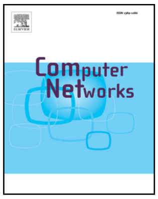
Computer Networks 114 (2017) 1–12

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet



# Characterizing the behavior of handheld devices and its implications

Xuetao Wei[a,*], Nicholas C. Valler[b], Harsha V. Madhyastha[c], Iulian Neamtiu[d], Michalis Faloutsos[e]

[a] University of Cincinnati, United States
[b] Crowdcompass, Inc., United States
[c] University of Michigan, United States
[d] New Jersey Institute of Technology, United States
[e] University of California, Riverside, United States

ARTICLE INFO

*Article history:*
Received 3 February 2016
Revised 21 December 2016
Accepted 3 January 2017
Available online 4 January 2017

*Keywords:*
Handheld devices
Behavior
Measurement
Management

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 as they have no control over these devices and no solid understanding of the behavior of these emerging devices. In order to cope with the impact of these BYOHs on current existing network management infrastructures, we identify two tightly-coupled questions that network administrators need to answer: (a) how do these BYOHs behave? and (b) how can we manage them more effectively based on the understanding of their behaviors? In response, we design and deploy Brofiler, a framework that could enable network administrators to effectively manage BYOHs via behavior-aware profiling. Our behavior-aware profiling captures the behaviors of each individual BYOH and improves the visibility on managing these BYOHs. In detail, the contributions of our work are three-fold. First, we present Brofiler, a time-aware device-centric approach for grouping devices into intuitive behavioral groups from multiple perspectives, including data plane, temporal behavior, and the protocol and control plane. 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. For example, we find that 70% of the BYOHs generate 50% of their monthly data traffic in one day, while remaining mostly idle the rest of the month. In addition, 68% of BYOHs do not conform to DHCP protocol specifications. Third, we present the implications of our study based on the framework in DHCP management, bandwidth management and access control. Overall, our approach could enable network administrators better understand and manage these new emerging devices for their networks in the post-PC era.

## 1. Introduction

In the post-PC era, 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 worldwide PC shipments decline while smartphone sales grew rapidly [1]; hence the *Bring Your Own Handheld-device* (BYOH) practice is going to increase. However, though the emergence of these devices changes the society rapidly, current network management infrastructures evolve slowly to accommodate them, which creates the challenges for the network administrator to effectively manage these devices in their networks. We use the term BYOH to describe only smartphones and tablets, in accordance with the National Institute of Standards and Technology's definition [2]. 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 as the loss of visibility into BYOHs brings the challenges to network administrators and add another layer of complexity on network management [3]. In detail, 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 and

[*] Corresponding author.
*E-mail address:* weix2@ucmail.uc.edu (X. Wei).

http://dx.doi.org/10.1016/j.comnet.2017.01.003
1389-1286/© 2017 Elsevier B.V. All rights reserved.

various versions of each operating system compared to other computing devices, and (d) most importantly, the apps that can run on them introduce a slew of management challenges [3–7].

The fundamental problem we address here is: what does the network administrator need to know about BYOHs? Specifically, we identify two key questions we want to answer in this paper: (a) how do these devices behave? and (b) What are the implications to manage operational concerns, such as the stress exerted on network resources?

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 better manage them.

Most prior efforts have focused on studying either the aggregate network traffic incurred by smartphones and tablets, or performance and network protocol issues, such as TCP and download times or mobility issues [8–12]. Such aggregate network traffic behavior hides more useful and fine-grained information to network administrator, which hinders them to gain the visibility on these BYOHs. 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 makes it difficult to monitor and control the behavior of these devices [4] in their networks. This is even worse as mobile apps are blooming. Therefore, we first need an approach that could effectively capture and reveal the behavior of each individual BYOH for network administrators so that they can better manage their networks. To the best of our knowledge, no prior work has focused on understanding individual BYOH behavior in campus networks, with a view towards better managing and provisioning network resources on-the-fly. In this paper, we extended our earlier paper [13] by explaining our profiling framework with more details on the background and architecture, extending the multi-dimensional behavior analysis towards operational concerns of BYOH management based on our framework, revising the operations of DHCP address allocation with informed profile information of BYOHs, recommending data usage quotas for the BYOHs, and enhancing the access control with fine-grained behavior profiles. We discuss related work in Section 6.

