Characterization of Wireless Multi-Device Users

Aveek K. Das*, Parth H. Pathak*, Chen-Nee Chuah†, Prasant Mohapatra*
*Computer Science Department, †Electrical and Computer Engineering Department,
University of California, Davis, CA, USA.
Email: {akdas, phpathak, chuah, pmohapatra}@ucdavis.edu

***Abstract*—There has been a huge growth in the number of wireless-enabled devices possessed by a user. Over two third of adults in United States currently own three devices - laptop, smartphone and tablet. In this paper, we provide a first look at the network usage behavior of today's multi-device users. Using the data collected from a large university campus, we provide a detailed measurement-based characterization study of over 30,000 users. Our objective is to understand how existence of multiple wireless devices affect the network usage behavior of users. Specifically, we study the usage pattern of devices, how the usage of difference devices overlap in time, user's preferences of accessing sensitive content and device-specific factors that govern their choice of WiFi encryption type. The study reveals numerous interesting findings such as how current DHCP configurations are oblivious to multiple devices which results in inefficient utilization of available IP address space.**

I. Introduction

The number of US adults who own a trio of laptop, smartphone and tablet have increased from 26% to 37% over the past year - an increase of 42% in one year [1]. New WiFi enabled wearable devices like smart watches and smart glasses are becoming increasingly popular. As a result, we can expect the aforementioned percentages to keep growing consistently. Recent wireless network measurement studies like [2], [3] have mostly focused on traffic characterization of only one device (e.g. smartphone) of users. Albeit important, such studies do not provide information about how users use their other devices and what are the dependencies in usage patterns. With the ever-increasing number of multi-device users, it has become essential to address several questions such as how such users use their different wireless devices, what content they access on them, what their security preferences and expectations are, etc. This work is first-of-its-kind attempt to answer these questions using real network traces of multi-device users.

Understanding the network usage pattern of different wireless devices for multi-device users is crucial in many ways. For a network service provider, it is useful in resource allocation and planning. For example, to cope with the increasing number of online devices that results in IP address space exhaustion, delaying or revoking IP addresses based on usage pattern can be beneficial to the providers. From the perspective of content providers, the usage pattern can provide information about which devices are being actively used, so that redundant content delivery to multiple devices of a user can be avoided. The same usage pattern information can be gathered by the advertisers and online analytics providers to get a more complete view of user's online activities beyond the partial view of what is available currently through one device. Last but not the least, different applications on user's devices can exploit this information in order to carry out intelligent multi-device coordination that can save energy by turning wireless radio on and off, depending on usage pattern. Although there have been recent efforts [4] in this direction, most applications on todays devices are more or less oblivious to the existence of other devices of the same user.

978-1-4673-7331-9/15/$31.00

There are a lot of potential applications for understanding the network usage pattern but, acquiring real-world network traces for multi-device users itself is a challenge. This is because network traces collected from the access or core networks rarely have any information about user's ownership of devices. In this work, we present a characterization of study of multi-device users using wireless network traffic traces collected from a large university campus. We combine the packet traces with user-device logs to associate traffic with users, which allows us to monitor fine-grained network usage activity. The characterization study described in this paper is based on data collected for nearly 1,000 access points from a university campus for approximately 30,000 users with the total network packet traces of 23 Terabytes. We classify user's wireless devices into three device types: smartphone, laptop and tablet.

The major findings of our work are as follows:

1) Device utilization of multi-device users: When a user owns more than one wireless enabled device, the overall network usage increases proportionally to the number of devices, rather than the usage being spread across the multiple devices. At the same time, the overall amount of time in which a particular device type is used, hardly varies based on the other devices owned by the user. Another interesting observation shows that when users own a tablet, the percentage packets generated by laptops decrease whereas the smartphone usage remains more or less constant.

2) ON-OFF usage patterns of devices and efficient DHCP assignment: We study the varying pattern in which a specific device type is used and how it is affected by other devices. A study of the "ON-OFF" usage of different wireless devices, show the usage remains specific to a particular device type and does not change depending on other devices being carried by the user. Also, the amount of time a hand-held (smartphones and tablets) device is continually ON is much lower than the DHCP lease times assigned.

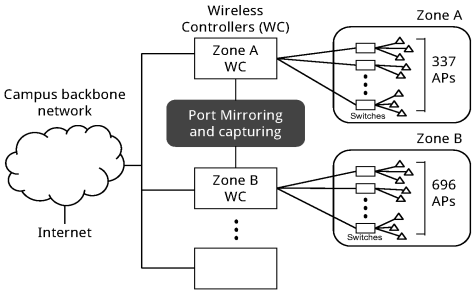


Fig. 1: Campus data capture setup

A study of inactivity of devices after they are assigned IP addresses shows that about 9000 handheld device sessions create a packet at least a minute after the assignment of an IP. The inactivity time in laptops are much smaller.

3) Security in multi-device users: General access of websites that reveal information personal to users are accessed more frequently from smartphones and as a result, among the multiple devices of a user, protecting a smartphone against security attacks is most important. The selection of WiFi network type (encrypted vs. unencrypted) is found to be more correlated to the device-type rather than specific user preferences. We observe that device-specific factors such as convenience of connection to specific network type from certain type of devices significantly affect user's choice.

In the rest of the paper, we introduce the dataset and our methodology for device detection in section II. In section III and IV we study in details multi-device utilization characteristics and the security aspects of multi-device users, respectively. Section V includes a discussion on the major findings. After presenting the related work in section VI, we conclude the paper in section VII.

