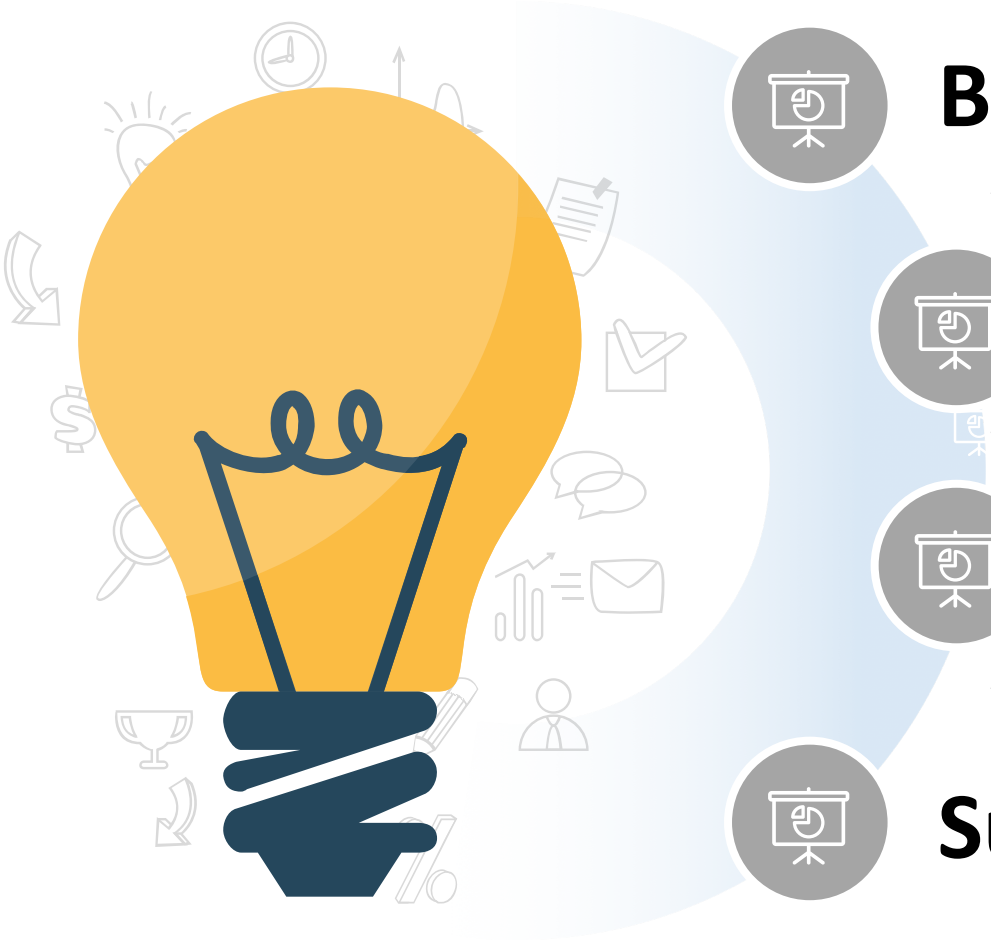
Finding the Stars in the Fireworks: Deep Understanding of Motion Sensor Fingerprint

Huiqi Liu, Xiang-Yang Li, Lan Zhang, Yaochen Xie, Zhenan Wu, Qian Dai, Ge Chen, Chunxiao Wan

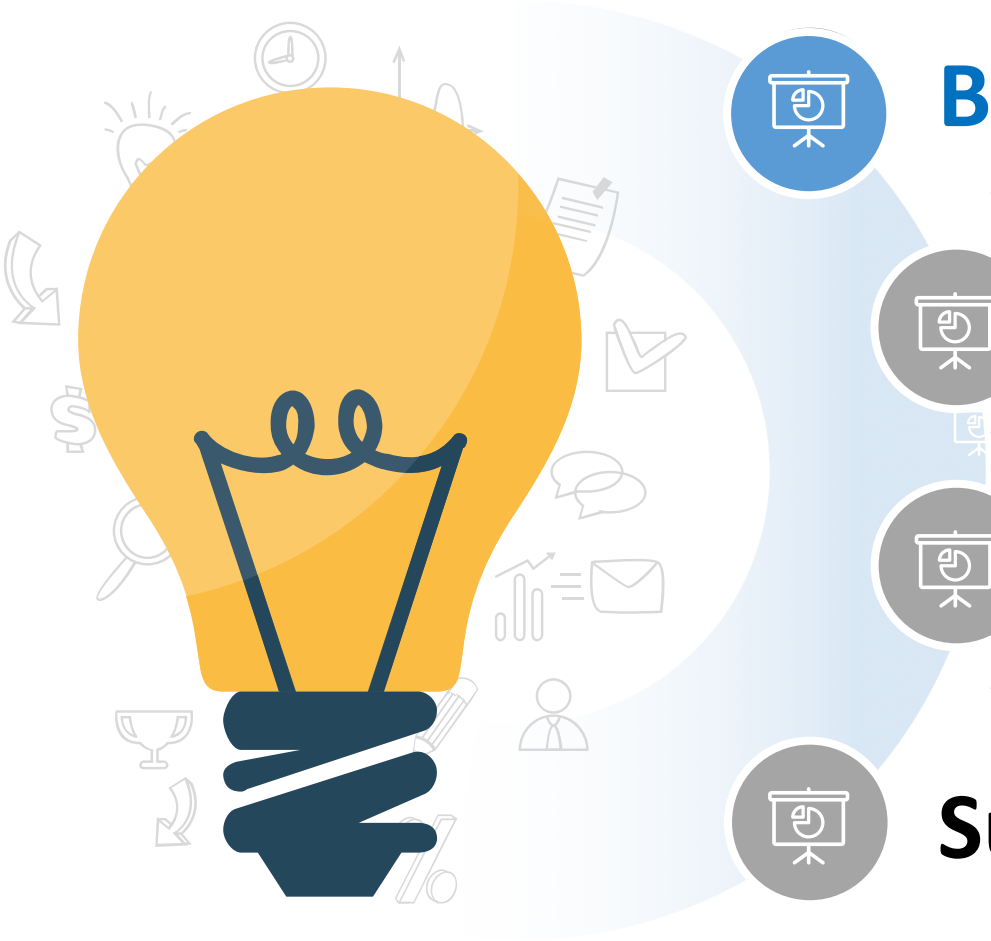
University of Science and Technology of China (USTC), Tencent

Outline



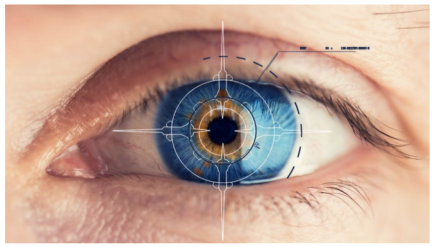
- **Background**
- **Fingerprinting Capacity Model**
- **Fingerprinting & De-fingerprinting**
- **Summary**

Outline



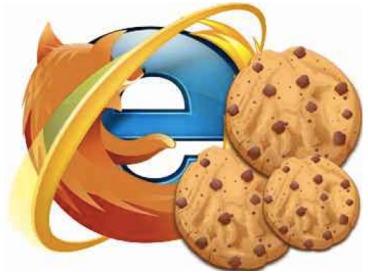
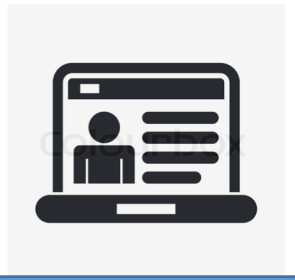
- **Background**
- **Fingerprinting Capacity Model**
- **Fingerprinting & De-fingerprinting**
- **Summary**

