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

Motivation

Measurement

Results

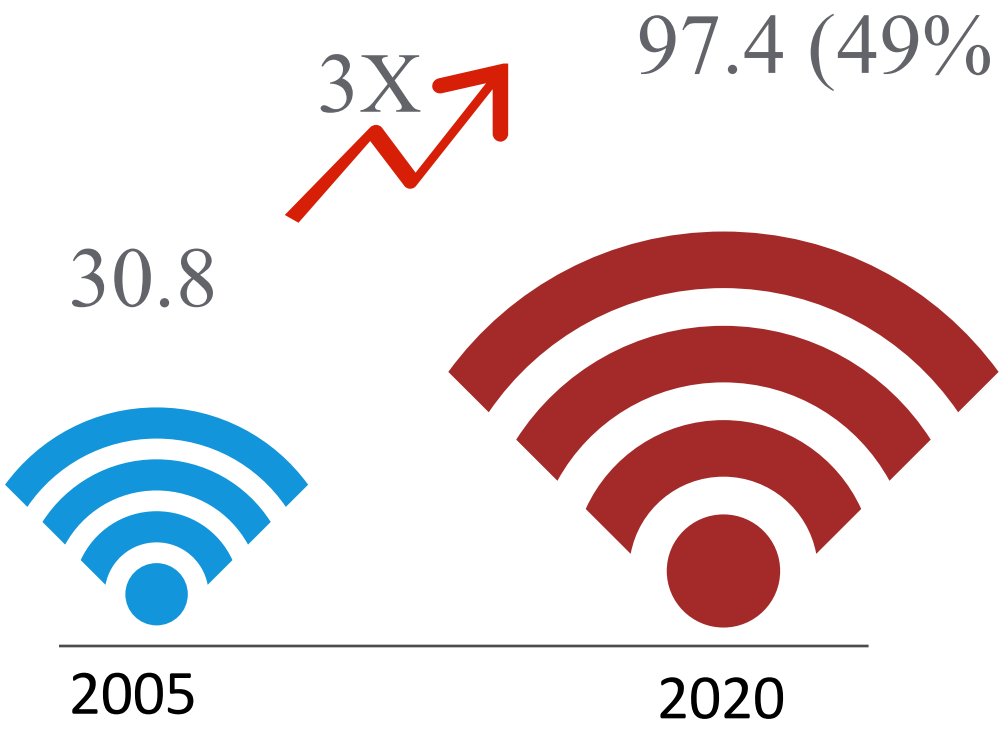
Correlation

Analysis

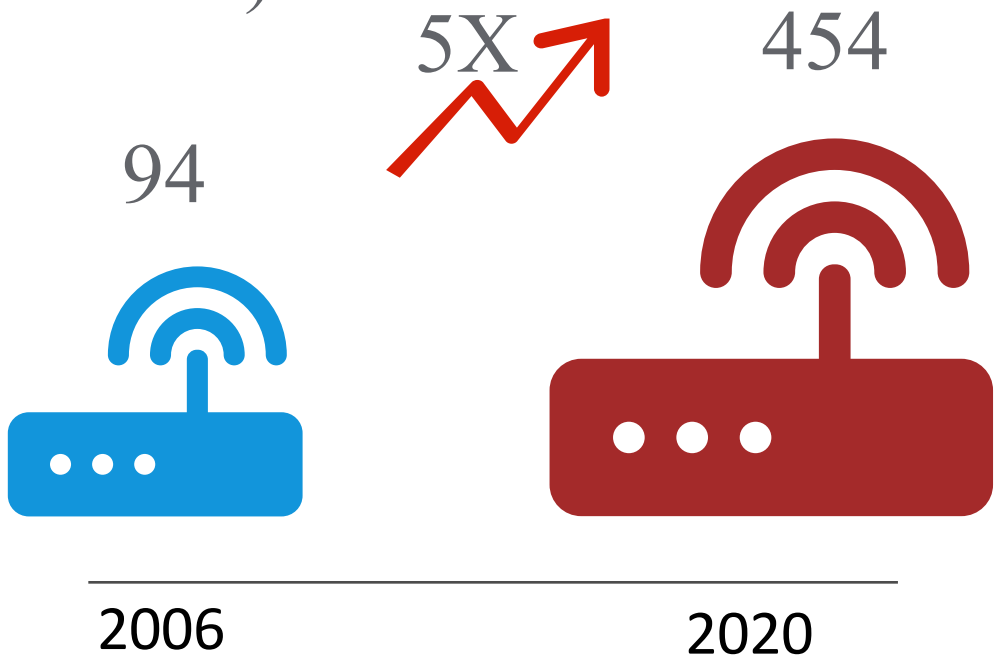
Modeling

WiFi is indispensable in our daily lives

Source: Cisco VNI Mobile, 2017

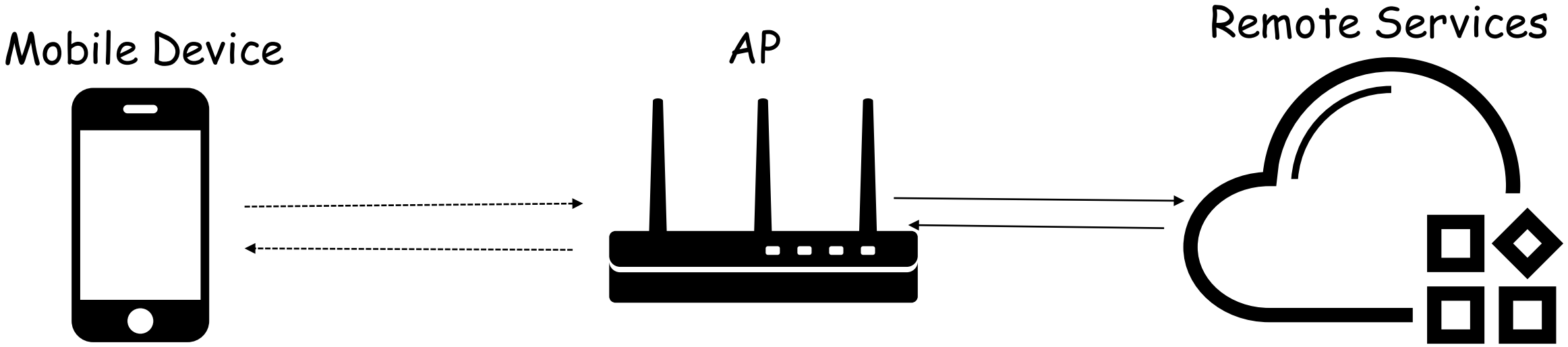


Overall WiFi Traffic Growth
(Exabytes)

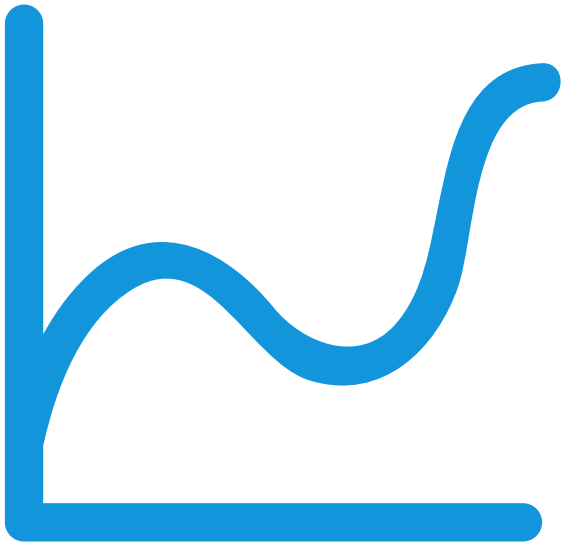


Hotspots
(Millions)