II. Dataset and Methodology

A. WiFi Network Traces

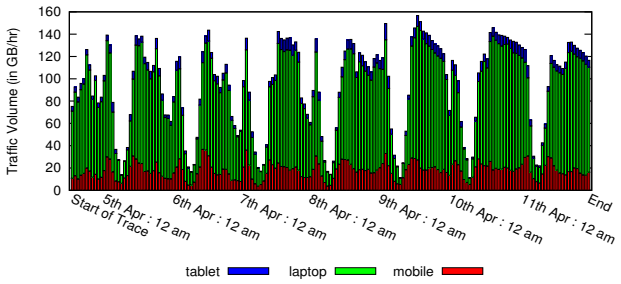
We collect the network packet traces from wireless controllers which connect to WiFi access points (APs). On the controller, we mirror the port, through which traffic is forwarded to and from the backbone network to capture the data. The setup for the wireless data capture is shown in Fig. 1. We collect data from the APs of two different areas:

1) Zone A: includes residential dormitories
2) Zone B: includes offices, classrooms, cafeterias

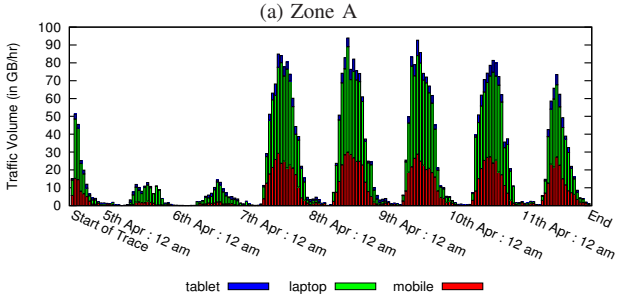
The network traces are collected for 8 days for the two zones. A detailed description of the traces (user, packets, size, etc.) can be found in Table I. As seen in the table, in Zone A, the total amount of data is much higher even with significantly lower number of users as compared to Zone B. This signifies that devices at the residential dormitories are connected to the network for a longer durations, which is expected. There is an overlap of 5280 users among the user-sets at two different locations - which indicates our dataset contains network data created by 32581 unique users. All the network data collected at the controllers comprised of both upstream and downstream traffic of the user devices as we are focused on the overall traffic for each device. Note that in this study, we only characterize user's *wireless* devices. A multi-device user may also have a wired device such as a desktop computer but the focus of our work are devices that connect to the WiFi network.

TABLE I: Dataset for characterization study

Location	Zone A	Zone B
Number of Users	7936	29925
Number of Devices	13729	48284
Number of Packets	19.9 billion	4.8 billion
Number of Access Points	337	696
Total Size	18.821 TB	4.942 TB



(a) Zone A



(b) Zone B

Fig. 2: Traffic volume in GB/hr at the two locations over the entire time of trace

Fig. 2 represents the total data volume (GigaBytes per hour) from the multiple devices of a user variation over the entire duration of our capture. A comparison of figs. 2a and 2b show that the data volume at Zone A has two peaks, one after midnight and one during noon, as compared to Zone B, which has one peak around noon. This is indicative of the location category as Zone B includes offices, classrooms, etc. and is expected to have high traffic only during office hours. For the same reason, the traffic in weekends (5th and 6th April) in Zone B is significantly low. Even though its expected that weekends will have higher overall traffic than weekdays - at Zone A - we see similar trend over the entire 8 days. This is because, each of the figures represent the trend of users' devices at that specific location, and not the overall trend of a user. We observe that laptops produce the largest volume of traffic, followed by smartphones and tablets - something that is quite intuitive.

B. Network Logs

Since our focus in this work is to understand the characteristics of multi-device users, we also acquire various logs to associate each packet with a user and a device. For this purpose we have two sets of logs:

1) Network Session Logs: The session logs record the association and dissociation times of each device to an AP. The log entries also contain the username, device MAC address, currently assigned IP address and the AP name to which the

TABLE II: Device count distribution

Location	Zone A		Zone B	
No. of Devices	User Count	% Users	User Count	% Users
1	3675	46.2	14919	49.9
2	3018	38	12158	40.6
3	992	12.6	2463	8.3
4	195	2.6	313	1
$\geq$ **5**	44	0.6	72	0.2

device is connected. These logs allow us to match each packet with to a user and her device using the IP address.

2) Network Address Translation (NAT) Logs: In certain areas, port-based NAT is used for handheld devices on campus. In such cases, we first map packet's public IP address and port to the corresponding private IP address and port using the NAT logs. After the mapping, the network session logs allow us to associate the packet with a user and a device.

Apart from the aforementioned logs, we also use the DHCP association logs. The DHCP association logs provides the *device name* for certain MAC addresses. As we show later, we use this information for detection of device type (smartphone, tablet or laptop).

With the use of the network session logs and the NAT logs, we do a packet-by-packet matching to associate each packet in the network packet traces described in Table I with a unique MAC address and a unique user.

C. Data Anonymization

The collected packet traces and the network logs are anonymized to remove any information that is specific to an individual. Specifically, we anonymize the IP addresses, the MAC addresses, the usernames, device-names and names of the access points. We employ prefix-preserving anonymization as proposed in [5]. The anonymization methods and parameters are kept consistent over all traces and logs in order for us to match packets, users and devices.

D. Device Count of Users

After associating each device to a specific user we calculate the number of devices a user owns. The device count variation of users at both locations is represented in Table II. We observe that about 50% of all users have more than one devices, which shows that there is a valid case for multi-device user study in a campus network. However, due the presence of visitors at Zone A and due to transient mobility patterns of users in Zone B, many users show up in our dataset with just one device, increasing the percentage of users with one device type.

E. Device Type Detection

One of the most important steps in our study is the detection of the type of a user device. We limit our observations to three device types - smartphone, laptop and tablet. To accomplish this we combine two different approaches:

TABLE III: Keywords for device type detection

Detection Method	DHCP Device Name	User Agent Parsing
Mobile Keywords	iPhone, Nokia HTC	Windows Phone, Dalvik Blackberry, Nexus 5
Laptop Keywords	Macintosh, PC Dell, Vaio	amd64, Fedora Ultrabook, Chrome OS
Tablet Keywords	iPad	iPad, Nexus 7, Surface

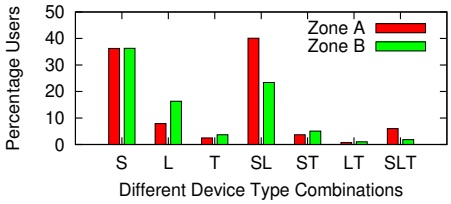


Fig. 3: Device Type Distributions at the two locations

TABLE IV: Different Device Types

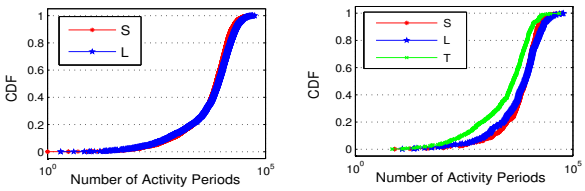
Device Type Combinations	Zone A	Zone B
S	2876	10861
L	626	4890
T	197	1106
SL	3182	6995
ST	293	1516
LT	56	307
SLT	463	556

1) DHCP Device Name Mining: The DHCP request message from the device to the server contains device's hostname. In most of the current platforms such as Windows and Mac OS, the hostname is the device name given by the operating system e.g: John-PC. As a result, the DHCP log file mentioned in section II-B includes the device-name for some MAC addresses. Device names like "John-PC", "Andy's MacBookPro" or "Trudy-iPhone" have keywords, the presence of which mean that the device is a laptop (in the first two cases) or a smartphone (in the last case). We do a keyword-based search on the DHCP host-names which predicts the device type of the MAC address. Some example keywords are shown in Table III.

2) User Agent Parsing and Mining: The user-agent field present in the HTTP GET Request header contains useful information about the device type. We use a combination of the information available (for e.g: CPU architecture, OS name, browser name, model name, etc.) [6] along with the user agent string for device type detection based on keyword-search. A set of keywords are shown in Table III.

Either of two approaches of device detection, by themselves, is not enough to detect the device type. For certain devices, the user-agent field has no useful device related information, whereas for some users, the DHCP host-name is blank or useless for our purpose. For example, in Android devices, the hostname is hashed for protection of user privacy and has no keywords which can contribute towards device detection. Overall, 57.4% of device type information are detected using the user-agent fields and the rest of 42.6% uses the DHCP hostname information. In certain cases, where there are fewer user-agent fields and there is no DHCP device name available, the device type remains unclassified. As a result, the device type distribution is slightly different in behavior than the device count distribution we showed above. The percentage of unclassified devices are 3.06% in Zone A and 12.22% in Zone B.

The device type distribution and the number of users in each device type combination is represented in Fig. 3 and Table IV. For our analysis, we divide the entire user set into 7 distinct groups: S, L, T, SL, ST, LT and SLT, where a user in set SLT owns smartphone, laptop and tablet. A user set is determined



(a) Zone A - user set SL (b) Zone A - user set SLT

(c) Zone B - user set SL (d) Zone B - user set SLT

Fig. 4: Activity period distributions for user sets SL and SLT at zones A and B - the distributions remain the same for specific a device type of a user, irrespective of other user devices

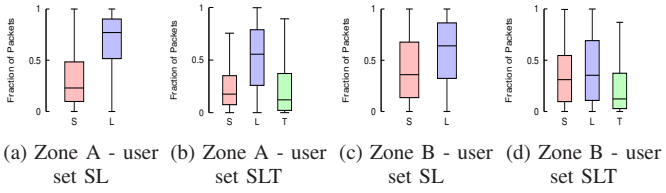
based on the number and type of devices a user owns. The highest number of occurrences of multiple devices is for users with a smartphone and a laptop. In a residential setting (Zone A), the number of users with all three device types (SLT) are higher as not all users carry out all their devices. The number of users with no mobile phones are almost negligible, which is expected as, in present scenario, almost every individual uses and carries around a smartphone. We use the same notation as seen in Fig. 3 to represent the different user sets. In addition, "S(SL)" is a representation of smartphone behavior among the user set having smartphones and laptops, and so on.

III. Multi-Device Utilization

The first question that we address in our multi-device user study is how do the users use their different devices to access the network. We answer the question using two levels of characterization. First, we provide a high-level aggregate characteristics of device usage in terms of time, packets and bytes. We then look at more fine-grained intermittent usage activity (such as ON-OFF usage) in Section III-B. Note that for all our analysis we consider the network usage as an indication of device usage, as it is known that the maximum network traffic volume is created when a device screen is on [7]. We also consider all the packets created by the devices (including TCP control packets, etc.)

A. Time and Packet Characteristics

1) Activity Period per Device Type: One of the primary indicators of device-usage is the amount of time for which the device generated network packets. A specific network session is not continuous network usage - it is a combination of many activity periods. We define one activity period as a 10-second time interval during which at least one packet was created by the device. As seen in [8], activity period determined using a 10 second window is a significant representative property of wireless network traffic. To understand how the total time usage of various device types varies in presence of other devices, we calculate the number of activity periods created by each device of a user. Fig. 4 shows Cumulative Density



(a) Zone A - user set SL (b) Zone A - user set SLT (c) Zone B - user set SL (d) Zone B - user set SLT

(e) Zone A - user set SL (f) Zone A - user set SLT (g) Zone B - user set SL (h) Zone B - user set SLT

Fig. 5: Division of traffic volume among different owned devices (a-d: packets, e-h: bytes). We see that inclusion of tablets result in significant decrease in laptop.

Function (CDF) of the activity periods of devices for users with smartphones and laptop (SL) and users with all three device types (SLT).

• From the CDF representations of activity periods of smartphones at Zone A in different user sets, we see the distributions for smartphones in Fig. 4a (S(SL)) and Fig. 4b (S(SLT)) are identical. Similar trend is seen for all other device types at both Zone A and B.

• As the distributions for different devices remain same for different user sets, we can combine the trends observed at both locations for all device types and claim that when a user has more than one device the overall time the wireless network is accessed increases rather than the total time getting divided between devices. As a direct result, the amount of time a specific device is used is independent of the presence of other devices. Overlapping activity of different devices, e.g. a user is using her laptop but her phone is also exchanging some traffic, is discussed in details in Section III.A3.

2) Percentage Traffic generated per Device Type: Similar to time, the traffic generated by a device is a definitive indicator of the usage of that device. Calculation of the amount of packets and bytes created by each device of a user shows how the overall generated traffic by a user is divided between her devices. The distributions of the fraction of packets generated by each device type for different user sets (SL and SLT) are shown in Fig. 5(a-d) in the form of a box-plot. Similarly, Fig. 5(e-h) show the distribution of the fraction of bytes created by each device type. In the plot, each bar represents the distribution that is specific to a particular device type in a unique user set.

• All the representations in Fig. 5 show that laptops create significantly higher traffic compared to the other device types. This follows intuitively (also mentioned in [9]) from the fact that data-extensive websites (like videos, file downloads etc.) are mostly accessed in laptops.

• At the residential dormitories of Zone A (Figs. 5c and 5d) the difference in generated traffic between laptop and handheld devices are much more prominent, as compared to Zone B (which includes classrooms, offices and cafeterias). This is another intuitive location based characteristic that is observed,

TABLE V: Keywords for website detection

Interest Category	Keywords
Social Networks	facebook, twitter, friends, social, plus.google
Entertainment	youtube, netflix, itunes, mp3, video, music
Games	zynga, xbox, games, puzzles, trivia, aws
News and Reading	nytimes, bbc, cnn, blogspot, news, magazine
Sports	espn, mlb, soccer, olympics, fifa, ncaa, nba
Education and Career	.edu, stackoverflow, github, courseera, school
Shopping	craigslist, amazon, ebay, target.com, groupon
Portals	yahoo, google, bing, msn

as handheld devices are used more in a non-residential setting.

• Due to the common set of apps in smartphones and tablets (e.g. Android or iOS apps), it is expected that a presence of tablet will reduce the percentage of traffic created by the smartphone, as similar content is expected to be accessed in both. However, a look at Figs. 5b and 5d and their comparison with Figs. 5a and 5c will reveal that the inclusion of a tablet device results in a significant drop in the percentage of packets created by laptops but does not , substantially decrease the packets created by smartphones. Similar trend is observed when we look at the bytes created by the different device types in each user set.

• In order to study different categories of content accessed from each device type, we do a keyword based search to classify the information available in packet headers - specifically full request URI available in HTTP GET requests and DNS queries - into different application categories. Fig. 6 shows the websites of different categories as accessed by the three device types among all users as a percentage of all websites accessed. Table V gives an example of keywords for the different categories. From the representation, we observe that the utilization of tablets and laptops are, in a way, interchangeable, that is, the content accessed from both these devices are similar to each other and in turn, different from the content accessed in smartphones.

3) Device Usage Overlap: Does the presence of more than one device mean that a user accesses the Internet with all her devices at the same time? In this section, we address that question by calculating the total amount of time there is an overlapped usage of two user devices. We calculate the number of activity periods when both user devices were simultaneously active. Fig. 7 shows activity periods of each device types and the overlap times between two pairs of devices. The simultaneous representation helps to compare the overlap times with the actual usage times.

• The overall overlap amount is very low (maximum being 1/4th of the entire time of device usage) as compared to the use of each device type. Comparing between the two locations, we observe more overlap in a residential setting as compared to Zone B. In Zone B, users in many cases, are in motion, and hence instances of overlapped usage is low.

• The maximum overlap of usage occurs for laptops and mobile phones. This, in a way is intuitive, and shows that a user has a normal tendency to use smartphones even when a laptop device is in use.

• The maximum value of activity period is much higher in Zone A, which follows directly from the fact that usage of devices happen for longer periods in a residential setting.

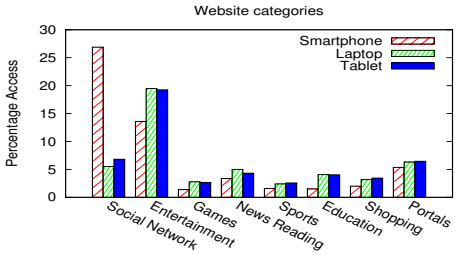
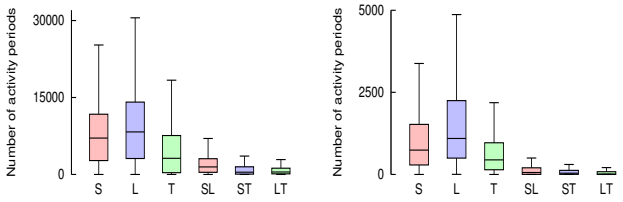


Fig. 6: Access of website categories in different device types: usage in laptops and tablets are almost the same and much different from smartphones.



(a) Zone A (b) Zone B

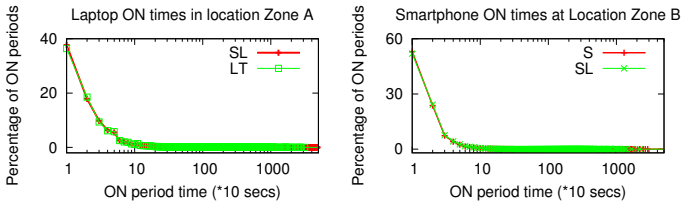
Fig. 7: Overlap of activity periods: overall very low overlap. Maximum overlap in smartphones and laptops

Findings: *From the time and overall traffic characteristics of multi-device users, we observe that the presence of additional user devices does not alter the duration of usage of a specific device. Another important observation is that the usage of a tablet causes a decrease in the percentage traffic in laptops, whereas the percentage traffic in smartphones remains unaltered. Based on traffic volume and content, we can say that the content usage of laptop and tablet are interchangeable whereas the mobile usage remains unaffected.*

B. Intermittent Network Usage Characteristics of Devices

1) ON-OFF Network Usage Pattern: We have studied the total amount of time a device was being accessed by users and observed that the behavior is independent of the presence of other devices, in most cases. However, the total usage time does not reveal any information about how a device is used, intermittently. As mentioned before, in our study, we consider the network usage as an indicator of device usage. In most cases a device is not used continuously, but follows an alternating on and off usage behavior. We refer to this behavior as "ON-OFF" device usage pattern, in this paper. During a WiFi connection, if a packet is created in a 10 second interval, we call the device "active" in that period. Continuous periods of activity constitute an "ON" period, and similarly, periods of inactivity constitutes an "OFF" period. We study how the presence of other devices have an effect on this intermittent user behavior, by calculating the ON-OFF times. Fig. 8 shows the probability mass functions (PMFs) of the "ON" times for laptops in user sets SL and LT and the "ON" times for smartphones in the user set S and SLT.

• The results in the Fig. 8 show that the ON-OFF usage of a device is not affected by the presence of other user devices. The PMF of laptop (and smartphone) ON-OFF times is almost identical across both user sets. This substantiates the claim that once a device is connected to the network and in use



(a) (b)

Fig. 8: "ON" period distributions: intermittent usage is also independent of other user devices present

by a user - the other devices owned by the same user - does not have an effect on the usage of that device. This is in a way, counter-intuitive, as we would expect the presence of a smartphone affecting the use of laptop (or vice versa), but the observations tell otherwise.

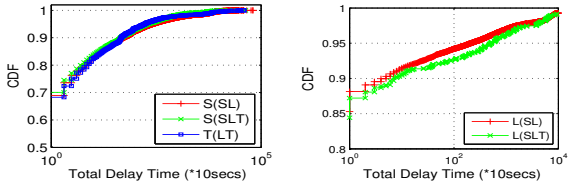
• However, the ON-OFF use of a smartphone is different from that of other device types. This is indicative of the fact that each device type has its own independent way of usage.

Based on the values of "OFF" times, the average inactivity time was calculated to be 100, 170 and 50 seconds for smartphones, laptops and tablets, respectively. Using these "OFF" period values, we recalculate the "ON" period distributions. In this case, we call a device inactive only if the continuous inactivity duration is greater than the average "OFF" duration for that device type. The recalculated "ON" have an average of 6 minutes for smartphones, 15 minutes for laptops and 2 minutes for tablets. The standard DHCP lease time provided on our traces is 900 seconds (15 minutes), which is less than half of the average ON times for handheld devices, as calculated above. A shorter DHCP lease duration assignment, for smartphones and tablets, can help in better utilization of the IP address space [10].

2) Delayed IP Address Assignment: Whenever a device revisits a previously known WiFi network, the device is automatically connected to the wireless network and is assigned an IP address. This IP address is assigned even if a user is not actively using the device. In this section we study, the amount of delay that exists between the time a user is assigned an IP address and the first packet created by the user. We observe that there is a distinct similarity in behavior in this respect between all handheld devices (smartphones and tablets). However, handheld devices behave much differently from the laptop devices. Fig. 9 shows the CDFs of the total delay times, in the case of laptops and handheld devices for a few representative user sets.

• We observe that nearly 17K sessions in handheld devices have a delay of at least one minute between IP assignment and the creation of the first packet. Overall, in 30% of the occasions, a handheld device has a delay of at least 10 seconds between the assignment of IP address and the first packet created. However, for laptops this number is as low as 12%.

• As there are a large number of sessions with significant delay in handheld devices, there can be certain conditions when the IP is not assigned directly to a handheld device on entering vicinity of the access point. Ultimately, when the user actually uses the device, a new DHCP request is sent to the server and the IP address is consequently assigned, thus



(a) Handheld devices (b) Laptop devices

Fig. 9: Delay between IP assignment and usage: hand-held devices have higher delay times as compared to laptops

avoiding auto-connection for handheld devices and in turn this can lead to better IP space utilization.

Findings: *(i) The "ON-OFF" device usage pattern of a specific device type remains unchanged irrespective of her other devices. (ii) We find that the average duration of continuous activity of handheld devices is much smaller compared to usual DHCP lease time which indicates that shorter DHCP lease times can be used for handheld devices. (iii) It is also observed that handheld devices have noticeable difference between the times of IP assignment and creation of first packet (due to auto-connection to WiFi networks) which, if corrected, can lead to efficient usage of IP address space*

IV. Security Aspects of Multi-Device Users

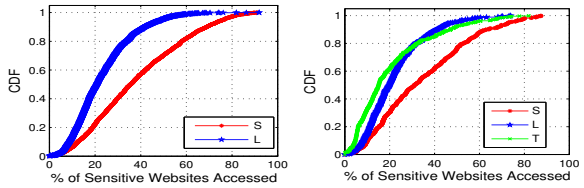
In the wireless networks, there are always security threats, with attacks ranging from D-DOS to spoofing, from malware spread to phishing attacks. From the multi-device perspective, we look at how users of different device types are vulnerable to attacks based on the websites they access or the choice of unencrypted wireless networks.

A. Access of Sensitive Websites

In this section, we study how the device type of a user governs the users' choice of accessing specific websites, specially websites with content sensitive to users. "Sensitive websites" are defined as websites which reveal information about users preferences or which contains user-sensitive personal information. In addition, websites which require a user to provide log-in information (username and password) are also considered in this category. Major categories of sensitive websites in our study are: health, finance, professional, social, productivity and preference. We identify sensitive websites using keyword-based search on information contained in the packet headers. We represent the statistics as a percentage of sensitive URLs and DNS queries from a device, among all those that were accessed by that device.

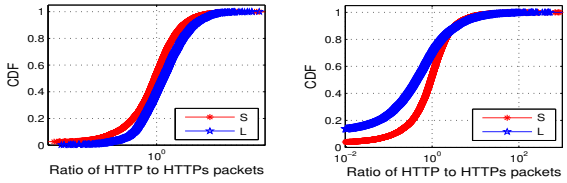
Fig. 10 represents the percentage of sensitive websites accessed across all the URLs and DNS queries at Zone A. We represent the CDF of sensitive website access for different user sets, based on the device types they carry. In addition, we also look at the ratio of HTTP and HTTPS packets created by smartphones and laptops. Such a representation is shows in Fig. 11. From this study the major observations include:

• The general pattern of access of sensitive websites, as seen in Fig. 10 shows that smartphone devices access sensitive websites more than the other devices. A large part of the smartphone traffic consists of social-networking websites,



(a) Zone A: user set SL (b) Zone A: user set SLT

Fig. 10: Access of sensitive websites: percentage sensitive websites are highest in smartphones - mainly due to social-networking, financial, education and email



(a) Zone A (b) Zone B

Fig. 11: Ratio of HTTP to HTTPS packets

banking related websites and emails. This attributes to the above mentioned observation.

• Another interesting observation is that specific device types have a consistent amount of access to sensitive websites, irrespective of the presence of other devices. This is in agreement with the observation in section III-B, where we see that once a device is being used, the presence of other devices does not alter its behavior.

• Comparison of figs. 11a and 11b show that Zone A has more HTTP packets than Zone B. This is a contextual location based characteristic, as in the office and work atmosphere of Zone B the websites with HTTPS enabled will be more than in a residential setting.

• Fig. 11a shows that smartphones have consistently more HTTPS traffic than laptops. HTTPS websites can be considered to be user-sensitive and thus, this result is consistent with our observation in Fig. 10 that smartphones have more access to sensitive websites than other device types.

Findings: *Sensitive websites constitute a higher percentage of overall content accessed in smartphones as compared to other device types - which indicate that protecting smartphones against security attacks is of utmost importance. At the same time, we observe that the behavior of each device is independent of other device types. In addition, we observe that smartphones have higher HTTPS traffic and the HTTP traffic proportion is higher in a residential location as compared to a work/university location.*

B. Choice of Encryption in Wireless Network

The campus wireless network provides two network options - one is an open wireless network (provides no WiFi encryption), while the other is encrypted. The two different wireless network options are provided from the same access point - so coverage of both network types is never an issue on campus. In this section we study how the use of wireless network type depends on user's device type and preference.

First, we study the amount of usage of each SSID type. We calculate the percentage usage of a specific wireless network

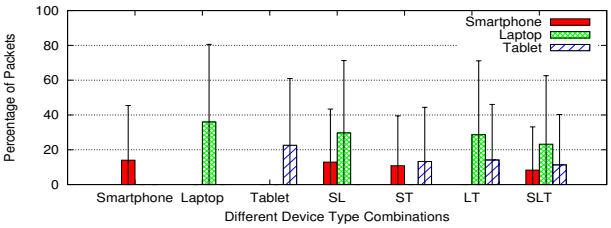


Fig. 12: Packets in unencrypted network: device behaviors are proportional to their screen size

type out of the overall network access. The results shown in Fig 12 represent the percentage of packets created via the non-encrypted SSID from different device types in all the seven representative user sets in the form of an error plot showing the variation (mean $\pm$ standard deviation). Fig. 12 show the access of the unencrypted network is consistent for smartphones, laptops and tablets across different user sets.

On further scrutiny, we can observe the percentage use of open network is directly proportional to the screen size of the device type. For access to the open wireless network on campus, students have to provide their login credentials on a portal after connecting to the network and they have to reconnect everytime they move to a new access point. For the encrypted network, the password is remembered by the devices and is automatically reconnected everytime (without any portal). As expected, the interface portal is easier to use and information can be conveniently filled in for devices with bigger screens, which explains the higher usage pattern of the access of unencrypted SSID for devices with bigger screen sizes.

1) Selection of Wireless Network Type: Once a device enters a WiFi network, the choice of network type can be made on the basis of a number of factors. We look at the dependence on device-type and user preference.

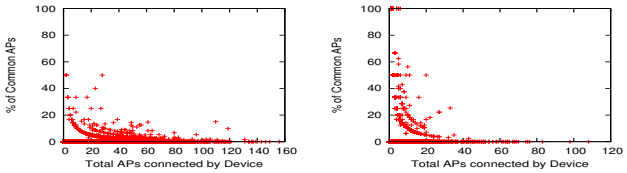
Device type dependence: Here, we want to study if the use of the network type in different device types, are interrelated with each other. We consider, as the null hypothesis, the distributions of network type access belong to the same underlying distribution for different device types. The alternate hypothesis is that their behaviors are independent. For this purpose, we calculate the two-sample K-S statistic for two empirical distributions(e.g.: smartphone and laptop), say S and L, based on the following equation:

$$\text{K-S statistic} = max(|S(i) - L(i)|), \tag{1}$$

where $S(i)$ denotes the fraction of elements in S with value less than or equal to i and $s(j)$ denotes the fraction of elements in S with values equal to j: $S(i) = \sum_{\forall j \leq i} s(j)$ and $\sum_{\forall j}(j) = 1$. We then compute the p-value, which defines the probability that the null hypothesis is true. A p-value less than the pre-selected significance level(α = 0.5) indicates the two distributions are different. Another way of interpretation is based on the value of the K-S Statistic - if greater than the critical value - the hypothesis is rejected. The critical

TABLE VI: K-S Statistic and p-Value

X	Y	K-S Stat	p-Value	CV	Hypothesis
smartphone	laptop	0.212	≈0	0.015	Rejected
smartphone	tablet	0.032	0.0037	0.02	Rejected
laptop	tablet	0.182	≈0	0.025	Rejected



(a) Handheld (b) Laptops

Fig. 13: Percentage of common access points

value(CV) is calculated as follows:

$$\text{Critical Value} = c(\alpha)\sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}, \quad (2)$$

where $c(\alpha)$ is based on the value of α and is equal to 1.36 and n_1 and n_2 are the number of datapoints in each distribution.

The calculated values and conclusions are shown in Table VI, where X and Y are the two distributions being considered. We observe that the null hypothesis is rejected in all the three cases, hence concluding that the usage of network type in different devices are independent of other devices of a user.

User Dependence: The next factor we study is whether the personal choice of users govern the selection of a particular network type for all her users - for example, if a user is security conscious she will ensure to connect all her devices to the encrypted WiFi network at all times. To quantify the dependence we calculate the Pearson's correlation coefficient between the network type usage distributions of different device types for multi-device users. Table VII shows the correlation values for different user sets.

TABLE VII: Correlation between packet distributions of different user sets

	S	L	T
S	1	0.37	**0.48**
L	0.37	1	0.30
T	**0.48**	0.30	1

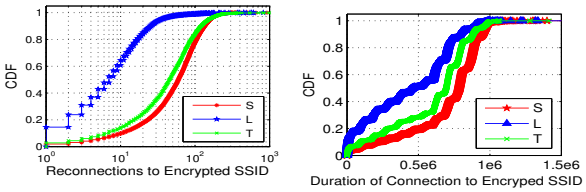
(a) Users with 3 devices

	S	L	T
S	1	0.35	**0.57**
L	0.35	1	0.39
T	**0.57**	0.39	1

(b) Users with 2 device

The results show a higher correlation between the distribution of packets created in each network type for handheld mobile devices as compared to the correlation between other device types. This is observed for all the different multi-device user sets. From Fig. 12, we see the major usage in these cases are of the encrypted network. Handheld devices are in use even when the user is moving around (users with high mobility), thus making the use of open network inconvenient as users are required to login via the portal whenever they move to a new access point. Thus, users prefer to use a network (the encrypted one) which authenticates automatically in their handheld devices, which explains the comparatively high correlation. Laptop devices do not have such high mobility and hence users' do not have a pre-determined choice of network type in those devices.

2) Characteristics of Switching of Wireless Network Type: The primary question we address in this section is: once a particular device type connects to a wireless network, does it change its network type? Overall statistics show that almost 95% of all devices have no change in the wireless device encryption type over the entire duration of our dataset. The dependence of load and location behind change of wireless network type has been discussed in the previous subsection.



(a) Number of reconnections before switching (b) Time of access before switching

Fig. 14: Encrypted network usage before switching

Common Access Points for both network types: In this section, we calculate the number of access points where a user connects to both network types as a percentage of total number of access points to which the user connects. From the scatter plot in Fig. 13 we see the most devices have a low number of common access points. A high percentage is only seen for devices connecting to a low number of access points overall.

We observe similar behavior patterns in smartphones and tablets which further strengthens the claim that handheld usage pattern of a user is correlated. These devices have a lower trend of common access points as they usually keep connected to the encrypted network and does not switch often. Laptops on the other hand have many instances where the number of common APs are a significant amount.

Re-connections and Time in same network: Once connected to a network type, we look at how many times and for how long does the user keep connecting back to the same network type. Fig. 14 shows the cumulative distribution of number of reconnections and total time before switching to the other network for the encrypted network type. Smartphones and tablets do not switch from the encrypted network often, as seen by the higher values of reconnection instances in Fig. 14a. The cumulative distribution of laptops show substantially less number of reconnections and shorter times spent in the encrypted network type. Similar behavior trends for smartphones and tablets, as claimed before, is reconfirmed from Fig. 14.

In general, the number of reconnections and time spent in the unencrypted network is lower (as seen in Fig. 12). Around 40% of users do not reconnect to the unencrypted network more than once, proving that users are in some cases concerned about the security of their devices.

Findings: *The choice of encrypted or unencrypted WiFi network shows loose correlation among different devices of the same user which shows low dependence on user's preferences. On the other hand, the choice is more correlated to the device type which indicates that device-specific factors such as auto-connect on handheld devices and ease of portal login on laptops play an important role in choosing the network type. Use of the encrypted network in handheld devices can be attributed more to the flexibility of usage of the encrypted network, rather than to reasons of security.*

V. Discussion of Results

We gained numerous insights through our characterization of multi-device users that can be useful to many entities. For instance, even though most tablet apps are similar to the ones on smartphones, we observe that the network access of tablets and laptops are interchangeable. This shows that

tablet app development should be more pertinent to laptop-type tasks as users prefer to offload their laptop access to tablet when mobile. Because of the same reason, from the perspective of online analytics and advertisers, we observed that mobile combined with laptop or tablet provide a more complete view of user's online footprint as opposed to laptop and mobile/tablet. We confirm that any online analytics should span across multiple devices of a user as the usage of multiple devices is more additive in terms of overall network access. Apart from this, since a user with more devices consumes more data overall, schemes that can address content redundancy for content providers as well as device platform developers should be actively investigated. We also inferred that network operators can improve the IP space utilization by assigning shorter lease times to handhelds as well as potentially delaying the IP assignment to the devices of multi-device users. Although expected, we verified that access to sensitive content on mobile platforms is significantly higher, which means protecting against mobile malware is extremely important. Also, we learned that users do not necessarily make an informed decision about the choice of encrypted or unencrypted network, but instead other factors such as convenience of connection to one type on network from a device type affect their choices.

The characterization study in this paper is based on a campus-wide dataset and the observations are directly applicable to a student population in a university campus. However, for understanding multi-device usage patterns for a larger population and for more generalized inferences, a similar study is required on other representative locations where the daily timelines and behaviors are different from a university campus. Our dataset includes data from WiFi access points. Thus, our study does not represent user behavior for people who have maximum internet usage using cellular data.

VI. Related Work

In recent years, there has been a number of research studies on smartphone characterization. [2] and [3] looked at the usage of smartphones among users, with focus on browsing patterns of users, protocol overhead, radio power usage and management. Other research efforts have studied the effects of mobility and interaction of users with smartphones at different locations [11], and tried to profile users based on their smartphone usage [12]. There has also been studies [13], [14] that investigated the diversity in users' smartphone interaction patterns. All of these studies are primarily device-centric as they only focus on smartphones and its usage characteristics. Our focus in this work is to explore *user-centric* patterns of network access for multiple devices of the user. Also, as opposed to collecting the data from a single device of volunteers, we have investigated a dataset that can capture network usage pattern of multiple devices of a user.

Gember et.al have provided a comparative study of overall usage of handheld and non-handheld devices in a campus network in [15]. Different from our work, their study mostly characterizes and compares network traffic of handheld and non-handheld devices. On the other hand, our objective in this work is to look at the characteristics of multi-device *users* and how the network access pattern changes for one device in the presence of other user devices. Works like [10], [16] have analyzed device connection session lengths for different types of devices. The inferences are shown to be useful in efficient DHCP lease time allocation. In our work, we extend the device-centric view of session lengths to a user-centric view whereby we claimed that delayed IP assignment for devices of multi-device users can be an effective mean to improve IP address space utilization.

VII. Conclusions and Future Work

In this paper, we presented a detailed characterization of multi-device users in a campus wireless network based on the network traces collected for 32,581 users over 8 days. We provide many insights regarding how the characteristics of multi-device users can be useful to various entities such as content providers, advertisers, network operators and application developers. As an extension of this work, we plan to design schemes that can provide improved coordination between the multiple wireless devices of a user and increase the energy efficiency as well as decrease the amount of redundant content delivered to all her devices.

References

[1] "Digital Omnivores Craving More Content Across Devices - Digital Democracy Survey." www.deloitte.com.

[2] H. Falaki, D. Lymberopoulos, R. Mahajan, S. Kandula, and D. Estrin, "A first look at traffic on smartphones," in *ACM IMC '10*.

[3] J. Huang, Q. Xu, B. Tiwana, Z. M. Mao, M. Zhang, and P. Bahl, "Anatomizing application performance differences on smartphones," in *ACM MobiSys '10*.

[4] "iPhone, iPad, and Mac. Connected like never before." www.apple.com/ios/ios8/continuity.

[5] 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," *Comput. Netw.*, vol. 46, pp. 253–272, Oct. 2004.

[6] G. Maier, F. Schneider, and A. Feldmann, "A first look at mobile handheld device traffic," in *Passive and Active Measurement '10*.

[7] J. Huang, F. Qian, Z. M. Mao, S. Sen, and O. Spatscheck, "Screen-off traffic characterization and optimization in 3g/4g networks," in *ACM Internet Measurement Conference '12*.

[8] A. K. Das, P. H. Pathak, C.-N. Chuah, and P. Mohapatra, "Contextual localization through network traffic analysis," in *IEEE INFOCOM, 2014*.

[9] "Cisco Visual Networking Index." http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html.

[10] I. Papapanagiotou, E. M. Nahum, and V. Pappas, "Configuring dhcp leases in the smartphone era," in *ACM IMC '12*.

[11] I. Trestian, S. Ranjan, A. Kuzmanovic, and A. Nucci, "Measuring serendipity: Connecting people, locations and interests in a mobile 3g network," in *ACM Internet Measurement Conference '09*.

[12] R. Keralapura, A. Nucci, Z.-L. Zhang, and L. Gao, "Profiling users in a 3g network using hourglass co-clustering," in *ACM MobiCom '10*.

[13] Q. Xu, J. Erman, A. Gerber, Z. Mao, J. Pang, and S. Venkataraman, "Identifying diverse usage behaviors of smartphone apps," in *ACM Internet Measurement Conference '11*.

[14] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin, "Diversity in smartphone usage," in *ACM MobiSys '10*.

[15] A. Gember, A. Anand, and A. Akella, "A comparative study of handheld and non-handheld traffic in campus wi-fi networks," in *PAM '11*.

[16] 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 *IEEE IPCCC '13*.