A Behavior-aware Profiling of Handheld Devices

Xuetao Wei* Nicholas C. Valler† Harsha V. Madhyastha⋆
Iulian Neamtiu⊕ Michalis Faloutsos×
*University of Cincinnati †Crowdcompass, Inc. ⋆University of Michigan
⊕University of California, Riverside ×University of New Mexico
weix2@ucmail.uc.edu nvaller@crowdcompass.com harshavm@umich.edu neamtiu@cs.ucr.edu michalis@cs.unm.edu

***Abstract*—The Bring-Your-Own-Handheld-device (BYOH) phenomenon continues to make inroads as more people bring their own handheld devices to work or school. While convenient to device owners, this trend presents novel management challenges to network administrators. Prior efforts only focused on studying either the comparative characterization of aggregate network traffic between BYOHs and non-BYOHs or network performance issues, such as TCP and download times or mobility issues. We identify one critical question that network administrators need to answer: how do these BYOHs behave individually? In response, we design and deploy BROFILER, a behavior-aware profiling framework that improves visibility into the management of BYOHs. The contributions of our work are two-fold. First, we present BROFILER, a time-aware device-centric approach for grouping devices into intuitive behavioral groups. Second, we conduct an extensive study of BYOHs using our approach with real data collected over a year, and highlight several novel insights on the behavior of BYOHs. These observations underscore the importance of that BYOHs need to be managed explicitly as they behave in unique and unexpected ways.**

I. INTRODUCTION

Smartphones and tablets are becoming ubiquitous in companies and universities. These devices are used more and more to complement, or even replace, desktops and laptops for computational needs: Gartner market research indicates that in the second quarter of 2013 worldwide PC shipments declined by 10.9%, while smartphone sales grew by 46.5% [13], [14]; hence the *Bring Your Own Handheld-device* (BYOH) practice is going to increase. We use the term BYOH to describe only smartphones and tablets, in accordance with the National Institute of Standards and Technology's definition [29]. In other words, we consider a device as BYOH if it runs a mobile OS, such as Android, iOS, or BlackBerry OS.

We argue that BYOHs deserve to be studied as a new breed of devices. First, every time a new technology or a new killer app emerges, IT departments must re-evaluate the way they manage their networks. Network administrators must understand the behavior of BYOHs in order to manage them effectively. Second, it is clear that BYOHs introduce different technologies and user behaviors: (a) BYOHs join and leave the network frequently, (b) their form factor enables novel uses compared to desktops and laptops, (c) they run different operating systems compared to other computing devices, and (d) most importantly, the apps that can run on them introduce a slew of management challenges [6], [10], [19], [37], [38].

The problem we address here is: how do these BYOHs behave individually? Given our interest in the network administrator's point of view, we have consulted with administrators of two different large networks, and our study has been largely shaped by their concerns and feedback. Both administrators admitted that there is a great need to better understand what BYOHs do, in order to devise better policies to manage them.

Most prior efforts have focused on studying either the comparative characterization of ***aggregate*** network traffic between BYOHs and non-BYOHs, or performance and network protocol issues, such as TCP and download times or mobility issues [4], [5], [12], [15], [16], [28], [34], [36]. How ***individual*** BYOH behaves has not been studied yet. In addition, existing approaches for managing traffic assume certain software installations on devices or embed tracking libraries in enterprise architectures. However, in practice, network administrators usually have no control over the software running on BYOHs, which results that we do not have deep visibility into BYOHs. This makes it difficult to control the behavior of these devices [10]. To the best of our knowledge, no prior work has focused on understanding ***individual*** BYOH behavior in campus networks, with a view towards managing and provisioning network resources. We discuss related work in Section V.

In this paper, we propose BROFILER (**B**YOH p**rofiler**), a systematic approach to profiling the behavior of BYOHs in a device-centric way. In addition, we arguably provide the first multi-dimensional study on the behavior of BYOHs from a network administrator's point of view. Our contributions are twofold: (a) we describe BROFILER, a time-aware device-centric approach for grouping BYOHs into intuitive behavioral groups, and a hierarchical framework for profiling individual user behavior based on multiple dimensions, (b) we conduct an extensive profiling study to understand BYOH behavior and show that our framework can provide useful insights. We use real network traces from a large campus: device access logs collected over the entire year, involving 22,702 BYOHs, and traffic data logs during the month of May involving 6,482 BYOHs.

A key advantage of our approach is that it is easy to deploy as shown in Figure 1: it learns BYOH behavior on-the-fly, and it does not require software installed on the device or device registration. Further, we argue that the intuitive profiles of BROFILER can help administrators: (a) form a conceptual view of what their BYOH user-base does, (b) help them troubleshoot issues by providing meaningful groups of users, and (c) provide an informed starting point for establishing reasonable and effective policies.

Our major contributions are highlighted below:

a. The BROFILER approach. We present BROFILER and describe how it can form the foundation of an effective BYOH profiling system (Section III). BROFILER analyzes and

Device Type	Count	Percentage
BYOHs	22,702	43.2%
Android	10,756	47.4%
iOS	11,328	50%
BlackBerry OS	618	2.6%
non-BYOHs	29,861	56.8%

TABLE I. DISTRIBUTION OF DEVICES IN DATASET DHCP-366.

profiles BYOHs across multiple dimensions, and we show how it can help us identify groups of users with interesting behaviors. For example, nearly half of the BYOHs are "mobile zombies", which acquire IP addresses without transferring any data over the campus network because they cannot advance past a captive portal. This behavior wastes resources, because zombies claim an address and possibly hit the captive portal log-in page, but never successfully log-in. Furthermore, a group of more than 32% of these mobile zombies (discussed in Section IV-D) appear *only one day* in the month of observation, which indicates ephemeral visitors with no impact on the network other than occupying an IP address; we refer to these as **vagabonds**.

b. An extensive profiling study. Using our approach, we conduct an extensive profiling study using real traces (in Section IV). We identify many unexpected behaviors and interesting groups of users. For example, we find that 68% of BYOHs do not conform to DHCP protocol specifications (reportedly due to a software bug [1]). Among the BYOHs that produce traffic, 94% of them generate less than 100MB in a month. At the same time, only 6% of BYOHs generate 82% of the total BYOH traffic.

