137.2	TLSv1	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	114
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TLSv1	796

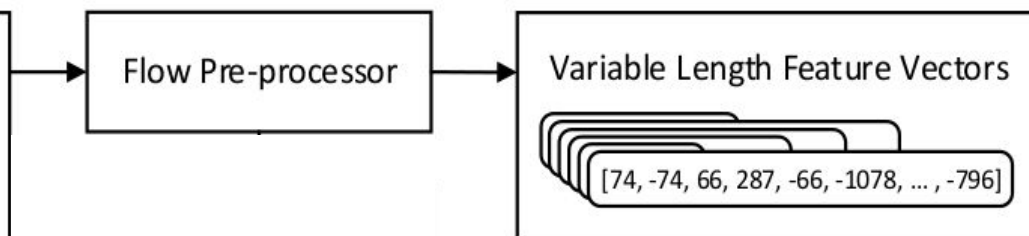




## Building the dataset

### TCP Packets captured

SOURCE_IP	DEST_IP	PROTO	LEN
192.168.137.2	23.23.162.140	TCP	74
23.23.162.140	192.168.137.2	TCP	74
192.168.137.2	23.23.162.140	TCP	66
192.168.137.2	23.23.162.140	TLSv1	287
23.23.162.140	192.168.137.2	TCP	66
23.23.162.140	192.168.137.2	TLSv1	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	114
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TLSv1	796



### Per Flow approach (1)

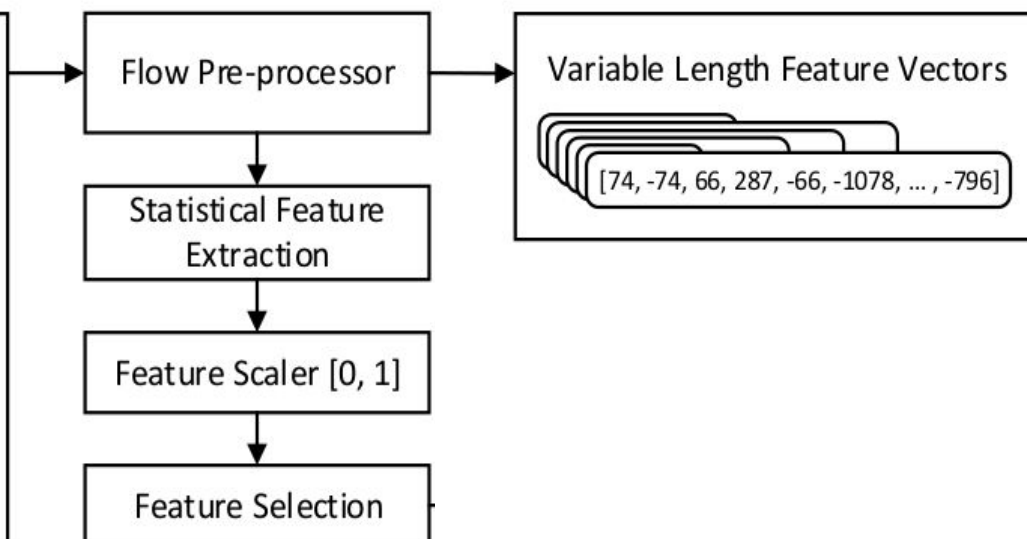
Variable Length Feature Vectors

[74, -74, 66, 287, -66, -1078, ..., -796]

## Building the dataset

### TCP Packets captured

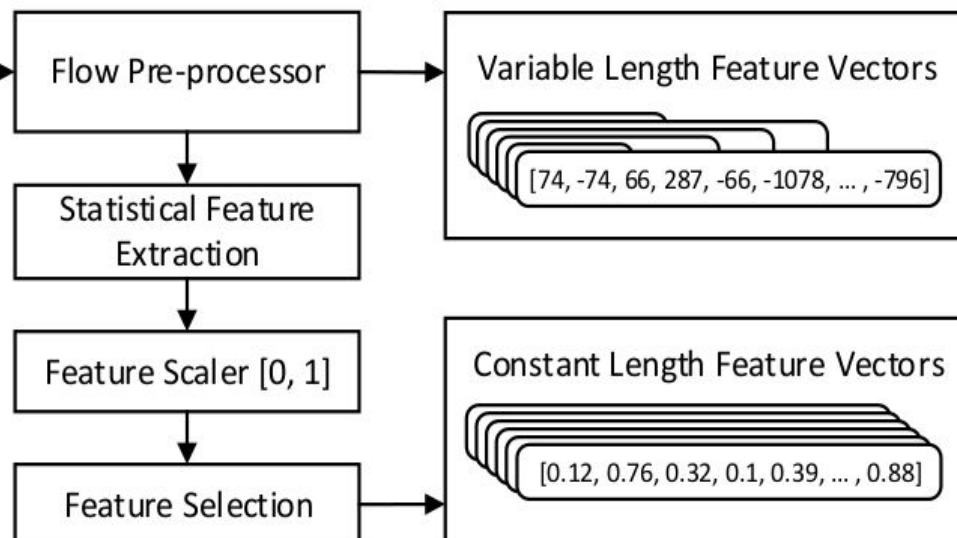
SOURCE_IP	DEST_IP	PROTO	LEN
192.168.137.2	23.23.162.140	TCP	74
23.23.162.140	192.168.137.2	TCP	74
192.168.137.2	23.23.162.140	TCP	66
192.168.137.2	23.23.162.140	TLSv1	287
23.23.162.140	192.168.137.2	TCP	66
23.23.162.140	192.168.137.2	TLSv1	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	114
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TLSv1	796



## Building the dataset

### TCP Packets captured

SOURCE_IP	DEST_IP	PROTO	LEN
192.168.137.2	23.23.162.140	TCP	74
23.23.162.140	192.168.137.2	TCP	74
192.168.137.2	23.23.162.140	TCP	66
192.168.137.2	23.23.162.140	TLSv1	287
23.23.162.140	192.168.137.2	TCP	66
23.23.162.140	192.168.137.2	TLSv1	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	114
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TLSv1	796



## Building the dataset

### TCP Packets captured

SOURCE_IP	DEST_IP	PROTO	LEN
192.168.137.2	23.23.162.140	TCP	74
23.23.162.140	192.168.137.2	TCP	74
192.168.137.2	23.23.162.140	TCP	66
192.168.137.2	23.23.162.140	TLSv1	287
23.23.162.140	192.168.137.2	TCP	66
23.23.162.140	192.168.137.2	TLSv1	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TCP	114
23.23.162.140	192.168.137.2	TCP	1078
23.23.162.140	192.168.137.2	TLSv1	796

Flow Pre-processor

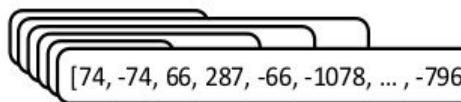
Statistical Feature  
Extraction

Feature Scaler [0, 1]

Feature Selection

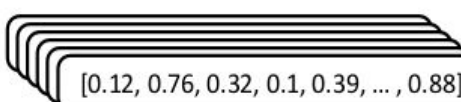
### Per Flow approach (1)

Variable Length Feature Vectors



[74, -74, 66, 287, -66, -1078, ..., -796]

Constant Length Feature Vectors



[0.12, 0.76, 0.32, 0.1, 0.39, ..., 0.88]

### Statistical approaches (2, 3)

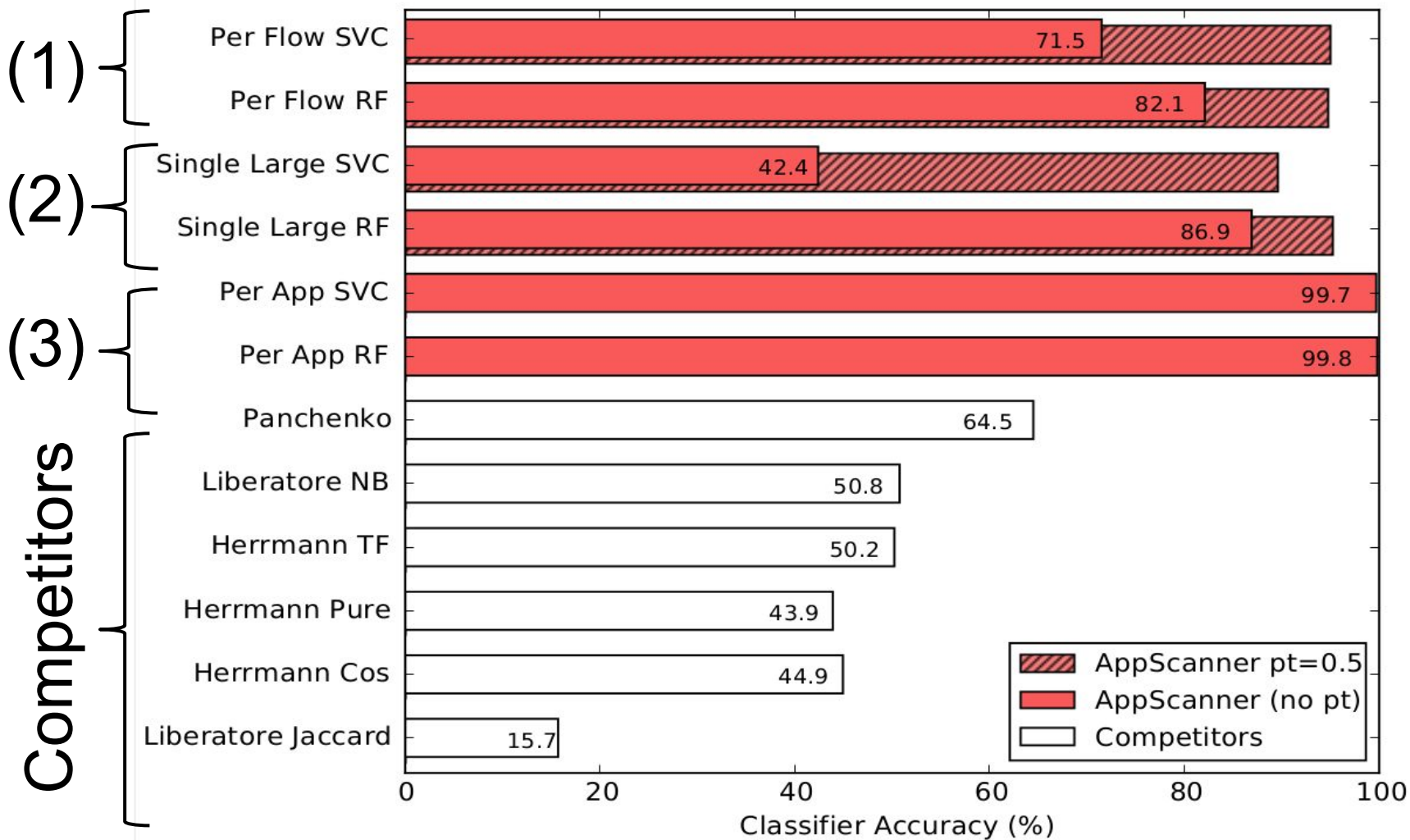
## Improving the accuracy of AppScanner

- Classification performed on **each** network traffic flow
- We aim to identify an app → many flows available
- Flow → Classifier prediction → (App, Probability of prediction)
- Applying a **probability threshold (PT)**
  - Filter out flows with **uncertain predictions**
  - Increase classification accuracy tuning PT





## Performance and Comparison





- Covert and Side Channels 101
- Network Traffic Analysis
  - *As a side channel: app and sensitive data inference*
- **Energy Consumption**
  - *As a side channel: user and app inference*
  - *As a covert channel: data exfiltration*
- Device Movement
  - *As a side channel: smartphone user authentication*
  - *Attacks against biometric authentication*
- Keystroke Timing
  - *As a side channel: text typed on keyboards*
- Acoustic Emanations
  - *As a side channel: text typed on keyboards*





SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA





M. Conti, M. Nati, E. Rotundo, R. Spolaor.

**Mind The Plug! Laptop-User Recognition Through Power  
Consumption.**

*In ACM AsiaCCS 2016 workshop IoTPTS 2016*

# Power Consumption Side Channel



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



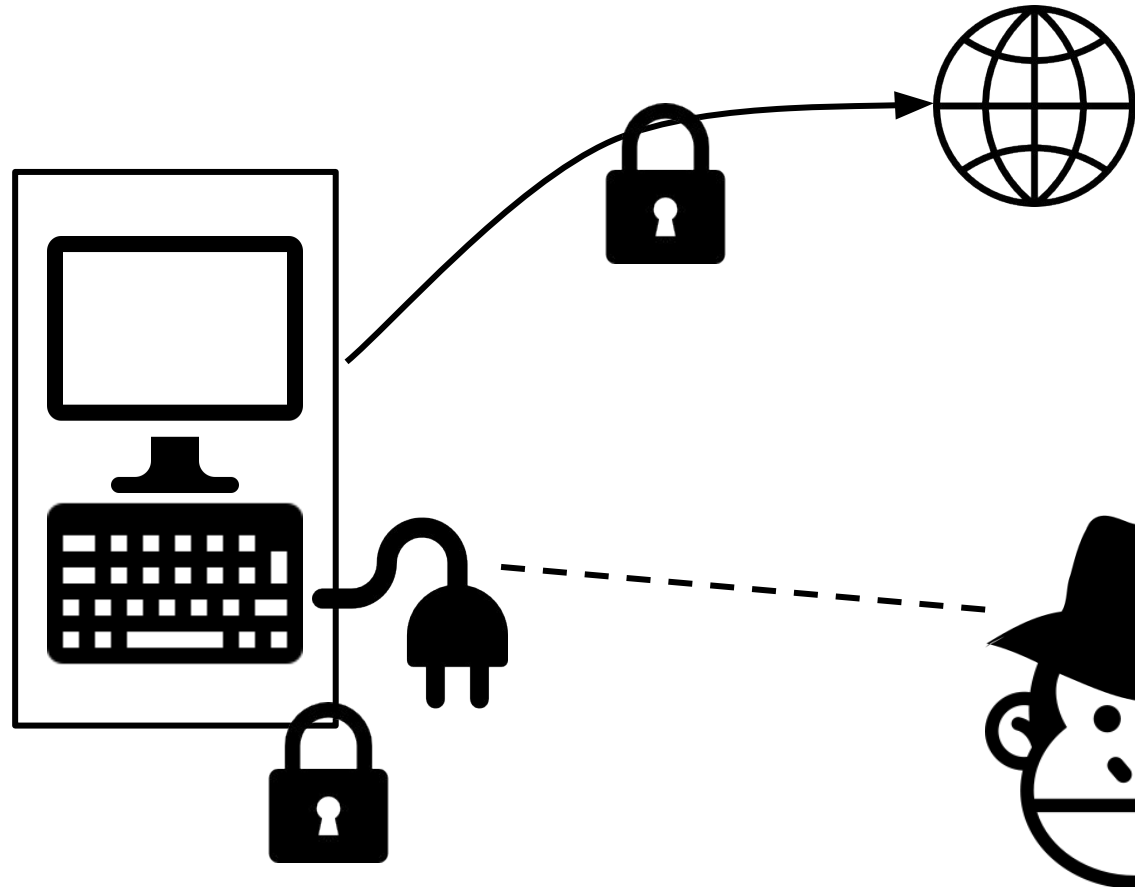
UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

Power consumption

*Can reveal what we are doing!*

Device drains different power  
depending on our actions

Works on **laptops** and  
**mobile**

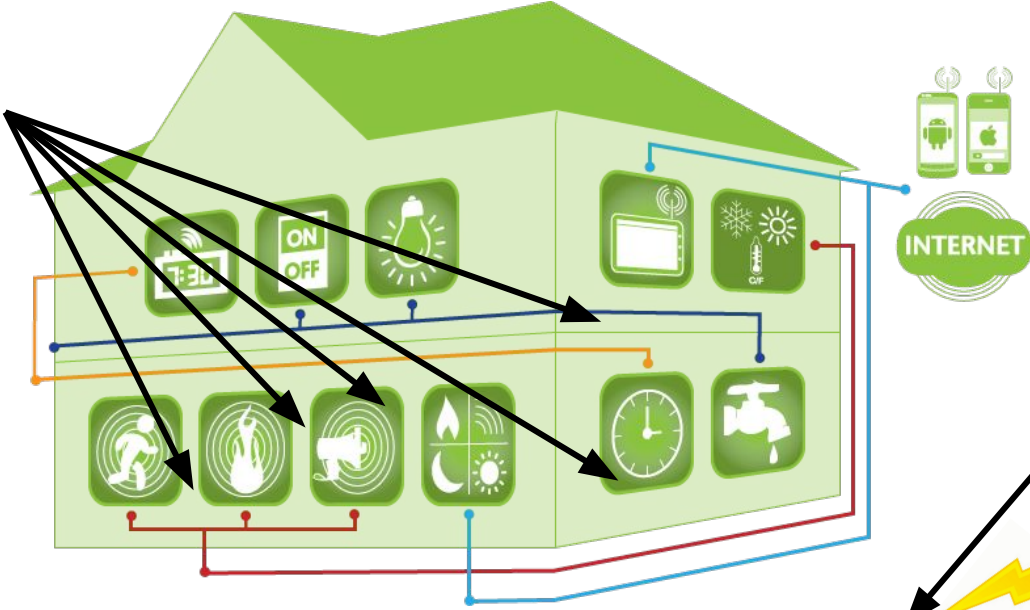




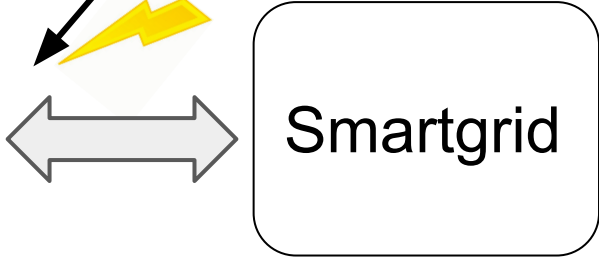
## Smartbuilding

Internet of Things applied not only to industry, but also to buildings, such as houses and **offices**

Wall-socket  
level  
sensors



household  
level  
sensors



## Wall-socket smartmeters

- Smartmeters are able to measure the electric quantities of the plugged appliances
  - **Reactive Power**
  - **RMS Current**
  - **Voltage**
  - **Phase**
- IoT testbed in University of Surrey (UK)
- Limitation:
  - only **1Hz** of sampling rate



## Definition of “Laptop-User”

A **Laptop-user** is made of the **combination** of:

- Laptop
- Software installed and running
- User behavior





## Goal & Motivation

Is it possible to recognize a **Laptop-user** from its energy consumption?

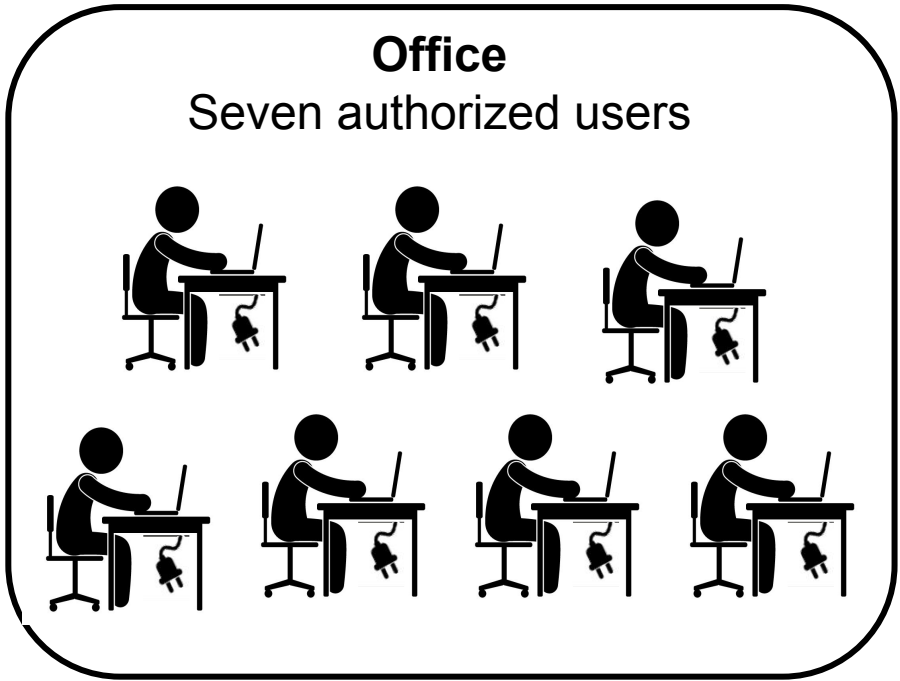
This can bring:

- **Benefit on smartbuilding automation,**
  - context-aware environments can automatically adjust and trigger predefined actions or services
    - e.g., according to the presence of a specific user
  - Detect un-authorized users
- **Threat to user privacy,**
  - it is possible to locate and trace a user

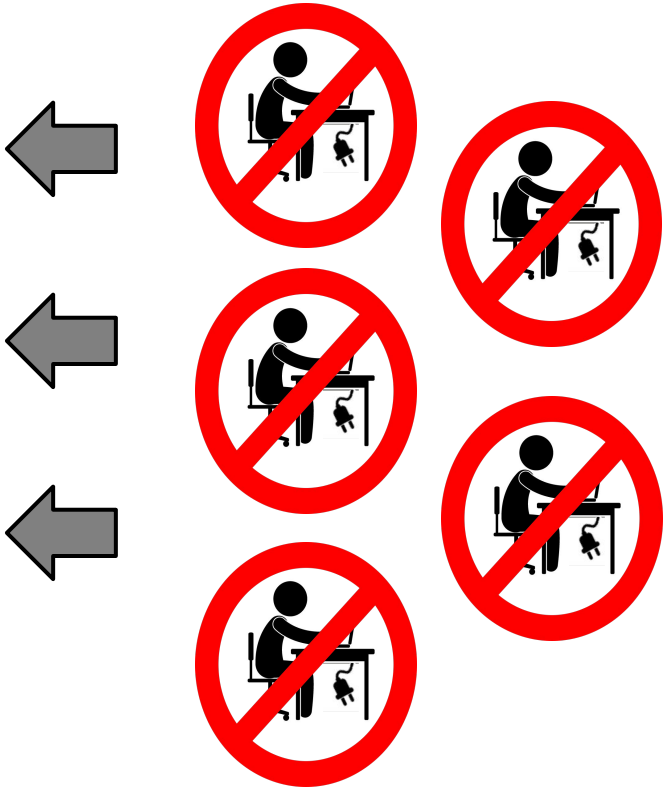




## Threat Model



Twenty unauthorized users



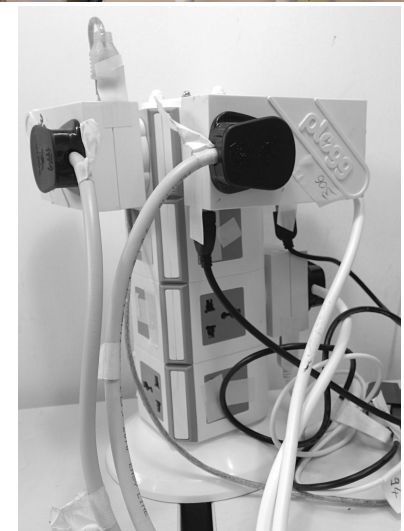
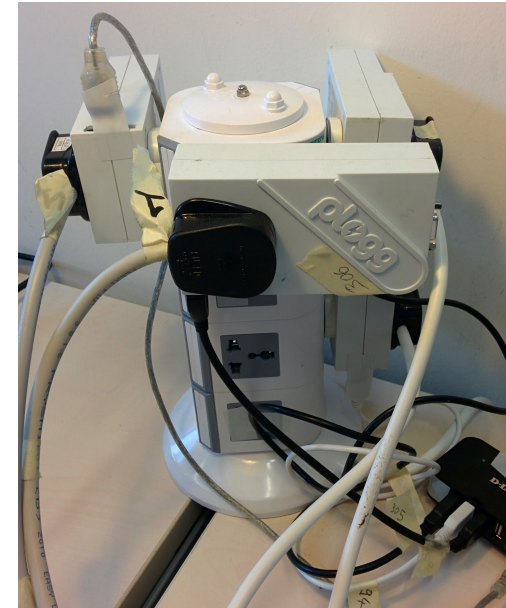
We aim to:

- Recognize whether the user is in the “authorized” set
- Identify the specific user in the “authorized” set

## Laptop-users Recognition

Multiclass classification (8 classes)

- The **seven authorized** laptop-users
- The **intruders** (as a single class)



Classification in three steps:

1. 10-fold cross validation for **parameters selection**
2. Performance **evaluation** on a disjoint test set
3. Classification **validation**

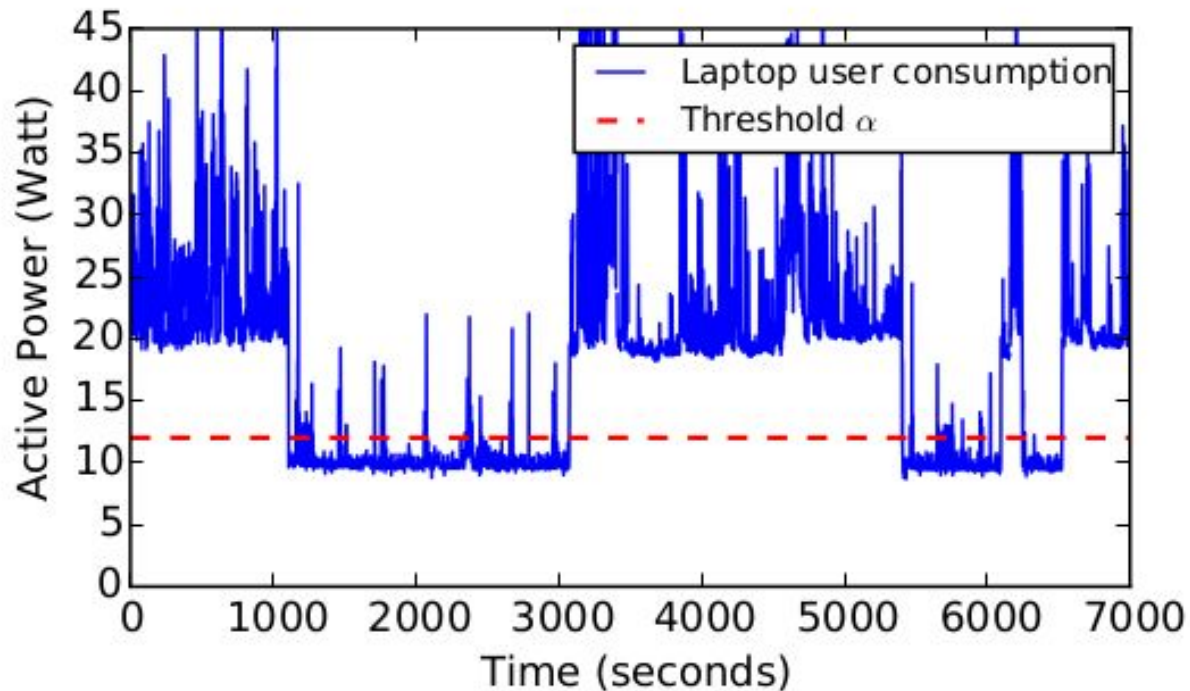
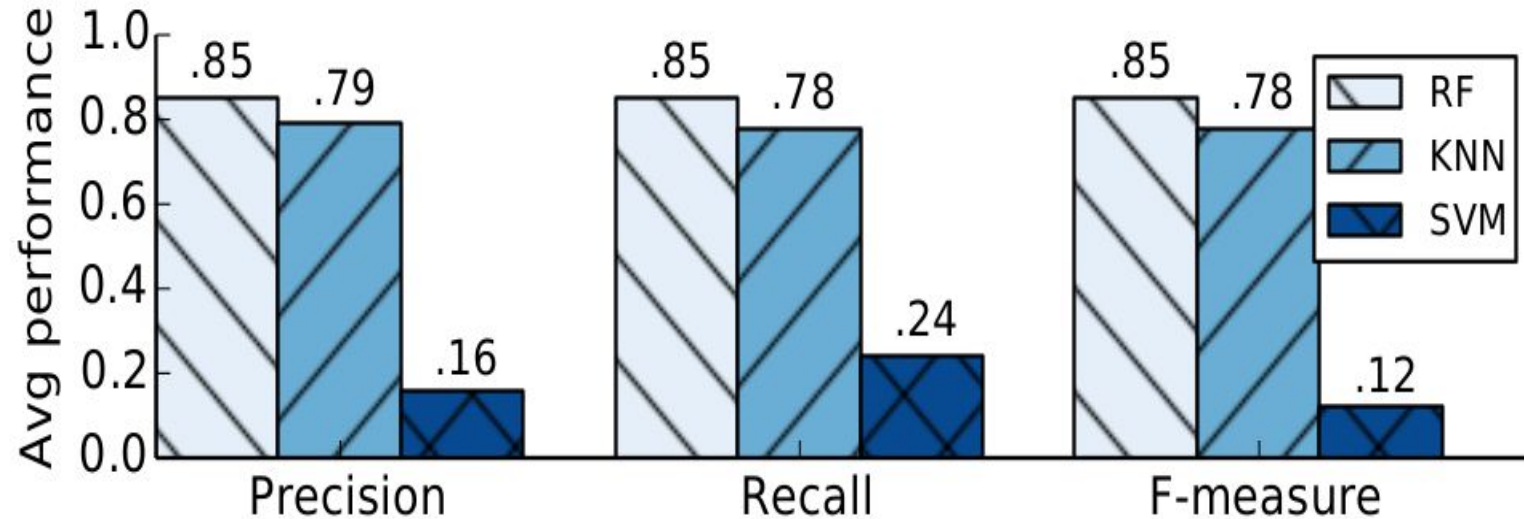


Figure 2: Example of *Active Power* trace (continuous blue line) and the lower-cutting threshold  $\alpha = 12$  Watt (dashed red line). Samples under  $\alpha$  are low-energy timespans in which the user does not use the laptop.



**85%** of F-measure with Random Forest classifier



## Classification validation

Classifiers label all segments in the testset

- **Bad for False Positive rate (FPR)**

We can leverage also the prediction probability

- Since classifiers output also their **confidence**

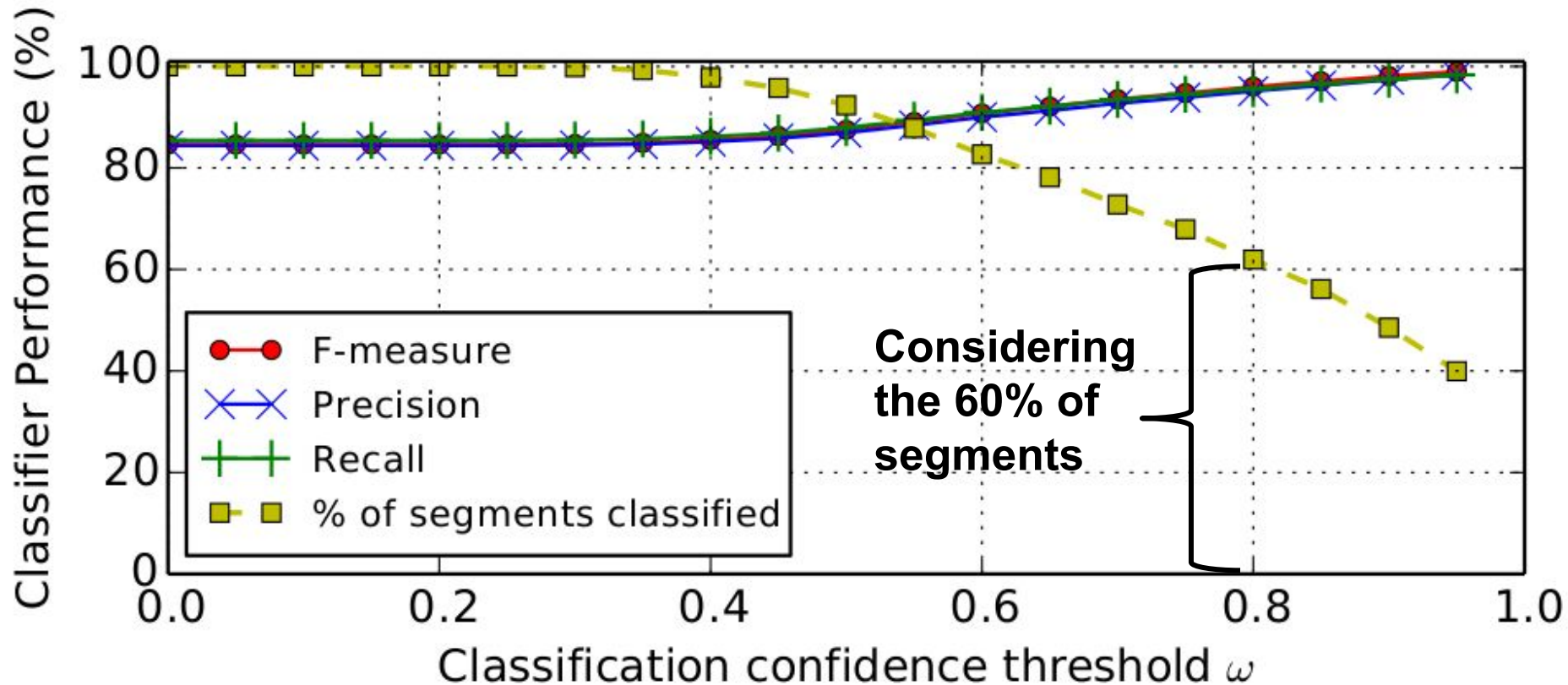
Tuning prediction probability threshold

- **It can reduce False Positives**

Other implications:

- MTPlug can be more conservative
- May take more segments to identify some laptop-user

## Classification validation results







## Limitations and Future work

### Structural limitation:

The plogg wall-socket sensors have a low sampling rate

### Solution:

Adopt a new generation wall-socket sensors

### Data limitation:

we collected data of seven users (office)

### Solution:

Collect more data in order to assess the feasibility of authentication system based on energy consumption





- Covert and Side Channels 101
- Network Traffic Analysis
  - *As a side channel: app and sensitive data inference*
- **Energy Consumption**
  - *As a side channel: user and app inference*
  - **As a covert channel: data exfiltration**
- Device Movement
  - *As a side channel: smartphone user authentication*
  - *Attacks against biometric authentication*
- Keystroke Timing
  - *As a side channel: text typed on keyboards*
- Acoustic Emanations
  - *As a side channel: text typed on keyboards*



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA





R Spolaor, L Abudahi, V Moonsamy, M Conti, R Poovendran.

**No Free Charge Theorem: a Covert Channel via USB Charging Cable  
on Mobile Devices.**

*In ACNS 2017*

*Presented at Black Hat Europe 2018*



# Power Consumption Covert Channel



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



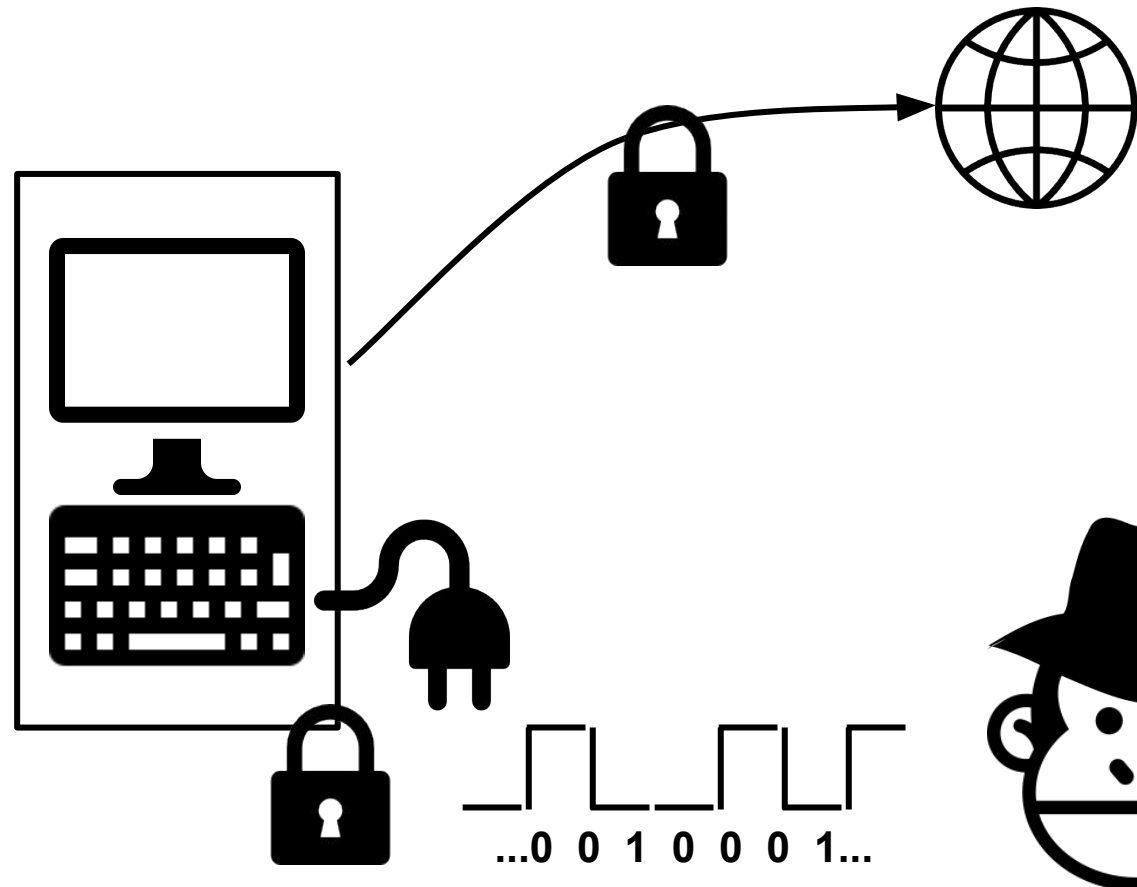
UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

Power consumption

*Can be used as a covert channel*

Malware makes device drain more/less power to communicate with a malicious power outlet

Thus exfiltrating secrets



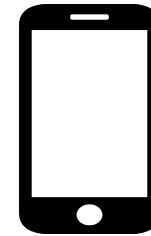
# No Free Charge Theorem: a Covert Channel via USB Charging Cable on Mobile Devices



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA





# USB protection...



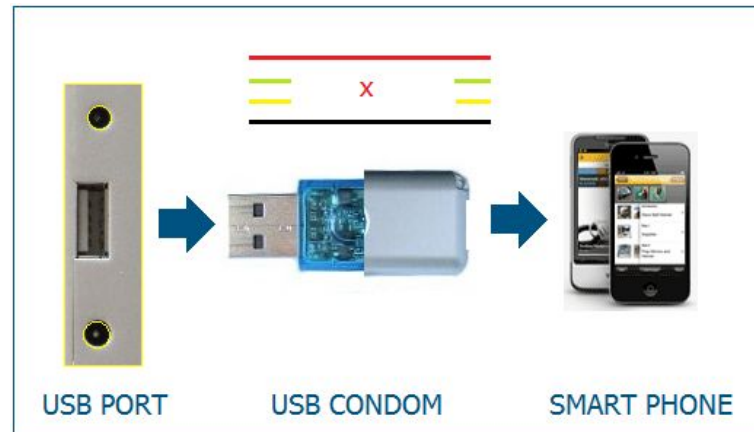
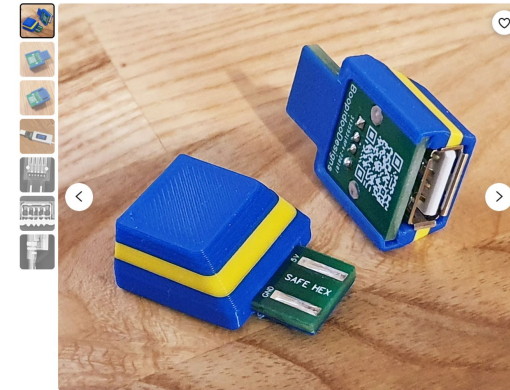
## Protect your data



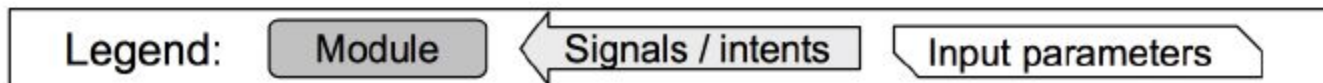
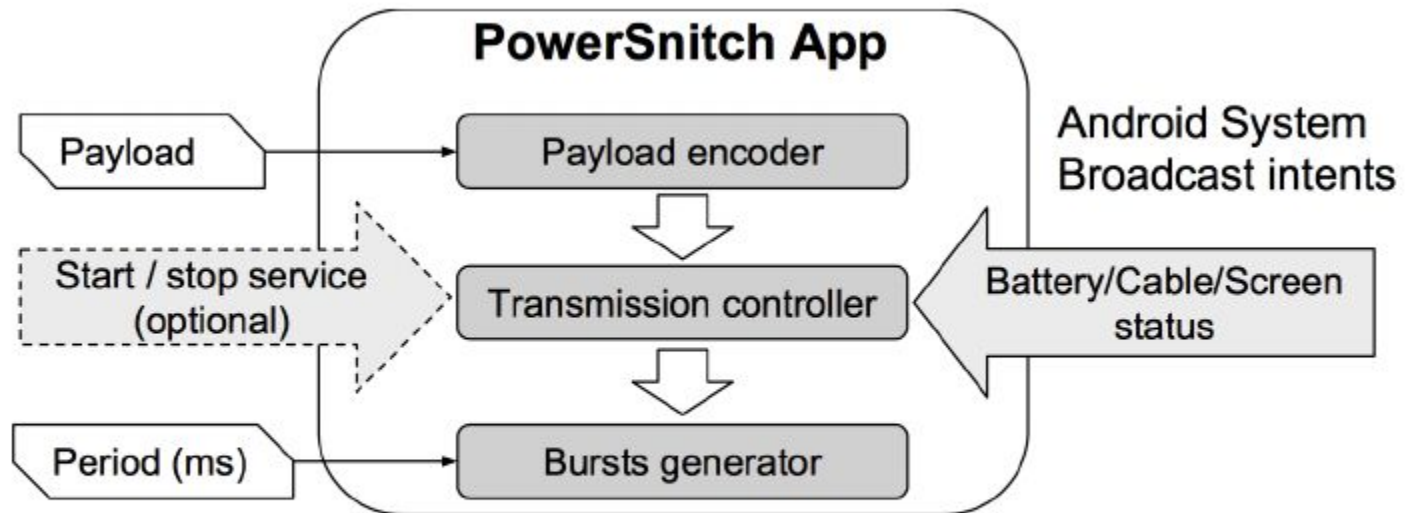
SyncStop prevents accidental data exchange when your device is plugged into someone else's computer or a public charging station. SyncStop achieves this by blocking the data pins on any USB cable and allowing only power to flow through. This minimizes opportunities to steal your data or install malware on your mobile device.

