

A Multi-dimension Measurement Study of a Large Scale Campus WiFi Network

Congcong Miao^{*†‡}, Jilong Wang^{†‡}, Hui Wang^{†‡}, Jun Zhang^{*}, Weiwei Zhou^{*}, Shengchao Liu^{*†‡}

^{*}Department of Computer Science and Technology, Tsinghua University

[†] Institute for Network Sciences and Cyberspace, Tsinghua University

[‡]Tsinghua National Laboratory for Information Science and Technology

Email: mcc15@mails.tsinghua.edu.cn, wjl@cernet.edu.cn, hwang@cernet.edu.cn

Abstract—The growing trend of wireless devices and WiFi networks poses significant management challenges to network administrators. Characterizing WiFi user behavior and understanding WiFi network usage pattern are helpful to identify the management challenges so that network administrators could manage WiFi networks more efficiently. In this work, we collect comprehensive datasets, i.e. DHCP dataset, AAA dataset, SNMP dataset of ACs in a large campus WiFi network. We provide a detailed measurement study from multiple dimensions, i.e. server plane, temporal plane, spatial plane and traffic plane. We observe that the WiFi network under study is far from optimal. First, the phenomenon of IP waste is severe due to the isolation between DHCP server and AAA server. Second, current deployment of network infrastructure resources is based on network administrators' experience and it results in that the WiFi performance varies a lot across different areas. Furthermore, we also study the user behavior with different types of devices and in different kinds of buildings. Our observations indicate that the WiFi network could be improved and managed more efficiently from multiple dimensions. We believe that this measurement study is helpful for network administrators and researchers to understand more about large scale WiFi networks.

Index Terms—User Behavior, Measurement, Multiple Dimension, Network Management

I. INTRODUCTION

In recent years, wireless devices, *e.g.* smartphones, tablets, and laptops, are becoming more ubiquitous due to the convenience for mobility. The International Data Corporation (IDC) Forecast indicates that smart connected device market shared by wireless devices will be accounting for more than 90% by 2018 and the penetration continues to grow in the future [1]. As wireless devices are more pervasive, 802.11 (WiFi) networks are becoming increasingly popular, especially in campuses and enterprises. Compared to cellular networks, WiFi networks are more preferred by wireless devices for mobile Internet access because they provide higher bandwidth and lower delay [2], [3].

The growing trend of wireless devices and WiFi networks poses significant management challenges to network administrators. Compared to wired desktop PCs, the wireless devices connecting to WiFi networks introduce quite different user behaviors. For example, some WiFi devices have already owned an IP address, however they may re-apply a new IP address when they roam from one place to another. As a result, these users own two IP addresses at the same time.

This phenomenon may exhaust all IP addresses even if the number of IP addresses in IP pool is more than the number of devices. Furthermore, understanding network resource demand of WiFi devices is also helpful for planning and deploying infrastructure resources of WiFi networks. For example, the AP resource would be wasted if the service capability of APs is much more than the demand of clients, and the service performance will be seriously degraded if the service capacity of APs does not meet the demand of clients. In summary, network administrators should understand the behavior of WiFi users in order to manage WiFi networks more efficiently.

In this paper, in order to study the behavior of WiFi users comprehensively, we use real network trace from a large campus with more than 50,000 users and 100,000 public IP addresses. We deploy SNMP manager to capture the SNMP data of ACs. We also collect the IP lease log and IP authentication log from DHCP server and AAA server respectively. We define these network datasets as SNMP dataset, DHCP dataset and AAA dataset respectively. These datasets provide very comprehensive information for us to study campus WiFi network.

Based on these datasets, we are able to characterize the WiFi user behavior and understand the WiFi network usage pattern from network administrators' point of view. Specifically, we study the behavior of WiFi users from multiple dimensions, *i.e.* server plane, temporal plane, spatial plane and traffic plane. To best of our knowledge, no prior work has focused on characterizing WiFi user behavior from multiple dimensions, with network administrators' point of view towards managing WiFi network more efficiently. Based on our work, we find that the WiFi network under study is far from optimal and it could be improved in multiple ways. Particularly, our main observations and contributions can be summarized as follows:

- 1) DHCP server and AAA server should cooperate to reduce the waste of IP addresses. We observe that 66% of lease sessions are *Squatters* (acquiring IP addresses but never launching authentication) and 31% of authentication sessions are *Idlers* (taking more than one lease time to authenticate after acquiring IP addresses) and 8% of users are *Greedy* users (re-applying a new IP address although they have already owned an IP address). These phenomena would cause severe waste of IP addresses and may exhaust all IP addresses in the pool.
- 2) Network administrators should take the WiFi user be-

havior into consideration while deploying the network infrastructure resources. We find that the deployment of AP resource which is planned based on their experience is not satisfactory. Specifically, the APs are intensively deployed in dorm buildings, and each AP in dorms only serves 2.3 devices on average. The resource is wasted because the service capacity of APs is much more than user demand. However, each AP in cafeteria should serve more than 70 devices which causes service performance degradation in that area.

3) Network administrators could deploy network resources and manage WiFi networks more efficiently based on our quantified observations. Specifically, everyday users are likely to be high traffic users, so prioritizing network resources to everyday users is desirable. The authentication session traffic patterns could be classified into 3 major traffic levels (low, middle and high) with different types of devices and in different kinds of buildings. Network administrators would be able to predict the traffic demand and manage network at a fine-grained level.

The rest of this paper is organized as follows. Section II describes the campus WiFi network under study and the collection and pre-processing of our measurement datasets. In section III, we study in details the characteristics of WiFi user behavior from multi-dimension and *explore policies for managing WiFi network efficiently*. Section IV presents the related work and we conclude the paper in Section V.