II. DATASETS AND INITIAL STATISTICS

Our study is based on two datasets collected at a monitoring point inside a large, educational, campus network. One dataset, denoted DHCP-366,[1] consists of the campus WLAN's year-long DHCP logs from January to December. Another dataset (denoted as Traffic-May) is network flow-level traffic for BYOHs during the month of May, which is obtained as follows. First, WLAN traffic is filtered by the WLAN IP address pool. We then identified those IP addresses associated with BYOHs from DHCP logs during the month May (we use DHCP-May to denote the DHCP logs from the month May). For each BYOH, we use the mapping between its IP addresses and MAC address to identify the network traffic flows associated with the device in the flow-level traffic dataset. In total, our year-long DHCP dataset (DHCP-366) comprises 22,702 BYOHs and 29,861 non-BYOHs. The month-long BYOHs' traffic dataset (Traffic-May) comprises 6,482 BYOHs.

BYOH vs. non-BYOH. We identified BYOHs by examining the device's operating system keywords and MAC address as captured by the DHCP log file. First, we extracted each device's manufacturer; the MAC address contains an OUI (Organizationally Unique Identifier) which identifies the manufacturer [18]. Next, we use the operating system and manufacturer information to distinguish between BYOH and non-BYOHs. We identify BYOHs based on keywords (e.g., Android, iPad, iPhone, or BlackBerry) in their operating system name [17], [18]. Table I shows the number of devices in each category in the dataset DHCP-366. Note that BYOHs represent 43.2% of WLAN-using devices during one year, thus constituting a significant presence on the campus network.

[1]The 366 stands for the days of a leap year.

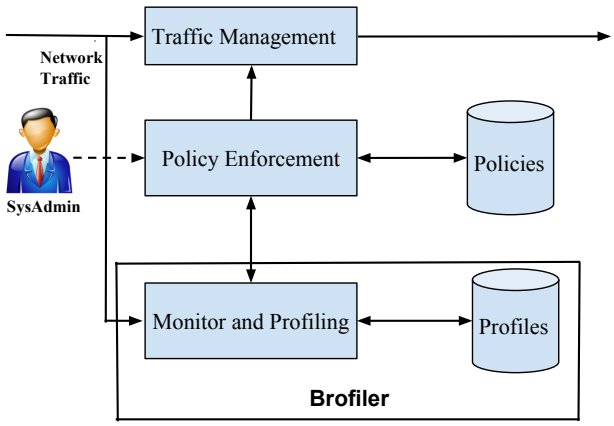

Fig. 1. Deployment of BROFILER.

Mobile platforms. We observe three mobile platforms in our DHCP-366 dataset: Android, iOS, and BlackBerry. As expected, Android and iOS are dominant and together, they account for roughly 97.4% of BYOHs.

III. BROFILER: SYSTEMATIC PROFILING

We propose a systematic approach to profile BYOHs based on their behavioral patterns. The goal is to develop a classification that is: *intuitive* and *useful*, so that network administrators can monitor, manage, and reason about groups of BYOHs. Our framework focuses on profiling user behavior based on multiple dimensions such as frequency of appearance, data usage, and IP requests.

We first present our classification approach using three dimensions, and then we combine multiple dimensions.

a. Data plane. In this dimension, we profile devices based on the traffic that they generate. Clearly, there are many different aspects and properties of traffic; in this work, we focus on traffic intensity. First, we determine whether the BYOH has any network traffic. Note that we define network traffic as the traffic that goes over the institution's network, not over the mobile wireless carrier.

We define two categories of BYOHs: (a) **Zero traffic BYOHs** or **mobile zombies**, that do not generate any network traffic, and (b) **Non-zero traffic BYOHs**, that generate traffic. Later, we show how we further study traffic behavior based on traffic intensity. In our dataset, there are 3,040 zero traffic BYOHs and 3,442 non-zero traffic BYOHs. We present the details in Section IV-B.

b. Temporal behavior. In this dimension, we profile devices based on temporal behavior, focusing on device appearance frequency on the campus network. A human-centric way to define frequency is by counting how many distinct weeks the device appeared on campus. The intuition is that regular employees and diligent people appear every week

on the campus network. Clearly, profiling criteria depend on the context and nature of the network, e.g., campus versus enterprise or a government network. Here, we use the datasets `DHCP-May` and `Traffic-May`. Note that the month May began on a Monday and spanned five weeks, labeled as follows: Week 1 (May 1 to May 5), Week 2 (May 6 to May 12), Week 3 (May 13 to May 19), Week 4 (May 20 to May 26), and Week 5 (May 27 to May 31).

We define the following terms. If a BYOH appears in at least four of the five weeks, we label it as `REG` (short for **regular**). Otherwise, we label the BYOH as `NRE` (short for **non-regular**). This applies to both zero and non-zero traffic BYOHs. We present the details in Section IV-C.

c. Protocol and Control plane. This dimension captures the operational properties of every BYOH. There are many interesting aspects such as the OS it runs, whether it conforms to protocol specifications, and whether it could pose security concerns, e.g., using encryption. In this work, we mostly focus on: (a) the behavior of the BYOH from a DHCP point of view, i.e., how it behaves in terms of acquiring an IP address, and (b) the use of encryption in terms of HTTPS. We present details in Section IV-A.

d. Multi-dimensional grouping using the H-M-L model. We propose a profiling framework using an H-M-L model, which groups devices based on intensity measures across different dimensions using three levels per dimension: H (High), M (Medium), and L (Low). Though we could use a different number of levels, we have opted for a three-level model because (a) it is more intuitive and thus easier to use, and (b) three levels are statistically suitable for capturing the distribution of the users on the dimensions of interest. Specifically, we used the X-means clustering algorithm [8] on our data to identify the three clusters and derive the thresholds, which correspond to our levels.

Flexibility and customizability. The main point here is to provide an initial framework and showcase its usefulness. Clearly, our framework can be customized and extended. Note that one could consider different or multiple metrics from each dimension and appropriately define thresholds for defining the H-M-L levels. The selection of metrics and thresholds could be dictated by: (a) what the network administrator wants to identify, and (b) the nature of the traffic under scrutiny. For example, in a military setting, devices could be expected to be present every day and a single unjustified absence could be a cause for concern.

The value of an intuitive model. The rationale behind our H-M-L model is that, often, relative and contextualized metrics are more useful than raw performance numbers, depending on the task at hand. For example, reporting that a user generates 100MB of data in a month is more precise, but arguably less useful than knowing that a user belongs to the network's heavy-hitters. We argue that an intuitive model can help administrators form a conceptual picture and then dive deeper into more fine-grained and quantitative analysis, as needed.

A. The Utility of our Approach

To showcase how BROFILER helps us identify interesting groups of users, we use two dimensions: days of appearance and daily average traffic. **Days of appearance** is the number of days that each BYOH shows up in the campus network. **Daily average traffic** is the ratio of total traffic per BYOH during one month over the number of days it shows up. We argue that the metrics and dimensions defined above are sufficient to give interesting results, and help administrators improve or devise new policies.