SyncStop is the 'cased' version of the original USB Condom. We listened and spent some time designing and manufacturing our own enclosure.

SyncStop works with any mobile device:

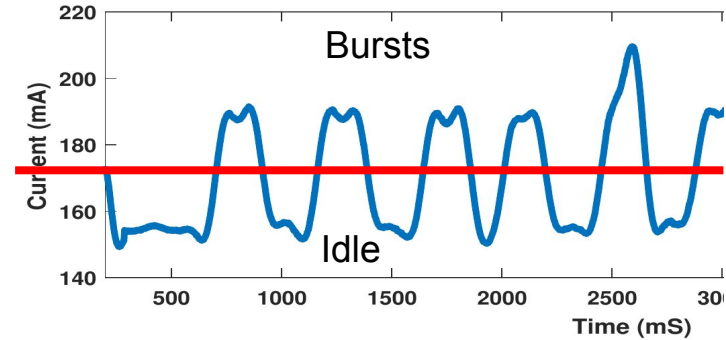
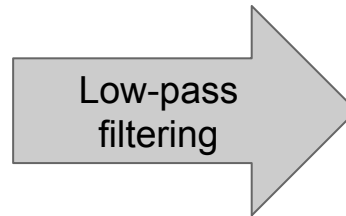
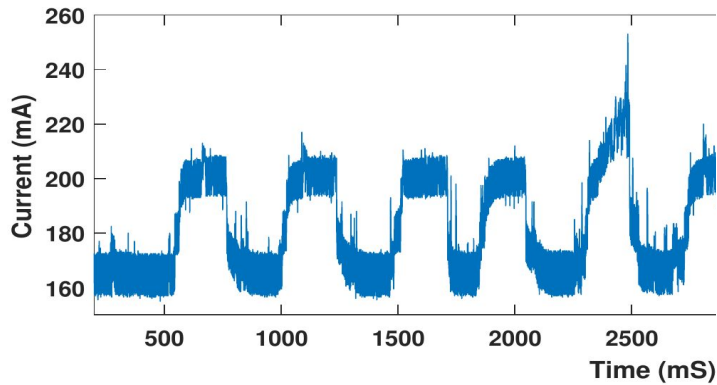


# PowerSnitch Application



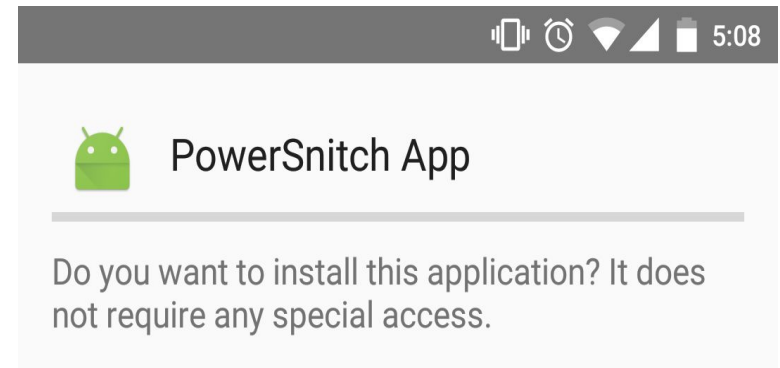


# No Free Charge Theorem: a Covert Channel via USB Charging Cable on Mobile Devices



## Results in terms of Bit Error Ratio (BER)

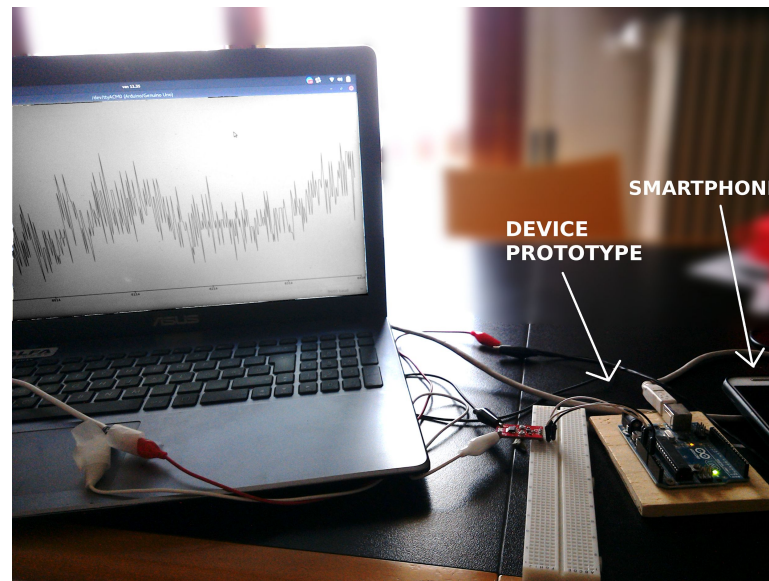
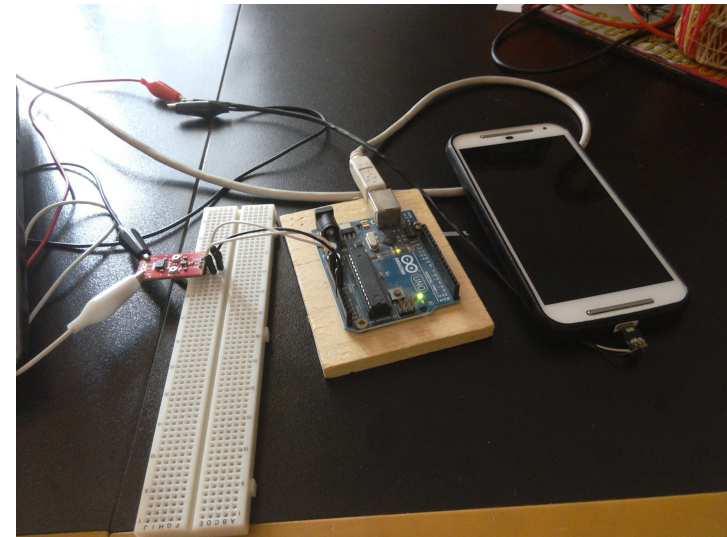
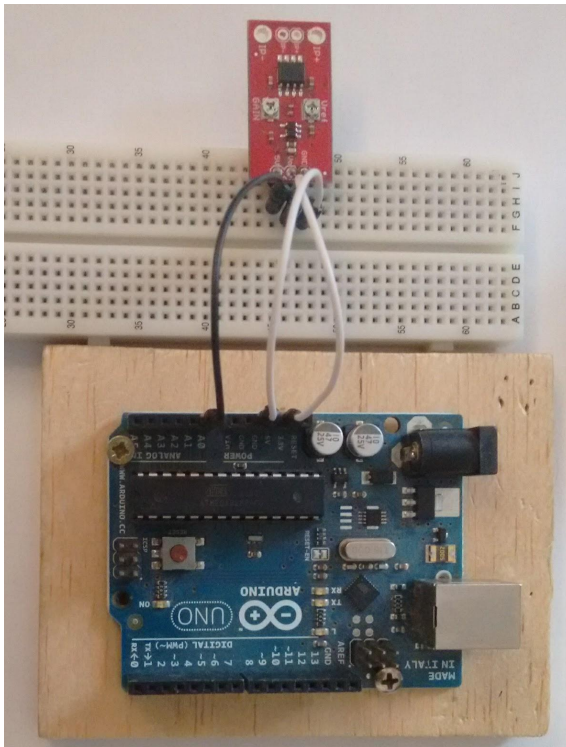
Device	Period (milliseconds)					
	1000	900	800	700	600	500
Nexus 4	13.5	0.78	0.0	0.0	13.33	16.21
Nexus 5	21.0	0.0	0.95	36.82	40.35	13.4
Nexus 6	1.07	0.0	0.21	0.0	4.05	7.42
Samsung S5	12.5	13.5	13.31	16.33	17.9	21.42



**PowerSnitch app does not require any permission !!!**



# Power Bank Prototype



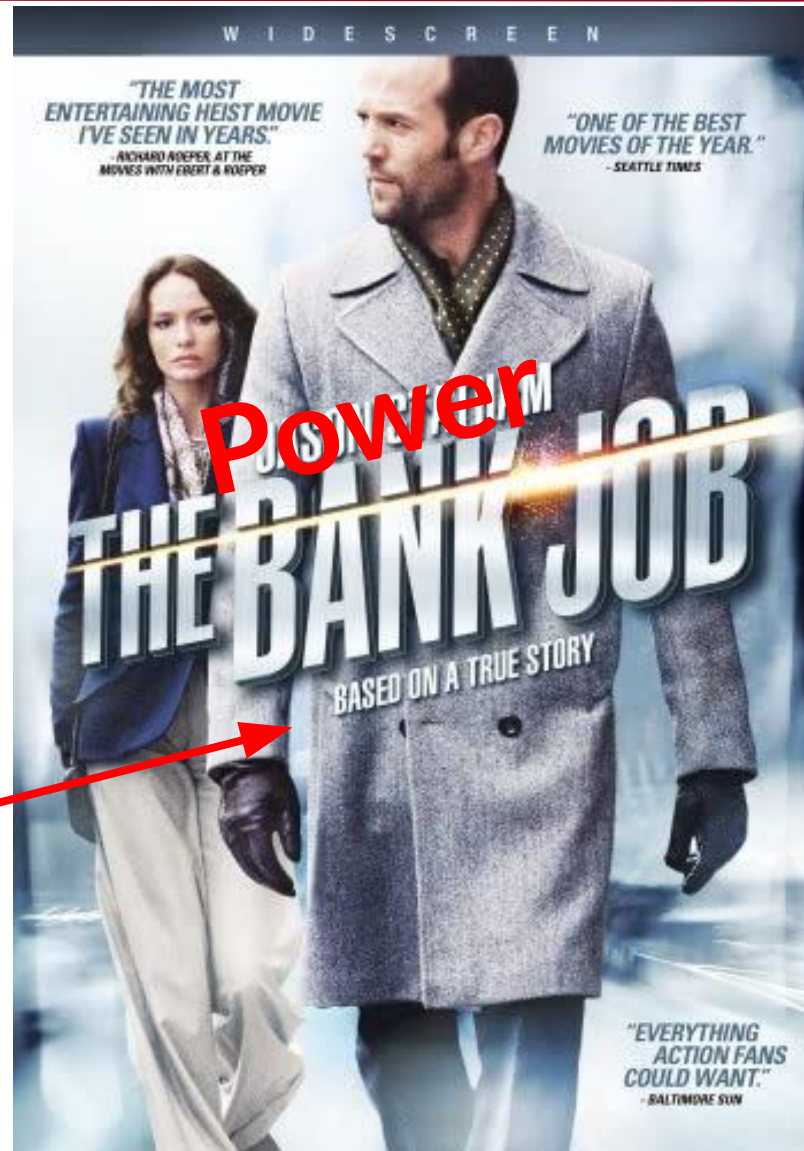
# Power Bank - DEMO TIME



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



<https://drive.google.com/file/d/1JXzoyOM3xpQqaM8exWF07htp67G5m82v/view?usp=sharing>



- Covert and Side Channels 101
- Network Traffic Analysis
  - *As a side channel: app and sensitive data inference*
- Energy Consumption
  - *As a side channel: user and app inference*
  - *As a covert channel: data exfiltration*
- **Device Movement**
  - ***As a side channel: smartphone user authentication***
  - *Attacks against biometric authentication*
- Keystroke Timing
  - *As a side channel: text typed on keyboards*
- Acoustic Emanations
  - *As a side channel: text typed on keyboards*



# Keystroke Dynamics 101



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



# Keystroke Dynamics 101



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



# Keystroke Dynamics 101



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA





# Keystroke Dynamics 101



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



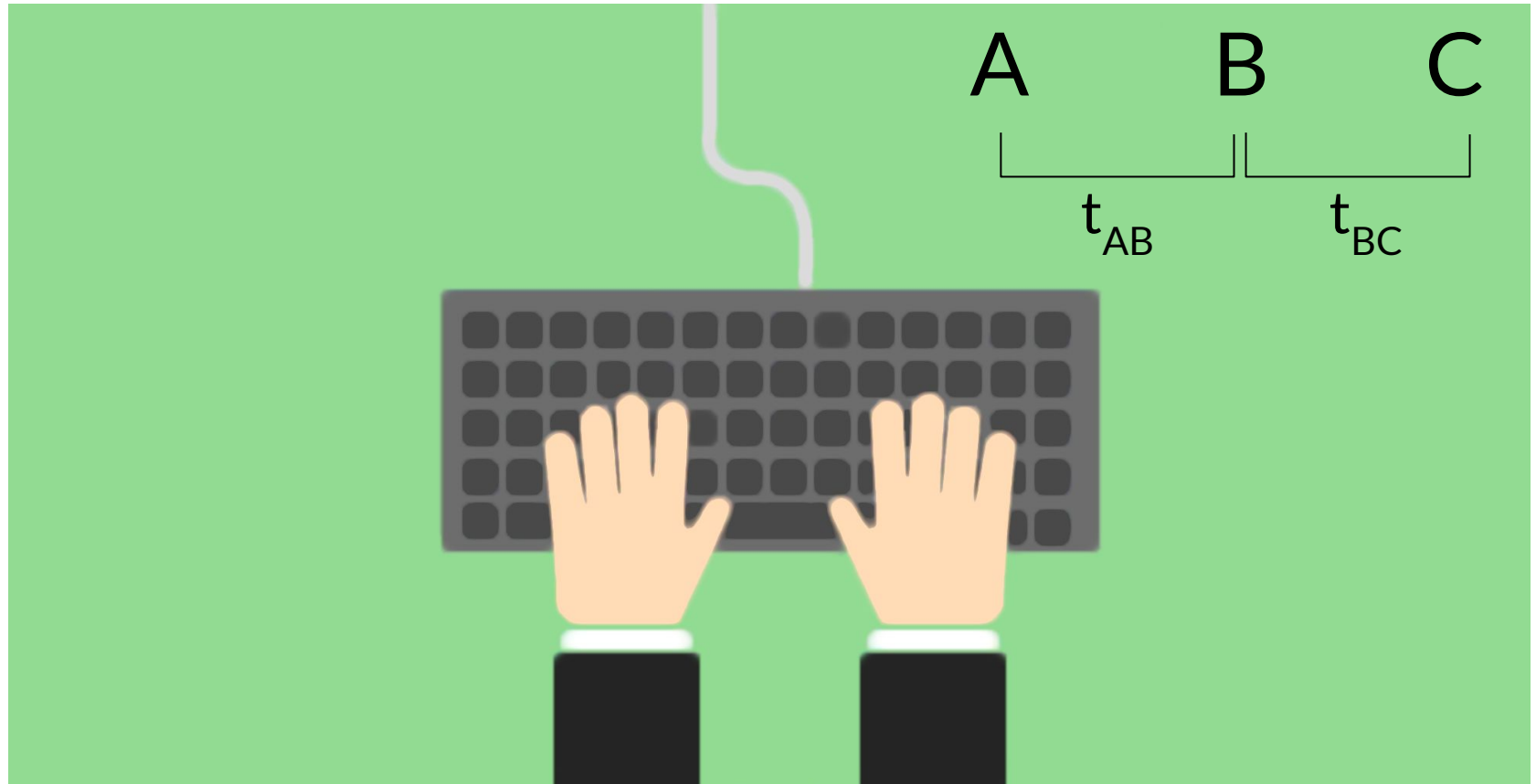
# Keystroke Dynamics 101



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



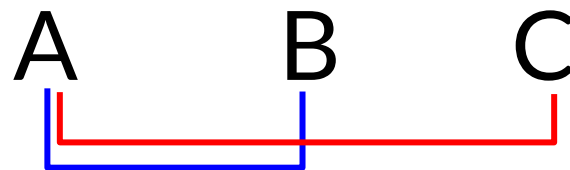
# Keystroke Dynamics 101



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



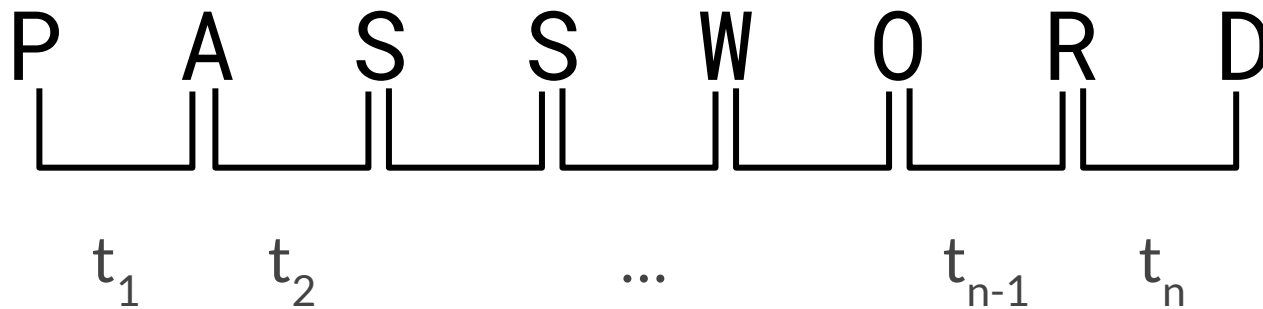
Digram

$t_{AB}$

$t_{AC}$

Trigram

# Keystroke Dynamics 101



- Inter-keystroke times as a personal *signature*
- Used as biometric in authentication systems

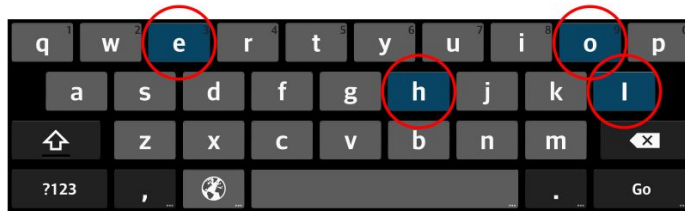


Kamil Majdanik, Cristiano Giuffrida, Mauro Conti, Herbert Bos.  
***I Sensed It Was You: Authenticating Mobile Users with  
Sensor-enhanced Keystroke Dynamics.***

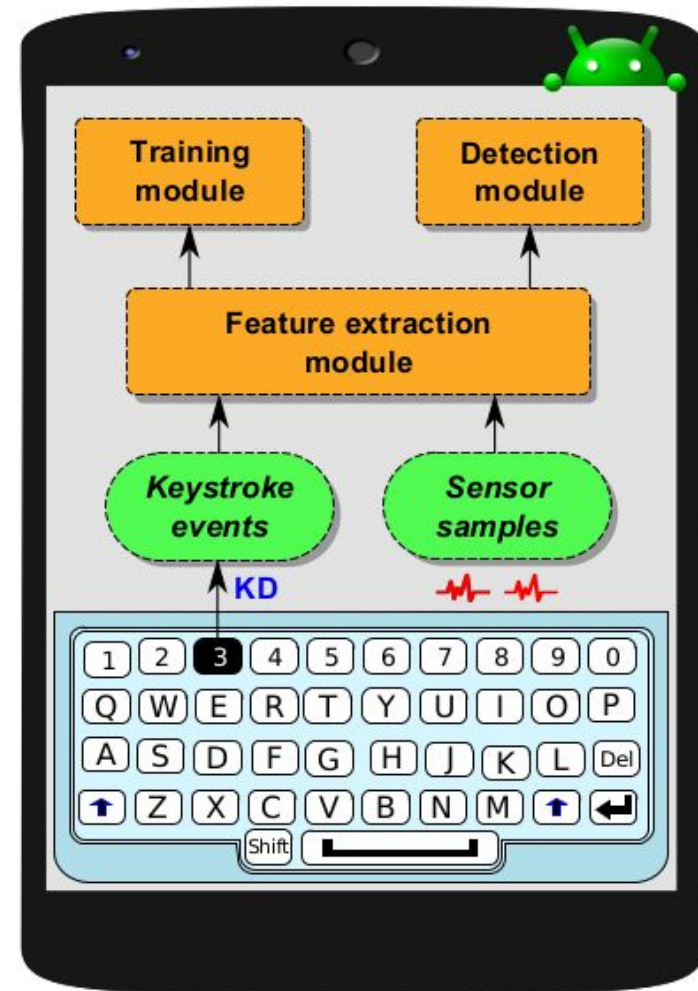
*In DIMVA 2014*

## Our system: Unagi

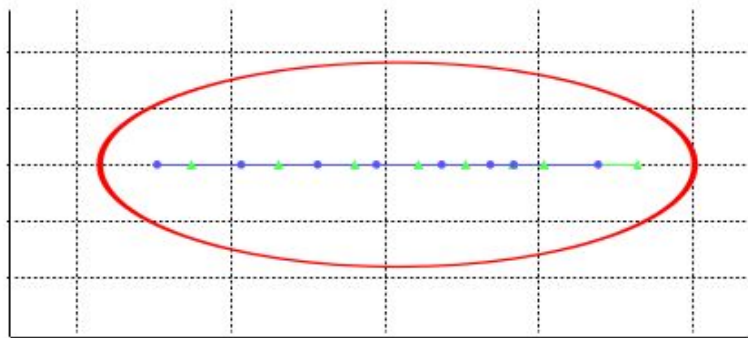
User authentication with  
Sensor enhanced  
Keystroke Dynamics



Scenario: User typing 'HELLO'



# I Sensed It Was You



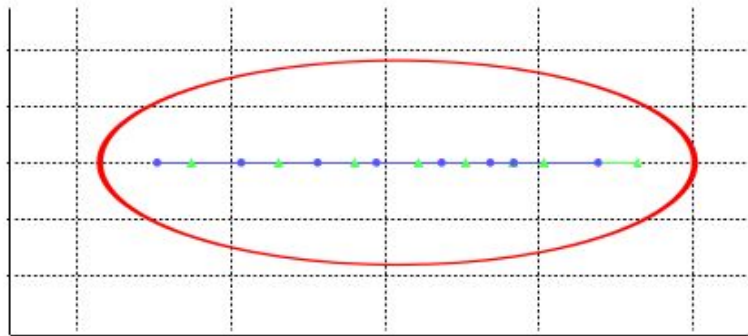
• User 1 KeyDowns → User 1 KeyUps



## Keystroke dynamics



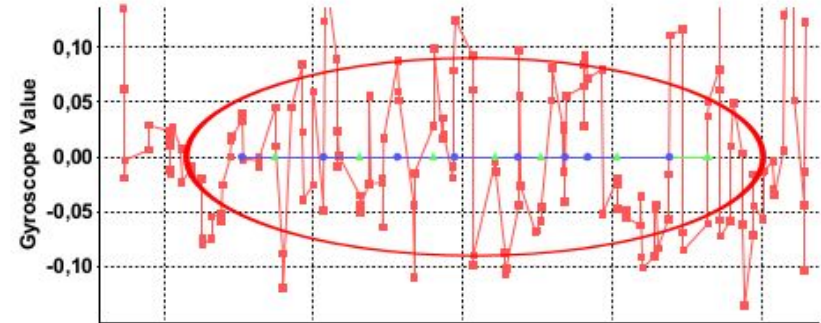
# I Sensed It Was You



→ User 1 KeyDowns → User 1 KeyUps



## Keystroke dynamics

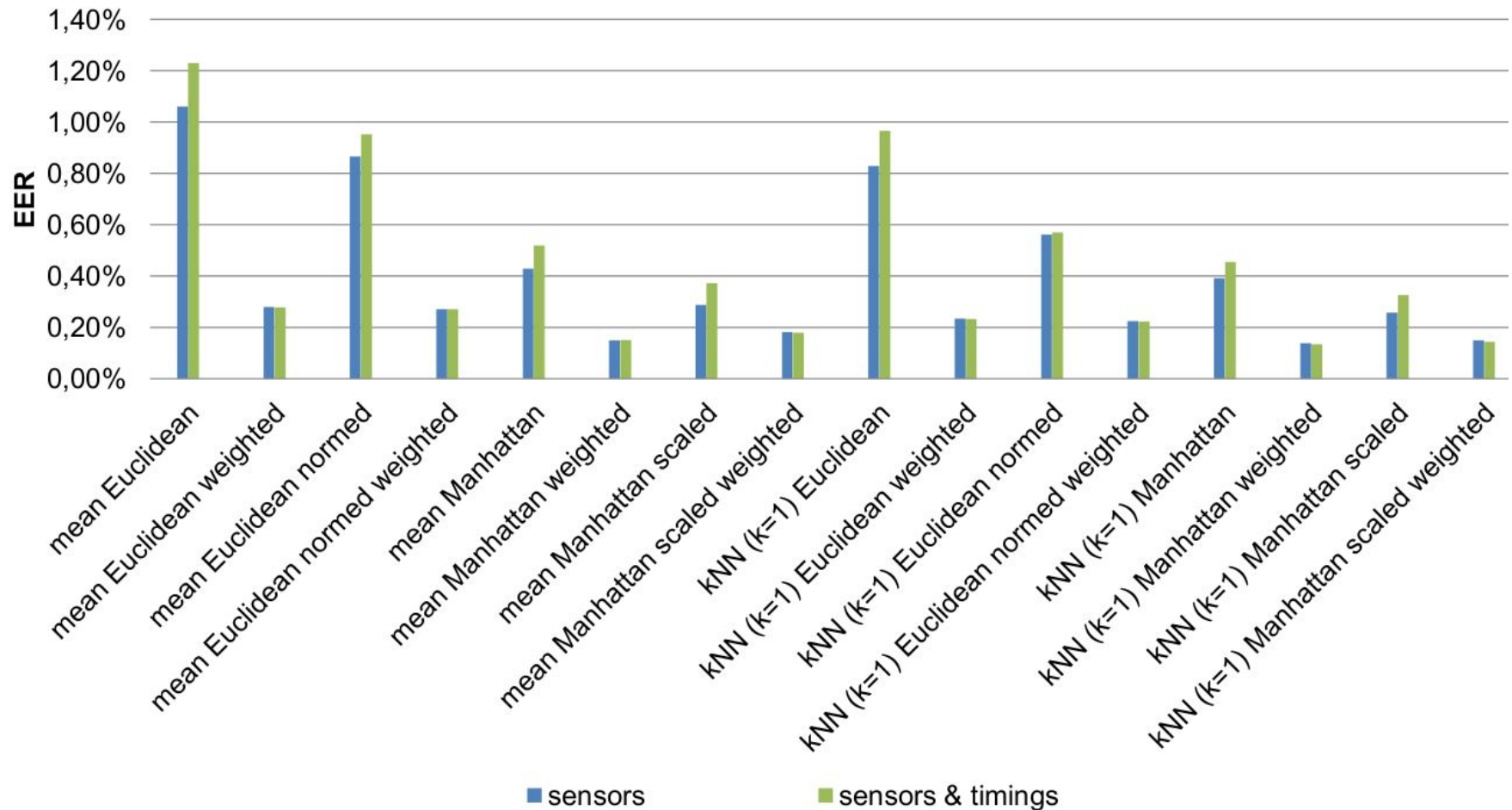


→ User 1 → User 1 KeyDowns → User 1 KeyUps

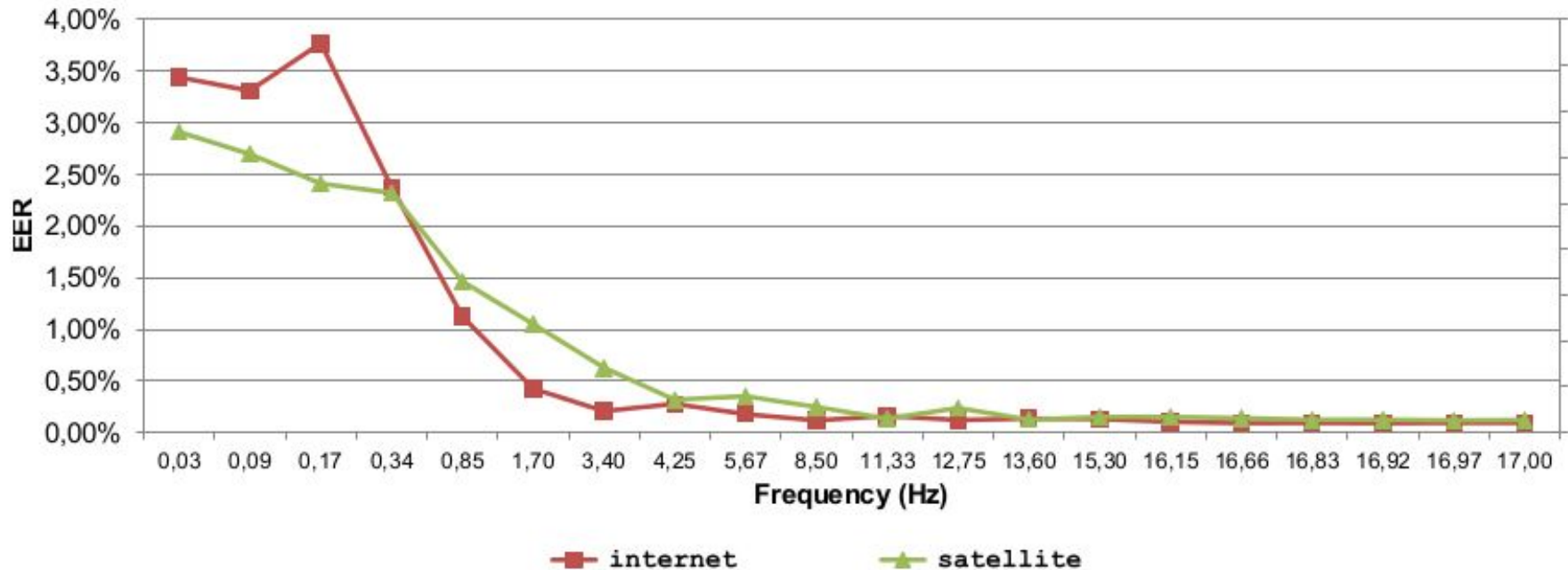


## Sensor-enhanced keystroke dynamics

## Accuracy (EER) for different considered algorithms



## Accuracy vs. Sensors Sampling Frequency



EER - Equal Error Rate (rate at which both acceptance and rejection errors are equal)



## Key Results

- Movement sensors are suitable for biometric authentication
- Sensors can dramatically enhance keystroke dynamics accuracy
- Effective even with short passwords and low sampling frequencies

## Future work

- Applicability to free-text authentication
- Robustness against statistical attacks



- Covert and Side Channels 101
- Network Traffic Analysis
  - *As a side channel: app and sensitive data inference*
- Energy Consumption
  - *As a side channel: user and app inference*
  - *As a covert channel: data exfiltration*
- **Device Movement**
  - *As a side channel: smartphone user authentication*
  - **Attacks against biometric authentication**
- Keystroke Timing
  - *As a side channel: text typed on keyboards*
- Acoustic Emanations
  - *As a side channel: text typed on keyboards*



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

V. D. Stanciu, R. Spolaor, M. Conti, C. Giuffrida

**On the Effectiveness of Sensor-enhanced Keystroke Dynamics**  
**Against Statistical Attacks**

*in ACM CODASPY 2016*





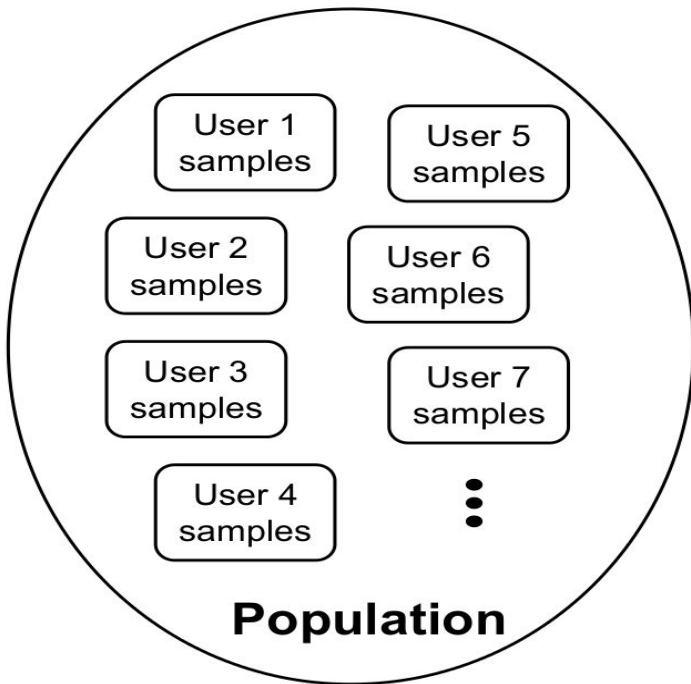
The previous **behavioral biometric authentication** system relies on:

- Secret of the password
- **Keystroke dynamics** (touch gestures)
- **Accelerometer** and **Gyroscope** sensors data

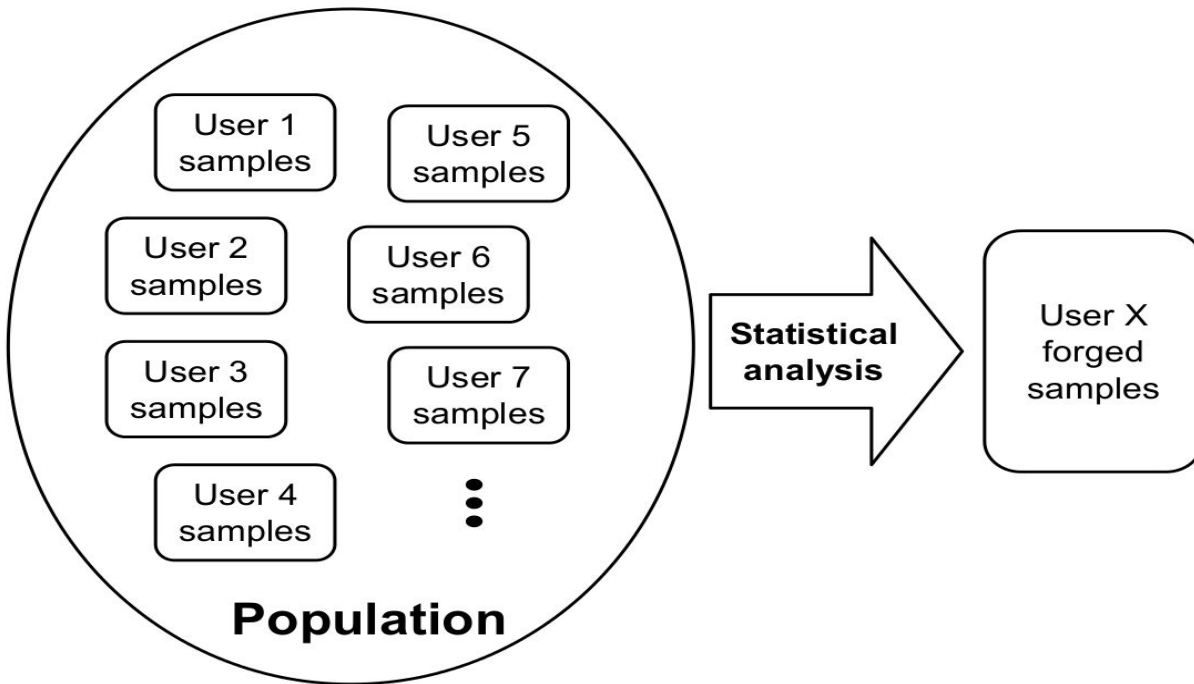
Previous work: we used kNN (with  $k=1$ ) and mean values combined with several metrics (e.g., euclidean, Manhattan)

**Question:** is our system resilient to **Statistical attacks**?

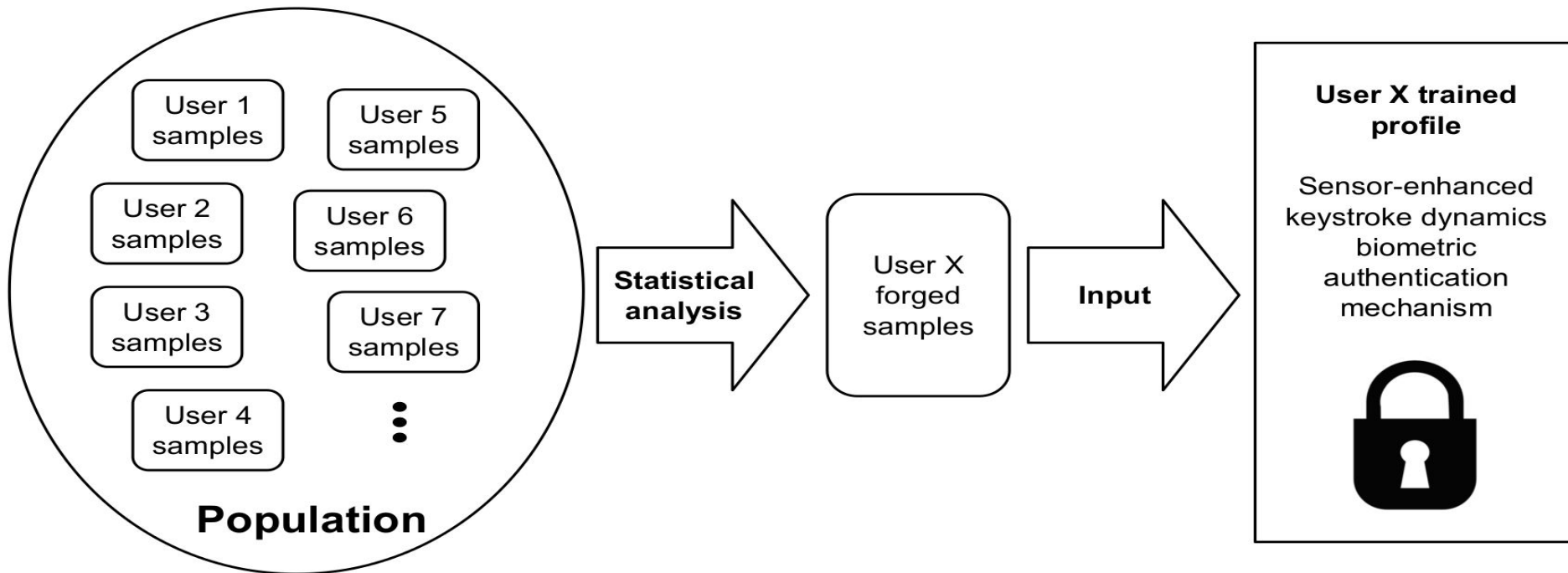
# Statistical Attack



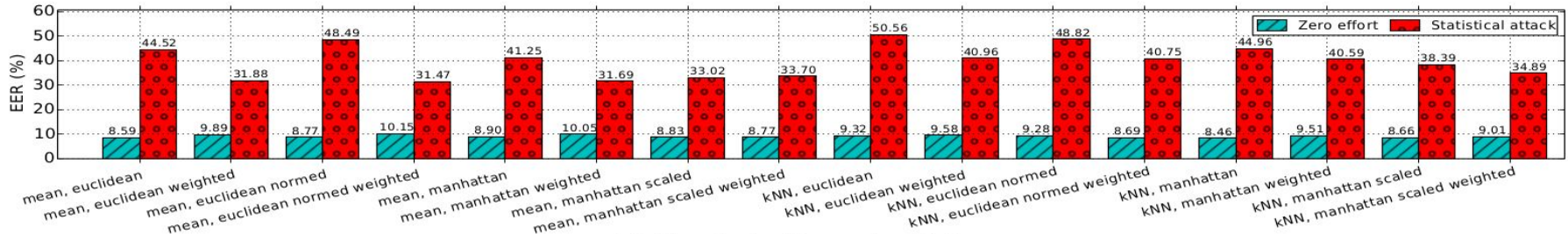
# Statistical Attack



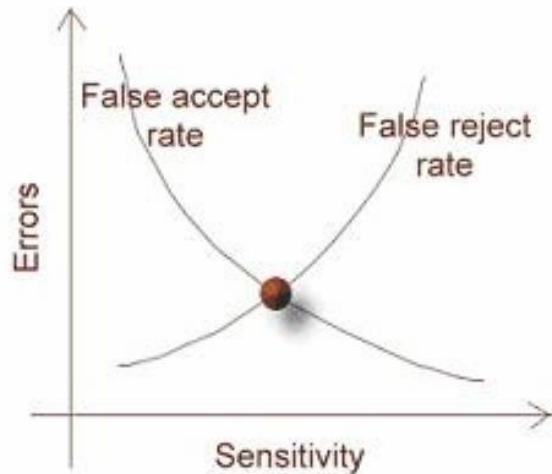
# Statistical Attack



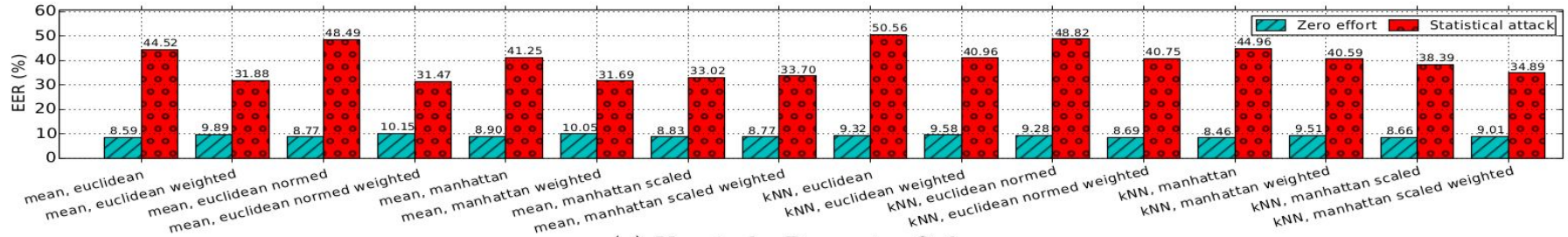
low Equal Error Rate (EER) == accurate authentication method



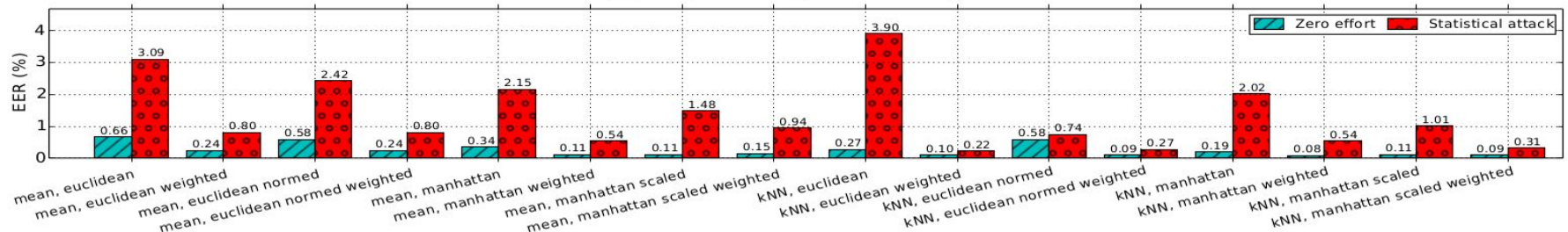
(a) Keystroke-Dynamics Only.



low Equal Error Rate (EER) == accurate authentication method



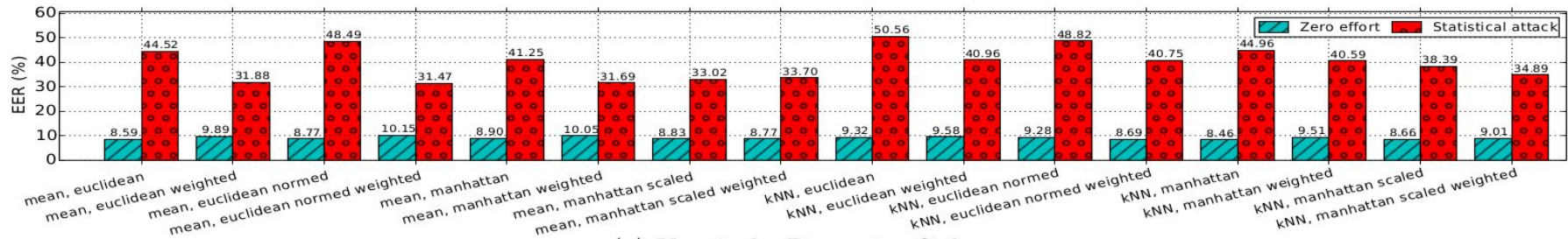
(a) Keystroke-Dynamics Only.



