Università degli Studi di Padova

New York Institute of Technology

SPRITZ Security & Privacy Research Group

GFT

Your PIN Sounds Good! Augmentation of PIN Guessing Strategies Through Audio Leakage

Matteo Cardaioli

Topics:

Authentication, Side-Channel, Behavioral

About Me

Master's Degree in Bioengineering

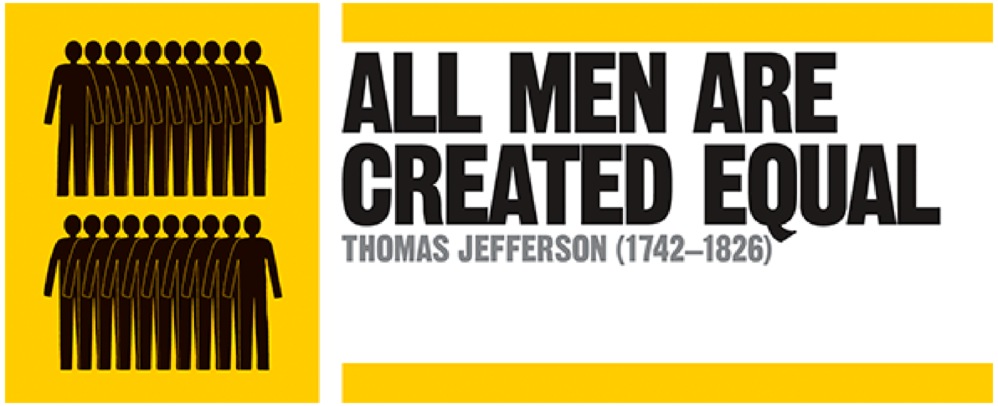
Ph.D. student in Brain, Mind and Computer Science at University of Padova (Italy)

SPRITZ (Security and Privacy Research Group) member

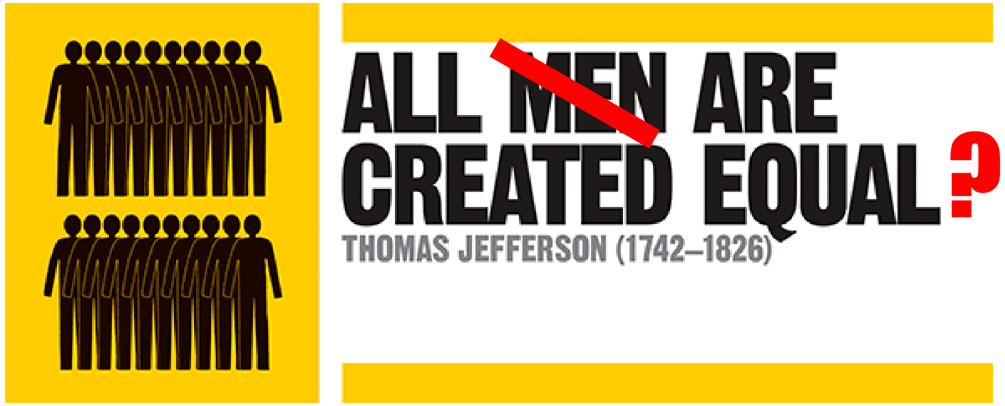
Project Developer @GFT Italy

Research Activities

My research activity focuses on the study and the development of new methods for analysing and ***detecting patterns of suspicious behaviors****,* ***authentication*** *procedures to detect and mitigate identity fraud and on the implementation of the so called "****predictive security****" of critical financial infrastructures.*


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

PIN



Your PIN Sounds Good!

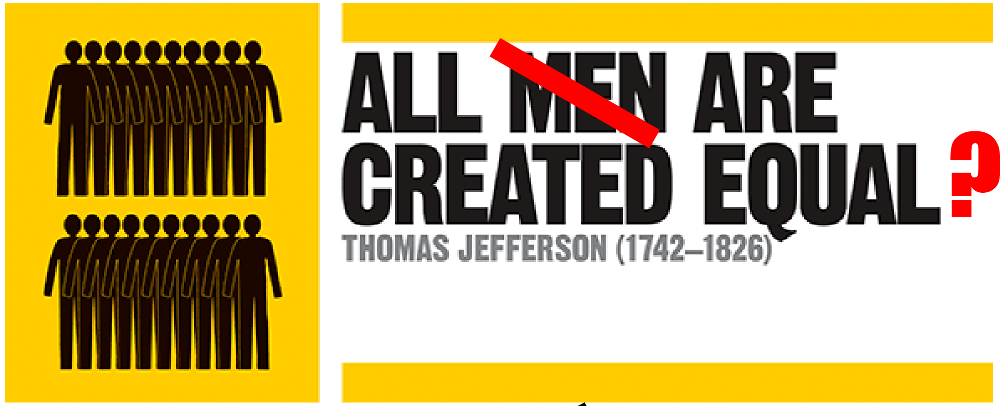
Università degli Studi di Padova

New York Institute of Technology

SPRITZ Security & Privacy Research Group

GFT

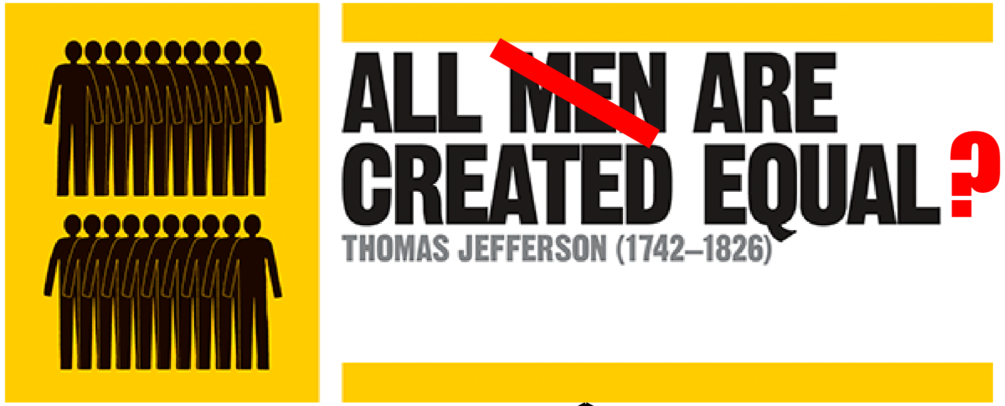
PIN



User chosen

Università degli Studi di Padova — New York Institute of Technology — SPRITZ Security & Privacy Research Group — GFT

PIN

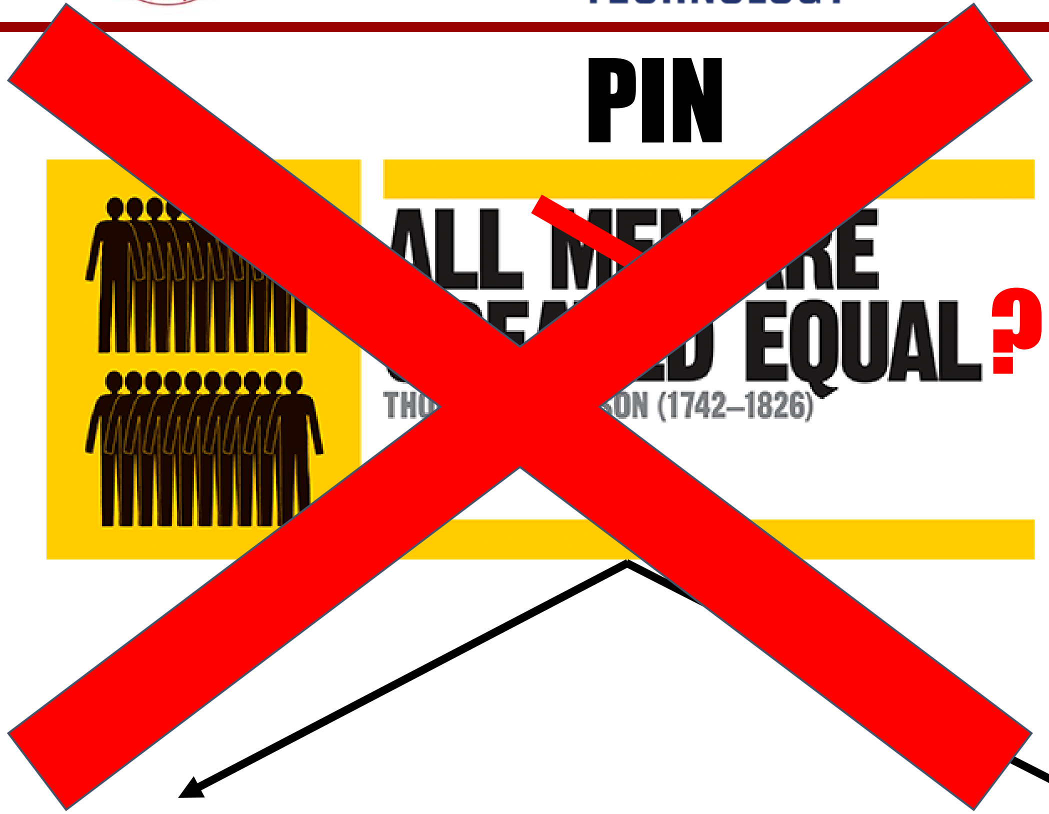


User chosen

Randomly assigned

Your PIN Sounds Good!

PIN



User chosen

Randomly assigned

DEFINITELY… NOT!

User chosen passwords

In December 2017 4iQ discovered, scanning the dark web, a single file containing a database of **1.4 billion credentials in clear**.

None of the database passwords are encrypted and **most of them have been verified as true**.

This collection contains 252 already known breaches (like Linkedin) and other new ones like Netflix, Last.FM, Zoosk or YouPorn.

https://4iq.com/wp-content/uploads/2018/05/2018_IdentityBreachReport_4iQ.pdf

UNIVERSITÀ DEGLI STUDI DI PADOVA

NEW YORK INSTITUTE OF TECHNOLOGY

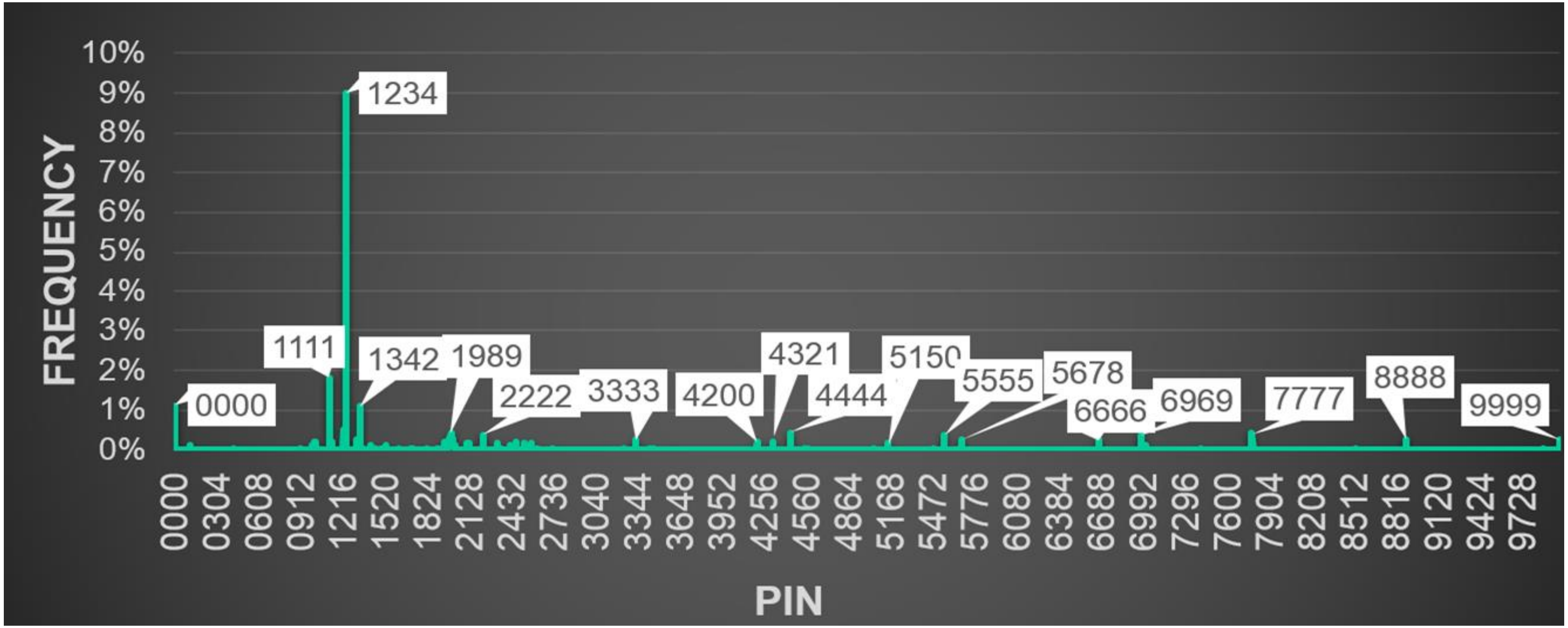
SPRITZ SECURITY & PRIVACY RESEARCH GROUP

GFT

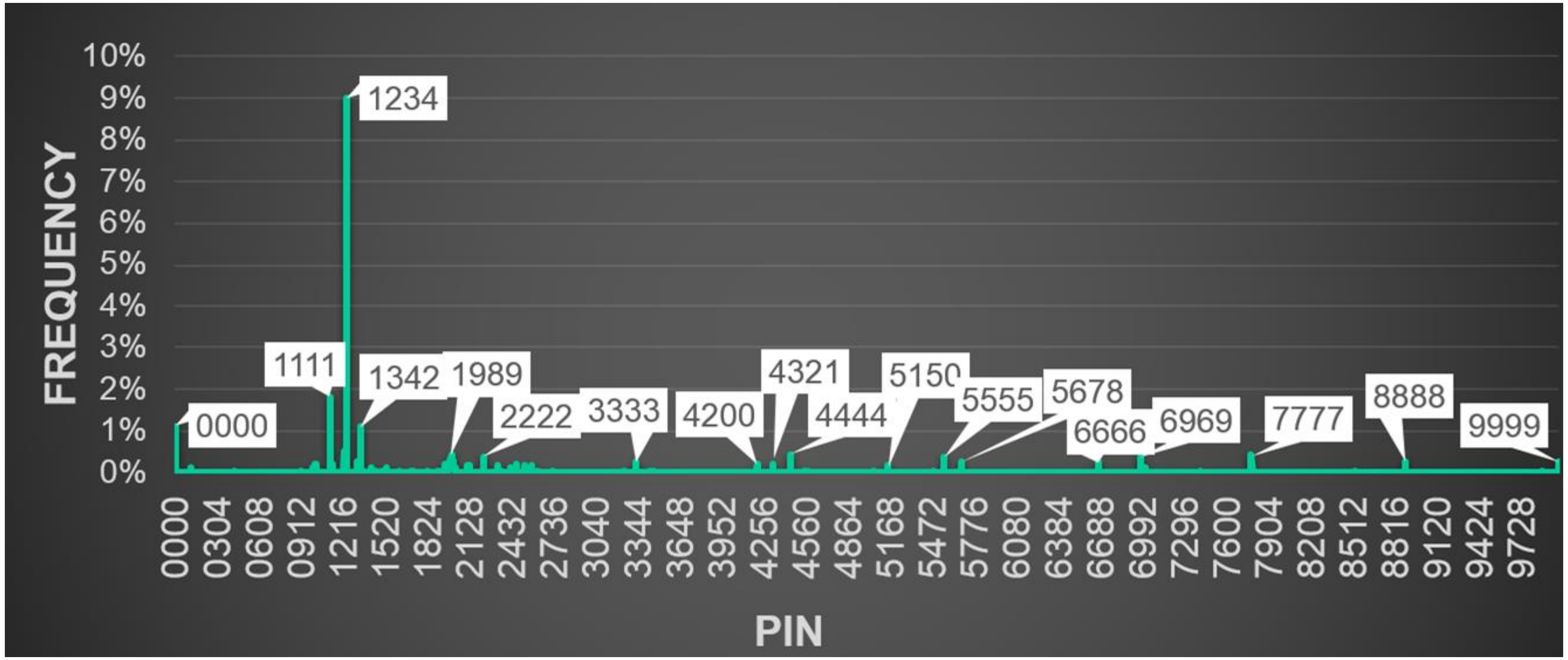
What about PINs?

UNIVERSITÀ DEGLI STUDI DI PADOVA
NEW YORK INSTITUTE OF TECHNOLOGY
SPRITZ SECURITY & PRIVACY RESEARCH GROUP
GFT

User chosen PINs



User chosen PINs



WORST 10 PINs

PIN	FREQ
1234	9.00%
1111	1.83%
0000	1.13%
1342	1.10%
1212	0.50%
4444	0.43%
1989	0.43%
1986	0.42%
7777	0.41%
2222	0.37%

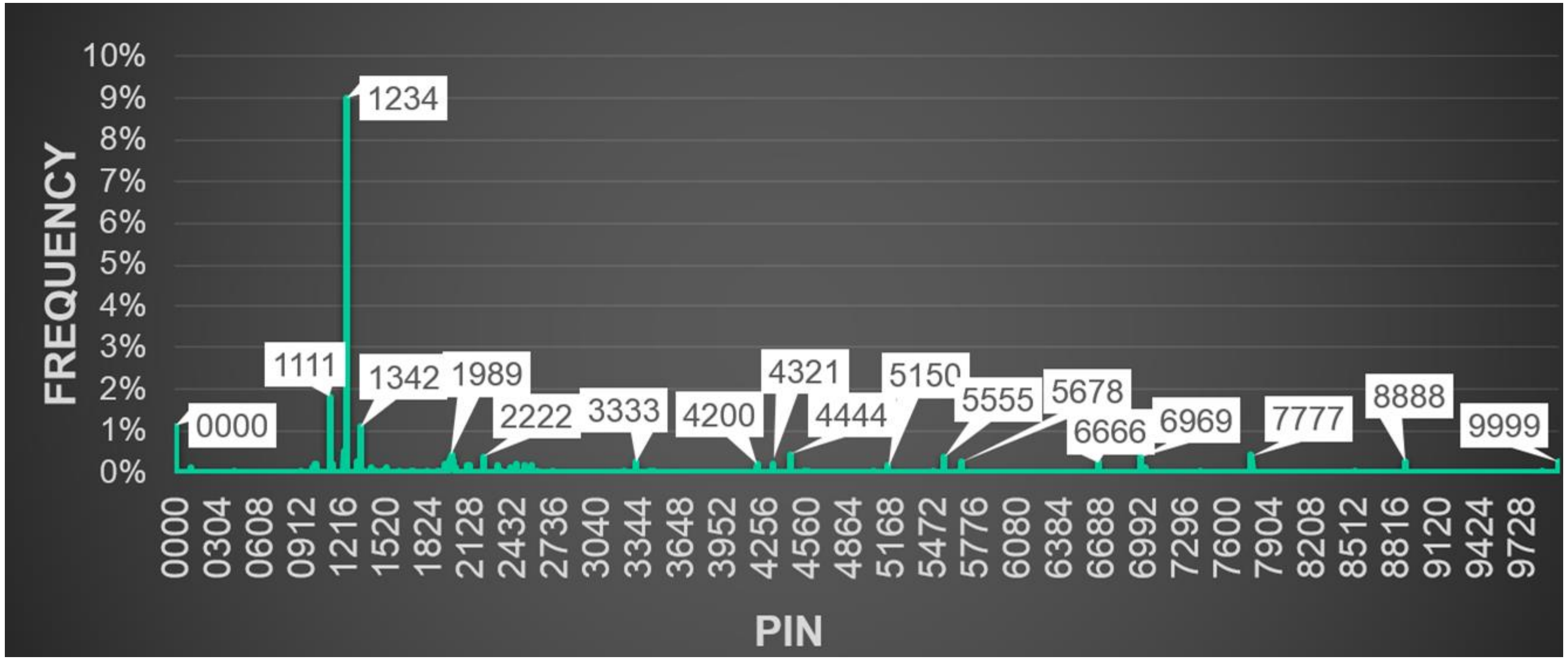
UNIVERSITÀ DEGLI STUDI DI PADOVA

NEW YORK INSTITUTE OF TECHNOLOGY

SPRITZ Security & Privacy Research Group

GFT

User chosen PINs



BEST 10 PINs

PIN	FREQ
0835	0.0014%
0849	0.0014%
0739	0.0014%
0639	0.0014%
0736	0.0013%
0938	0.0013%
0837	0.0013%
0939	0.0013%
0738	0.0012%
0839	0.0009%