Fingerprints Everywhere

Human Fingerprints

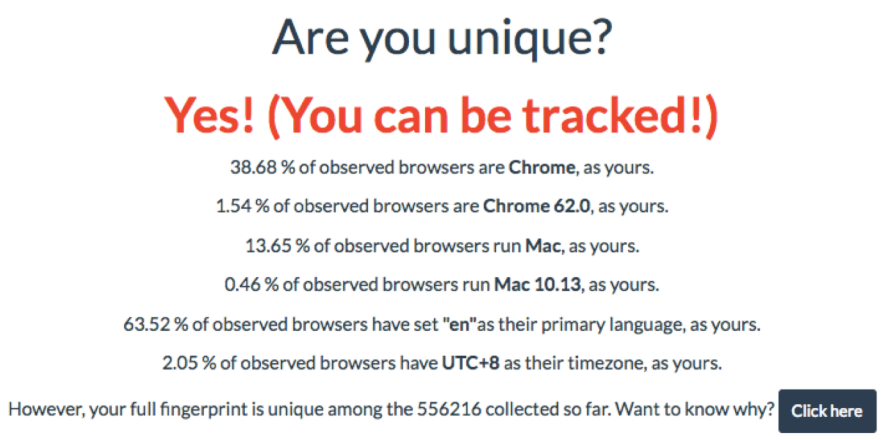
Digital World Fingerprints

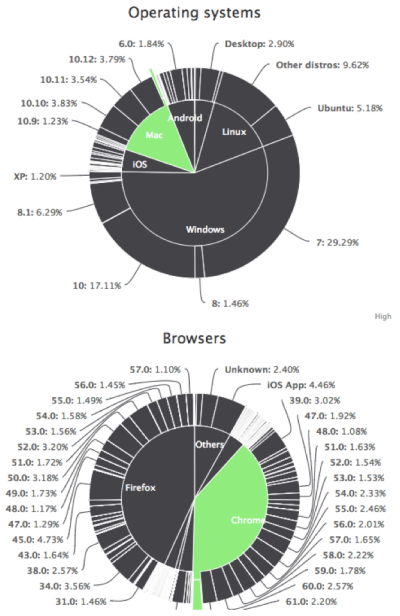



How are they tracking devices?

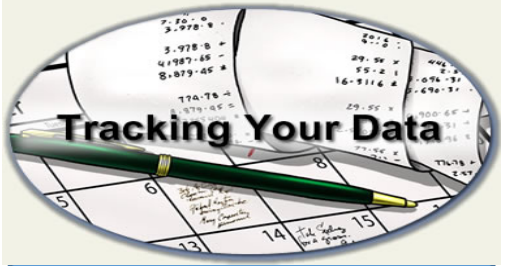
Device Fingerprint: An Example

- Tracking exists in the real world!





Data Tracking

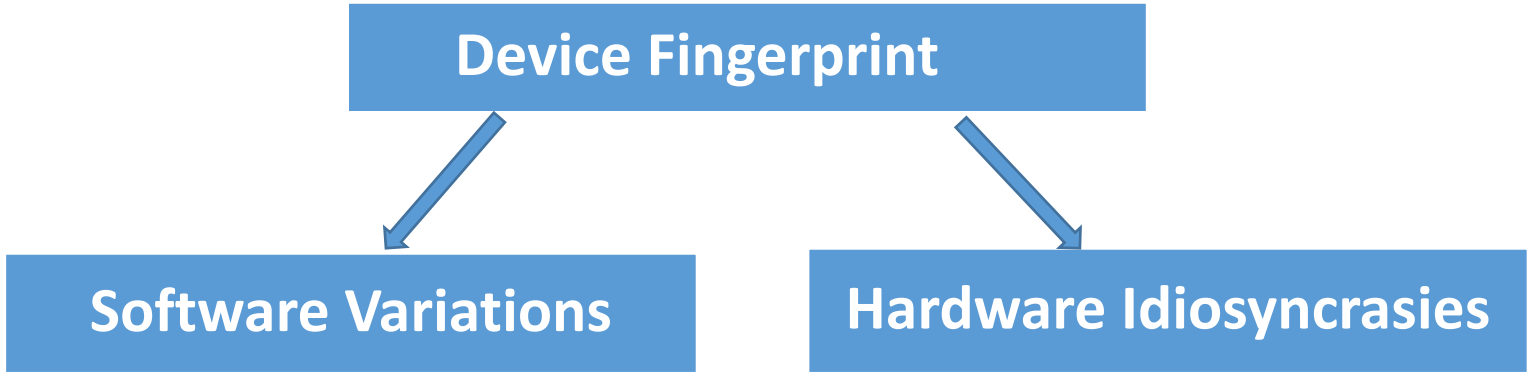


User Tracking

Privacy Leakage

Device Fingerprints Techniques

- How are device fingerprints generated?



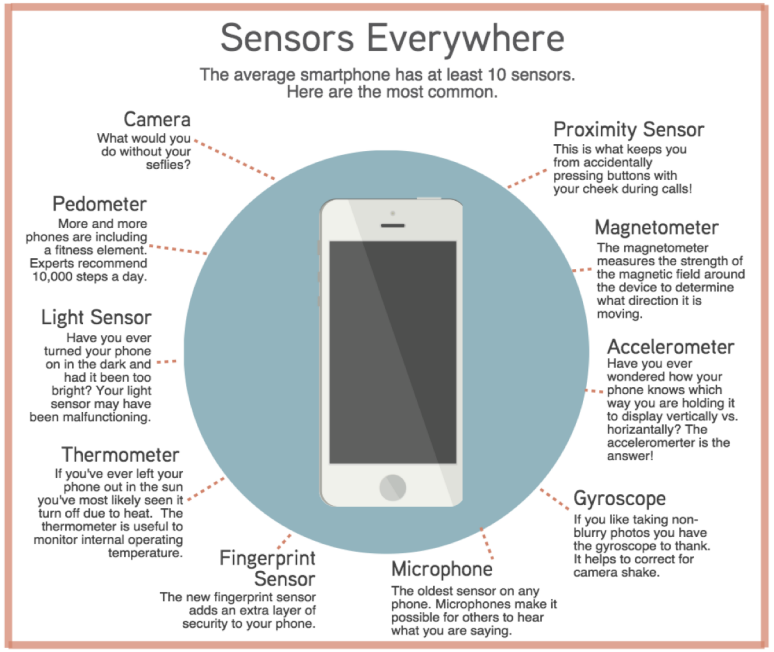
- Difference in Protocol Stack/Network Stack
- Difference in Firmware and Device Driver
- Difference in installed Software
- MAC Headers

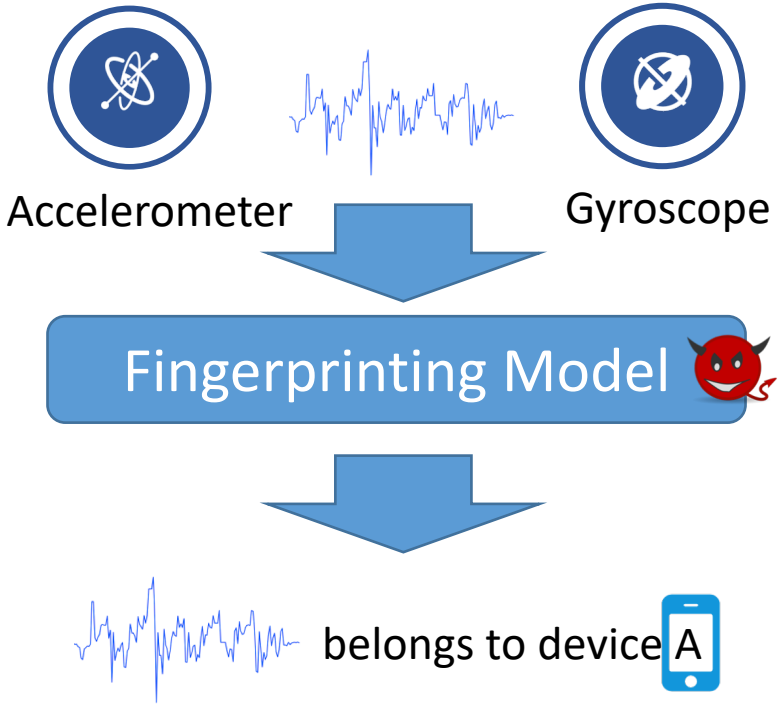
- Difference in spectral property of Radio Signal Transmitters
- Difference in emitted radio frequency of NIC
- Unique and constant clock skews in network devices

We exploit small deviations in mobile devices from **hardware**.