II. OUR CAMPUS WiFi NETWORK AND DATASETS

The architecture of T campus WiFi network is shown in Figure 1. Our measurement study is based on real network trace collected from three collection points inside T campus WiFi network. Our campus WiFi network is based on widely used Access Controllers (ACs) and fit Access Points (APs) structure that 15 ACs control more than 8,000 APs in 131 buildings. All APs share the same Service Set Identifier (SSID), allowing WiFi users auto-connection when roaming in campus. There are two steps, *i.e.* the step of acquiring IP address and the step of authenticating IP address, when WiFi users connect to the WiFi network to access the public Internet.

Acquiring IP Address: Dynamic host configuration protocol (DHCP) [4] is widely used for automatically managing IP address in the WiFi network. The network operators deploy three DHCP servers which manage 100,000 public IP addresses in our campus WiFi network. Each DHCP server provides DHCP service for one geographical area. When one mobile device roams to one area and detects the campus WiFi network signal, if its "auto-connection" is *ON* for this SSID, the device would automatically send *request* message to the DHCP server of this area to get an IP address. The DHCP server then assigns an IP address to the user with a default IP lease time. In T campus, the default lease time is set to be 30 minutes. After getting an IP address lease from one DHCP server, the client needs to extend its lease by sending a *renew* message periodically after half of a lease time. The extension of lease is usually done by the operating systems on the mobile devices. The release of IP addresses can happen

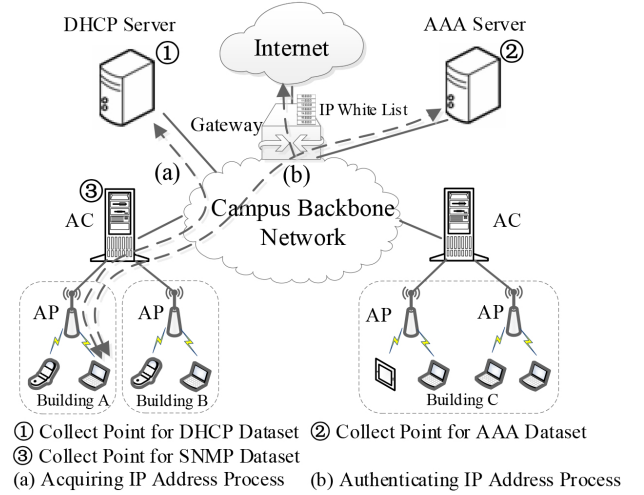


Fig. 1. Architecture of T Campus WiFi Network

in two kinds of cases. First, the WiFi user sends an explicit *release* message to the DHCP server when it does not need to occupy that IP address any more. Second, if the WiFi user leaves the service area of DHCP, it would not extend its lease time periodically. In this case, the lease time expires, and the server reclaims that IP address and logs an *expire* message. In this paper, we refer to the time period from the acquirement of an IP address to the release of this IP address as the *IP lease session*.

Authenticating IP Address: Authentication, Authorization, Accounting (AAA) [5] is used for controlling access to public Internet in campus T. After acquiring an IP address, the client should authenticate in AAA server to access the public Internet. The campus WiFi network has a border gateway to control the data packet flows to the public Internet using an IP white list. Only data packets of IP addresses in the list can go through the border gateway. Therefore, one WiFi user has to send authenticating request to our AAA server with its user account and its IP address before it visits the public Internet. Then all data packets of this IP address would be accounted into this user account. This binding of IP address and user account will be unbound in two cases. First, the user decides to stop accessing the public Internet and it sends a *logout* message to the AAA server explicitly. Second, if the user has not generated traffic for latest 6 hours (default set by network administrators), the AAA server will automatically log out that IP address. In this paper, we refer to the time period from the binding of the IP address and user account to the unbinding as the *IP authentication session*.

A. dataset

The datasets contain a typical week records from Saturday May 20 to Friday May 26, 2017. We collect IP lease log and IP authentication log from DHCP server and AAA server respectively. We also deploy SNMP manager to capture the SNMP data of ACs. For auxiliary dataset, we consult with network administrators to acquire AP-building map to learn

TABLE I
BUILDING TYPE, BUILDING COUNT AND AP COUNT

Building Type	Building Count	AP Count
Classrooms	13	287
Departments	44	1183
Offices	9	108
Cafeterias	6	26
Gyms	4	48
Dorms	52	6390
Libraries	3	73
Total	131	8115

the location of each AP. The datasets include DHCP dataset (about 33,600,000 entries), SNMP dataset (about 15,600,000 entries), AAA dataset (about 500,000 entries) and auxiliary dataset (about 8,000 entries). These datasets provide very comprehensive information for us to study campus WiFi network. The datasets will be described below in detail.

a. DHCP dataset: Our DHCP dataset is collected from 3 DHCP servers. The dataset is in fact the log file which records all interactions between clients and DHCP servers. Each entry in the dataset logs a DHCP activity, *e.g.*, *Assign*, which means the DHCP server assigns an IP address to a client, *Renew*, which means the DHCP server receives a request from one client to extend lease time, and *Expire*, which means a DHCP server reclaims the IP address when a lease expires [6]. For each activity, our dataset also includes the detailed information of the activity, such as timestamp, IP address, and the device information, *i.e.* host name, MAC address and vendor name.

b. SNMP dataset: SNMP is widely deployed for network monitoring and it is already configured in AC and AP hardware [7]. We set up a SNMP manager program to poll SNMP data every 5 minutes from 15 ACs. Each AC would provide a list of clients that are connected with one of APs under the control of the AC. In detail, each timestamped entry in our dataset is in the format of (AP name, AP MAC, client IP, client MAC), which means the client is connected with the AP.