User chosen PINs

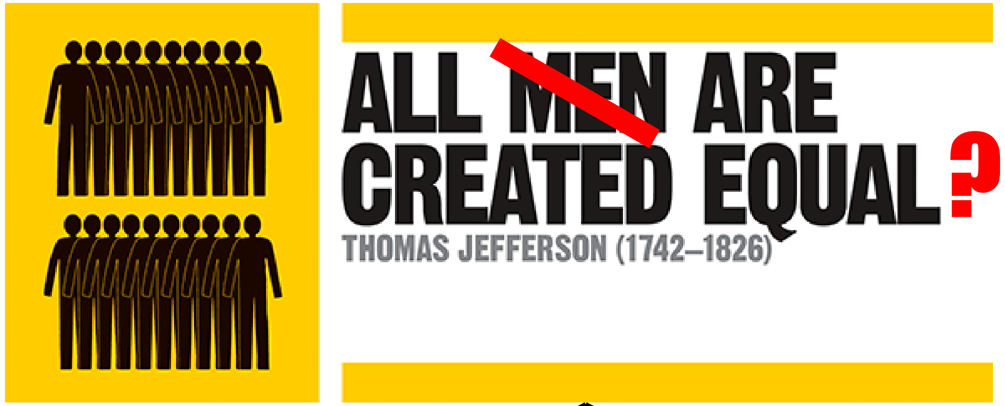
In this scenario, **20% of PINs** can be guessed by **trying the 20 most common combinations** chosen by the user

If these 20 4-digit PINs were distributed **uniformly and randomly**, they would represent only **0.2%** of the total

WORST 10 PINs

PIN	FREQ
1234	9.00%
1111	1.83%
0000	1.13%
1342	1.10%
1212	0.50%
4444	0.43%
1989	0.43%
1986	0.42%
7777	0.41%
2222	0.37%

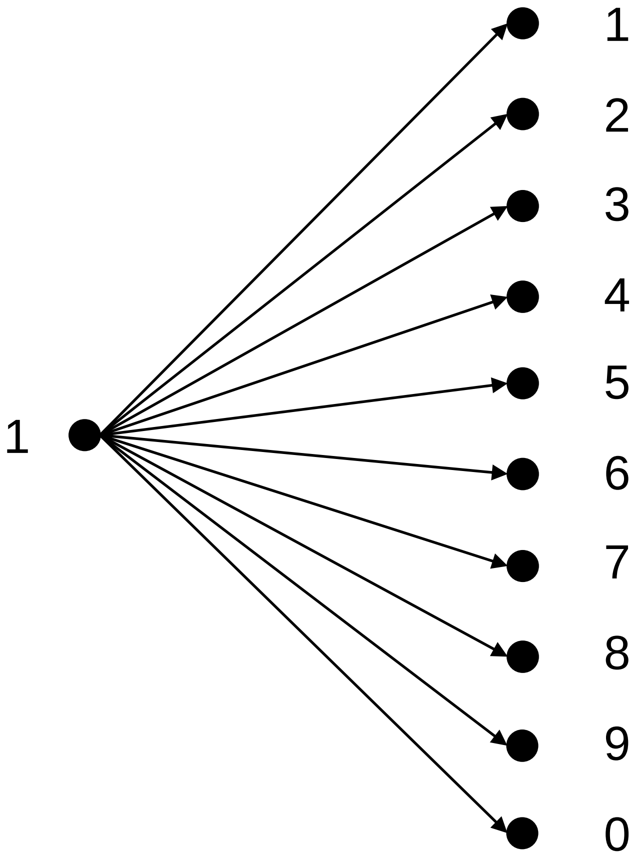
PIN

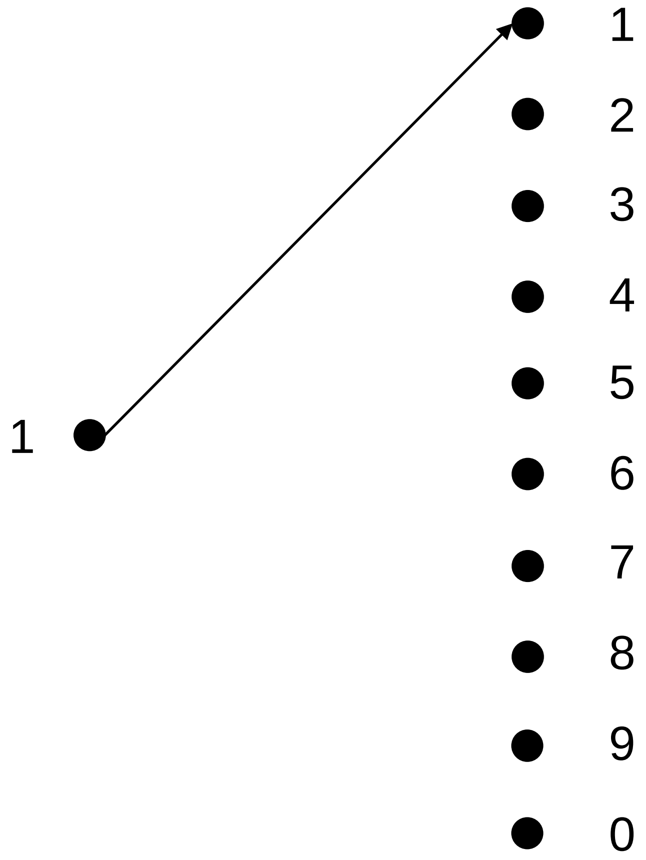


Randomly assigned

Università degli Studi di Padova
New York Institute of Technology
SPRITZ Security & Privacy Research Group
GFT

Your PIN Sounds Good!





Distance 0

11**

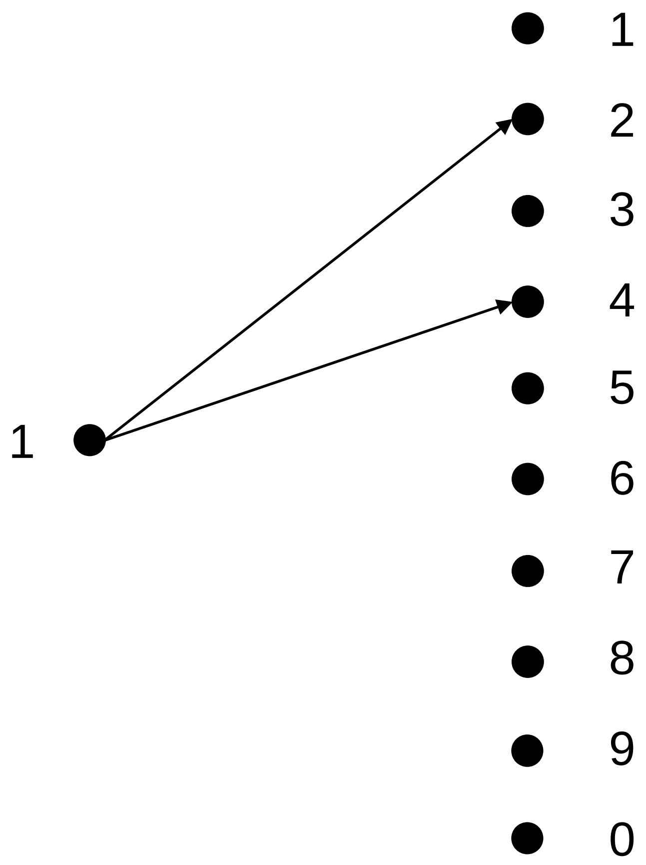
22**

55**

00**

Your PIN Sounds Good!

Università degli Studi di Padova
NEW YORK INSTITUTE OF TECHNOLOGY
SPRITZ Security & Privacy Research Group
GFT



Distance 1

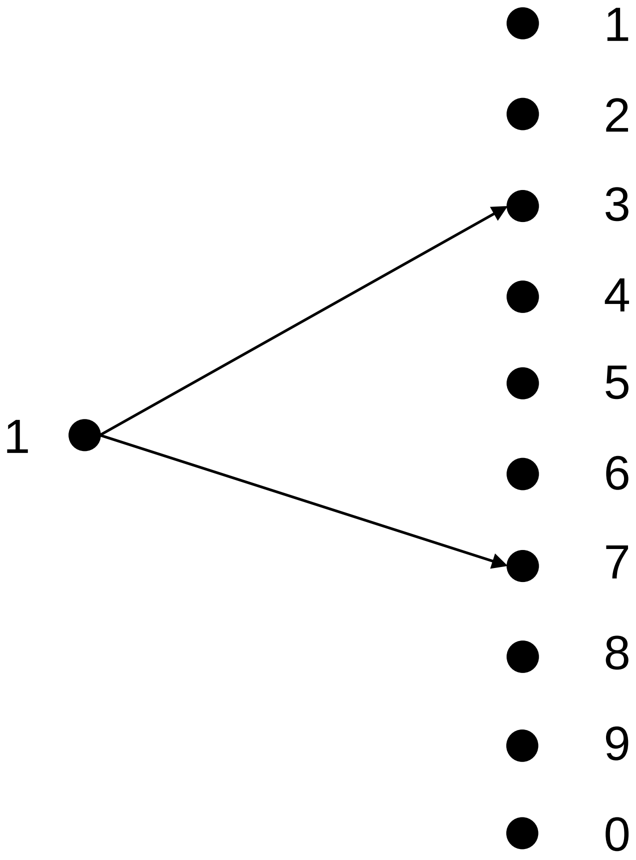
12**

14**

45**

69**

Your PIN Sounds Good!



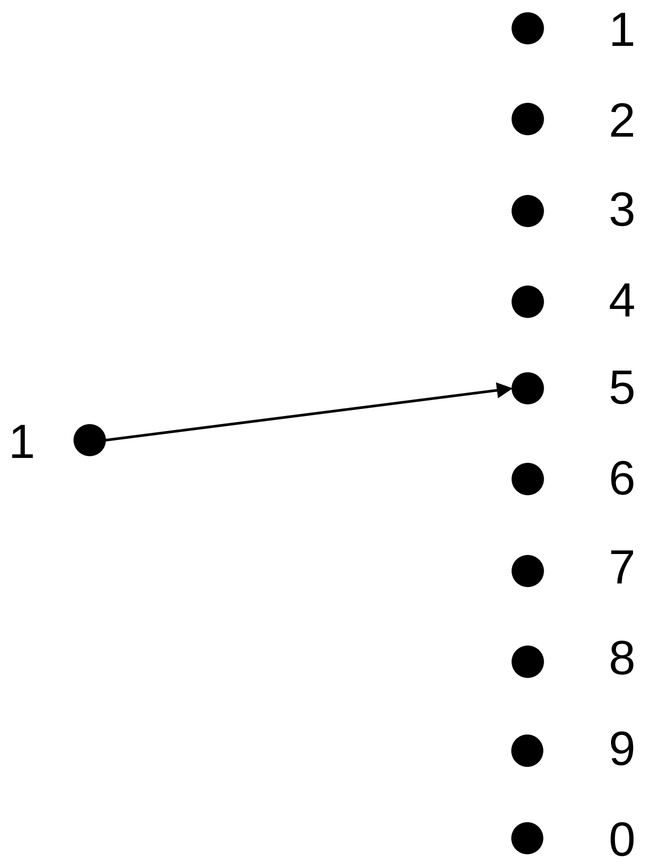
Distance 2

13**

17**

28**

79**



Distance D1

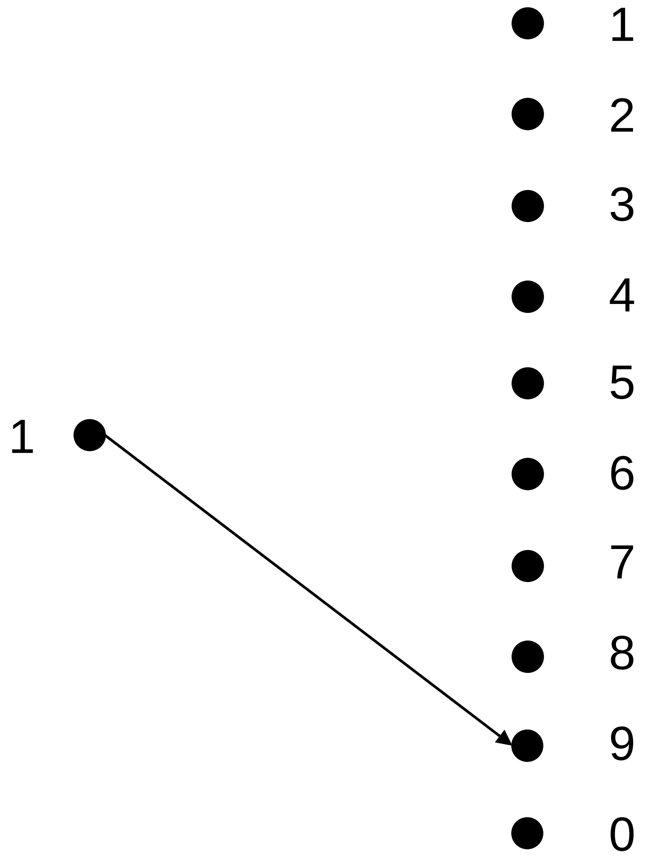
15**

57**

70**

26**

Your PIN Sounds Good!

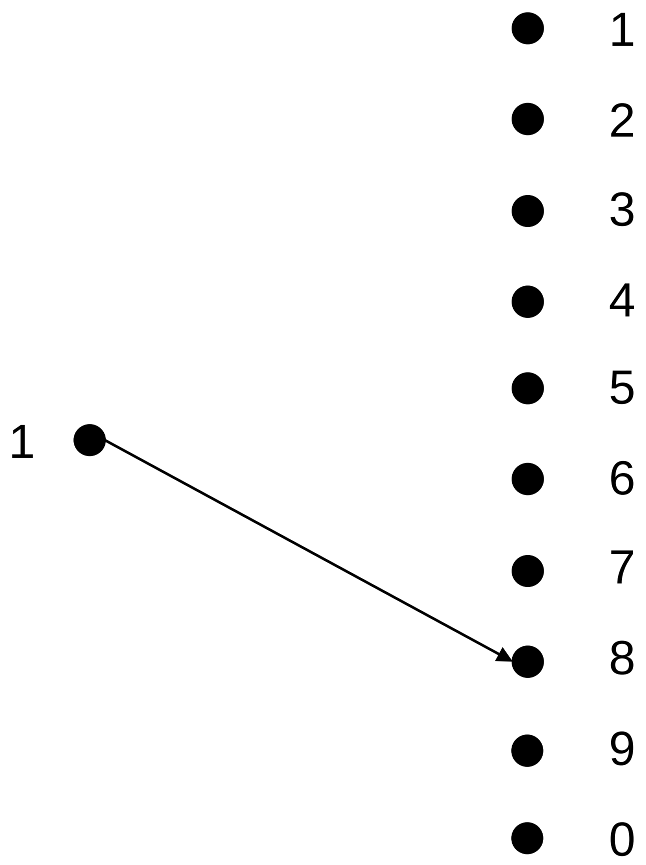


Distance D2

19**

73**

Your PIN Sounds Good!



Distance DS

18**

29**

40**

06**

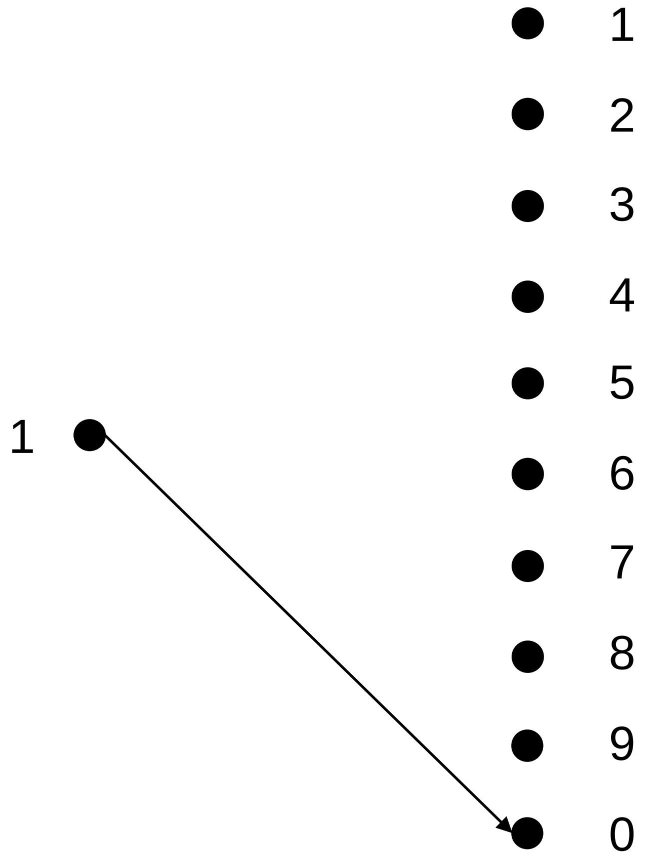
Your PIN Sounds Good!