How are Mobile Devices Different?

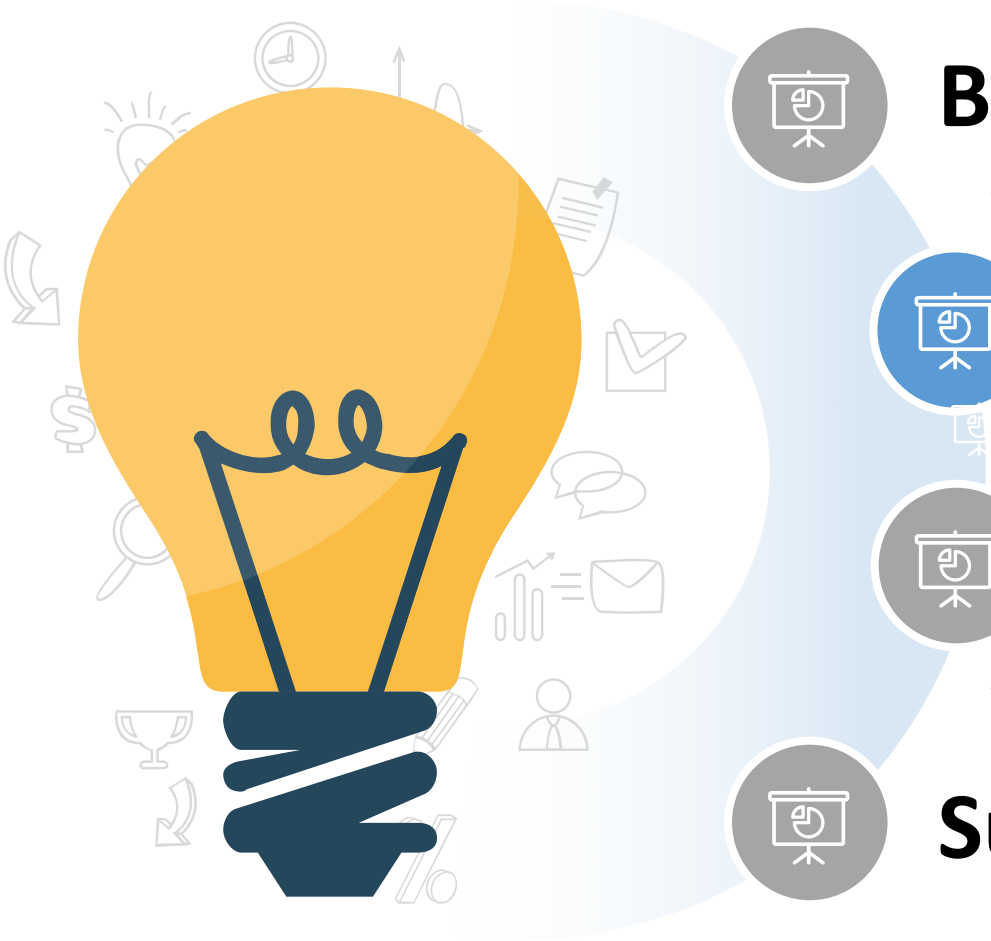
- Smartphones are equipped with a wide range of sensors.





We focus on motion sensors to generate fingerprints.

Outline



- Background
- **Fingerprinting Capacity Model**
- Fingerprinting & De-fingerprinting
- Summary

What is Fingerprinting Capacity Model?

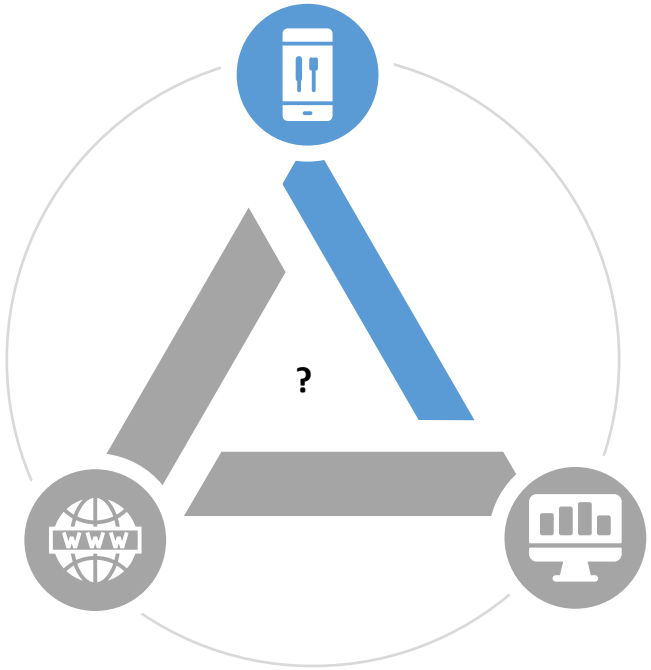
- The capacity means to estimate how many devices can be distinguished by their manufacturing variances.
- It is a theoretical (mathematical) model to estimate the capacity of motion sensor fingerprint.

Key Questions in Devices Fingerprint?

Fingerprinting Capacity

What is the capacity of device fingerprint?

The model, analysis and feature.



Fingerprinting Factors

Which factor infects the fingerprint most?

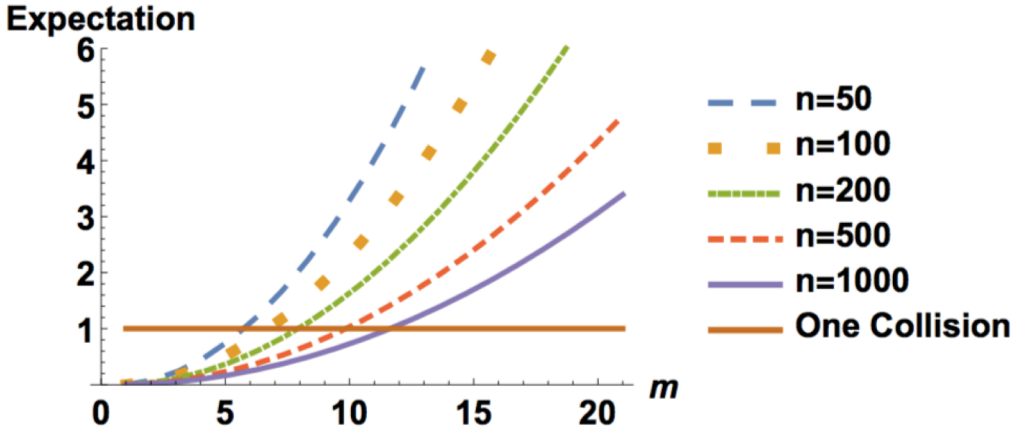
The user activity, device brand or device model.

De-fingerprinting Trade-off

How to anonymize sensor data while retaining utility?

Capacity Model

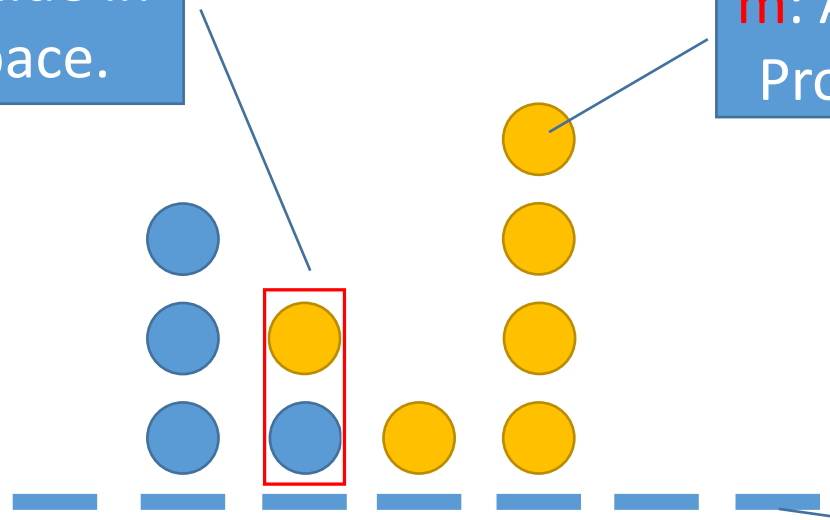
- Balls into bins in one dimension



Two devices collide in one feature space.