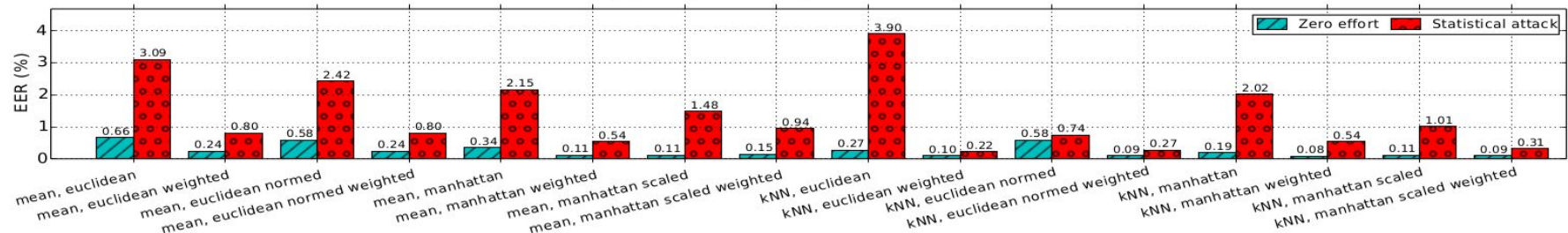
(b) Sensors only.



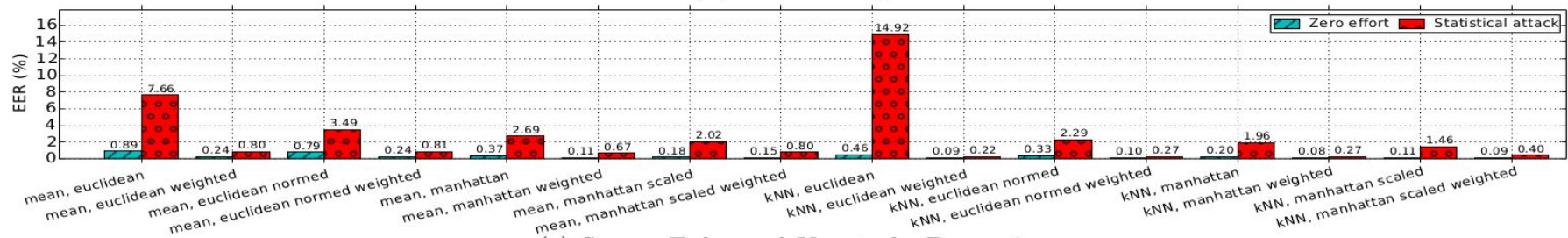
low Equal Error Rate (EER) == accurate authentication method



(a) Keystroke-Dynamics Only.



(b) Sensors only.



(c) Sensor-Enhanced Keystroke-Dynamics.





- Covert and Side Channels 101
- Network Traffic Analysis
  - *As a side channel: app and sensitive data inference*
- Energy Consumption
  - *As a side channel: user and app inference*
  - *As a covert channel: data exfiltration*
- Device Movement
  - *As a side channel: smartphone user authentication*
  - *Attacks against biometric authentication*
- **Keystroke Timing**
  - *As a side channel: text typed on keyboards*
- Acoustic Emanations
  - *As a side channel: text typed on keyboards*

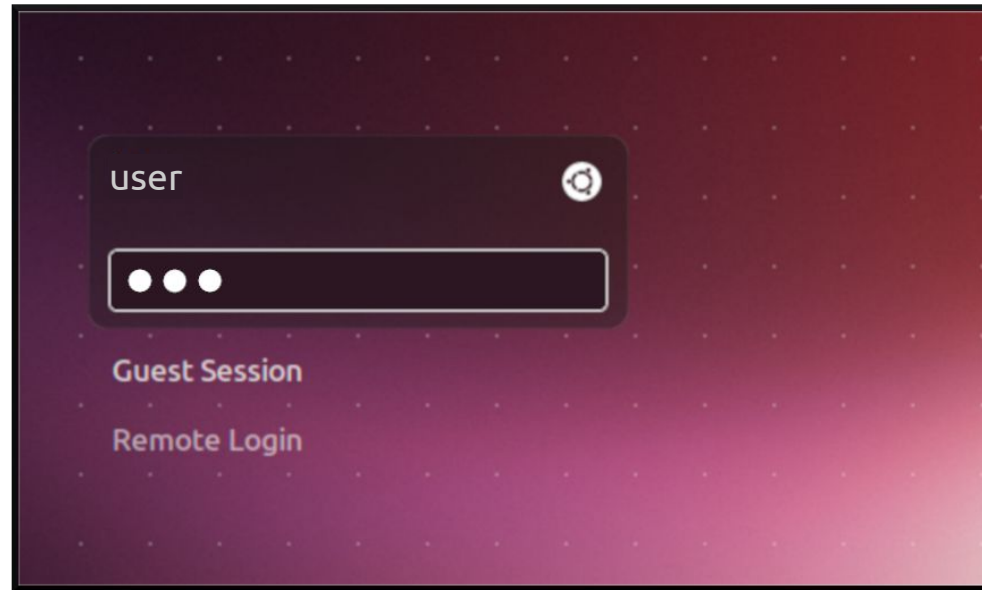
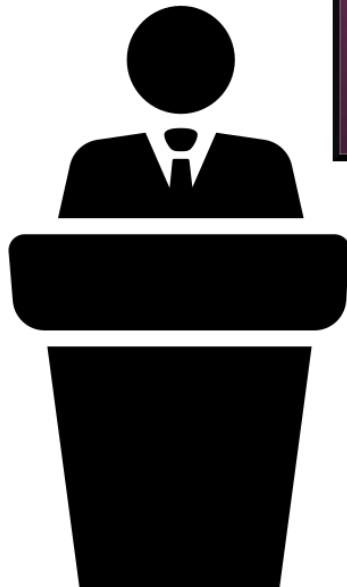


Kiran Balagani, Mauro Conti, Paolo Gasti, Martin Georgiev, Tristan Gurtler,  
Daniele Lain, Charissa Miller, Kendall Molas, Nikita Samarin, Eugen Saraci,  
Gene Tsudik, Lynn Wu

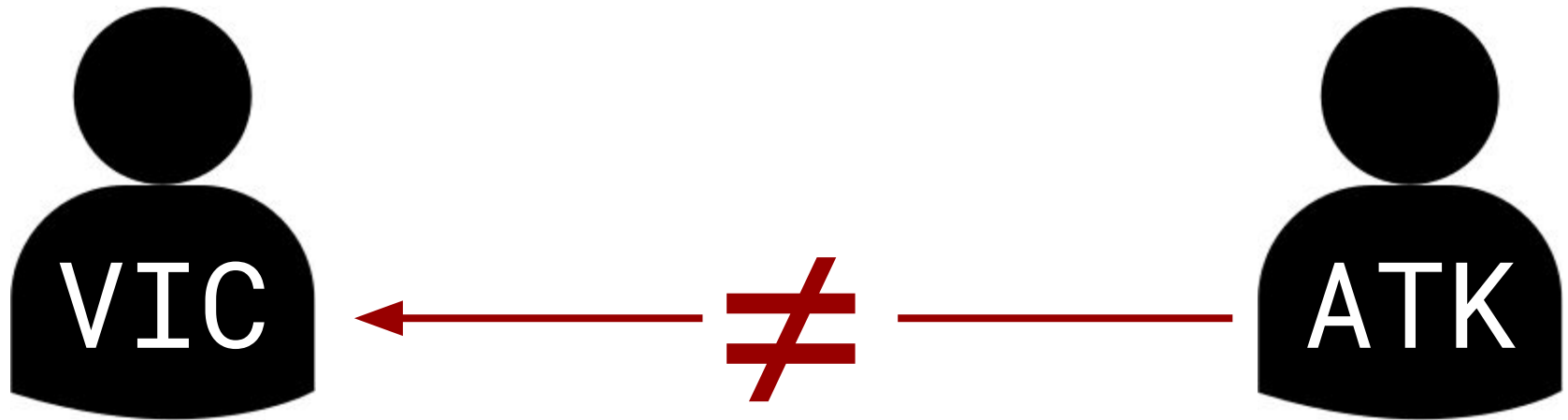
**SILK-TV: Secret Information Leakage From Keystroke Timing Videos.**

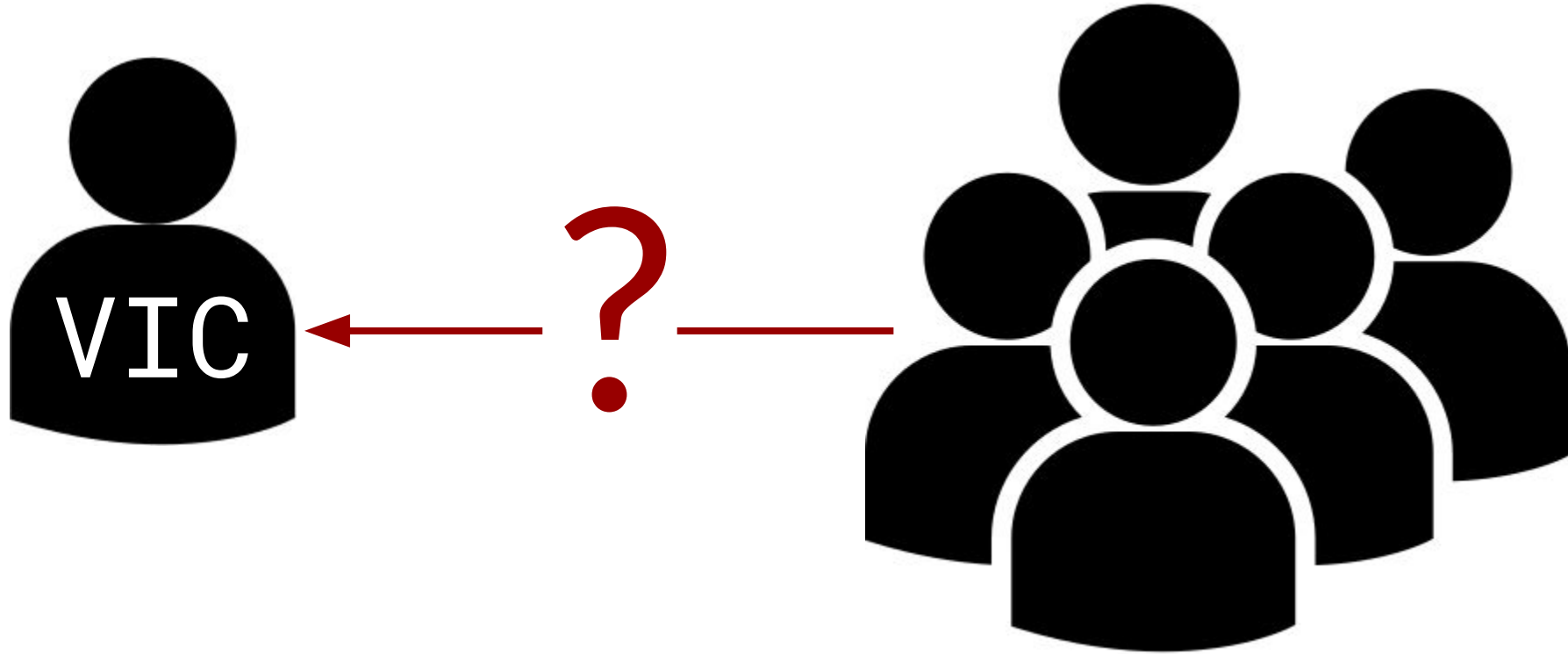
*In ESORICS 2018*

# Timing Information Leak - 1





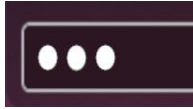


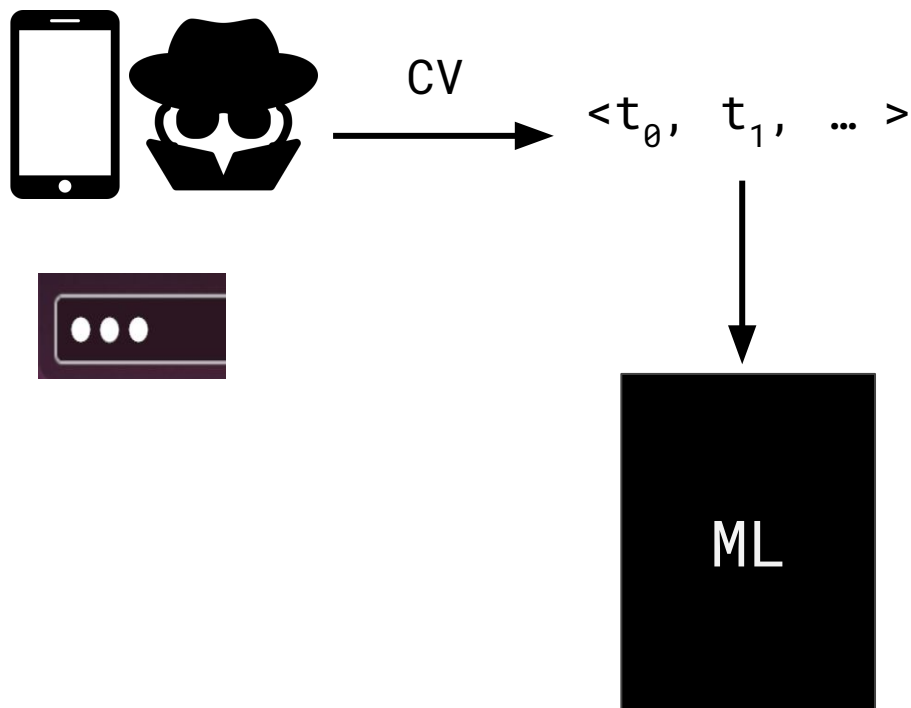


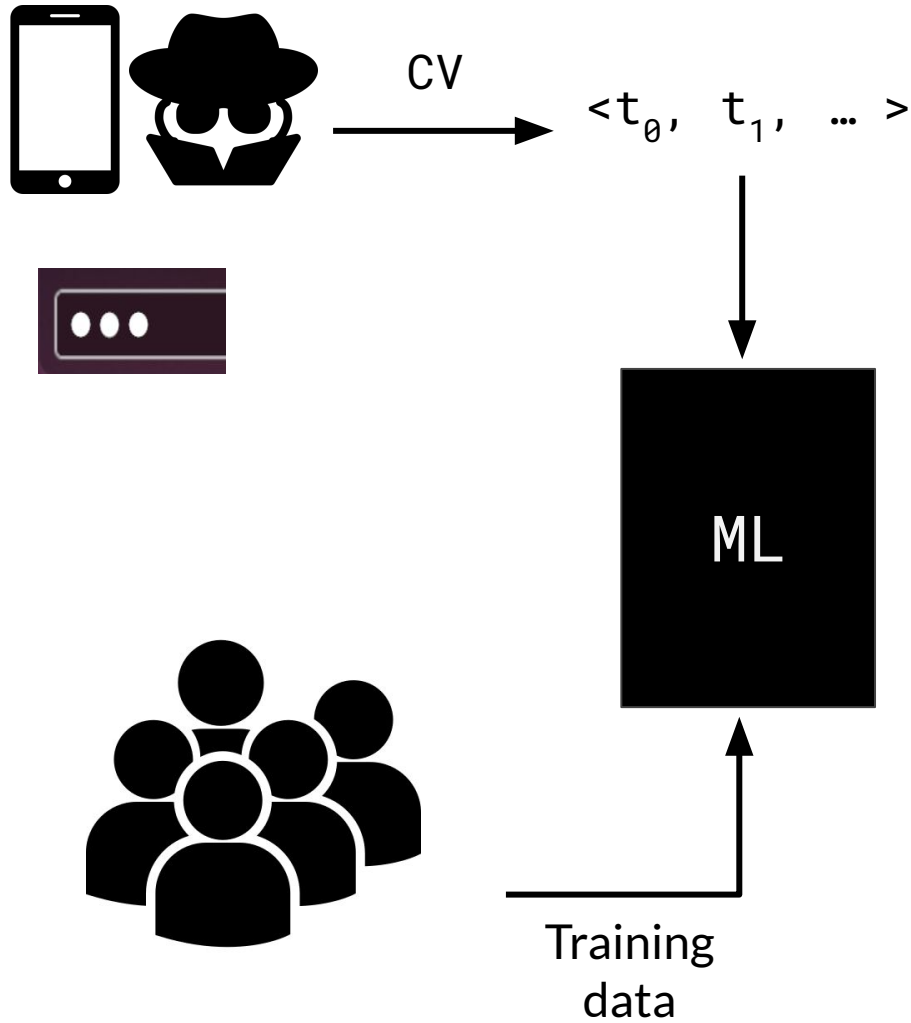


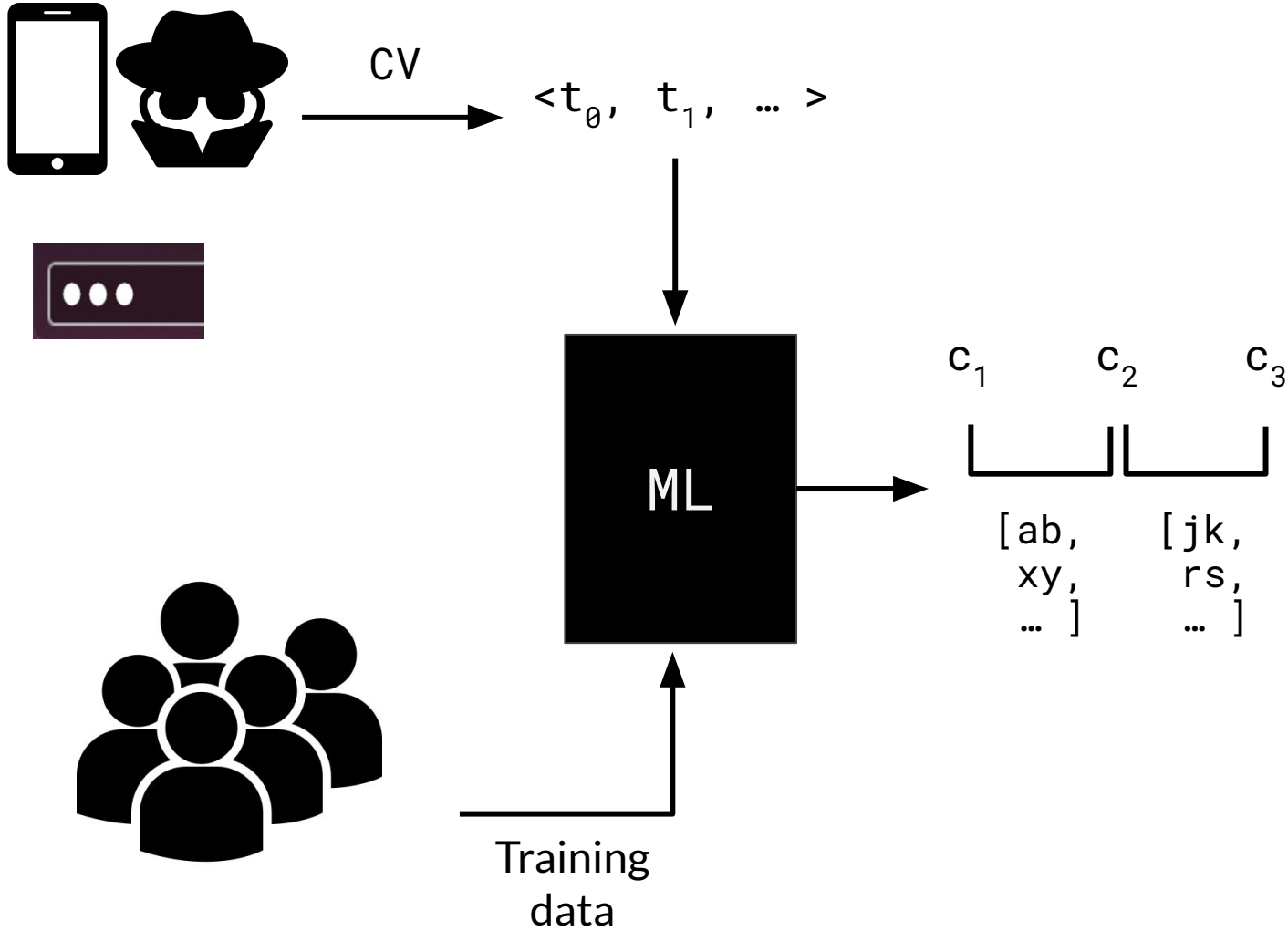
- Quantify information leakage of on-screen keystroke feedback
- Novel attack: *SILK-TV*
  - *Uses public datasets only from multiple sources (“population data”)*
  - *Machine Learning to guess typed text (passwords and PINs)*

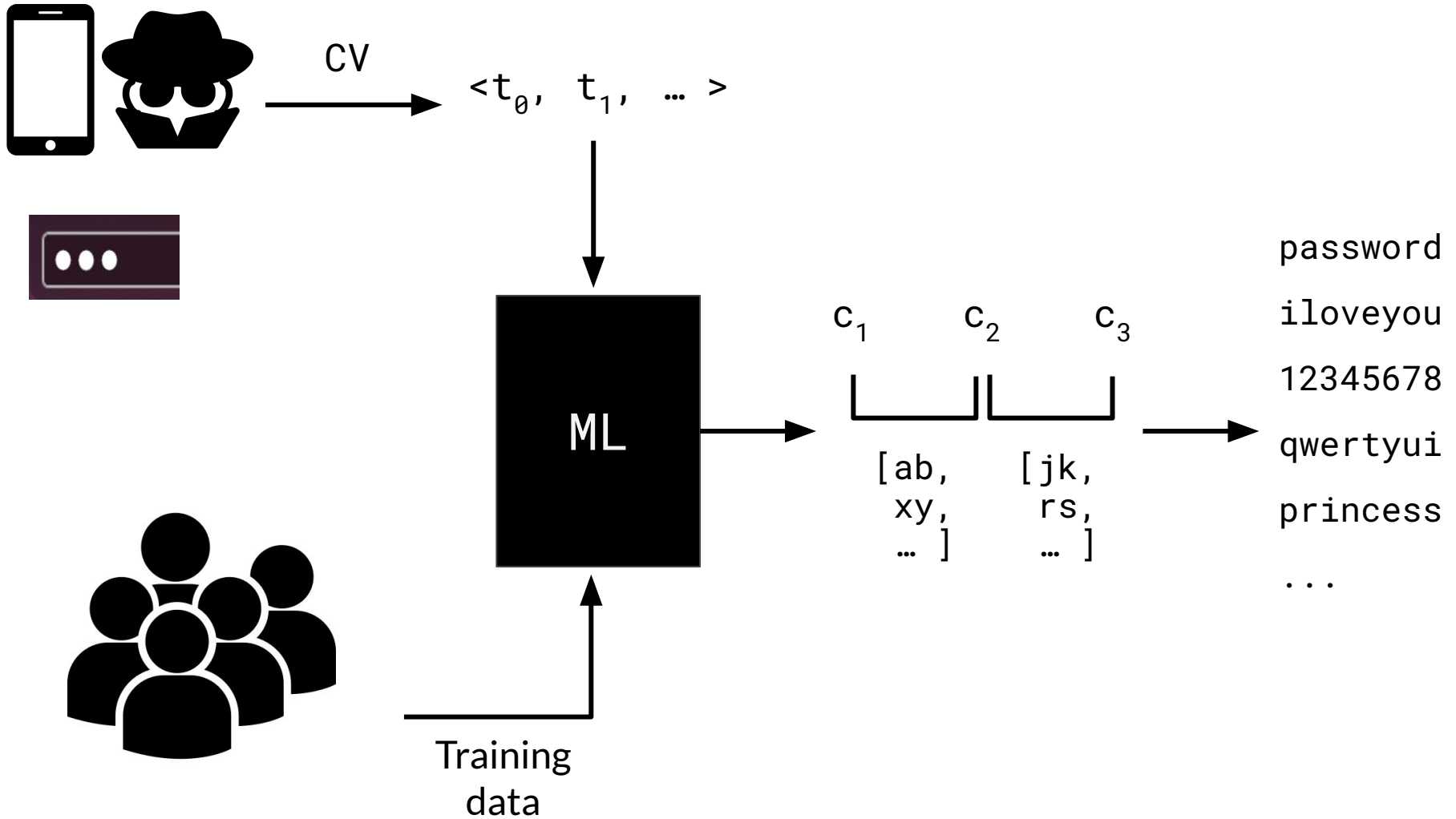












# Data Collection - Passwords



- Data from **projector** and **laptop screen @ 60Hz**
- Recorded with a smartphone
- 62 users - 3 times each pwd - **touch typing** on keyboard
- Randomly selected 4 passwords from `rockyou`<sup>1</sup>
  - `123brian`, `jillie02`, `lamondre`, `william1`



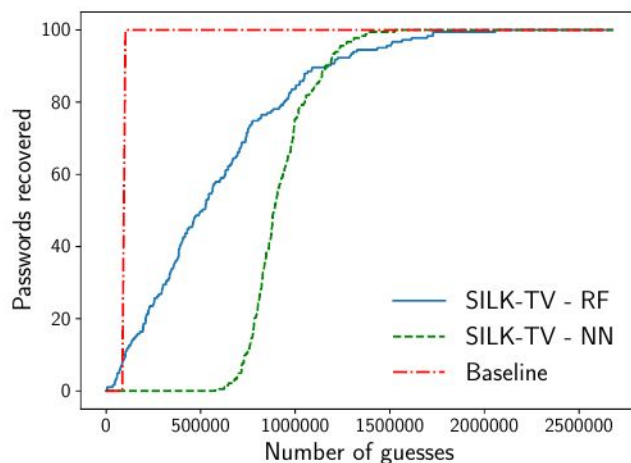
1 - <http://downloads.skullsecurity.org/passwords/rockyou.txt.bz2>



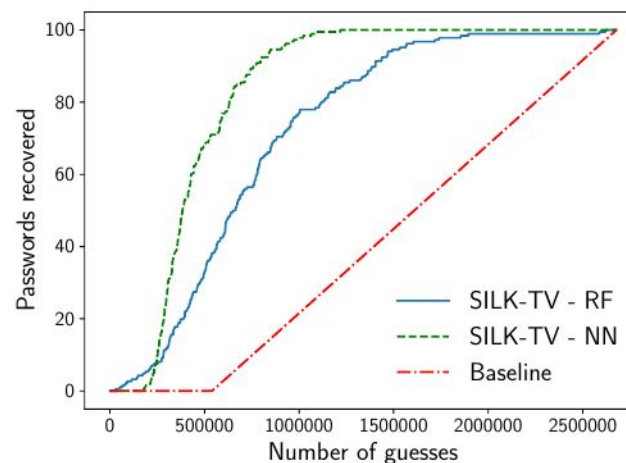
- Baseline: password list sorted by frequency
  - “Best” strategy for a zero-information attacker
  - *123brian* - 93,874<sup>th</sup>
  - *jillie02* - 1,753,571<sup>st</sup>
  - *lamondre* - 397,213<sup>rd</sup>
  - *william1* - 187<sup>th</sup> ← very frequent password
- Evaluation scenarios
  - “Single shot”
  - “Multiple recordings” (e.g., professor at lectures)



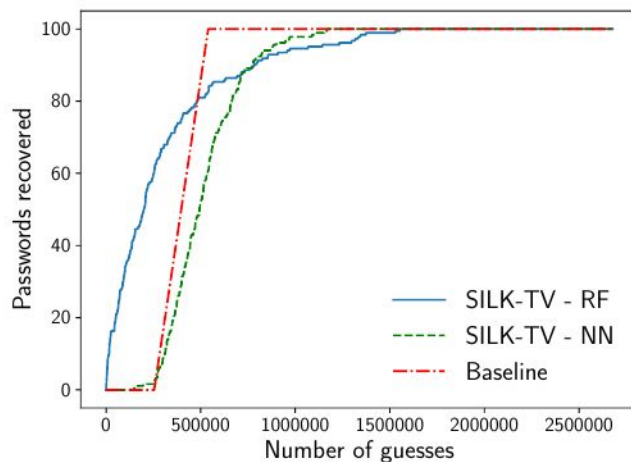
# Password - "Single Shot" results



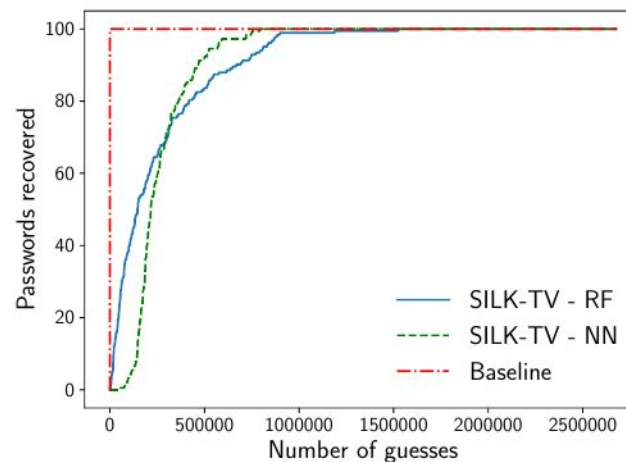
(a) 123brian (183 auth. attempts).



(b) jillie02 (186 auth. attempts).

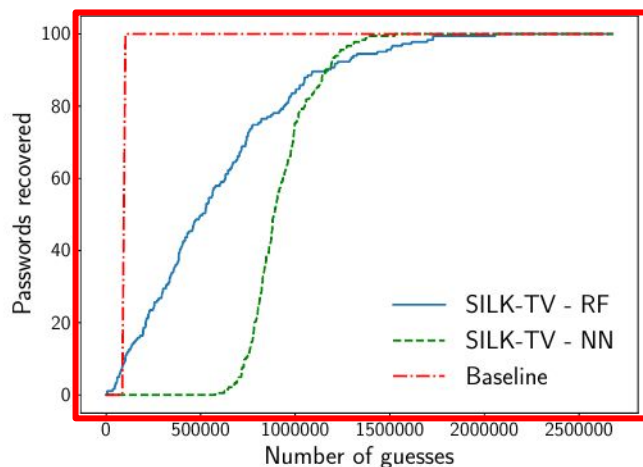


(c) lamondre (184 auth. attempts).

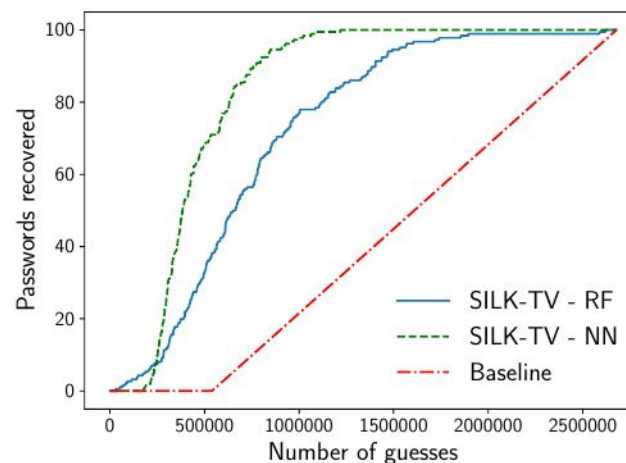


(d) william1 (183 auth. attempts).

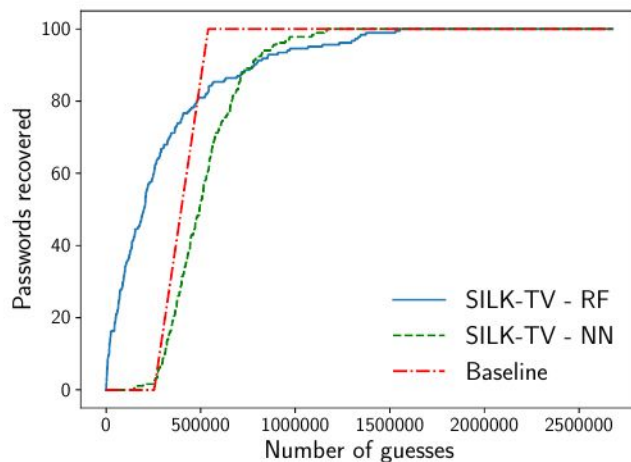
# Password - "Single Shot" results



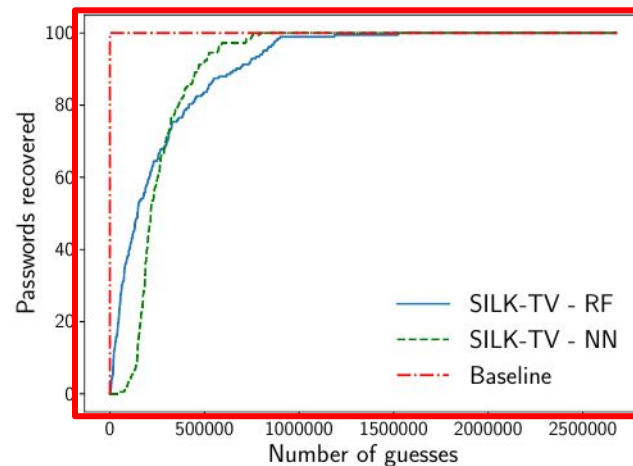
(a) 123brian (183 auth. attempts).



(b) jillie02 (186 auth. attempts).

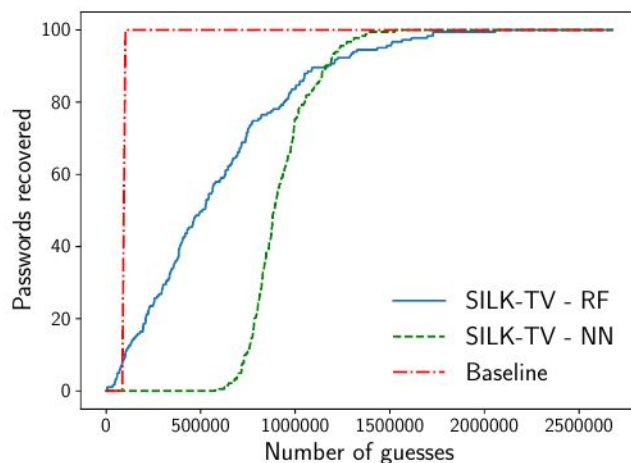


(c) lamondre (184 auth. attempts).

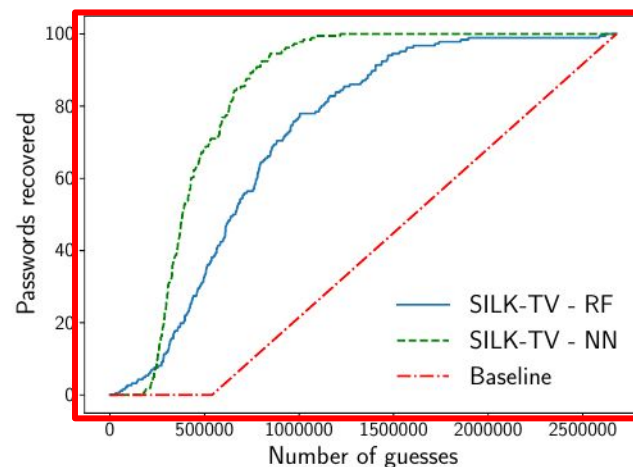


(d) william1 (183 auth. attempts).

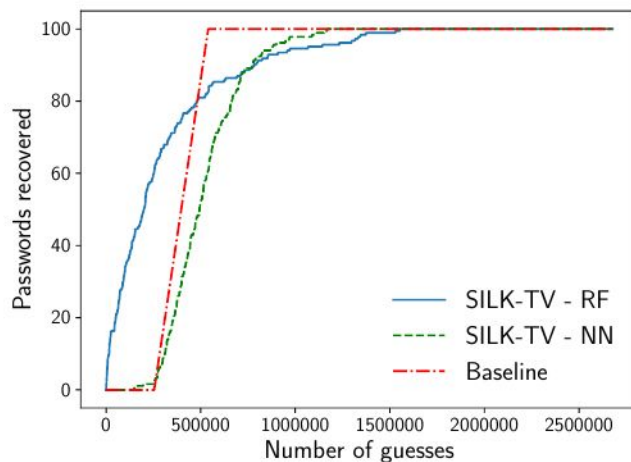
# Password - "Single Shot" results



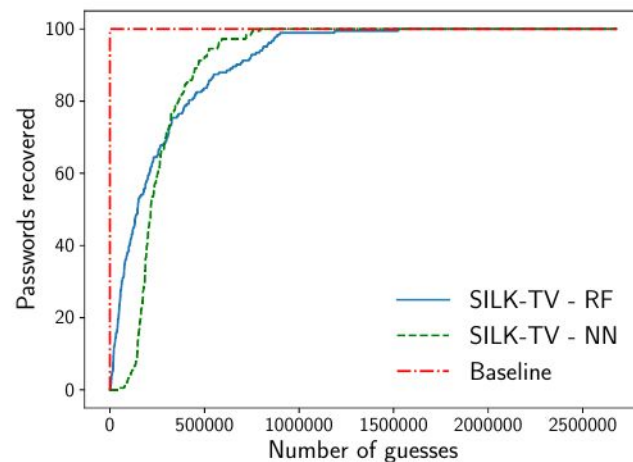
(a) 123brian (183 auth. attempts).



(b) jillie02 (186 auth. attempts).



(c) lamondre (184 auth. attempts).



(d) william1 (183 auth. attempts).