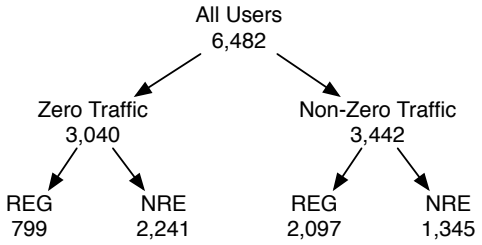

Fig. 2. A visualization of BROFILER's classification hierarchy: group designation and number of BYOHs in each group. We use the H-M-L model to further refine the leaves of the tree.

The classification (groups and number of devices in each group) is shown in Figure 2. We further profile the `REG` and `NRE` group devices with the H-M-L model. We present a more detailed discussion and related plots that lead to the observations below in Section IV. Note that we use data from the month of May, where we have both DHCP, `DHCP-May`, and traffic information, `Traffic-May`. We now turn to presenting some of the findings enabled by BROFILER.

1) In the `Traffic-May` dataset, nearly half of the BYOHs are **mobile zombies**, which we define as BYOHs that hold IP addresses without transferring any data through the campus network. Note that the data transferred while interacting with the captive portal does not count; rather we mean no data is transferred after the captive portal exchange.
2) We find that 23% of the BYOHs in `Traffic-May` are **vagabonds**, a term we use to refer to BYOHs that appear only one day during that month. Vagabonds is a sub-category of non-regular BYOHs, that we defined earlier.
3) We found that 3% of non-zero traffic BYOHs show low frequency of appearance and high traffic (denoted as LH), which is an uncommon behavior. We investigated this further and found the cause to be the use of video and streaming.
4) 26% of the mobile zombies appear frequently, each for more than 10 days in a month. This group unnecessarily and repeatedly occupies IP addresses, and should be managed accordingly.
5) We identify a group with high frequency of appearance during the month and low traffic (denoted as HL in our H-M-L classification), which accounts for 4% of non-zero traffic BYOHs.

IV. STUDYING AND PROFILING BYOHS

We use BROFILER as a starting point towards a long-term study on real BYOH traces. We show how BROFILER can help us profile and classify BYOHs, and reveal performance and network management issues. The goal here is to highlight both

the usefulness of our approach, and interesting observations on BYOH behaviors. Even for the rather expected behaviors, such as diurnal pattern and bimodal usage, this is arguably the first study to quantify these behaviors for BYOHs in a systematic and comprehensive way.

Summary of observations. We highlight our results grouped by the four dimensions of our approach.

a. Protocol and Control Plane.

1) 68% of BYOHs misbehave, by not conforming to the DHCP protocol specifications.
2) 80.6% of the IP lease requests by BYOHs are non-conforming.
3) Most of the web data of BYOHs is not encrypted: less than 15% of web traffic uses HTTPS.

b. Data Plane.

1) Of the BYOHs that produce traffic, 94% generate network traffic of less than 100MB (in a month). However, just 6% of BYOHs generate 82.1% of total BYOHs' traffic.
2) Data generation is very bursty, with 70% of BYOHs generating half of their monthly traffic in just one day. Surprisingly, 28.8% of BYOHs are active (sending or receiving traffic) *only one day* during the month.
3) 42% of BYOHs talk to internal (campus) servers.

c. Temporal Behavior.

1) BYOHs' patterns of appearance on the network follow weekly and daily patterns.
2) Intra-day behaviors of BYOHs are anthropocentric.
3) 55% of BYOHs are `NRE` devices while 45% of devices are `REG` devices.
4) Over 23% of the BYOHs are vagabonds that appear on only one day.

d. Multi-level profiling. The key results were listed in Section III-A.

A. Protocol and Control Plane

There are many interesting aspects in this dimension. Here, we focus on the DHCP operations of BYOHs and the use of encryption.

Non-conforming IP Lease Requests: We examine the DHCP operations between BYOHs and DHCP servers. We find that 68% of BYOHs issue unnecessary IP lease requests; this behavior is largely limited to BYOHs. We define a `non-conforming` **IP lease request** as an IP lease request sent by a device which already has an IP address from an earlier, unexpired lease. Note that this process begins with `DHCPDISCOVER` and it is not the regular IP lease renewal process via `DHCPREQUEST`. In other words, clients behave as if the IP acquisition process has failed, and they go back to the initial IP discovery phase, as indicated by the `DHCPDISCOVER` message.

Roughly 80% of IP requests issued by BYOHs are `non-conforming`**.** This erratic behavior significantly increases DHCP server workload and overloads the networks' DHCP service. In contrast, we find that non-BYOHs never issue such requests. Recent anecdotal evidence suggests that software bugs (acknowledged by Google [1]) in BYOHs are responsible for this misbehavior and argues that this erratic behavior is not due to the events of disconnection, reconnection and roaming [1].This observation suggests that network administrators should monitor and diagnose protocol operation behaviors from BYOHs in order to detect malfunctioning devices.

Amazon	17.95%
Facebook	13.3%
MSN	13.3%
internal web-servers	13.2%
Google	11.36%

TABLE II. TOP 5 HTTPS DOMAINS IN OUR DATA BY PERCENTAGE OF HTTPS TRAFFIC.

Given the observation above, a question arises naturally: *Are BYOHs making more IP requests because of shorter IP lease times?* We show that this is not the case. BYOHs issue more IP lease requests, although they have longer lease times compared to non-BYOHs. We identify lease times by analyzing the `DHCPOFFER` and `DHCPACK` messages, which contain a variety of lease parameters, including IP address lease time. We compute the average IP lease for both types of devices and find that the average IP lease time of non-BYOHs is 28 minutes, whereas that of BYOHs is 2.6 hours. This rules out a short lease time as the cause for the large number of IP lease requests from BYOHs.

Encrypted Traffic: Our study confirms that HTTP traffic dominates BYOH traffic [4], [36]. However, we observe diverse HTTPS/HTTP ratios across BYOHs. We find that roughly 24% of BYOHs have network traffic in which the fraction of traffic that uses HTTPS is over 50%. Surprisingly, some BYOHs have 100% HTTPS traffic. We further investigate the HTTPS domains that BYOHs talk to (Table II). We see that most of the HTTPS traffic is from popular online service providers. This is natural, as traffic to these providers is privacy-sensitive. For example, Amazon provides shopping and cloud services, and maintains personal or business transaction information. Facebook, the popular social networking service, contains private content, such as personal messages and photos. We see that web servers internal to the campus are among the top five web servers in terms of HTTPS traffic volume, with 13.2% of the total HTTPS traffic; these correspond to secure enterprise services, such as financial services, employee credentials, and email. Though we find the percentage of HTTPS traffic to be small, it is not clear that the presence of unencrypted HTTP traffic is necessarily a security risk. To verify this, we need to do an in-depth analysis of the unencrypted traffic, which we could not perform with our current data trace (lack of access to HTTP headers or payload data).