m: A Device Which Process a feature

one-dimension

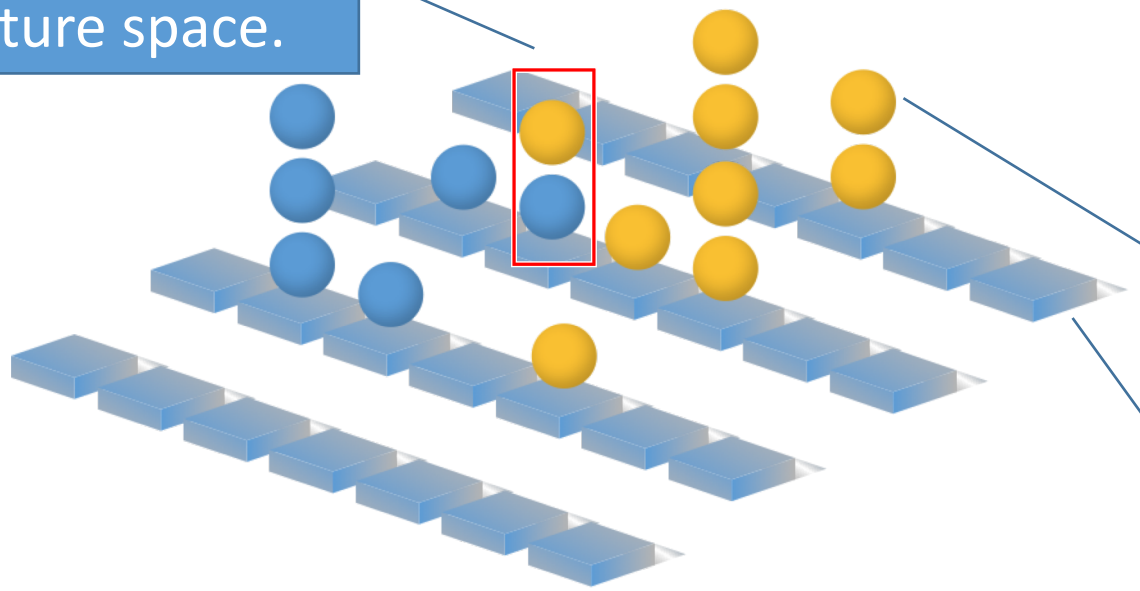


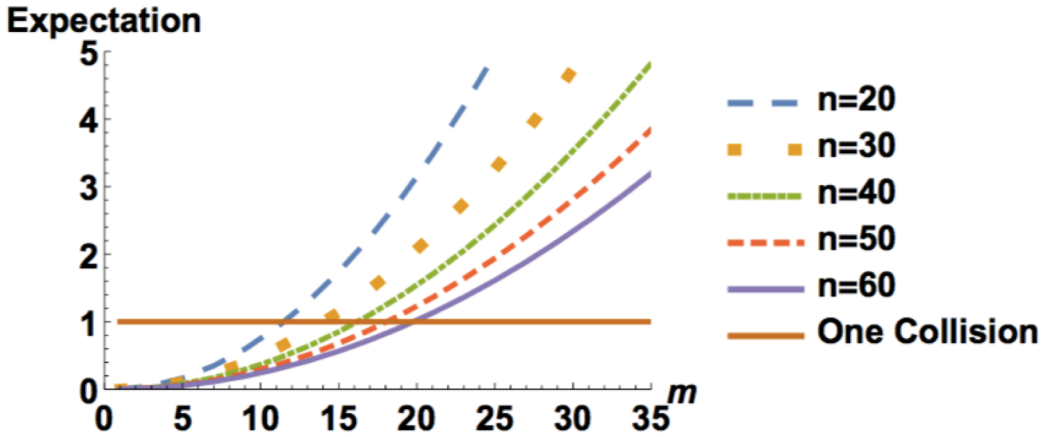
n: Device Feature Space

Capacity Model

- Balls into bins in multi-dimension

Two devices collide in one feature space.





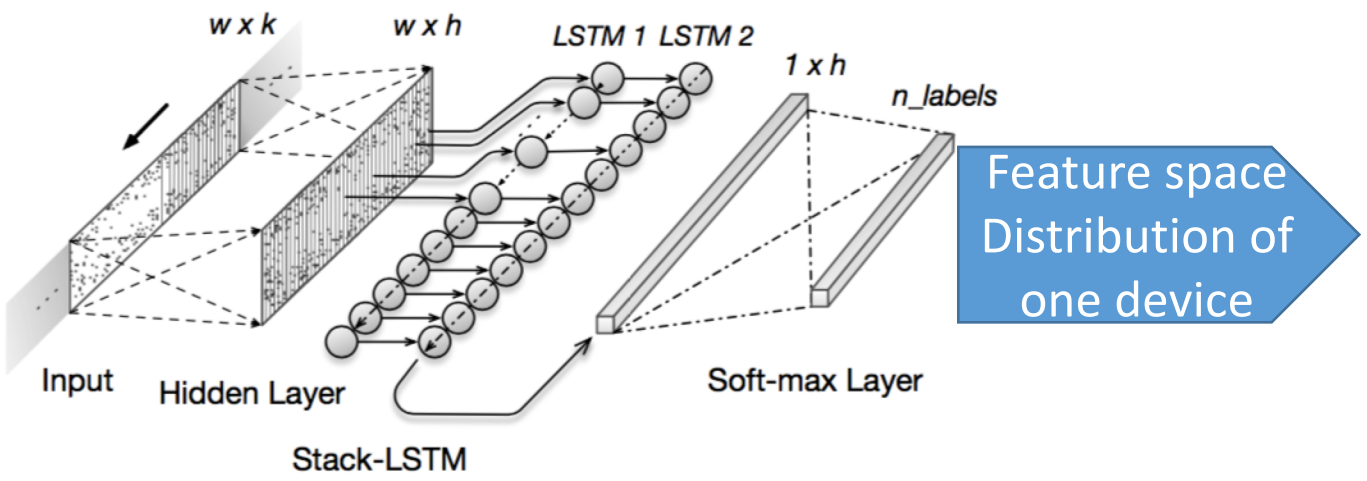
two-dimension

m: A Device Which Process a feature

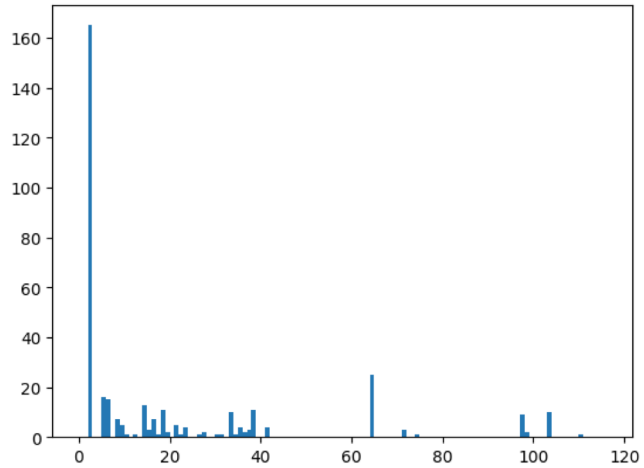
n: Device Feature Space

Capacity Model

- We treat the fingerprinting problem as a classification problem.
  - For ‘bin’, we use the classification layer as the feature space.
  - For ‘ball’, each data piece of a device is treated as a ball.
  - For ‘dimension’, two sensors (accelerometer and gyroscope) can be treated as independent dimensions for device fingerprint.