c. AAA dataset: AAA protocol [5] is widely used in today's campus network, however rarely used for studying campus WiFi network. To best of our knowledge, we are the first to use AAA logs to study the campus WiFi user behavior. Each entry in our AAA dataset records the information of an IP authentication session, which includes user login name, login time, logout time, and accumulated download bytes during this period. We use the accumulated download bytes to study traffic pattern of WiFi users in III-D.

d. Auxiliary dataset: We consult with network administrators to learn the locations of each AP, *i.e.* map each AP name to its located building. As demonstrated in previous work [7], [8], the characteristics of WiFi user behavior are aggregated and influenced by building functions. Therefore, we divide the 131 buildings in our datasets into 7 categories, *i.e.*, classrooms, departments, offices, cafeterias, gyms, dorms and libraries. Table I lists the building type, building count and AP count in each building type. We observe that the APs are intensively deployed in dorms while the number of APs in cafeterias is

TABLE II
KEYWORDS FOR DEVICE TYPE DETECTION

	Host Name Keywords	Vendor Name Keywords
Smart-phone	iPhone, Sumsung, Huawei, Blackberry, Xiaomi, Meizu	Android, Huawei, Blackberry, dhcpd
Laptop	MacBook, Lenovo, Dell, HP, ThinkPad, VAIO, ACER, MSI	MSFT
Tablet	iPad, iPod, Xiaomipad	-

very small. Combining with the AP name in SNMP dataset, we are able to get the location of each connected device.

B. Data Pre-processing

a. Data Anonymization and correlation: It is worth pointing out that privacy issues of these datasets are seriously considered and measurements are taken to protect the privacy of WiFi users. The collected datasets are anonymized to remove all sensitive personally identifiable information before they are used for our study. We employ prefix-preserving anonymization as proposed in [9]. Specifically, we anonymize the IP addresses, the user login names, the MAC addresses of devices and the MAC addresses of APs. The anonymization methods and parameters are kept consistent over all traces to match users. Since all network datasets contain time information and IP address information, we use these two fields to correlate the user in all datasets.

b. Device type detection: Device type detection is a very important step to study user behavior in campus WiFi network since the network behavior of WiFi users varies in different device types [10], [11]. We limit our work into three device types, *i.e.* smartphones, tablets and laptops, which are mostly used in our daily life.

In DHCP dataset, each entry records the information of associated device, including host name, MAC address and vendor name, which can be used to infer the type of device. First, we take advantage of host name to identify device type based on keywords searching. For example, host names like "songdeiPhone", "Dongde-iPad" or "LY-MacBook-Pro" have keywords, the presence of which means smartphone, tablet or laptop. Then it is assisted by vendor name if host name is blank or useless for our purpose. Vendor names like "Blackberry" or "MSFT" can be inferred as smartphone or laptop. Lastly, we use MAC address to determine the vendor of interface [12] and then identify device type if the vendor of device can be directly mapped to a specific type of device. Some example host name keywords and vendor name keywords are shown Table II.

Using the fields in DHCP dataset, we have identified 87% of devices (The total device number is 87,073). Among the 75753 identified devices, the number of smartphones, laptops and tablets are 46406, 26362, 2985, accounting for 61%, 35%, 4% respectively.

III. STUDYING THE CAMPUS WiFi NETWORK FROM MULTI-DIMENSION

With the comprehensive datasets, we develop a systematic approach to study the campus WiFi network from multiple dimensions. The goal is to characterize the WiFi user behavior and understand the WiFi network usage pattern from network administrators' point of view. Our approach provides insights of WiFi network and reveals the potential problems of campus WiFi network under study.

The procedure of our approach is as follows. First, we mirror the WiFi network datasets and store them into the local database. Then, we run analysis process on the local database and the process generates a series of corresponding profiles. The analysis process analyzes the data from four dimensions, *i.e.* server plane, temporal plane, spatial plane and traffic plane. Each plane contains several perspectives. The generated profiles are stored in profile module. According to the profiles, network administrators could devise new policies to improve the WiFi network. The whole procedure will be run periodically to update the corresponding profiles to perform policies in time due to the constantly changing WiFi network usage patterns.

We summarize our main findings as follows.

a. Server Plane

- 1) 66% of IP lease sessions are *Squatters* and 31% of authentication sessions are *Idlers* which will result in the severe waste of IP address.
- 2) 8% of IP lease sessions are *Greedy* users which may exhaust the IP addresses pool.

b. Temporal Plane

- 1) The users appearing on the network follow periodic patterns and 18% of users are *everyday users*.
- 2) The logout actions of users could be divided into two parts: *active logout* and *non-active logout* and the laptop users are more likely to log out actively.

c. Spatial Plane

- 1) The WiFi performance varies a lot across different areas. Surprisingly, each AP in cafeterias should serve 71.2 devices during peak hour.
- 2) The proportion of smartphones is dominant in cafeteria and gym.

d. Traffic Plane

- 1) The one-month traffic generated by one user varies from 977 KB to 224 GB and 5% of heavy users generate more than 40 GB WiFi traffic. Surprisingly, almost all heavy users are *everyday users*.
- 2) The session traffic pattern could be classified into 3 major levels with different types of devices and in different kinds of buildings.

Now we describe our data analysis and the policies to improve the performance of WiFi network from four dimensions in detail.