# Password - "Single Shot" results



	Avg	Stdev	Med	Rnd	<Rnd	Best	<20k	<100k
<b>Random Forest</b>								
123brian	581,743	414,761	508,332	93,874	8.7%	5,535	1.1%	9.3%
jillie02	749,718	448,319	656,754	1,753,571	97.8%	28,962	0.0%	2.7%
lamondre	301,906	334,681	199,344	397,213	75.0%	145	13.0%	33.7%
william1	246,437	264,090	145,966	187	0.5%	68	10.9%	39.9%
<b>Neural Network</b>								
123brian	923,534	165,454	886,802	93,874	0.0%	577,739	0.0%	0.0%
jillie02	456,811	210,512	383,230	1,753,571	100.0%	164,754	0.0%	0.0%
lamondre	517,472	189,355	493,713	397,213	28.8%	148,403	0.0%	0.0%
william1	265,813	140,753	215,840	187	0.0%	45,176	0.0%	3.8%

Avg, Stdev, Median of SILK-TV cracking attempts

Rnd average baseline cracking attempts

<Rnd, Best, <20k, <100k highlights of SILK-TV performance

# Password - "Single Shot" results



	Avg	Stdev	Med	Rnd	<Rnd	Best	<20k	<100k
<b>Random Forest</b>								
123brian	581,743	414,761	508,332	93,874	8.7%	5,535	1.1%	9.3%
jillie02	749,718	448,319	656,754	1,753,571	97.8%	28,962	0.0%	2.7%
lamondre	301,906	334,681	199,344	397,213	75.0%	145	13.0%	33.7%
william1	246,437	264,090	145,966	187	0.5%	68	10.9%	39.9%
<b>Neural Network</b>								
123brian	923,534	165,454	886,802	93,874	0.0%	577,739	0.0%	0.0%
jillie02	456,811	210,512	383,230	1,753,571	100.0%	164,754	0.0%	0.0%
lamondre	517,472	189,355	493,713	397,213	28.8%	148,403	0.0%	0.0%
william1	265,813	140,753	215,840	187	0.0%	45,176	0.0%	3.8%

Avg, Stdev, Median of SILK-TV cracking attempts

Rnd average baseline cracking attempts

<Rnd, Best, <20k, <100k highlights of SILK-TV performance

# Password - "Single Shot" results



	Avg	Stdev	Med	Rnd	<Rnd	Best	<20k	<100k
<b>Random Forest</b>								
123brian	581,743	414,761	508,332	93,874	8.7%	5,535	1.1%	9.3%
jillie02	749,718	448,319	656,754	1,753,571	97.8%	28,962	0.0%	2.7%
lamondre	301,906	334,681	199,344	397,213	75.0%	145	13.0%	33.7%
william1	246,437	264,090	145,966	187	0.5%	68	10.9%	39.9%
<b>Neural Network</b>								
123brian	923,534	165,454	886,802	93,874	0.0%	577,739	0.0%	0.0%
jillie02	456,811	210,512	383,230	1,753,571	100.0%	164,754	0.0%	0.0%
lamondre	517,472	189,355	493,713	397,213	28.8%	148,403	0.0%	0.0%
william1	265,813	140,753	215,840	187	0.0%	45,176	0.0%	3.8%

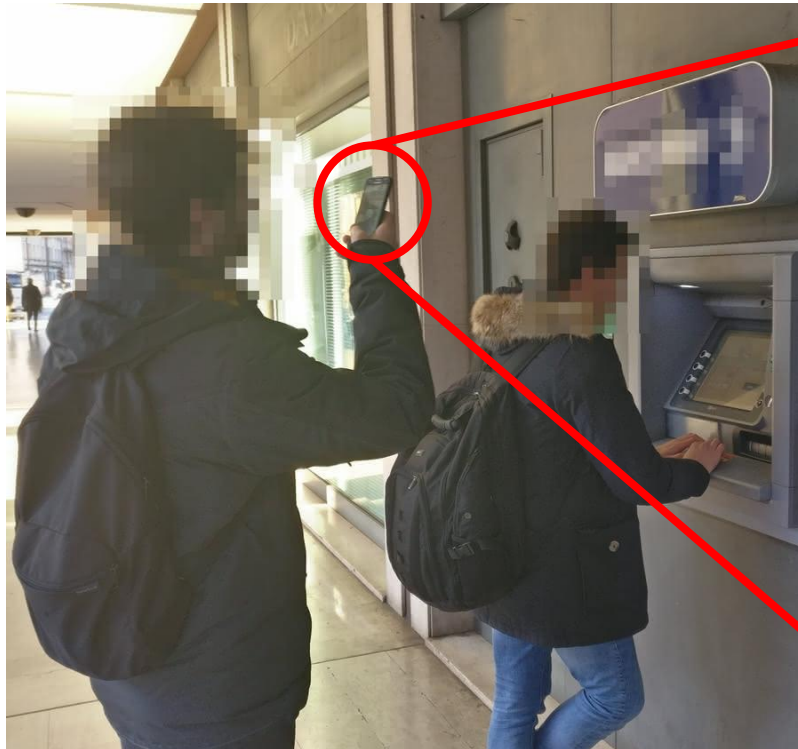
Avg, Stdev, Median of SILK-TV cracking attempts

Rnd average baseline cracking attempts

<Rnd, Best, <20k, <100k highlights of SILK-TV performance



# Timing Information Leak - 2



Keypad not visible - but the screen is!





- Timing information from videos is **accurate**
- Password masking leak timing → useful information
  - *Reduces number of attempts*
  - *More useful on **uncommon** passwords!*
- Performances on PIN... not great (close to random guess)



# PIN Salabim

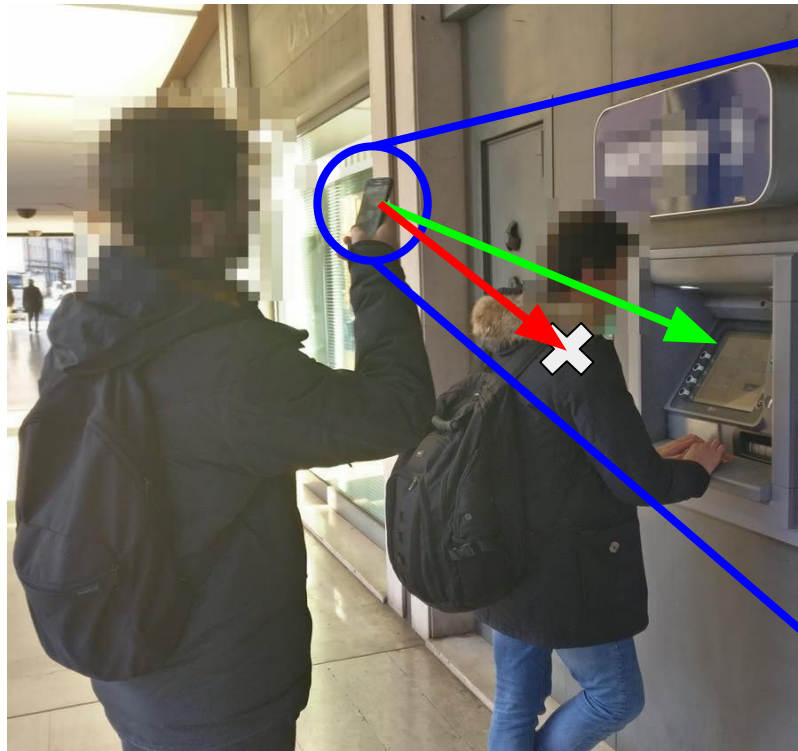


SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

GFT



Keypad not visible - but the screen is!



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

GFT ■

## PILOT

# Password and PIN Information Leakage from Obfuscated Typing Videos

Kiran Balagani, Matteo Cardaioli, Mauro Conti, Paolo Gasti, Martin Georgiev,  
Tristan Gurtler, Daniele Lain, Charissa Miller, Kendall Molas, Nikita Samarin,  
Eugen Saraci, Gene Tsudik, and Lynn Wu

*In Journal of Computer Security 2019*



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

**NYIT**

NEW YORK INSTITUTE  
OF TECHNOLOGY

GFT ■



**ETH** zürich

# PILOT

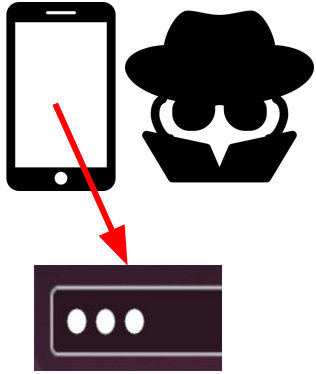


SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP

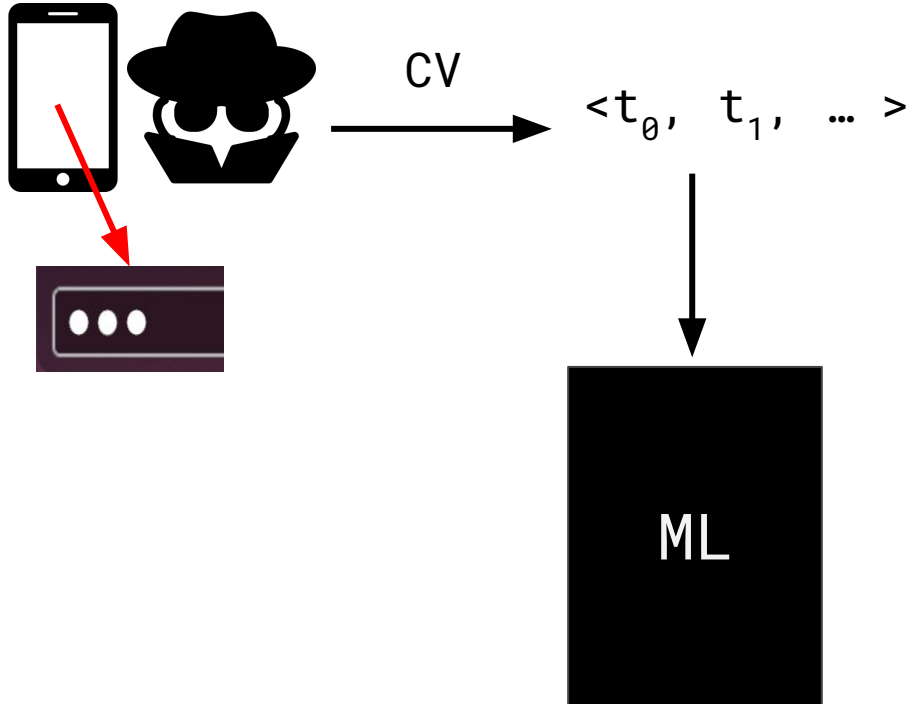


UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

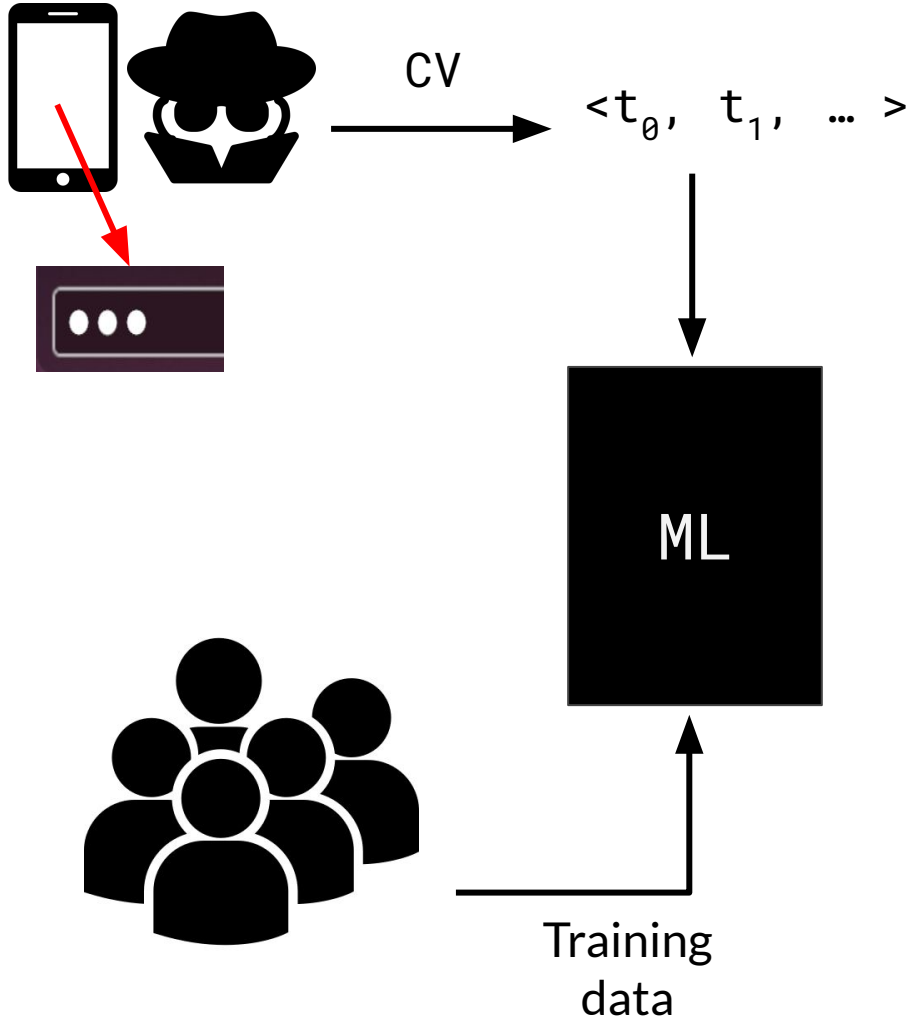
GFT ■



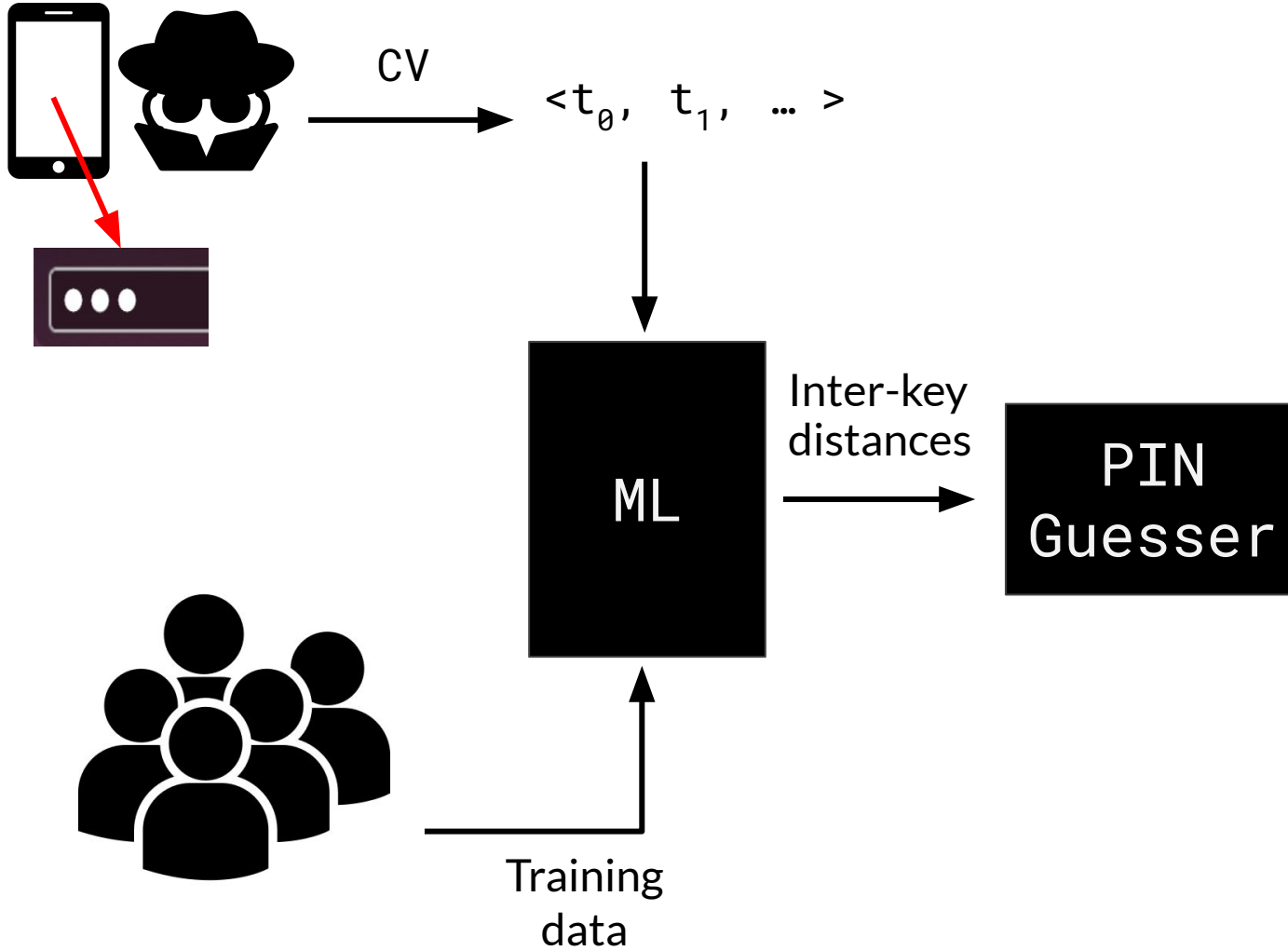
# PILOT



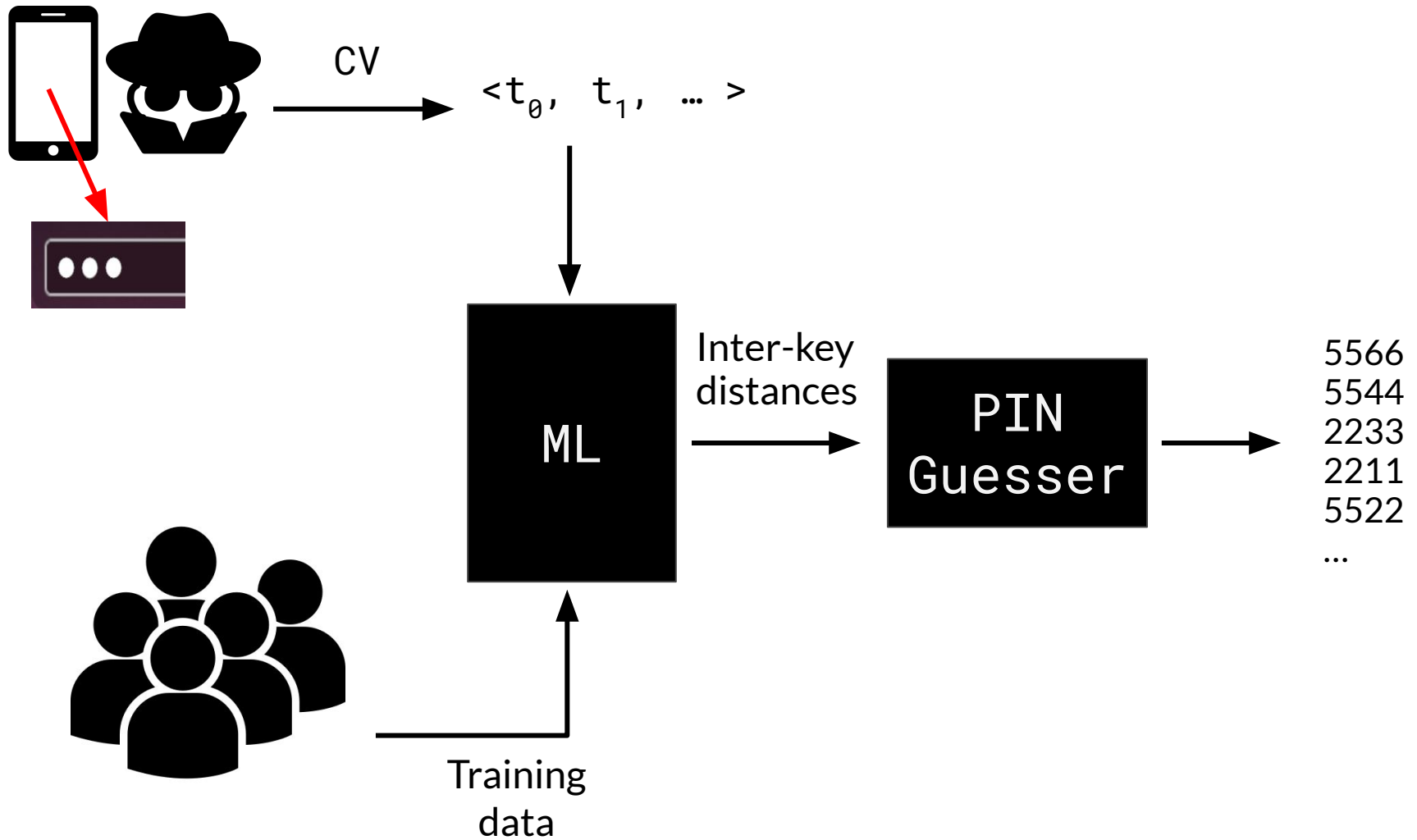
# PILOT





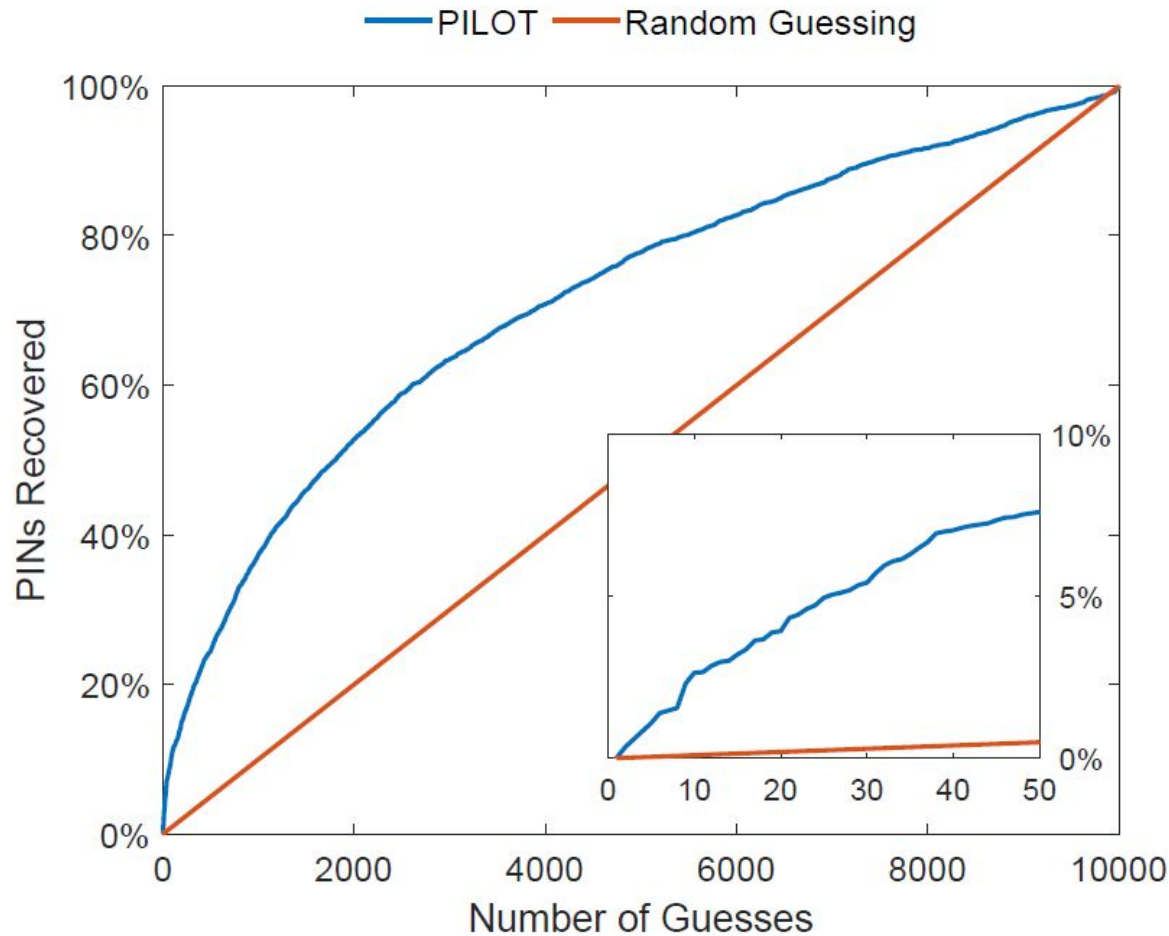


# PILOT



## Percentage of PINs recovered with PILOT vs Random Guessing

- 4 digit PIN (USA ATM card)





SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

GFT ■

**Your PIN Sounds Good!**  
**On The Feasibility of PIN Inference Through Audio Leakage**

Matteo Cardaioli, Mauro Conti, Kiran Balagani, and Paolo Gasti

IEEE Transactions on Information Forensics and Security 2019 (Submitted)

<https://arxiv.org/abs/1905.08742>

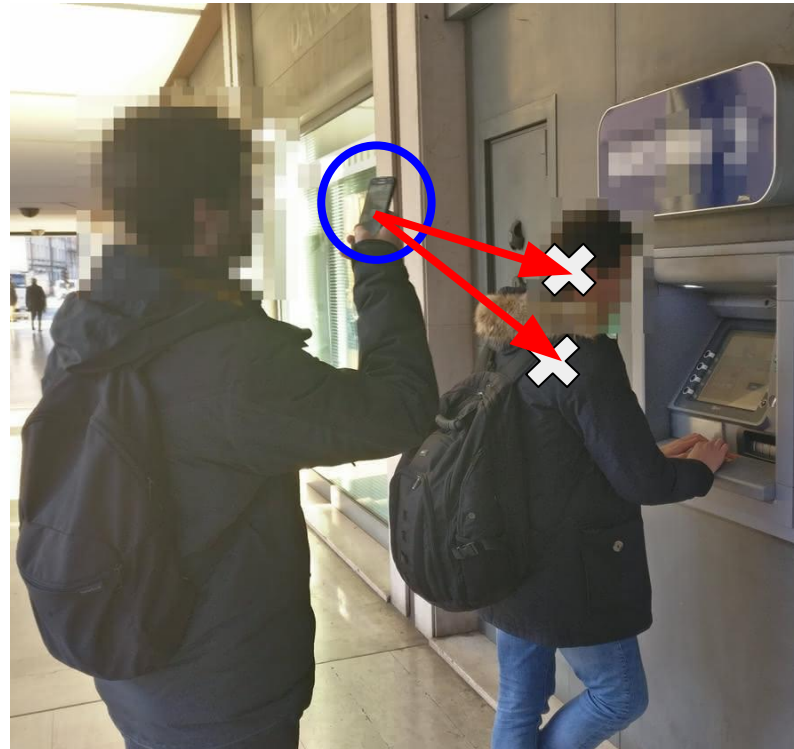


UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

**NYIT**

NEW YORK INSTITUTE  
OF TECHNOLOGY

GFT ■



Neither keypad nor screen are visible



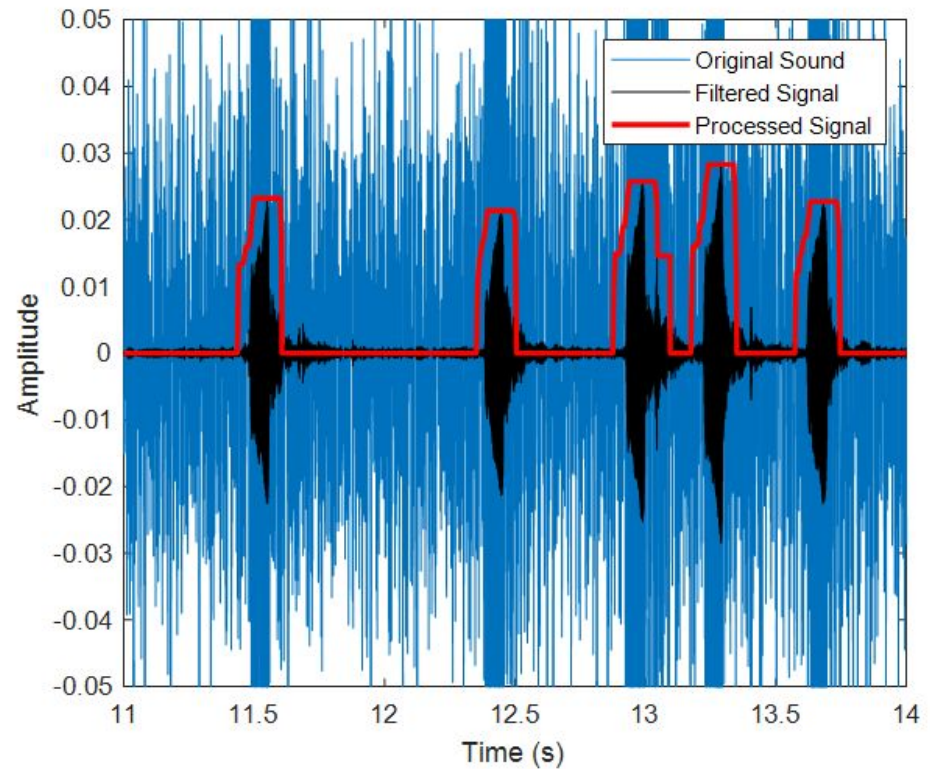
## Inter-keystroke timing identification through sound analysis

- Signal filtering

*To extract feedback sound characteristic frequency*

- Signal processing

*To remove residual noise and to identify time distance between peaks*





## Adversarial additional knowledge about the user or the PIN

- Knowledge of **typing behavior**

*Hunt-and-peck vs. touch typing*

- Knowledge of a **digit**

*Adversary knows one digit of the PIN*



- **Heatmap**

*Adversary performs a **thermal attack***

- *Better on plastic and rubber*  
*Not so good on metal*



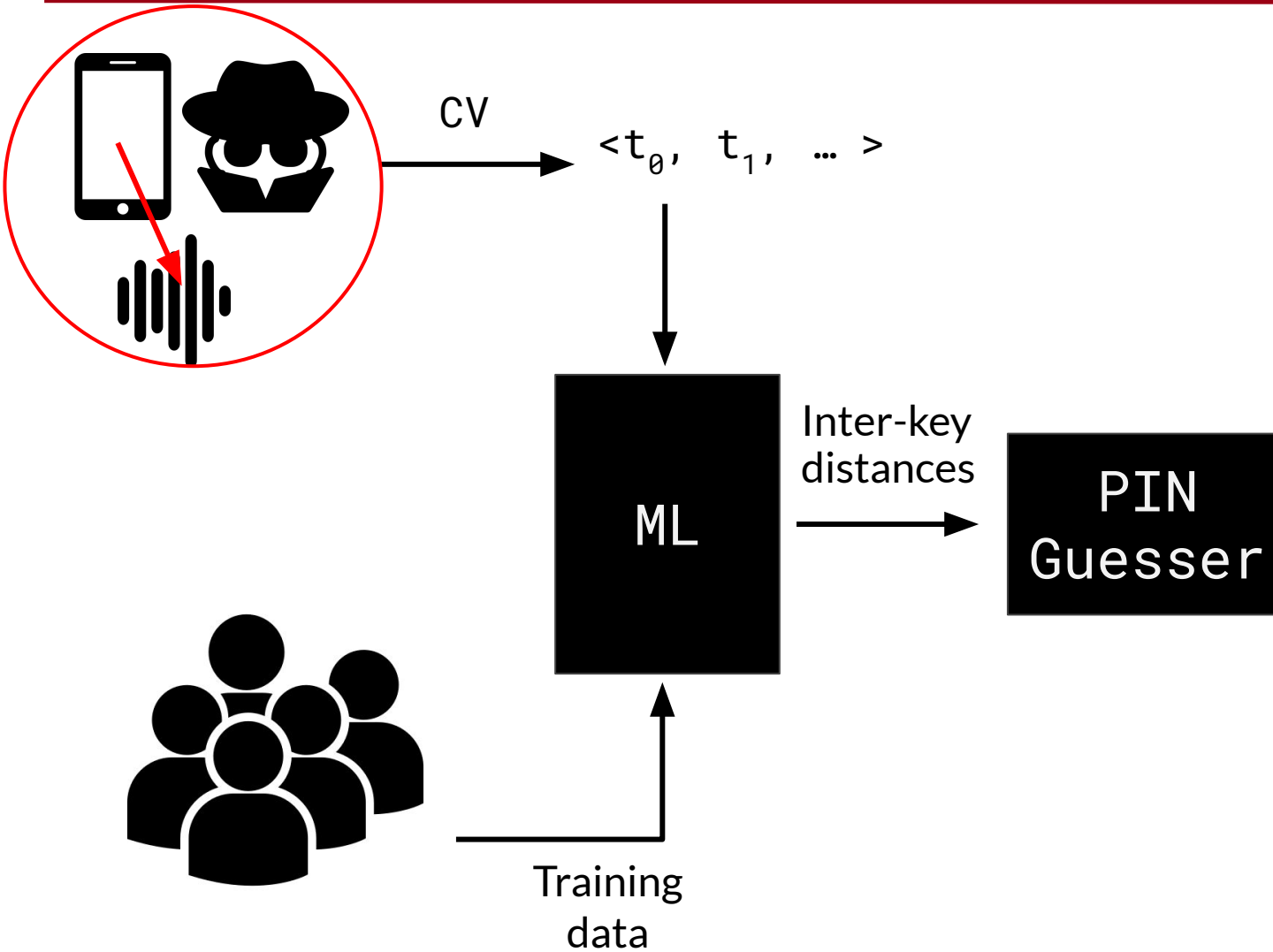
FLIR One PRO  
Lt iOS...

252 €

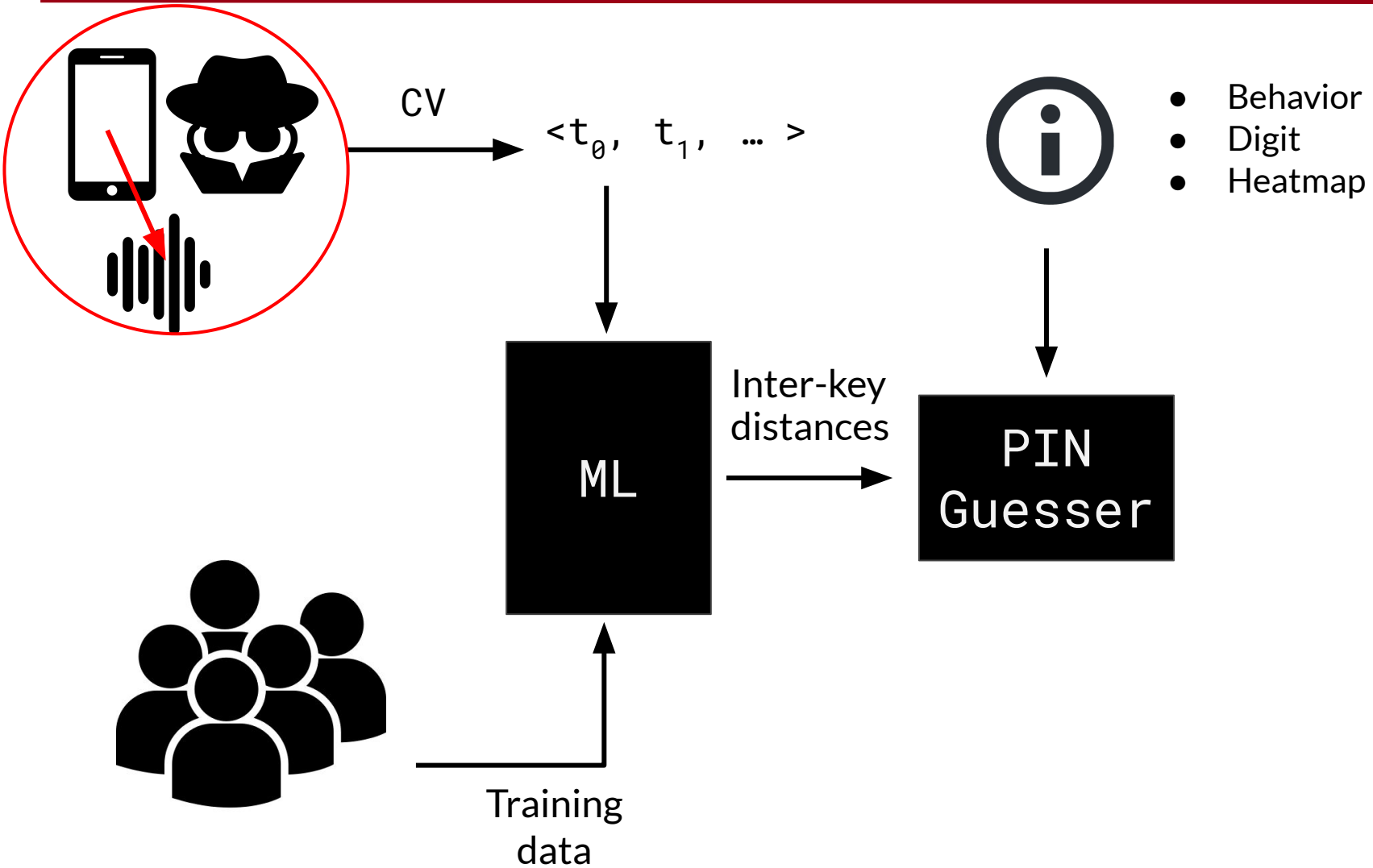
amazon



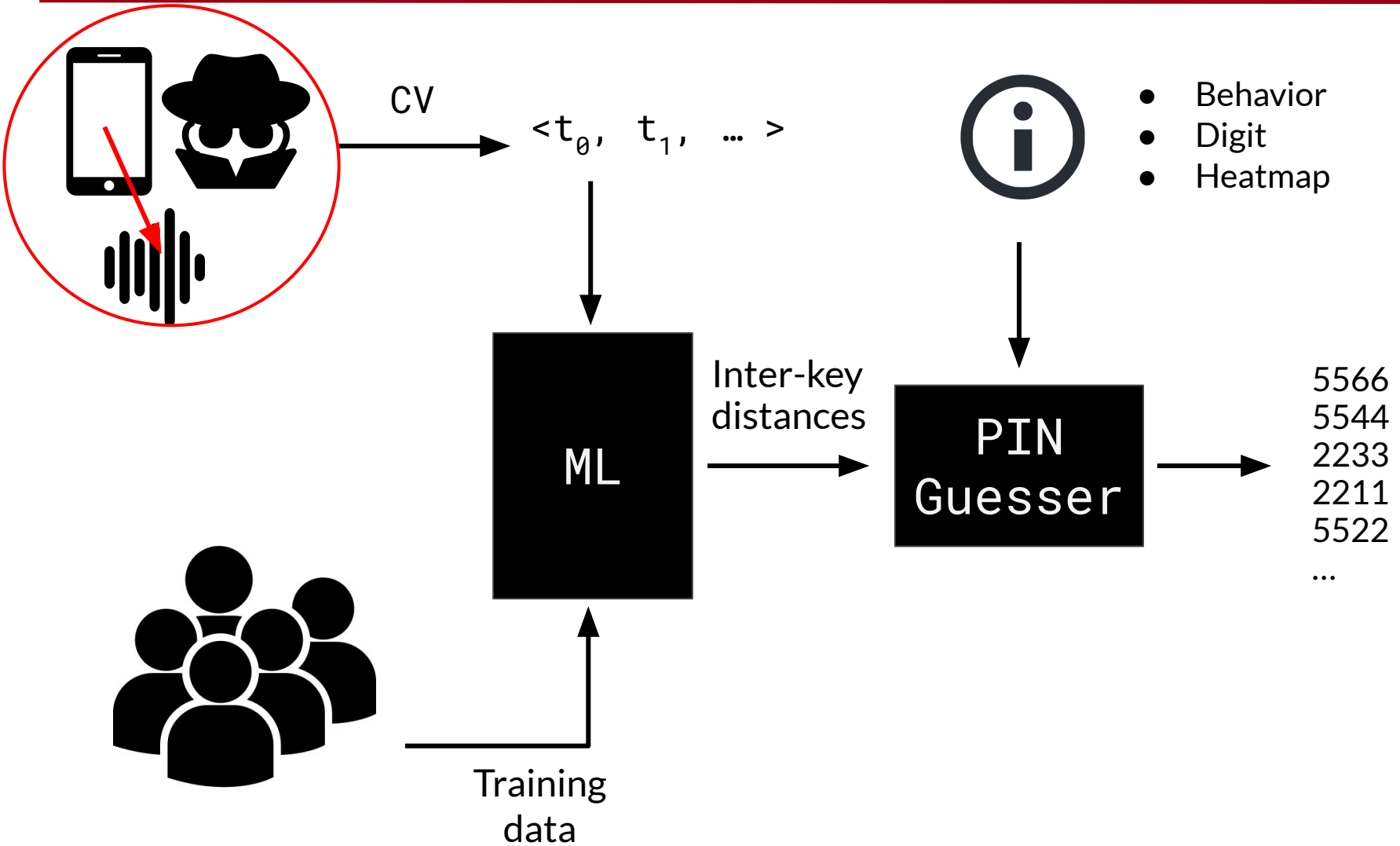
# Your PIN Sounds Good!



# Your PIN Sounds Good!



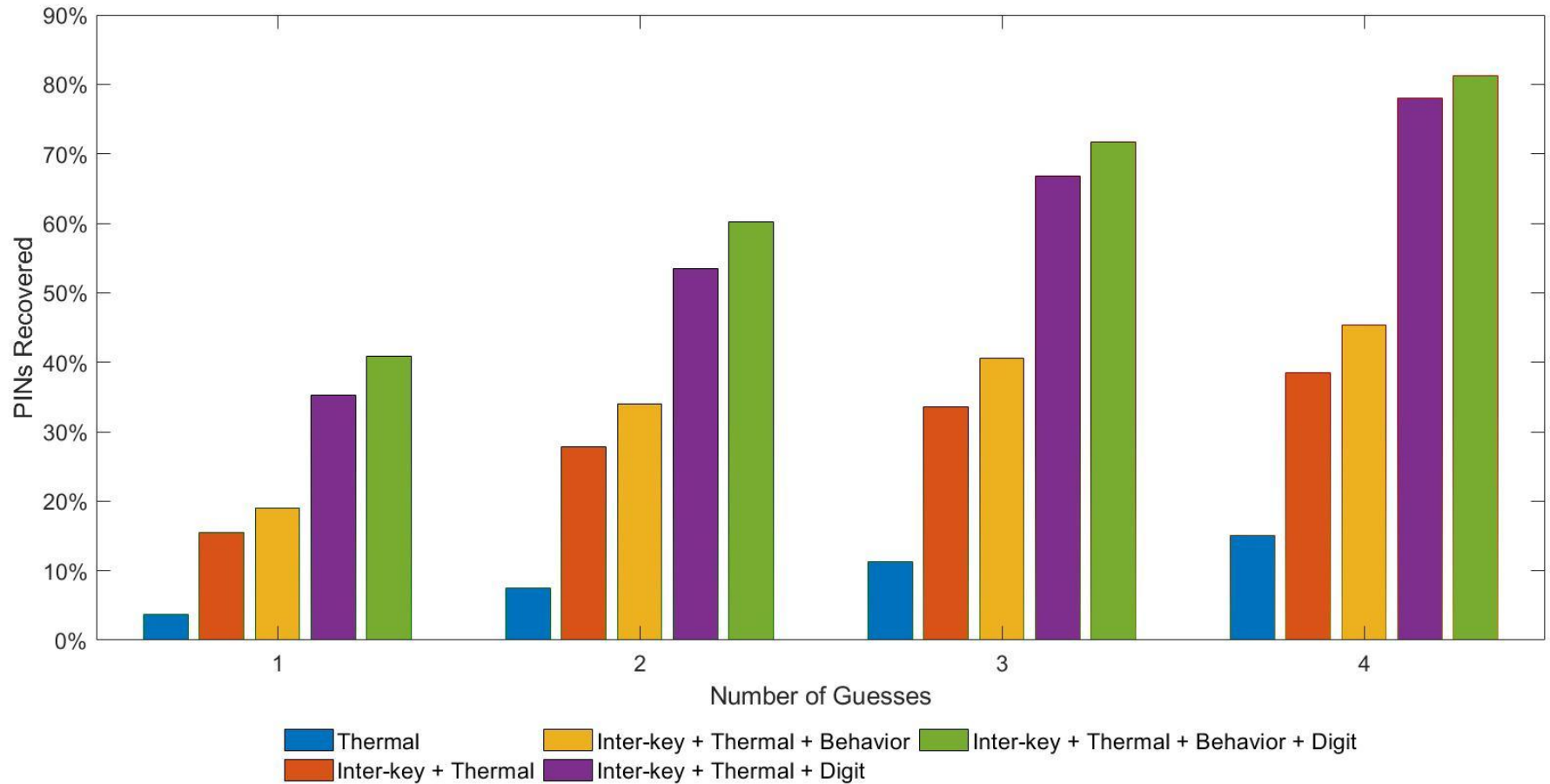
# Your PIN Sounds Good!