Università degli Studi di Padova

NEW YORK INSTITUTE OF TECHNOLOGY

SPRITZ Security & Privacy Research Group

GFT



Distance DL

01**

30**

1 ●

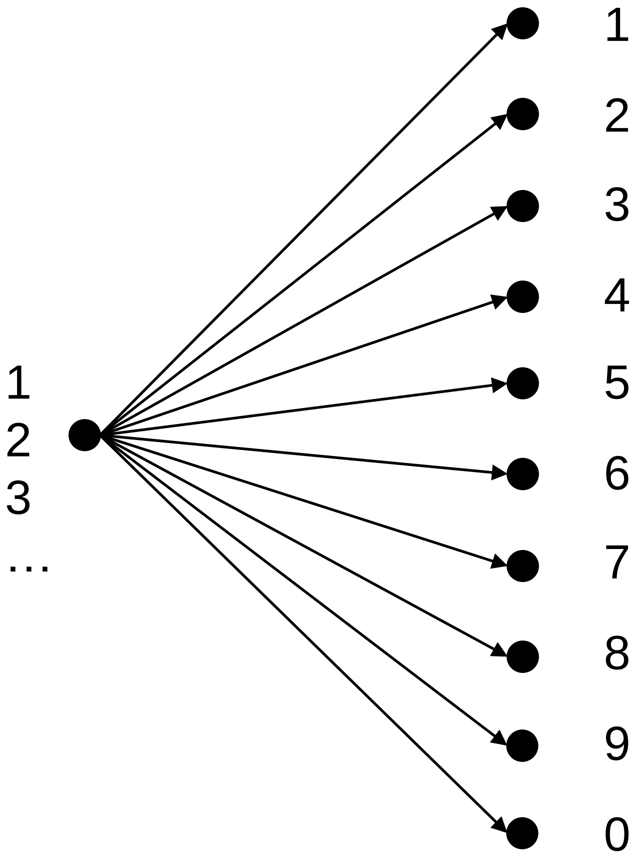
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 0

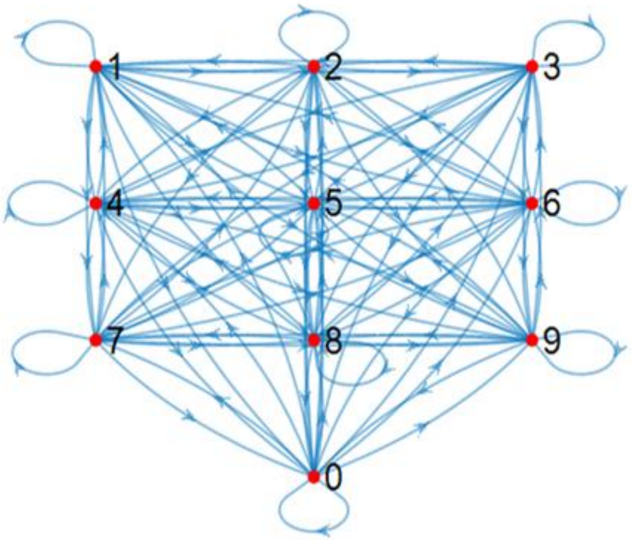
Distance 3

20**

Your PIN Sounds Good!

Università degli Studi di Padova
NEW YORK INSTITUTE OF TECHNOLOGY
SPRITZ Security & Privacy Research Group
GFT





Università degli Studi di Padova

New York Institute of Technology

SPRITZ Security & Privacy Research Group

GFT

How much does the knowledge of the distance affect PIN security?

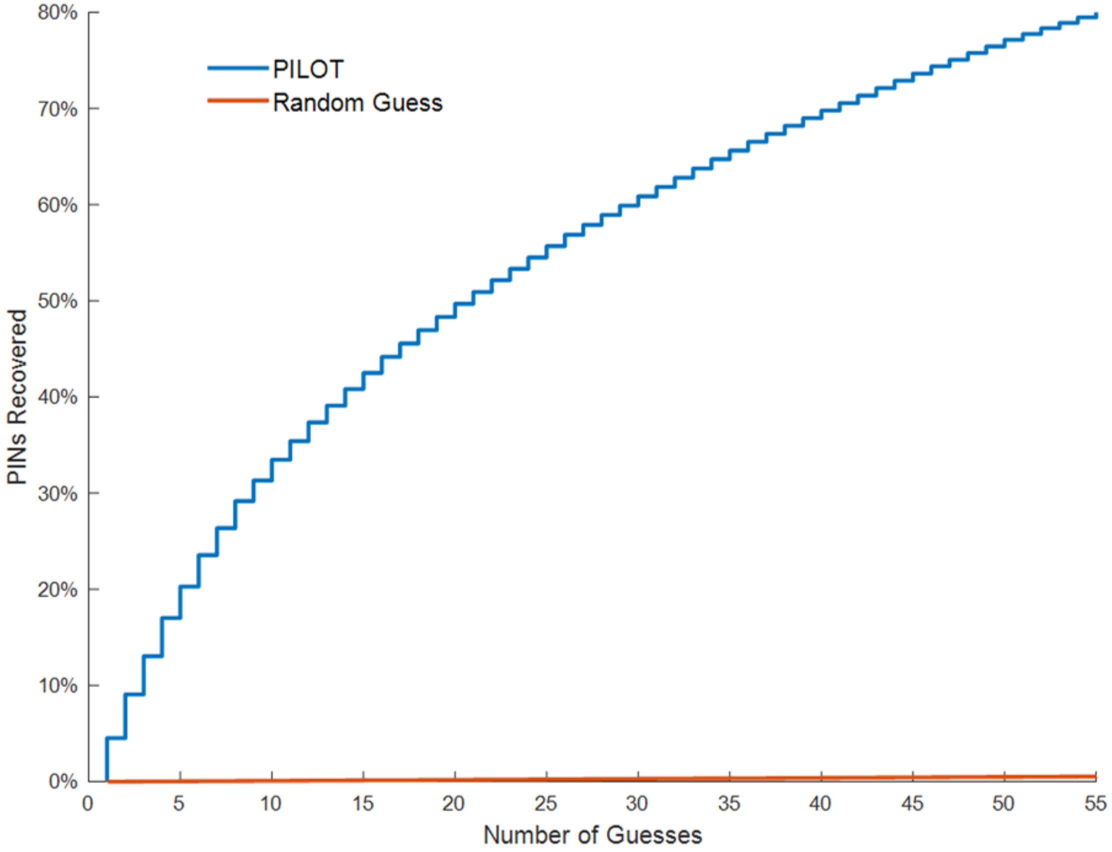
UNIVERSITÀ DEGLI STUDI DI PADOVA

NEW YORK INSTITUTE OF TECHNOLOGY

SPRITZ SECURITY & PRIVACY RESEARCH GROUP

GFT

How much does the knowledge of the distance affect PIN security?

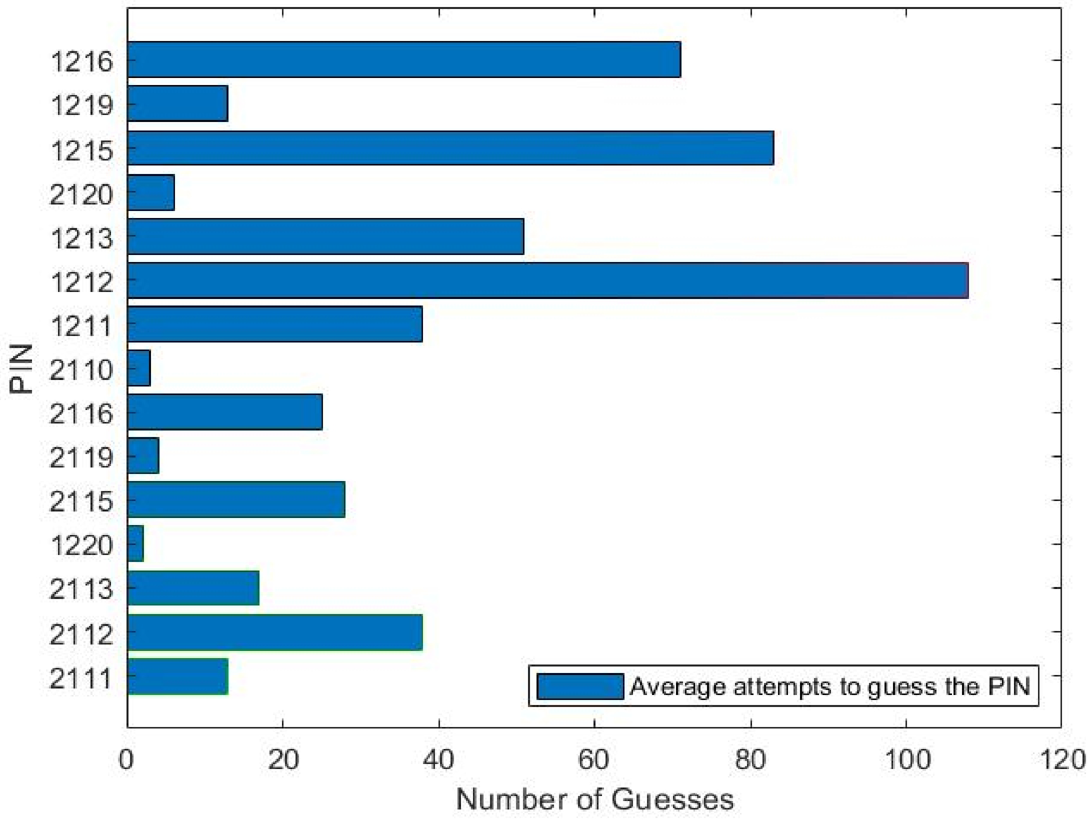


Balagani, Kiran, et al. "PILOT: Password and PIN information leakage from obfuscated typing videos." Journal of Computer Security.

Knowing the physical distance between keys poses another significant security problem ...

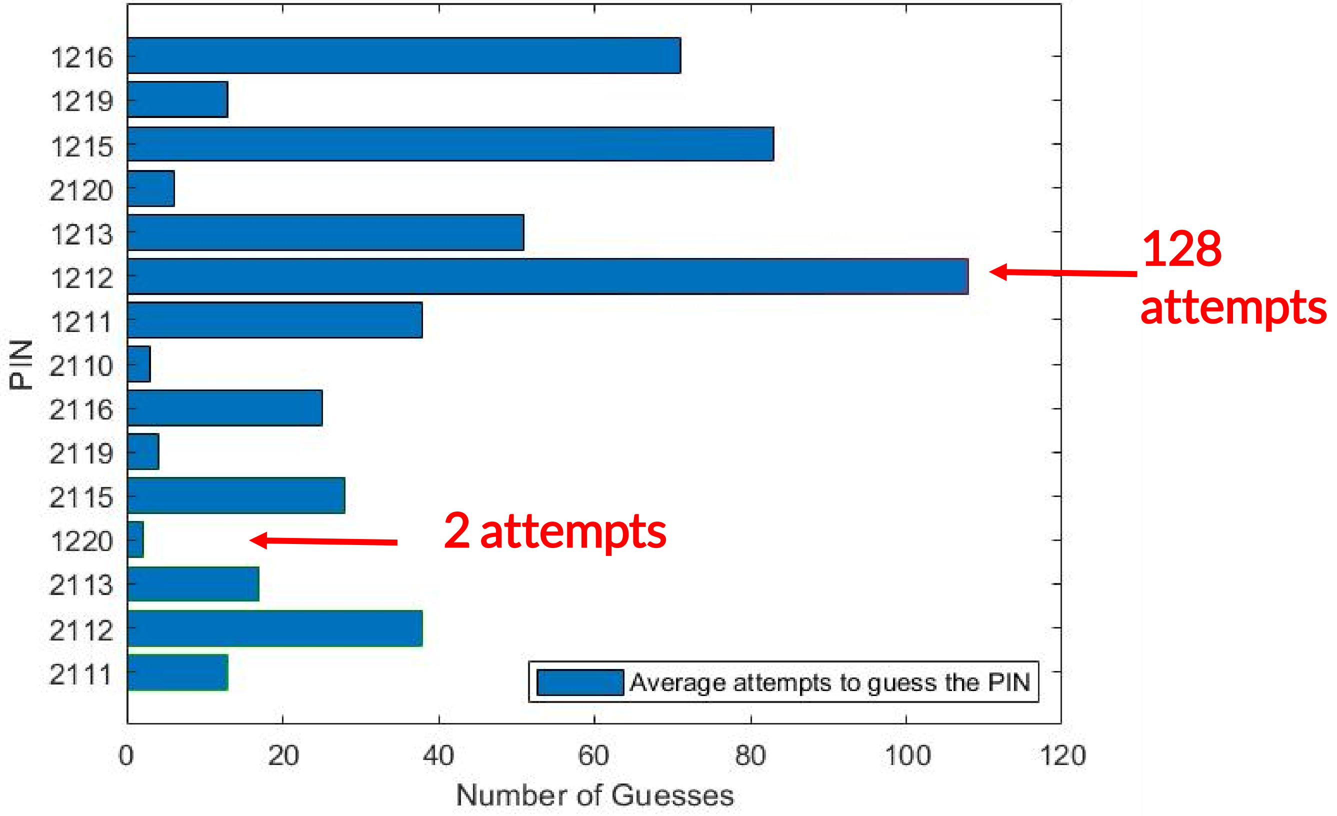
Your PIN Sounds Good!

Although distributed randomly and uniformly, some subsets of PINs may be more likely than others



Your PIN Sounds Good!

Although distributed randomly and uniformly, some subsets of PINs may be more likely than others

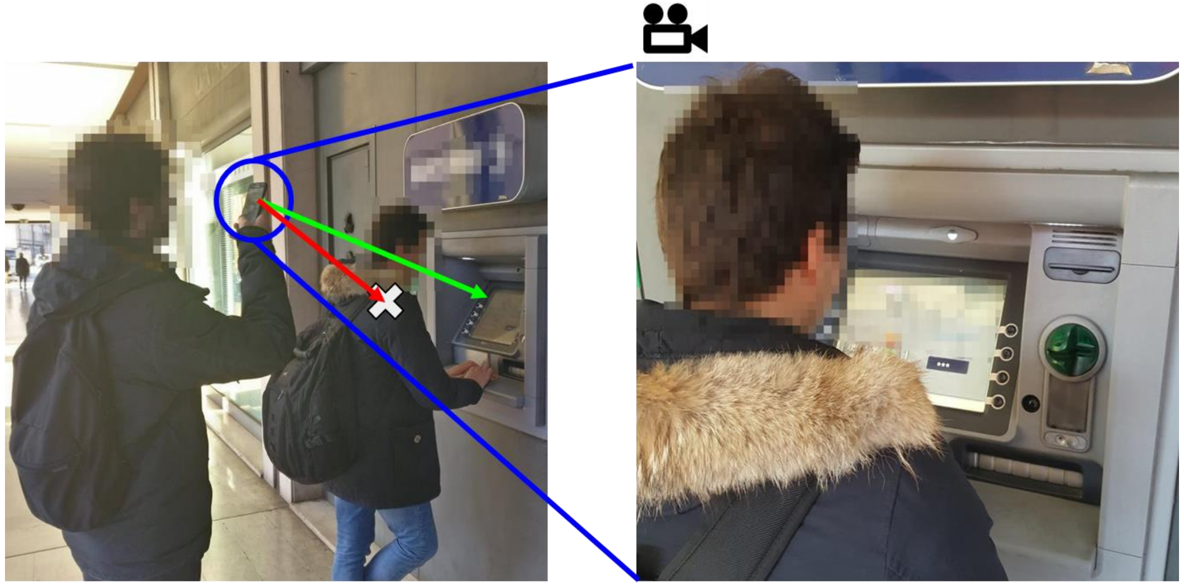


Your PIN Sounds Good!

Is it possible to deduce the distances between keys using environmental information?

ATM
$$$

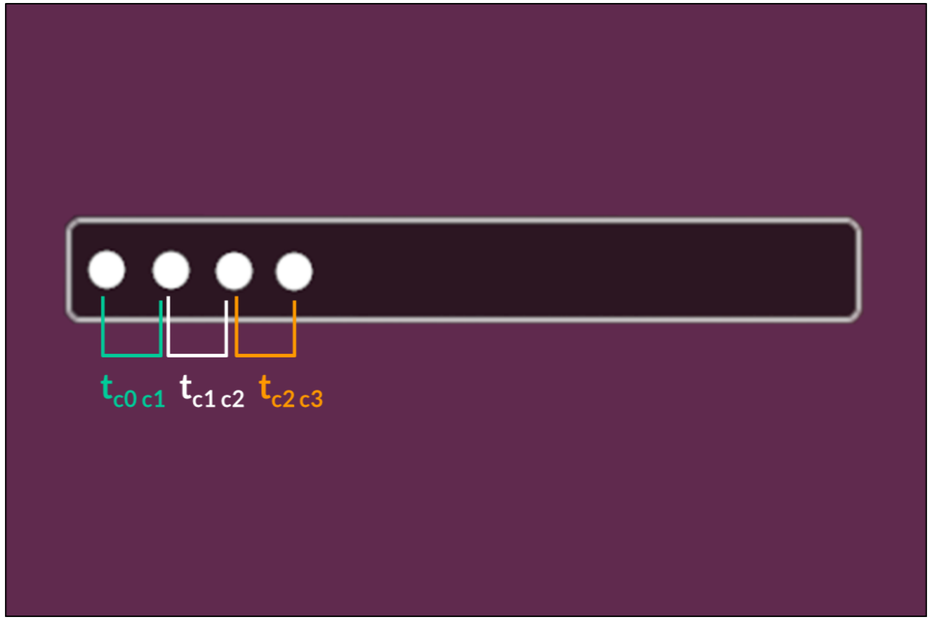
If the screen is visible, we can record the asterisk sequence ...



Università degli Studi di Padova
New York Institute of Technology
SPRITZ Security & Privacy Research Group
GFT

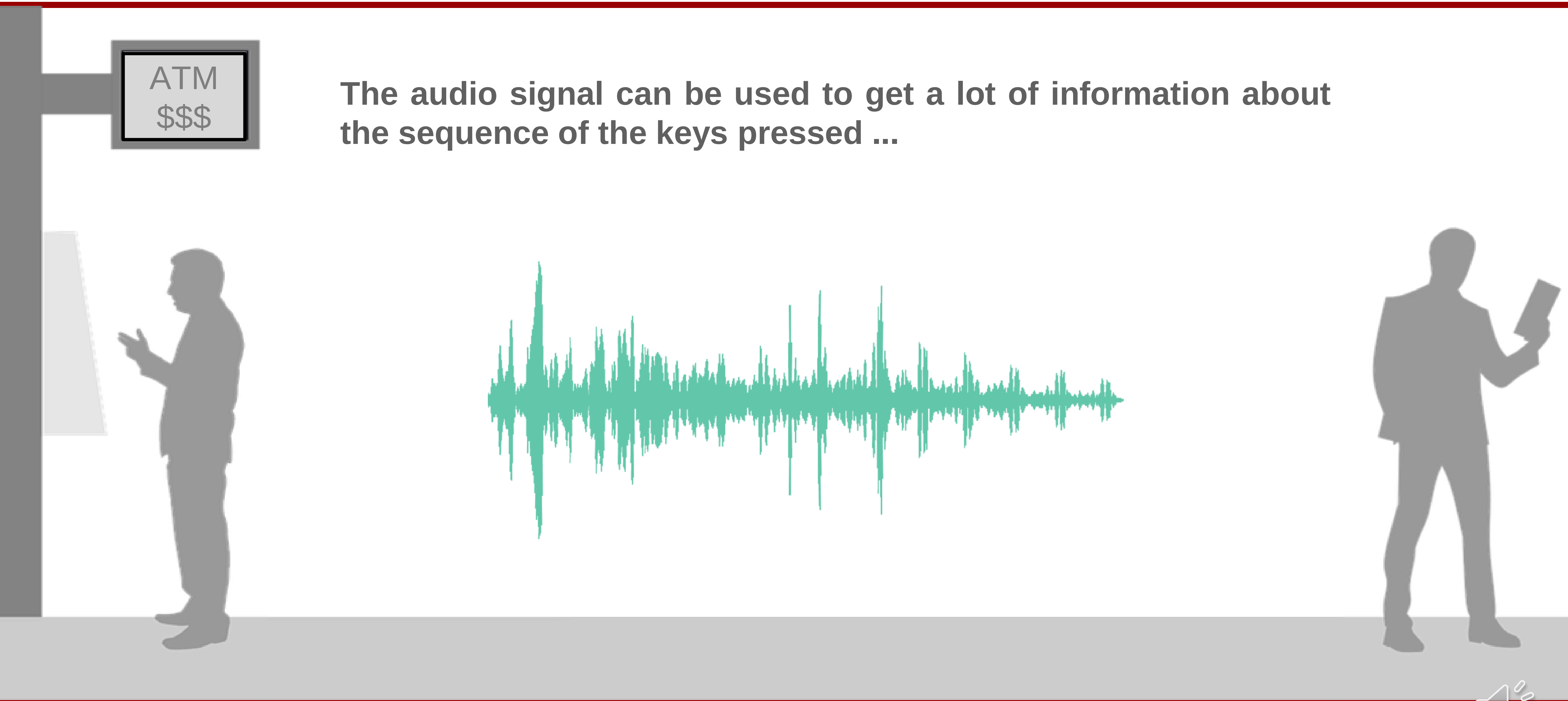
ATM
$$$

… and deduce the inter-keystroke timing



Your PIN Sounds Good!