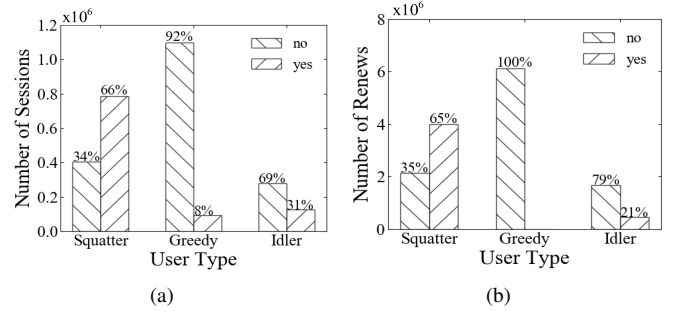


Fig. 2. (a). The session number and proportion of squatters, greedy users and idlers. (b). The number and proportion of *renew* messages sent by squatters, greedy users and idlers.

A. Server Plane

In the server plane, we mainly study the procedure of clients getting the IP addresses from DHCP server and authenticating the IP addresses they have got to access public Internet on AAA server. The results are shown in Figure 2, and will be described below in detail.

Squatter: Squatters are defined as clients who acquire an IP address, however, do not transfer any data over the campus WiFi network because they have not authenticated. The existence of squatters brings out two problems. First, it results in the severe waste of IP addresses. In Figure 2(a), we surprisingly find that there are about 780,000 IP lease sessions that are not authenticated, occupying 66% of total IP lease sessions. The IP authentication session only accounts for 34% of the total IP lease session. Second, it seriously increases the load in DHCP servers. The squatters periodically send *renew* messages to DHCP servers to extend the IP lease. In Figure 2(b), the invalid *renew* messages account for 65% of total messages which are about twice as many as messages sent by authenticated users.

Therefore, we propose that it is valuable to introduce a MAC white list that records client MAC addresses which are often authenticated in our WiFi network. The DHCP servers only automatically assign IP addresses to these devices in the white list. In this way, we can avoid assigning IP addresses to potential squatters, such as visitors who can never authenticate through AAA server.

Greedy: Greedy users are defined as those who have already owned an IP address, however re-apply a new IP address when they roam from one place to another although the previous IP address is not expired. In other words, greedy users own two or more than two IP addresses at the same time. In Figure 2(a), we observe that the number of greedy users is about 100,000, taking up 8% of total IP lease sessions. To some extent, the IP address waste caused by greedy users is more serious than other phenomena, because theoretically greedy users can exhaust all IP addresses even if the number of IP addresses in a large IP pool is more than the number of online devices.

Hence, we suggest that the DHCP servers in the campus WiFi network should cooperate to avoid the waste of IP

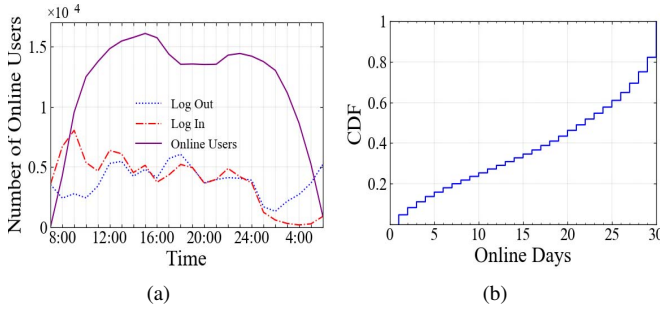


Fig. 3. (a). The intra-day behavior of users in each hour. (b). The number of days that user appears on WiFi network for one-month dataset.

addresses. Specifically, network operators should design a central controller which maintains a MAC-IP address list to monitor the IP requests of all clients. Once one DHCP server receives an IP request from a client, the central controller checks the MAC-IP address list of that device. If the device has already occupied an IP address, the controller notifies the other DHCP server to release the IP address. Finally, the DHCP server assigns a new IP address to that device. This method guarantees that one device could only own one IP address at the same time.

Idlers: Idlers refer to authenticated users who do not launch authentication immediately after acquiring IP addresses. It means they always take a relatively long period between acquiring IP addresses and launching authentication. Here we use one lease time (30 minutes) as a threshold. In other word, if the time interval between the client acquiring an IP address and authenticating that IP address is more than one lease time, the client is regarded as idler. In Figure 2, we find that there are about 130,000 idlers, accounting for 31% of total authentication sessions, and these idlers introduce the 21% invalid *renew* messages.

As a result, we suggest that DHCP server and AAA server should communicate with each other. Specifically, DHCP server sends a notifying message to AAA server after assigning an IP address to the client. The AAA server checks that IP address regularly and notifies DHCP server if that IP address has not been authenticated for one lease time. If the DHCP server receives that message, it then reclaims that IP address.

Summary: We find that there are three types of users, i.e. *squatters*, *greedy users* and *idlers*, which cause the waste of IP addresses and increase the server load. Based on these observations, we suggest that 3 DHCP servers in our campus WiFi network should cooperate to avoid the waste of IP addresses, and AAA server and DHCP server should exchange messages to reduce the waste of IP addresses.

B. Temporal Plane

In temporal plane, we mainly focus on the temporal characteristics of user behavior, including the daily patterns and regularity and session durations. To study weekly behavior of users, we extend AAA dataset to cover one month from May 15th (Monday) to June 14th (Wednesday). Our observations can be summarized as follows.

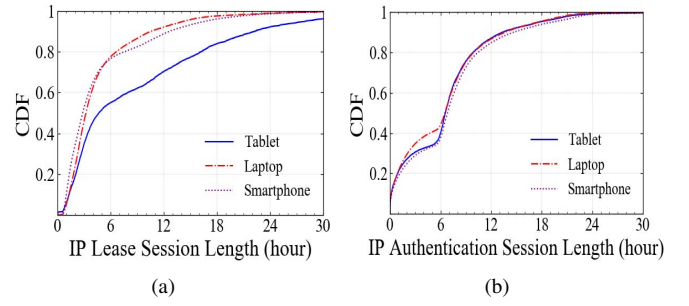


Fig. 4. (a). CDF of the IP lease session length with different device types. (b). CDF of the IP authentication session length with different device types.