Experience of WiFi Network

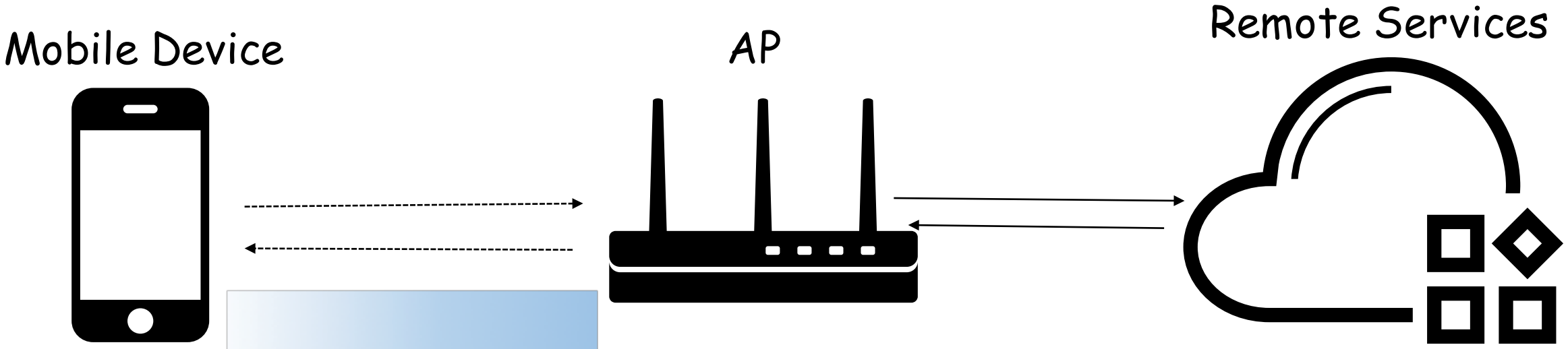


Throughput

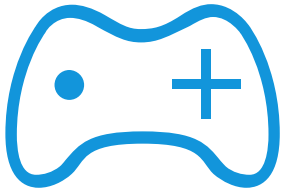


Downloading

Experience of WiFi Network

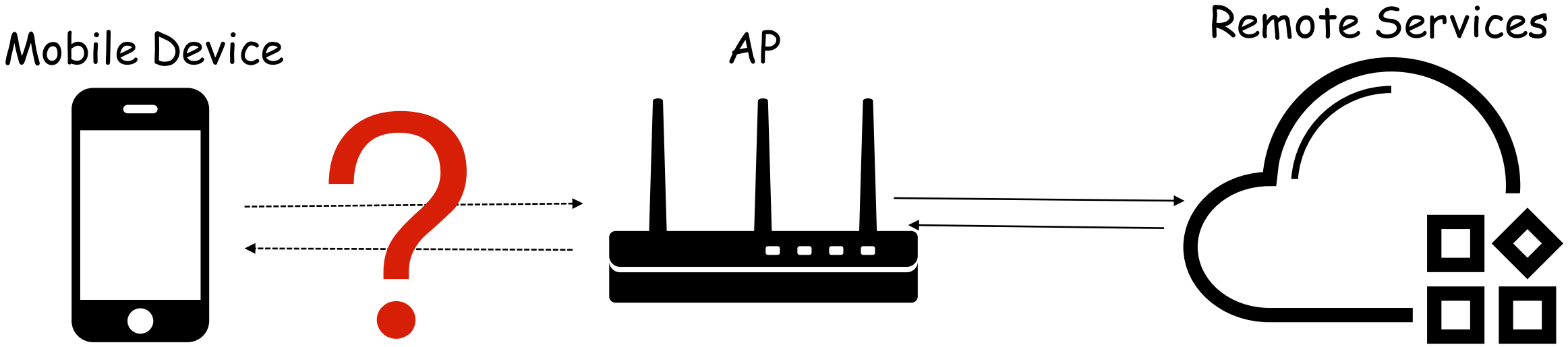


WiFi Hop Latency

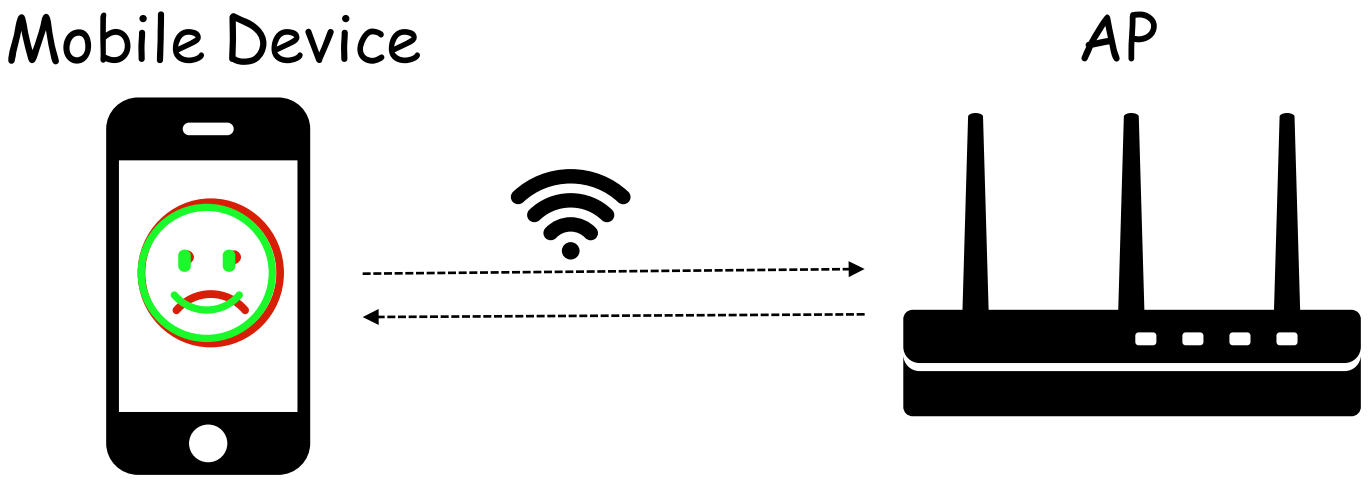


Online Gaming

Experience of WiFi Network



Experience of WiFi Network

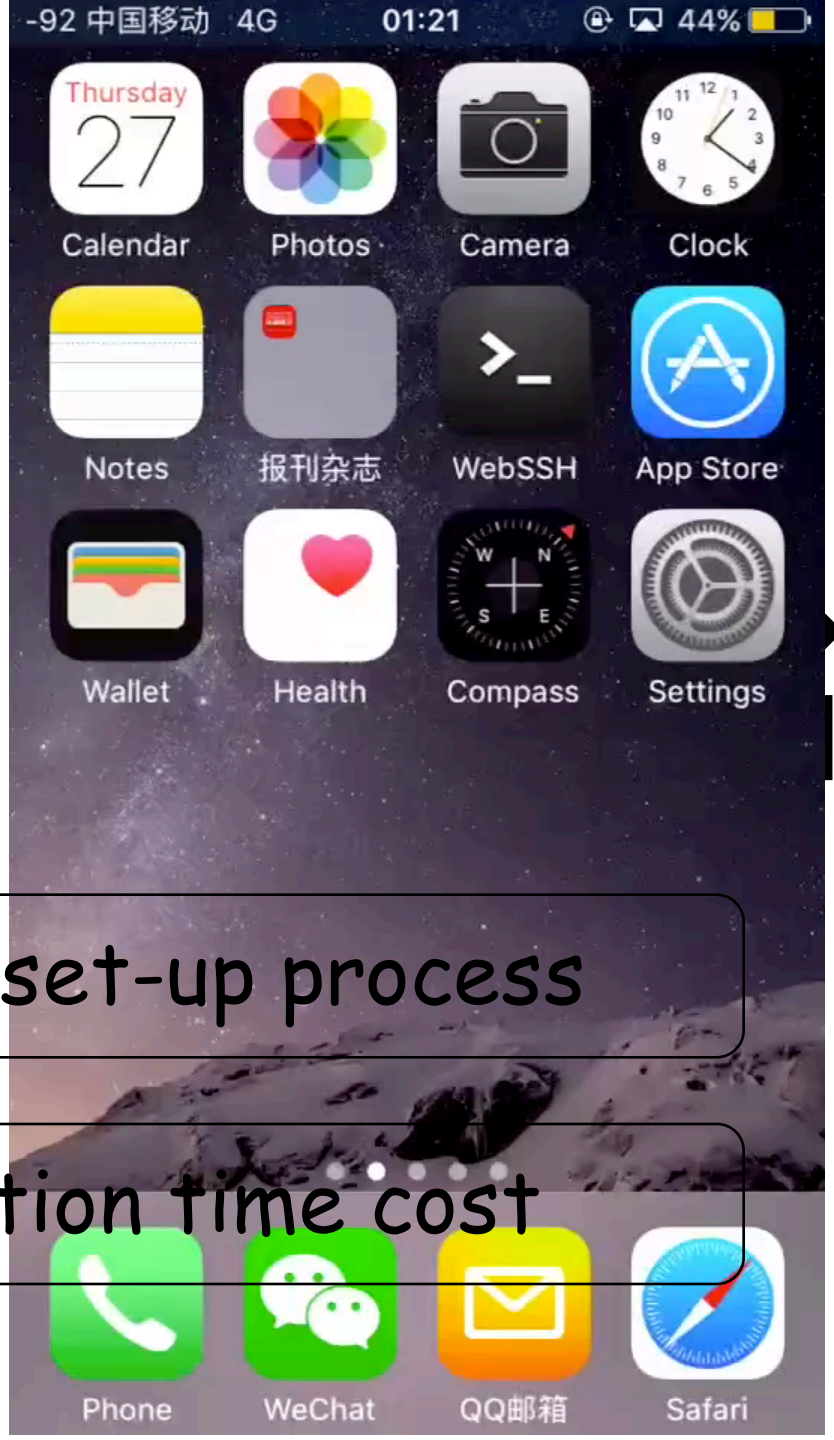


I want to connect to the AP!

Now I can talk to the Internet ^^

Connection set-up process

Connection time cost



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

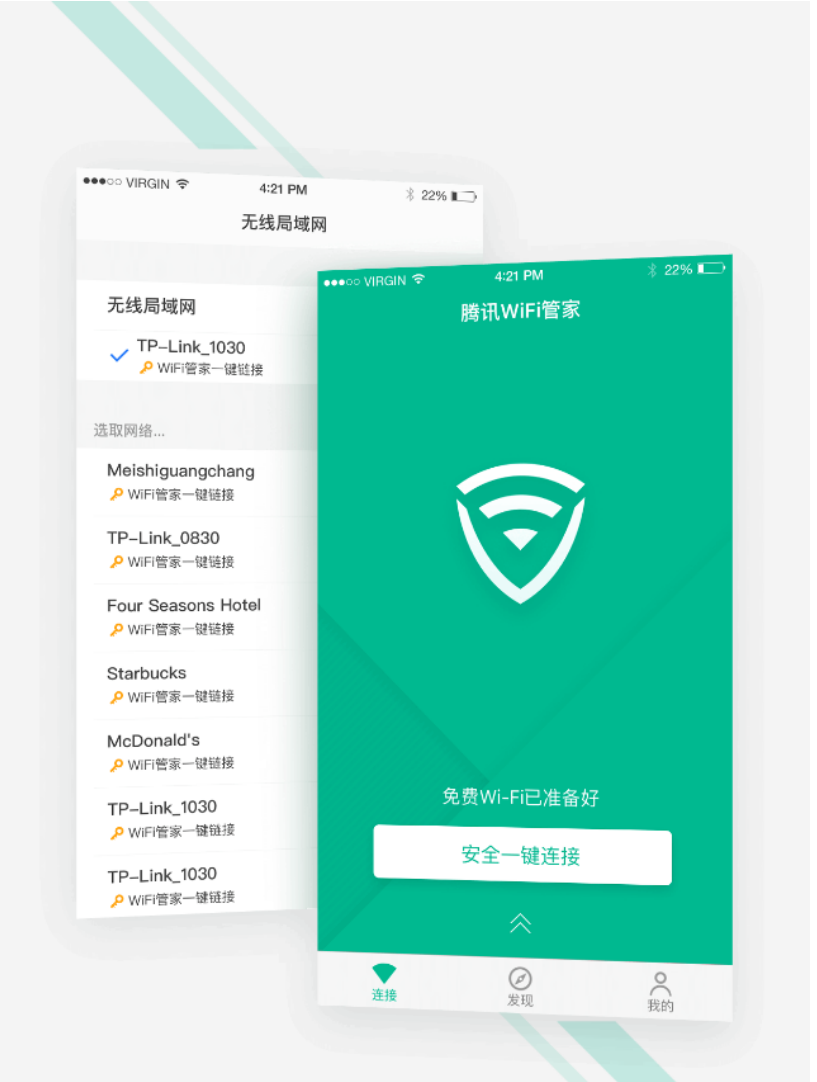
Correlation

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.