The audio signal can be used to get a lot of information about the sequence of the keys pressed ...



UNIVERSITÀ DEGLI STUDI DI PADOVA

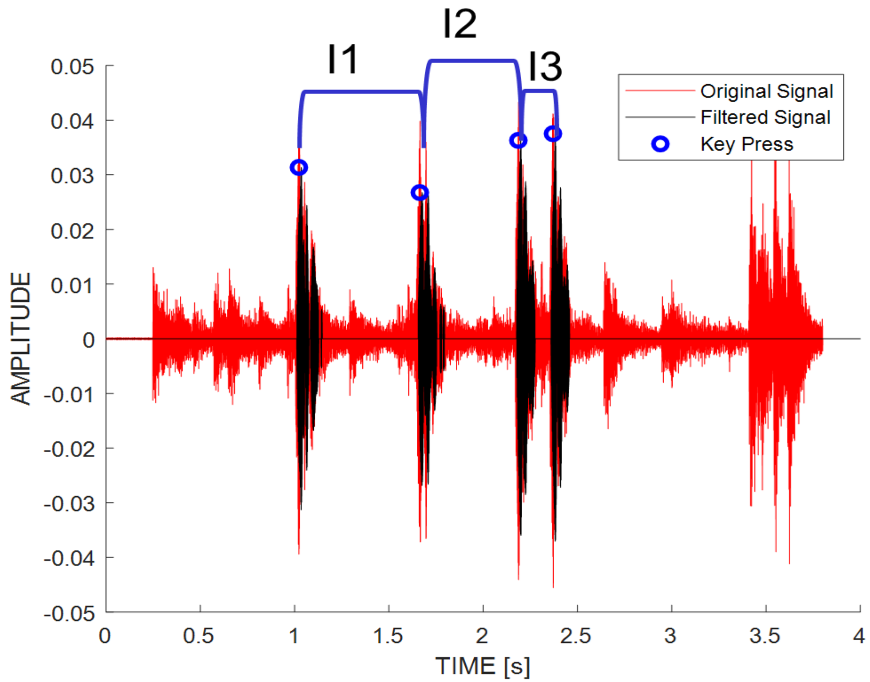
NEW YORK INSTITUTE OF TECHNOLOGY

SPRITZ SECURITY & PRIVACY RESEARCH GROUP

GFT

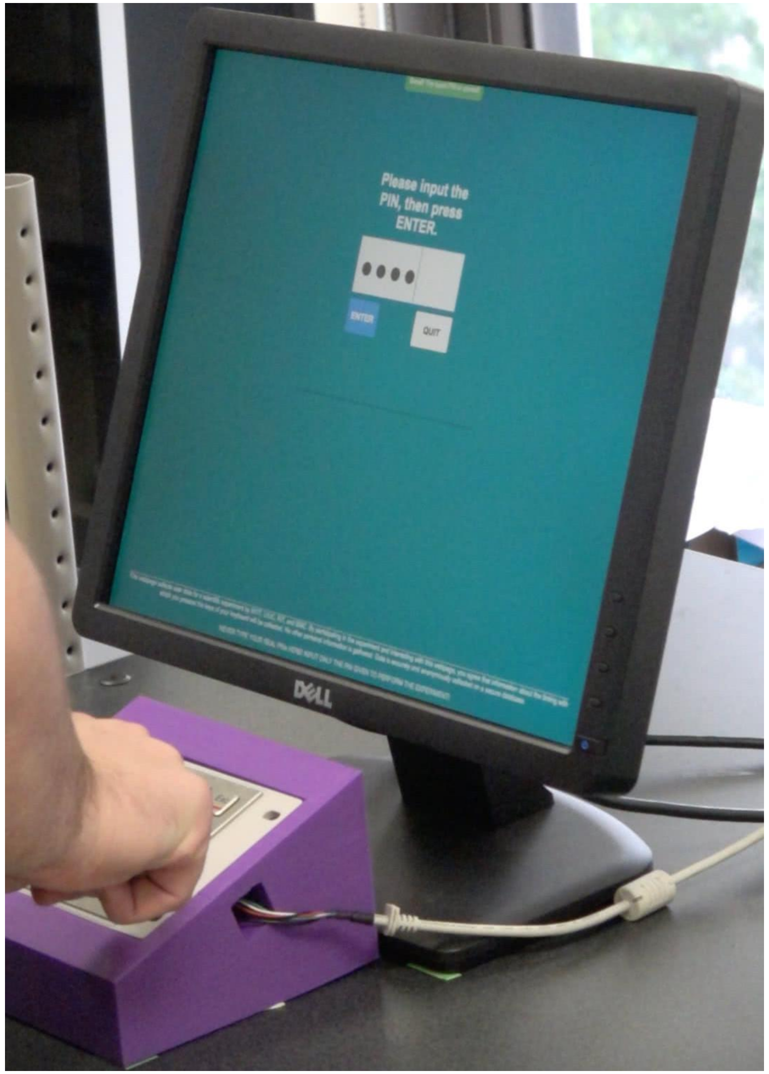
ATM
$$$

Filtering the audio it is possible to trace back to the instant in which the key was pressed, even in noisy environments



Your PIN Sounds Good!

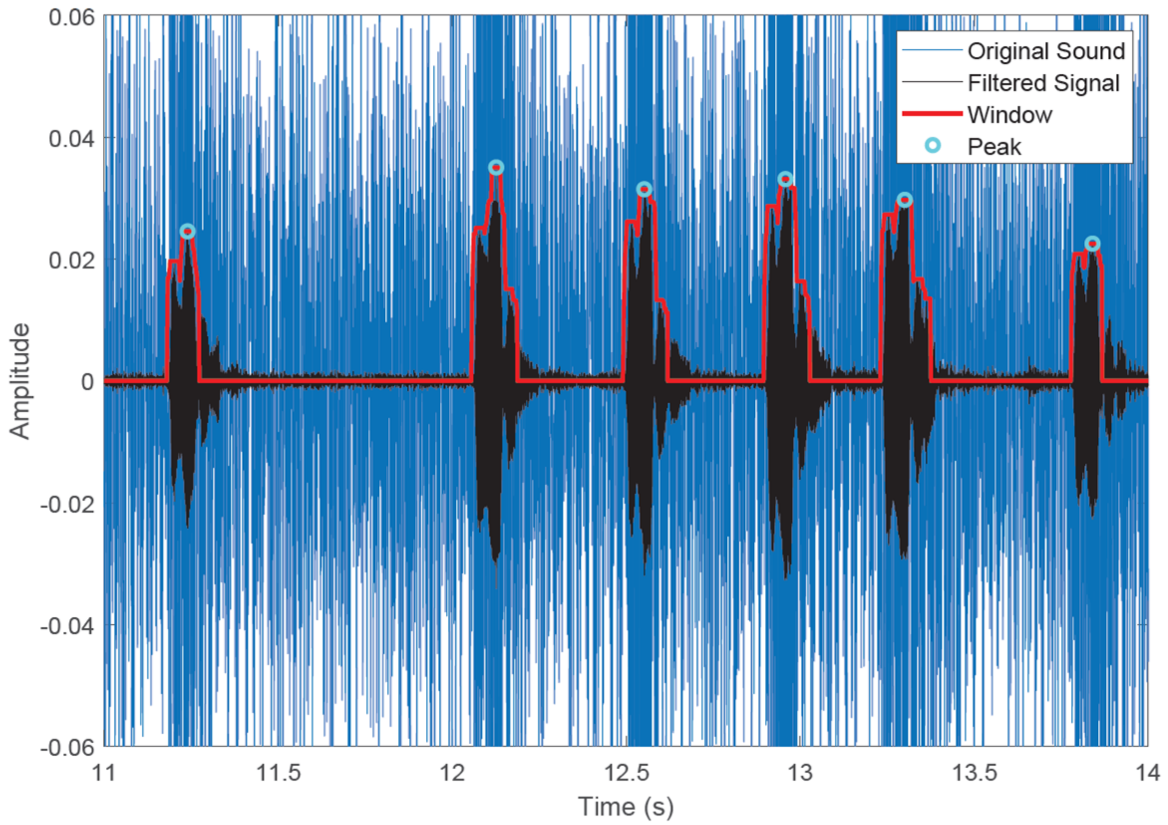
Università degli Studi di Padova
New York Institute of Technology
SPRITZ Security & Privacy Research Group
GFT



Dataset

- **22 participants** recorded with a camera located at a distance of 1.5 meters
- 15 different 4-digit PINs entered 12 times per session
- 19 participants completed 3 sessions
- 3 participants completed 1 session

Inter-keystroke timing extraction



- Linear **normalization** of the audio recording amplitude in the interval [-1; 1]
- 16-order **Butterworth band-pass filter** centered in 5.6 kHz to isolate the feedback sound frequency
- **Samples** with an amplitude below 0.01 were “**muted**”
- **Maximum samples amplitude** in a sliding window of 1200 samples (25 milliseconds)

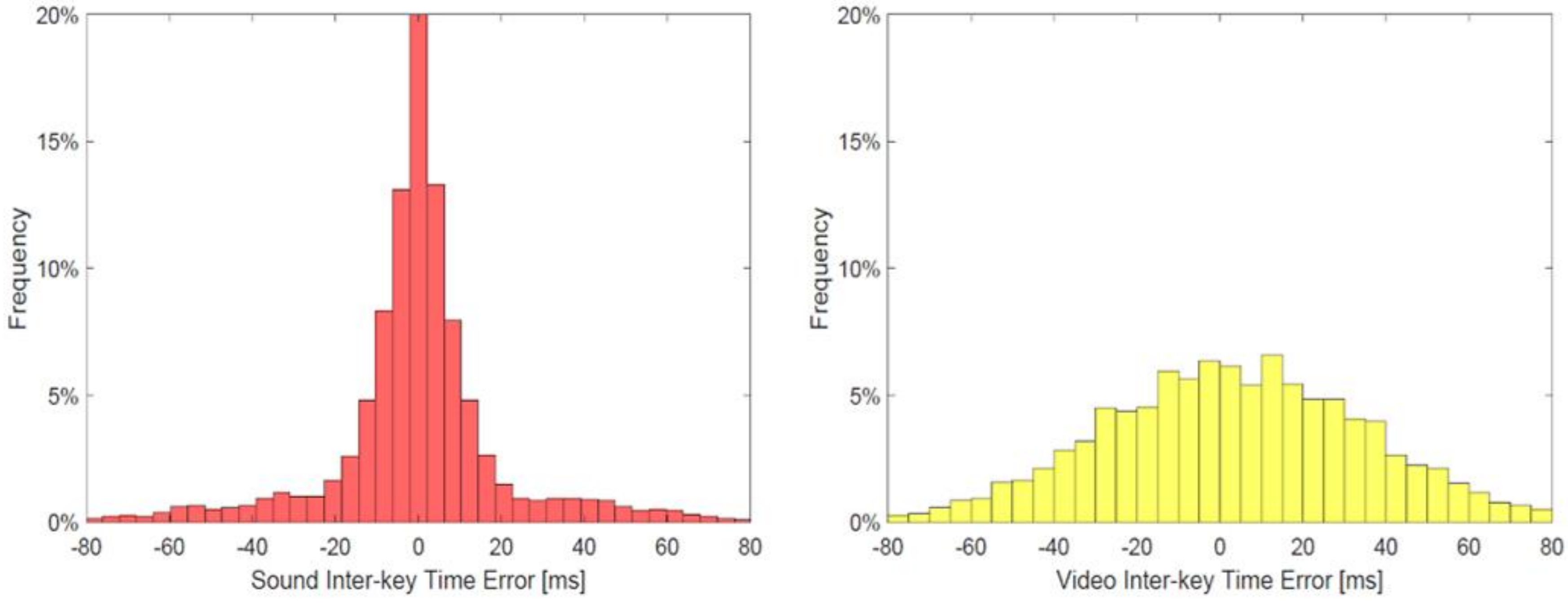
Università degli Studi di Padova

NEW YORK INSTITUTE OF TECHNOLOGY

SPRITZ Security & Privacy Research Group

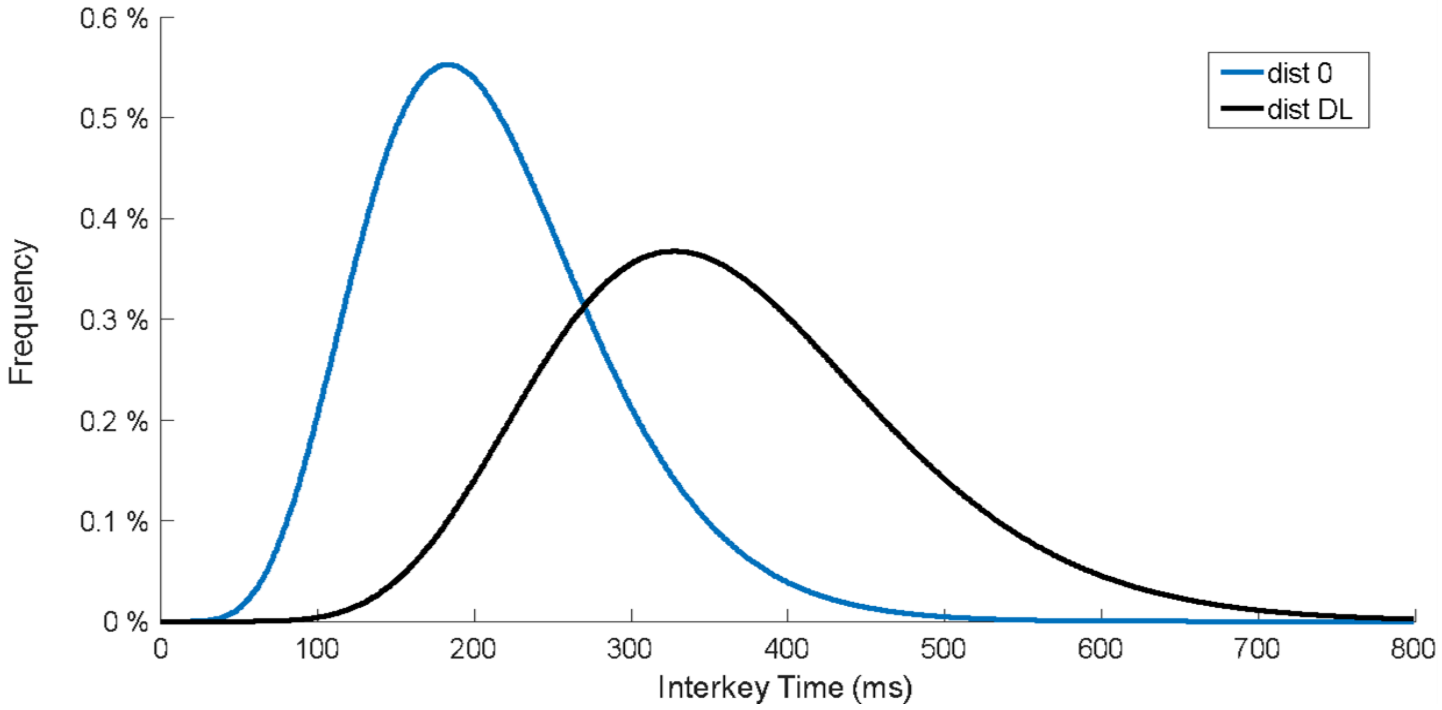
GFT

Sound vs Video Inter-keystroke timing



ATM
$$$

Inter-keystroke timing difference between distance 0 (11**, 00**) and distance DL (10**, 30**)



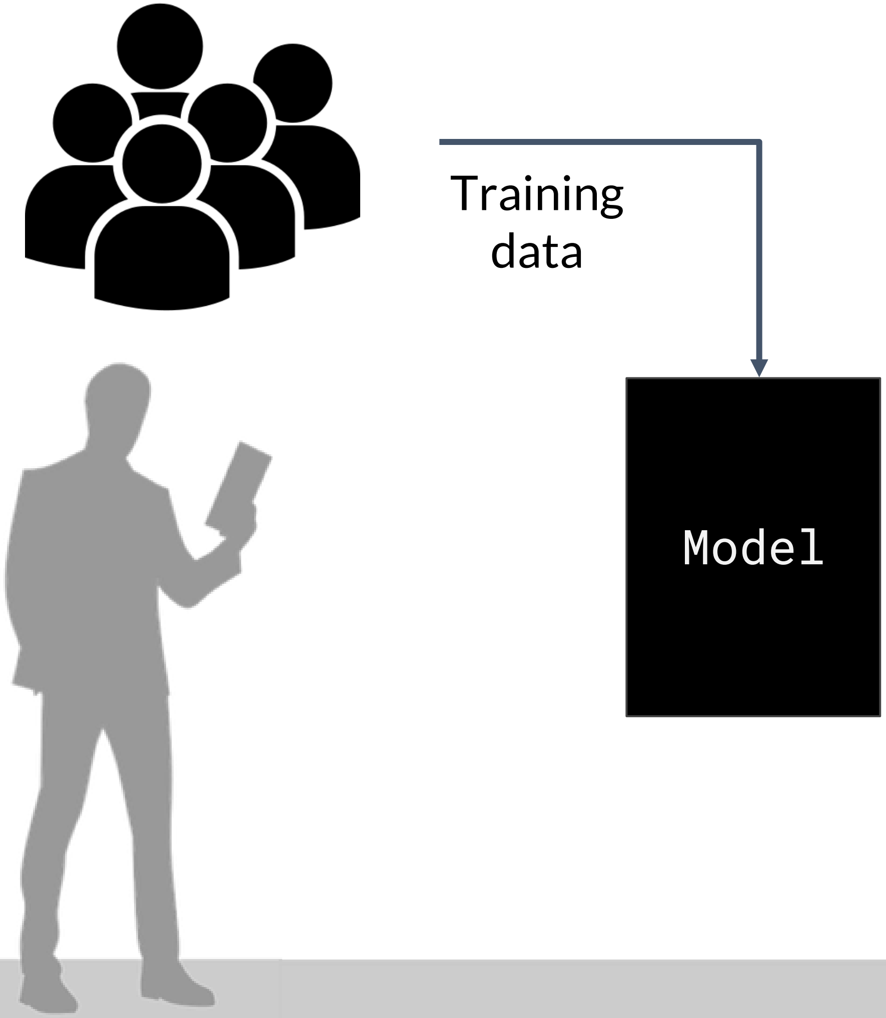
Your PIN Sounds Good!