B. Data plane

In this dimension, we focus on the traffic behavior of BYOHs. We first profile and classify the BYOHs by looking at the traffic volume generated by each BYOH, then further look at the traffic dynamics, and whether these BYOHs talk to internal servers and malicious sites.

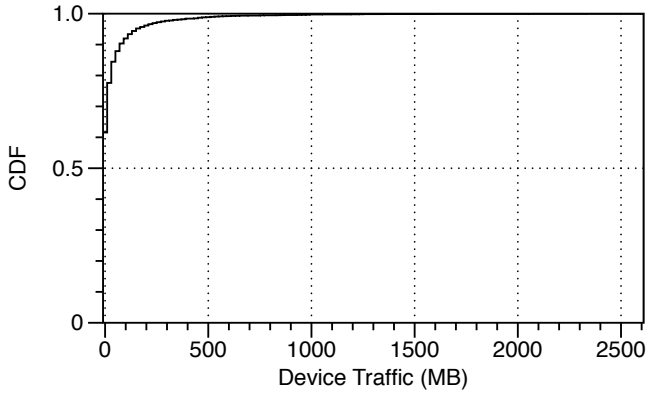

Fig. 3. Distribution of traffic volume per BYOH.

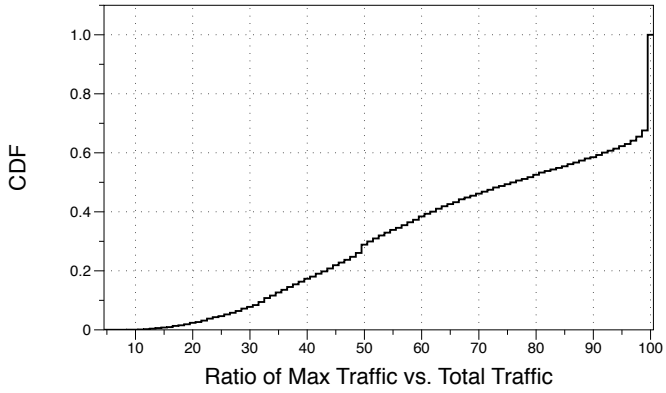

Fig. 4. Ratio of maximum daily traffic volume over total monthly traffic for each device.

Traffic Volume: In Figure 3, we plot the distribution of traffic volume across BYOHs, over the entire month. The distribution is highly skewed as roughly 94% of BYOHs generate less than 100MB during the month. The traffic volume per BYOH varies significantly across BYOHs, e.g., traffic volume ranges from as little as 72 bytes to as large as 2.5GB. In fact, we find that 6% of BYOHs generate 82.1% of the total traffic from BYOHs. This strongly indicates that a small fraction of BYOHs consumed most of the network bandwidth, hence classifying such groups of users and prioritizing network resources accordingly are desirable.

Traffic Dynamics: A natural question to ask is whether the traffic behavior is consistent day to day. We find that it is not. In Figure 4, we plot the CDF of the ratio between the maximum daily traffic over the total volume of the BYOH for the month. If the traffic was equally distributed among the days of the month, then the maximum daily traffic over the total monthly volume would be around 3.33% (100% divided by 30 days), hence the CDF would rise abruptly around the 3.33 point on the x-axis. Instead, we see that more than 70% of BYOHs consume half of their total monthly traffic in a single day ($x = 50$, $y = 0.3$). Surprisingly, 28.8% of BYOHs are active (sending or receiving traffic) only one day in the entire month. The above observations are helpful guidelines for managing and provisioning the network. At a high level, the observations suggest that traffic volumes: (a) vary across devices significantly, and (b) are very bursty in time. An effective management policy will need to consider these factors.

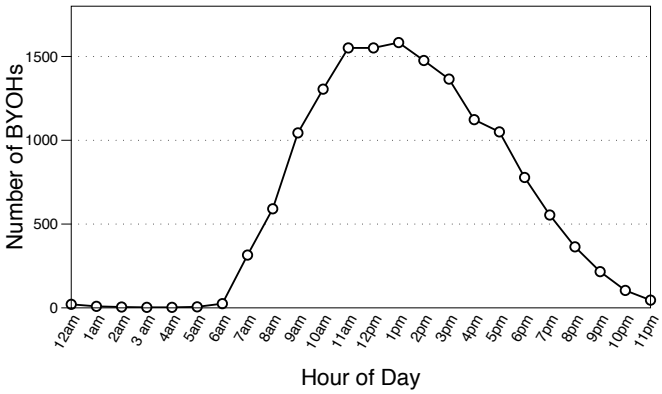

Fig. 5. Active BYOHs at each hour.

Number of time regions	Devices appearing (%)
1	39.4
2	42.27
3	17.69
4	0.64

TABLE III. TIME REGIONS VS. PERCENTAGE OF DEVICES.

Talking to internal servers and malicious sites. We found that 42% of BYOHs talk to internal servers (i.e., servers within the campus network) and 58% talk only to outside servers. We also examine the traffic sources to see if any BYOHs are connecting to blacklisted websites and IPs—we found no such devices. Overall, understanding the typical behavior of users could provide profiles and patterns that could help identify outliers and surprising behaviors.

C. Temporal behavior

We now study the temporal behavior of BYOHs.

Weekly and Daily Patterns: Our study indicates that BYOHs' patterns of appearance on the network follow weekly and daily patterns. Our daily observations along the entire month indicate that the number of BYOHs exhibits weekly periodicity: the number of devices increases on Monday, reaches its peak point on Tuesday and Thursday, and then decreases from Friday to Sunday. By considering these weekly and daily patterns, network operators have an opportunity to provision and use network resources more efficiently.

Intra-Day Behavior: To manage traffic on a per-hour basis, we need to understand the intra-day behavior of BYOHs. In Figure 5, we plot the number of active devices at each hour of the day. We observe that the number of active BYOHs (sending or receiving traffic) is low before 6 a.m. After 6 a.m., the number of active BYOHs increases and reaches a peak point during 11 a.m.–1 p.m. After 1 p.m., the number of active BYOHs decreases steadily until 11 p.m.