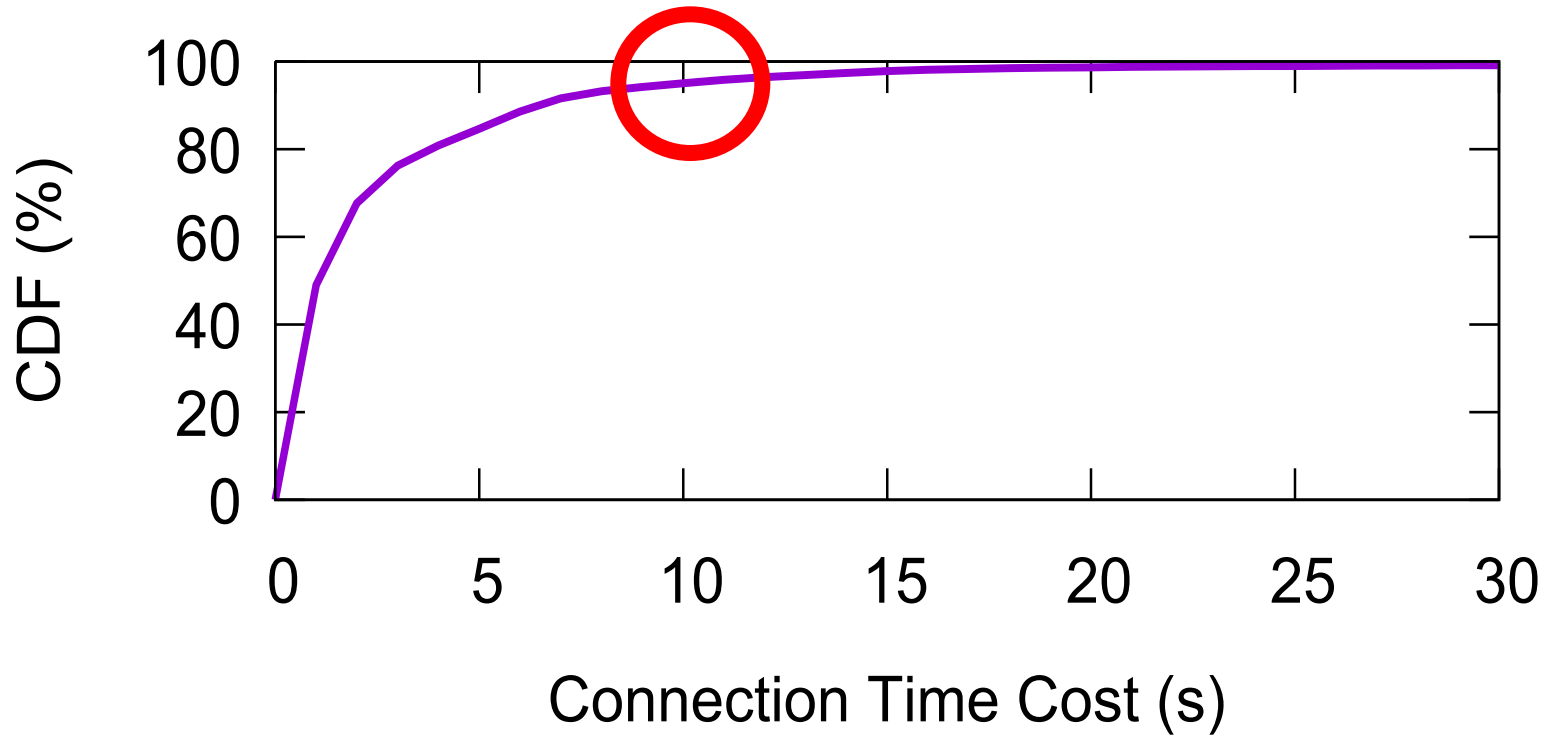


DATASET

Connection Log Dataset

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

CDF of the connection time cost



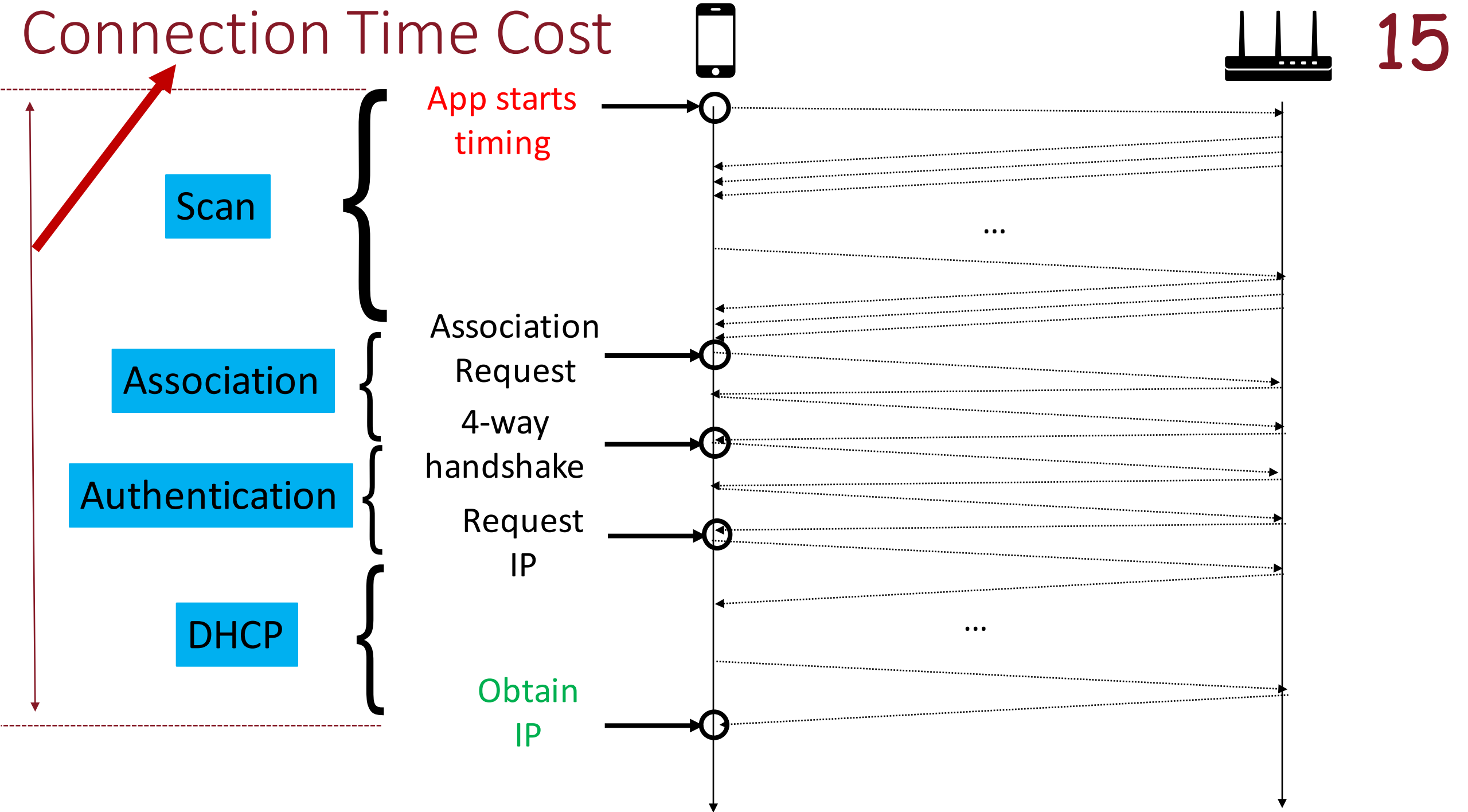
15% (5%) successful connections consume over 5 (10) seconds!