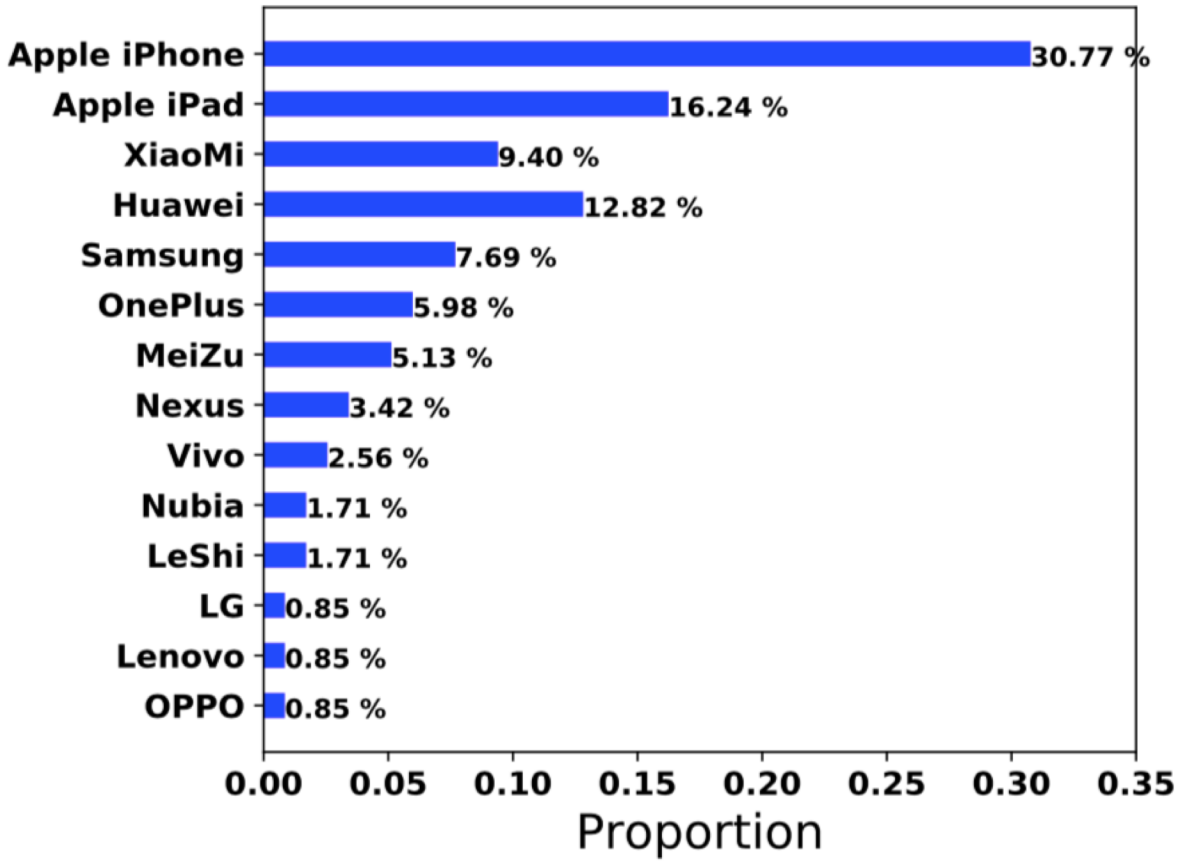


Feature space Distribution of one device



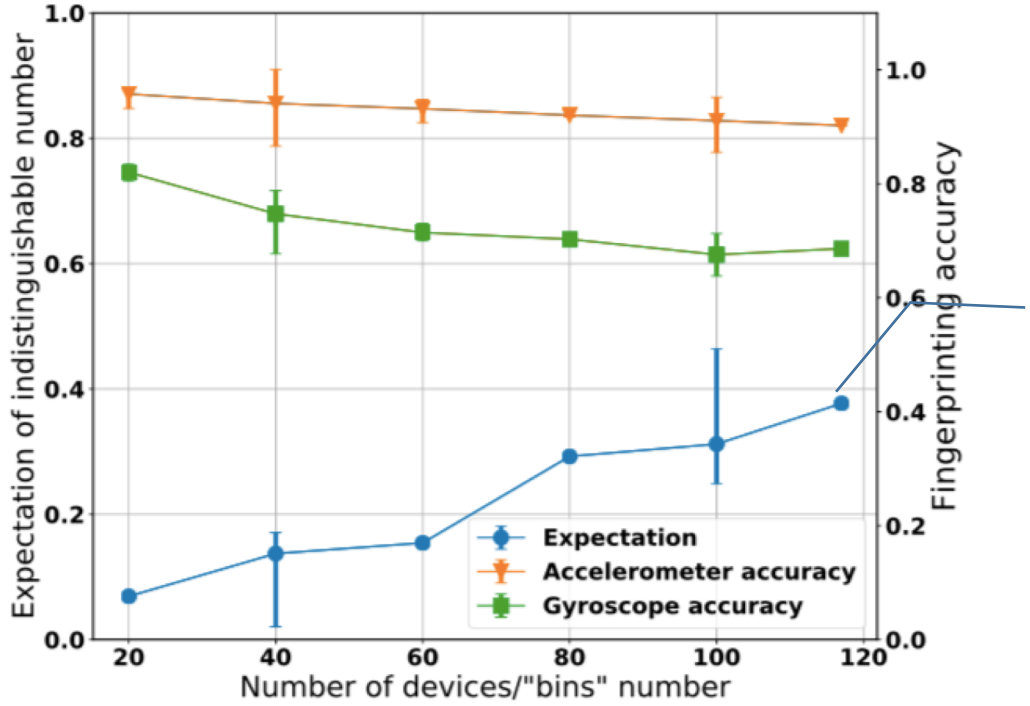
Dataset

- With users' permission, we collect motion sensor (accelerometer and gyroscope) data from **total 117 mobile phones** with 13 different brands.



Capacity Model to Validate the NN Result

The expectation of collided devices are increasing when the number of devices increases in our experiments.

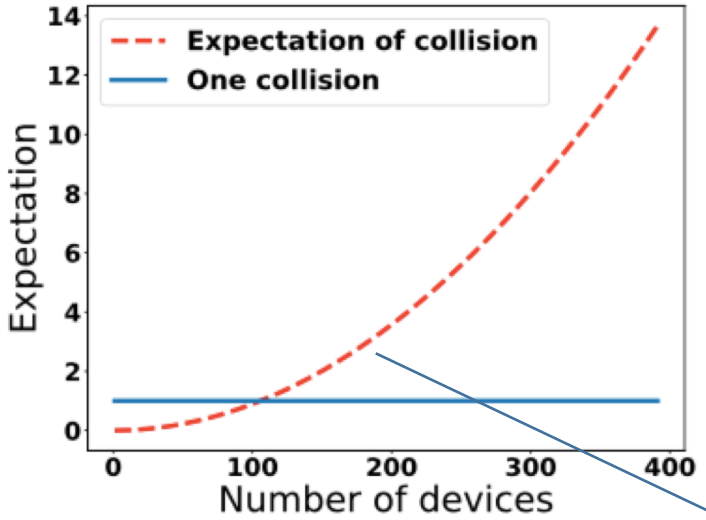


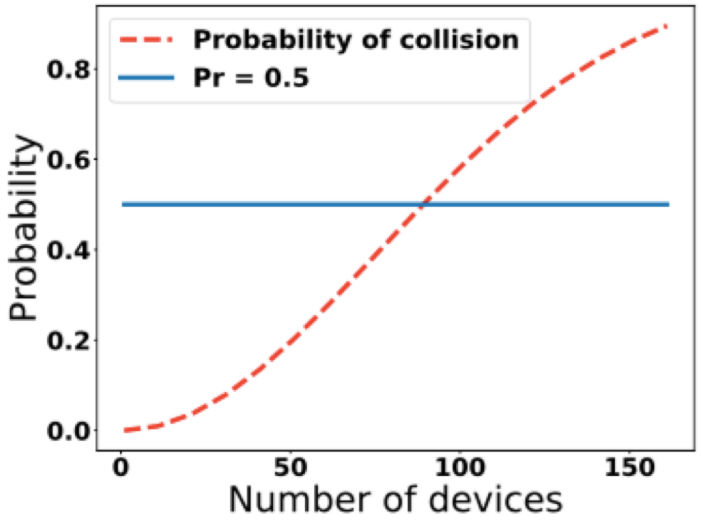
The expectation=0.38 when the device number=117

This is **consistent** with our experiment result that for 117 devices given 20 seconds data of each devices the **fingerprinting accuracy is 99%.**