We further examine for how long devices are present during a day to enable a more "anthropocentric" analysis. Based on this observed behavior, which was consistent with other days,

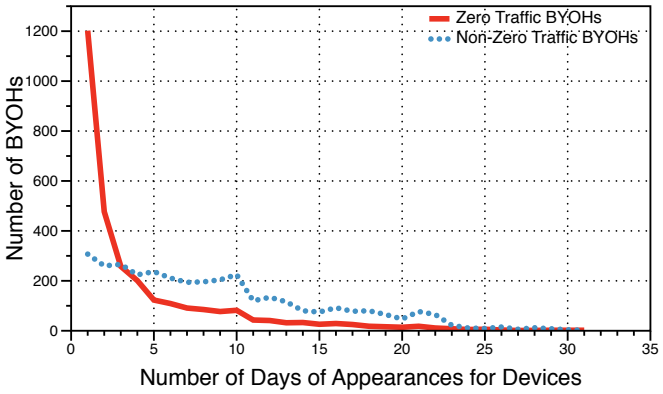

Fig. 6. Distribution of *days of appearance*.

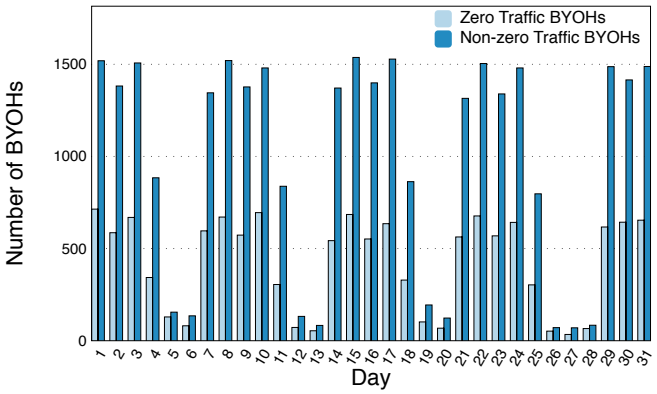

Fig. 7. Number of BYOHs per calendar day.

Group	Avg. # IP requests
Non-zero Traffic BYOHs	66.8
REG Non-zero Traffic BYOHs	95.7
NRE Non-zero Traffic BYOHs	21.7
Zero Traffic BYOHs	34.3
REG Zero Traffic BYOHs	84.1
NRE Zero Traffic BYOHs	16.6

TABLE IV. AVERAGE IP REQUESTS PER BYOH FOR EACH GROUP.

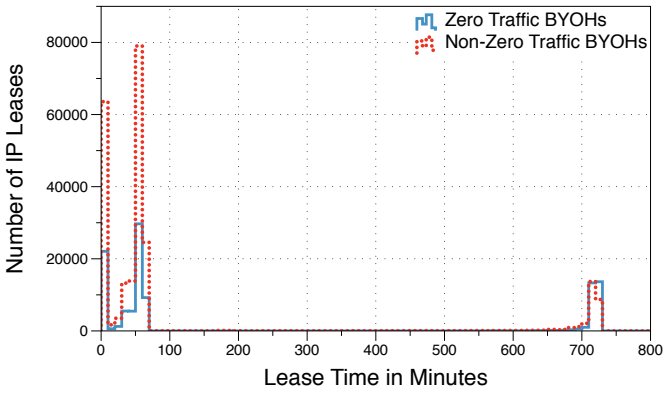

Fig. 8. Number of IP leases vs. lease time.

we define four distinct *time regions* during a day: Night (12 a.m.–6 a.m.), Morning (6 a.m.–12 p.m.), Afternoon (12 p.m.–6 p.m.), and Evening (6 p.m.–12 a.m.). In Table III, we show how many time regions devices appear in. We can see that most devices appear in 1 or 2 time regions, with 3 time regions being rare and 4 time regions uncommon. We further find that among the 1-time-region devices, Afternoon is the most popular. Among all devices that appear on two time regions, most devices appear during Morning and Afternoon, as expected. Note that while this behavior is unsurprising, we are the first to *quantify* these aspects.

Regularity of appearance: For every BYOH, we determine whether it appears regularly on campus. A human-centric way to define frequency is by counting how many distinct weeks the BYOH has appeared on the network—the intuition is that regular employees appear every week. This social behavior could allow us to estimate which group of devices are used by regular employees, and which group of devices are used by visitors, part-time contractors, and vagabonds. Recall that we classify BYOHs into REG and NRE, as discussed earlier in Section III. We apply this classification to both BYOHs with zero and non-zero traffic, and identify 2,896 REG BYOHs and 3,586 NRE BYOHs.

Vagabonds: In Figure 6, we see that over 23% of the BYOHs are vagabonds that appear only one day. Furthermore, 32% of mobile zombies (Zero-traffic BYOHs, see definition in Section IV-D), i.e., more than 1,000 BYOHs, belong to this group. Identifying this group could prompt several actions at the operational level. First, we could manage them separately, as they may not be employees. Second, we may want to give them short IP leases, until they prove that they actually need them for sending data.

D. Multi-level profiling and H-M-L model

We find that nearly half of the BYOHs are mobile zombies. The mobile zombie behavior can have significant implications for management purposes. First and foremost, this behavior is potentially problematic as IP addresses are often a limited resource. As a result, there is a need to allocate IPs in a more efficient way, for example, by not allocating IPs to known zombie devices. Second, it is a useful observation in estimating the required bandwidth for a group of BYOHs and defining user profiles. We highlight how our profiling method helps us identify interesting groups of BYOHs.

Days of appearance of both Zero Traffic and Non-zero Traffic BYOHs: We present the distribution of devices by number of days of appearance in Figure 6. We can see that most of the zero traffic BYOHs appear on few days, typically one or two. Furthermore, in Figure 7, we plot the number of non-zero and zero traffic BYOHs that appear on each calendar day. We observe that both non-zero traffic and zero traffic BYOHs have similar distributions in terms of days of appearance within a month, although there are fewer zero traffic BYOHs.