DATASET

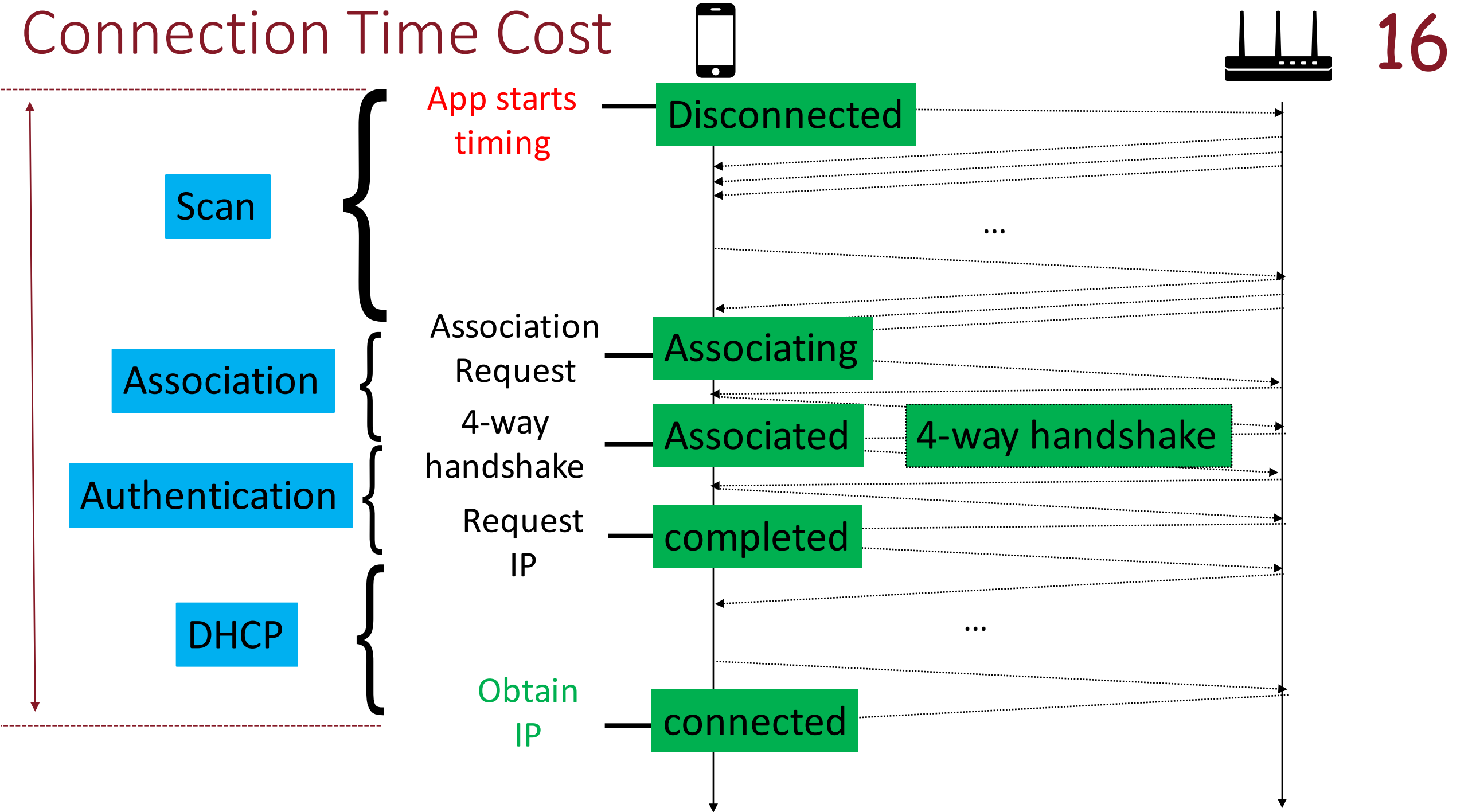
Breaking Down Dataset

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

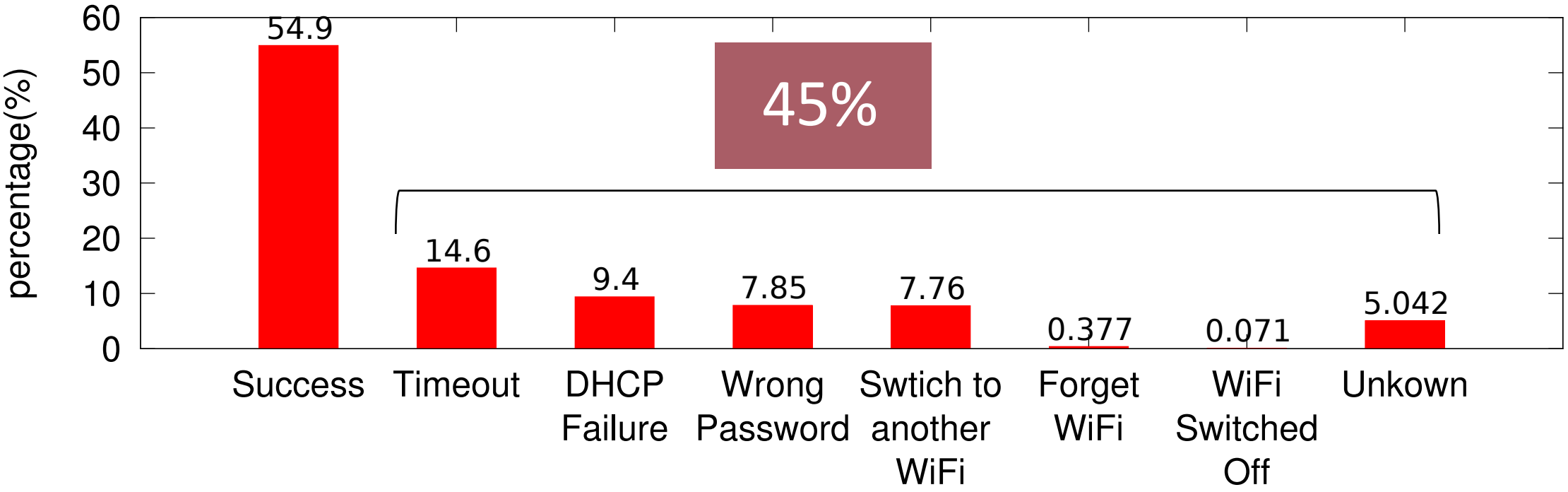
Connection Time Cost



Connection Time Cost



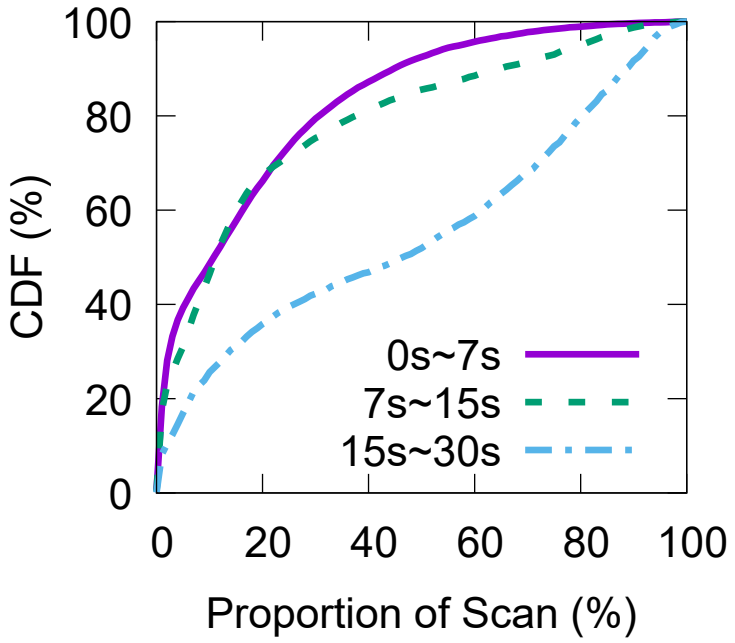
WiFi Association: Success vs. Failure



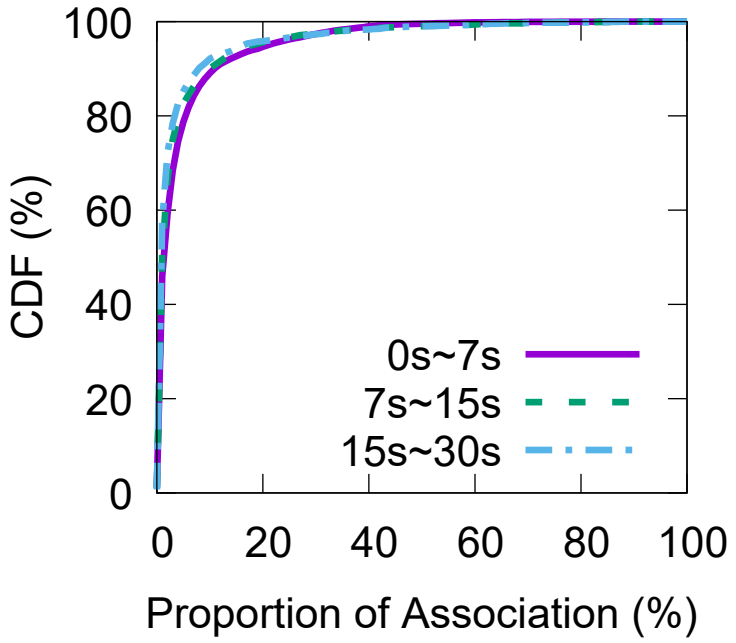
Based on the *breaking down dataset*.

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

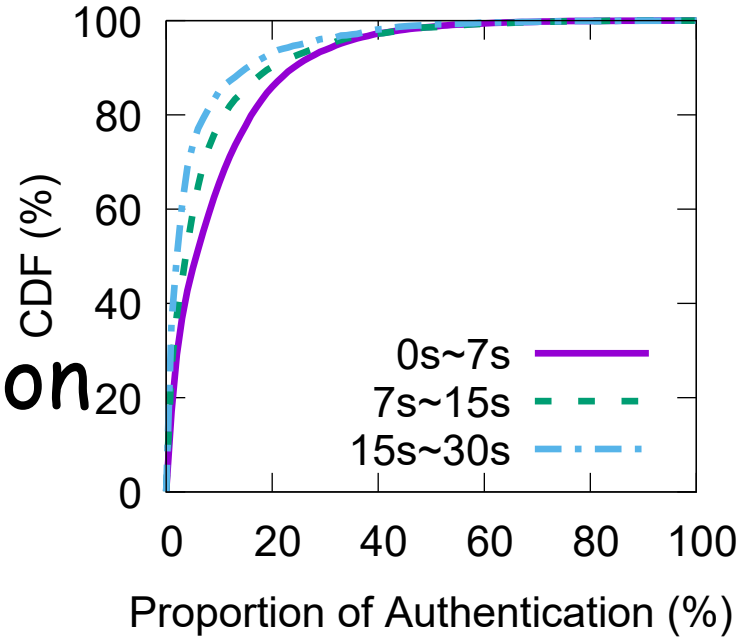
Scan



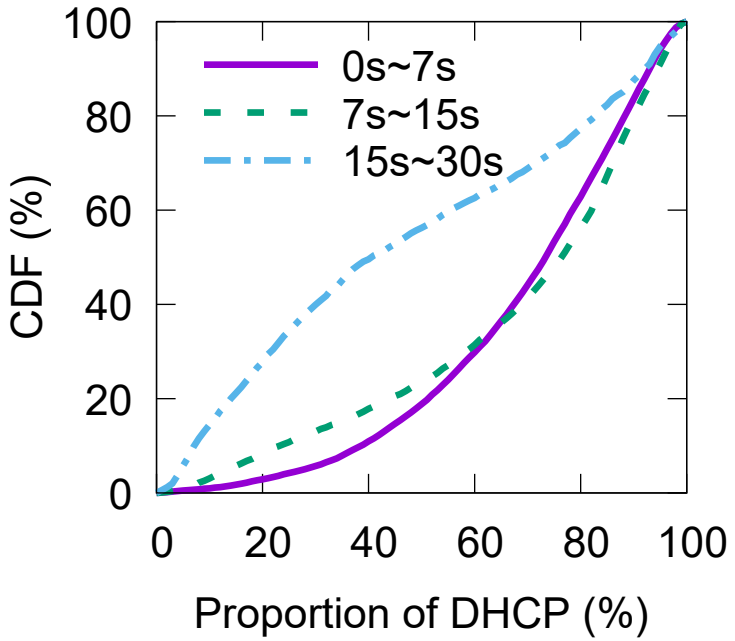
Association



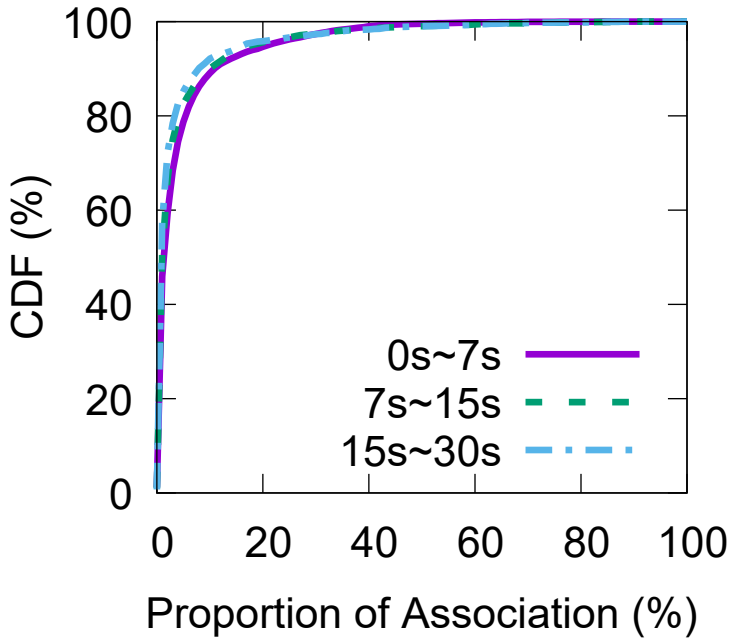
Authentication



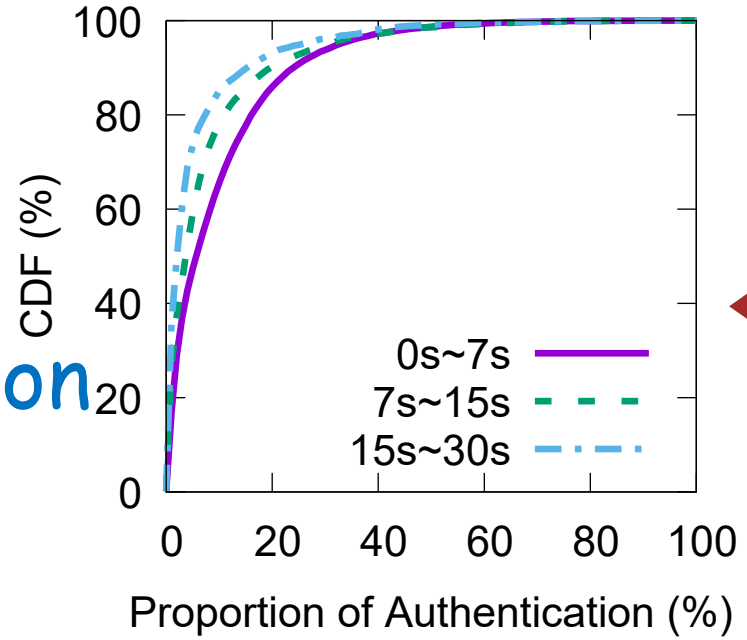
DHCP



Association and Authentication do not take too much time.