# Your PIN Sounds Good!



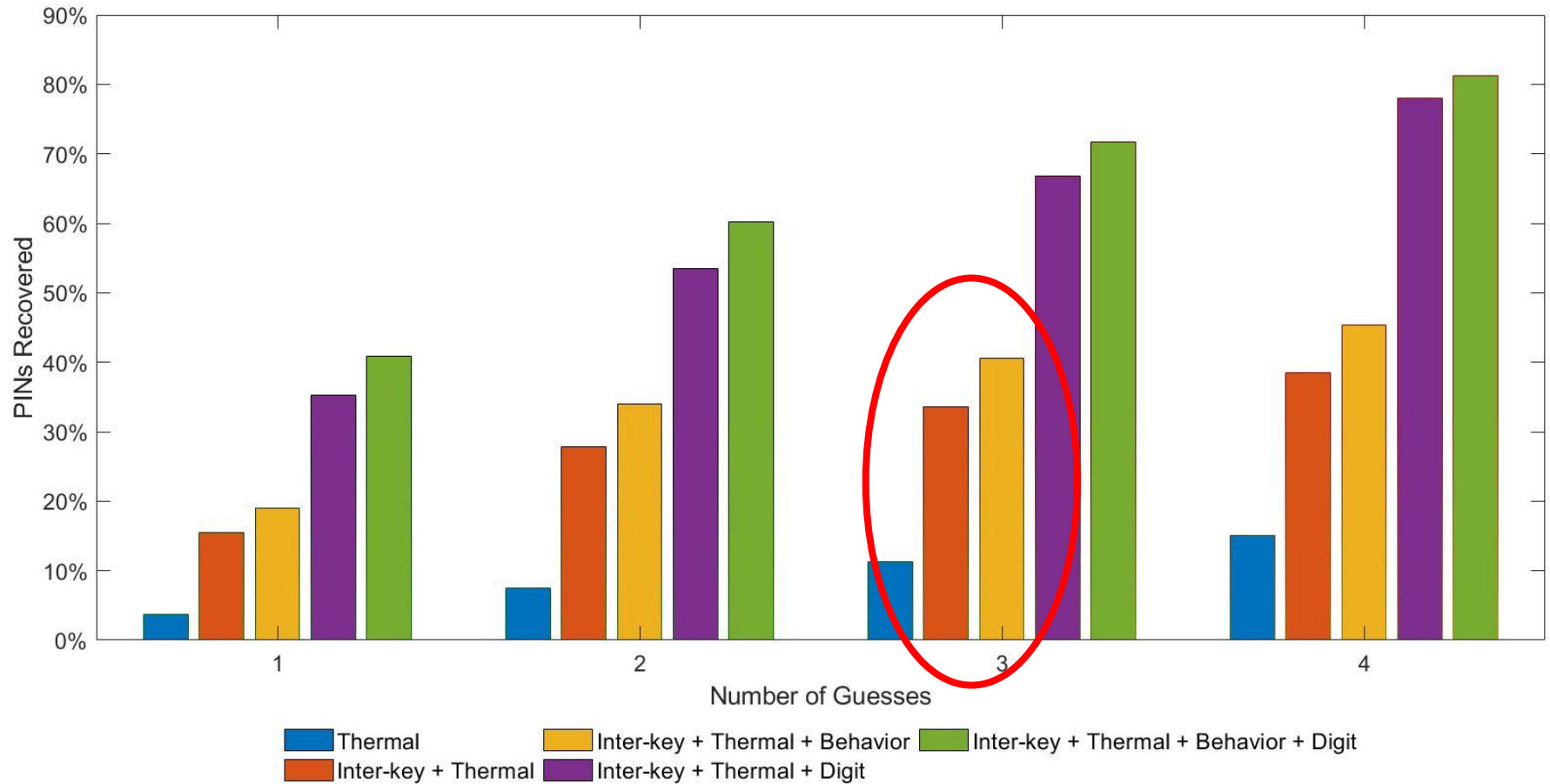
## % PINs recovered: inter-keystroke timing + other informations



# Your PIN Sounds Good!



## % PINs recovered: inter-keystroke timing + other informations



Your PIN Sounds Good!

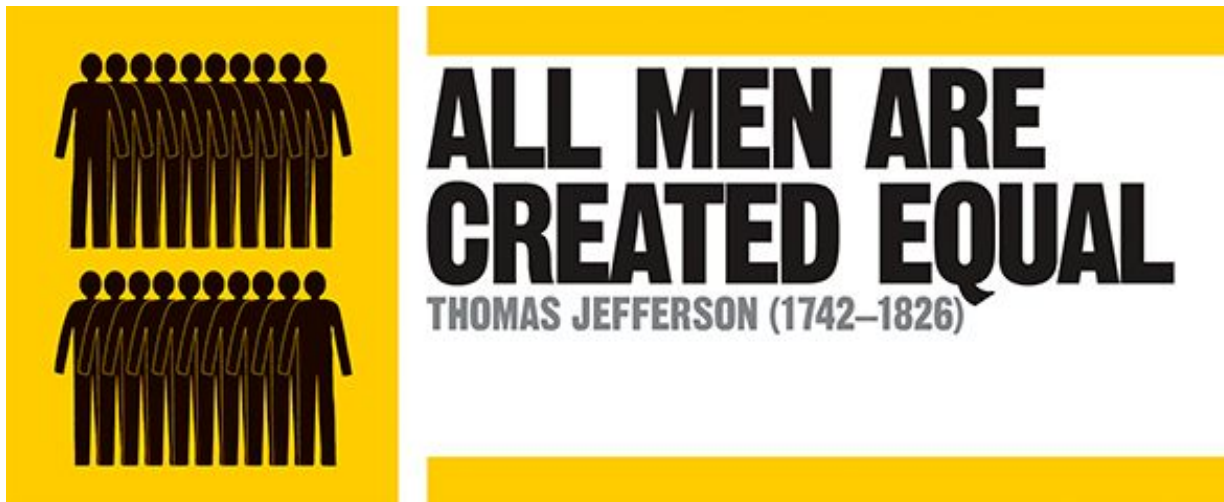


SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

GFT ■



Your PIN Sounds Good!



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

GFT ■

**PIN**

**ALL MEN ARE  
CREATED EQUAL?**

THOMAS JEFFERSON (1742–1826)



Your PIN Sounds Good!



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP

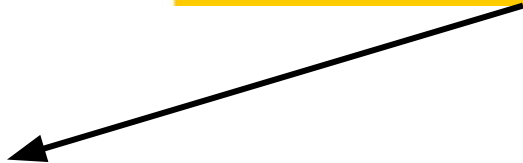


UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

GFT ■

**PIN**

**ALL MEN ARE  
CREATED EQUAL?**  
THOMAS JEFFERSON (1742–1826)



User Chosen





# PIN

# ALL MEN ARE CREATED EQUAL?

THOMAS JEFFERSON (1742-1826)

User Chosen



Random



Your PIN Sounds Good!

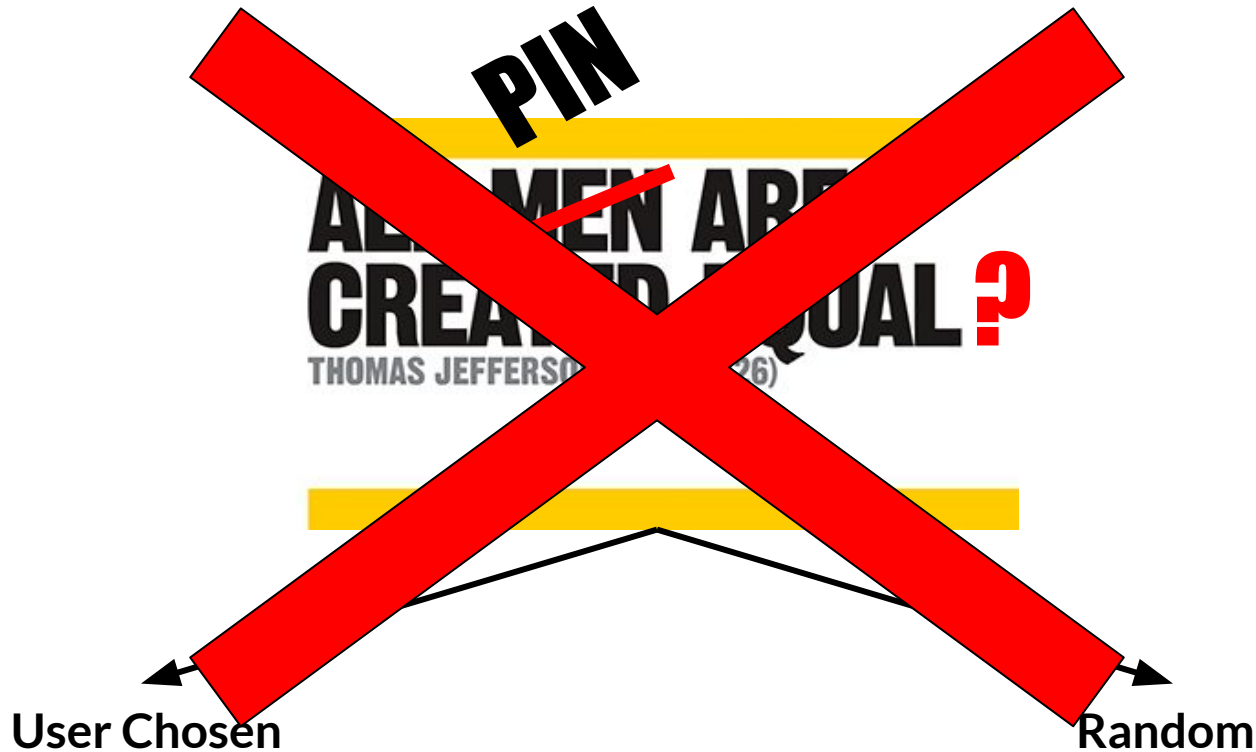


SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

GFT



**DEFINITELY... NOT!**

1122 5555 4321  
0000  
3333 1313 1010

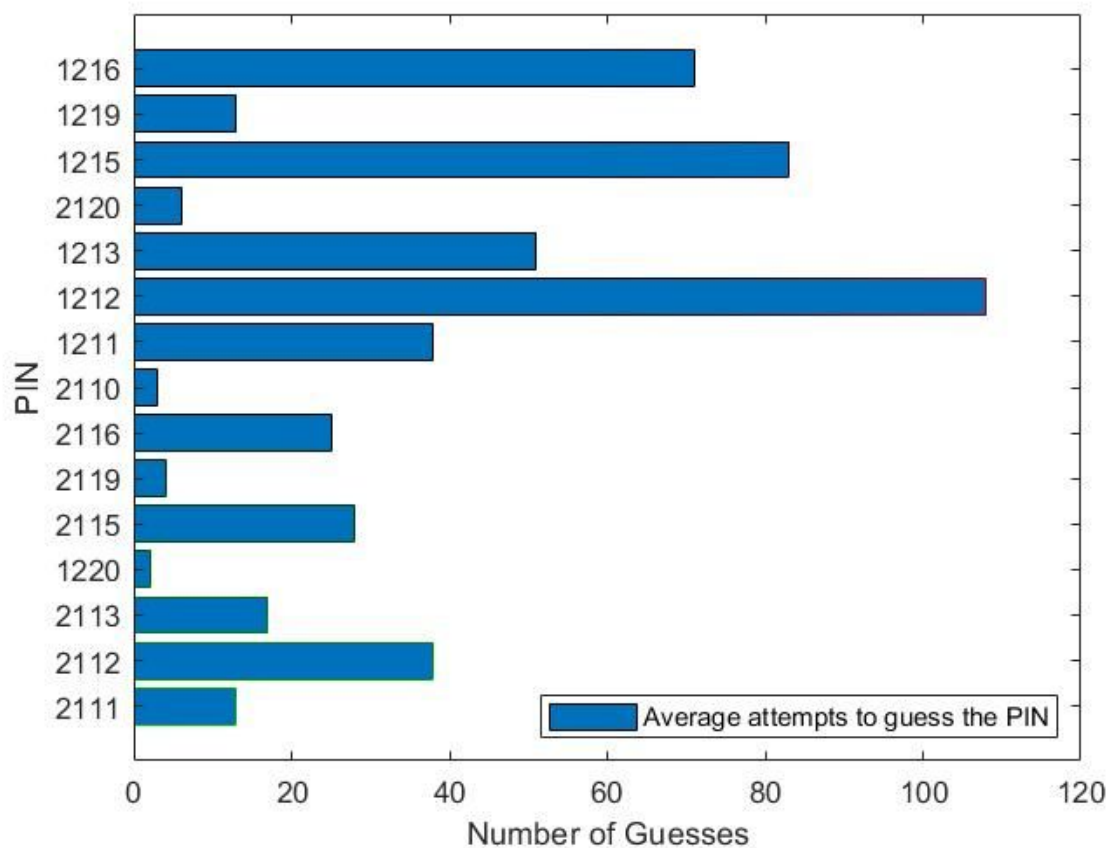


# Your PIN Sounds Good!



## Not all PINs are born the same

*Knowing inter-key distance only*



# Your PIN Sounds Good!

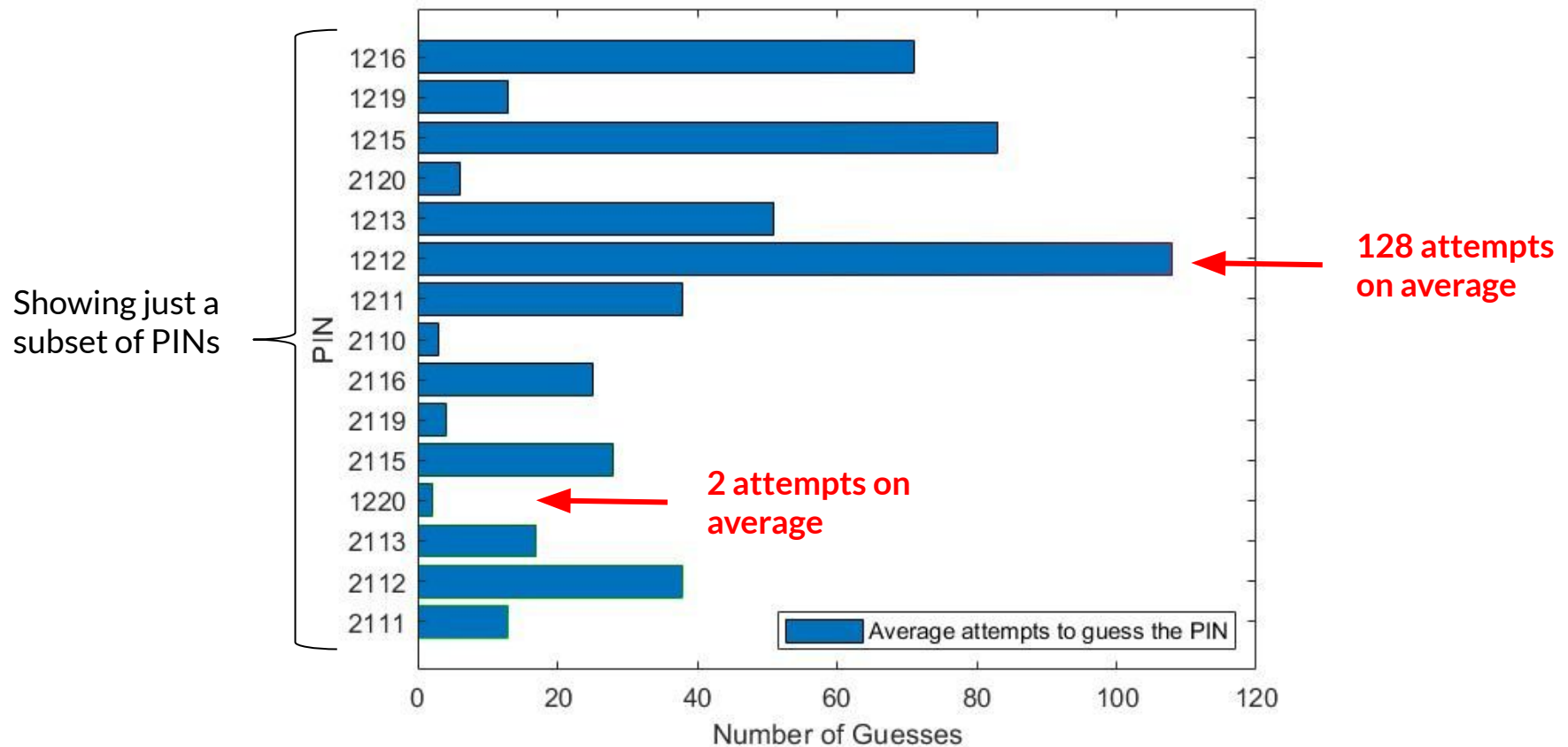


## Not all PINs are born the same

Knowing *inter-key distance* only



*PINs probability distribution is no longer uniform*



DEMO time!



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

GFT ■





SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

GFT ■

**Hand Me Your PIN!**  
**Inferring ATM PINs of Users Typing with a Covered Hand**  
Matteo Cardaioli, Stefano Cecconello, Mauro Conti, and Simone Milani

*In USENIX Security Symposium 2022*



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



Delft University of Technology

GFT ■



# Hand me Your PIN



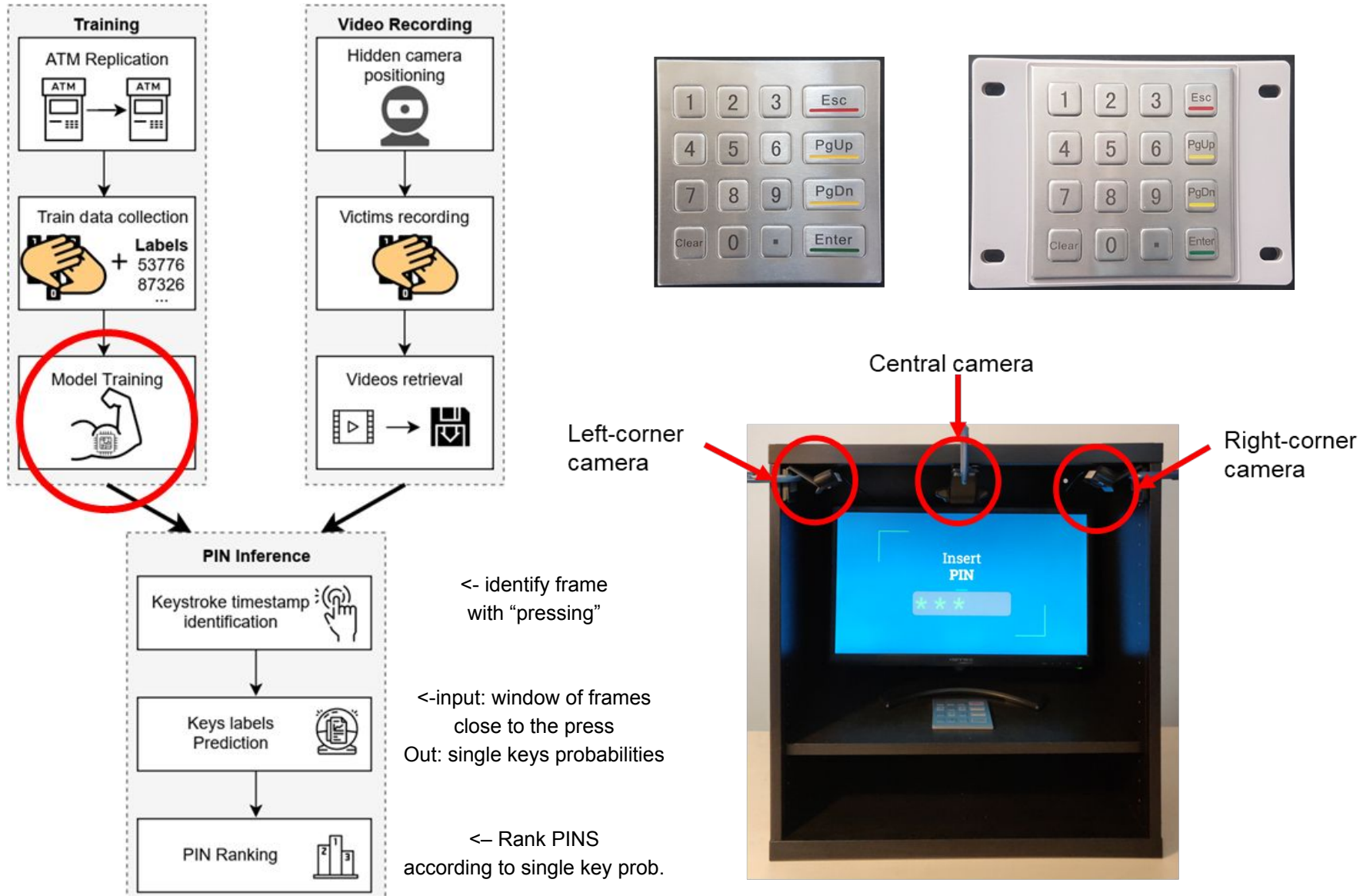
Hidden camera recording the PIN pad



Victim's covering strategy to avoid shoulder surfing attacks

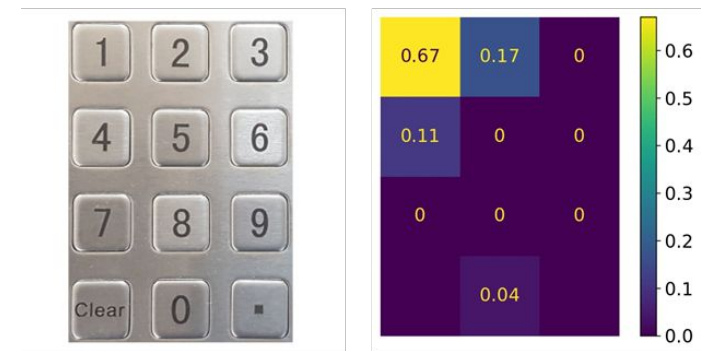
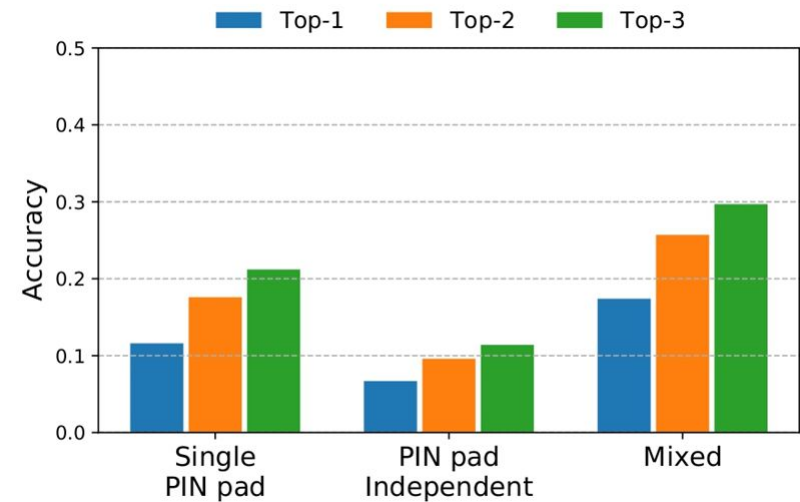


# Experimental Setting



## Attack Scenarios

- **Single PIN pad:**
  - the adversary knows the target PIN pad model and owns a copy
- **PIN pad Independent:**
  - the adversary trains the machine learning model on a PIN pad with a similar (but not the same) layout to the target one.
- **Mixed:**
  - the adversary owns both a copy of the target PIN pad and a PIN pad similar to the target one



heatmap for prediction of Digit "1"

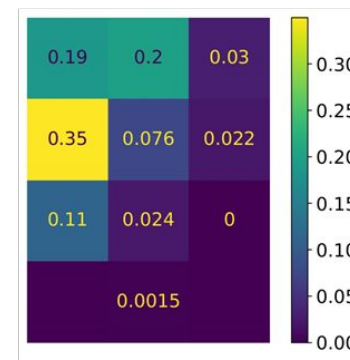
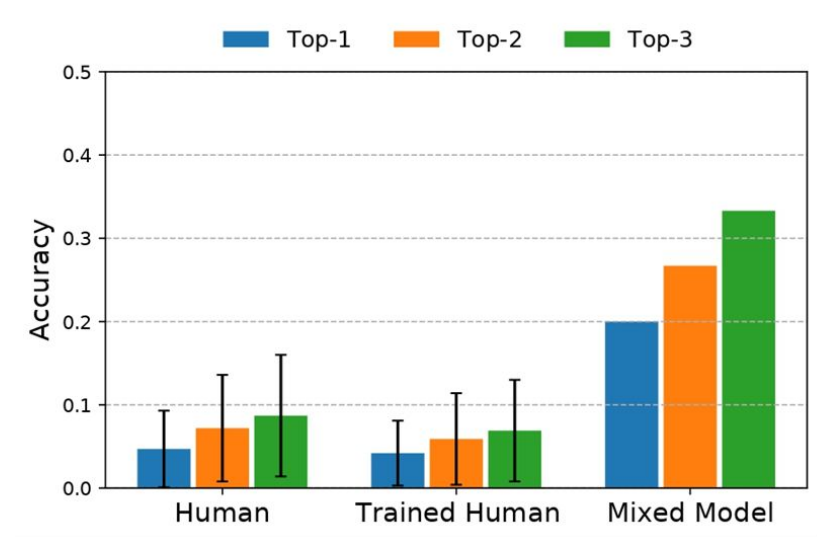
## Human Vs Machine assessment

### Survey:

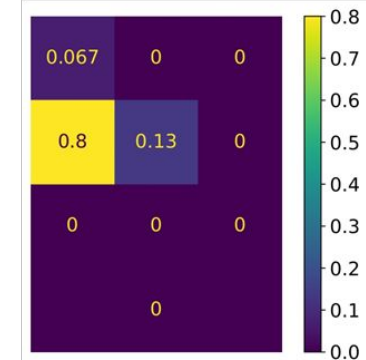
- 30 videos of people entering 5-digit PINs by covering the PIN pad with the non-typing hand
- Videos from the **Mixed scenario** test set (i.e., the only one including both PIN pads)
- Participants had to indicate **the three most likely PINs**

### Participants:

- 45 participants performed the questionnaire **without any training**
- 33 participants **pre-trained** on other covered PIN videos from the test set.



Human  
(non trained)



Mixed Model



# Results



(a) *Side: hand resting on the side of the palm.*



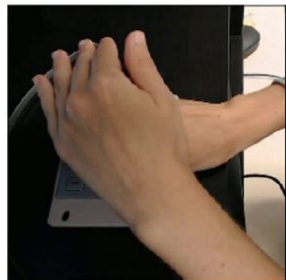
(b) *Over: raised hand not touching the surface.*



(c) *Top: hand resting on fingers and vertically covering the PIN pad.*

Covering strategy	Scenario	Key accuracy	PIN TOP-3 accuracy
Side	Single	0.64	0.30
	Independent	0.42	0.12
	Mixed	0.77	0.53
Over	Single	0.52	0.12
	Independent	0.31	0.10
	Mixed	0.46	0.07
Top	Single	NA	NA
	Independent	0.41	0.13
	Mixed	NA	NA

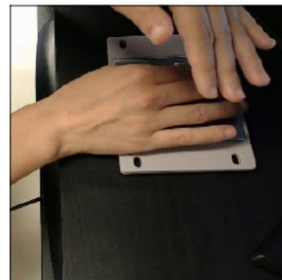
Experiment	Key accuracy	PIN TOP-3 accuracy
Input resolution 125 x 125	0.55	0.23
Input resolution 64 x 64	0.47	0.15
Left-corner camera	0.46	0.10
Right-corner camera	0.62	0.31
Multi-camera training	0.53	0.22
No data augmentation	0.44	0.11
Blacklisted excluded in training	0.54	0.18



(a) *Left-corner camera.*

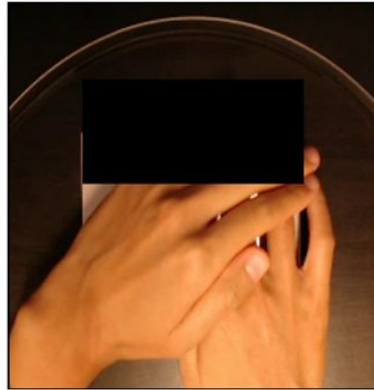


(b) *Center camera.*



(c) *Right-corner camera.*

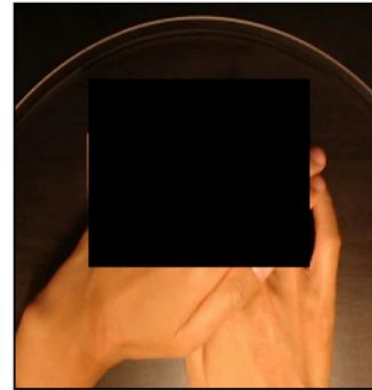
# Countermeasures



(a) 25% of PIN pad surface covered (i.e., digits form 1 to 3).



(b) 50% of PIN pad surface covered (i.e., digits form 1 to 6).



(c) 75% of PIN pad surface covered (i.e., digits form 1 to 9).



(d) 100% of PIN pad surface covered (i.e., no digit is visible).



Coverage percentage	Key accuracy	PIN TOP-3 accuracy
25%	0.54	0.22
50%	0.55	0.22
75%	0.50	0.17
100%	0.33	0.01



- Covert and Side Channels 101
- Network Traffic Analysis
  - *As a side channel: app and sensitive data inference*
- Energy Consumption
  - *As a side channel: user and app inference*
  - *As a covert channel: data exfiltration*
- Device Movement
  - *As a side channel: smartphone user authentication*
  - *Attacks against biometric authentication*
- Keystroke Timing
  - *As a side channel: text typed on keyboards*
- **Acoustic Emanations**
  - ***As a side channel: text typed on keyboards***





SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

GFT ■

**For your voice only**  
**Exploiting side channels in voice messaging for environment detection**

Matteo Cardaioli, Mauro Conti, and Arpita Ravindranath

*In ESORICS 2022*



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

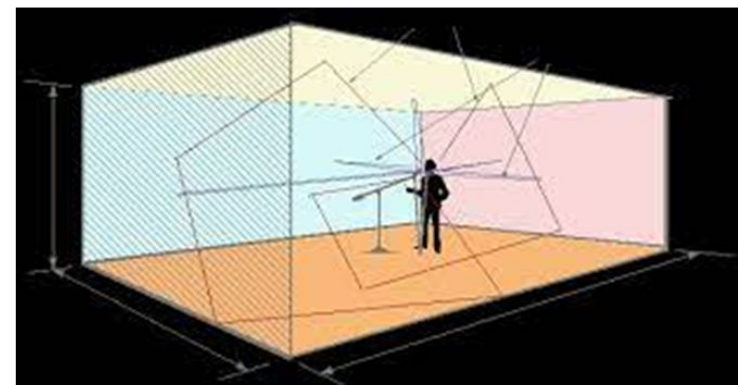
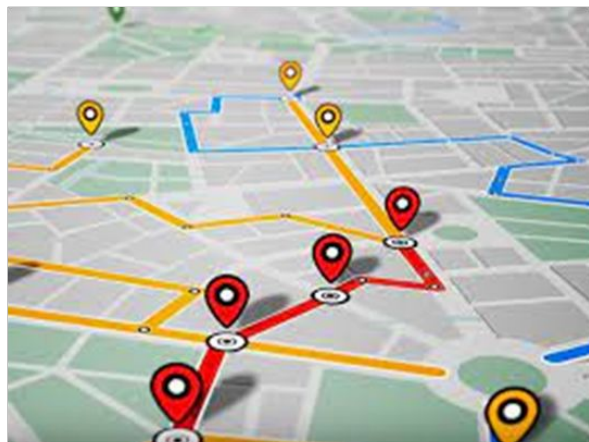


Delft University of Technology

GFT ■

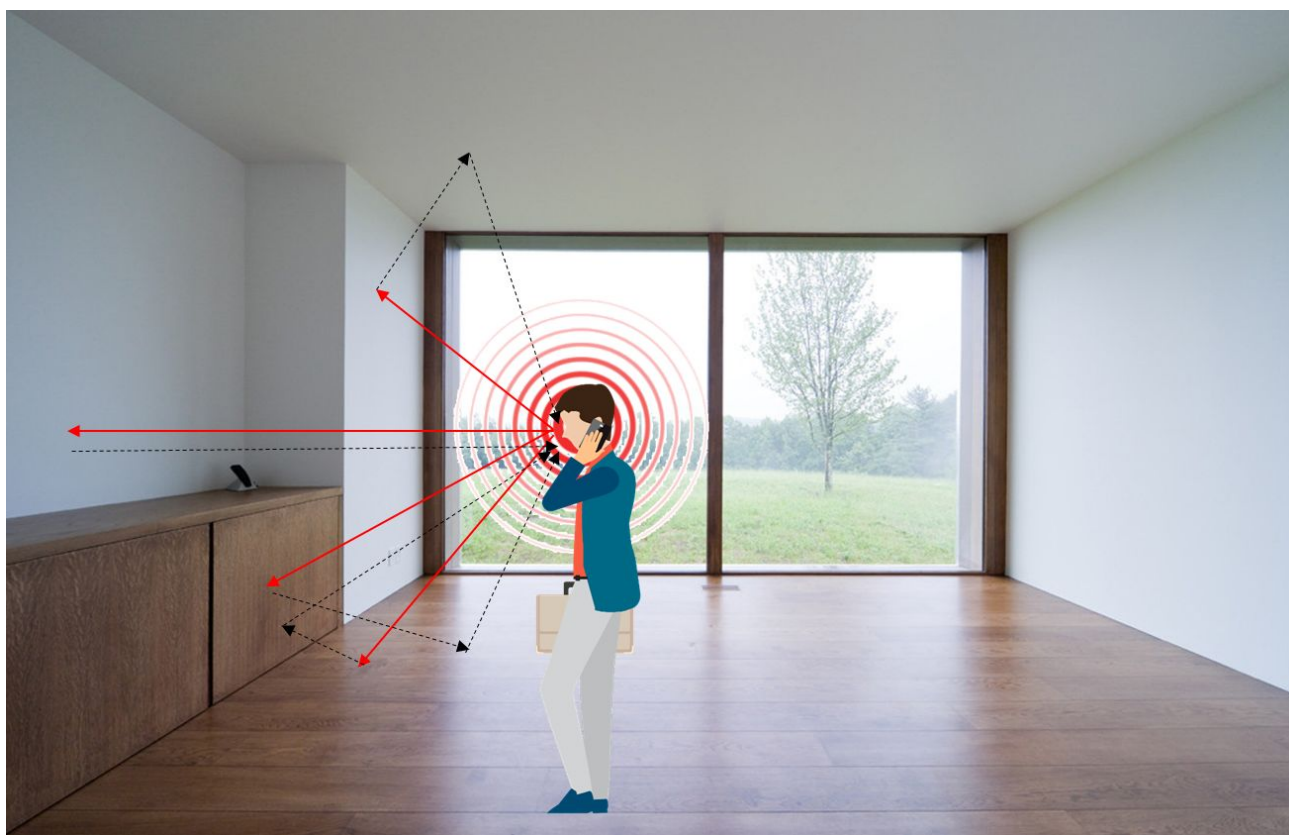


# Is GPS the only way to locate you?

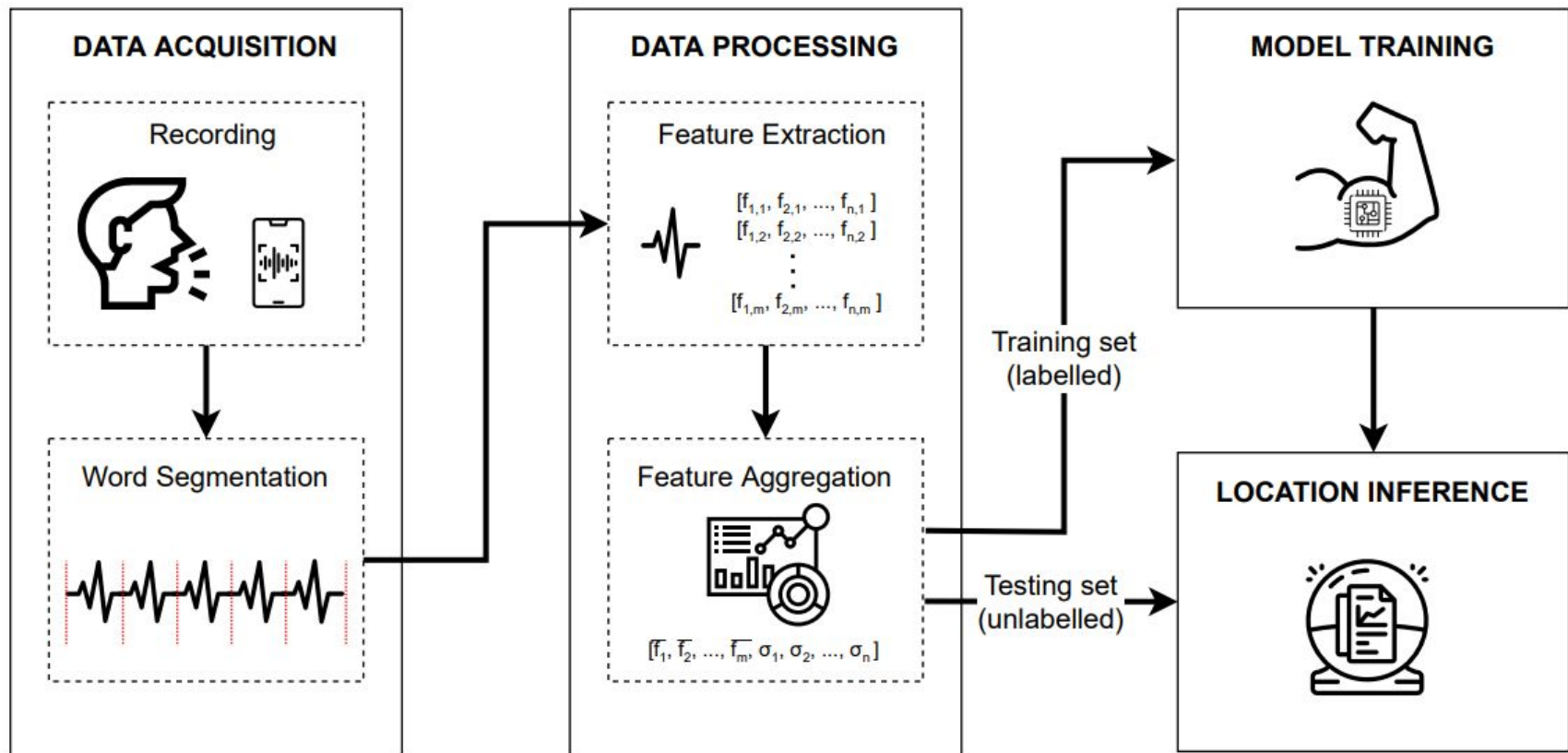




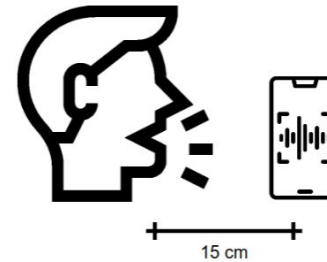
# Can audio messages be used in identification of the location/room?