### *1.1. Contributions*

In this paper, we propose Brofiler (BYOH profiler), 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. A key advantage of our approach is that it is easy to deploy: 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 developing effective management of BYOHs. Specifically, our work has two implications: (1) BYOHs need to be managed explicitly as they behave in unique and unexpected ways, and (2) there are significant opportunities from tailored management strategies.

Our major contributions are highlighted below:

a. The Brofiler approach. We present 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, including the data plane, temporal behavior, the protocol and control plane, and the combined multiple dimensions. We also describe how it can form the foundation of an effective BYOH management system (Section 3). Brofiler analyzes and profiles BYOHs across these 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 4.4) 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 from a large campus (in Section 4): device access logs collected over the entire year, involving 22,702 BYOHs, and traffic data logs during one month involving 6482 BYOHs. 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 [14]). 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.

c. The Implications of Our Study. Based on our profiling, we present the implications of our study that go a long way toward improving device management, network usage and operations, and ultimately user experience. These implications could help network administrator better understand the operational issues of BYOHs and manage their networks. Finally, in a more open-ended case-study, we show how our profiling can help start the discussion towards security enhancement, using access control as an example.

## 2. 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. Another dataset (denoted as Traffic-May) is network flow-level traffic for BYOHs during the month 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 6482 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 [15]. 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 [15,16]. Table 1 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 the year, which was a leap year.

**Table 1**
Distribution of devices in dataset DHCP-366.

| 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% |

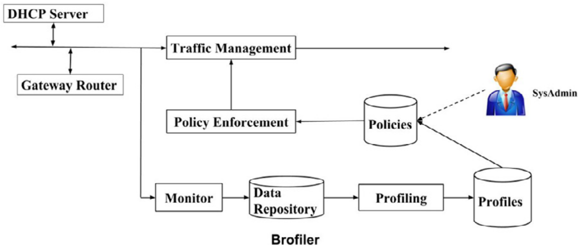


**Fig. 1.** System architecture of Brofiler.

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.

## 3. 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.

### *3.1. Architecture*

At the same time, our profiling system can be integrated in a policy and traffic management system as shown in Fig. 1. The monitor module of Brofiler first monitors the traffic, which includes the network flow-level traffic and DHCP traffic, and stores the traffic data into the data repository. The traffic data could be collected both at the gateway router (flow-level traffic data) and from DHCP server logs in the target campus, and stored in the data repository module. Then, Brofiler runs the profiling module on the collected data and generates the corresponding profiles. Our profiles include the profiling results from three dimensions, data plane, temporal behavior, and the protocol and control plane, and also the results of combining these dimensions. The information of profiles could be used by network administrators to develop the management policies. Then, the policy enforcement module could configure these policies to the traffic management module [17]. Network administrator could manage the traffic via the traffic management module with these configured policies. This whole process will be performed periodically to update the corresponding profiles due to the evolving device behavior.

### *3.2. Profiling*

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 3040 zero traffic BYOHs and 3442 non-zero traffic BYOHs. We present the details in Section 4.2.

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 students 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–May 5), Week 2 (May 6–May 12), Week 3 (May 13–May 19), Week 4 (May 20–May 26), and Week 5 (May 27–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 4.3.

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. In Section 5.3, we also examine whether a BYOH communicates with internal servers, which could be benign or raise security concerns. We present details in Section 4.1.

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 [18] on our data to identify these three clusters and derive the thresholds, which correspond to our levels. Note that, the major benefit of using X-means clustering algorithm is that network administrators don't need to supply the number of clusters in advance; the number of clusters is derived based on the target dataset automatically. However, if the network administrator wants to have the control on the thresholds, our framework can easily replace X-means clustering algorithms with the K-means clustering algorithm or other clustering algorithms.