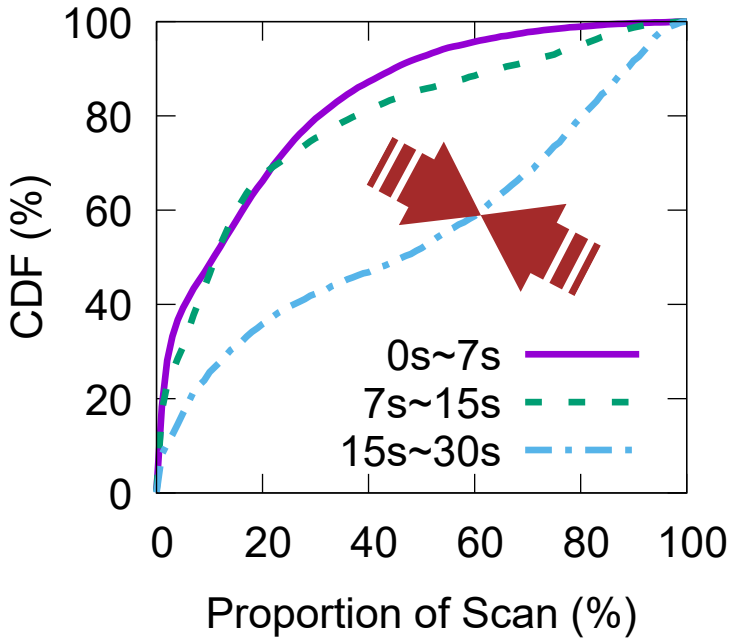
Association



Authentication

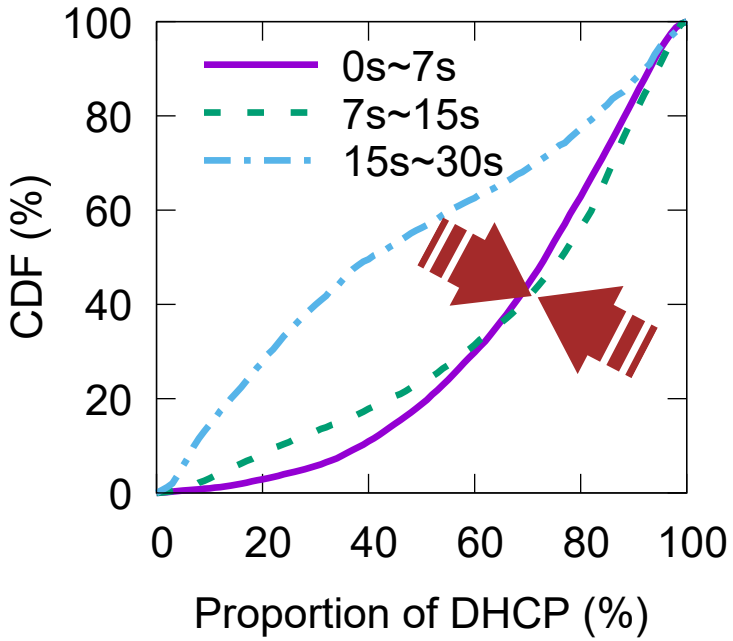
Fixed number of MAC-layer packets exchange in Association or Authentication.

Scan



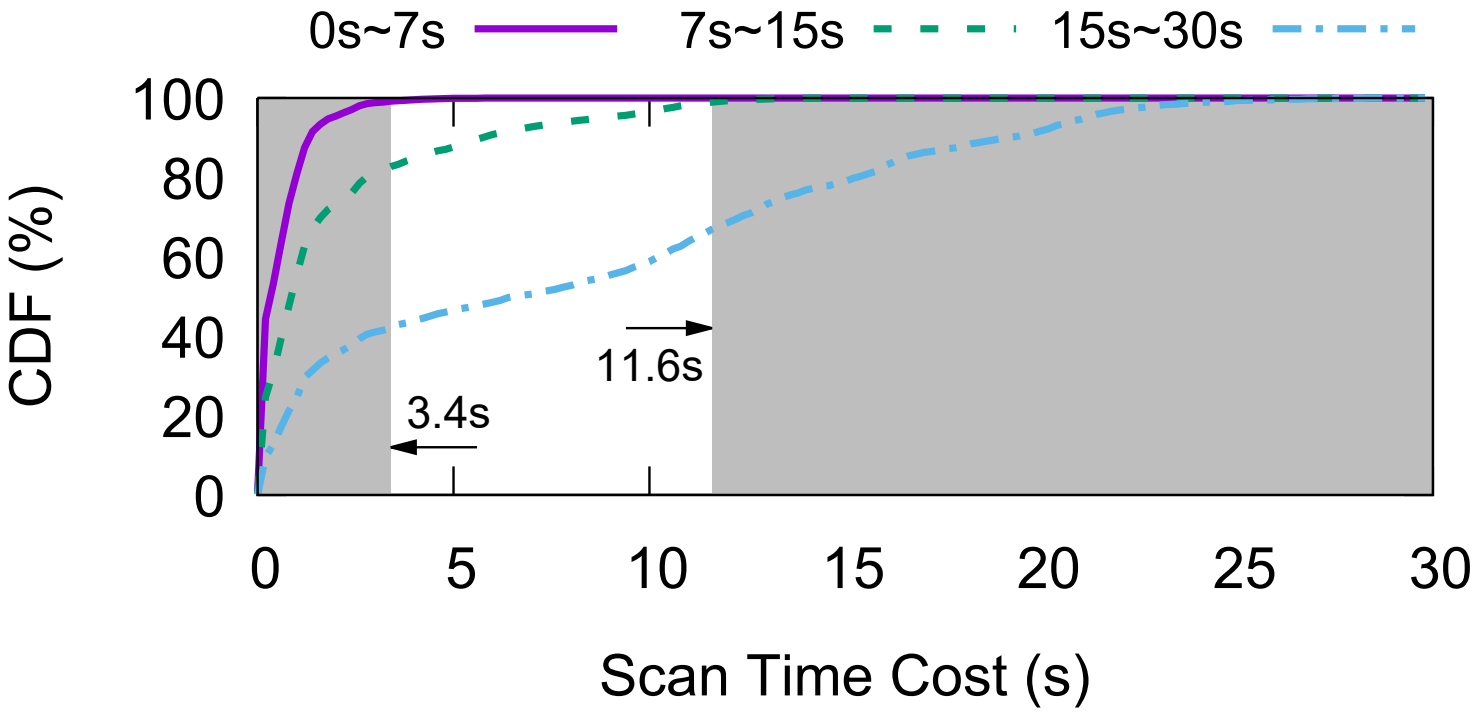
15-30s: Scan phase consumes more than DHCP phase for those > 15s.

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



DHCP

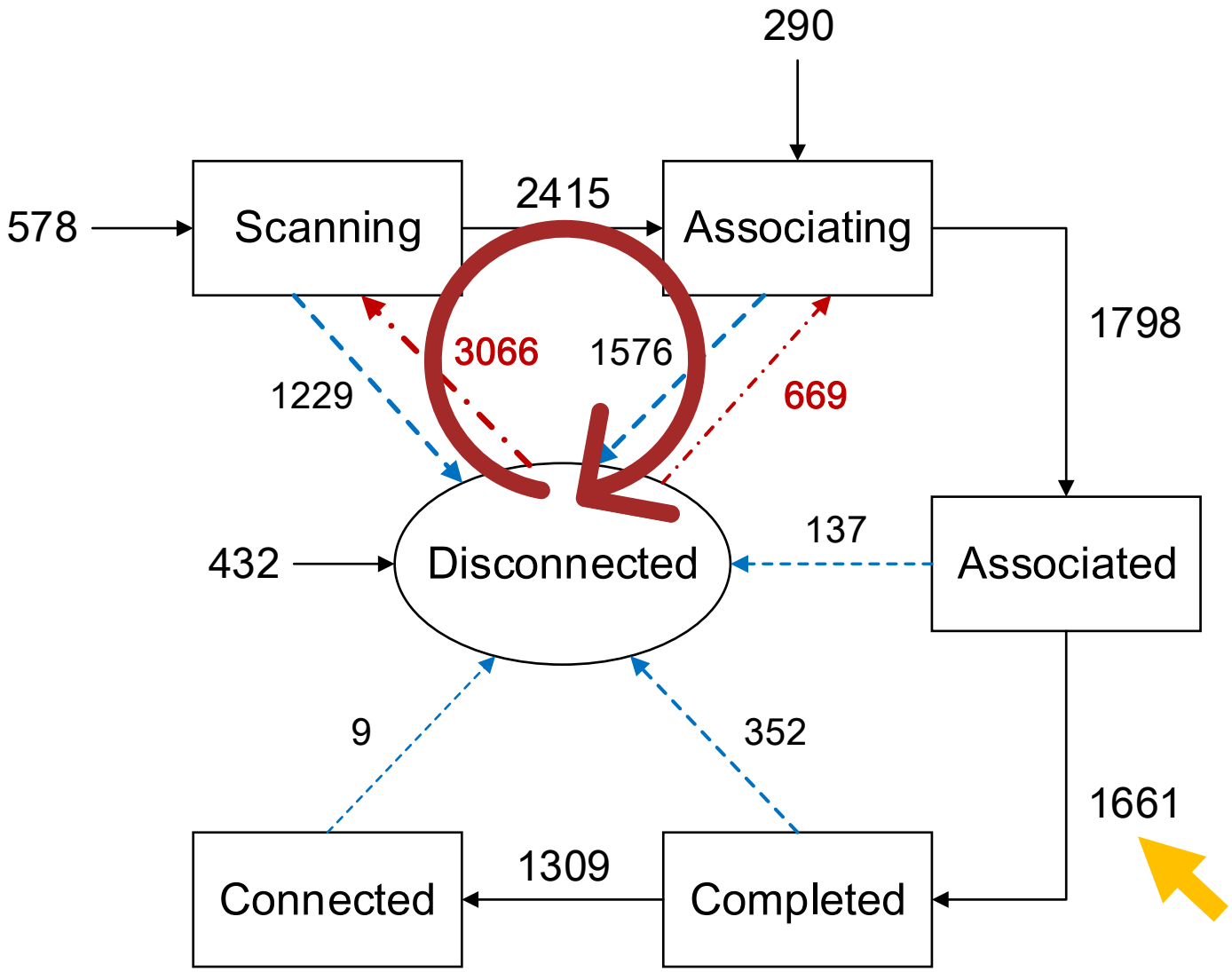
Scan Phase dominate the connection



- 0s-7s class: Scan consume less than 3.4s.
- 15s-30s class: For more than 40% processes, Scan phase consume more than 11.6s.

• Why does this happen?

Anomalous transitions cause long scanning



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

Total number of transitions in the dataset.

Take-away messages:

1. For those connection whose time cost > 15s, Scan is the dominate sub-phase.
2. Scan dominates the whole process because there are anomalous transitions.

Motivation

Measurement

Results

Correlation

Analysis

Modeling

Which feature affect the connection time cost the most?