UNIVERSITÀ DEGLI STUDI DI PADOVA

NEW YORK INSTITUTE OF TECHNOLOGY

SPRITZ SECURITY & PRIVACY RESEARCH GROUP

GFT



Training set

- 11 Users

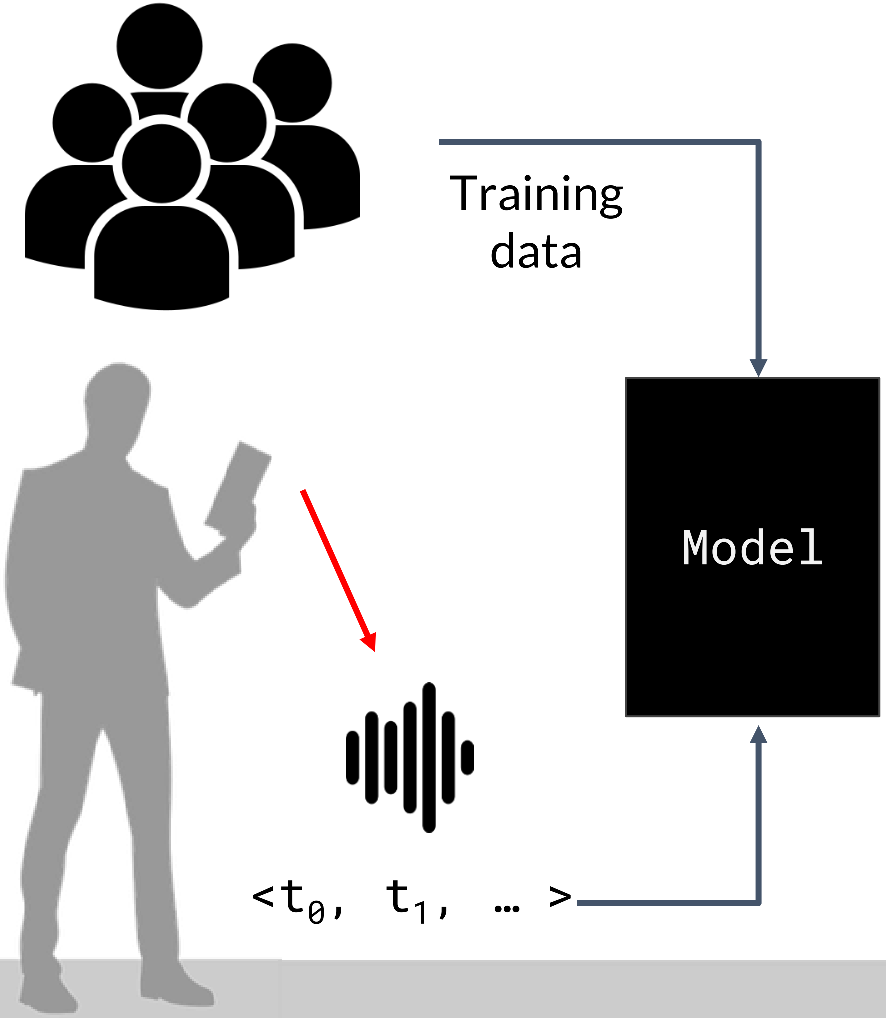
- 5195 PINs

Test set

- 11 Users

- 5135 PINs

Your PIN Sounds Good!



Training set

- 11 Users

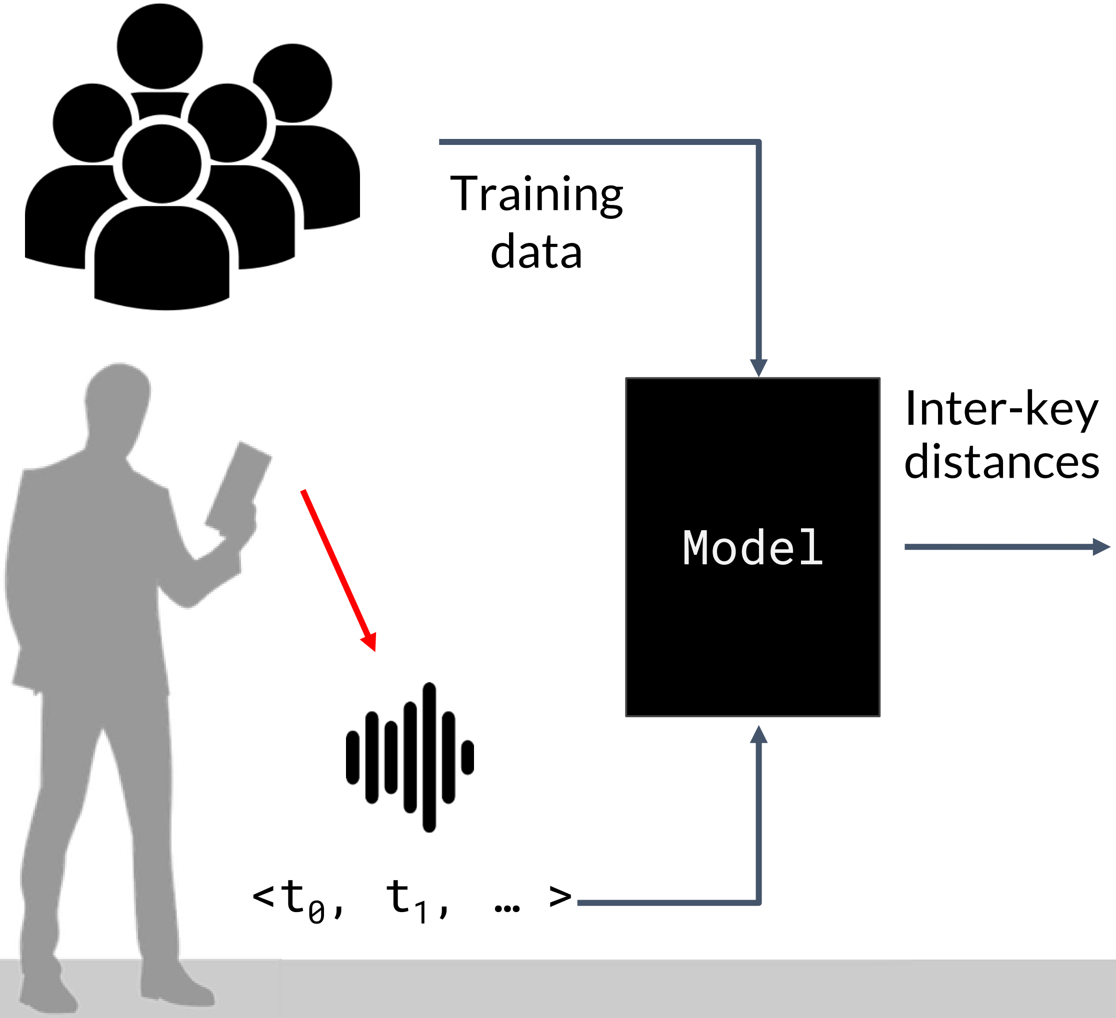
- 5195 PINs

Test set

- 11 Users

- 5135 PINs

Your PIN Sounds Good!



Training set

- 11 Users

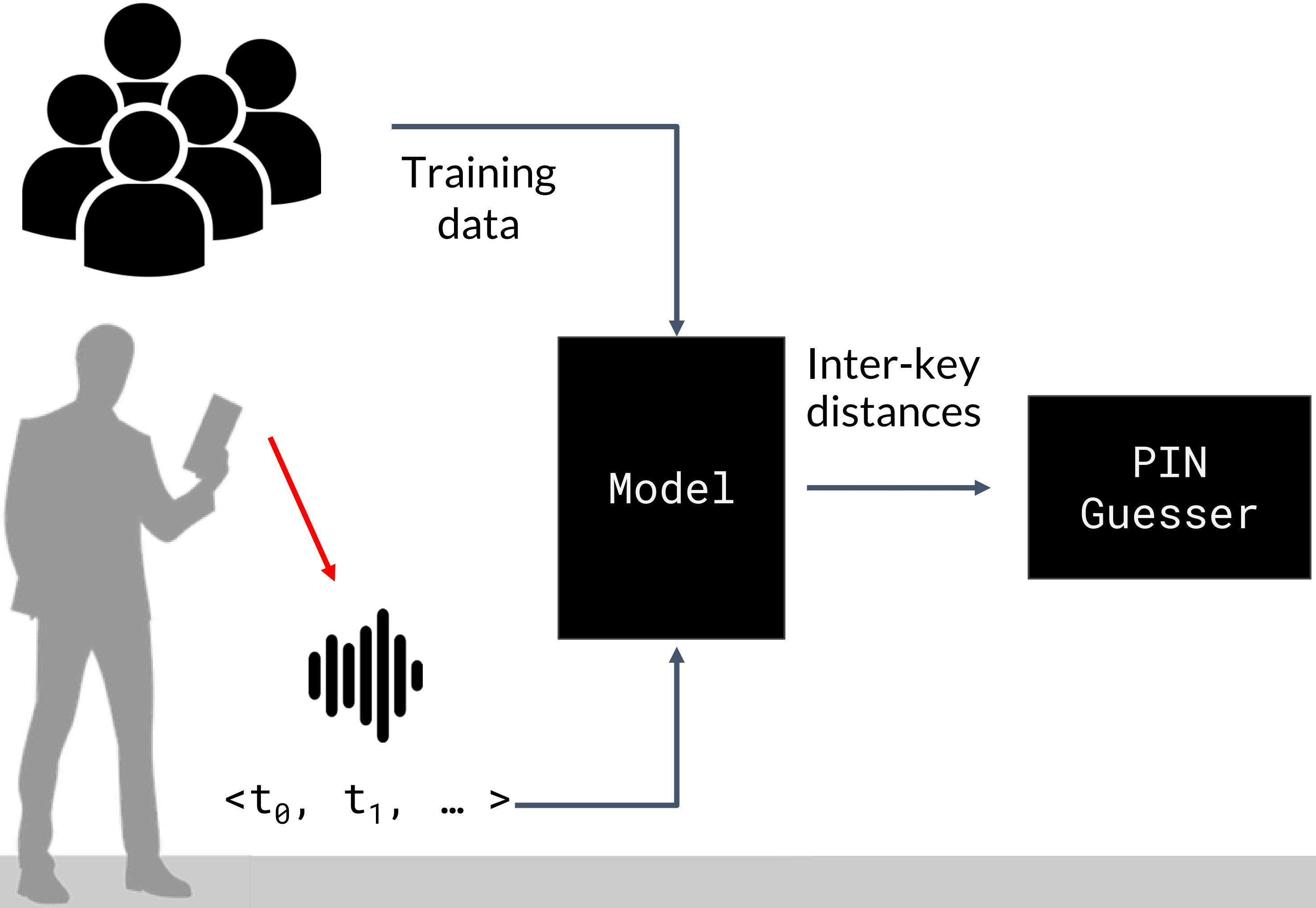
- 5195 PINs

Test set

- 11 Users

- 5135 PINs

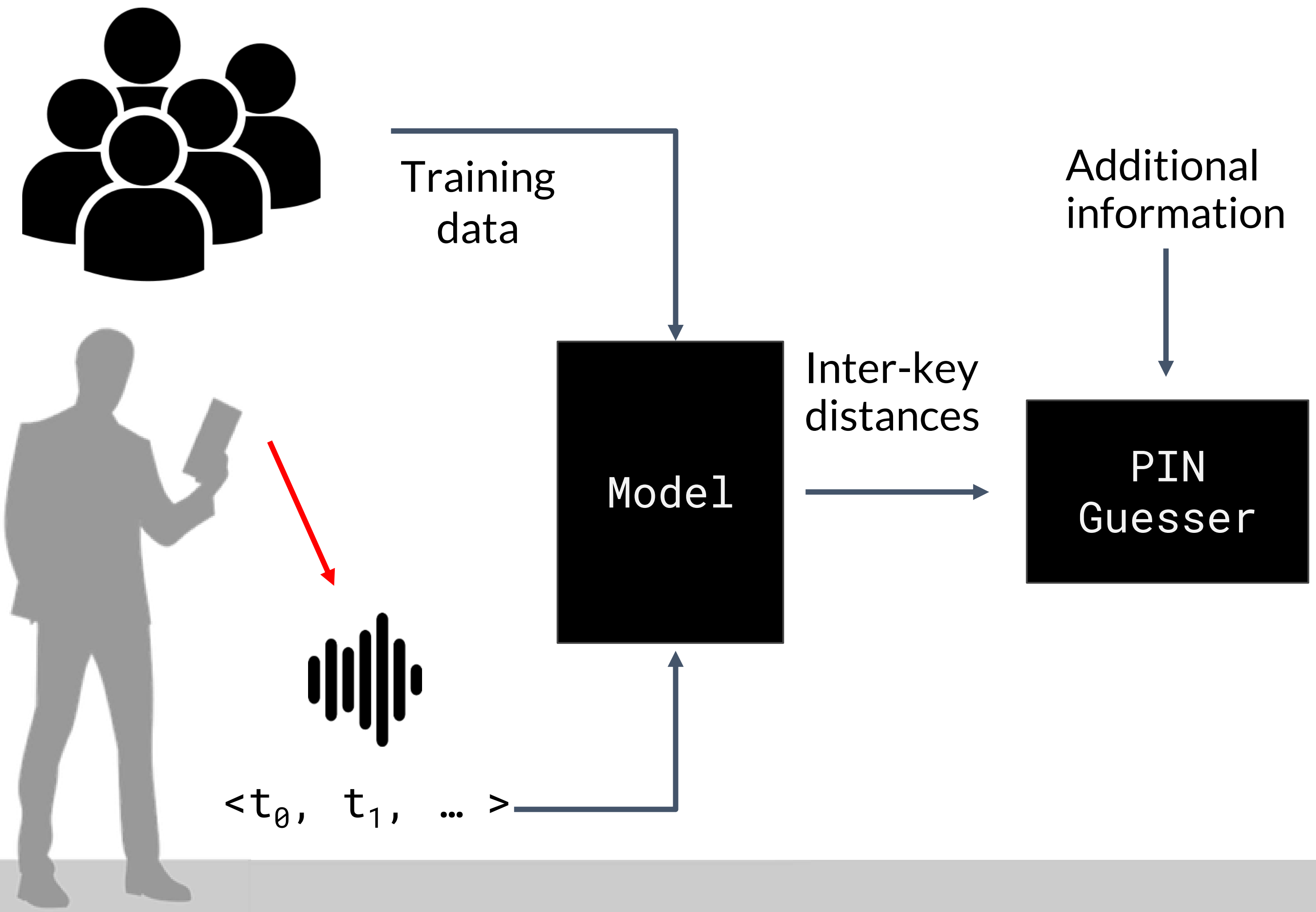
Your PIN Sounds Good!



Training set
- 11 Users
- 5195 PINs

Test set
- 11 Users
- 5135 PINs

Your PIN Sounds Good!



Training set

- 11 Users

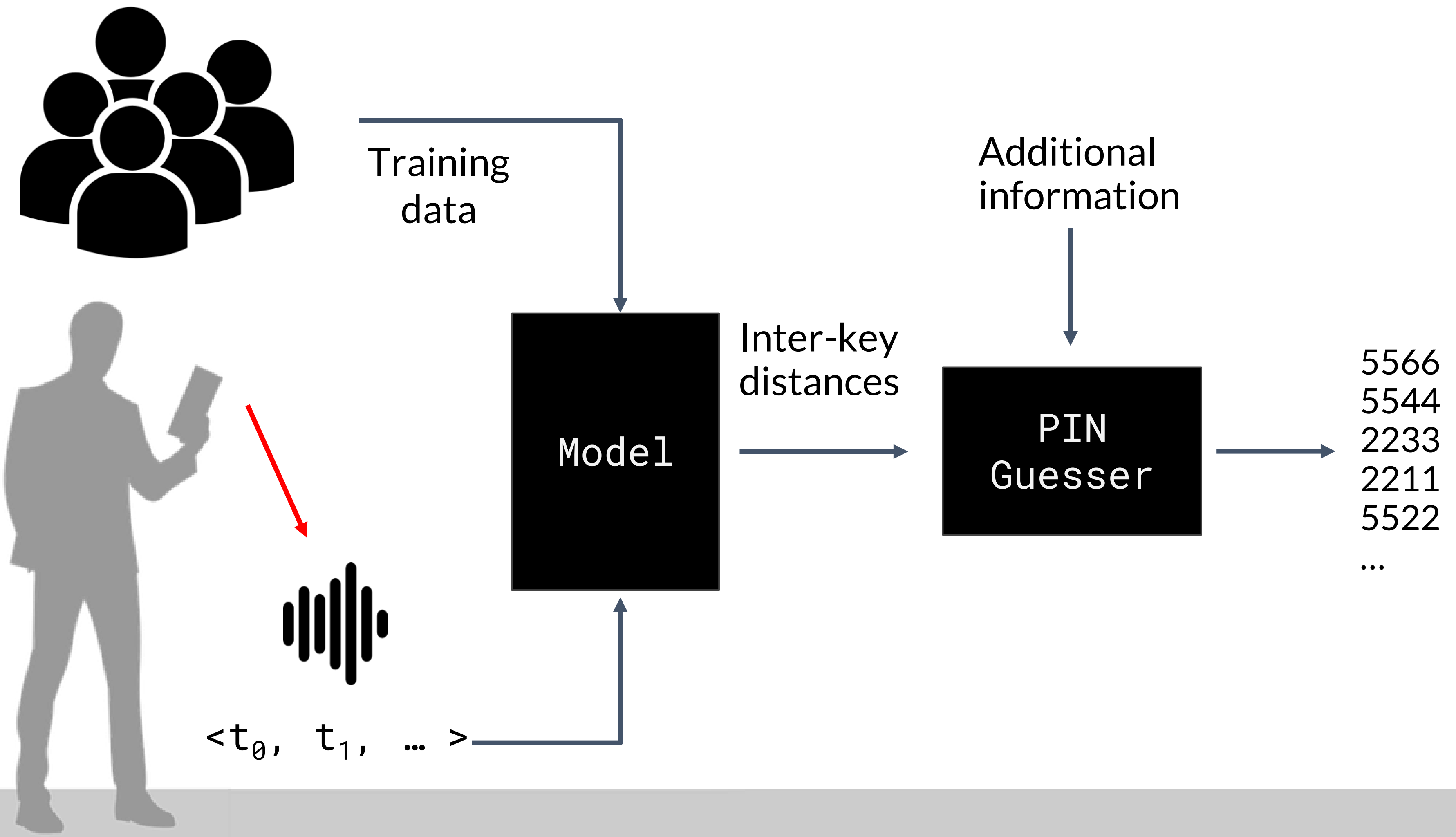
- 5195 PINs

Test set

- 11 Users

- 5135 PINs

Your PIN Sounds Good!