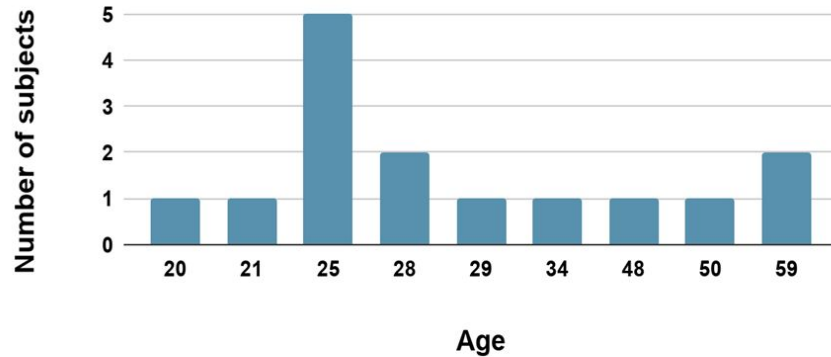
# Experimental Setting



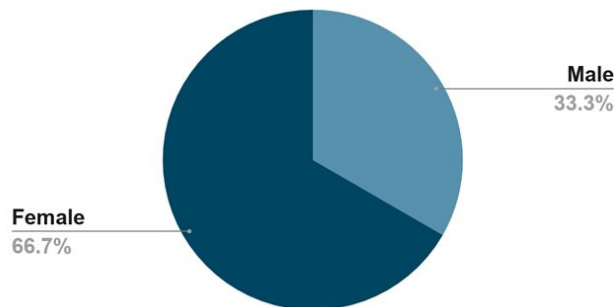
# Experimental Setting



## Age distribution



## Gender Ratio



**LOCATIONS :**  
3 INDOOR  
1 OUTDOOR



**AUDIO CONTENT/SYLLABLES :**



**DEVICE: 14 DIFFERENT DEVICE MODELS**  
{SAMSUNG, ONEPLUS, IPHONE, MOTO}



# Experimental Setting



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

GFT



# Experimental Setting

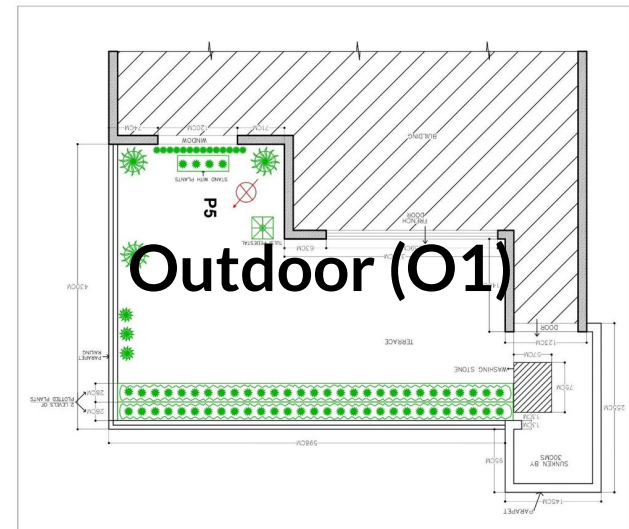
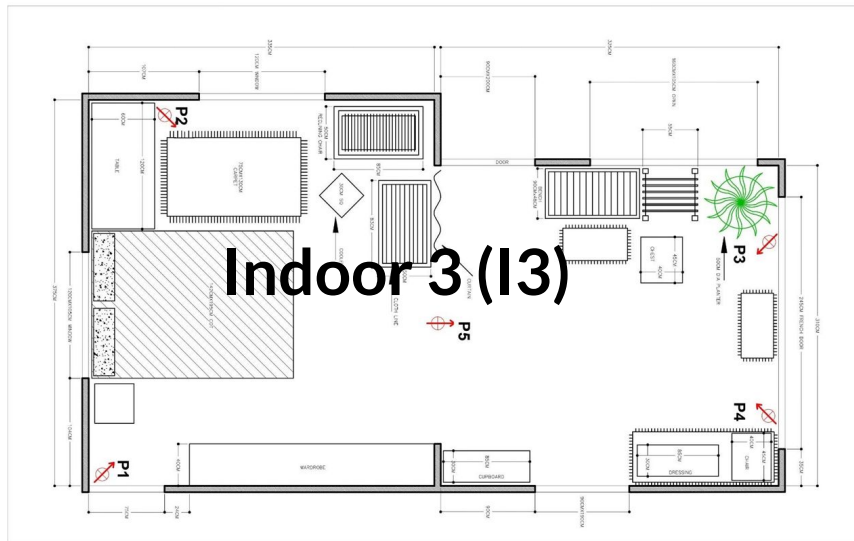
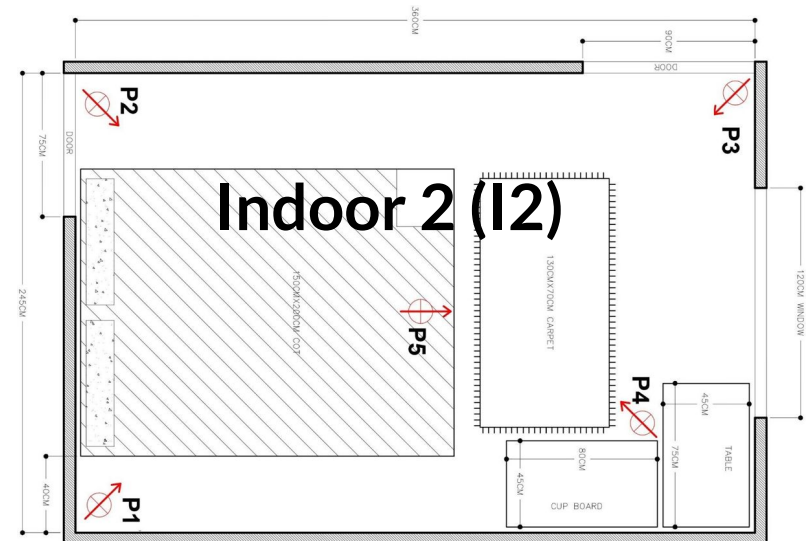
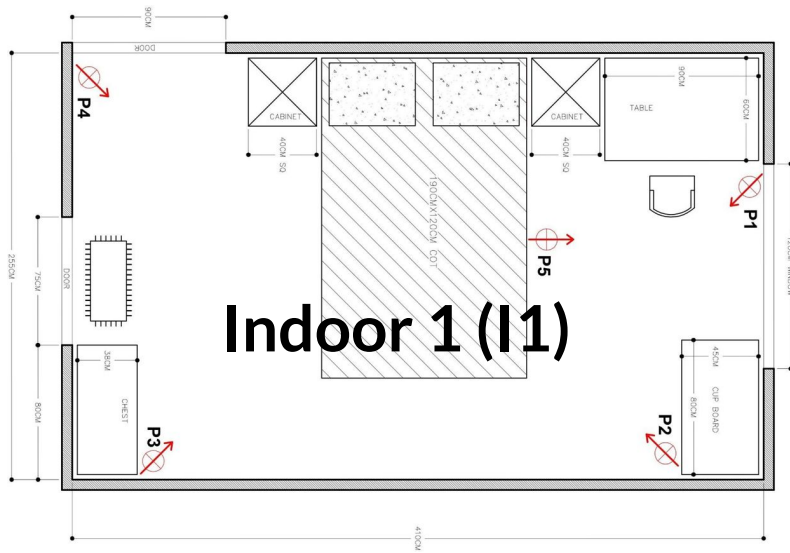


SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

GFT





## SPEAKER KNOWN

Complete Profiling  
("Investigator" case)

LOCATION  
KNOWN



## SPEAKER UNKNOWN

Location Profiling



ADV has samples on

- ALL locations (room and specific position inside)
- but NOT from Victim

LOCATION  
UNKNOWN



User Profiling

ADV has samples

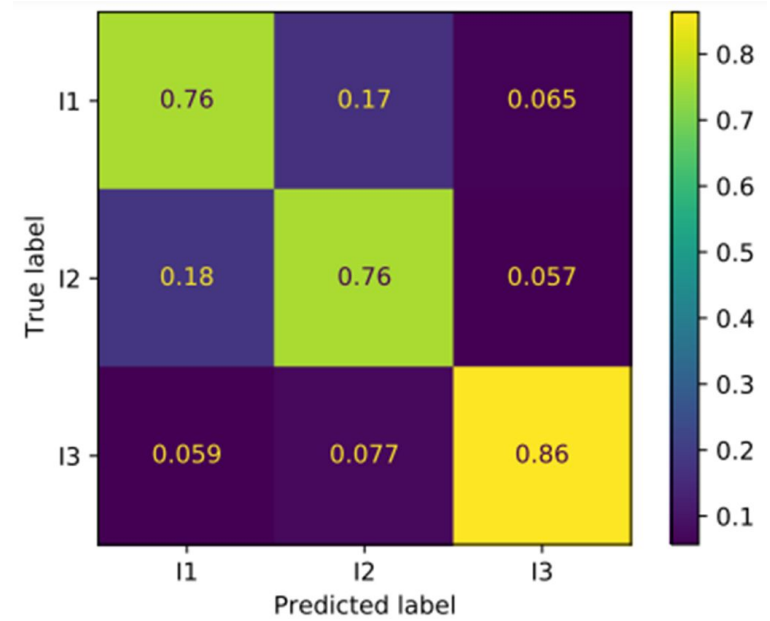
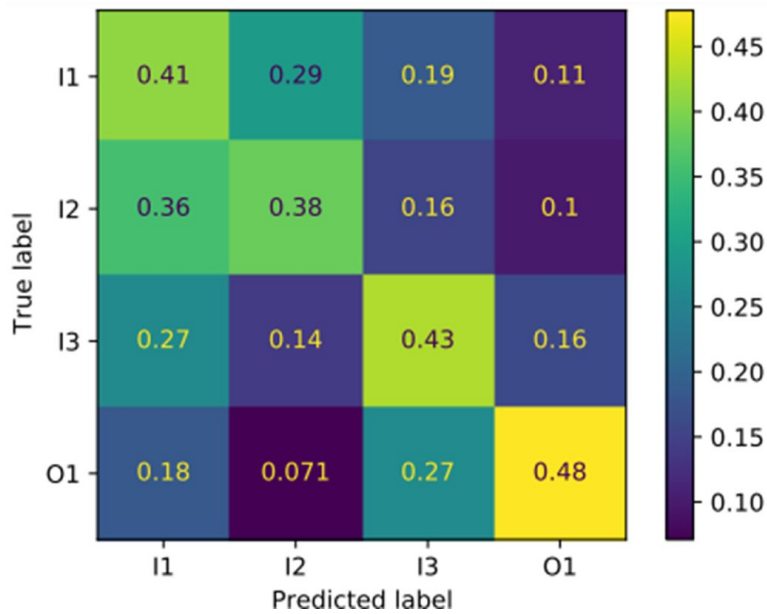
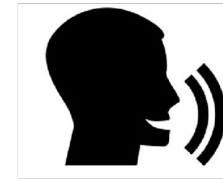
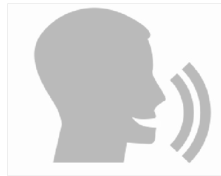
- OF the victim
- In the correct room but in unknown specific position (inside that known room)

# Scenarios (just in case...)

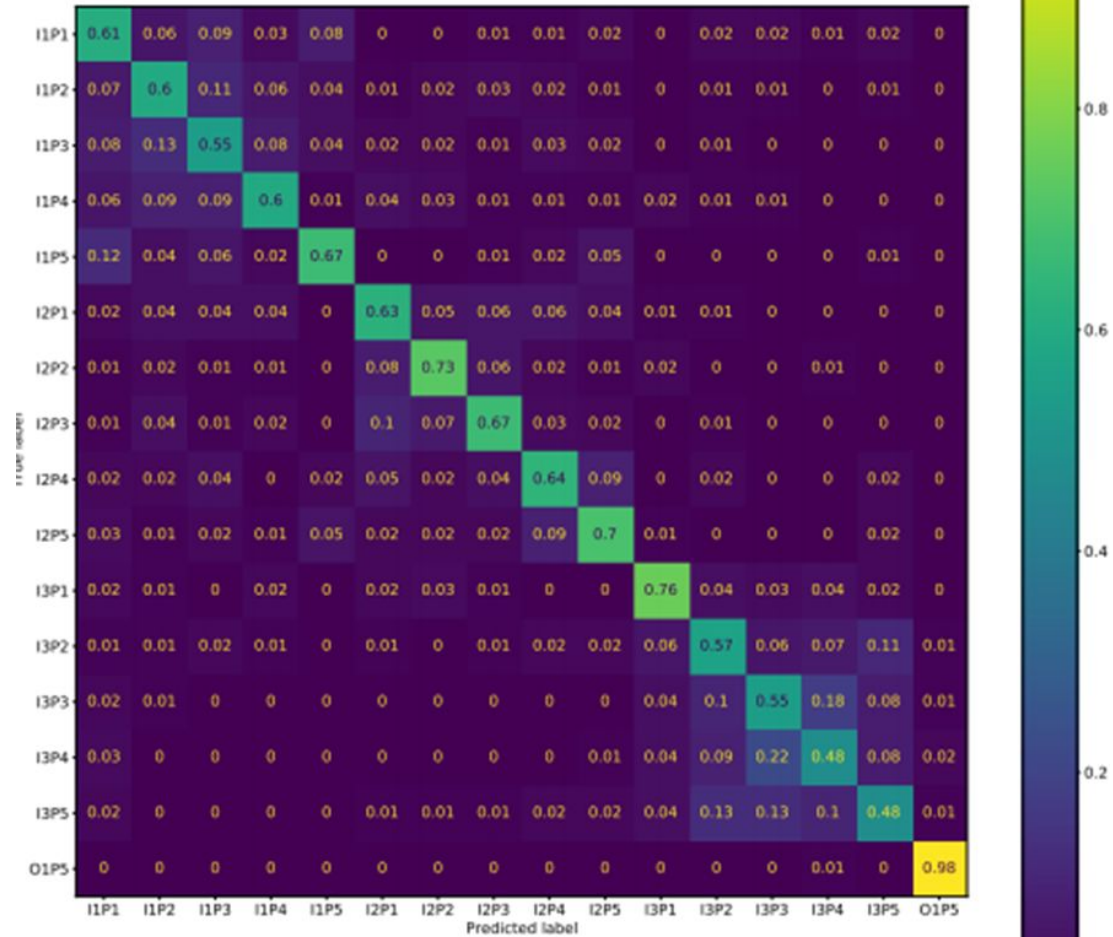
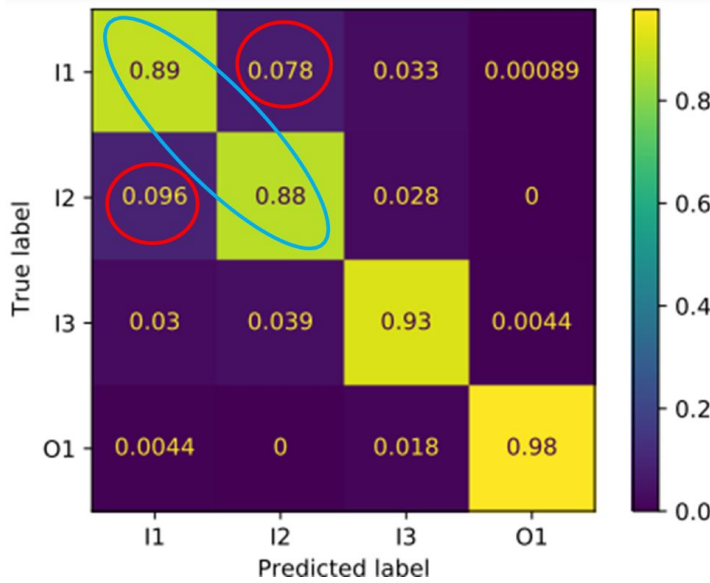


- *Complete Profiling*: This scenario occurs when the attacker asks the victim to send voice messages from specific locations. For example, an investigator (i.e., the attacker) might ask a suspect (i.e., the victim) to stand in a specific part of a room to verify that the suspect was there or elsewhere at the time a voice message was sent. In this scenario, the attacker has recordings of the victim in all the selected locations. Moreover, the attacker also knows the victim's specific position in the selected locations (e.g., a room corner). In this scenario, the attacker has the highest knowledge to execute his attack.
  - *Location Profiling*: In this scenario, the attacker cannot access any of the victim's voice messages other than the one he wants to infer the location. The attacker knows that the victim has sent the voice message from a selected location (e.g., the attacker knows that the victim is in a specific building). Therefore, the attacker can have WhatsApp audio recordings of different speakers but the victim. The speakers are assumed to have recorded their messages at the same locations where the victim is sending the voice message. Hence, the victim is “unknown” while the location position is “known” to the attacker.
    - *User Profiling*: This scenario occurs when the attacker owns the victim's voice messages and knows the recording location but does not know the specific position in the location (e.g., a corner of a room) from which they were recorded. The attacker wants to infer the location of a new voice message sent by the victim. Different from the *Complete Profiling* scenario, the attacker cannot ask the victim to send more voice messages from specific positions of the selected locations (e.g., the victim is no longer reachable). The victim is “known” while the position is “unknown” to the attacker in this situation.

# Results



# Results







SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA





F. Marchiori, M. Conti

**Your Battery Is a Blast!**

**Safeguarding Against Counterfeit Batteries with Authentication**

*In ACM Conference on Computer and Communications Security (CCS' 23)*

# Battery Authentication



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

How many Lithium-ion batteries are around you right now?

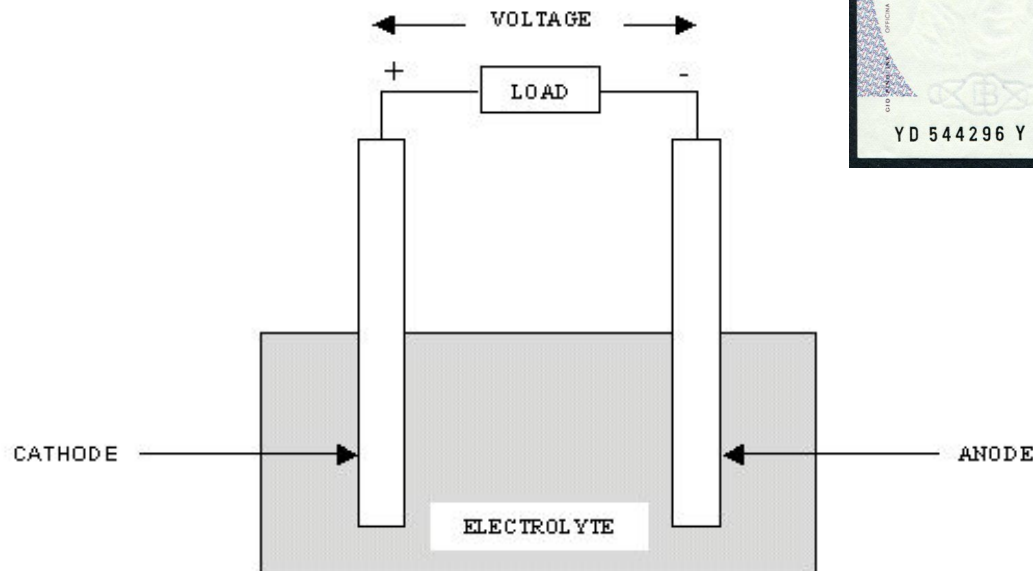




# Battery Authentication



- Store as chemical energy -> turned into electrical energy



# Battery Authentication



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

How many safe Lithium-ion batteries are around you right now?



Lithium-ion (Li-ion) batteries market was estimated to be up to **48 billion U.S. dollars in 2022**

In 2003, roughly **5 million counterfeit cellular phone** batteries were seized worldwide.

[https://www.wilsonelser.com/files/repository/PL\\_eNews0308\\_LithiumIonBatteries.pdf](https://www.wilsonelser.com/files/repository/PL_eNews0308_LithiumIonBatteries.pdf)

In 2016, in a case related to hoverboards with counterfeit batteries, the U.S. customs and border protection agency seized over 16 thousand counterfeit hoverboards with an estimated value of over **USD 6 million**

<https://www.cbp.gov/newsroom/local-media-release/cbp-seizes-record-amount-counterfeit-hoverboards>

How have we checked it until now?  
(tick means defence is successful)

Method	Attacks			
	<i>Cloning</i>	<i>Replay Attacks</i>	<i>Unscalability</i>	<i>Rewrapping</i>
Markings		✓	✓	
External Features		✓	✓	
Form Factor		✓	✓	
Resistor		✓	✓	
Chip	✓			
CR (in clear)	✓			
CR (encrypted)	✓	✓		
<b>DCAuth</b>	✓	✓	✓	✓
<b>EISthentication</b>	✓	✓	✓	✓

CR = Challenge and Response Protocols



## Our contribution

**DCA**uth

**EIS**thentication

- Leverage only internal characteristics of the batteries
- Scalable to many models and architectures
- Small computational cost

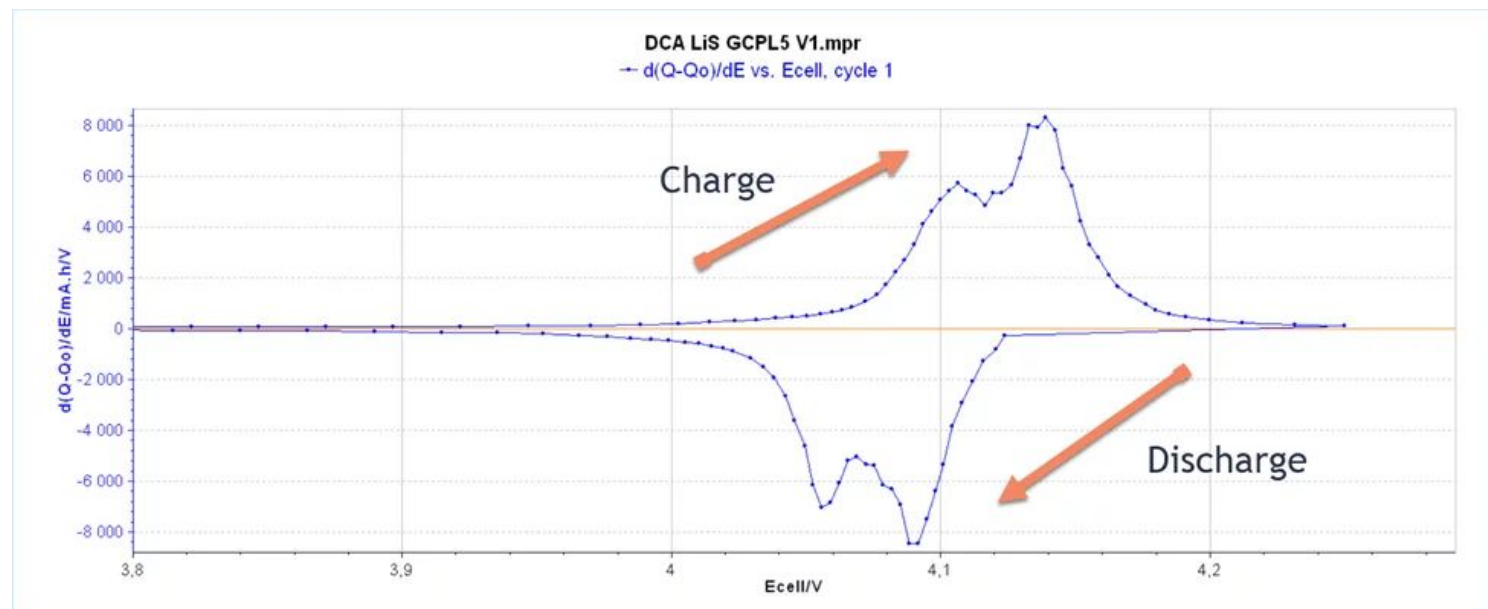
We make dataset and code available.

<https://github.com/Mhackiori/DCAuth>

<https://github.com/Mhackiori/EISthentication>

## Differential Capacity Analysis (DCA)

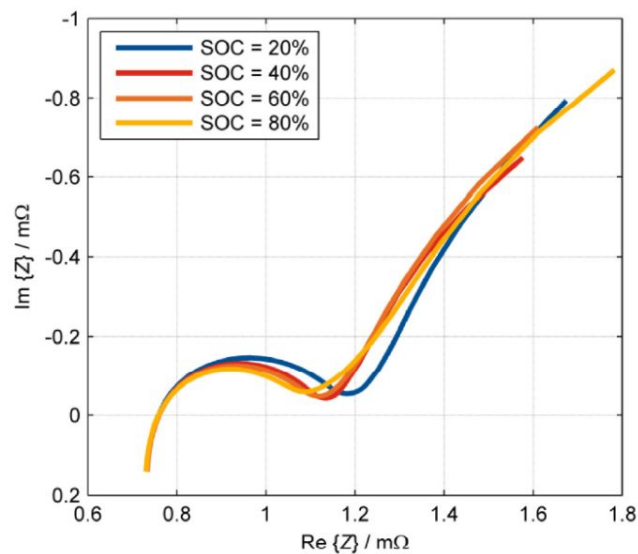
- Measuring change in capacity response in the electrodes
- It tracks increase/decrease in capacity when charged/discharged
- Plot of differential capacity versus voltage



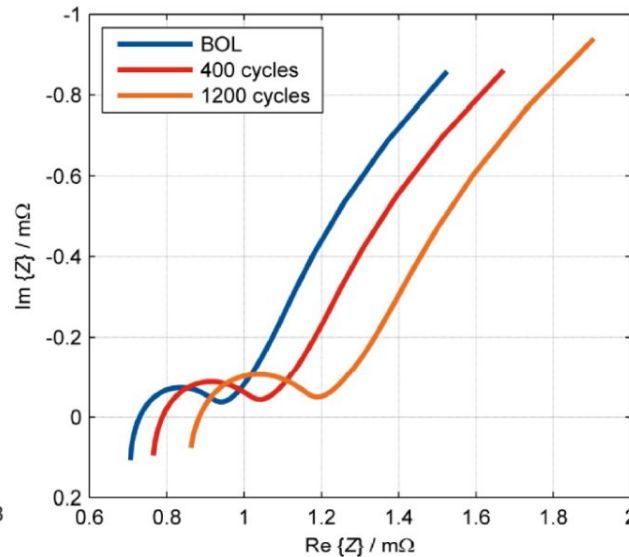


## Electrochemical Impedance Spectroscopy (EIS)

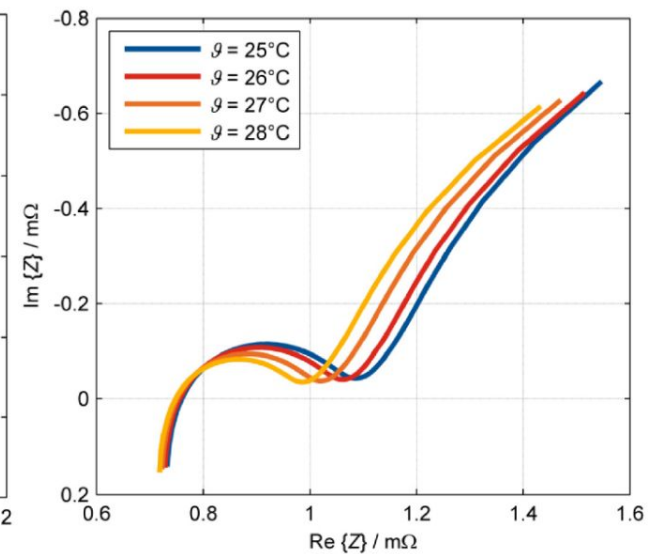
- Analytical technique for electrochemical system characterization
- Measures the electrical impedance
- Dependence on several environment/external factor



(a): Dependence on SOC.



(b): Dependence on SOH.

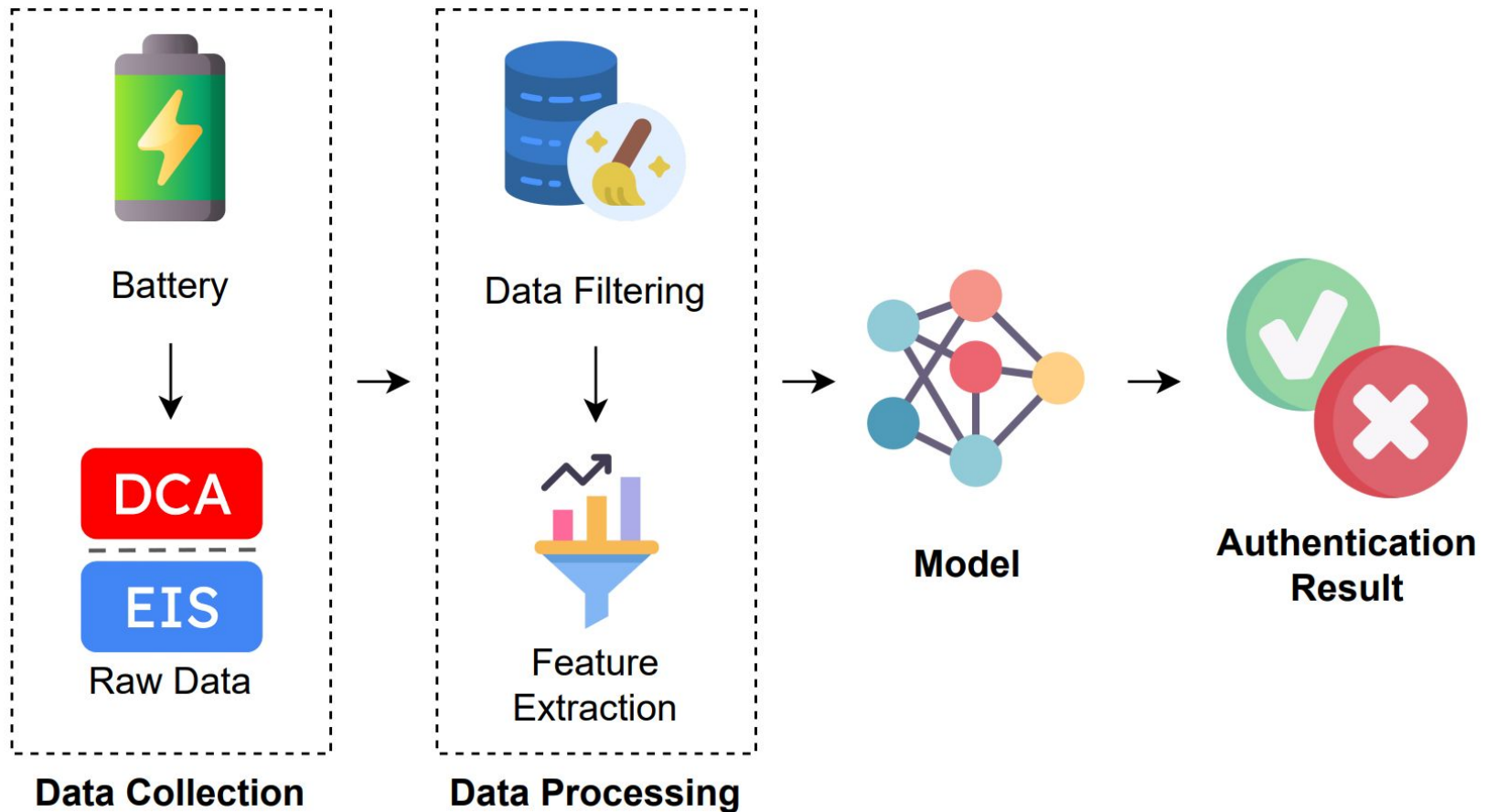


(c): Dependence on temperature.

# Battery Authentication



## System Model

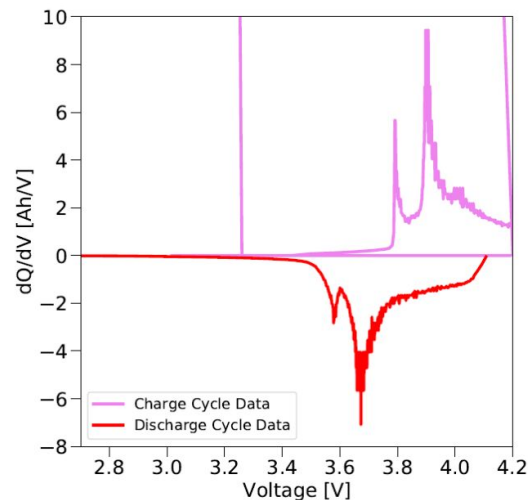




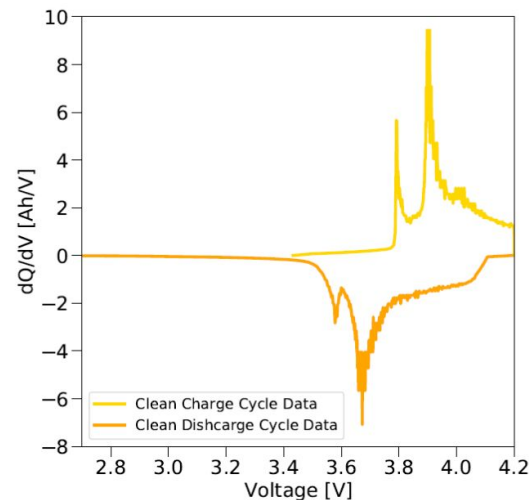
## Datasets

- Issues in finding collaborations with companies or organization
- Collection of available datasets
- 20 datasets (17 for DCA, 3 for EIS)
  - *That includes 11 different models, 5 different architectures*

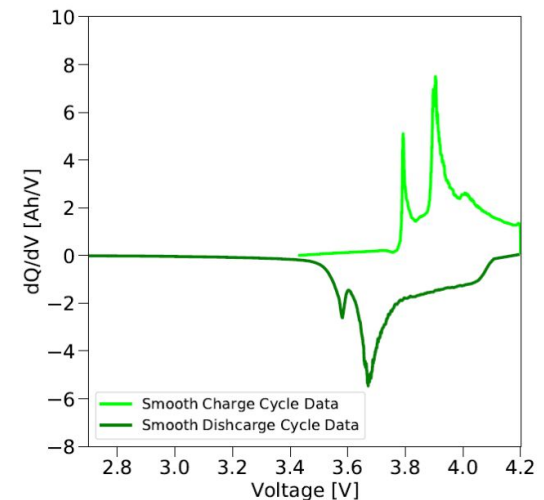
## Processing (available on GitHub)



(a): Raw DCA plot.



(b): Clean DCA plot.



(c): Smooth DCA plot.

## Models

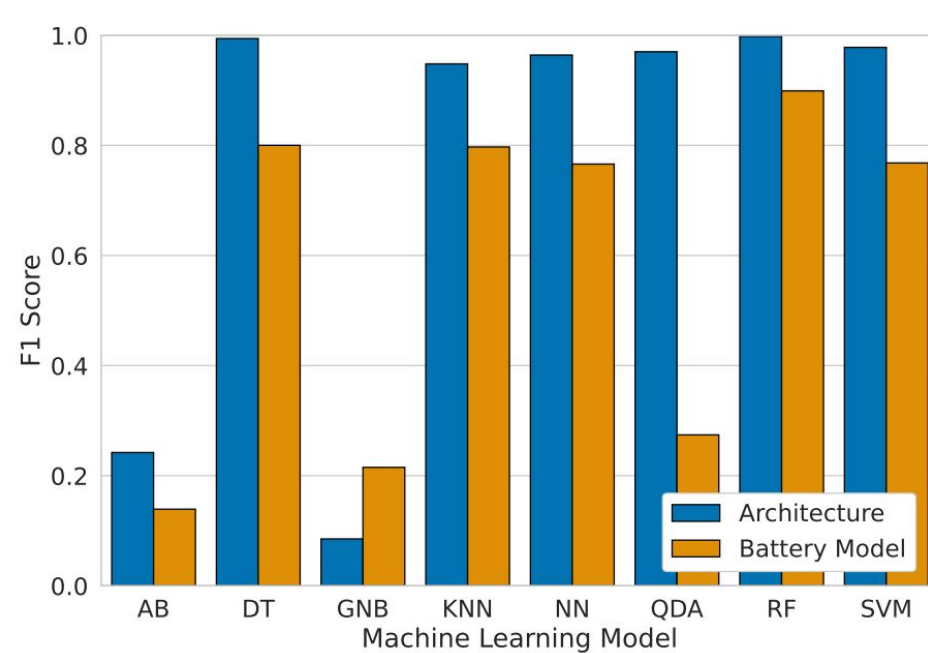
- Machine Learning
- Avoiding complex DL to keep low computational cost
- Commonly used in literature

## Evaluation Metrics

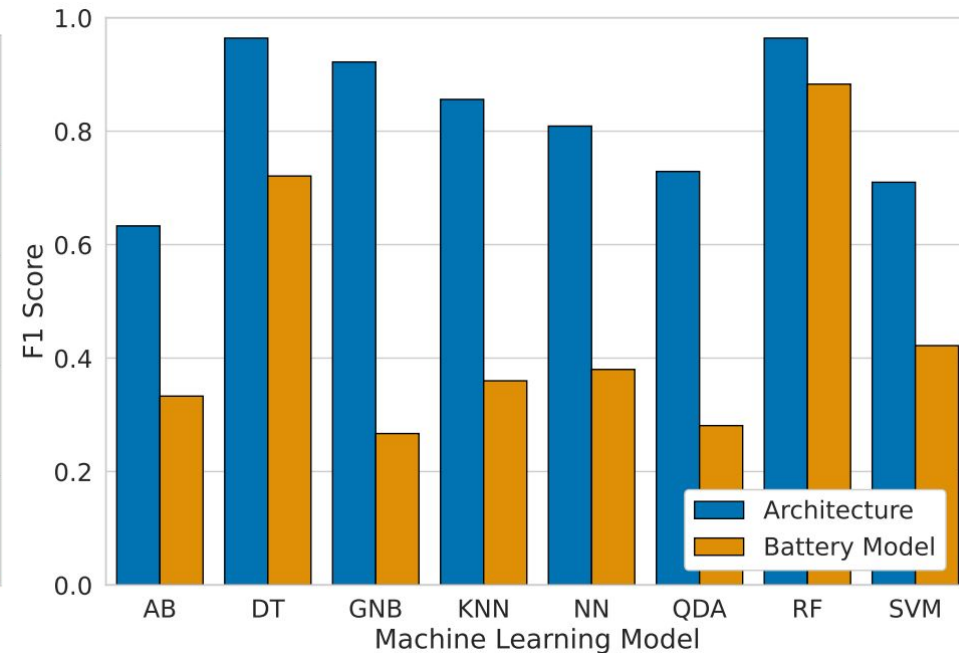
- Precision
- Recall
- F1 Score
- False Acceptance Rate (FAR)
- False Rejection Rate (FRR)

Models	Hyperparameters
AdaBoost (AB)	<ul style="list-style-type: none"><li>• Number of estimators</li></ul>
Decision Tree (DT)	<ul style="list-style-type: none"><li>• Criterion</li><li>• Maximum Depth</li></ul>
Gaussian Naive Bayes (GNB)	<ul style="list-style-type: none"><li>• Variance Smoothing</li></ul>
Nearest Neighbors (KNN)	<ul style="list-style-type: none"><li>• Number of neighbors</li><li>• Weight function</li></ul>
Neural Network (NN)	<ul style="list-style-type: none"><li>• Hidden layer sizes</li><li>• Activation function</li><li>• Solver</li></ul>
Quadratic Discriminant Analysis (QDA)	<ul style="list-style-type: none"><li>• Regularization Parameter</li></ul>
Random Forest (RF)	<ul style="list-style-type: none"><li>• Criterion</li><li>• Number of estimators</li></ul>
Support Vector Machine (SVM)	<ul style="list-style-type: none"><li>• Kernel</li><li>• Regularization parameter</li><li>• Kernel coefficient</li></ul>

## Results - Identification

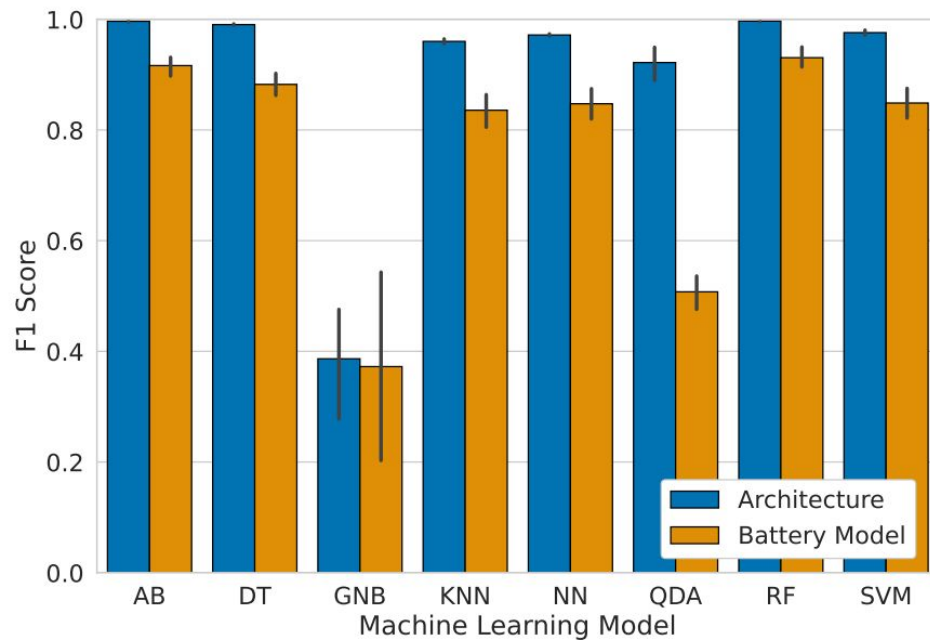


DCAuth

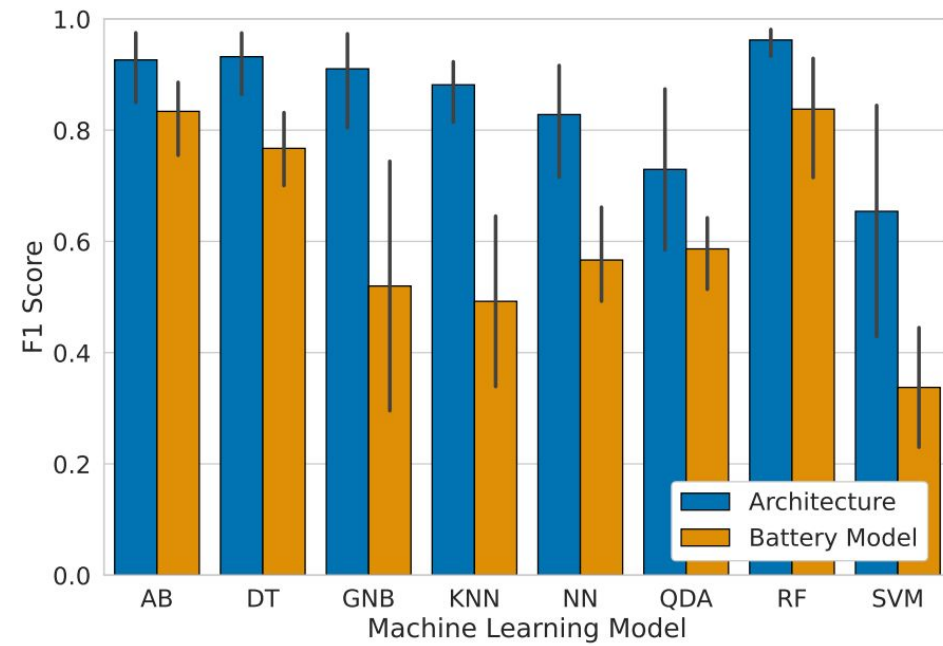


EISthentication

## Results - Authentication

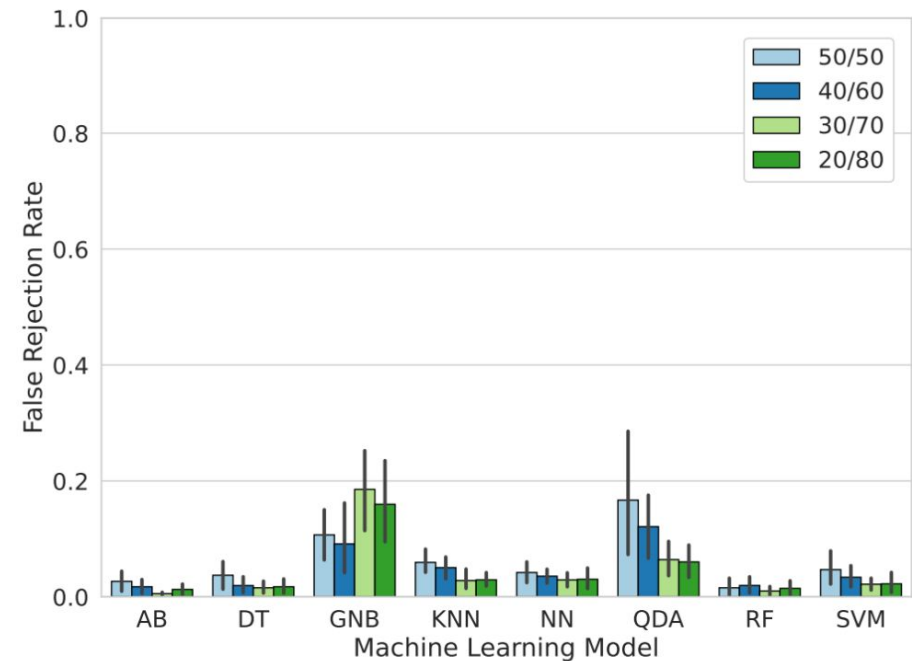
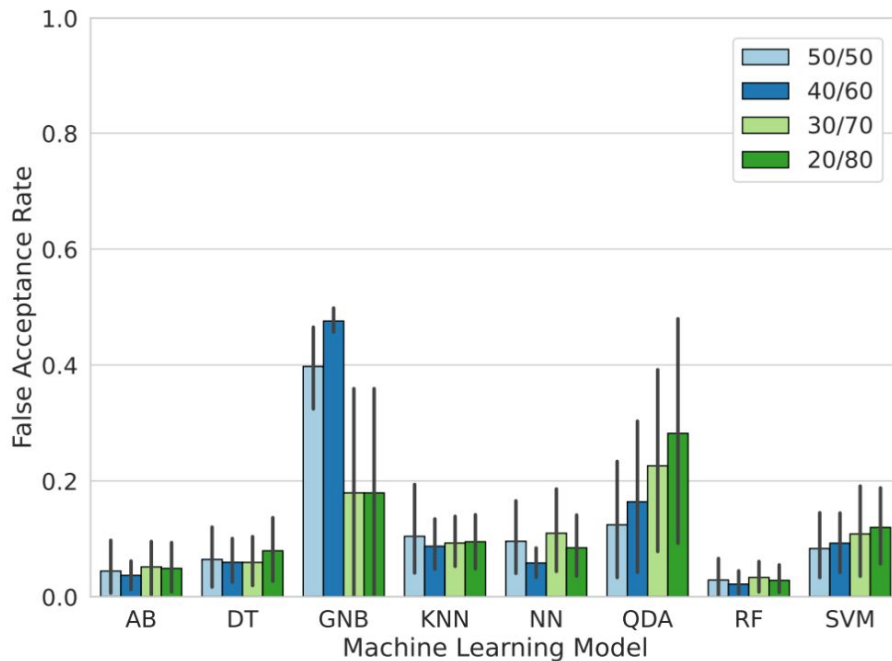