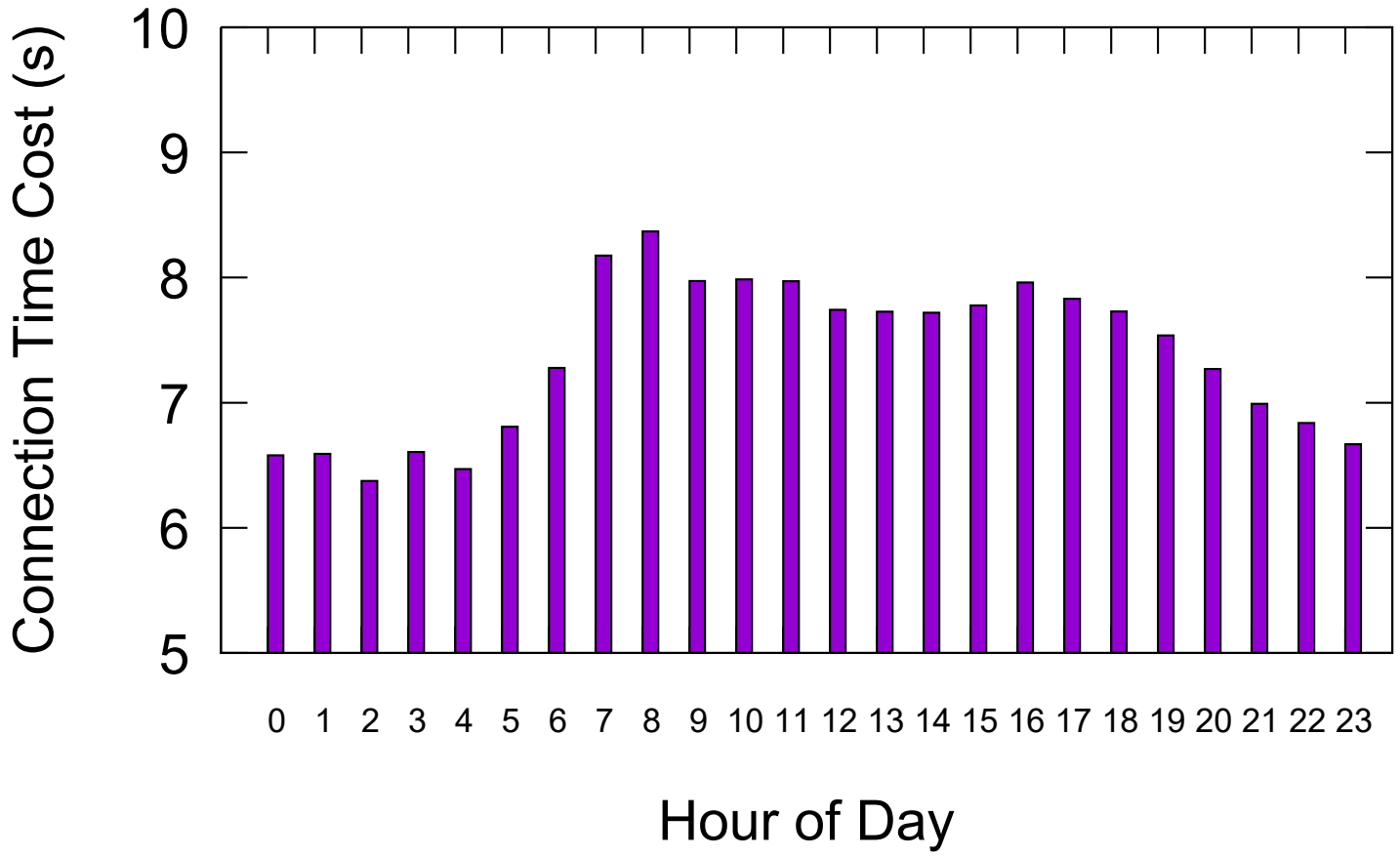
1. Give Intermediate results to gain some intuitions before the ML model.
2. Help feature selection.

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.

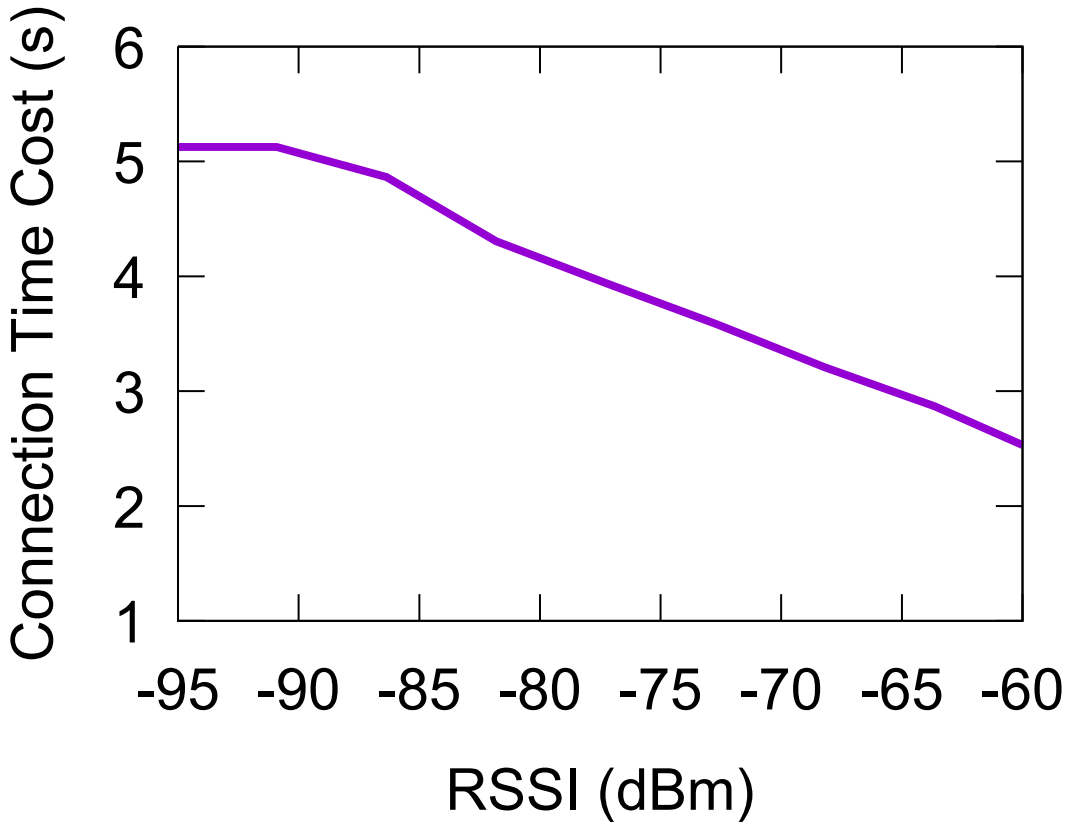
Visualization analysis

Aggregated results of all the APs.



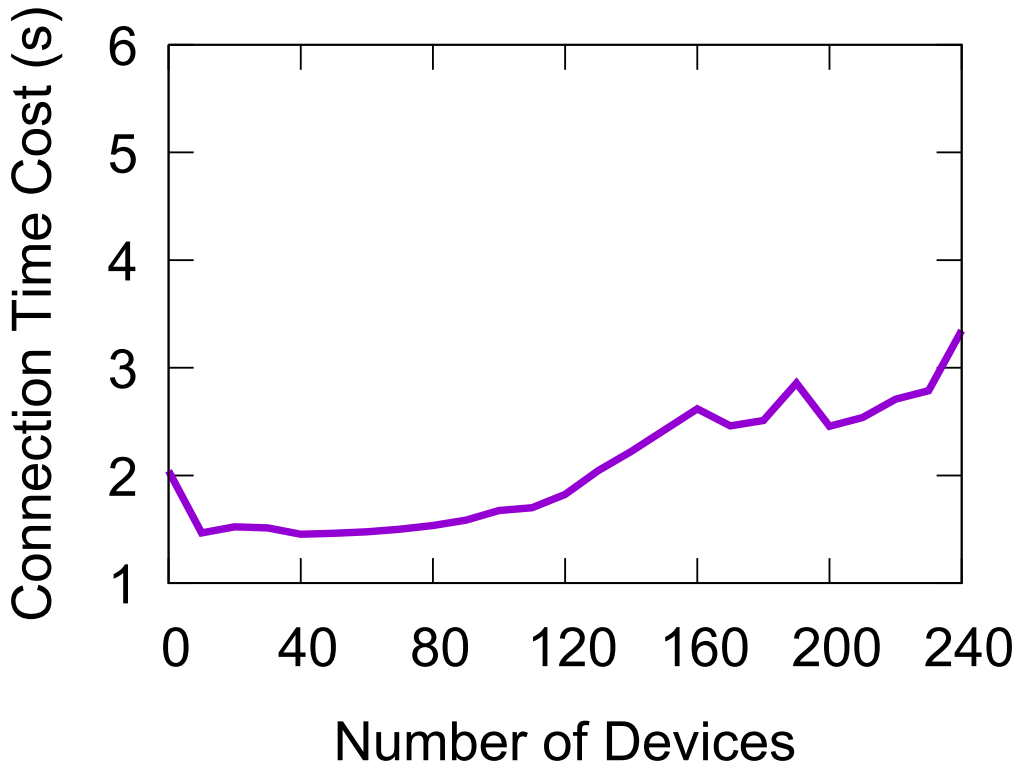
- Association timing affects the connection time cost.

Visualization analysis



- Connection with higher RSSI tend to have smaller average connection time costs.

Visualization analysis



- The larger the number of associated devices is, the higher average connection time cost.

Correlation Analysis

- Kendall correlation: (rank correlation)

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

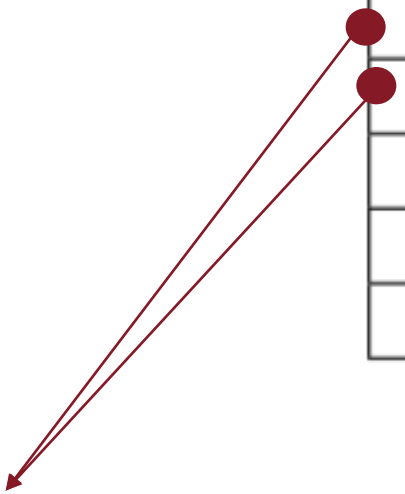
- Relative Information Gain: (RIG)

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

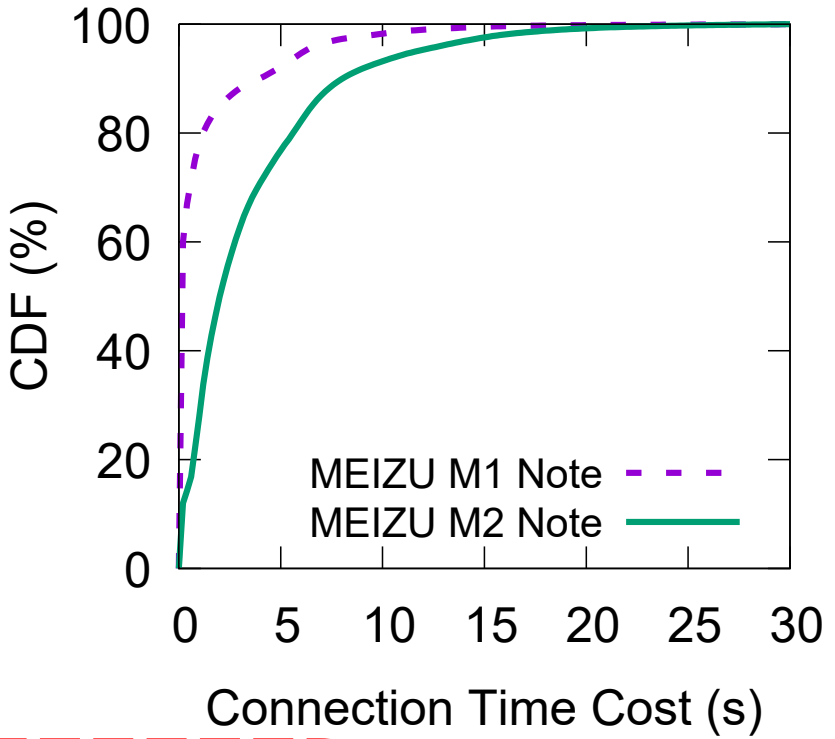


TABLE III: The software and hardware parameters for different kinds of mobile device models.

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.