Intrigued, we investigated further and found that zero-traffic BYOHs that appear on only one day have a similar distribution across different weeks during the month. In other words, there is a fairly consistent presence of vagabond devices on a daily basis. In Table IV, we show the average number of

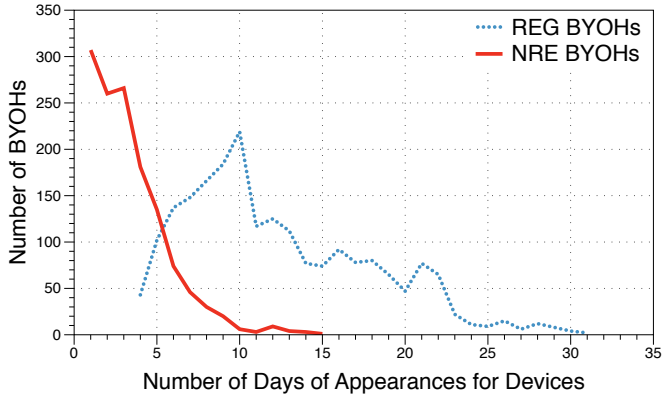

Fig. 9. Number of days that each REG and NRE BYOH appears.

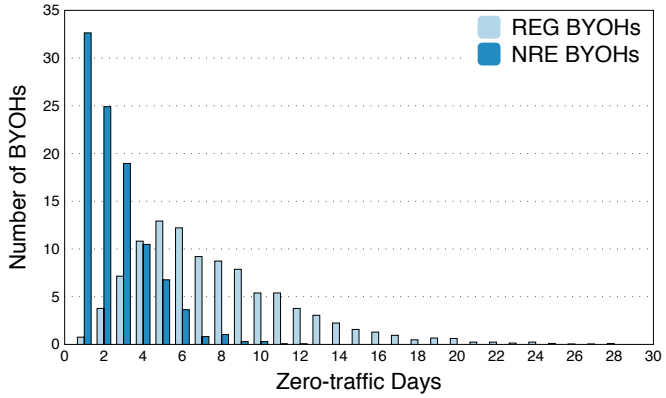

Fig. 10. Number of zero-traffic days in REG and NRE non-zero traffic BYOHs.

IP requests for each group (for the month of May). Non-zero traffic BYOHs have a higher intensity of IP requests than zero traffic BYOHs, as expected. In fact, non-zero traffic BYOHs place, on average, twice as many IP requests as zero traffic BYOHs. Such an observation can help administrators estimate the number of DHCP requests, which indicates a potential use of our device-centric profiling techniques.

Given this difference, we investigated whether there is a correlation between traffic volume and IP lease time. In Figure 8, we show the distribution of IP lease times for non-zero traffic and zero traffic BYOHs. The durations of IP lease time between zero traffic and non-zero traffic BYOHs are similar, which shows that a single IP allocation strategy is being used across all devices. This is an inefficient use of scarce IP resources, and a differential group-based IP allocation is necessary.

Regularity of Non-zero Traffic BYOHs: We now proceed to further profile non-zero traffic BYOHs in more detail, in a way that will help us define the thresholds for our H-M-L model. We focus this analysis on non-zero BYOHs to understand how device traffic, and to an extent user behavior, changes from day to day.

In Figure 9, we present the number of days of appearance for REG and NRE BYOHs. As expected, REG BYOHs appear more frequently than NRE BYOHs and most of the NRE BYOHs show up on fewer than 8 days. In addition, we

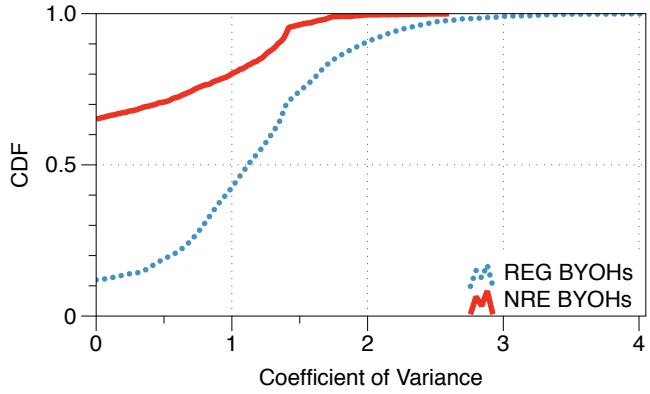

Fig. 11. Coefficient of variance of normalized traffic between REG and NRE BYOHs.

	L	M	H
Days of appearance	[0,8)	[8,20)	[20,+)
Daily average traffic (MB)	[0, 1.13)	[1.13, 10.01)	[10.01, +)

TABLE V. GROUP DEFINITIONS IN THE H-M-L MODEL.

see that 20 days seems to also be an important threshold in this distribution, that aligns with users appearing more than four days a week, every week, pointing to full-time campus employees. This higher frequency of appearances of REG BYOHs on campus networks results in a higher number of IP lease requests to the DHCP server. In Table IV, we can see that, in the categories of non-zero traffic BYOHs, the intensity of IP requests from REG BYOHs is significantly larger (by a factor of four) compared to that of NRE BYOHs. Table IV shows similar results when comparing REG with NRE in zero traffic BYOHs. Again, these observations can be helpful for estimating and provisioning purposes. *An* NRE *BYOH is more likely to have a zero-traffic day*, a term we use to describe a day on which a BYOH is present but with no traffic activity. In Figure 10, we see that the number of zero-traffic days in most REG BYOHs is greater than 2, largely skewed towards more days. This indicates that even non-zero traffic BYOHs do not necessarily use the network every day they appear. This is another opportunity for improving the efficiency of IP address usage, assuming the ability to identify such days. REG *BYOHs exhibit more variable daily traffic behavior.* In Figure 11, we plot the distribution of the coefficient of variance of the daily traffic volume for REG and NRE BYOHs. We see that roughly 58% of REG BYOHs have a coefficient larger than 1 ($x = 1$, $y = 0.42$) which indicates high variability.

In summary, **there are significant differences between the behaviors of** REG **and** NRE **BYOHs**. This suggests that: (a) our classification can identify groups with distinct behaviors, and (b) establishing different management policies can help optimize resource utilization.

Using the H-M-L model for a deeper investigation: Table V shows the thresholds that we identify using our H-M-L based classification of BYOHs. In Table VI, we show the distribution of non-zero traffic REG BYOHs (in percentages) for all possible groups in these two dimensions. The table

Days of appearance	Daily traffic		
	L	M	H
L	17%	9%	3%
M	29%	22%	8%
H	4%	5%	3%

TABLE VI. DAYS OF APPEARANCE V. DAILY TRAFFIC INTENSITY IN REG NON-ZERO TRAFFIC BYOHS.

HL BYOHs	LH BYOHs
Google (22.09%)	Google (21.09%)
Facebook (8.18%)	Amazon (16.03%)
Amazon (7.25%)	Level3 (12.15%)
Twitter (4.76%)	LimeLight (9.24%)
NTT (4.4%)	Akamai (7.11%)

TABLE VII. TOP 5 DOMAINS FOR HL AND LH BYOHS IN THE REG GROUP (PERCENTAGE IS THE TRAFFIC FRACTION OF TOTAL TRAFFIC FROM THAT GROUP OF DEVICES).

provides a quick and intuitive snapshot of the activity. For example, we can identify a specific group of interest that we want to monitor and analyze further, or we can observe a surprising change in the size of a group. Such a change could signal a new trend in the user base. For example, an increase in the LH groups could indicate the emergence of a new high-bandwidth application used by low-appearance users.

As a case-study of our model, we further analyze two of the resulting groups. We find that 3% of REG BYOHs are in group LH: low days of appearance and high daily average traffic. In addition, 4% of REG BYOHs form group HL: high days of appearance and low daily average traffic. These two groups of BYOHs have rather counter-intuitive behaviors, which we investigate next by examining the applications that these two different groups use. To do that, we resolve the IP addresses to domain names, as we do not have access to the HTTP headers. In Table VII, we present the top five domain names for LH and HL BYOHs. We observe that most of the traffic in either group is with Google. This is not surprising, as Google is one of the most frequently accessed web sites and Google applications (e.g., Google Maps, Google Voice, Gmail) are widely used by BYOHs. Similarly, Amazon's cloud services serve many popular smartphone applications (e.g., Hootsuite and Foursquare). In the HL group, we can see that a sizable fraction of traffic goes to Facebook and Twitter, which are the most popular social network applications. Facebook typically uses Akamai to serve sizable static content (e.g., video), and uses its own servers to serve dynamic content directly (e.g., wall posts). However, in the LH group, a lot of traffic goes to content delivery networks (CDNs), such as Limelight and Akamai, that deliver large volume traffic (e.g., video). These domain differences between LH and HL groups could explain why LH devices generate a lot of traffic, while HL devices do not. At the same time, it also provides an indication of the interests of end-users in that group.

V. RELATED WORK

No prior efforts have focused on comprehensively understanding the behavior of *individual* BYOH on multiple dimensions, with a view of BYOH management on campus networks.

Campus network studies. Prior research on DHCP has focused on studying and optimizing DHCP performance [27], [35]; these are earlier studies, around 2007, when smartphones and tablets were not widely used. A fingerprinting technique was proposed to classify devices by type and to manage IP lease time according to device type [17]. Here, we use a DHCP point of view to capture the operational properties of BYOHs in the protocol and control plane. Very few prior efforts focus on BYOH management over campus WiFi networks, which is our main focus here, and those efforts had largely different goals, from the characterization of traffic [9], [33], network performance [2] to mobility [26]. However, smartphones were only widely adopted recently. Later, Afanasyev et al. [25] indicated that the number of smartphone users significantly increased in WiFi networks. Deshpande et al. [30] compared the performance between 3G and WiFi networks and found that significant benefits could be obtained through the hybrid network design. Gember et al. [4] have studied the user-perceived performance differences between handheld devices and non-handheld devices(e.g., laptops) in campus networks. They found that smartphones tend to have smaller flow size and smaller range of flow durations. Chen et al. [36] have studied the network performance of smartphones in campus networks, focusing on delay and congestion. In contrast, we focus on BYOH management from the point of view of the network administrator and focus on *individual* BYOH behavior, behavior-based profiles, which are not addressed in the aforementioned studies.

General smartphone studies. In the broader area of smartphone studies, several studies focus on general modeling of wireless and smartphone traffic characterization focusing on public WiFi, 3G cellular networks, or residential networks. These studies do not look at the BYOH management problem from the point of view of a network administrator. Falaki et. al. [15] have analyzed network traffic from 43 smartphones and focused on TCP transfer performance, network congestion, and delay issues. The same group [16] also analyzes the diversity of smartphone usage, e.g., user interactions with devices and smartphone application usage patterns, in an effort to improve network and energy usage. Maier et al. [12] have analyzed smartphone traffic from the home, by analyzing DSL line traces. Huang et al. [22] have studied smartphones on 3G networks, and focused on application performance issues. Shafiq et al. [28] have studied the traffic of smartphones as aggregated over backbone Internet links. Livelab [7] is a measurement tool implemented on iPhones to measure iPhone usage and different aspects of wireless network performance. Sommers et al. [23] compare the performance of cellular and WiFi in metropolitan areas. Gember et al. [3] developed guidelines to accurately assess smartphone performance from the perspective of in-context. PROTEUS was developed to passively collect network information and forecast future network performance [31]. Qian et al. [11] investigated Redundancy Elimination techniques to achieve the reduction of smartphone traffic. Huang et al. [21] studied the impact of protocol and application behaviors on the network performance based on a large-scale LTE measurement. Nikravesh et al. [5] observed significant performance differences of mobile devices across different carriers, different access technologies, different geographic regions and over time. Erman et al. [20] investigated the impact of large events, e.g., Super Bowl, on the resource

provision of wireless networks. Fukuda et al. [24] studied the effectiveness of mobile traffic offloading in the wild. A tool QoE Doctor was proposed to measure and analyze the mobile app Quality of Experience, and better understand QoE problems across multiple layers [32]. MAPPER was developed to enforce management policies on diverse smartphones apps [6].

VI. Conclusions and future work

Taking a network administrator's point of view, the key contribution of our work is BROFILER, a systematic approach for profiling the behavior of BYOHs along four dimensions: (a) Protocol and Control Plane, (b) Data Plane, (c) Temporal behavior, and (d) across dimensions using the H-M-L model by considering the different levels of intensity in each dimension. We arguably provide the first multi-dimensional study of BYOHs, which shows how our profiling can provide interesting insights. Our work is a key step that could eventually lead to a complete picture of BYOH behavior from the network administrator's perspective. In the future, we want to expand BROFILER to: (a) reveal more interesting behaviors, by introducing more metrics within each dimension, and (b) apply our approach to address more BYOH-related management problems under the guidance of network administrators.

Acknowledgements

We would like to thank the anonymous reviewers for their feedback. This work was supported in part by funds from the University of Cincinnati CECH, by National Science Foundation awards CNS-1064646, SaTC 1314935 and NETS 0721889. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the National Science Foundation or the U.S. Government.

References

[1] Android Rapidly Repeats DHCP Transactions Many Times, October 2013. https://code.google.com/p/android/issues/detail?id=33590.

[2] A. Balachandran, G. Voelker, P. Bahl, and V. Rangan. Characterizing User Behavior and Network Performance in a Public Wireless LAN. In *ACM Sigmetrics*, 2002.