Training set

- 11 Users

- 5195 PINs

Test set

- 11 Users

- 5135 PINs

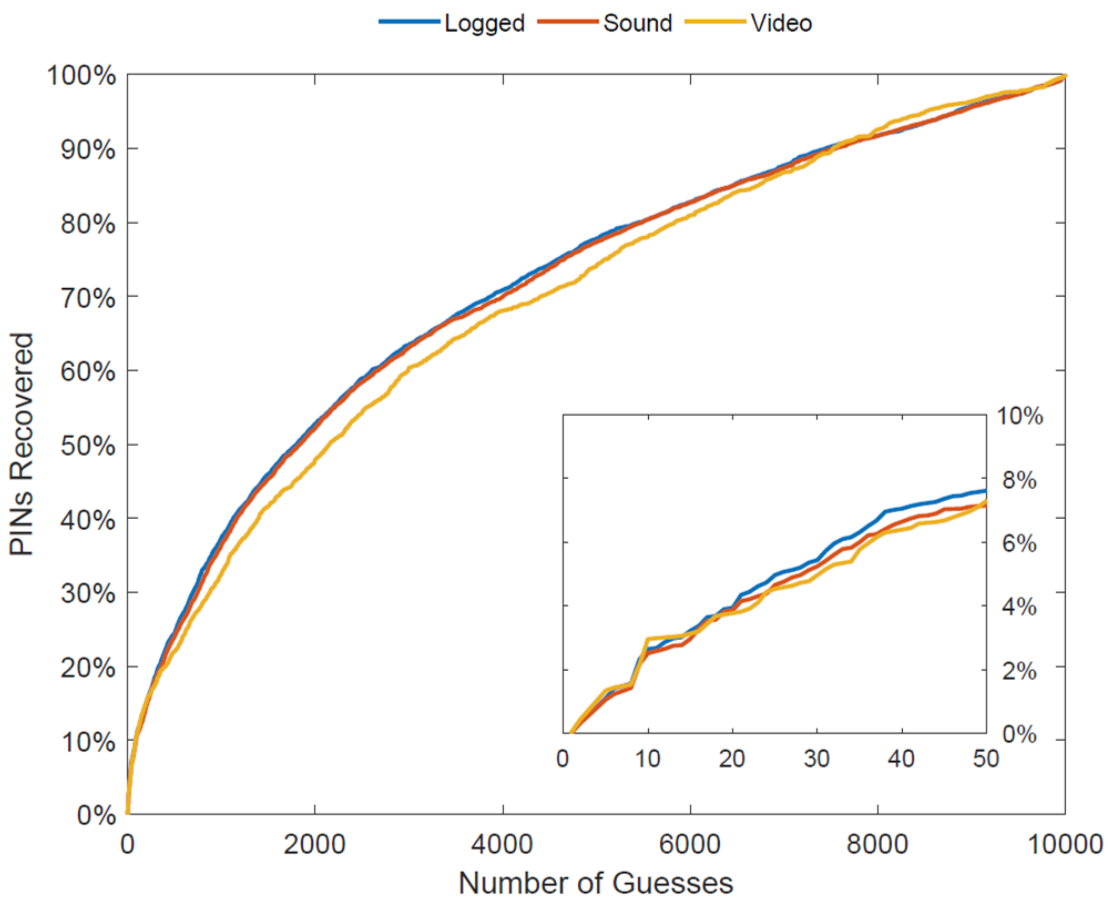
Your PIN Sounds Good!

UNIVERSITÀ DEGLI STUDI DI PADOVA

NEW YORK INSTITUTE OF TECHNOLOGY

SPRITZ Security & Privacy Research Group

GFT

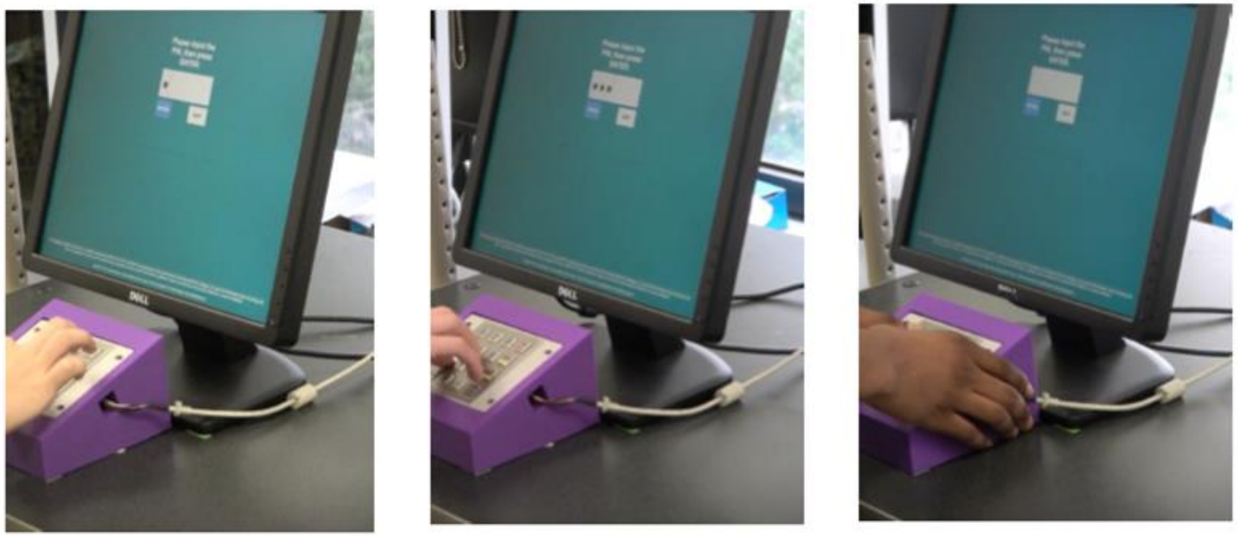


Your PIN Sounds Good!

Does the way we type the PIN provide useful information?

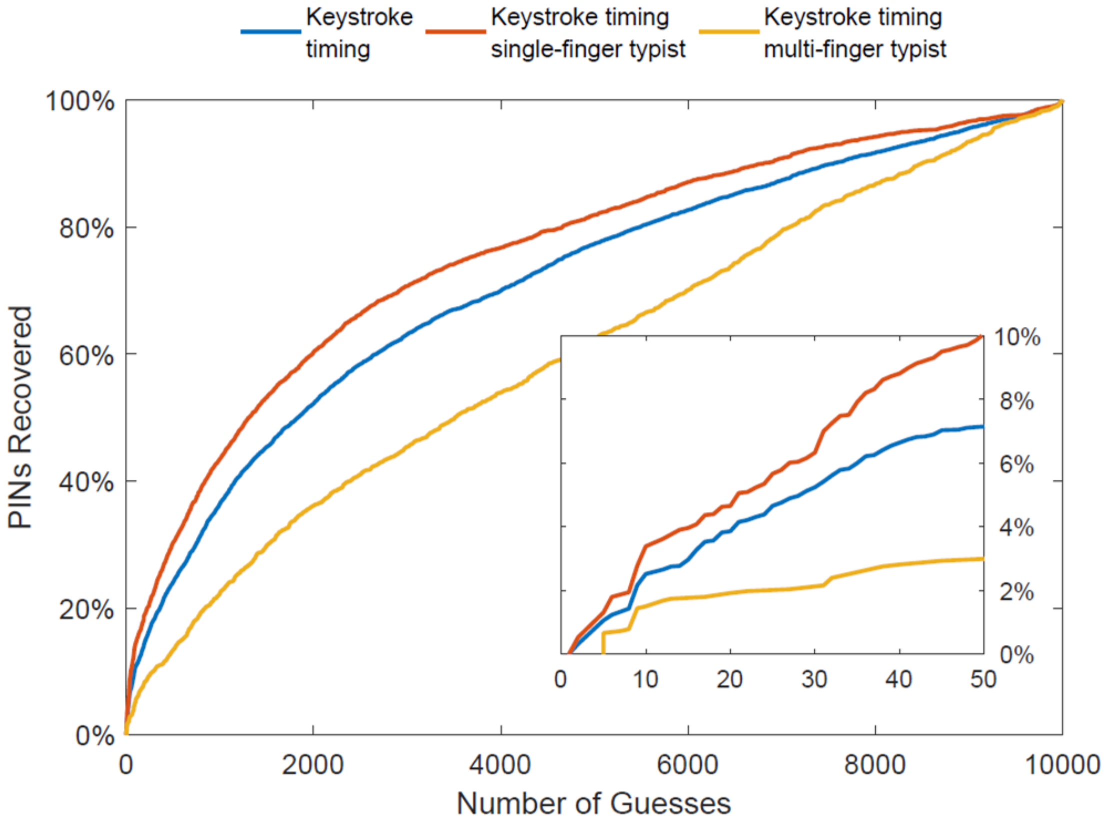
Your PIN Sounds Good!

"**Single finger PIN**" (SFP) represent **70% of the total**, **92%** of them is entered using the **index** while **8%** is typed with the **thumb**.



Your PIN Sounds Good!

There is a correlation between distance and PIN entry method

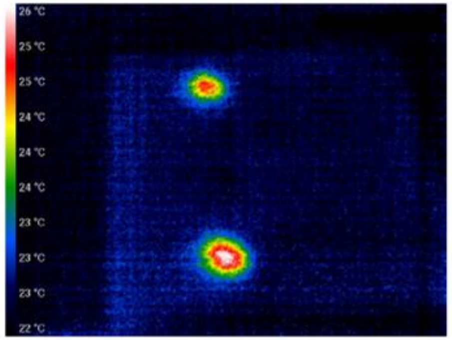


Can we get information even **after** entering the PIN?

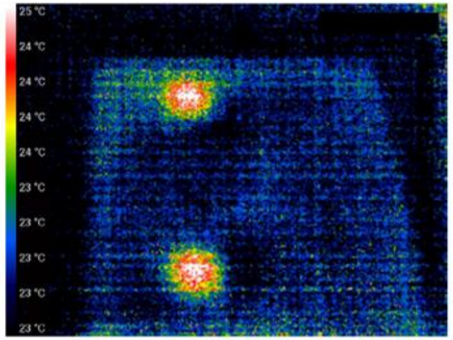
Università degli Studi di Padova | New York Institute of Technology | SPRITZ Security & Privacy Research Group | GFT

Università degli Studi di Padova
New York Institute of Technology
SPRITZ Security & Privacy Research Group
GFT

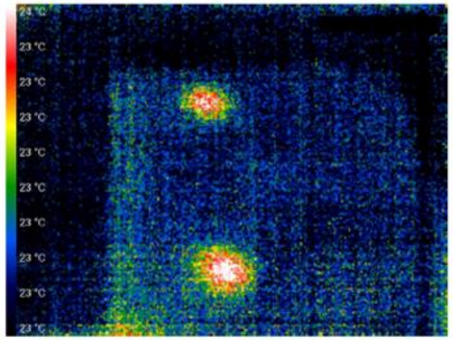
Using a thermal camera it is possible to identify the **thermal trace** left on the keypad



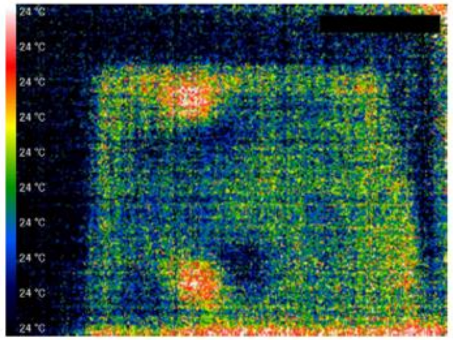
(a) Thermal trace after 2 seconds.



(b) Thermal trace after 7 seconds.



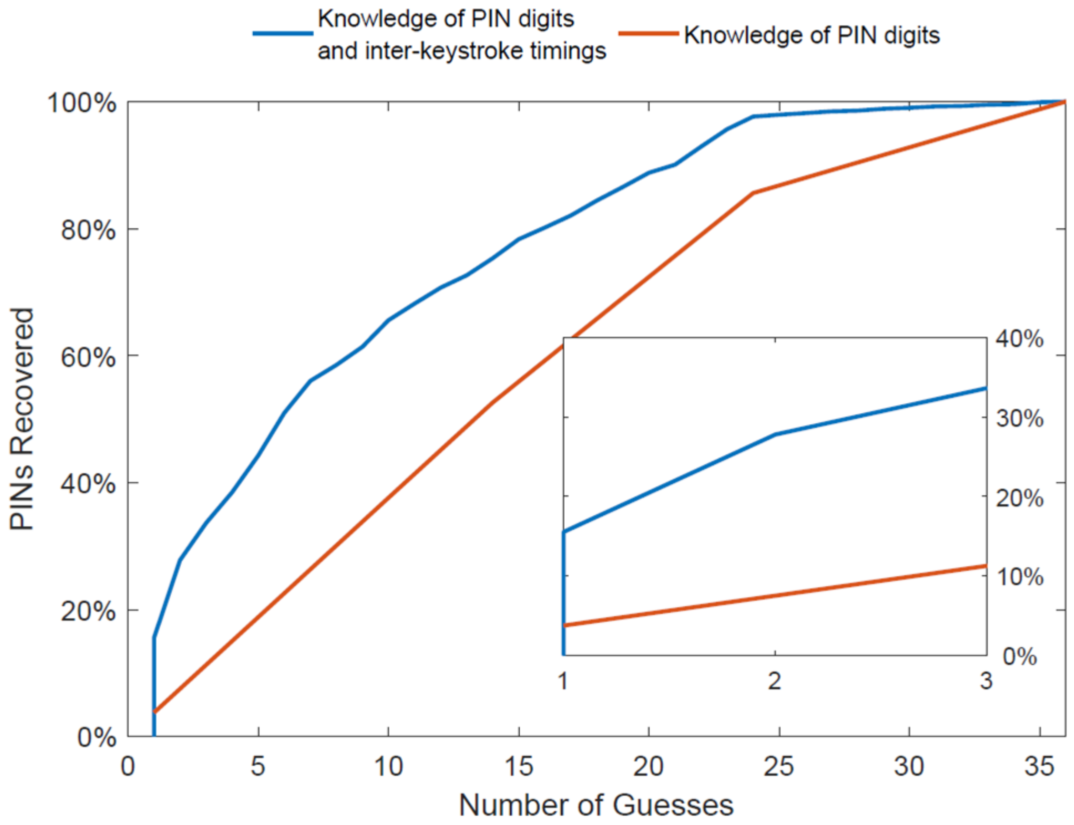
(c) Thermal trace after 10 seconds.



(d) Thermal trace after 15 seconds.

Your PIN Sounds Good!

Using a thermal camera it is possible to identify the **thermal trace** left on the keypad



Your PIN Sounds Good!

Information				PINs Guessed Within Attempt				
Keystroke Timing	Single Finger	First Digit	PIN Digits	1	2	3	5	10
				0.01%	0.02%	0.03%	0.05%	0.10%
		o		0.10%	0.20%	0.30%	0.50%	1.00%
o				0.02%	0.31%	0.70%	1.05%	2.51%
o	o			0.03%	0.52%	0.91%	1.30%	3.38%
o		o		3.02%	3.72%	4.36%	6.97%	11.04%
o	o	o		3.73%	4.13%	5.43%	8.73%	14.01%
			o	3.76%	7.52%	11.28%	18.80%	37.60%
o			o	15.54%	27.79%	33.63%	44.25%	65.57%
o	o		o	19.04%	34.01%	40.60%	50.74%	71.31%
		o	o	13.27%	26.62%	39.88%	66.40%	92.80%
o		o	o	35.27%	53.46%	66.84%	86.76%	98.99%
o	o	o	o	40.86%	60.24%	71.77%	89.19%	99.28%

Conclusion

- It is possible to retrieve **accurate inter-keystroke** timing information **from audio** in a real context
- **Inter-keystroke timing** inferred from audio feedback emitted by standard PIN pads can be effectively used to **reduce the attempts to guess a PIN**
- Compared to prior sources of keystroke timing information, **audio feedback is easier to collect**
- The **typing behavior** affects the adversary's ability to guess PINs
- **Inter-keystroke** timing can be **combined with other information** to drastically reduce the number of attempts required to guess a PIN

Thank you!

Matteo Cardaioli

matteo.cardaioli@phd.unipd.it

Backup slides

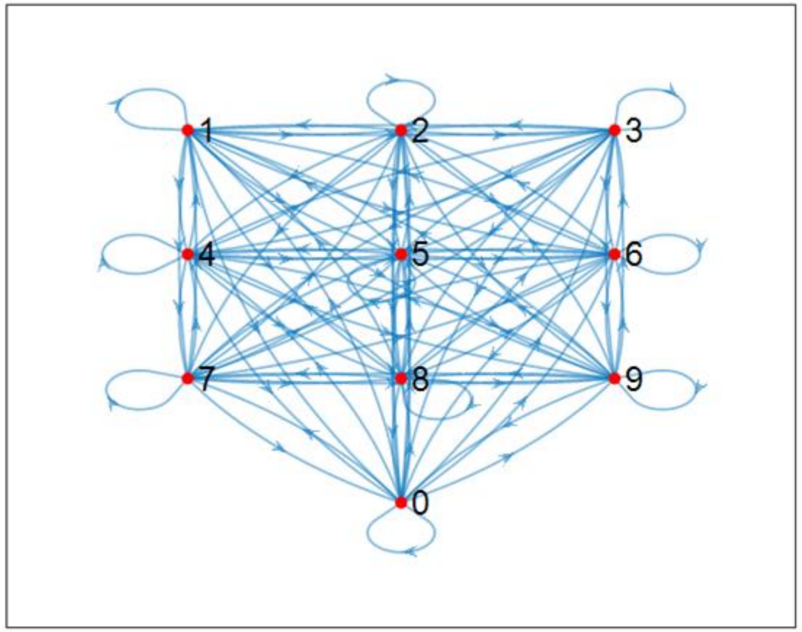
Use case…

PIN Salabim

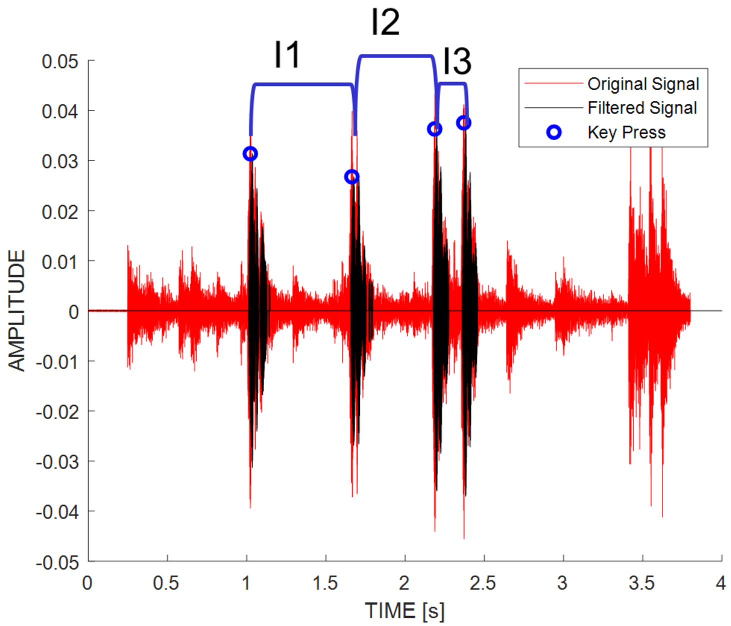
Use case…
Let's try to guess the PIN 1077