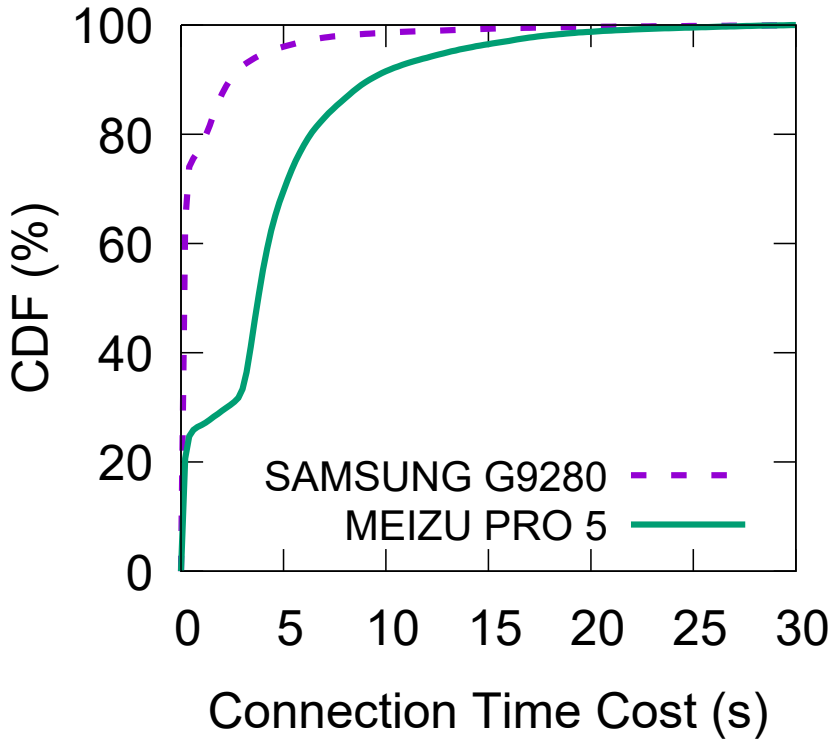


TABLE III: The software and hardware parameters for different kinds of mobile device models.

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

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 in average 1.5x larger than Samsung.
- RSSI has large RIG and the highest Kendall.

Correlation Analysis

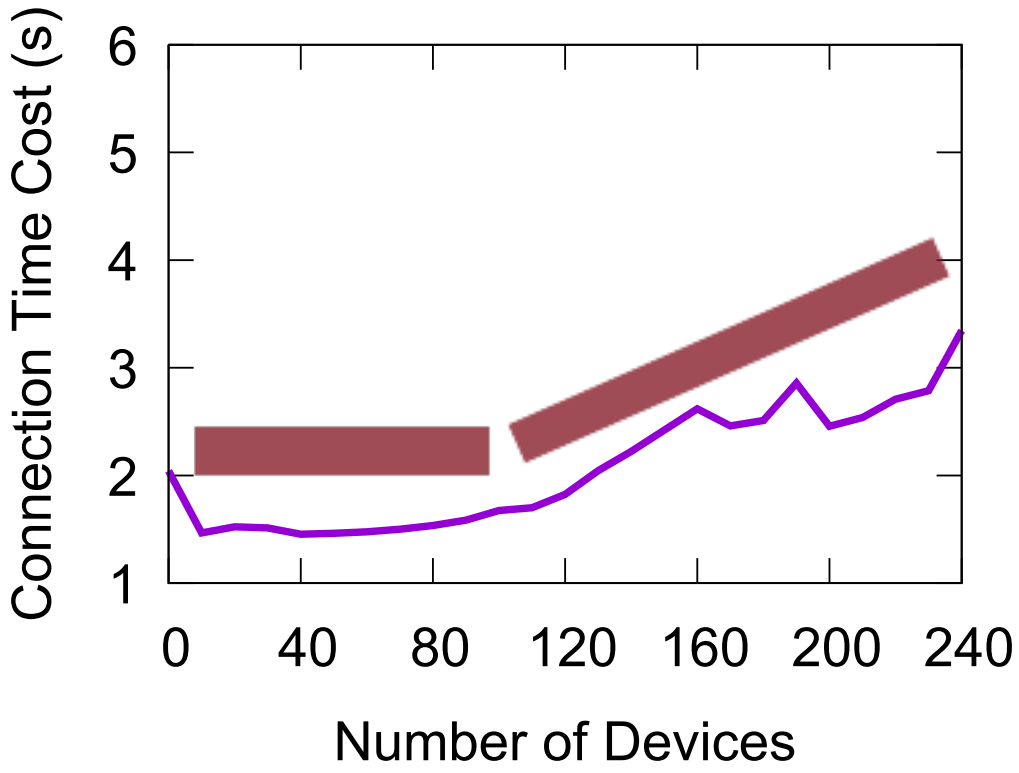
Features	RIG	Kendall
mobile device model	0.156	/
AP model	0.078	/
RSSI	0.020	-0.395
number of devices	0.006	0.208
hour of day	0.005	/

- Number of devices helps little.
  - Step function of number of devices

- Mobile devices and AP model has the highest RIG.
  - HTC on average 1.5x larger than Samsung.

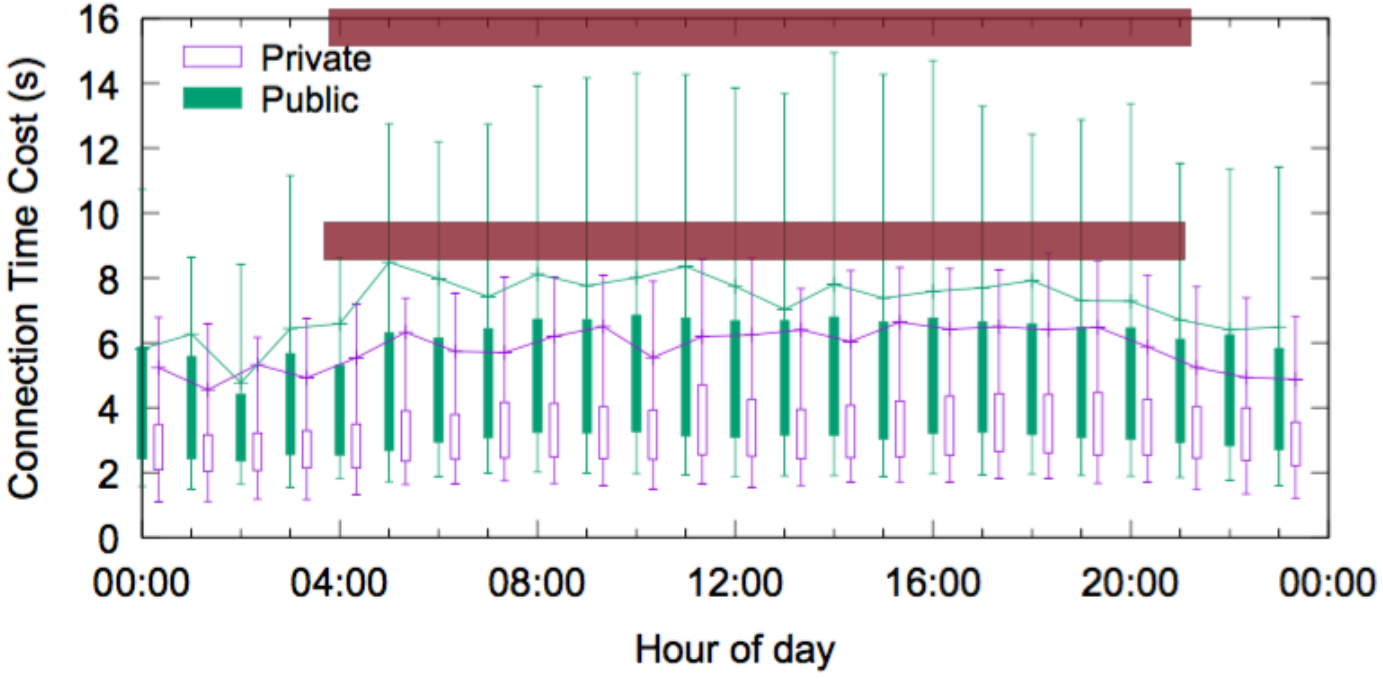
- RSSI has large RIG and the highest Kendall.

Visualization analysis



- The larger the number of associated devices is, the higher average connection time cost.

AP Model



- Private APs: APs which provide private WiFi services for a relatively small number of users.
- Public APs: APs which provide public/open WiFi services.
- Manually label 200K APs.

Motivation

Measurement

Results

Correlation

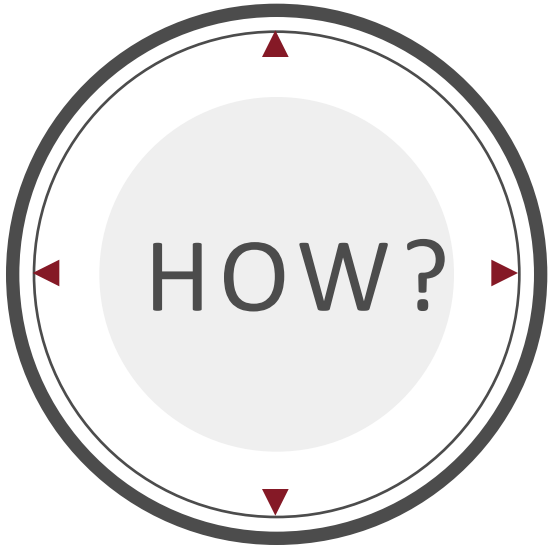
Analysis

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.
  - Online Learning
  - Prediction speed.

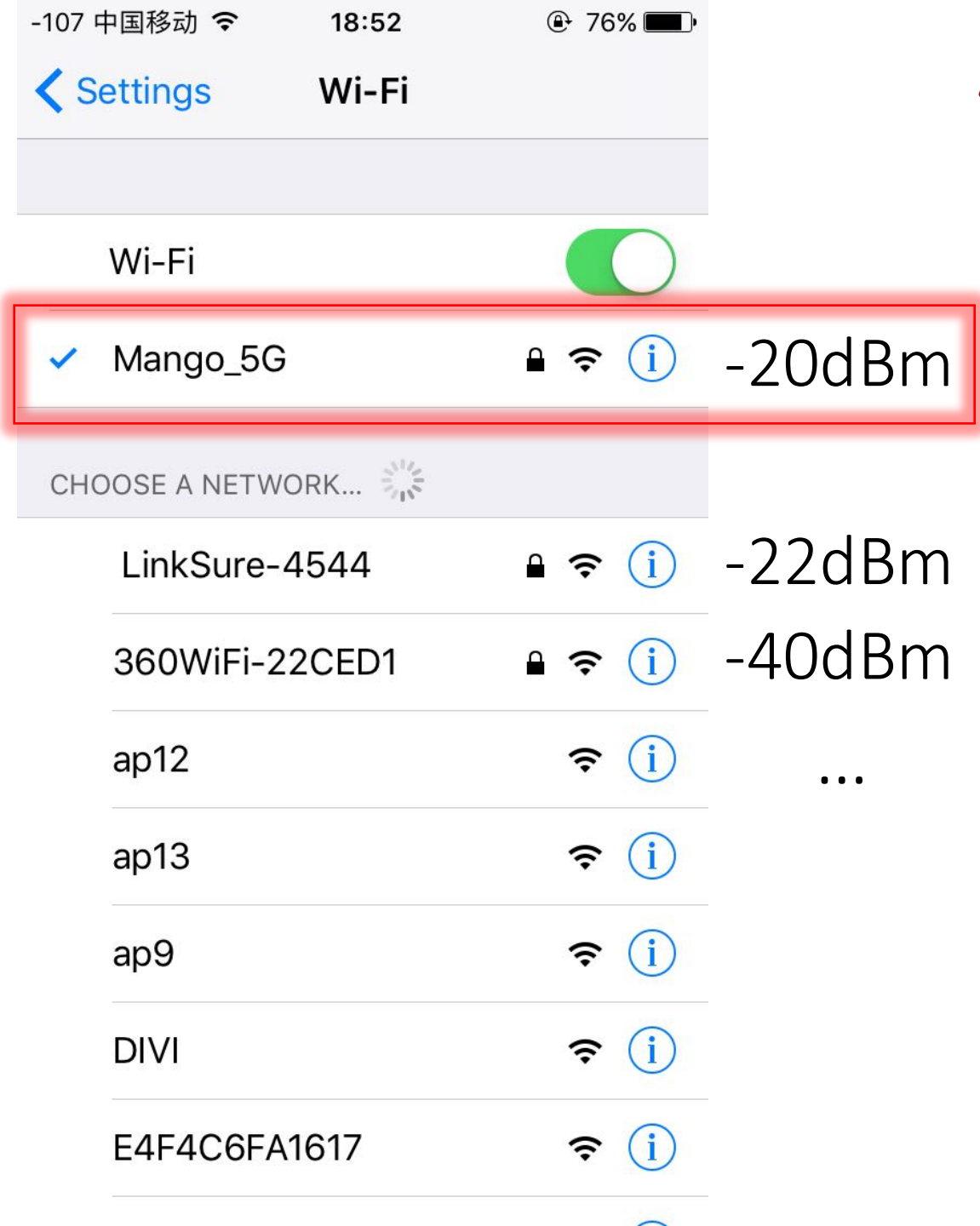
Machine Learning based Model

- Feature selection
  - All the features should be easily measured by mobile devices
  - Use as few features as possible under acceptable accuracy