DCAuth



EISthentication

## Results - FAR/FRR on Dataset Balance



DCAuth

**Table 12: Complexity.**

<b>Model</b>	<b>Time<sub>DCA</sub></b>	<b>Size<sub>DCA</sub></b>	<b>Time<sub>EIS</sub></b>	<b>Size<sub>EIS</sub></b>
AB	15.492 ms	75 kB	8.523 ms	59 kB
DT	3.892 ms	31 kB	2.881 ms	20 kB
GNB	4.687 ms	53 kB	3.192 ms	33 kB
KNN	12.951 ms	4800 kB	7.1 ms	263 kB
NN	4.595 ms	2600 kB	3.204 ms	1200 kB
QDA	7.856 ms	3100 kB	4.435 ms	271 kB
RF	13.661 ms	348 kB	13.288 ms	221 kB
SVM	9.854 ms	500 kB	2.99 ms	158 kB



## Conclusions and Follow-ups

- Important issue to address for user safety
- More data can improve the methodology
- Collecting data in various condition can enhance the adaptability of the system

<https://arxiv.org/abs/2309.03607>







SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



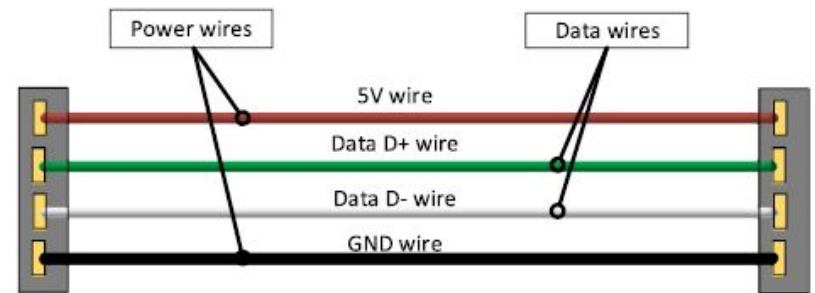


R. Spolaor, H. Liu, F. Turrin, M. Conti, X. Cheng

**Plug and Power: Fingerprinting USB Powered  
Peripherals via Power Side-channel**

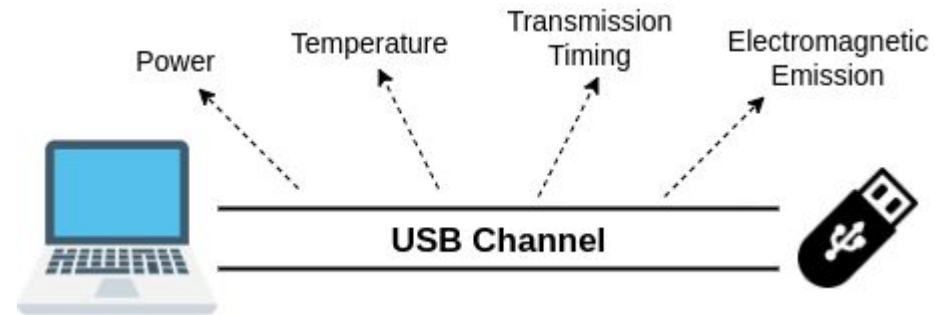
*In IEEE International Conference on Computer Communications (INFOCOM) 2023*

- Widely used in everyday life
  - Peripheral devices, smartphone, IoT
- Data Transfer + Power supply
- **No security** measure by design
- Common attack vectors
  - Malware, BadUSB, USBkill



## Exploit Power Side-Channel to identify authorized devices

- Identification of **legitimate devices**
- Recognize **legitimate actions**
- Detect **malicious devices**



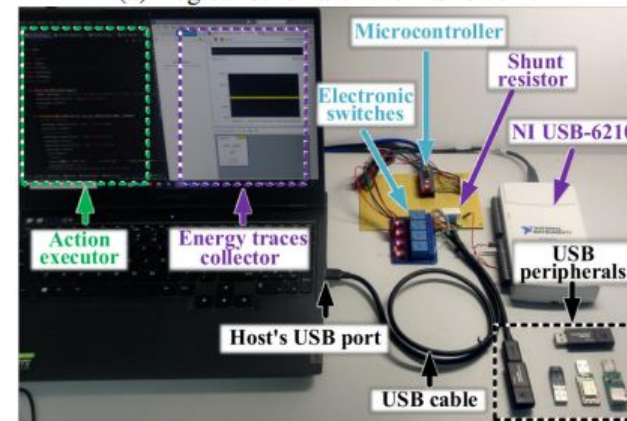
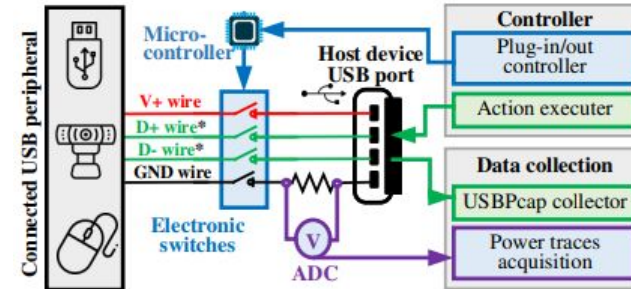
## Use cases

- End-user Personal Protection
- Organization Assets Protection



## USB Power traces collection

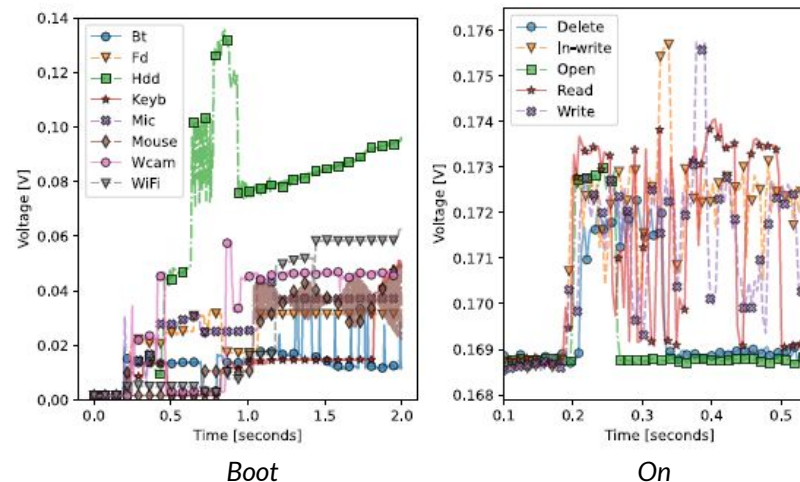
- 82 different devices
  - 8 types
    - HDD, USB stick, WiFi & Bluetooth adapters, mouse, keyboard, webcam, microphone
  - 35 models
- Automated collection
- Different action
  - *Boot*
  - *On* (operating mode)
  - *Actions* (e.g., read, write, connect)
- Univariate time series



# Analysis Goals



1. **Type** (during *Boot* and *On* states)
2. **Model** (*Boot* and *On*)
3. Specific **Device** among the ones with same model
4. **Action** given a device type
5. Given a type, **Device via action**
6. **Good vs. Bad** (malicious USB peripherals)



## 1) Traces Preprocessing

- a) Segmentation: sliding window (1 second with a 75% overlap)
- b) Feature extraction with tsfresh libraries (740 features per segment)



## 2) Model tuning

- a) Random Forest classifier (each task)
- b) 70% training, 10% validation, and 20% test (stratified)
- c) SMOTE to balance classes

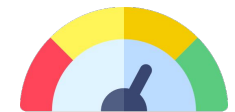


## 3) Classification approaches

- a) Multiclass with “Other” class
- b) Binary (One-vs-All strategy) with Unknown devices in test



## 4) Evaluation Metrics: Precision, Recall, F1-Score, G-Mean, AUC



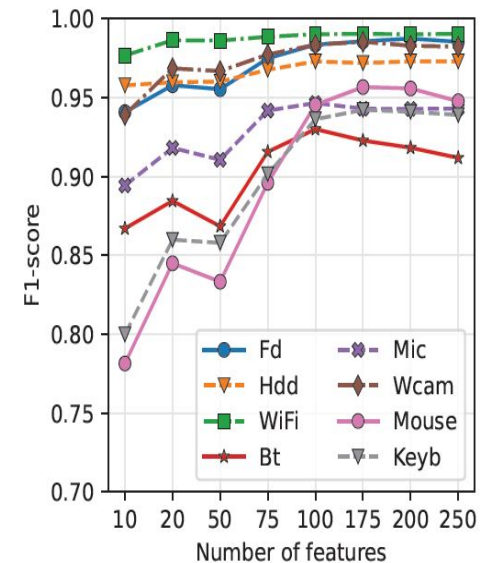


# Type Recognition - Results (1/6)

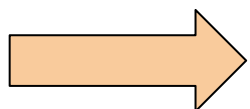
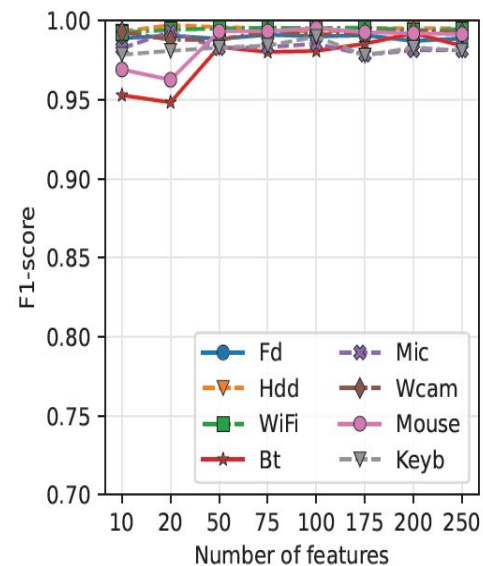


- Recognize the type during *Boot* and *On* states
- *Multiclass* approach
  - 8 classes
  - *Other* includes random traces
- *Boot*: Mouse and Keyb (upon visual inspection)
  - Very quick (below 0.5 second)
  - LEDs may introduce noise
- *On*: simple to detect

State Boot



State On



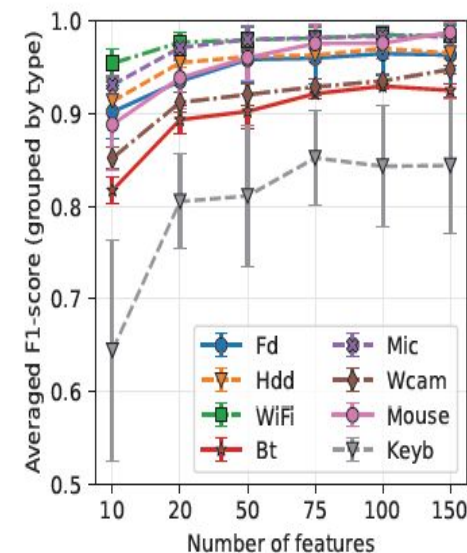
We can discriminate USB type for *Boot* and *On*

# Model Recognition - Results (2/6)

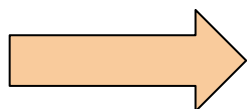
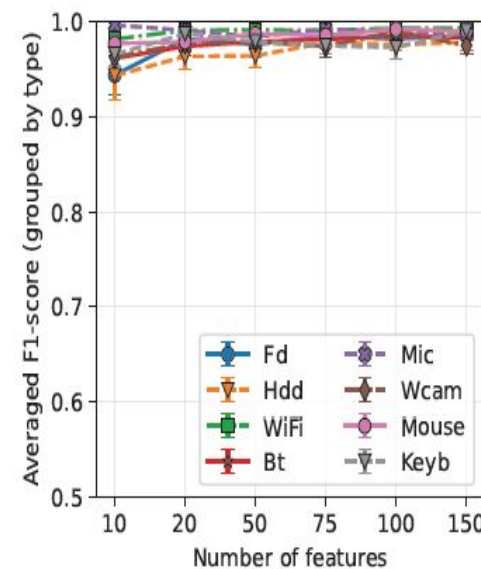


- Recognize the model during *Boot* and *On* states
- *Multiclass* approach
  - 35 classes
  - *Other* includes random traces
- *On*: high classification performance
- Keyb3 and Fd8 perform worst
  - *Very quick (below 0.5 second)*
  - *LEDs may introduce noise*
- Accurate fingerprint with 75 features both *Boot* and *On*

State Boot



State On



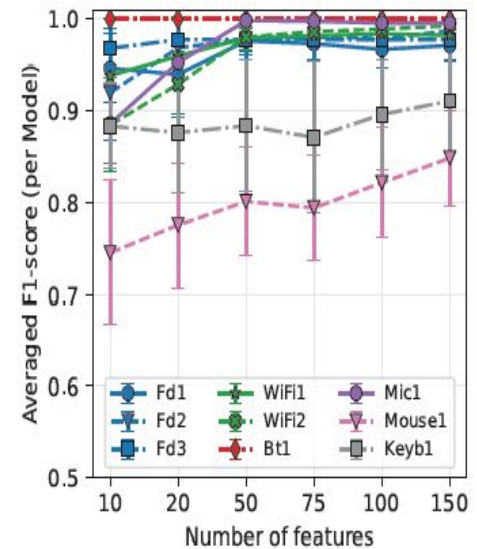
**We can discriminate USB model for Boot and On**

# Device Recognition - Results (3/6)

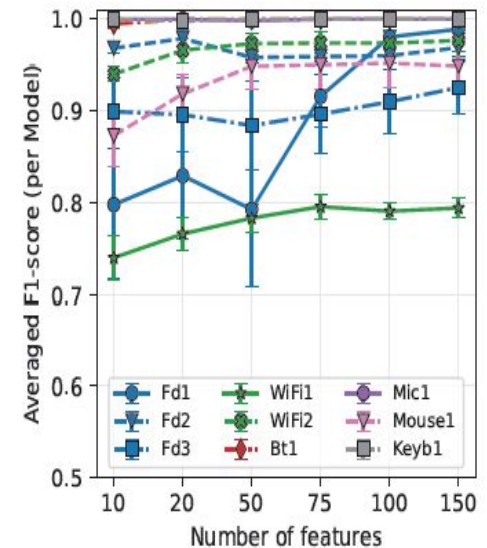


- Given peripherals of the same model identify the specific device
  - Models with  $\# \geq 4$  individual devices
- Binary approach
  - One random class not in Training set
- No good results on Mouse1 and Keyb1 state *Boot*
- WiFi1 model has the lowest score on state *On*
  - Models' traces are very similar

State Boot



State On



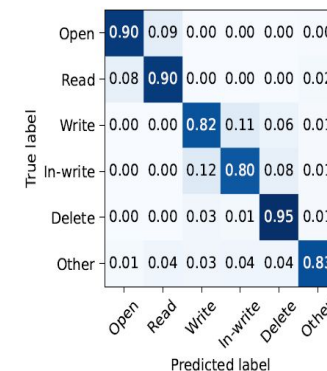
We can almost discriminate the specific USB device

# Action Recognition - Results (4/6)

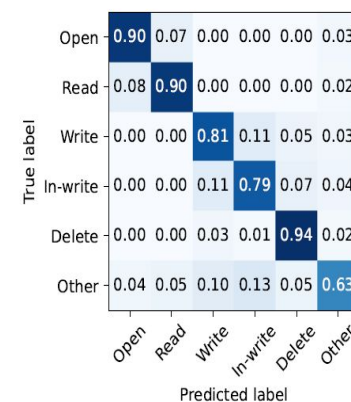


- Recognize an ongoing action given a device type
- *Multiclass* approach
  - *Fd, Hdd, and WiFi*
  - *Other* includes random actions
- WiFi type have a clear fingerprint
- Miss-classification between Write and In-Write
  - *In-Write is derived by the combination of Read and Write*

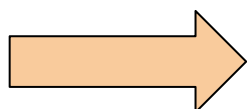
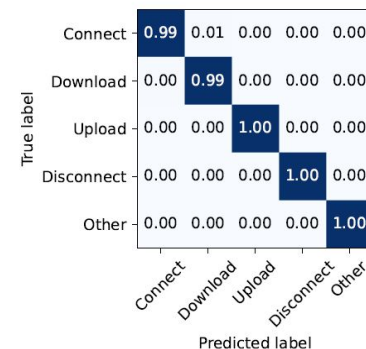
Flash Drive



HDD



WiFi adapter

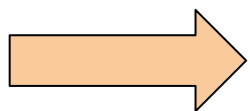
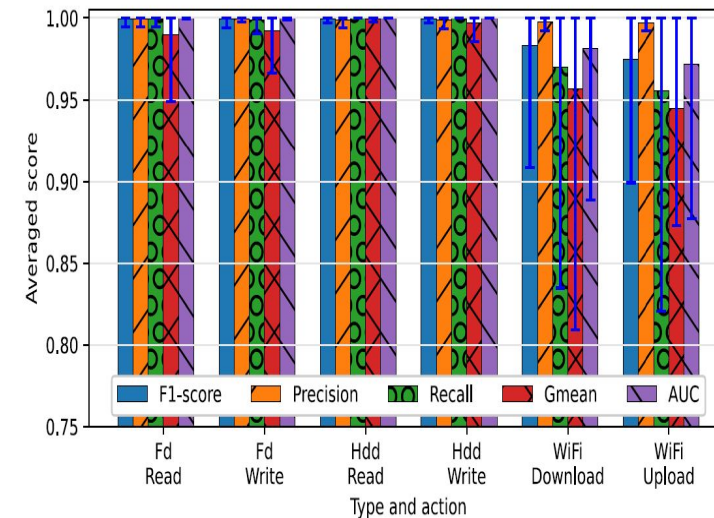


**We can discriminate action given a type**

# Device via Action - Results (5/6)



- Given an action for a type, identify specific device
- *Binary* approach
  - Fd, Hdd, and WiFi types (46, 10, and 38 classes)
- Good performance for all the types and actions
- Fd and Hdd actions are distinguishable
- WiFi slightly lower performance (similar behavior)



**We can fingerprint an individual device from its actions**

# Bad vs. Good - Results (6/6)

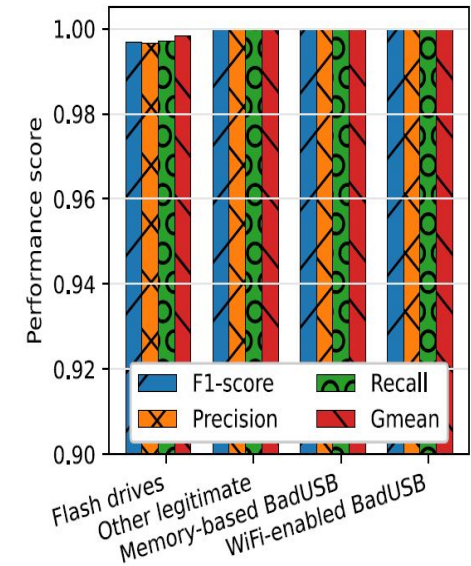


- Discriminate between
  - Flash Drives
  - Bad USBs
- *Multiclass* approach
  - 3 classes
  - *Other legitimate* includes other legitimate peripherals
- While collecting traces we run several attacks
  - command injection, WiFi scanning and connection
- Good scores according to all metrics

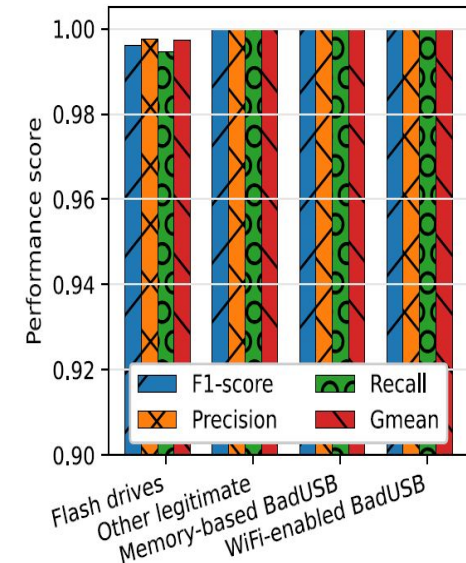


**We can discriminate Bad USBs**

State Boot



State On







- USB devices are still a common attack vector
- Evolution of the standard did not include any security
- Power consumption allows USB fingerprinting
  - *State*
  - *Type*
  - *Model*
  - *Specific device*
  - *Action*
  - *Malicious devices*
- Protect the host from USB-based threats
  - *Non Intrusive*
  - *Privacy preserving*





SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA





M. Conti, E. Losiouk, A. Visintin

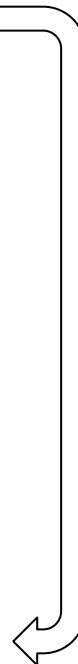
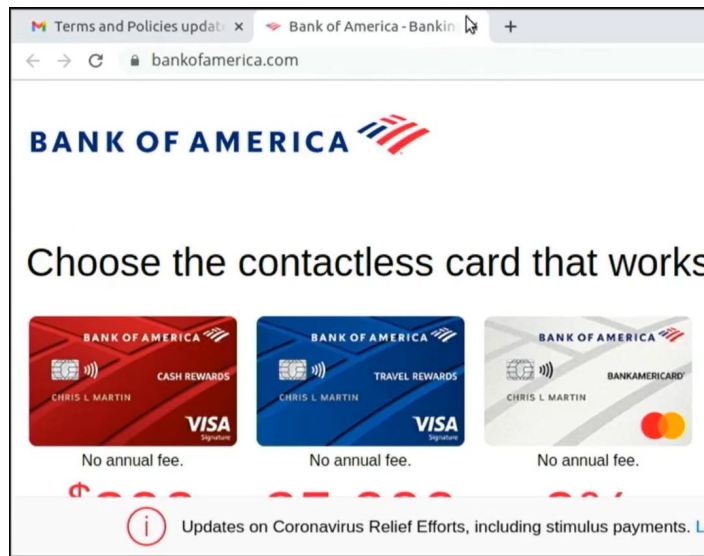
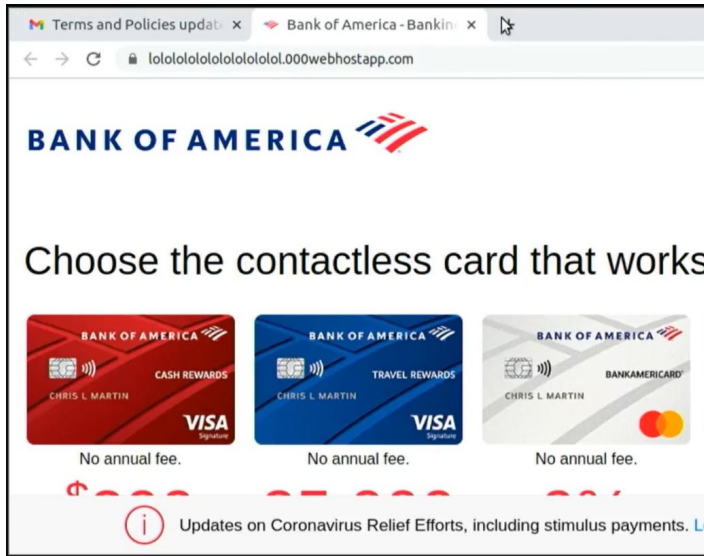
**What You See is Not What You Get**  
**A Man-in-the-Middle Attack Applied to Video Channels**

*In ACM/SIGAPP Symposium On Applied Computing 2022*



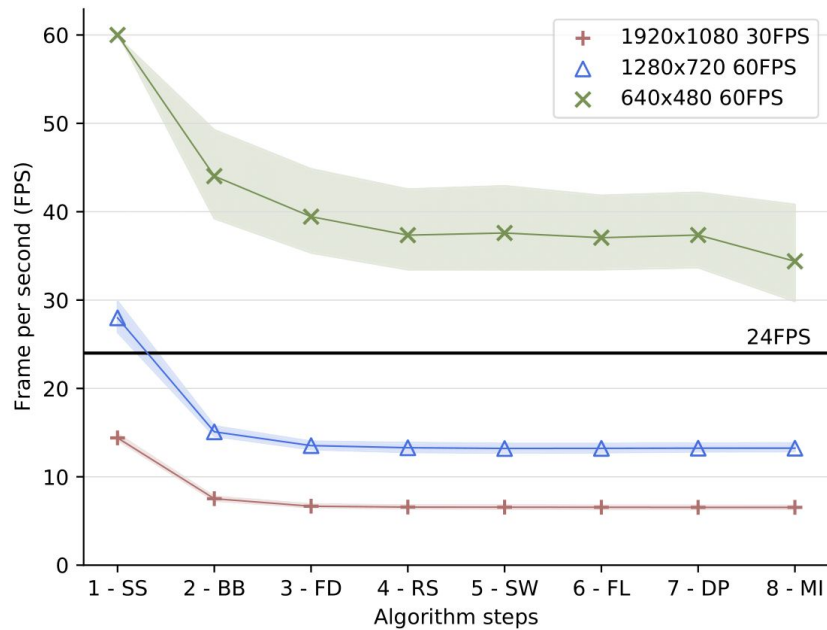
## Man-in-the-Middle attack on a video channel.

*Using a Raspberry PI to modify in real-time the HDMI output before it is displayed.*



Phishing replica of Bank of America website.

*Raspberry PI detects and modify the URL into a legit one.*



Measured performances show the practicality of the attack.  
*The frame rate can be substantially improved using dedicated hardware.*



Attack demo available online.

[https://www.youtube.com/watch?v=lvsoJdpNsZA&ab\\_channel=SPRITZResearchGroupvideos](https://www.youtube.com/watch?v=lvsoJdpNsZA&ab_channel=SPRITZResearchGroupvideos)



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

A. Compagno, M. Conti, D. Lain, G. Tsudik

**Don't Skype & Type! Acoustic Eavesdropping in Voice-over-IP.**

*In ACM SIGSAC AsiaCCS 2017*

*Presented at Black Hat USA 2017*





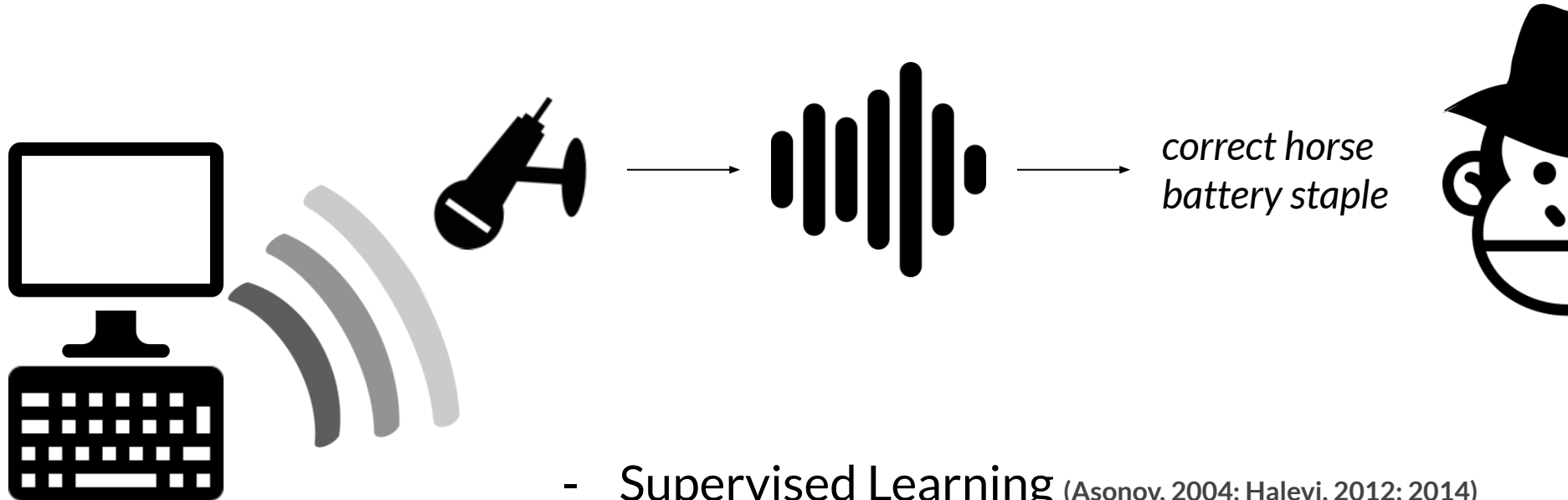
# Keyboard Acoustic Eavesdropping



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP

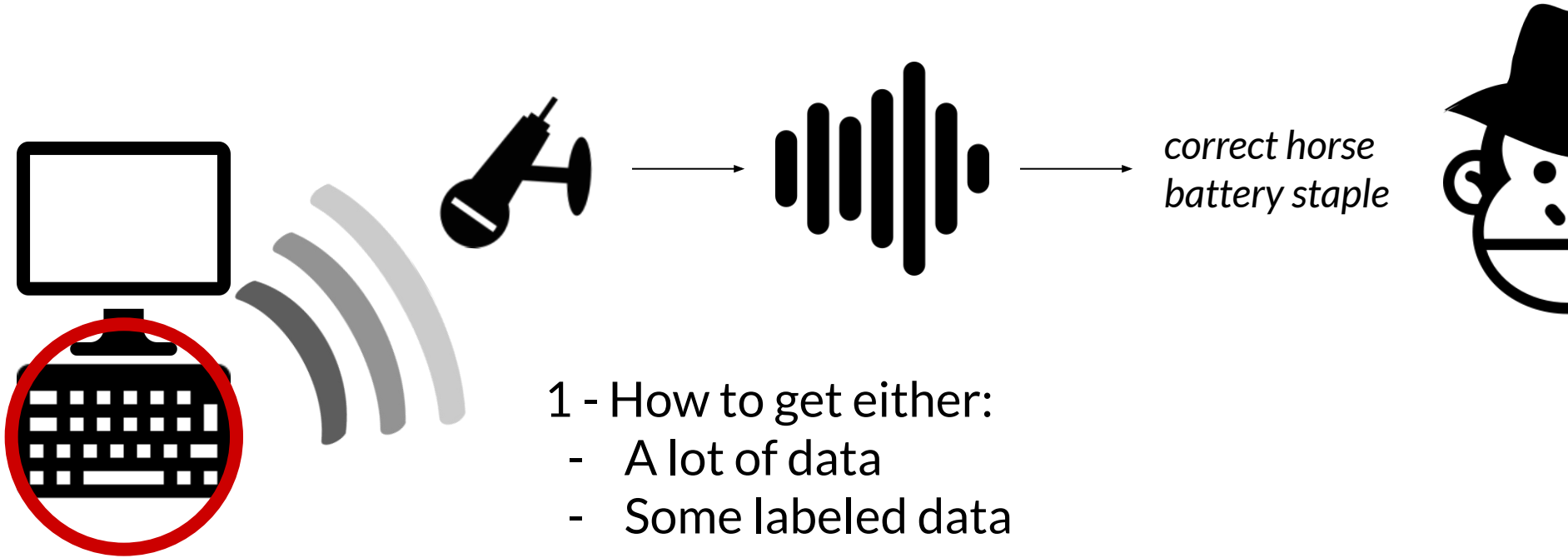


UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



- Supervised Learning (Asonov, 2004; Halevi, 2012; 2014)  
*Less input assumptions, more specific*
- Unsupervised Learning (Berger, 2006; Zhuang, 2009)  
*More input assumptions, more general*

# Keyboard Acoustic Eavesdropping



# Keyboard Acoustic Eavesdropping



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



- 1 - How to get either:
- A lot of data
  - Some labeled data

2 - How to place a compromised microphone close to my victim?

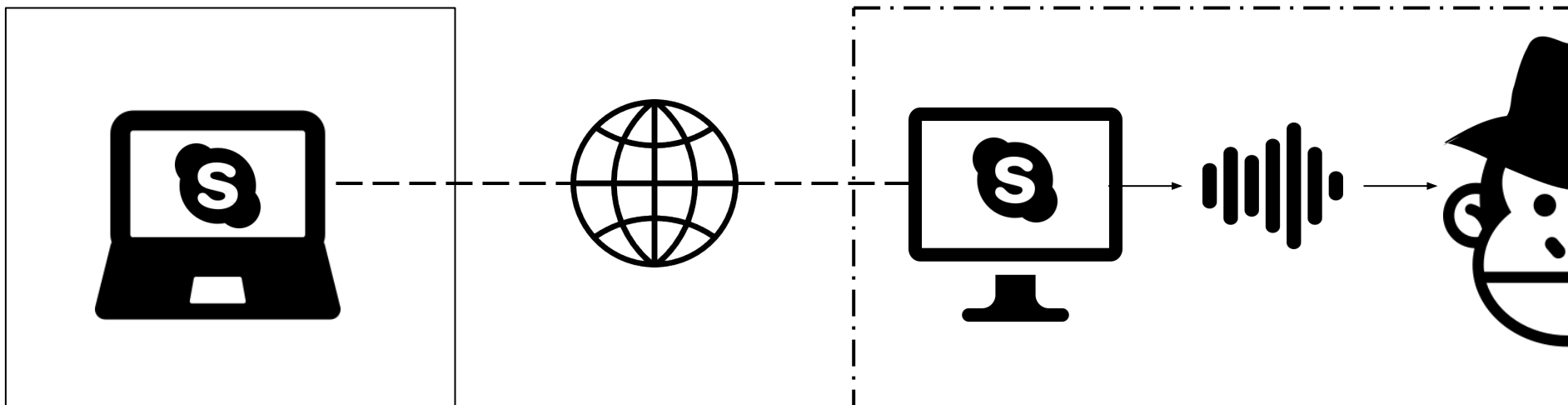
VoIP → one of the most used software: in academia, industry, at home

People type private stuff during Skype calls - it happens!

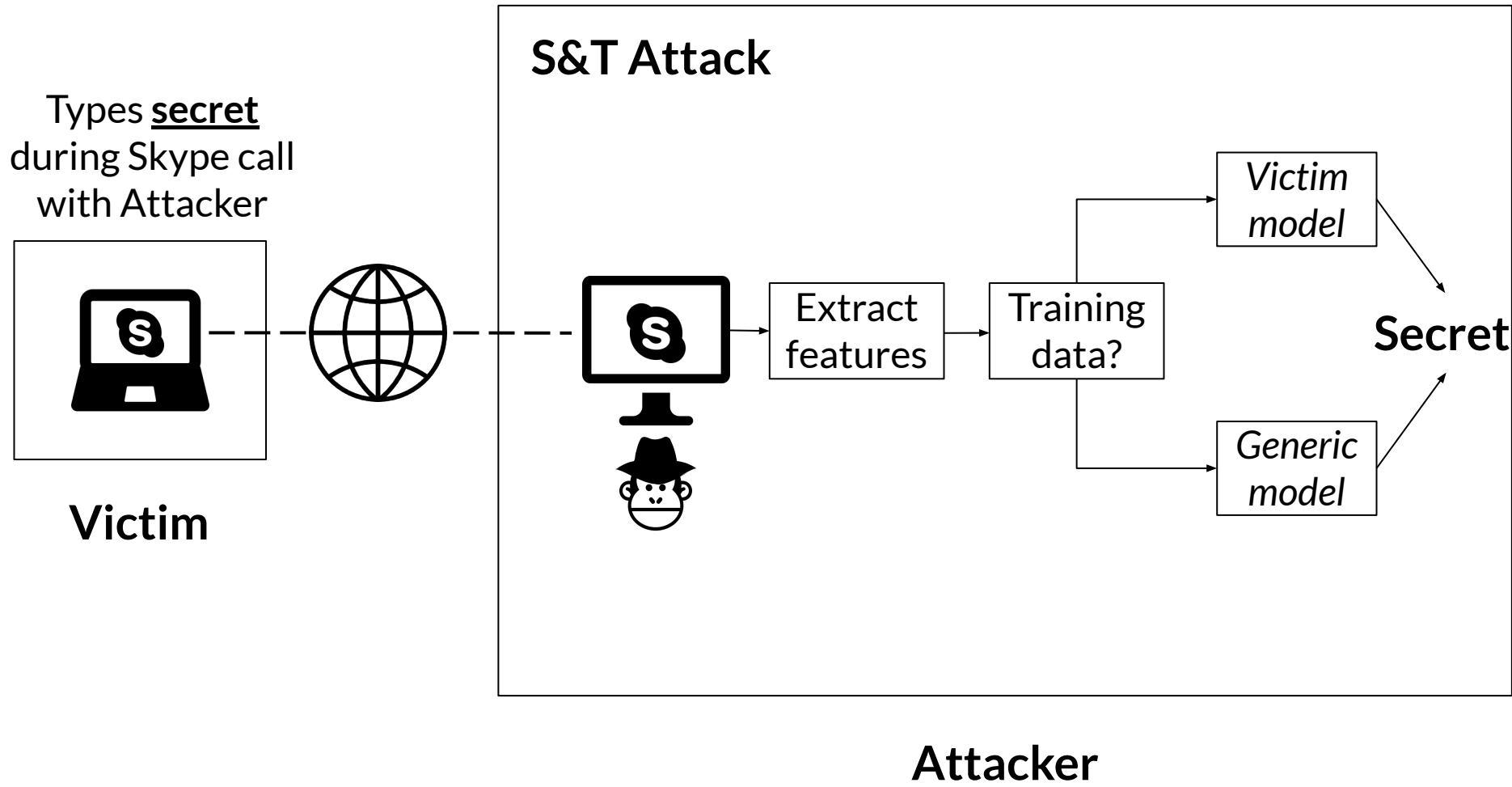
- *Login to websites*
- *Write a sensitive email*
- *Take notes*

We hear the keys' noise and use it to understand typed text

- *Victim is willingly giving us access to his microphone*



# Skype&Type Attack



- Data windowing and segmentation

*To extract sound samples*

- Mel frequency cepstral coefficients

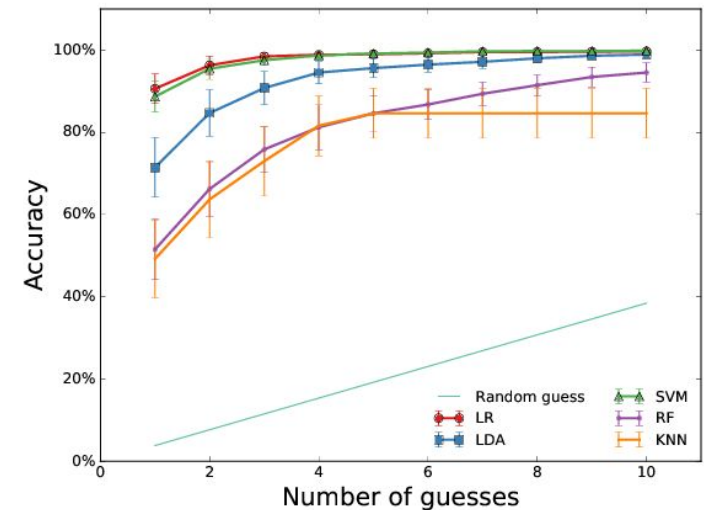
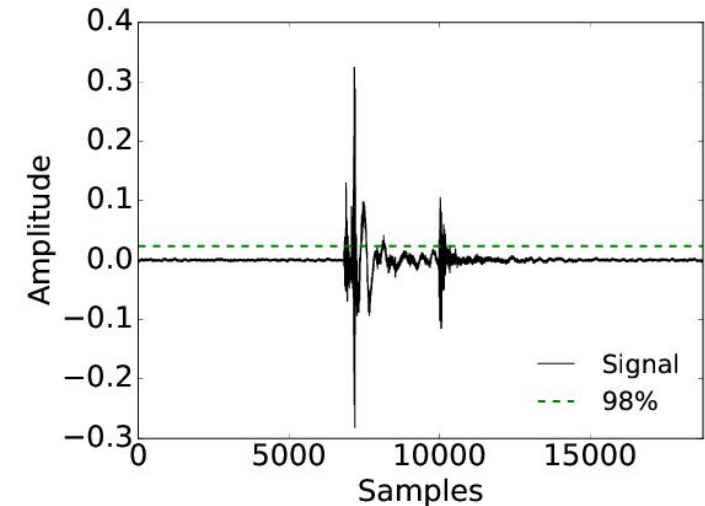
*Best performing and robust*

- Supervised learning paradigm

*Target text can be possibly:*

- Short (no clustering)
- Random (no dictionary)

- Logistic Regression classifier





- Try S&T in many scenarios
  - With 5 different users over **Skype** (Google Hangouts also vulnerable)
  - Using 3 different common laptops: Macbook Pro, Lenovo, Toshiba
  - With 2 typing styles: single finger, and natural “touch” typing
- Evaluate top-n accuracy of character recognition
  - as a function of the number of guesses, focus on top-1 and top-5 accuracy*
- Against a “dumb” random guess
  - Might be a random password -- we can not use “smarter” approaches*



## Evaluate the attack on two realistic scenarios

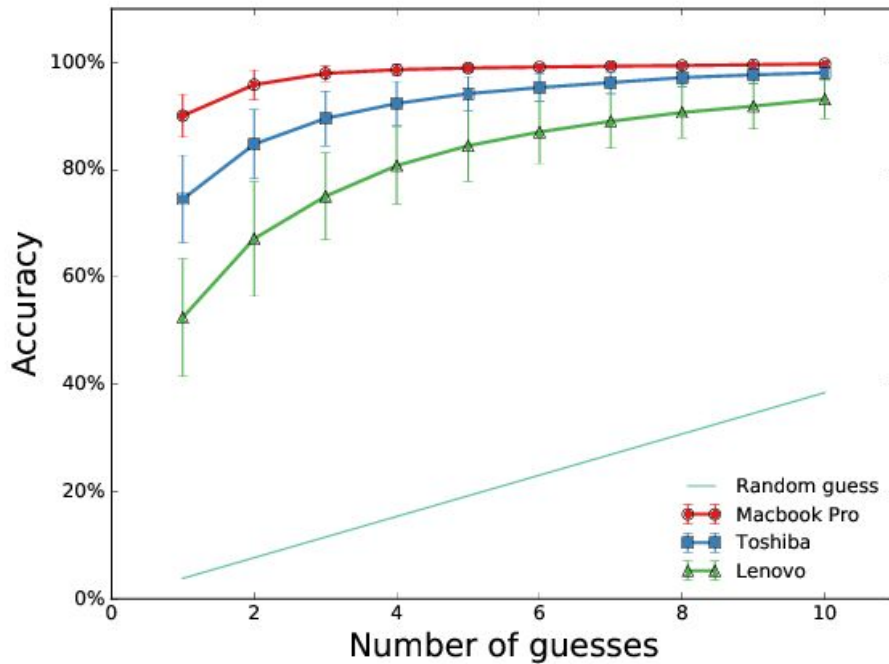
- Complete Profiling Scenario (Asonov, 2004; Halevi, 2012; 2014)
  - Profiled the user on his laptop → specific training set
  - Ground truth disclosure, e.g., a short chat message
  
- (Laptop-)Model Profiling Scenario
  - Profiled a laptop of the same model on some users
  - Victim is/can be unknown!