Daily patterns and regularity:

From the one-month AAA dataset, we observe that the appearance of devices on the WiFi network follows weekly pattern: the number of devices increases rapidly on Monday, reaches peak on Thursday, and starts to decrease on Friday. We then study the intra-day behavior to manage the network resource demand on an hourly basis. Figure 3(a) depicts the number of login users, logout users and online users in each hour. We observe that the number of online users reaches a peak point at 15:00 and drops from 16:00 to 21:00. Then the number of online users rises again after 21:00 and reaches the other peak point at 23:00. After that, the number of users keeps dropping until the next day morning.

We further investigate the number of days that the users appear on the campus WiFi network. We find that there are 18% of users who appear on the WiFi network every day. We then categorize users into two types: *everyday users* and *non-everyday users*. We will further observe the traffic behavior of the everyday users in subsection D.

Understanding periodic online pattern and hourly network usage pattern, network administrators could take advantage of these patterns to predict network resource demand and manage network resources on an hourly basis.

Session Length of Different Device types: In order to study IP lease session and IP authentication session, we plot CDF curves of session length with different device types in Figure 4(a) and Figure 4(b).

For IP lease session, we observe that smartphones and laptops have the similar characteristics that in 20% of occasions, the session length is more than 6 hours, however, for tablets, the session length is more than 16 hours. For IP authentication session, there is an interesting observation that authentication session length could be divided into two types with 6 hours as a boundary. Note that the default set of logout time in AAA server is 6 hours. We can infer that the authenticated users could be classified into two parts: *active logout* and *non-active logout*. We observe that smartphones and tablets have the similar characteristics of IP authentication session. The users are more likely to logout actively on laptops than other device types. In 45% of occasions, the authentication session length is less than 6 hours for laptops. We find in both lease and authentication sessions, there are a small part of outliers whose session lengths reach more than one day.

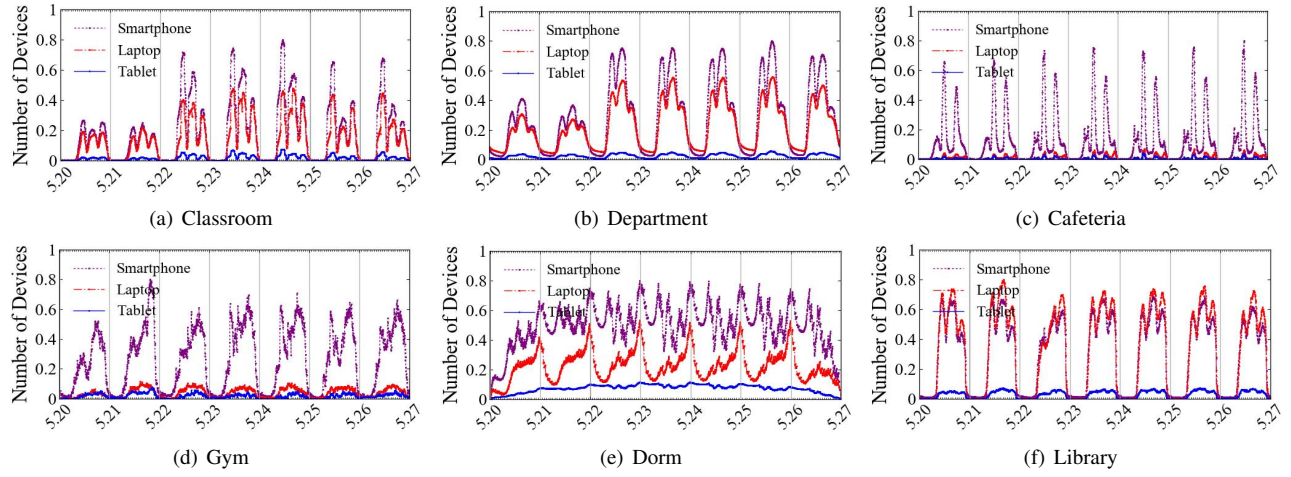


Fig. 5. The proportion of three device types in different building types (normalization)

Identifying outliers and understanding the session length with different device types, network administrators could devise the policies pointedly. For example, the network administrators could establish a (MAC, patter) table to record the user behavior, and then reduce the default logout time to the users who do not log out actively.

Summary: (i) We find that the temporal behavior of WiFi users follows the periodic patterns and every-day users account for 18% of total users. (ii) For IP authentication session, we find that the users are classified into two parts: active logout and non-active logout, and users are more likely to logout actively on laptops. Based on these observations, we suggest that network administrators could predict network resource demand, and manage network on an hourly basis. They could set up shorter lease time to the outliers and shorter logout time to non-active logout users.

C. Spatial Plane

In the spatial plane, we study the WiFi performance and behavior of devices in different building types. Different from traditional work [7], [8], we study the number of each device type in different building types, defined as *Building Device Proportion*. We also propose to use the number of devices that one AP serves to study the campus network, defined as *AP Service Quality*.

Building Device Proportion: We would like to know whether the proportion of three device types is the same in different building types. In Figure 5, we plot the number of three devices in different buildings. Y axis represents the number of devices in each hour, which is normalized by the maximum value. Since the proportion of three device types in office is similar to the proportion in departments, we omit the figure in offices due to page limit.