PIN Salabim

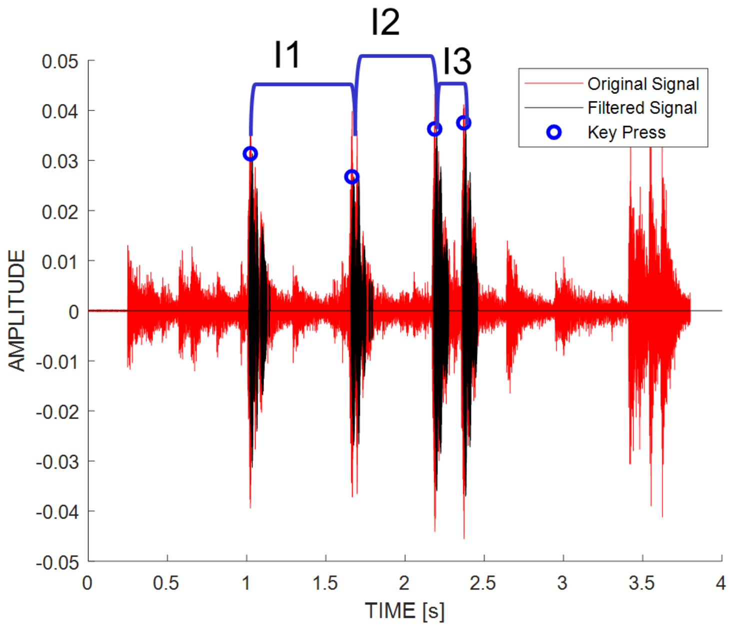
Use case…
Let's try to guess the **PIN 1077**

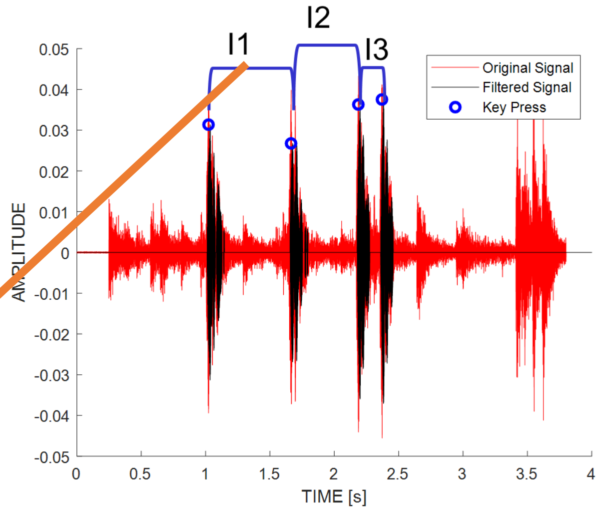


The first step is to **filter the audio signal** to extract the inter key timing

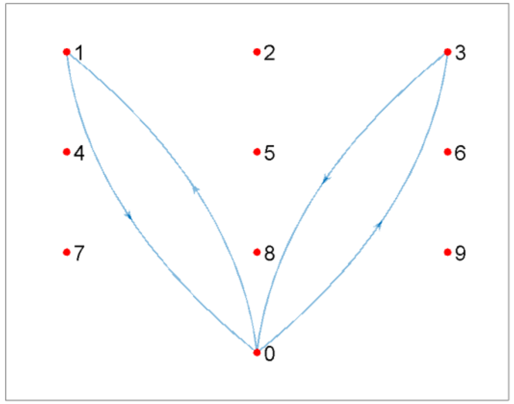


From the inter-keystroke timing, the ML model tells us which physical distances are most likely for each consecutive pair of keys

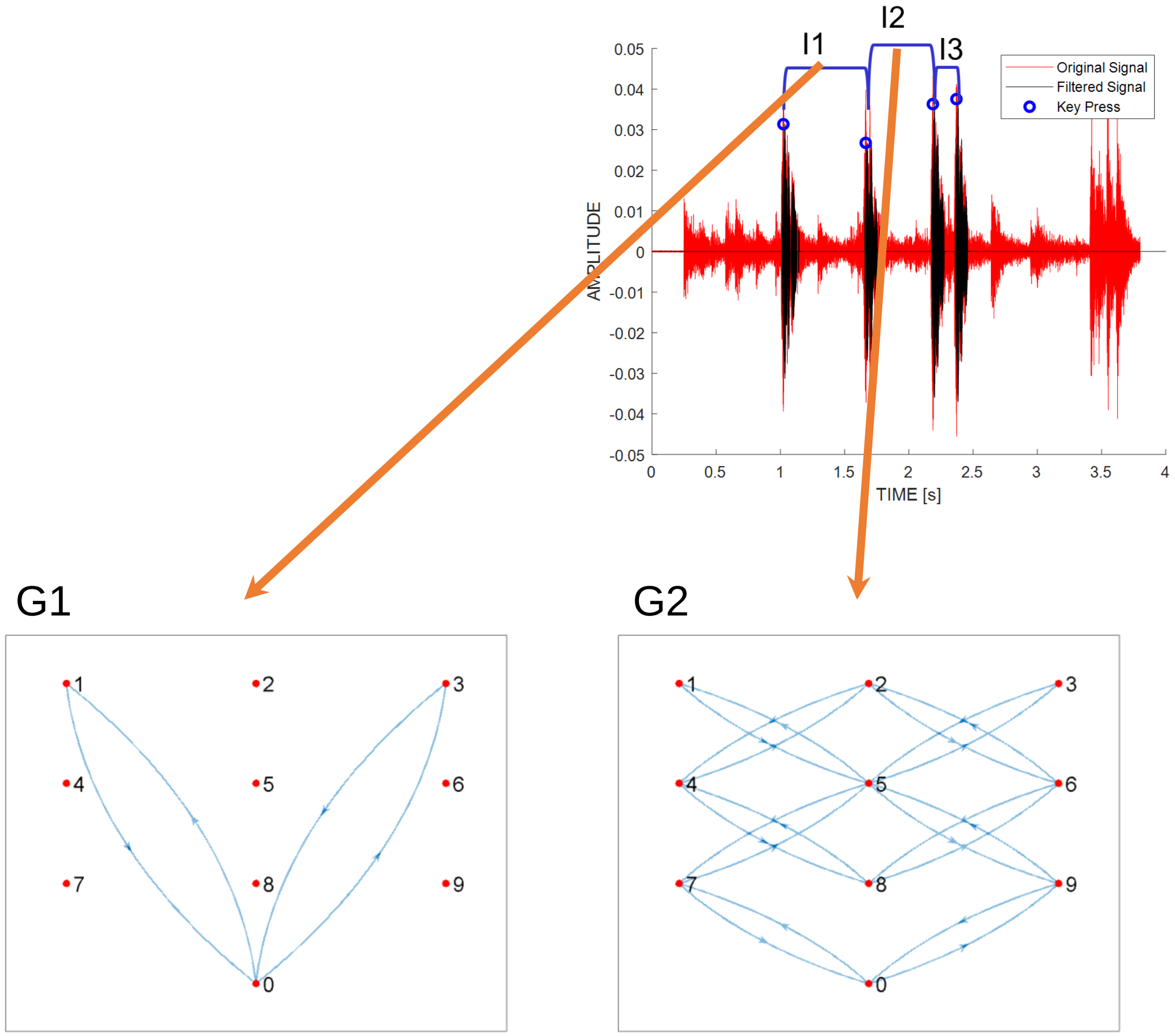



I1
I2
I3
Original Signal
Filtered Signal
Key Press
AMPLITUDE
TIME [s]
0.05
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
-0.05
0
0.5
1
1.5
2
2.5
3
3.5
4

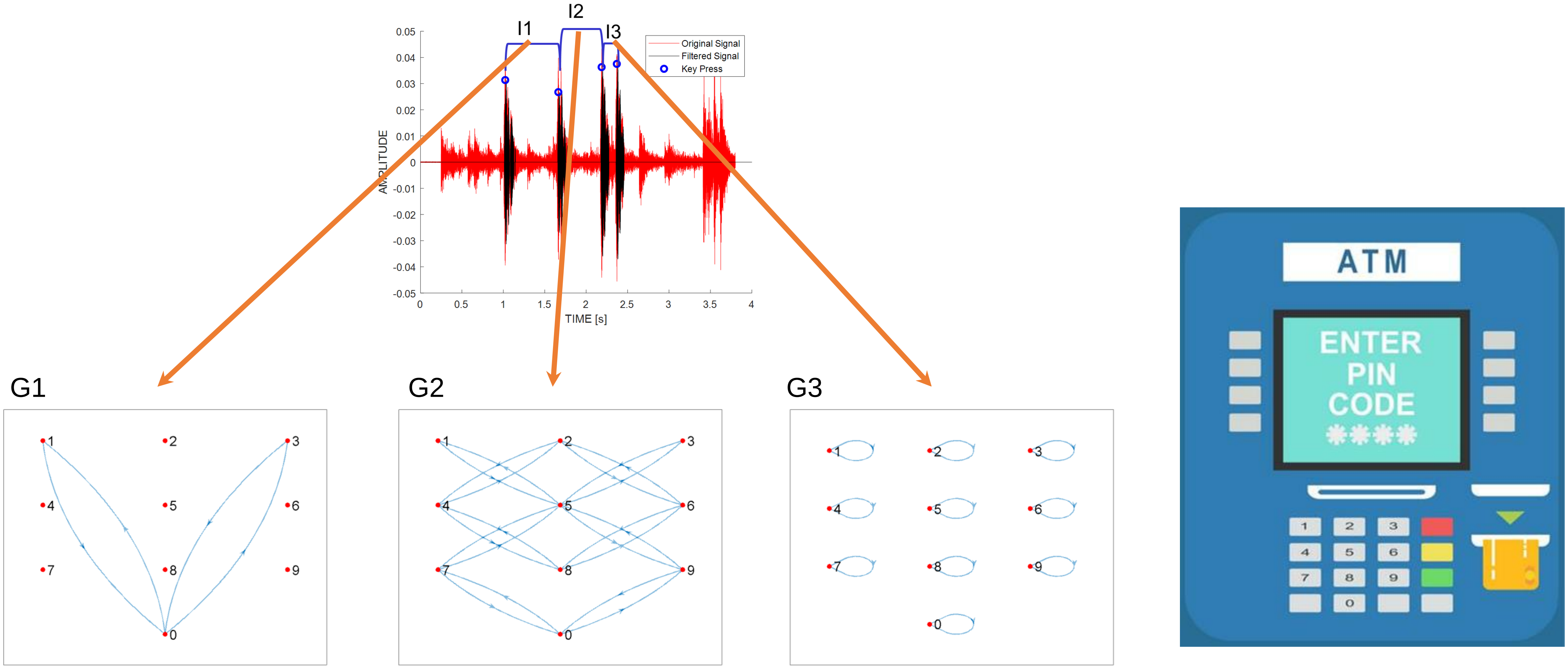
G1


1
2
3
4
5
6
7
8
9
0


ATM
ENTER
PIN
CODE
1
2
3
4
5
6
7
8
9
0

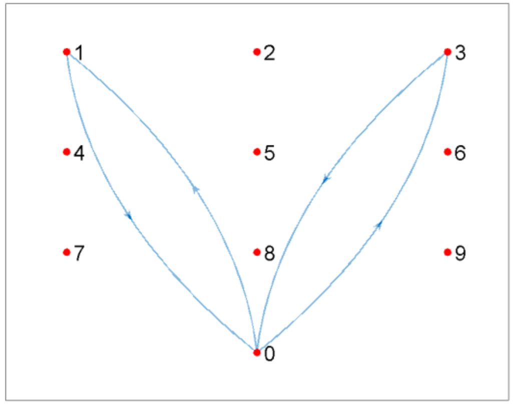

I1
I2
I3
Original Signal
Filtered Signal
Key Press
AMPLITUDE
TIME [s]
0.05
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
-0.05
0 0.5 1 1.5 2 2.5 3 3.5 4
G1
G2
1 2 3
4 5 6
7 8 9
0


ATM
ENTER PIN CODE
****
1 2 3
4 5 6
7 8 9
0

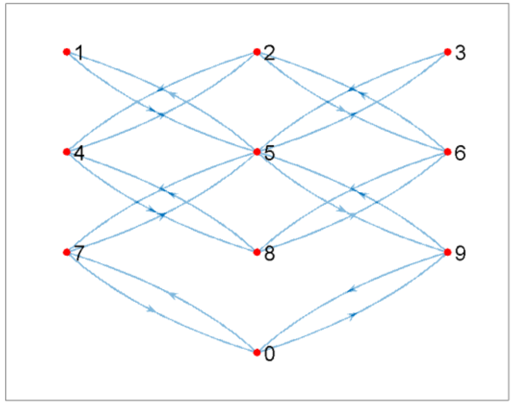

I1
I2
I3
Original Signal
Filtered Signal
Key Press
AMPLITUDE
TIME [s]
G1
G2
G3
ATM
ENTER PIN CODE
****

The subgraphs obtained are processed by an algorithm that identifies **all possible patterns**

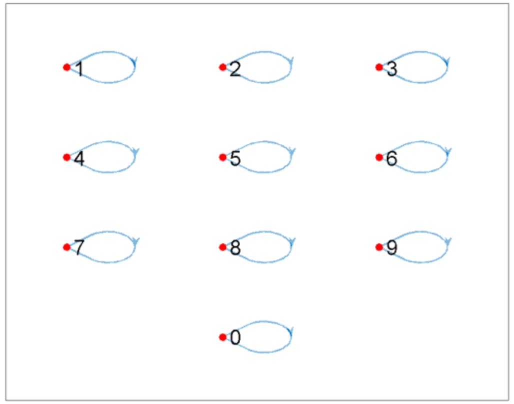
G1



G2

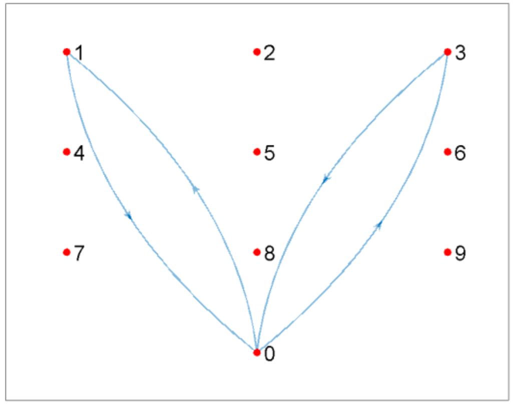


G3

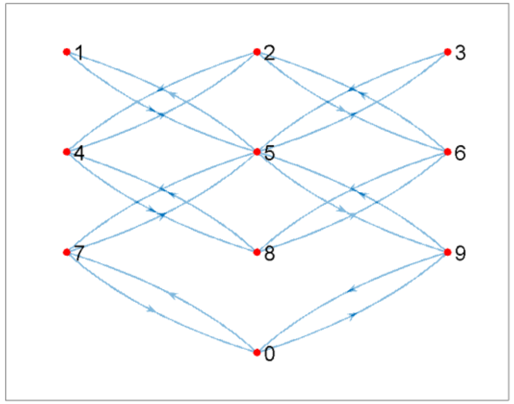


From G1 we know that the first two digits will be 1-0, 0-1, 0-3 or 3-0

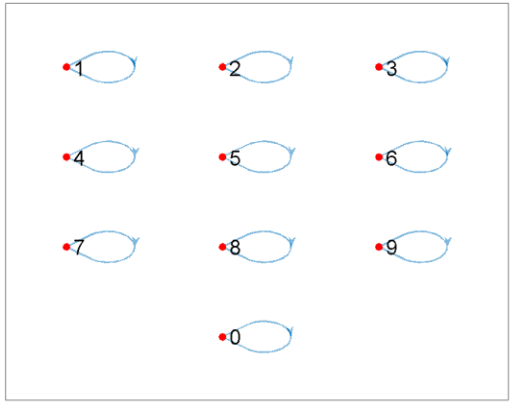
G1



G2

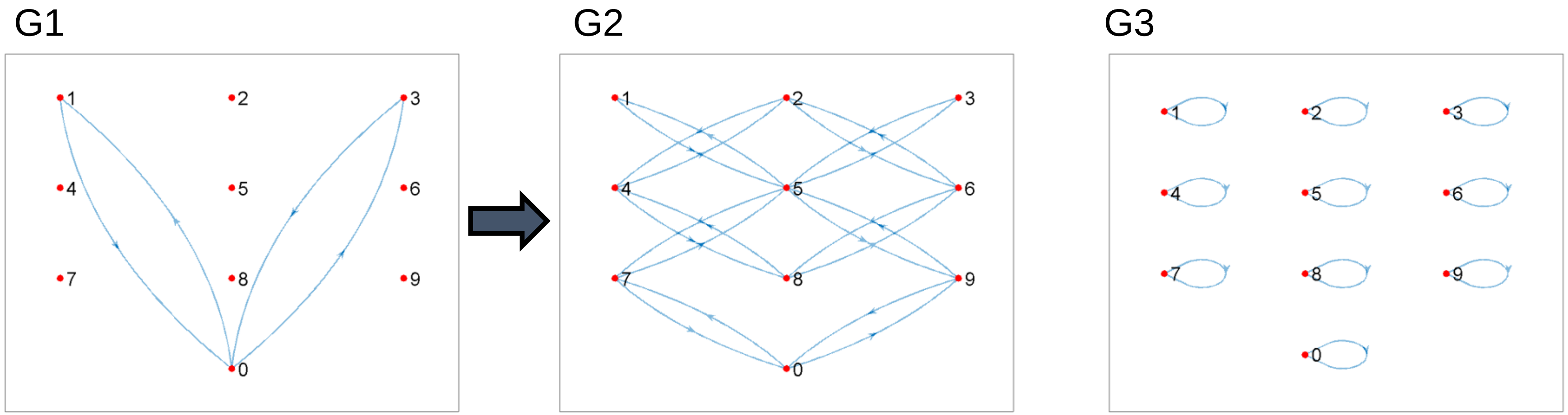


G3



From **G1** we know that the **first two digits** will be **1-0, 0-1, 0-3 or 3-0**

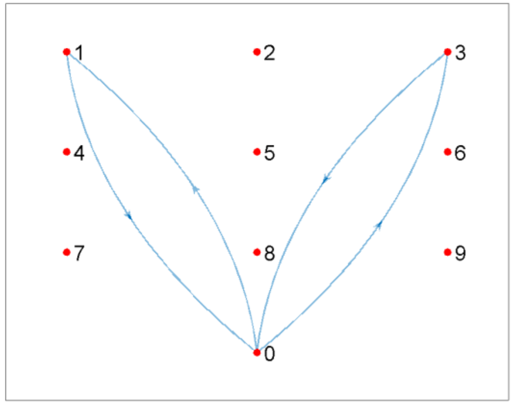
Merging this information with that provided by **G2**, we can **exclude some combinations** like 1-0-5 or 0-3-6



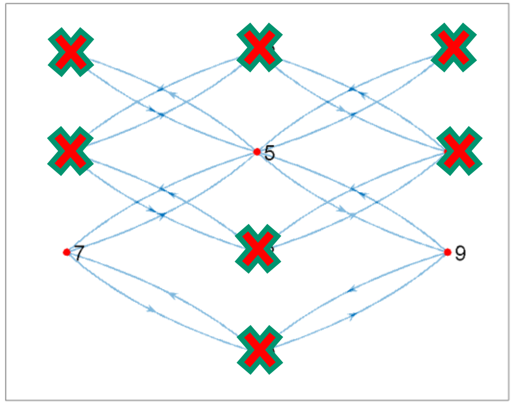
From G1 we know that the first two digits will be 1-0, 0-1, 0-3 or 3-0