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.

The Utility of our Approach. In order to showcase how Brofiler helps us identify interesting groups of devices or 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 develop the effective management of BYOHs, as we do in Section 5.

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 as stated in Section 4. Note that, we use data from the month of May, where we have both DHCP, DHCP-May, and traffic information, Traffic-May. We now start 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.

## 4. 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 3.2.

In the following, we will first present the profiling study based on three dimensions used in Brofiler, namely, the protocol and control plane, data plane, and temporal behavior, and then the multi-dimensional grouping using the H-M-L model.

### 4.1. Protocol and control plane

The dimension of the protocol and control plane 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 our paper, 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 [14]) in BYOHs are responsible for this misbehavior and argues that this erratic behavior is not due to the events of disconnection, reconnection and roaming [14]. This observation suggests that network administrators should monitor and diagnose protocol operation behaviors from BYOHs in order to detect malfunctioning devices. Given our profiling information on each BYOH, we could revise the DHCP protocol to react more intelligently to these malfunctioning devices and ensure the normal operation of IP allocation.

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,

**Table 2**
Top 5 HTTPS domains in our data by percentage of HTTPS traffic.

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

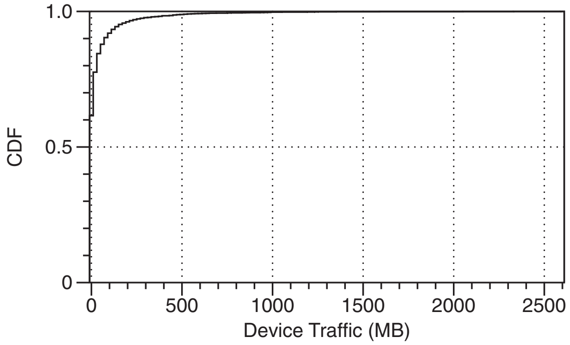


**Fig. 2.** Distribution of traffic volume per BYOH.

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 min, whereas that of BYOHs is 2.6 h. 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 [10,19]. 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 2). We can 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 can also 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. Such fine-grained information could provide more context-aware knowledge on improving the access control. 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).

### 4.2. Data plane

The data traffic generated by the devices is an important dimension to capture the behavior of BYOHs when they interact with the networks. There are many different aspects and properties of traffic. 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.

Traffic Volume: In Fig. 2, 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 (see Section 5.2).

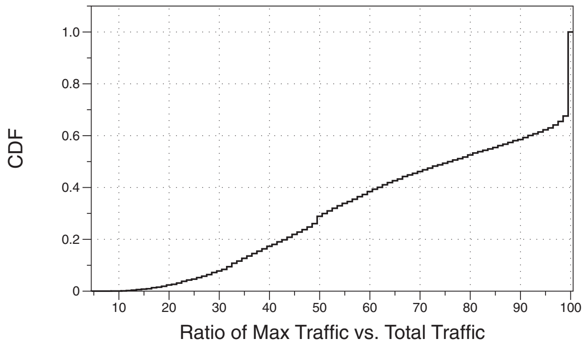


**Fig. 3.** Ratio of maximum daily traffic volume over total monthly traffic for each device.

Traffic Dynamics: A natural question to ask is whether the traffic behavior is consistent day to day. We find that it is not. In Fig. 3, 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 strategy will need to consider these factors. For example, different data usage quotas could be assigned to different BYOHs in order to ensure the effective and fair operation of data usage. In fact, we will see how these implications can help network administrators to improve the management of BYOHs in Section 5.2.

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, in return, to enhance the access control with more fine-grained profile information.