From the figure, we have the following observations. First, the number of devices follows daily pattern in different building types. Second, we can see that the proportion of three device types is very different across different building types. The number of tablets is very small in all building types. Let

TABLE III
THE MAX AND AVE RATIO OF DEVICES AN AP SHOULD SERVE AND THE RUSH HOUR IN DIFFERENT BUILDING TYPES

Building types	AP count	Max count (rush hour)	Ave count	Max ratio	Ave ratio
Classrooms	287	5968 (10:00-11:00)	1849	20.8	6.4
Departments	1183	10442 (15:00-16:00)	4998	8.8	4.2
Offices	108	661 (11:00-12:00)	292	6.1	2.7
Cafeterias	26	1853 (12:00-13:00)	341	71.2	13.1
Gyms	48	248 (17:00-18:00)	122	5.2	2.5
Dorms	6390	27143 (23:00-24:00)	15003	4.3	2.3
Libraries	73	1207 (16:00-17:00)	473	16.5	6.5

us skip the tablets and focus on the number of smartphones and laptops. We observe that the number of smartphones is much larger than the number of laptops in cafeteria and gym. On the other hand, the number of laptops in turn is more than the number of smartphones in library.

Previous researches have known that traffic characteristics and online times of different device types are quite different [13], [14], [15]. Now we find that the device proportion of different buildings is diverse. So network operators can optimize their WiFi network performance for particular device type in different buildings. For example, in gym and cafeteria, operators can only consider to improve the WiFi network performance for smartphone applications. While in other five building types, both smartphone applications and laptop applications should be considered during network management.

AP Service Quality: To study AP service quality, we need to know the maximum number and average number of devices using WiFi service and the number of APs in each building. We define *max ratio* as the ratio of maximum number of devices to number of APs and *ave ratio* as the ratio of average number of devices to number of APs. Table III lists the max ratio and ave ratio of different building types.

We observe that the WiFi performance varies a lot across different areas. Specifically, the APs are deployed intensively in dorm buildings, so that each AP in dorms should only serve 2.3 devices on average. The AP resource and electricity energy

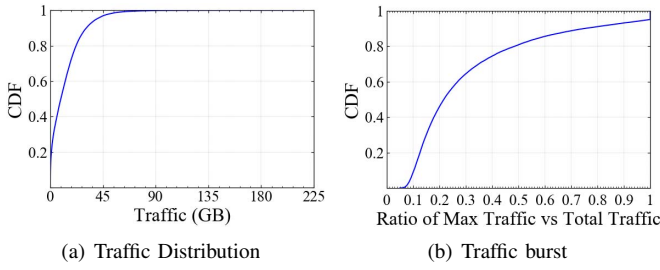


Fig. 6. The CDF of Traffic Volume and Traffic Burst

are wasted because the service capacity of APs is much more than user demand [16]. However, an AP in cafeteria should serve 71.2 devices in maximum and 13.1 devices on average. Many users connecting to an AP will lead to the AP overload, resulting in less bandwidth for each user and longer access delay due to CSMA-CA protocols [17].

We also study the time when the maximum concurrent occurs, shown in third column of Table III. There is one interesting observation that the peak hours of different building types are distributed with different time points. For example, the peak hours of classroom, office, cafeteria, gym and dorm are 10:00-11:00, 11:00-12:00, 12:00-13:00, 17:00-18:00 and 23:00-24:00 respectively, which follows the pattern of campus people daily life.

Based on these observations, we would like to pinpoint two possible ways to improve WiFi network performance. First, network administrators should deploy network infrastructure resources, *i.e.* APs, based on user behavior rather than their own experience. Second, since the network resource demand is staggered shifts among different building types, network administrators could provision network resources more efficiently.

Summary: (i) We observe that the proportion of three device types is quite different in different building types and smartphones is dominant in cafeteria and gym. (ii) The deployment of APs based on network administrators' experience is not satisfactory. It causes that the WiFi performance varies a lot across different areas. Based on our observations, we suggest that network operators could take the WiFi user behavior into consideration while deploying the network infrastructure resources. They could manage network resources pointedly in rush hours of different building types. Further more, they could optimize WiFi network performance for particular devices in different buildings.

D. Traffic Plane

In the traffic plane, we focus on the traffic behavior of users. We firstly analyze traffic volume and traffic distribution of each user, and then further step into the session traffic behavior in campus WiFi network. In Section III-B, we extend the AAA dataset to study the weekly behavior of users. In this part, we use the AAA dataset to study the traffic behavior of WiFi users.

Traffic Volume: We study the month-long accumulated traffic volume of every user and plot CDF curves in Figure

TABLE IV
THE DETAILED COMBINATION OF 3 TRAFFIC VOLUME PATTERNS, * MEANS MATCHING ALL DEVICE TYPES OR BUILDING TYPES

Index (#)	Traffic Pattern Combinations (Device, Building)	Traffic Volume Pattern (Level)
1	(Smartphone, Cafeteria), (Tablet, Classroom)	7 MB (Low)
2	(Laptop, Cafeteria), (Tablet, Department), (Smartphone, Classroom), (Smartphone, Department), (Smartphone, Gym), (Smartphone, Office), (Smartphone, Library), (Laptop, Classroom)	18 MB (Middle)
3	(* , dorm), (Laptop, Library), (Laptop, Department), (Laptop, Office), (Laptop, Gym), (Tablet, Office), (Tablet, Library)	45 MB (High)

6(a). We can see that the traffic volume is diverse significantly across users, *i.e.*, the traffic volume of one user ranges from as little as 977 KB to as large as 224 GB. About 5% of users generate more than 40 GB WiFi traffic for one month, defined as *heavy users*. We further study the traffic behavior of everyday users and find that traffic volume of everyday users is more than 25.0 GB on average, compared to 13.5 GB of all users on average. Furthermore, we observe that almost all heavy users are everyday users. Everyday users are likely to be high traffic users who require more networking resources than other users so that classifying such group of users and prioritizing network resources to everyday users are desirable.