Capacity Model to Predict Large-Scale Result

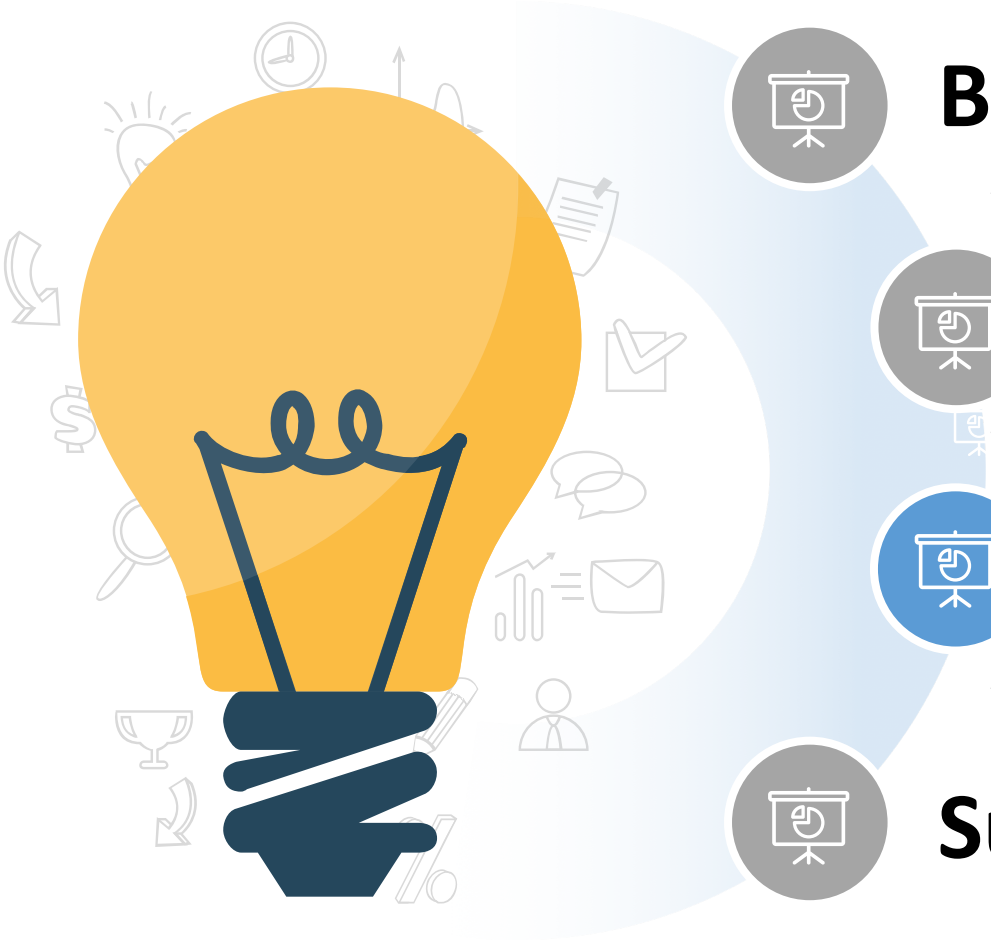
The expectation and the probability of indistinguishable devices.





For more than 200 devices, there is expected to be at least two collisions.

Outline

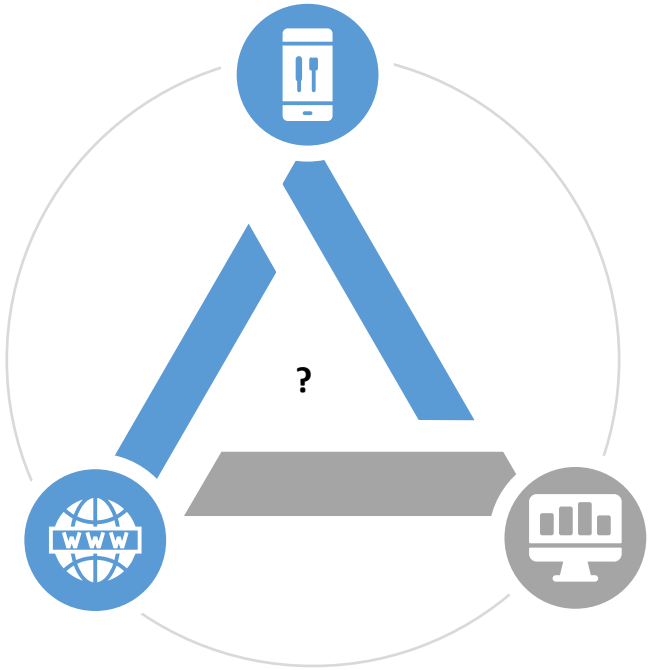


- Background
- Fingerprinting Capacity Model
- **Fingerprinting & De-fingerprinting**
- Summary

Key Questions in Devices Fingerprint?

Fingerprinting Capacity

What is the capacity of device fingerprint? The model, analysis and feature.



Fingerprinting Factors

Which factor infects the device fingerprinting?

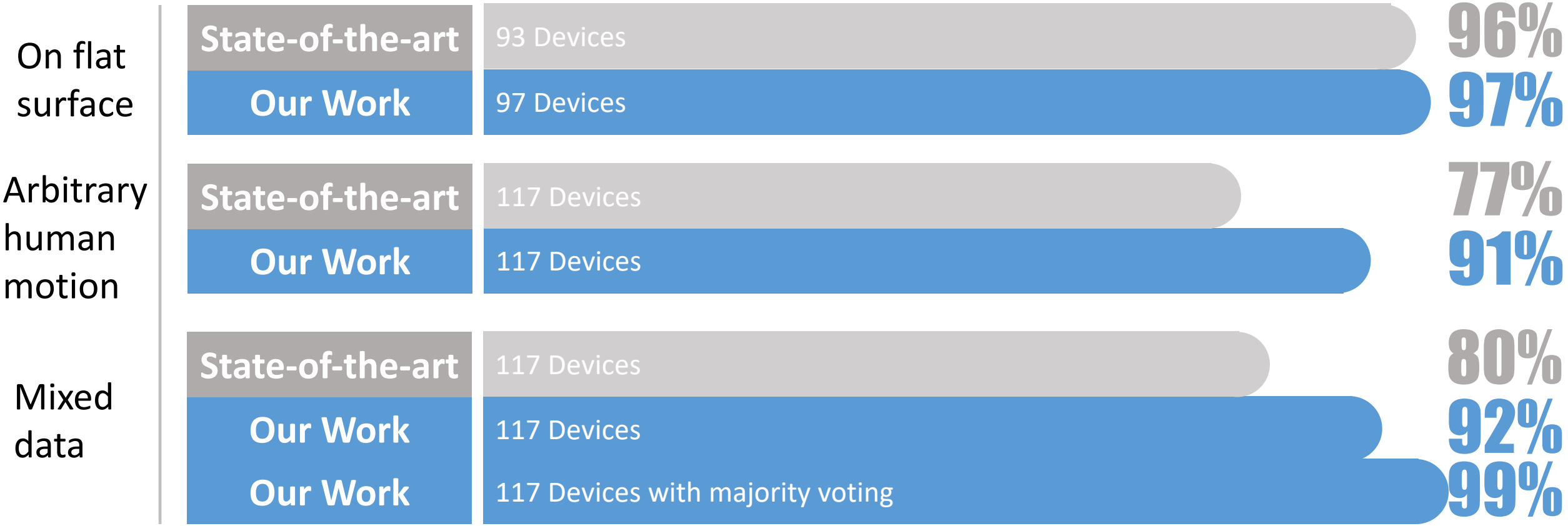
The user activity, device brand or device model.

De-fingerprinting Trade-off

How to anonymize sensor data while retaining utility?

Fingerprinting Analysis

- Influence of Static vs. Dynamic



Fingerprinting Analysis Cont'd

- Sensors and axes
  - First, we conduct LSTM-based fingerprinting on each sensor's data separately.
  - Each sensor has three axes, we fingerprint three axes of each sensor separately.