Merging this information with that provided by G2, we can exclude some combinations like 1-0-5 or 0-3-6

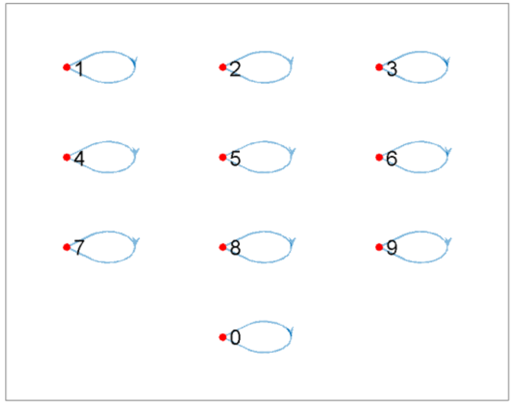
G1



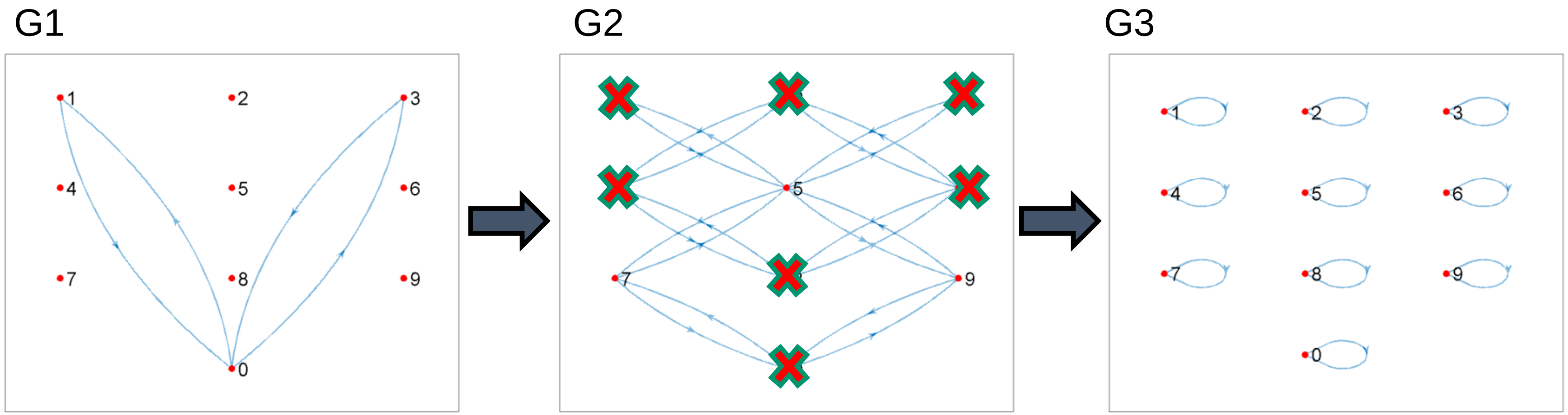
G2



G3

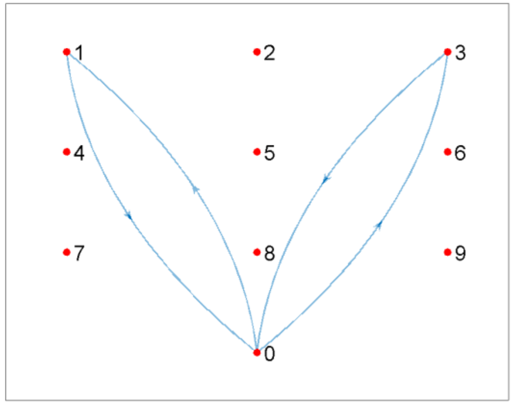


Similarly, the same process is done by combining the information derived from G1 and G2 with that provided by G3

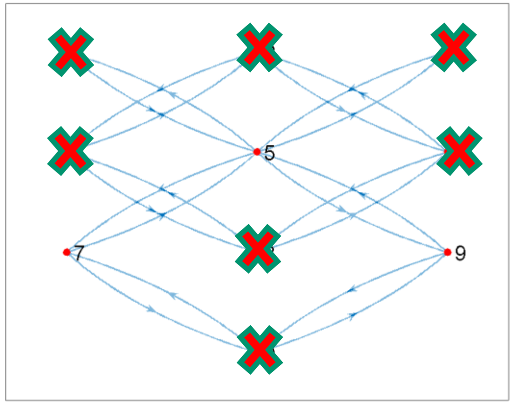


Similarly, the same process is done by combining the information derived from G1 and G2 with that provided by G3

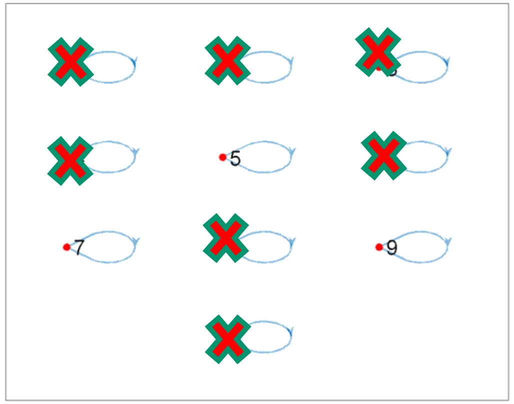
G1



G2

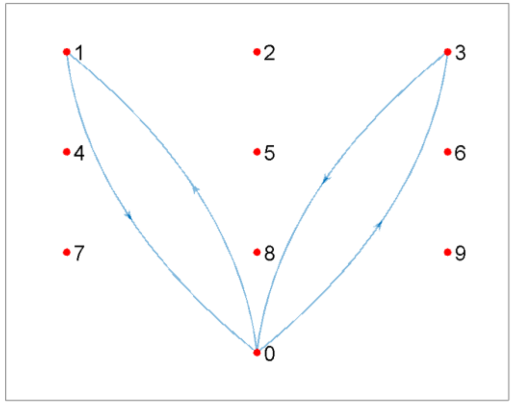


G3

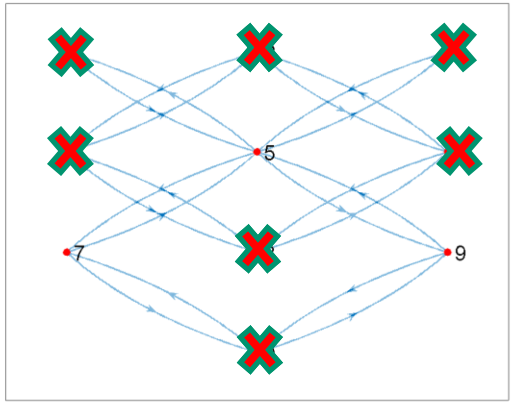


The possible combinations identified by the algorithm are: **0155, 0355, 1077, 3077, 1099, 3099**

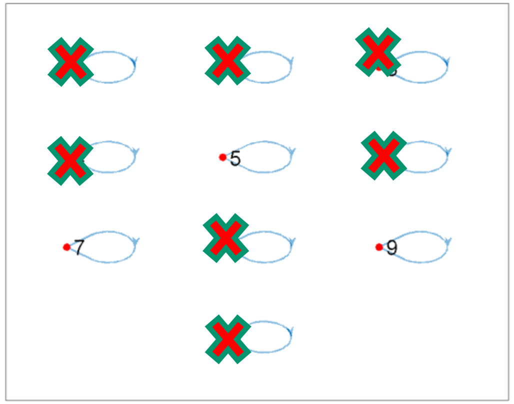
G1



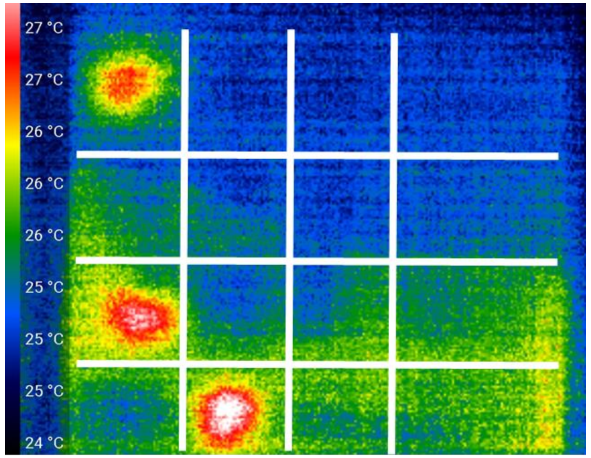
G2



G3

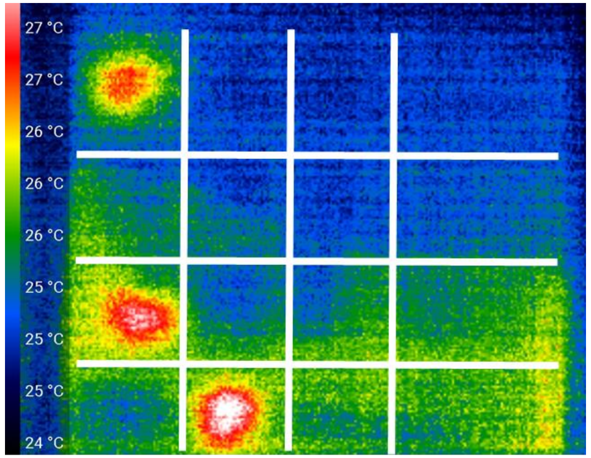


The possible combinations identified by the algorithm are: 0155, 0355, 1077, 3077, 1099, 3099



Analyzing the thermal trace we know that the user has entered the numbers 1,7,0 ... But not the order

The possible combinations identified by the algorithm are: 0155, 0355, 1077, 3077, 1099, 3099

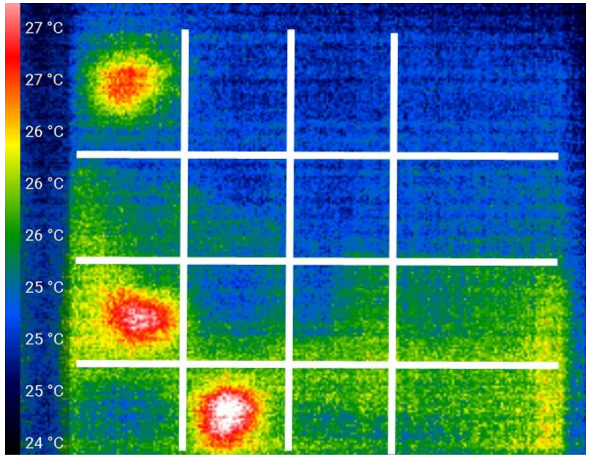


Analyzing the thermal trace we know that the user has entered the numbers 1,7,0 ... But not the order

We note that there is only one PIN that satisfies all the features

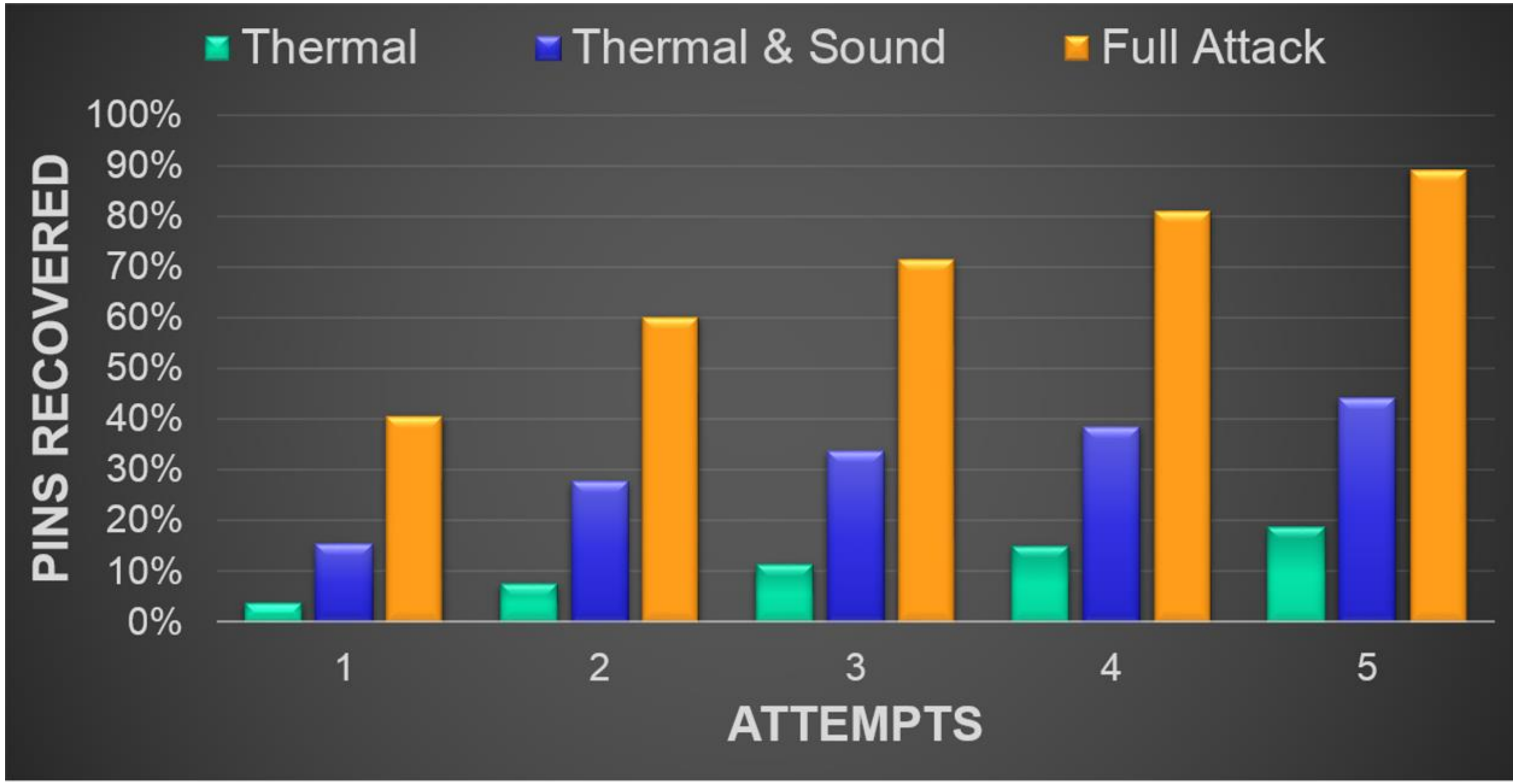
PIN Salabim

The possible combinations identified by the algorithm are: 0155, 0355, **1077**, 3077, 1099, 3099

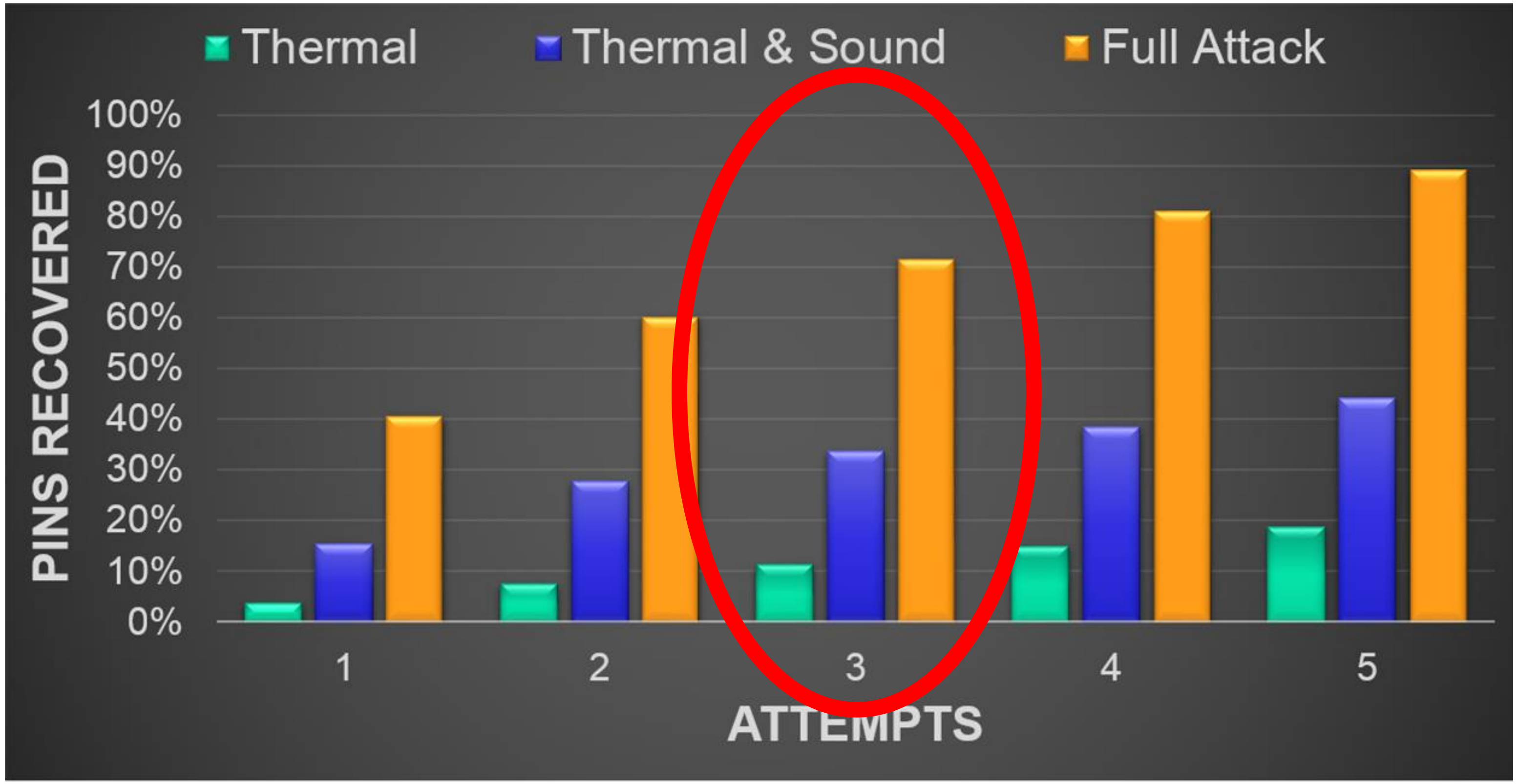


Analyzing the thermal trace we know that the user has entered the numbers 1,7,0 ... But not the order

We note that there is only one PIN that satisfies all the features


Thermal
Thermal & Sound
Full Attack
PINS RECOVERED
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1
2
3
4
5
ATTEMPTS


ATM
ENTER PIN CODE
****


Thermal
Thermal & Sound
Full Attack
PINS RECOVERED
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1
2
3
4
5
ATTEMPTS


ATM
ENTER PIN CODE
****

UNIVERSITÀ DEGLI STUDI DI PADOVA

SPRITZ SECURITY & PRIVACY RESEARCH GROUP

GFT

Thank you!

Matteo Cardaioli

Brain, Mind and Computer Science
University of Padua, Italy
matteo.cardaioli@phd.unipd.it

Topics:

Authentication, Side-Channel, Behavioral