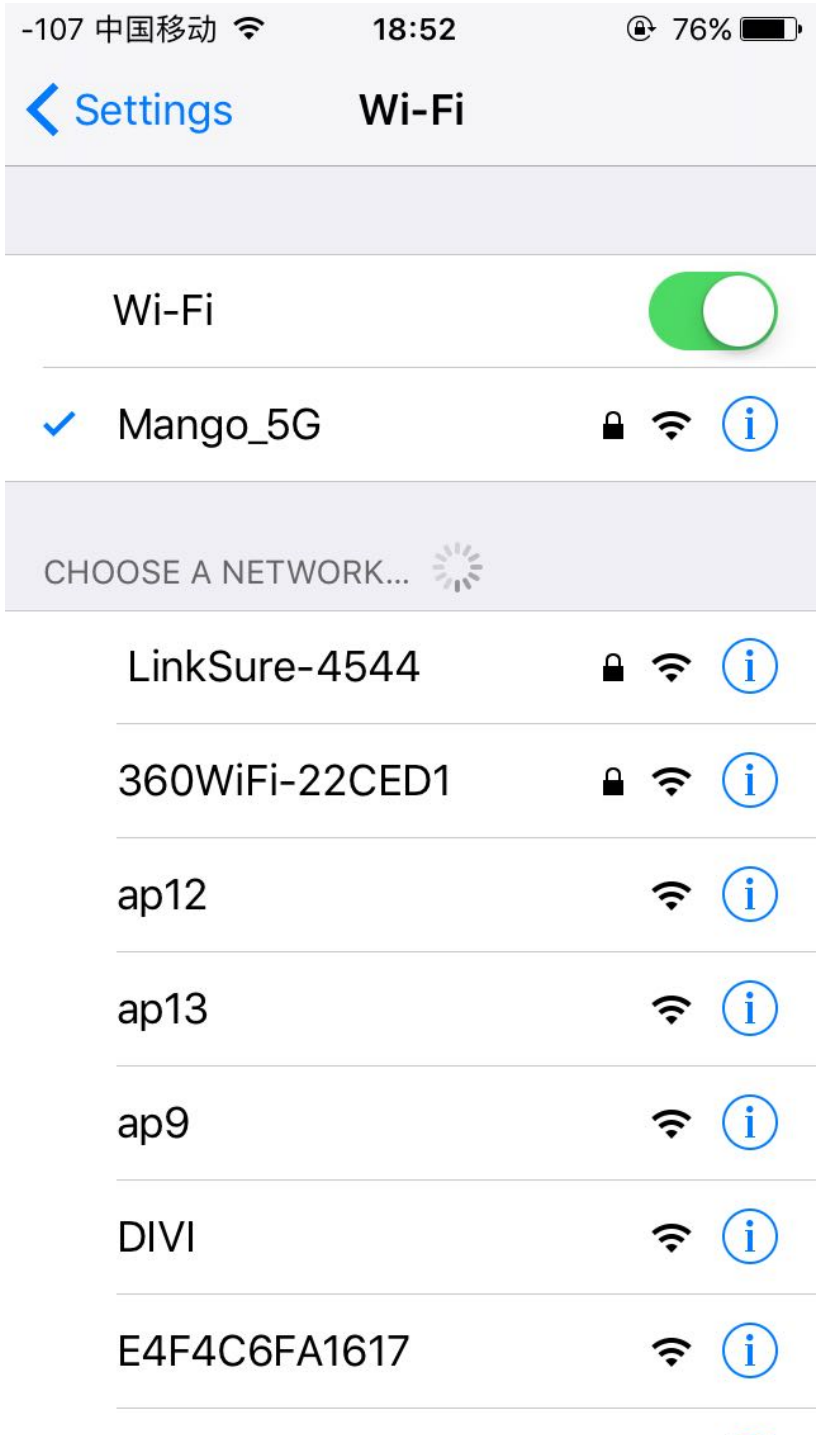
TABLE IV: Accuracy of random forest model. The parameters we use for this model are: Tree depth=90, #Tree=100, weight=0.3.

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**	

Strongest Signal Strength Algorithm



Enhanced AP Selection Algorithm



(12am, IPhone, Cisco, -20dBm, Yes)

(12am, IPhone, TP-Link,-40dBm, Yes)

(12am, IPhone, Hiwifi,-60dBm, No)

...

Enhanced AP Selection Algorithm

SLOW

SLOW

FAST

...

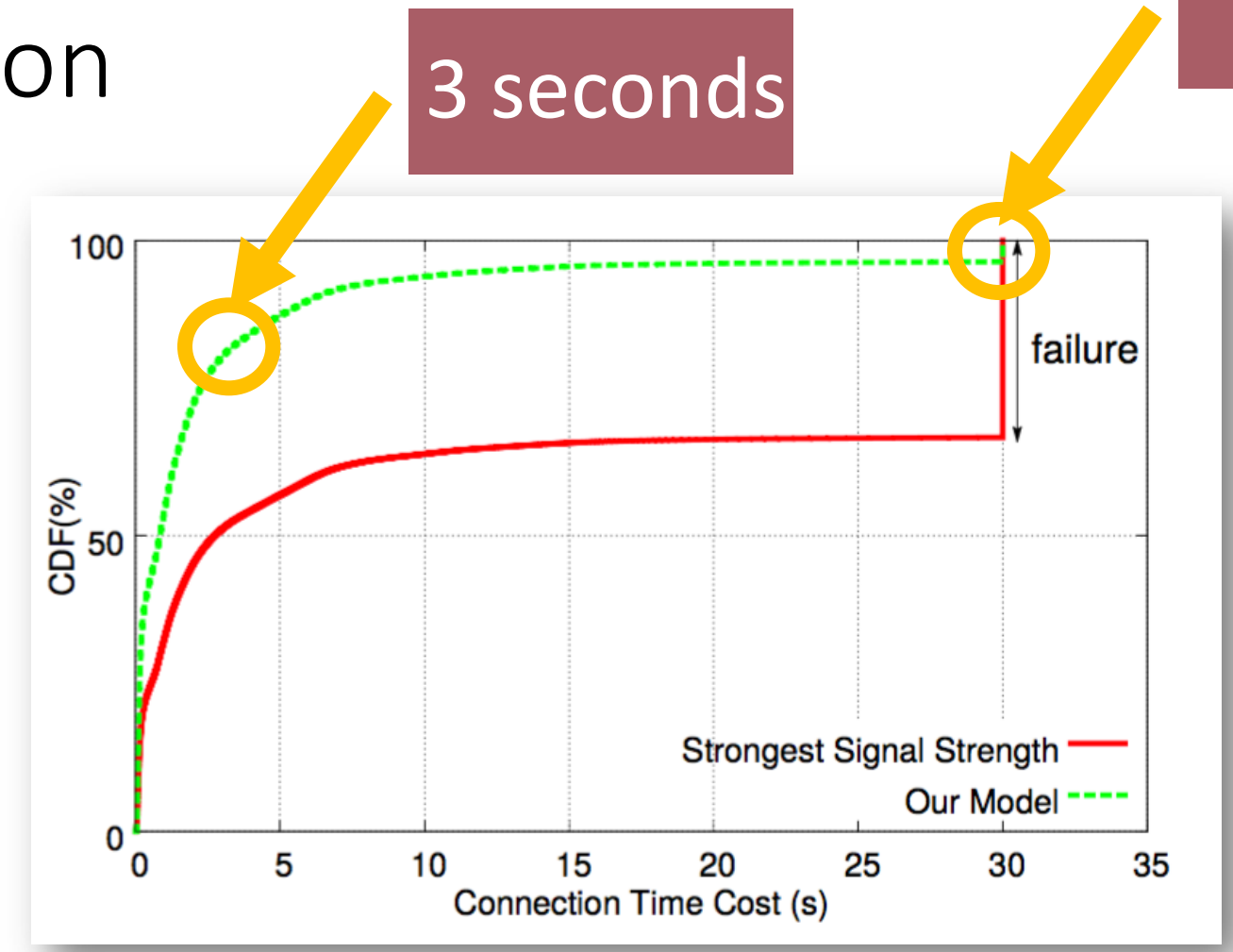
Evaluation

- Let the two algorithms work with the same dataset.
- Compare the time cost of the APs selected by different algorithms. (The cost is already known when certain device connects to certain AP in the dataset.)

Q: One device did not connect to all the neighbor APs!

A: We use the device whose 60 features are the same to approximate the connection time cost to each other!

Evaluation


3 seconds
3.6%
100
50
0
CDF(%)
failure
Strongest Signal Strength
Our Model
0
5
10
15
20
25
30
35
Connection Time Cost (s)

Conclusions

- WiFi connection set-up time cost is important but few works focus on it.
- Exhaustive real world measurement from a popular mobile WiFi manager App. 45% of the WiFi connection attempts fail.
- Using customized code to break down the whole process into different sub-phases for the first time.
- We propose a machine learning based AP selection algorithm to help users connect AP which shows great performance gain.

References

- [1]A. Patro, S. Govindan, and S. Banerjee. Observing home wireless experience through wifi aps. In *MobiCom*, pages 339–350. ACM, 2013.
- [2] S. Grover, M. S. Park, S. Sundaresan, S. Burnett, H. Kim, B. Ravi, and N. Feamster. Peeking behind the nat: an empirical study of home networks. In *IMC*, pages 377–390. ACM, 2013.
- [3] S. Sundaresan, W. De Donato, N. Feamster, R. Teixeira, S. Crawford, and A. Pescape`. Broadband internet performance: a view from the gateway. *SIGCOMM*, 41(4):134–145, 2011.
- [4] S.Sundaresan,N.Feamster,andR.Teixeira. Measuring the performance of user traffic in home wireless networks. In *PAM*, 2015.
- [5] Suranga Seneviratne, Aruna Seneviratne, Prasant Mohapatra, and Pierre Ugo Tournoux. Characterizing wifi connection and its impact on mobile users: practical insights. In *WiNTECH*, pages 81–88. ACM, 2013.

Thank you! Q&A?

changhuapei@gmail.com