We further study how the user traffic is distributed among all days in the month, and calculate the ratio of the maximum daily traffic to the total monthly traffic for each user. We plot the CDF of the ratios of all users in Figure 6(b). We find that 19.1% of users generate half of their total monthly traffic in a single day and more than 50% of users consume around 1/4 of total monthly traffic in a single day. Surprisingly, about 5% of users generate all their monthly traffic volume in a single day. This can be explained by the reason that the client download large softwares or videos which account for large amounts of traffic for a few days. The result strongly indicates that daily traffic generated by users is not steady and traffic burst is a common phenomenon. Network operators should take traffic burst into consideration for better WiFi network services.

Authentication Session Traffic Behavior: A natural question to ask is whether the traffic volume during the entire authentication session is the same for all authenticate sessions. We find that it is not. The session traffic volume is as large as 10 GB for the most and as small as 0 byte for the least during an authentication session. We define the *session traffic* to represent the traffic volume generated in that authentication session. Previous work has shown that device types are inclined to be used for different services and building types have significant influence on user behavior [8], [10]. In order to quantify the session traffic more specifically, we combine 3 device types and 7 building types which have 21 combinations in total. We omit 2 combinations, *i.e.* (Tablet, Gym) and (Tablet, Cafeteria), because they occupy a very small proportion of the total sessions. We then plot the CDF of the session traffic for 19 combinations respectively and find

some of combinations have a similar traffic patterns (The CDF figures are omitted due to page limit).

We use the median of session traffic in each combination to represent the traffic volume pattern for that combination and the distribution of traffic volume patterns is shown in Table IV. We observe that there are 3 major traffic levels with different types of devices and in different kinds of buildings. We observe that the session traffic generated by smartphones in cafeterias or by tablets in classrooms is inclined to be very low. However, the session traffic generated in dorm or by laptop (except in cafeteria area) is inclined to be very high. Note we are the first to quantify the session traffic volume and classify sessions to study the traffic behaviors.

Summary: (i) We observe that almost all heavy users are everyday users. (ii) Traffic volume generated by users is not steady day to day and traffic burst phenomenon is a common phenomenon. (iii) We quantify the session traffic behavior and observe that traffic volume could be classified into 3 major levels with different types of devices and in different kinds of buildings. Network administrators can learn the traffic pattern, predict the traffic demand of different devices in each area, and then provision network resources at a fine-grained level.

IV. RELATED WORK

A. Understanding User Behavior

With the explosive increase of WiFi devices, there have been a number of research studies on the user behaviors of WiFi devices. W.J. Hsu et al. [18] proposed a time-variant community mobility model to reveal the time dependent behavior at specific locations. U. Kumar et al. [15], [14] designed a Markov O(2) predictor considering past two locations to predict the next AP that device would access. S. Biswas [19] studied several thousand wireless networks and provided a deep insight in network usage, interference levels and spectrum utilizations. X. Wei et al. [20] profiled user behavior and proposed the H-M-L model between user behavior in days of appearance and daily traffic. Our work provides a systematic approach with the comprehensive datasets to study the large campus WiFi network and we characterize the WiFi user behavior and understand the WiFi network usage pattern from multiple dimensions.

B. Analyzing Network Performance

The network performance of different WiFi devices are very different. A. Gember et al. [13] studied multiple characteristics (e.g. protocols, flows, traffics) of handheld devices and non-handheld wireless devices through analyzing two separate multi-APs wireless networks. X. Chen et al. [21] measured the network performance of smart mobile handhelds and found the dominant factors that cause inefficient use of network and operating system resources. A. Nikraves [22] studied the network performance of more than 100 mobile devices by use controlled, active measurement experiments. The results showed that the performance is different across different carriers, different locations and different times. M. Z. Shafiq et al. [23] analyzed the cellular network performance

during crowded events and account for how and why cellular network performance degraded during crowded events. Q. Xu et al. [24] proposed PROTEUS system to passively collect network performance information from applications, and used regression trees to learn network performance patterns to forecast future network performance.

C. Studying Campus Network

There are some prior work on optimizing WiFi service. M. Khadilkar et al. [6] studied the DHCP usage patterns and proposed two strategies for dynamically adjusting DHCP lease times for each client: single adaptation and exponential adaptation. I. Papapanagiotou et al. [11] measured that DHCP proposed a device-type-based lease setting method without incurring much communication overhead. X. Chen et al. [25] built an analytical model to estimate concurrent IP usage based on session lengths and IP address usage. A. K. Das [8] proposed a multi-modal method to perform network load balancing and network resource switching (activating/deactivating) for energy savings. Our focus in this work is to provide insight of the WiFi network, reveal the potential problems and then devise efficient policies in each dimension to improve the performance of WiFi network.

V. CONCLUSION AND FUTURE WORK

In this paper, we study the WiFi user behavior and WiFi network usage pattern with the comprehensive network trace, i.e. SNMP dataset, DHCP dataset and AAA dataset, which are collected from a large campus WiFi network. Then, we conduct a detailed measurement study from multiple dimensions, i.e. server plane, temporal plane, spatial plane and traffic plane from network administrators' point of view. We provide insights of user behavior and pinpoint that the WiFi network under study is far from optimal. Our observations indicate that the WiFi network could be improved and managed more efficiently from multiple dimensions. We believe that this measurement study is helpful for network administrators and researchers to understand more about large scale WiFi networks. As a future work, we plan to design schemes to improve the WiFi network based on our observations. For example, we are now designing a message exchange method between DHCP server and AAA server to reduce the waste of IP addresses.