### 4.3. Temporal behavior

In the dimension of temporal behavior, our Brofiler profiles devices based on temporal behavior by 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 different users appear on the campus network in different temporal patterns. Note that, the profiling criteria depends on the context, complexity, the nature of the network, e.g., campus versus enterprise or a government network. 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

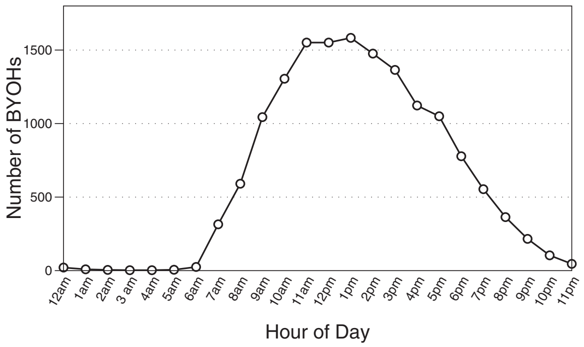


**Fig. 4.** Active BYOHs at each hour.

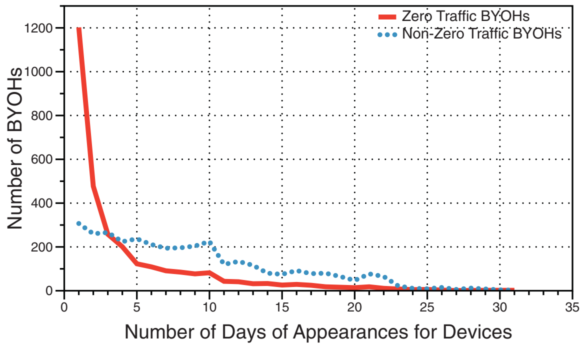


**Fig. 5.** Distribution of *days of appearance*.

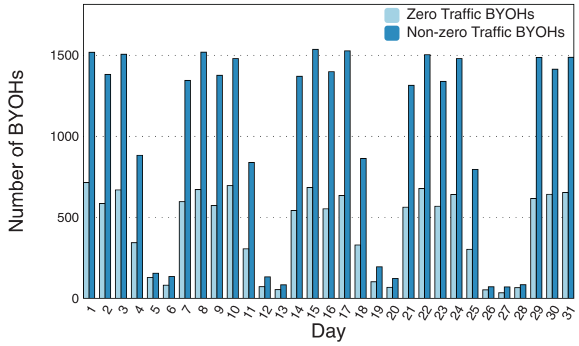


**Fig. 6.** Number of BYOHs per calendar day.

**Table 3**
Time regions vs. percentage of devices.

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

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 Fig. 4, 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, 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 3, 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 3. We apply this classification to both BYOHs with zero and non-zero traffic, and identify 2896 `REG` BYOHs and 3586 `NRE` BYOHs.

Vagabonds: In Fig. 5, 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 4.4), i.e., more than 1000 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. Overall, these temporal behavior analyses not only help network operators identify the operational issues caused by these devices, but also offer the opportunities to better manage the BYOHs.

### 4.4. Multi-level profiling and H-M-L model

In previous subsections, we only present the study on each individual dimension, which shows the interesting profiling results. In this section, we take a step further to conduct the cross-dimension analysis for a deeper investigation. 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 Fig. 5. We can see that most of the zero traffic BYOHs appear on few days, typically one or two. Furthermore, in Fig. 6, 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