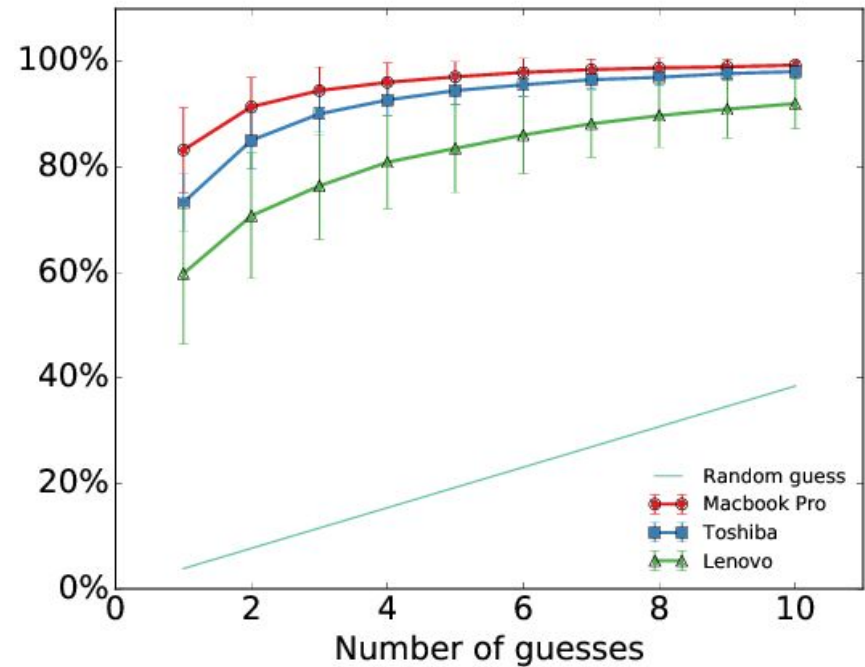
# Complete Profiling



Training set with the data the user disclosed



*Hunt&Peck typing, unfiltered data*

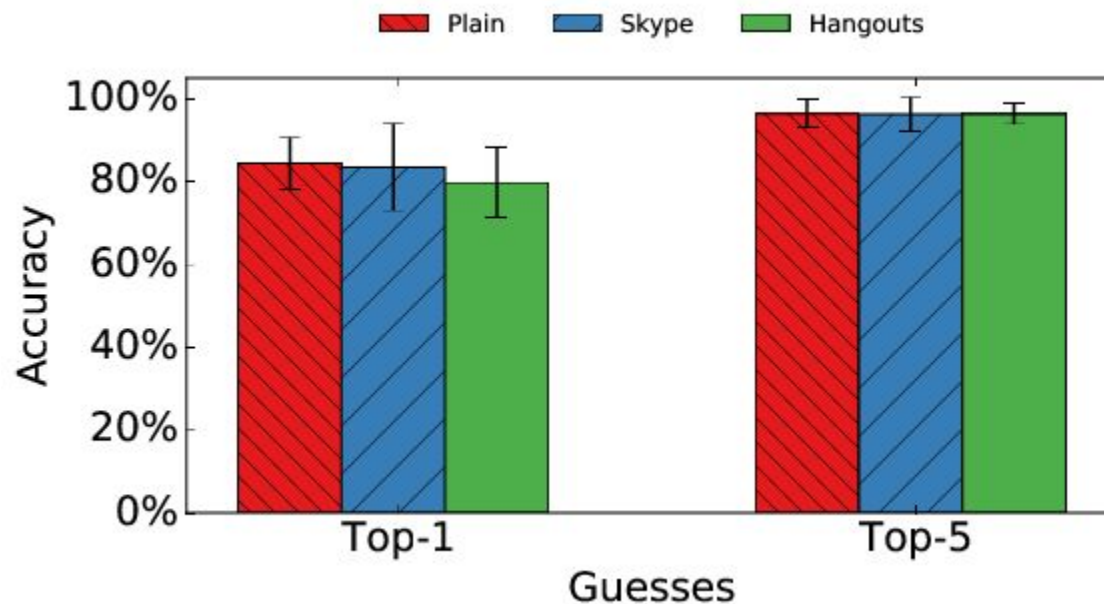


*Touch typing, Skype filtered data*

# Complete Profiling



Is only Skype vulnerable to our attack?



No! It looks like a common problem for VoIP software

On the *Model Profiling* Scenario, the victim can be unknown  
*Someone the attacker does not know personally*



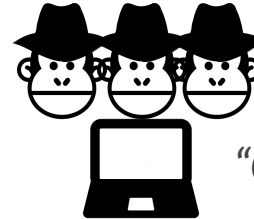
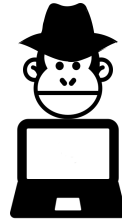
First need to understand the laptop of the victim  
→ match it with a database of model signatures

- Guess correctly **93%** of the times if the model is known
- Statistical measures if the model is unknown

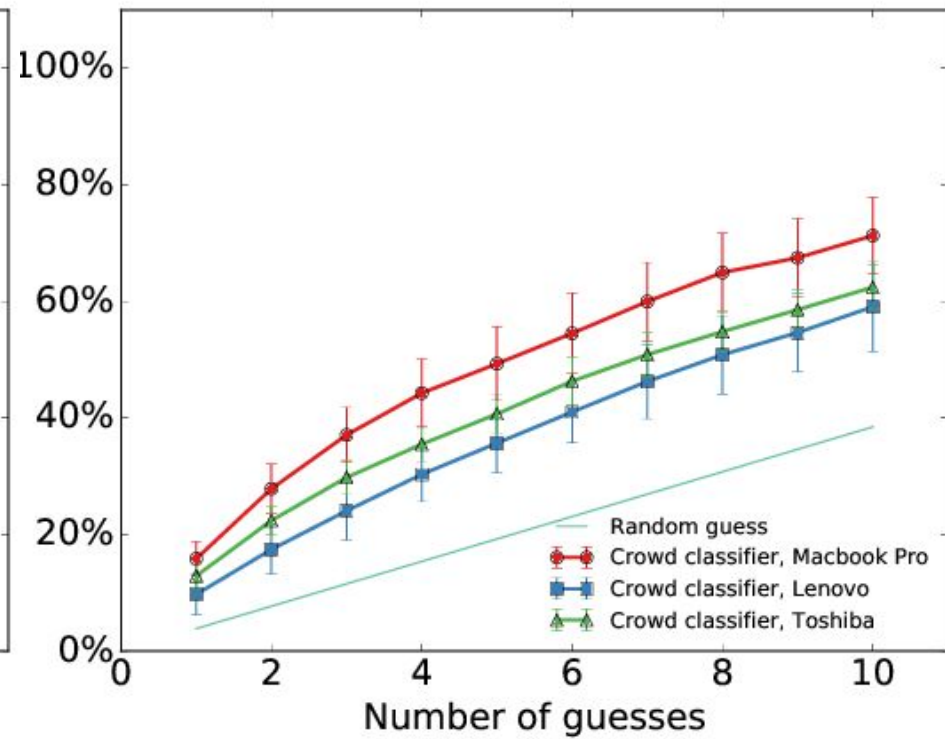
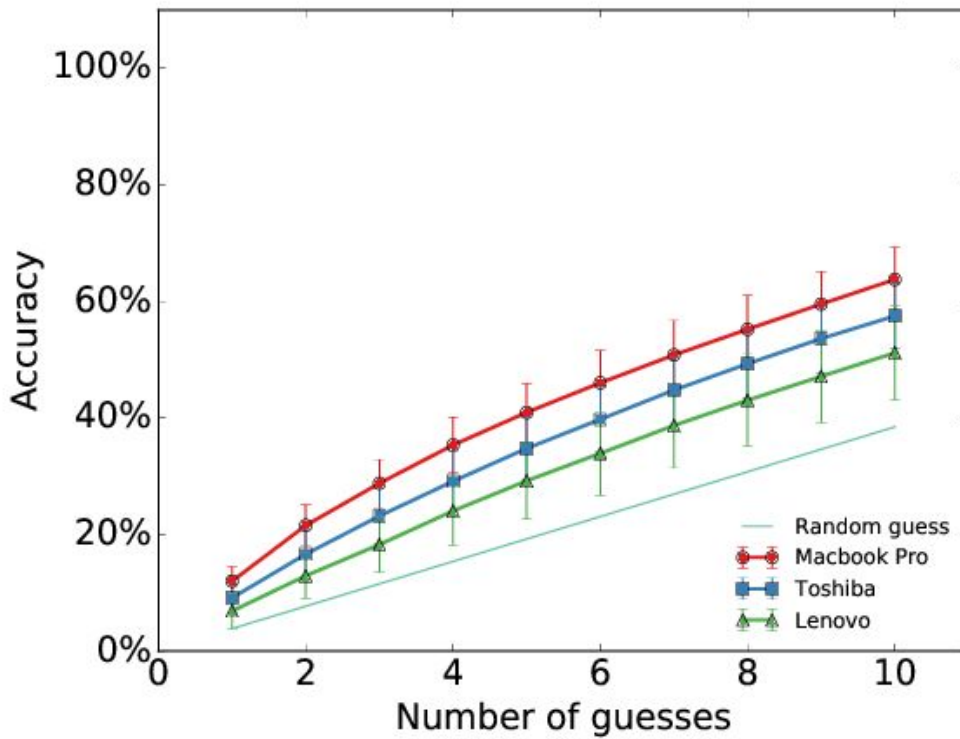
# (Laptop-)Model Profiling



One user



"Crowd" of multiple users



# Summing Up Our Results



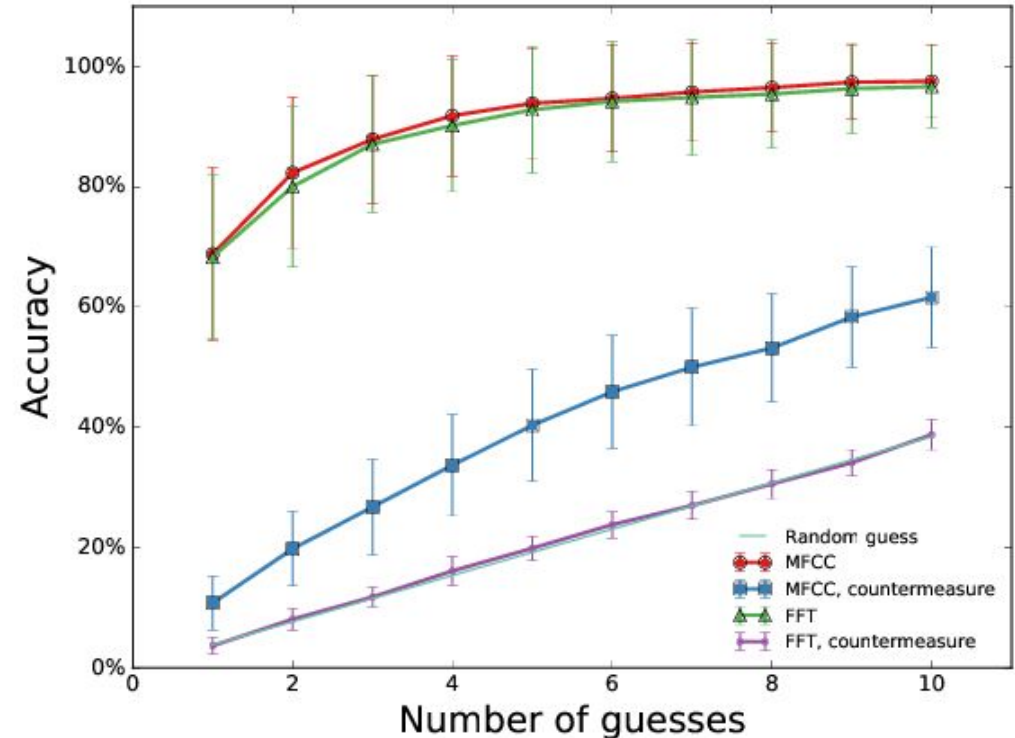
SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

- Recognize a single character
  - Complete Profiling: 90%+ accuracy
  - Model Profiling: 40%+ accuracy
- Recognize a single word
  - Complete Profiling: 98% correct letters
  - Model Profiling: 50% correct letters
- Recognize a random password
  - Improves 1-5 orders of magnitude time needed to guess the password
  - From 50 days to 42 seconds on a domestic PC

- Don't Skype & Type
- Remove volume when we detect a keypress sound
  - *Impacts voice, greatly degrades call quality*
- Disrupt spectral features with random equalization
  - *Assess impact on voice, real time feasibility*







- VoIP Keyboard acoustic eavesdropping a serious threat
- Feasible and accurate:
  - *Realistic attack scenarios*
  - **91.71% on Complete Profiling scenario**
    - *Halevi (2012; 2014): 85.78%*
  - **41.89% on Model Profiling scenario**
    - *Novel attack vs. unknown victims*
  - *Robust to degradation and to voice*
- Future work:
  - *Try more users and different keyboards, and on more VoIP software*
  - *Try to attack another user in the same room*
  - *Analyze and improve the countermeasures*

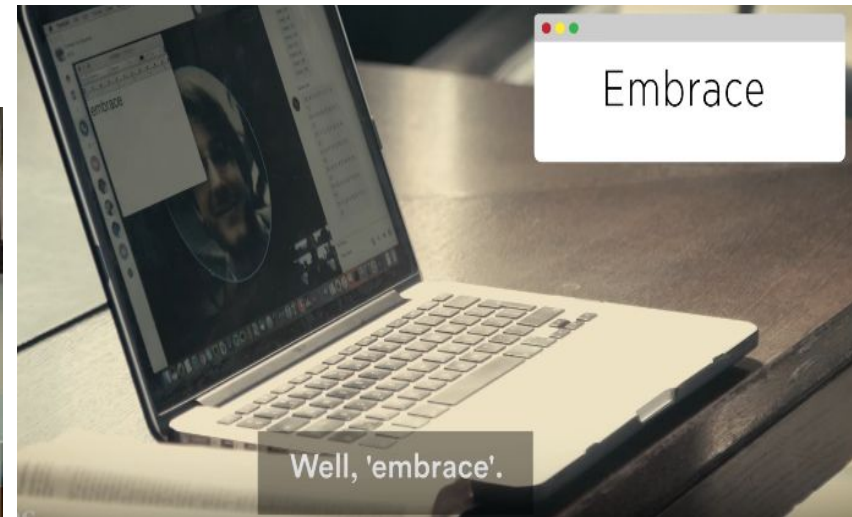
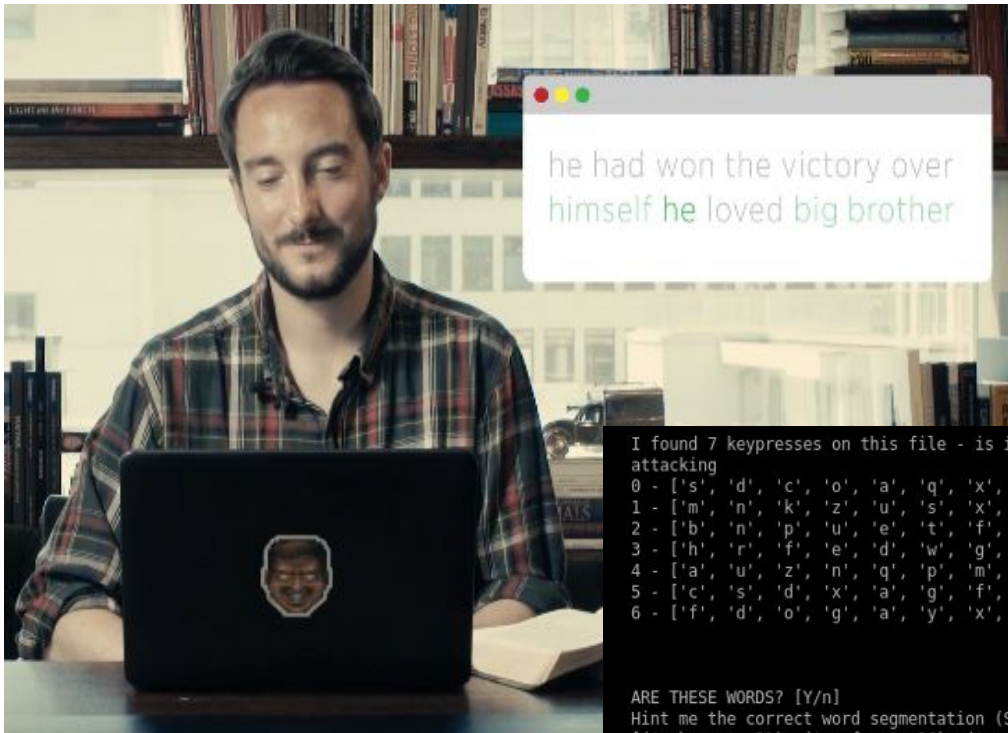


- VoIP Keyboard acoustic eavesdropping a serious threat
- Feasible and accurate:
  - *Realistic attack scenarios*
  - **91.71% on Complete Profiling scenario**
    - *Halevi (2012; 2014): 85.78%*
  - **41.89% on Model Profiling scenario**
    - *Novel attack vs. unknown victims*
  - *Robust to degradation and to voice*
- Future work:
  - *Try more users and different keyboards, and on more VoIP software*
  - *Try to attack another user in the same room*
  - *Analyze and improve the countermeasures*

# Does it really work?



## vs Forbes, 1984 & the Bible



```
I found 7 keypresses on this file - is it correct? [Y/n]
attacking
0 - ['s', 'd', 'c', 'o', 'a', 'q', 'x', 'f', 'g']
1 - ['m', 'n', 'k', 'z', 'u', 's', 'x', 'i', 'a']
2 - ['b', 'n', 'p', 'u', 'e', 't', 'f', 's', 'v']
3 - ['h', 'r', 'f', 'e', 'd', 'w', 'g', 'p', 'c']
4 - ['a', 'u', 'z', 'n', 'q', 'p', 'm', 'c', 's']
5 - ['c', 's', 'd', 'x', 'a', 'g', 'f', 'k', 'z']
6 - ['f', 'd', 'o', 'g', 'a', 'y', 'x', 'h', 'c']

ARE THESE WORDS? [Y/n]
Hint me the correct word segmentation (Suggested spaces in []):
[['embrace', 21), ('surface', 26), ('conduct', 28), ('disease', 29), ('attract', 30), ('courage', 31), ('fantasy', 32), ('contact', 33), ('intense', 33), ('library', 33), ('silence', 33), ('already', 34), ('average', 34), ('defense', 34), ('impress', 34), ('subject', 34), ('suppose', 34), ('discuss', 35), ('expense', 35), ('offense', 36), ('science', 36), ('storage', 36), ('absence', 37), ('stomach', 37), ('finance', 38), ('operate', 38), ('overall', 38), ('suspect', 38), ('century', 39), ('funding', 39)]]
```

[https://drive.google.com/file/d/1uLDDI\\_eESwOm6pQs59I4NeOjXwiOq1M/view?usp=sharing](https://drive.google.com/file/d/1uLDDI_eESwOm6pQs59I4NeOjXwiOq1M/view?usp=sharing)



Credits: <https://www.forbes.com/sites/thomasbrewster/2017/07/06/skype-and-type-attack-steals-passwords>

# Thank you!

# Questions?

(if you do not have one, please find some suggestions below)

**Security Questions**  
Select a security question or create one of your own. This question will help us verify your identity should you forget your password.

Security Question

Answer

Security Question

Answer

Security Question

Answer

Security Question

Answer

Security Question

Answer

This is the END!

Backup Slides  
after this point... ;-)

File Edit View Go Message Tools Help

Get Messages Write Tag

Reply Reply All Forward Archive Junk Delete More

From Sadeghi, Ahmad-Reza <ahmad.sadeghi@trust.informatik.tu-darmstadt.de>

To manuel@atug.de <manuel@atug.de>, Lejla Batina <lejla@cs.ru.nl> MORE

17/09/23, 23:50

Cc Kleffel, Petra <petra.kleffel@tu-darmstadt.de>

Subject **Your talks arrangement**

Dear Speakers,

We assumed that most of you want to use your own laptops during your talk. Please get ready short before your talk and prepare possible adaptors so that we do not loose much time when switching laptops.

In case you would use our laptop, we need an USB stick with your slides on it.

Best  
Ahmad

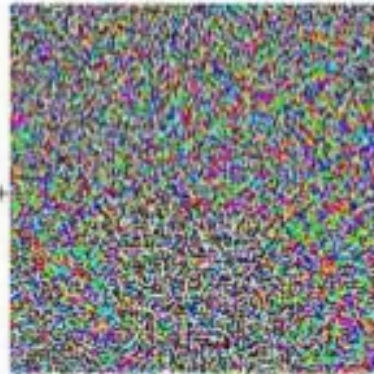
DELL



## Adversarial Examples (Deep Learning/CNNs)



Original image classified as a panda with 60% confidence.



Tiny adversarial perturbation.



Imperceptibly modified image, classified as a gibbon with 99% confidence.

<http://www.kdnuggets.com/2015/07/deep-learning-adversarial-examples-misconceptions.html>

<http://karpathy.github.io/2015/03/30/breaking-convnets/>





## Classification Accuracy

Classification Accuracy is what we usually mean, when we use the term accuracy. It is the ratio of number of correct predictions to the total number of input samples.

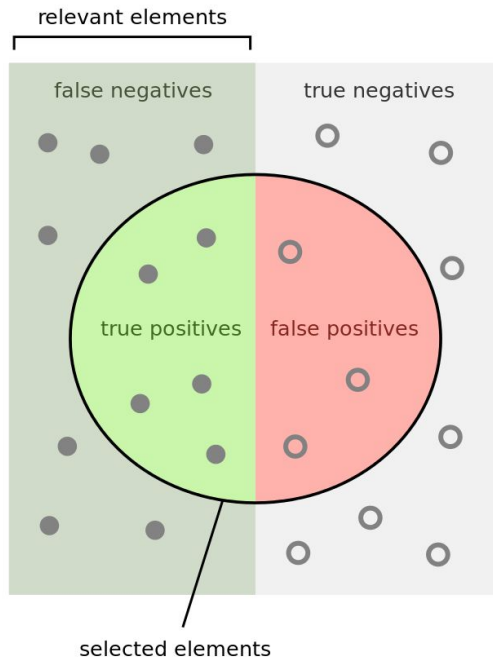
$$\textit{Accuracy} = \frac{\textit{Number of Correct predictions}}{\textit{Total number of predictions made}}$$

It works well only if there are equal number of samples belonging to each class.

For example, consider that there are 98% samples of class A and 2% samples of class B in our training set. Then our model can easily get **98% training accuracy** by simply predicting every training sample belonging to class A.

When the same model is tested on a test set with 60% samples of class A and 40% samples of class B, then the **test accuracy would drop down to 60%**. Classification Accuracy is great, but gives us the false sense of achieving high accuracy.

## - Precision, Recall, and F-measure



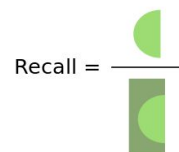
$$F_1 = 2 \cdot \frac{1}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

$$F_\beta = \frac{(1 + \beta^2) \cdot \text{true positive}}{(1 + \beta^2) \cdot \text{true positive} + \beta^2 \cdot \text{false negative} + \text{false positive}}$$

How many selected items are relevant?



How many relevant items are selected?



# Attack - Data Processing

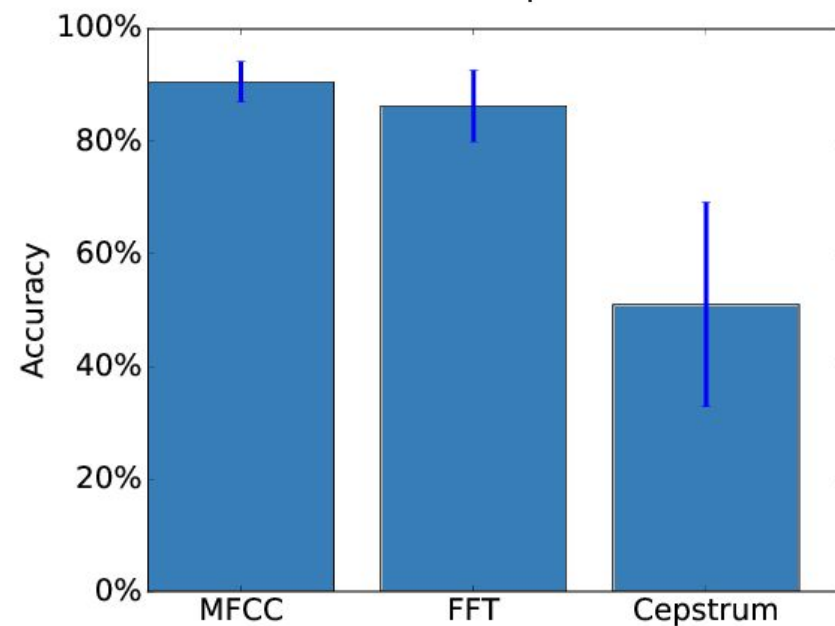
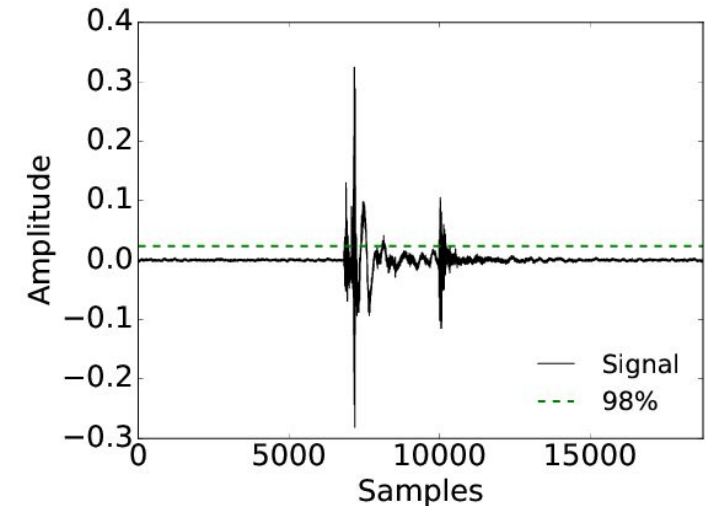


- Data windowing and segmentation

*To extract sound samples*

- Feature extraction: *mel frequency cepstral coefficients*

*Selected with a preliminary experiment*



## Evaluate the attack on three different realistic scenarios

- **Complete Profiling Scenario** (Asonov, 2004; Halevi, 2012; 2014)
  - *Profiled the user on his laptop → specific training set*
  - *Ground truth disclosure, e.g., a short chat message*
- **User Profiling Scenario**
  - *Profiled the user on a different laptop*
  - *Social engineering techniques*
- **Model Profiling Scenario**
  - *Profiled a laptop of the same model on some users*
  - *The victim can be unknown*



# Evaluation - Small Training Set

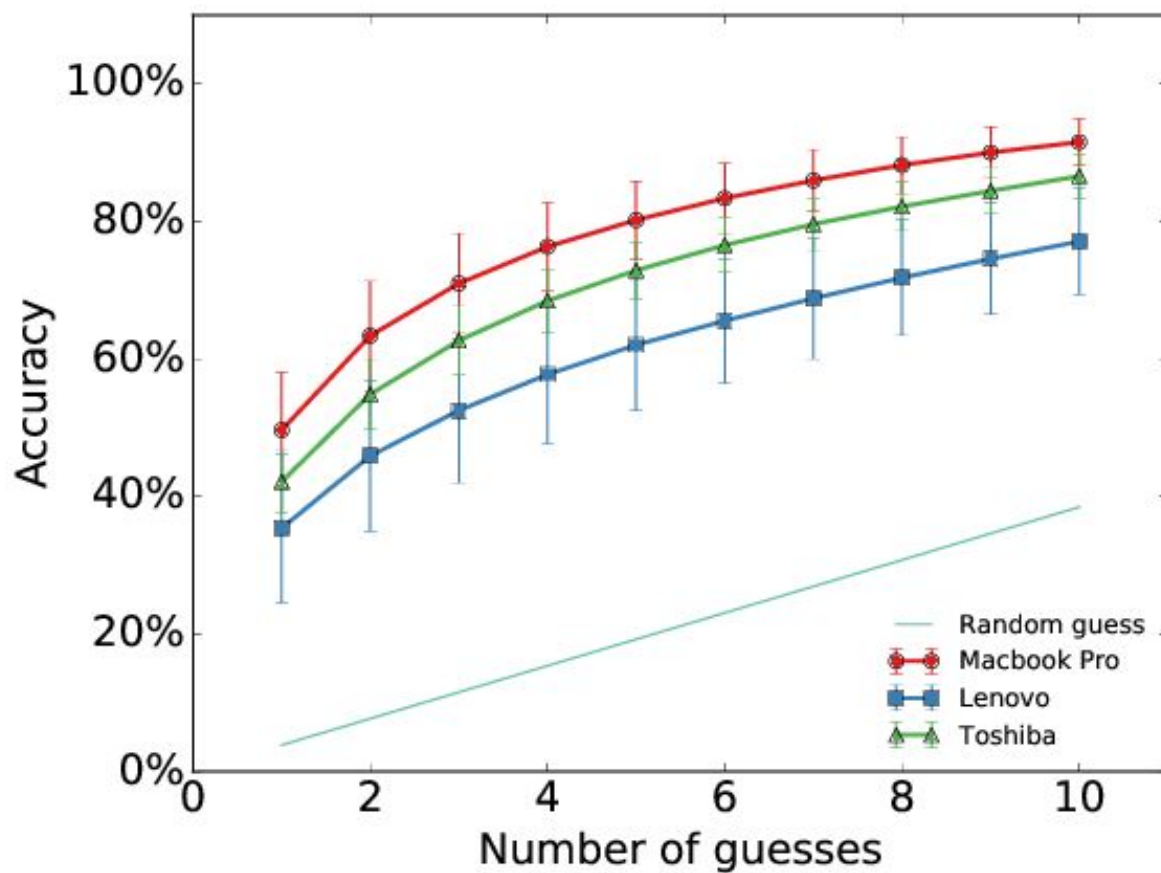


## 10 samples/character aren't your typical chat message

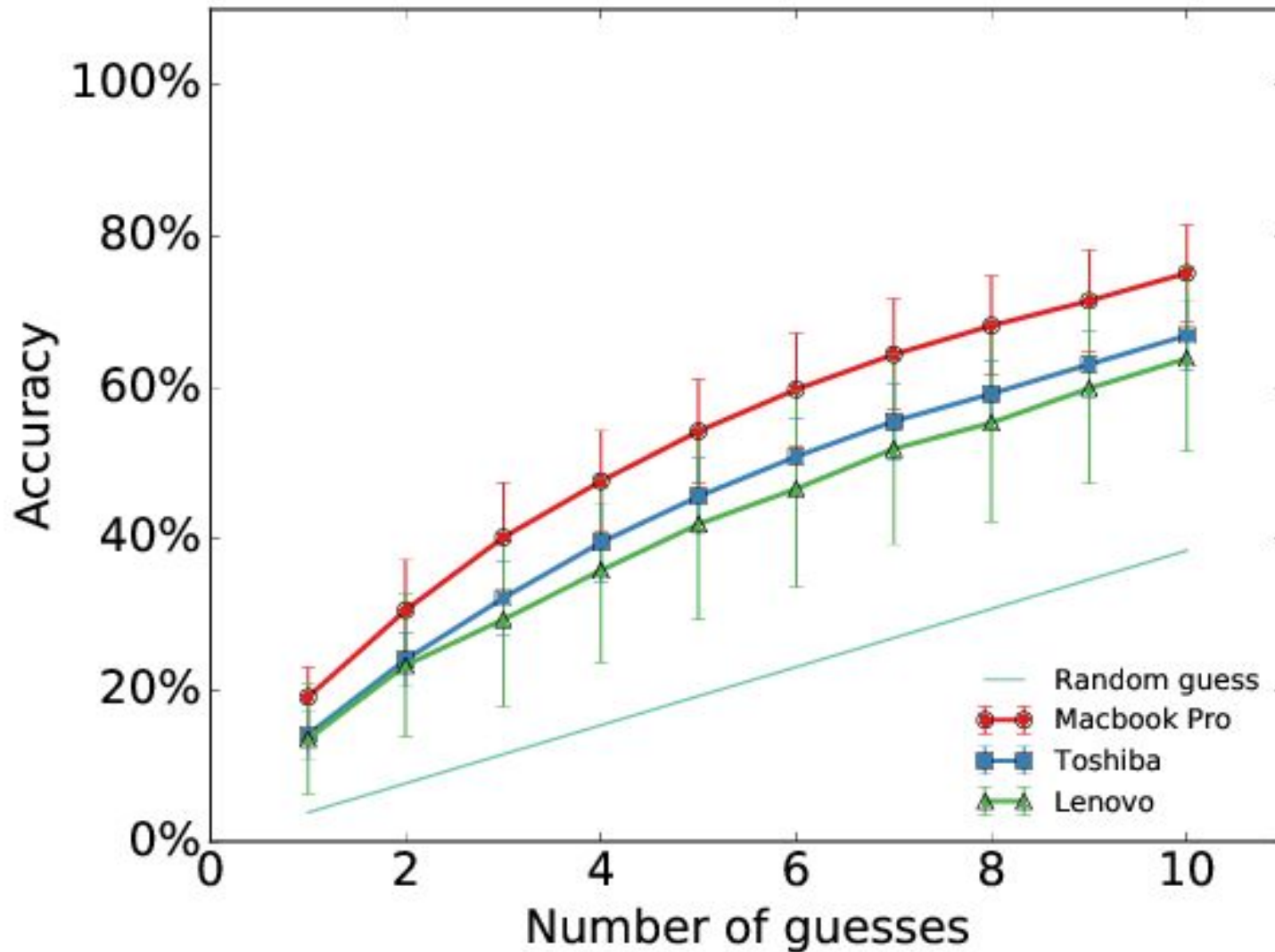
*Training set with realistic letter frequencies*  
*Test against random password*



Character	# Samples
E	10
A	9
R	7
J	1
Z	1



# Evaluation - User Profiling



The goal was to crack the victim's random password

→ We need bruteforce techniques

Random password of 10 lowercase letters

- $\log_2(26^{10}) = 47$  bits of entropy

On the Complete Profiling Scenario (high accuracy)

- $\log_2(5^{10}) = 23.22$  bits of entropy

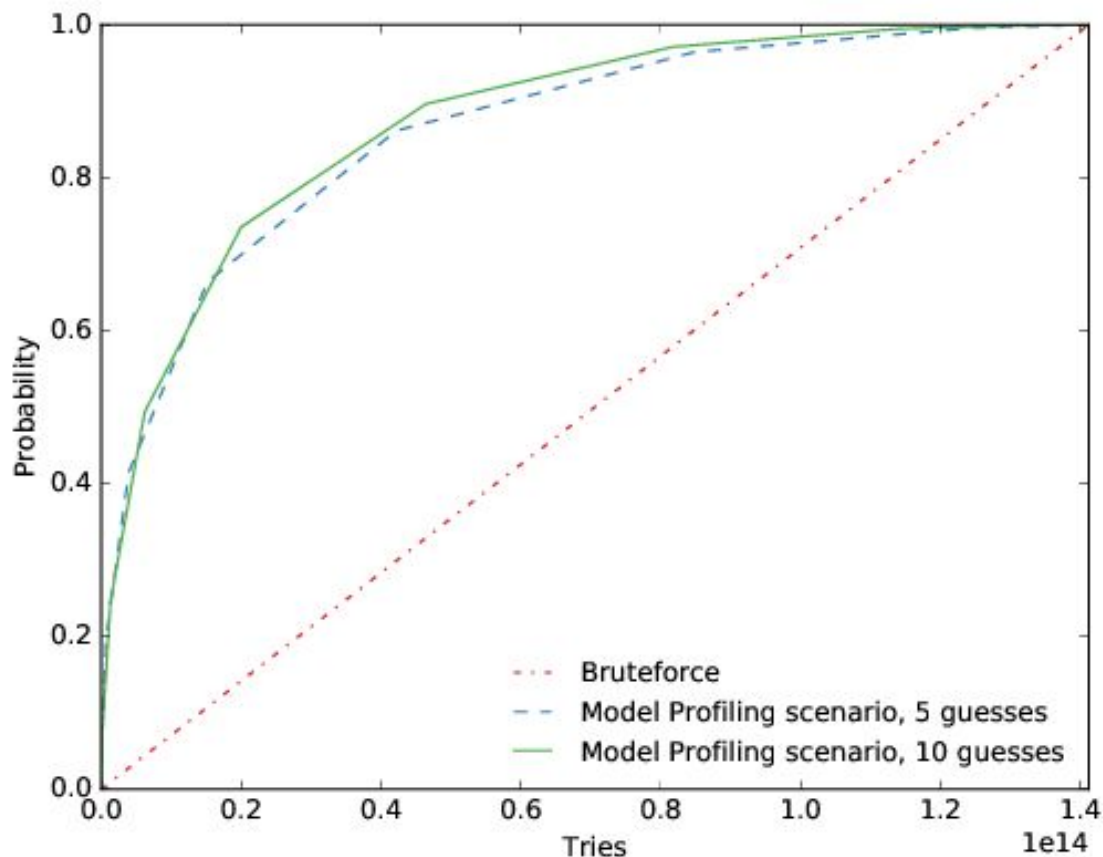
On the other scenarios - entropy is not meaningful



## Model Profiling Scenario → improved bruteforce

*Take into account character probabilities*

Evaluate the reduction of the average number of trials



## *Fast Fourier Transform coefficients*

$$S(f(t)) = 20 \log_{10} (|\mathcal{F}(f(t))|)$$

$f(t)$  = signal

$\mathcal{F}$  = Discrete Fourier Transform function

## *Cepstrum coefficients*

$$C(f(t)) = |\mathcal{F}^{-1}(S(f(t)))|^2$$

## *Mel frequency cepstral coefficients*

$$MFC(f(t)) = DCT(\log_{10}(\text{mel}\{|\mathcal{F}(f(t))|\}))$$

$$\text{mel}(f) = 2595 \log_{10} \left( 1 + \frac{f}{700} \right)$$

$DCT$  = Discrete Cosine Transform

# Side and Covert Channels: *the Dr. Jekyll and Mr Hyde of Modern Technologies*

Mauro Conti

**2020 WiseML @ WiSec**

July 13 2020



SPRITZ  
SECURITY & PRIVACY  
RESEARCH GROUP



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA