# Why it Takes so Long to Connect to a WiFi Access Point

**Changhua Pei**, Zhi Wang, Youjian Zhao, Zihan Wang, Yuan Meng, Dan Pei

Yuanquan Peng, Wenliang Tang, Xiaodong Qu



INFOCOM 2017

Motivation

Measurement

Results

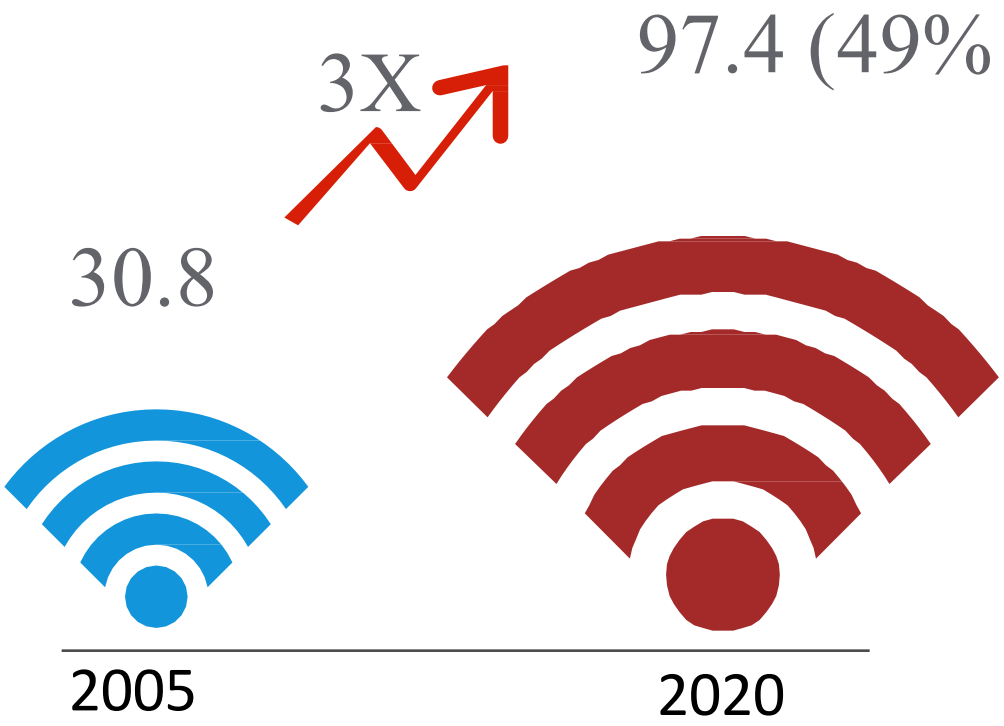
Correlation

Analysis

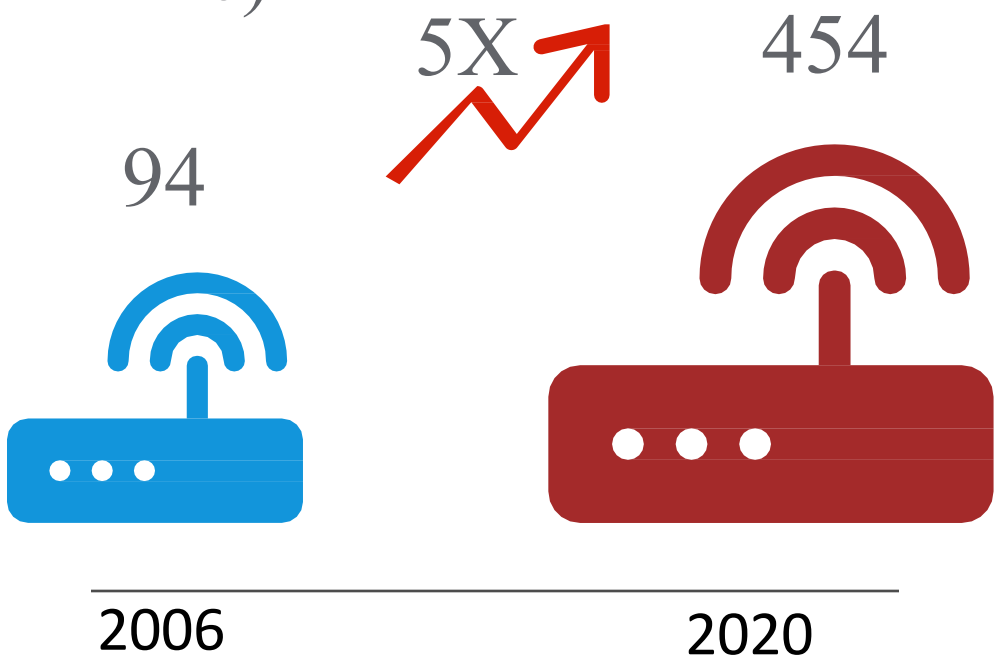
Modeling

# WiFi is indispensable in our daily lives

Source: Cisco VNI Mobile, 2017

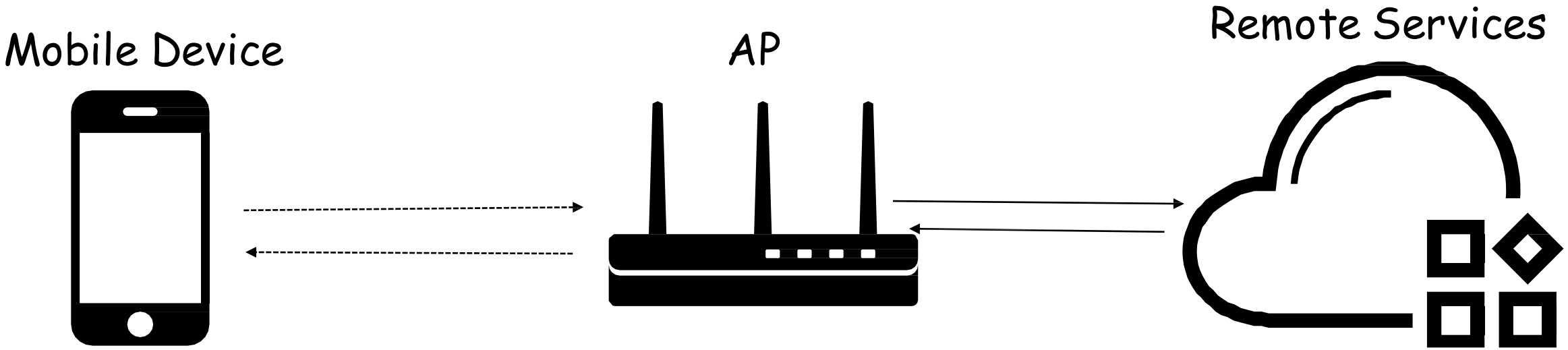


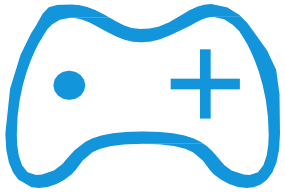

Overall WiFi Traffic Growth (Exabytes)




Hotspots (Millions)

# Experience of WiFi Network



Studies to improve user experience from Bandwidth,lantency….

Online Gaming

# Urgent need to study the connection set-up time

**Suranga [WiNTECH'13] is the first work focus on WiFi connection time cost :**

- The connection set-up process in the wild is unknown
- Lack thorough investigation in a larger scale.

**We focus on:**

- How about the connection time cost in the wild?
- What is the culprit of the high connection time cost?
- What can I do to reduce the connection time cost?

Motivation

Measurement

Results

Correlation

Analysis

Modeling

# DATASET

- WiFi Manager of Tencent Technology
- Provide Free WiFi service
- Top in the Android/iOS App market (China)
- About 50K downloads every day
- Continuously collect one week data from May 3 to May 9.

# DATASET

## Connection Log Dataset

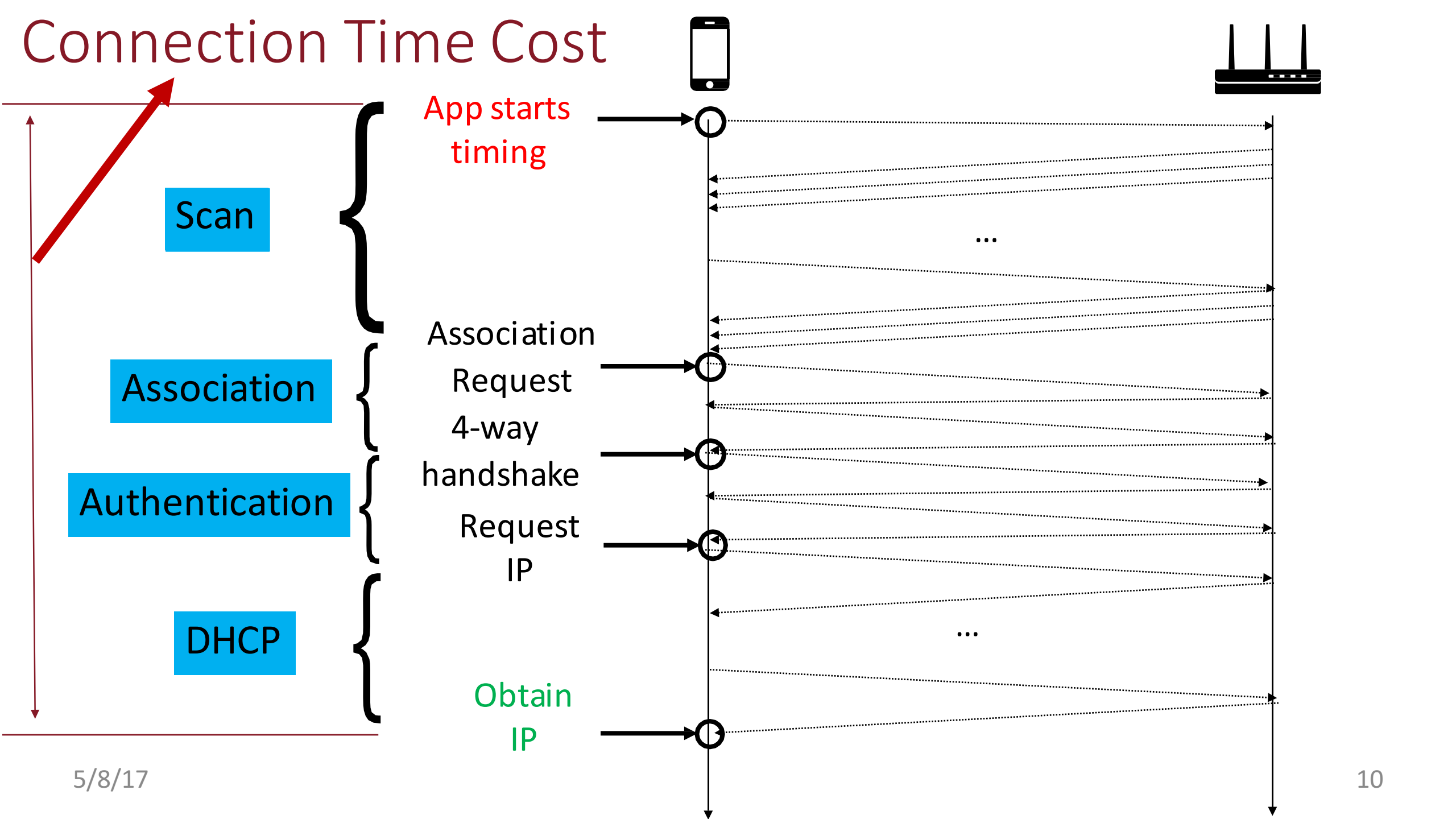
- 7 Million unique APs
- 5 Million unique mobile devices
- 4 different cities.
- 0.4 billion overall connection attempts.

# DATASET

## Breaking Down Dataset

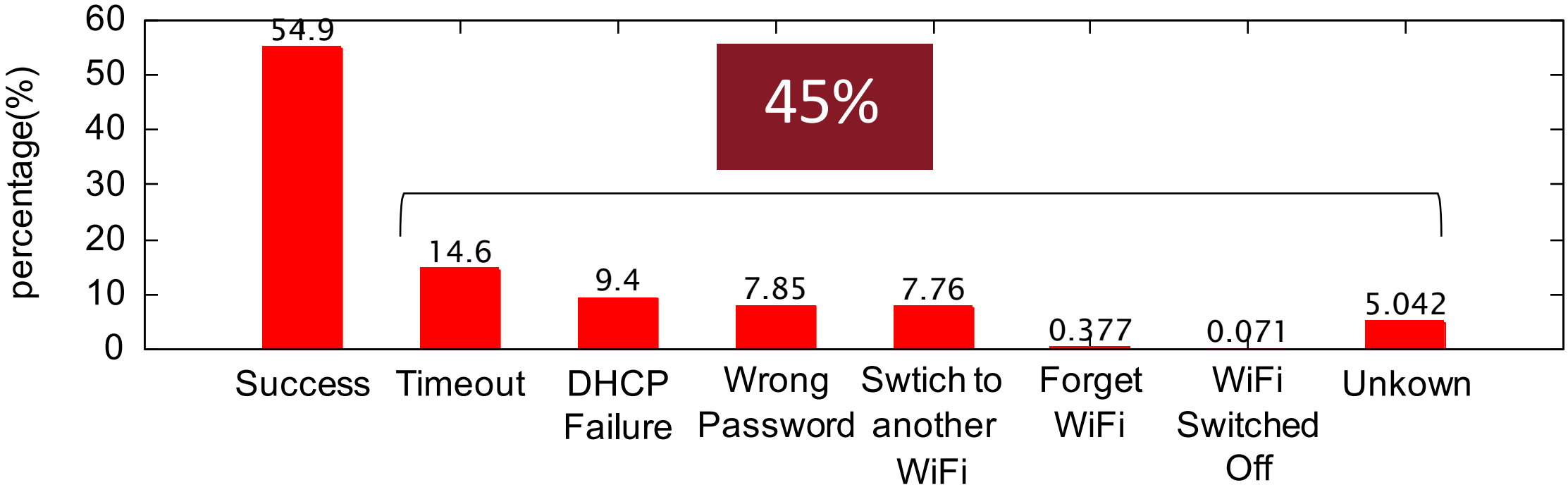
- 12,472 selected devices
- 706K connection attempts
- Spread over different places.

# Connection Time Cost



# WiFi Association: Success vs. Failure

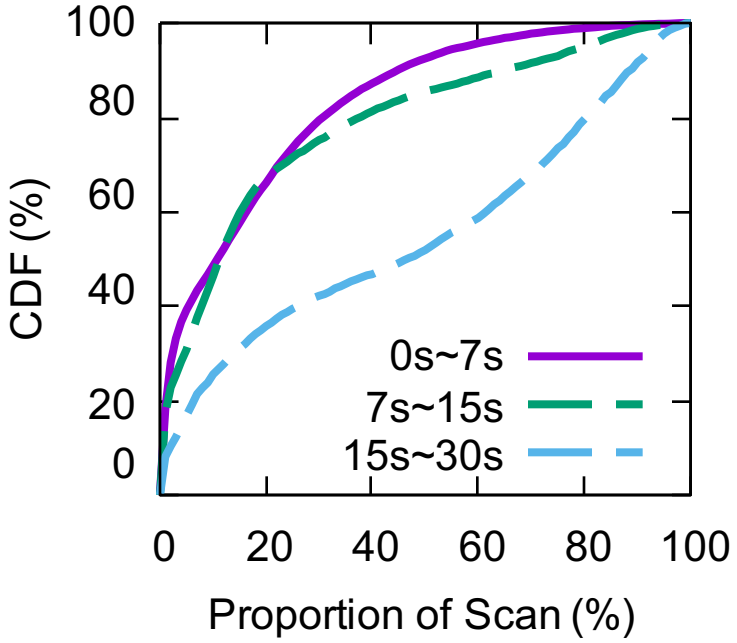
**Success**: set up successfully within 30 seconds

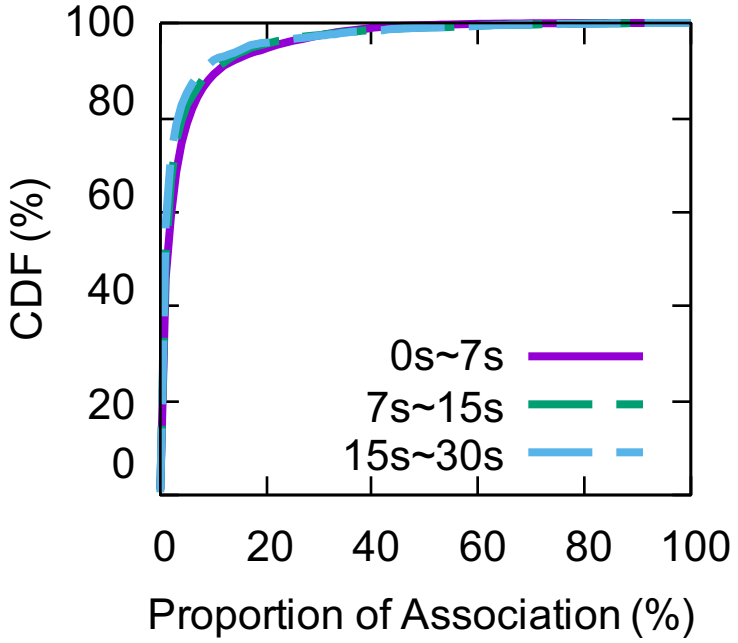



Based on the *breaking down dataset*.

Does there exist one sub-phase which dominates the overall connection set-up process?

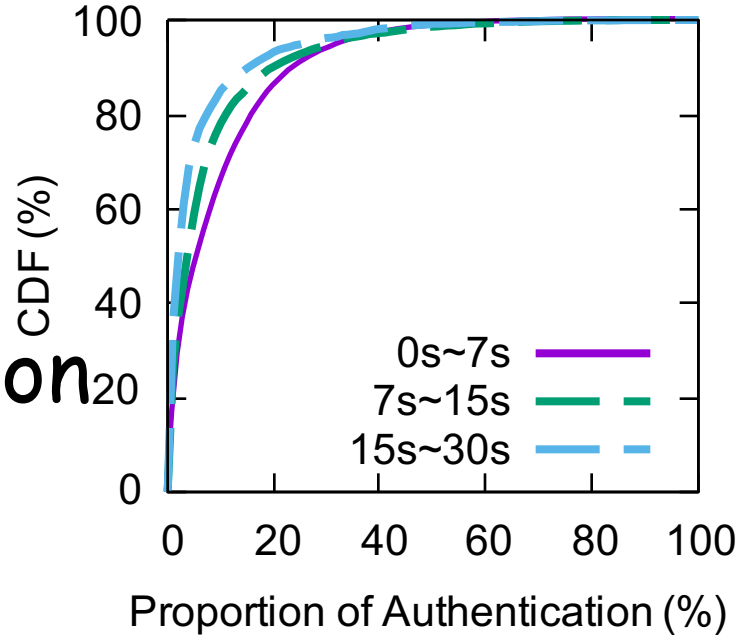
Scan

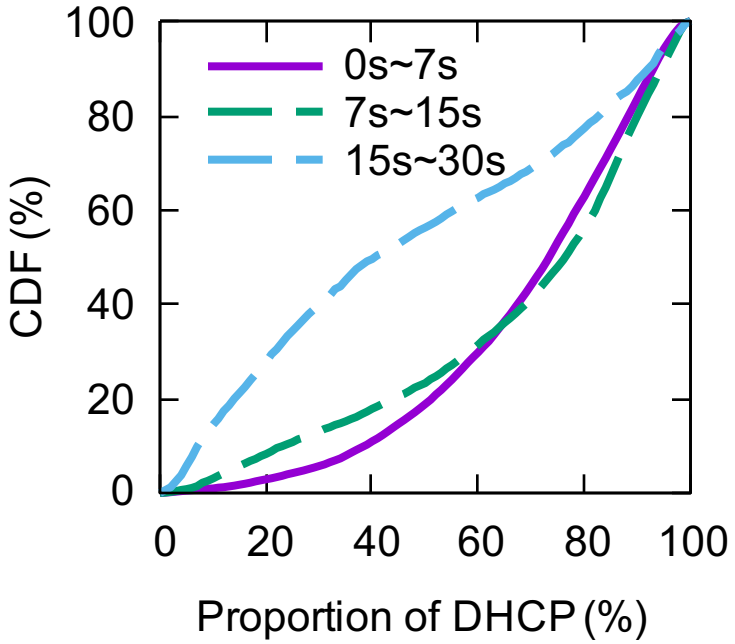



Association

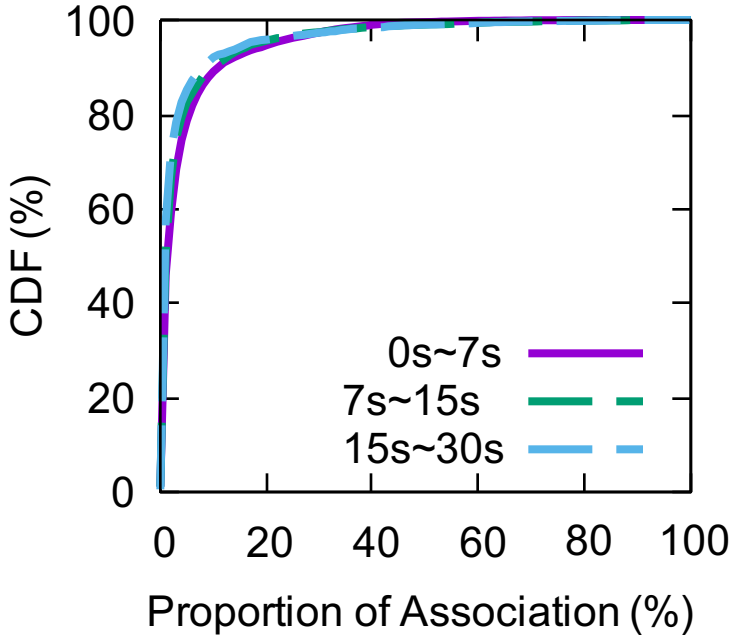
Authentication



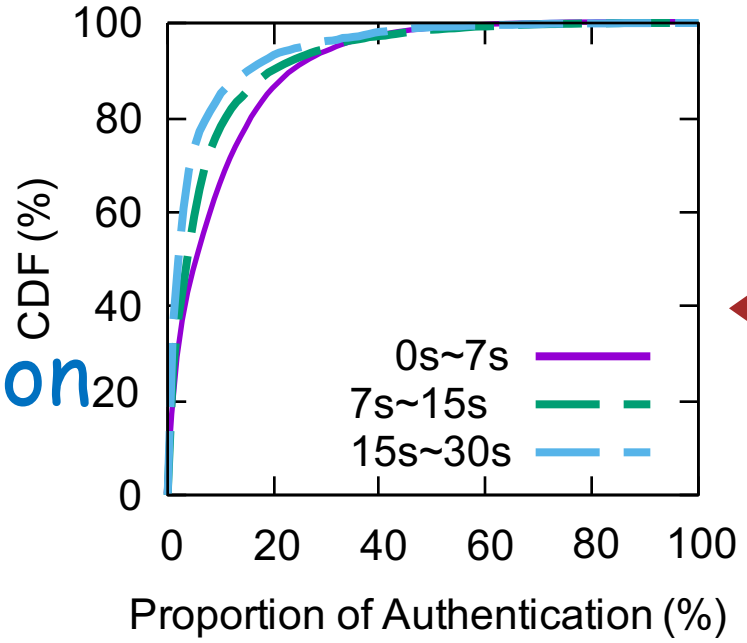




DHCP

Association and Authentication do not take too much time.



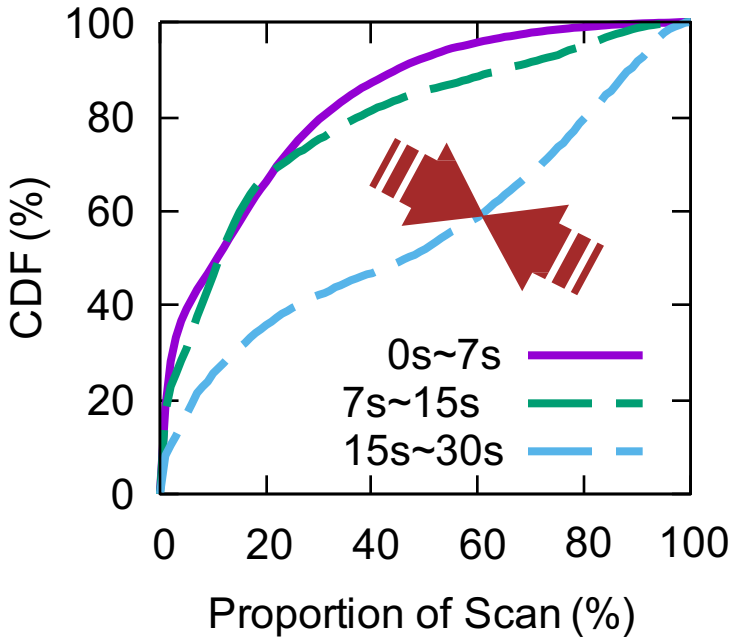

Association




Authentication

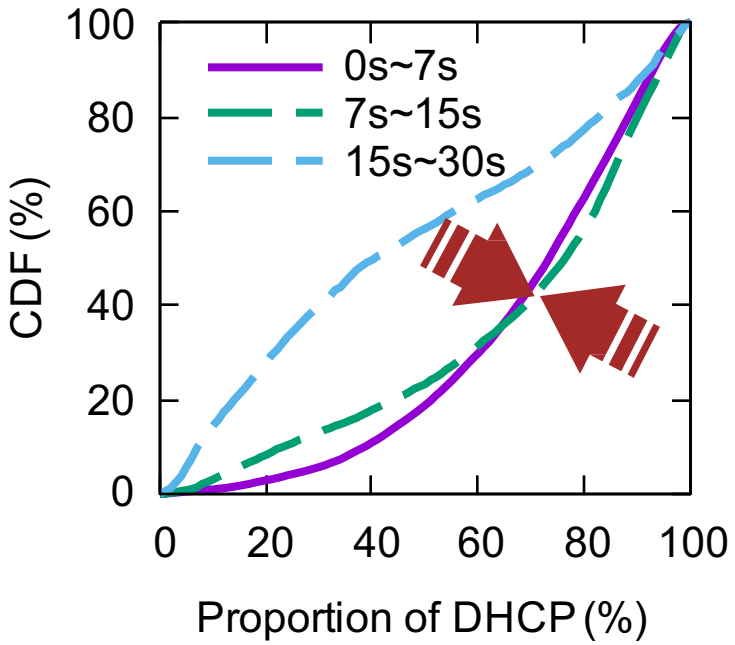
Fixed number of MAC-layer packets exchange in Association or Authentication.

Scan



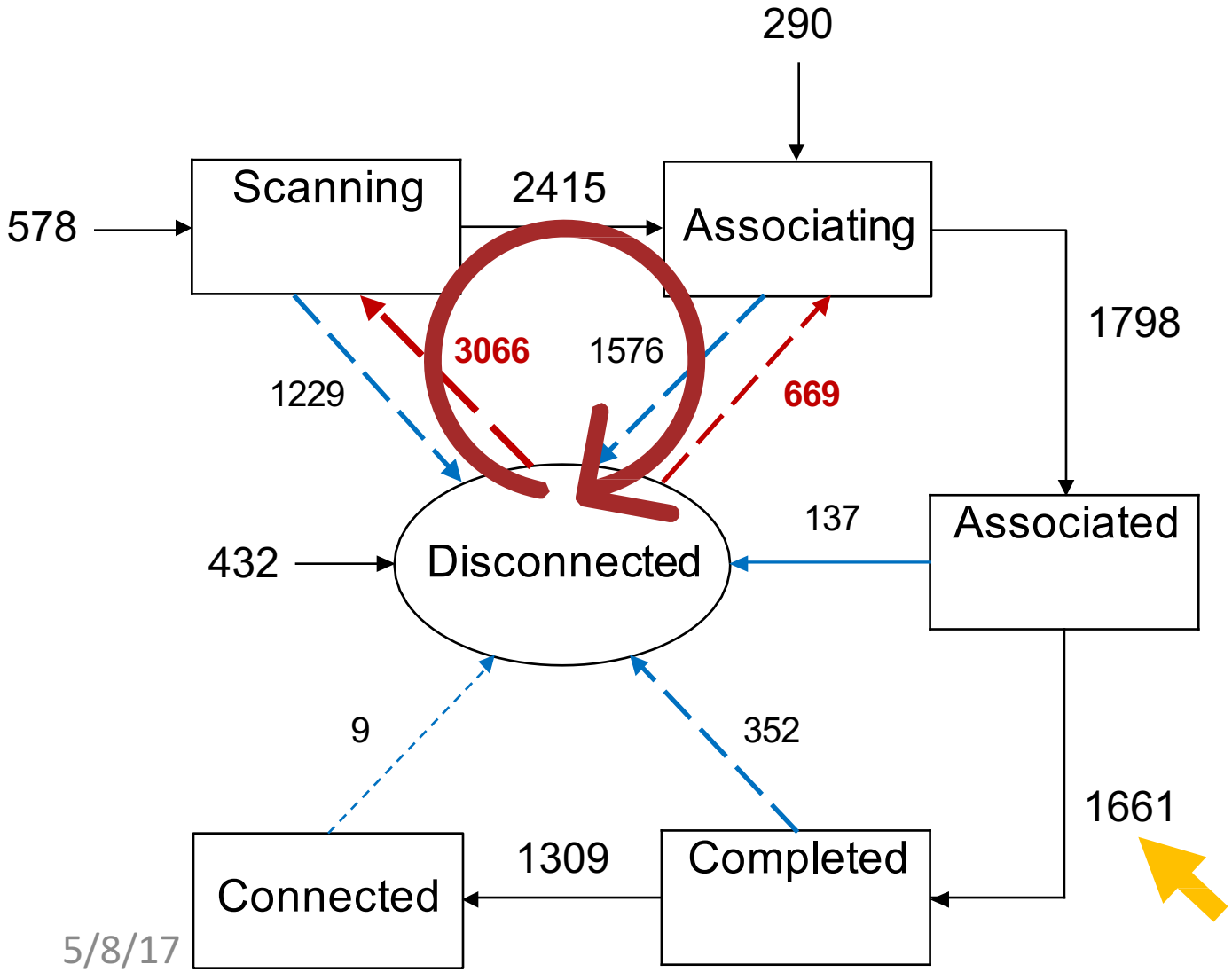

0s-7s, 7s-15s: DHCP phase occupies more than 80%, which is consistent with WiNTECH work.

15-30s: Scan phase consumes more than DHCP phase for those > 15s.




DHCP

# Anomalous transitions cause long scanning




- Anomalous transition to Disconnected state
  - Mobility
  - WiFi interference
  - System process delay
  - …

Total number of transitions in the dataset.

Motivation

Measurement

Results

Correlation

Analysis

Modeling

# Introduction of the Connection Log Dataset

| Abbreviation | Features |
|---|---|
| *hour of day* | Hour of day. |
| *RSSI* | Received Signal Strength Indicator. |
| *number of devices* | Number of associated devices. |
| *mobile device model* | Mobile device model. |
| *AP model* | AP model. |
| *Encrypted* | Encryption type of the AP. |
| *IsPublic* | Is public AP? |
| *result* | Connection result reported by the App. |
| *connection time cost* | Connection time cost. |

# Correlation Analysis

- Kendall correlation: (rank correlation)

$$tau = \frac{|concordant\ pairs| - |discordant\ pairs|}{n(n-1)/2}$$

- Relative Information Gain: (RIG)

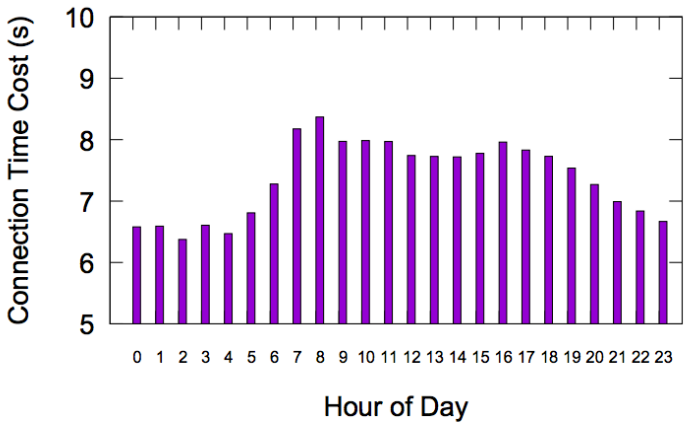
how much a factor X helps to predict the final latency Y

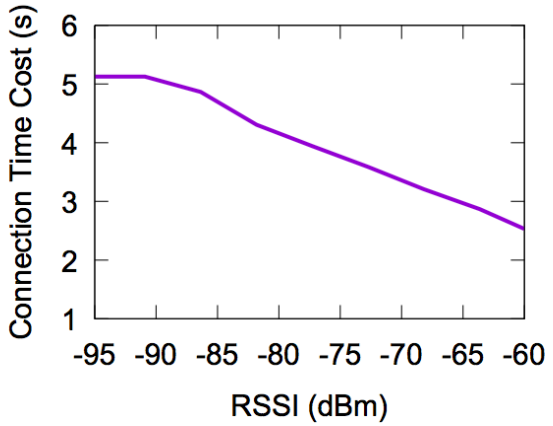
$$RIG = \frac{H(Y) - H(Y|X)}{H(Y)} \qquad H(Y) = \sum_i P[Y = y_i] log \frac{1}{P[Y = y_i]}$$

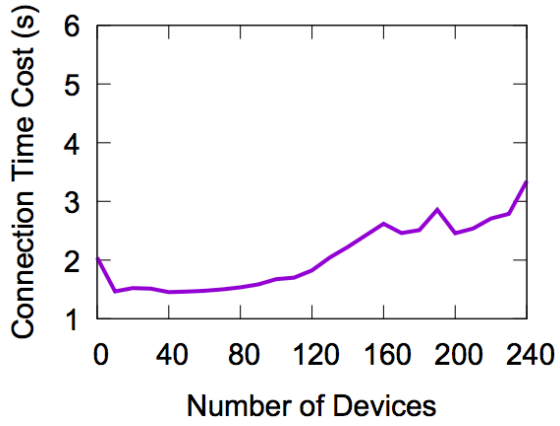
# Correlation Analysis

| Features | RIG | Kendall |
|---|---|---|
| *mobile device model* | 0.156 | / |
| *AP model* | 0.078 | / |
| *RSSI* | 0.020 | -0.395 |
| *number of devices* | 0.006 | 0.208 |
| *hour of day* | 0.005 | / |

- Mobile devices and AP model has the highest RIG.
  - HTC on average 1.3x larger than Samsung.



# Mobile Device Model

- Chipset matters.
  - Each model contains> 10K pieces of data
  - RSSI > -60 dBm
  - 500+ devices, 500+ APs, 7 days, 500+ places

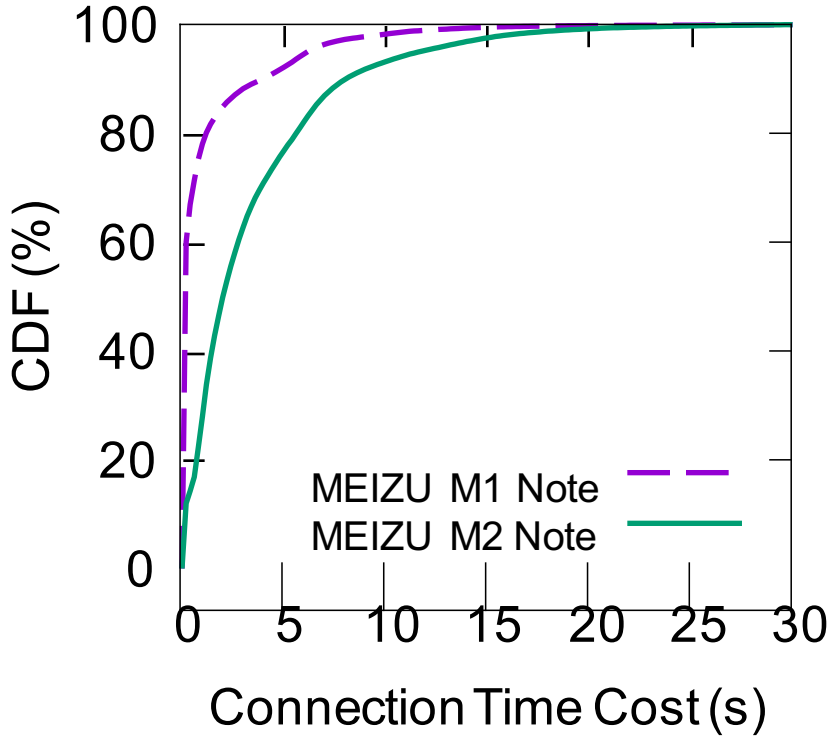



TABLE III: The software and hardware parameters for different kinds of mobile device models.

| Average *connection time cost* | Device model | Operating System | Chipset | CPU Frequency | RAM Size | Wireless Interface |
|---|---|---|---|---|---|---|
| 475ms | **MEIZU M1 Note** | Flyme | MediaTek 6752 | 1.7GHz | 2GB | IEEE a/b/g/n |
| ... | ... | ... | ... | ... | ... | ... |
| 2463ms | **MEIZU M2 Note** | Flyme | MetiaTek 6753 | 1.3GHz | 2GB | IEEE a/b/g/n |

# Mobile Device Model

- Operating system matters.

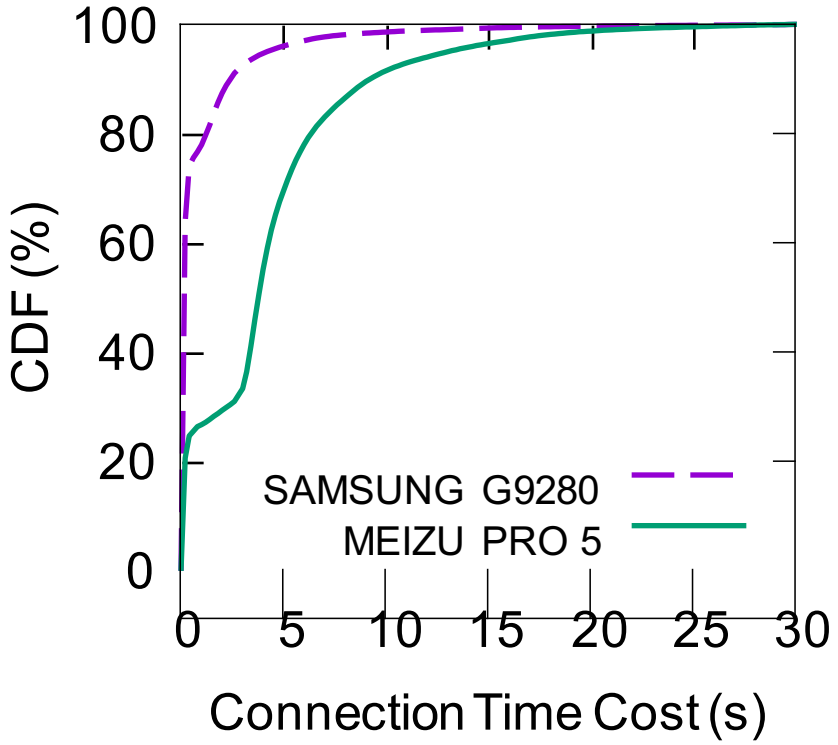



TABLE III: The software and hardware parameters for different kinds of mobile device models.

| Average *connection time cost* | Device model | Operating System | Chipset | CPU Frequency | RAM Size | Wireless Interface |
|---|---|---|---|---|---|---|
| 754ms | **SAMSUNG G9280** | Android OS | Exynos 7420 | 2.1GHz | 4GB | IEEE a/b/g/n/ac |
| ... | ... | ... | ... | ... | ... | ... |
| 3534ms | **MEIZU PRO 5** | Flyme | Exynos 7420 | 2.1GHz | 4GB | IEEE a/b/g/n/ac |

Motivation

Measurement

Results

Correlation

Analysis

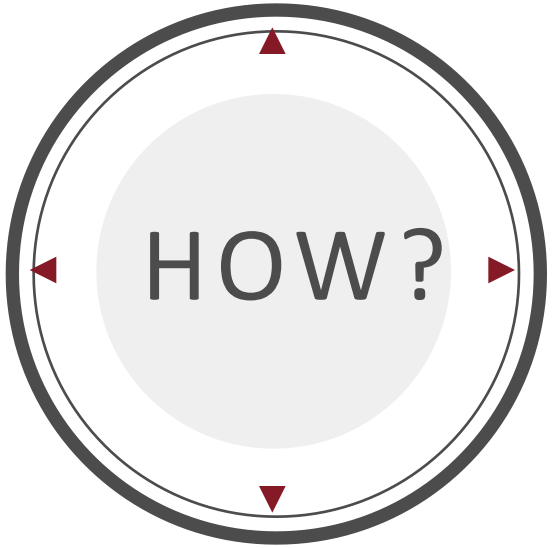
Modeling

# What can I do to reduce the connection time cost?

Machine Learning Based Model

Feature selection

Model selection




Enhanced AP Selection Algorithm

What-If Analysis

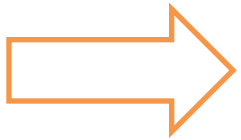
# Machine Learning based Model

- Labeling
  - Use **15 seconds** as the threshold to divide the process into **SLOW** and **FAST**.
- Model Selection
  - Highest accuracy: Random Forest.