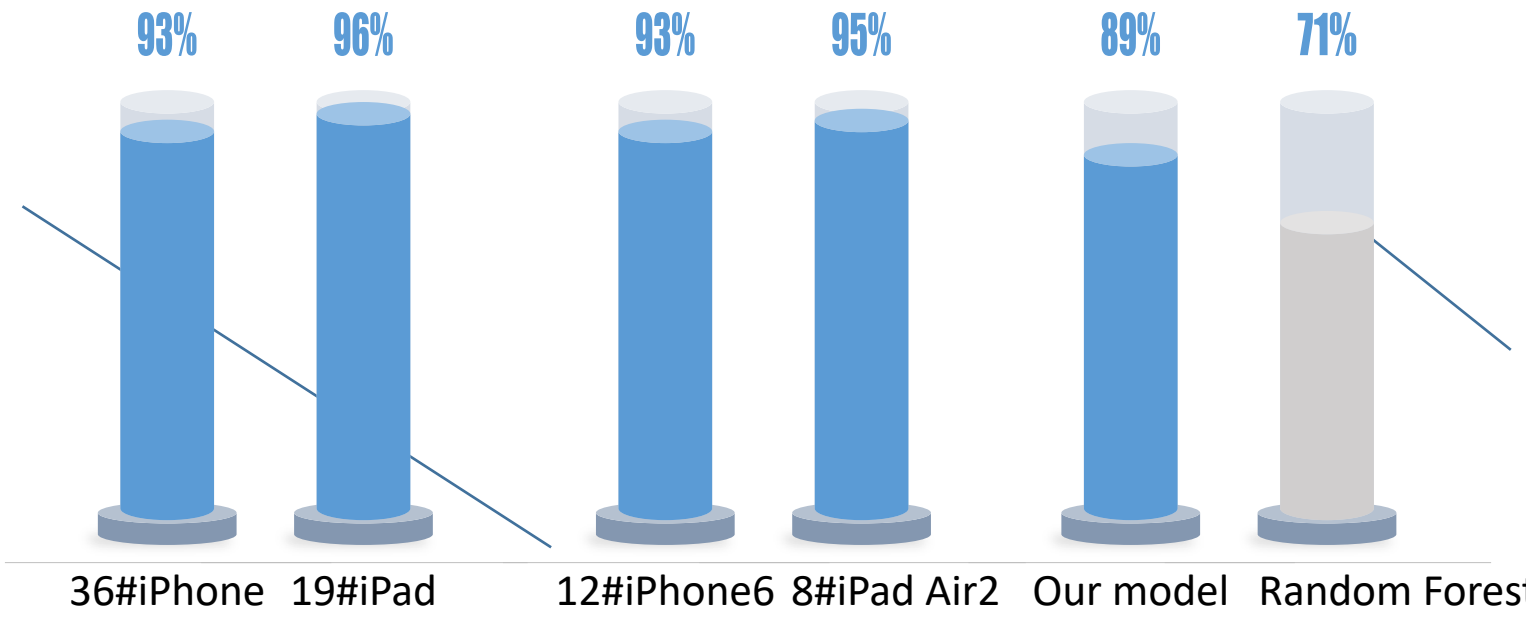
/%	Accelerometer	Gyroscope	Accelerometer			Gyroscope		
Axis	none	none	ax	ay	az	ax	ay	az
Accuracy	90.26	68.62	72.56	72.96	43.57	52.92	54.02	53.85
Fusing	91.41		87.01					

Implies there is more correlation among axes.

Fingerprinting Analysis Cont'd

- Influence of brands, models and human

The device distinguishability is slightly different among different brands or models.



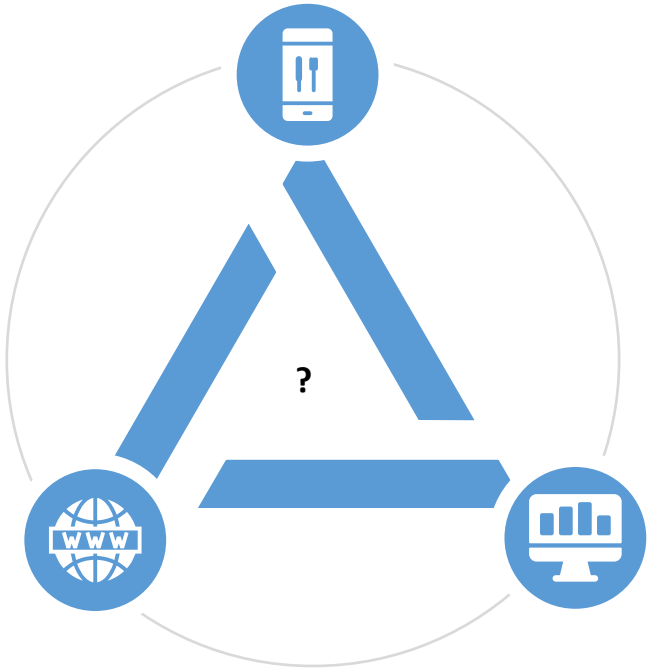
Granularity	One Brand, Different Devices	One Model, Different Devices	One User, 12 Different Devices

the human factor indeed increases the difficulty device fingerprinting.

Key Questions in Devices Fingerprint?

Fingerprinting Capacity

What is the capacity of device fingerprint? The model, analysis and feature.



Fingerprinting Factors

Which factor infects the fingerprint most?

The user activity, device brand or device model.

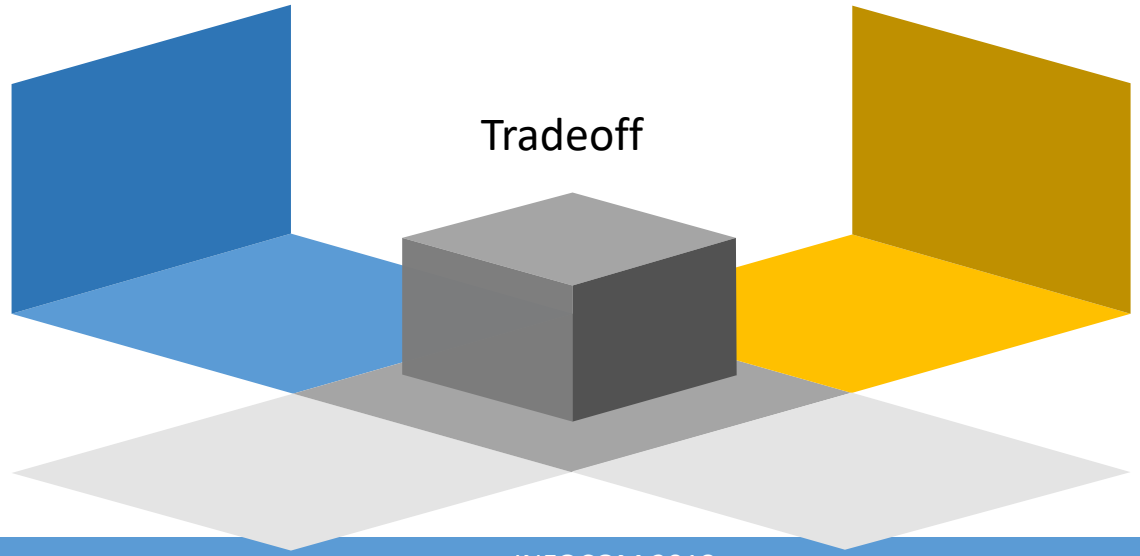
De-fingerprinting Trade-off

How to **anonymize** sensor data while retaining **utility**?