[3] A. Gember, A. Akella, J. Pang, A. Varshavsky, and R. Caceres. Obtaining In-context Measurements of Cellular Network Performance. In *ACM IMC*, 2012.

[4] A. Gember, A. Anand, and A. Akella. A Comparative Study of Handheld and Non-Handheld Traffic in Campus Wi-Fi Networks. In *PAM*, 2011.

[5] A. Nikravesh, D.R. Choffnes, E. Katz-Bassett, Z. Morley Mao, and M. Welsh. Mobile Network Performance from User Devices: A Longitudinal, Multidimensional Analysis. In *PAM*, 2014.

[6] A. Sapio, Y. Liao, M. Baldi, G. Ranjan, F. Risso and A. Tongaonkar. Per-user Policy Enforcement on Mobile Apps through Network Functions Virtualization. In *ACM MobiArch*, 2014.

[7] C. Shepard, A. Rahmati, C. Tossell L. Zhong and P. Kortum. LiveLab: Measuring Wireless Networks and Smartphone Users in the Field. In *HotMetrics*, 2010.

[8] D. Pelleg, A.W.Moore. X-means: Extending K-means with Efficient Estimation of the Number of Clusters. In *ICML*, 2000.

[9] D. Tang and M. Baker. Analysis of a Local-area Wireless Network. In *ACM MobiCom*, 2000.

[10] Enterasys. Trends in BYOD:Network Management and Security Are Leading Concerns, March 2013. http://blogs.enterasys.com/trends-in-byod-network-security-and-management-are-leading-concerns/.

[11] F. Qian, J. Huang, J.Erman, Z. Morley Mao, S. Sen, and O. Spatscheck. How to Reduce Smartphone Trafc Volume by 30%? In *PAM*, 2013.

[12] G. Maier, F. Schneider, and A. Feldmann. A First Look at Mobile Hand-held Device Traffic. In *PAM*, 2010.

[13] Gartner. , 2013. http://www.gartner.com/newsroom/id/2420816.

[14] Gartner. , 2013. http://www.gartner.com/newsroom/id/2573415.

[15] H. Falaki, D. Lymberopoulos, R. Mahajan S. Kandula and D. Estrin. A First Look at Traffic on Smartphones. In *ACM IMC*, 2010.

[16] H.Falaki, R.Mahajan, S. Kandula D.Lymberopoulos R.Govindan and D.Estrin . Diversity in Smartphone Usage. In *ACM MobiSys*, 2010.

[17] I. Papapanagiotou,E. M Nahum and V. Pappas. Configuring DHCP Leases in the Smartphone Era. In *ACM IMC*, 2012.

[18] IEEE Standards. Vendors of Mac Address, November 2012. http://standards.ieee.org/develop/regauth/oui/oui.txt.

[19] Increased Smartphone Usage Increases Network Complaints. http://www.telecompetitor.com/j-d-power-increased-smartphone-usage-increases-network-complaints/, March 2012.

[20] J. Erman, K.K. Ramakrishnan. Understanding the Super-sized Traffic of the Super Bowl. In *ACM IMC*, 2013.

[21] J. Huang, F. Qian, Y. Guo, Y. Zhou, Q. Xu, Z. Morley Mao, S. Sen, and O. Spatscheck. An In-depth Study of LTE: Effect of Network Protocol and Application Behavior on Performance. In *Sigcomm*, 2013.

[22] J. Huang, Q. Xu, B. Tiwana Z. M. Mao M. Zhang and P. Bahl. Anatomizing app Performance Differences on Smartphones. In *ACM MobiSys*, 2010.

[23] J. Sommers and P. Barford. Cell vs. WiFi: On the Performance of Metro Area Mobile Connections. In *ACM IMC*, 2012.

[24] K. Fukuda, K. Nagami. A Measurement of Mobile Traffic Offloading. In *PAM*, 2013.

[25] M. Afanasyev, T. Chen, G.M. Voelker, and A.C. Snoeren. Analysis of a Mixed-Use Urban WiFi Network: When Metropolitan becomes Neapolitan. In *ACM IMC*, 2008.

[26] M. Balazinska and P. Castro. Characterizing Mobility and Network Usage in a Corporate Wireless Local-area Network. In *ACM MobiSys*, 2003.

[27] M. Khadilkar, N. Feamster, M. Sanders and R. Clark. Usage-based DHCP lease time optimization . In *IMC*, 2007.

[28] M. Z. Shafiq, L. Ji, A. X. Liu and J. Wang. Characterizing and modeling Internet traffic dynamics of cellular devices. In *ACM Sigmetrics*, 2011.

[29] NIST (National Institute of Standards and Technology). Guidelines for Managing and Securing Mobile Devices in the Enterprise, July 2012. http://csrc.nist.gov/publications/drafts/800-124r1/draft_sp800-124-rev1.pdf.

[30] P. Deshpande, X. Hou, and S.R. Das. Performance Comparison of 3G and Metro-Scale WiFi for Vehicular Network Access. In *ACM IMC*, 2010.

[31] Q. Xu, S. Mehrotra, Z. Morley Mao, and J. Li. PROTEUS: Network Performance Forecast for Real-Time, Interactive Mobile Applications. In *ACM Mobisys*, 2013.

[32] Qi Alfred Chen, Haokun Luo, Sanae Rosen, Z. Morley Mao, Karthik Iyer, Jie Hui, Kranthi Sontineni, Kevin Lau. QoE Doctor: Diagnosing Mobile App QoE with Automated UI Control and Cross-layer Analysis. In *IMC*, 2014.

[33] T. Henderson, D. Kotz, and I. Abyzov. The Changing Usage of a Mature Campus-wide Wireless Network. In *ACM MobiCom*, 2004.

[34] T. Henderson, D. Kotz and I. Abyzov. The Changing Usage of a Mature Campus-wide Wireless Network. In *Computer Networks 52(14)*, pages 2690–2712, 2008.

[35] V. Birk, J. Stroik, and S. Banerjee. Debugging DHCP Performance . In *IMC*, 2004.

[36] X. Chen, R. Jin, K. Suh, B. Wang and W. Wei. Network Performance of Smart Mobile Handhelds in a University Campus WiFi Network. In *ACM IMC*, 2012.

[37] X. Wei, L. Gomez, I. Neamtiu and M. Faloutsos. ProfileDroid: Multi-layer Profiling of Android Applications . In *ACM Mobicom*, 2012.

[38] Y. Zhou and X. Jiang. Dissecting Android Malware: Characterization and Evolution. In *IEEE S&P*, 2012.