VI. ACKNOWLEDGEMENT

The research is supported by National Key Research and Development Program of China (No. 2016YFB0801301).

REFERENCES

- [1] "International data corporation forecast," <https://www.idc.com/research/forecasts.jsp>.
- [2] M. Ovum, "Smartphone & tablet usage trends & insights 4g lte and wi-fi powering data consumption," 2015.
- [3] J. Sommers and P. Barford, "Cell vs. wifi: on the performance of metro area mobile connections," in *Proceedings of the 2012 ACM conference on Internet measurement conference*. ACM, 2012, pp. 301–314.
- [4] R. Droms, "Dynamic host configuration protocol," 1997.
- [5] S. Glass, T. Hiller, S. Jacobs, and C. Perkins, "Mobile ip authentication, authorization, and accounting requirements," Tech. Rep., 2000.

- [6] M. Khadilkar, N. Feamster, M. Sanders, and R. Clark, "Usage-based dhcp lease time optimization," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. ACM, 2007, pp. 71–76.
- [7] K. Sui, Y. Zhao, D. Pei, and L. Zimu, "How bad are the rogues' impact on enterprise 802.11 network performance?" in *Computer Communications (INFOCOM), 2015 IEEE Conference on*. IEEE, 2015, pp. 361–369.
- [8] A. K. Das, P. H. Pathak, J. Jee, C.-N. Chuah, and P. Mohapatra, "Non-intrusive multi-modal estimation of building occupancy," 2017.
- [9] J. Fan, J. Xu, M. H. Ammar, and S. B. Moon, "Prefix-preserving ip address anonymization: measurement-based security evaluation and a new cryptography-based scheme," *Computer Networks*, vol. 46, no. 2, pp. 253–272, 2004.
- [10] A. K. Das, P. H. Pathak, C.-N. Chuah, and P. Mohapatra, "Characterization of wireless multidevice users," *ACM Transactions on Internet Technology (TOIT)*, vol. 16, no. 4, p. 29, 2016.
- [11] I. Papapanagiotou, E. M. Nahum, and V. Pappas, "Configuring dhcp leases in the smartphone era," in *Proceedings of the 2012 ACM conference on Internet measurement conference*. ACM, 2012, pp. 365–370.
- [12] "Ethernet number registration," <http://www.iana.org/assignments/ethernet-numbers>.
- [13] A. Gember, A. Anand, and A. Akella, "A comparative study of handheld and non-handheld traffic in campus wi-fi networks," in *Passive and Active Measurement*. Springer, 2011, pp. 173–183.
- [14] U. Kumar, J. Kim, and A. Helmy, "Comparing wireless network usage: laptop vs smart-phones," in *Proceedings of the 19th annual international conference on Mobile computing & networking*. ACM, 2013, pp. 243–246.
- [15] U. Kumar, J. Kim, and Helmy, "Changing patterns of mobile network (wlan) usage: Smart-phones vs. laptops," in *Wireless Communications and Mobile Computing Conference (IWCMC), 2013 9th International*. IEEE, 2013, pp. 1584–1589.
- [16] F. Ganji, L. Budzisz, F. G. Debele, N. Li, M. Meo, M. Ricca, Y. Zhang, and A. Wolisz, "Greening campus w lans: Energy-relevant usage and mobility patterns," *Computer Networks*, vol. 78, pp. 164–181, 2015.
- [17] R. Doost-Mohammady, M. Y. Naderi, and K. R. Chowdhury, "Performance analysis of csma/ca based medium access in full duplex wireless communications," *IEEE Transactions on Mobile Computing*, vol. 15, no. 6, pp. 1457–1470, 2016.
- [18] W.-J. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy, "Modeling spatial and temporal dependencies of user mobility in wireless mobile networks," *IEEE/ACM Transactions on Networking (ToN)*, vol. 17, no. 5, pp. 1564–1577, 2009.
- [19] S. Biswas, J. Bicket, E. Wong, R. Musaloiu-e, A. Bhartia, and D. Aguayo, "Large-scale measurements of wireless network behavior," in *ACM SIGCOMM Computer Communication Review*, vol. 45, no. 4. ACM, 2015, pp. 153–165.
- [20] X. Wei, N. C. Valler, H. V. Madhyastha, I. Neamtiu, and M. Faloutsos, "A behavior-aware profiling of handheld devices," in *Computer Communications (INFOCOM), 2015 IEEE Conference on*. IEEE, 2015, pp. 846–854.
- [21] X. Chen, R. Jin, K. Suh, B. Wang, and W. Wei, "Network performance of smart mobile handhelds in a university campus wifi network," in *Proceedings of the 2012 ACM conference on Internet measurement conference*. ACM, 2012, pp. 315–328.
- [22] A. Nikraves, D. R. Choffnes, E. Katz-Bassett, Z. M. Mao, and M. Welsh, "Mobile network performance from user devices: A longitudinal, multidimensional analysis," in *PAM*, vol. 14. Springer, 2014, pp. 12–22.
- [23] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, S. Venkataraman, and J. Wang, "A first look at cellular network performance during crowded events," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 41, no. 1. ACM, 2013, pp. 17–28.
- [24] Q. Xu, S. Mehrotra, Z. Mao, and J. Li, "Proteus: network performance forecast for real-time, interactive mobile applications," in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 2013, pp. 347–360.
- [25] X. Chen, L. Lipsky, K. Suh, B. Wang, and W. Wei, "Session lengths and ip address usage of smartphones in a university campus wifi network: Characterization and analytical models," in *Performance Computing and Communications Conference (IPCCC), 2013 IEEE 32nd International*. IEEE, 2013, pp. 1–9.