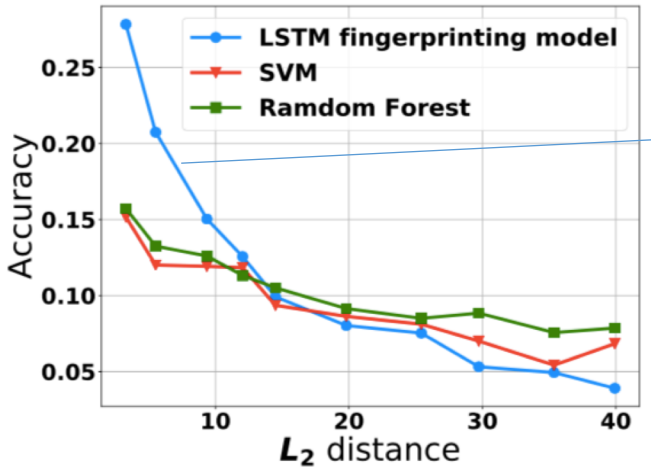
De-fingerprinting (anonymization)

Anonymization effect
- Fingerprinting model result

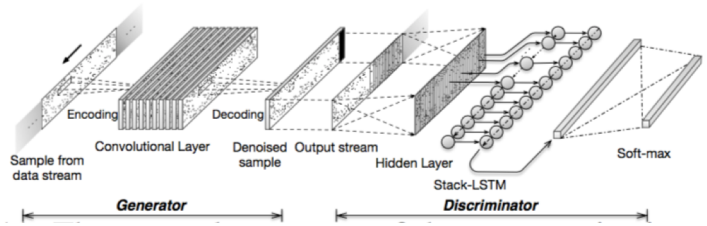
Data utility
- L2 distance
- Step counter result



De-fingerprinting (anonymization) Analysis

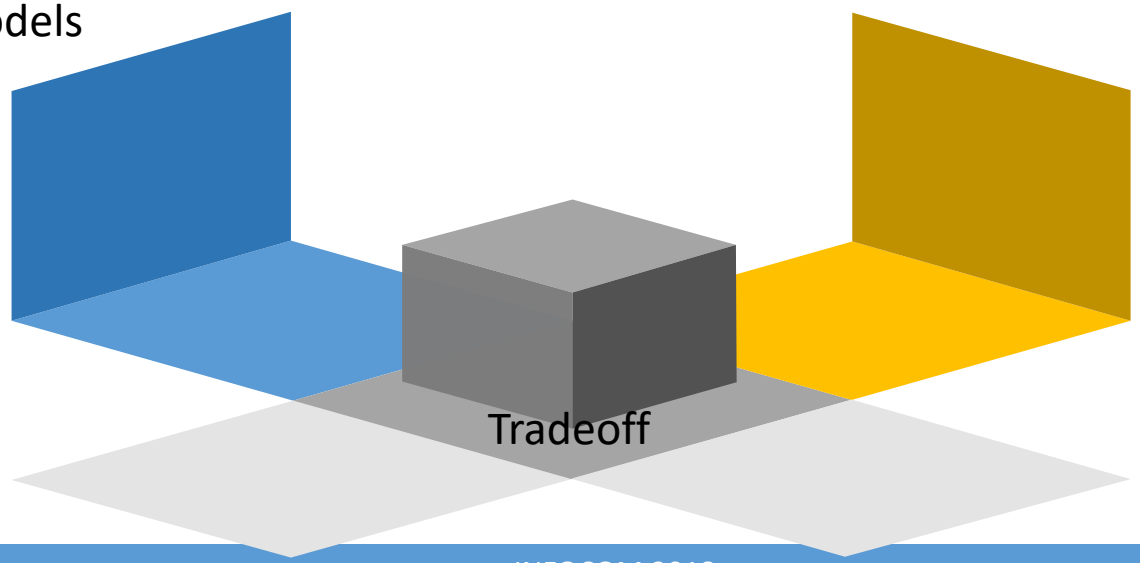


Anonymization effect measured by different fingerprinting models

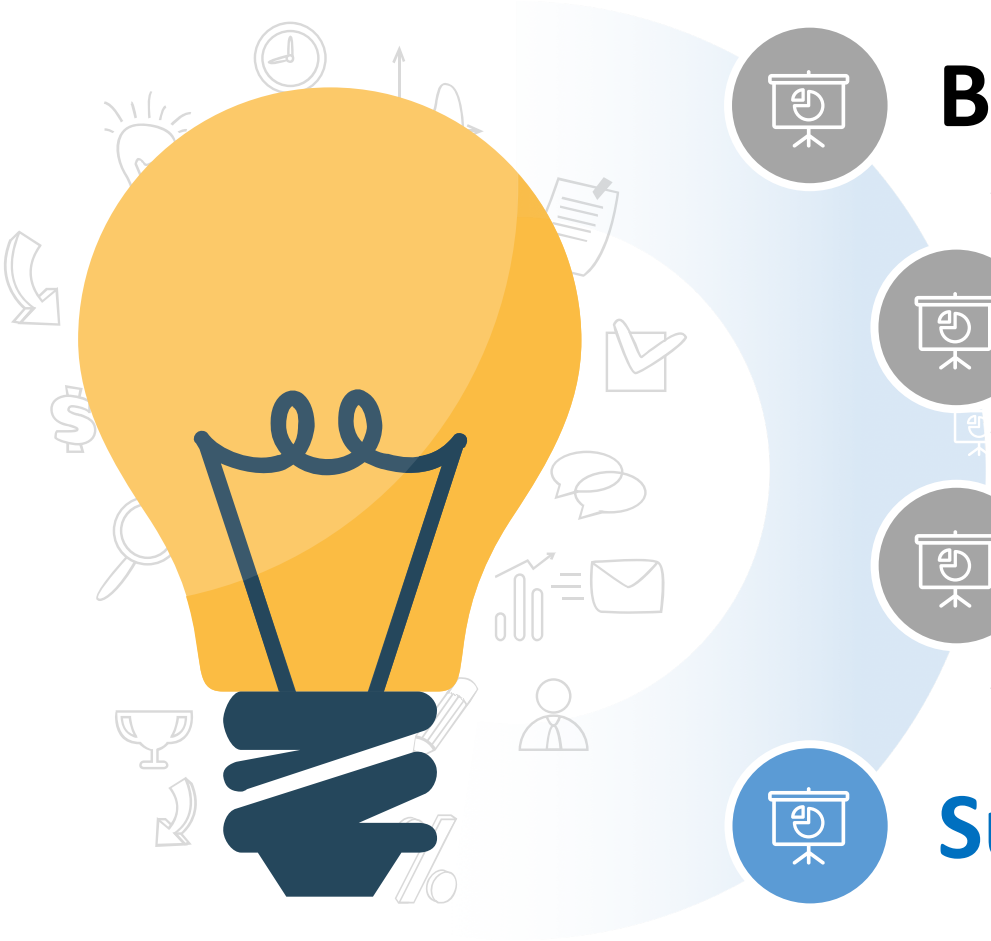


Ours model	L2(≈)	6
	Mean	30
	Stdev	0
Uniform Noise	L2(≈)	6
	Mean	29.86
	Stdev	0.51

Step Counter Result



Outline



- Background
- Fingerprinting Capacity Model
- Fingerprinting & De-fingerprinting
- Summary

Conclusions & Future Work

- We propose a **theoretical model** to understand the capacity of fingerprinting, it is a primary work and can be also used in other scenarios.
- We design a deep neural network based model to fingerprint mobile device sensors in real-life uses.
- We propose a novel generative model to anonymize sensor data while retaining good data utility, but it is still needed to deeply investigate that **how vulnerable are the de-fingerprinting models** against different types of fingerprinting attacks.

Datasets

- Our dataset is available, PLEASE feel free to DOWNLOAD it for fingerprinting research.
- Link: https://drive.google.com/open?id=14eYWdB-77NMUCui4MZQPxpbjwZeNLi94

Introduction

For each motion sensor, i.e., accelerometer or gyroscope, three data sequences are simultaneously generated with with timestamps by three axes. So, in our experiments, we obtain 6 data sequences from two motion sensors. Each sequence can be a channel of the neural network input. However, they are generated with unstable time intervals, which depends on the schedule of the mobile operating system according to the real-time system load. Hence, we conduct piece-wise cubic Hermite interpolation to obtain equally spaced data sequences as the inputs of neural networks. We divide continuous sensor data into 0.5~20-second segments as our dataset.

Data format

Recommanded method to load the dataset:

```
with gzip.open(filename, 'rb') as f:
    dataset = pickle.load(f)
    data = dataset['data']
    label = dataset['label']
```

Each file follows the format:

```
# dataset
{
    # data, shape (number of pieces, number of sample points, 3 axes, 2 sensors)
    'data': []
    # label, shape (number of pieces, )
    'label': []
}
```

Q & A

Thank you for listening~

- Huiqi Liu
- liuhuiqi@mail.ustc.edu.cn
- https://charlesliu7.github.io/