# Modeling

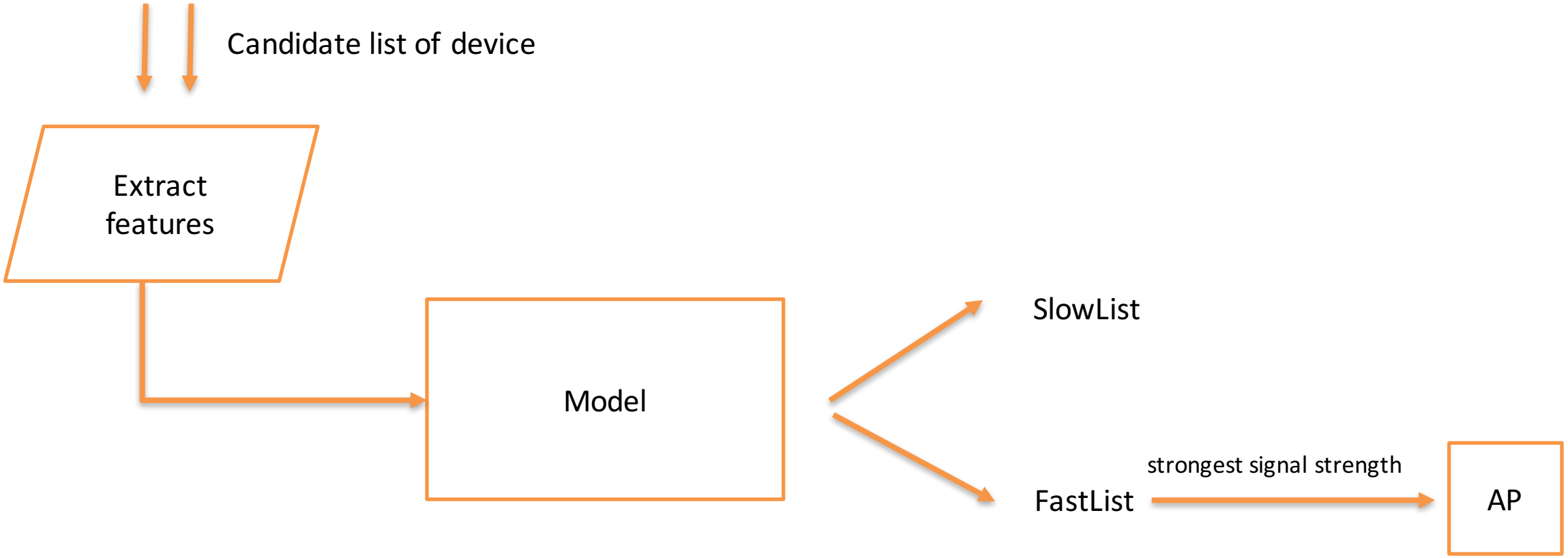
Features

- hour of day
- RSSI
- Mobile device model
- AP model
- Encrypted



Random Forest Model

# AP Selection Algorithm



# Evaluation

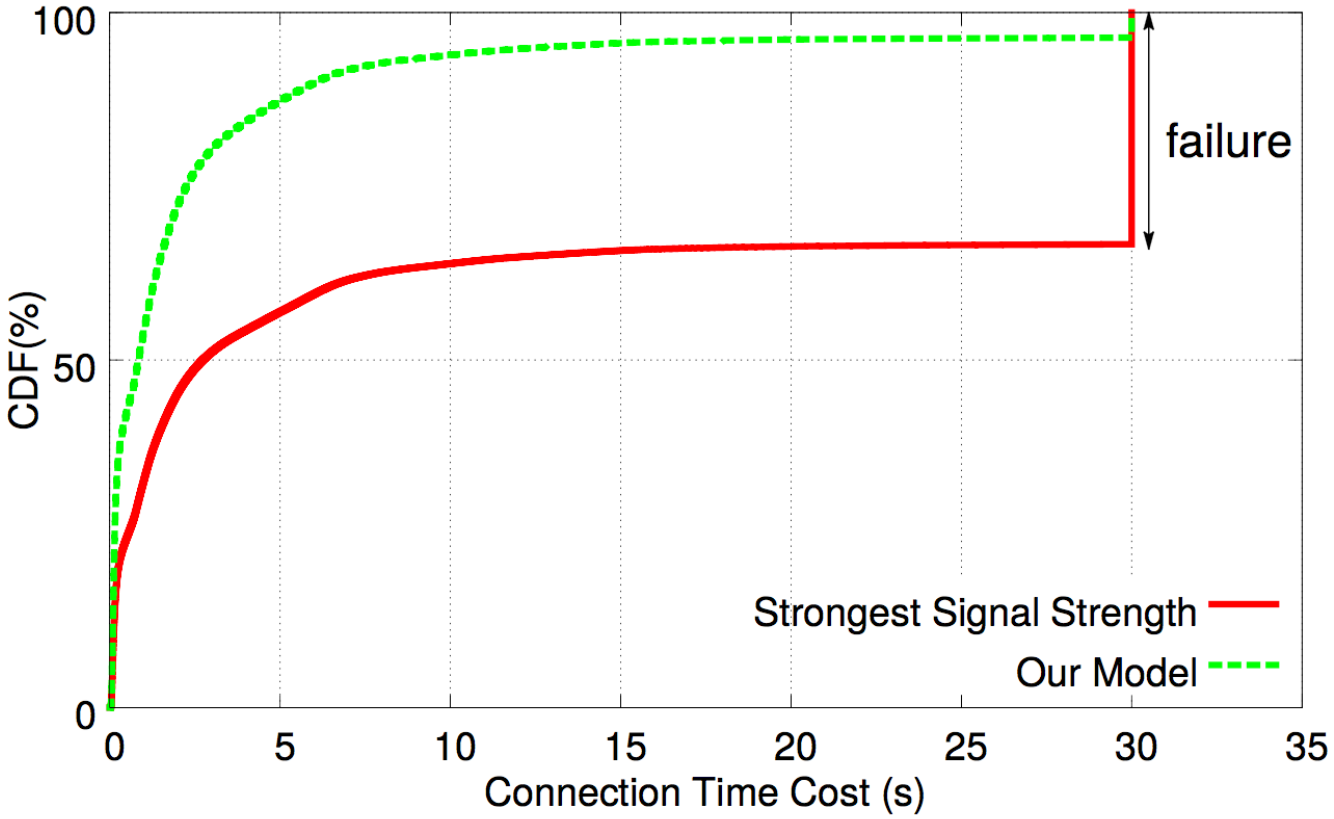



Fig. 10: Comparison of AP selection algorithms.

Connection attempt fail

33%

sss

3.6%

model

# Evaluation

| Precision(SLOW) | Recall(SLOW) | PoA |
|---|---|---|
| 0.40 | 0.98 | 0.15 |
| 0.43 | 0.96 | 0.21 |
| 0.49 | **0.90** | **0.33** |
| 0.54 | 0.84 | 0.40 |

PoA= $\frac{|FAST\ set|}{|FAST\ set|+|SLOW\ set|}$.

TABLE IV: Accuracy of random forest model. The parameters we use for this model are: Tree depth=90, #Tree=100, weight=0.3.

| Label. | Precision | Recall | Features Used |
|---|---|---|---|
| FAST | **0.91** | 0.49 | *hour of day, RSSI, AP model, mobile device model, Encrypted* |
| SLOW | 0.48 | **0.90** | |

# Conclusion

- Focusing on connection set-up time
- Large scale measurement compared with current work
- Data-driven method: set-up connection time
- Using deep learning method to solve the problem

# Shortcomings

Modeling selection method is not persuasive
Modeling index is not very high(Precision and recall)
The dataset doesnot cover the iphone users


# Thank you! Q&A?

changhuapei@gmail.com