**Table 4**
Average IP requests per BYOH for each group.

| 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 |

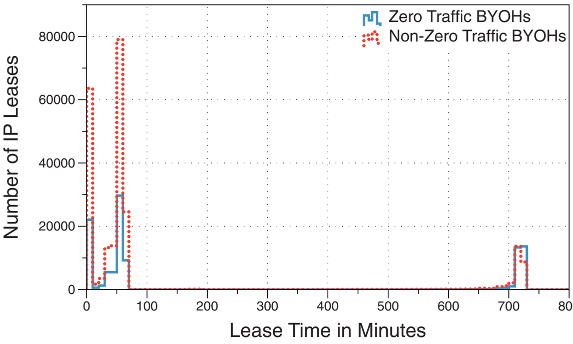


**Fig. 7.** Number of IP leases vs. lease time.

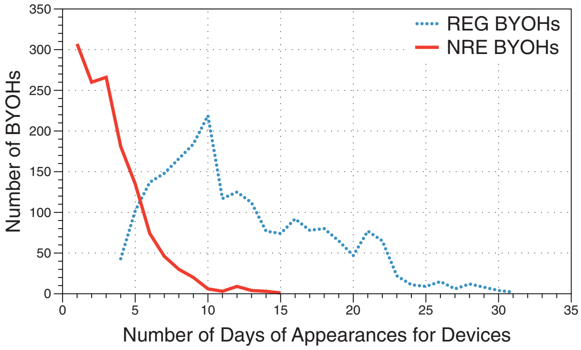


**Fig. 8.** Number of days that each REG and NRE BYOH appears.

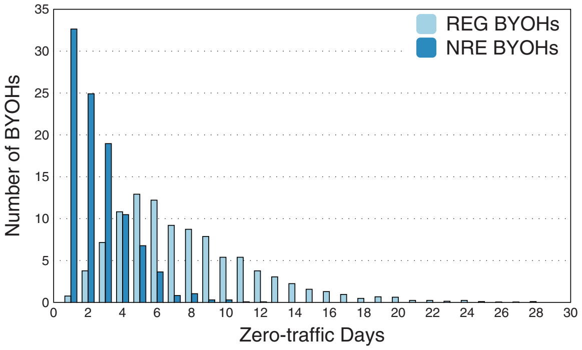


**Fig. 9.** Number of zero-traffic days in REG and NRE non-zero traffic BYOHs.

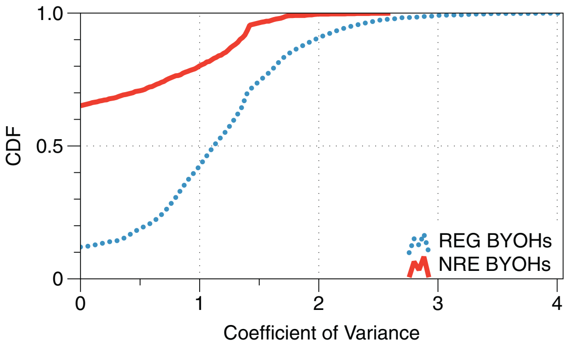


**Fig. 10.** Coefficient of variance of normalized traffic between REG and NRE BYOHs.

fairly consistent presence of vagabond devices on a daily basis. In Table 4, we show the average number of 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 Fig. 7, 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 Fig. 8, 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. As a point of reference, some students have classes on Tuesdays and Thursdays, which would lead to 10 days of appearance in our dataset. In addition, we 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 students and 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 4, 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 4 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 Fig. 9, 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 Fig. 10, we plot the distribution of the coefficient of variance of the daily traffic volume for REG and NRE BYOHs. We see that roughly 23% of REG BYOHs have a coefficient larger than 1 $(x = 1,\ y = 77)$ which indicates high variability.

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

Using the H-M-L model for a deeper investigation: Table 5 shows the thresholds that we identify using our H-M-L based classification of BYOHs. In Table 6, we show the distribution of non-zero traffic REG BYOHs (in percentages) for all possible groups in these two dimensions. The table provides a quick and intuitive

**Table 5**
Group definitions in the H-M-L model.

| | 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 6**
Days of appearance v. daily traffic intensity in REG non-zero traffic BYOHs.

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

**Table 7**
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).

| 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%) |

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. Overall, these fine-grained profiles could be correlated with the operational issues brought by BYOHs.

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 7, 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.

## 5. The implications of our study

So far, we have applied the Brofiler framework to profile BYOHs along multiple dimensions and to identify groups with common behaviors. Here, we close the loop by showing how Brofiler could help administrators better manage their network. We revisit the implications that our observations have, and propose some recommendations to some of the operational issues that we have identified focusing on the efficient use of resources.

A major advantage of our approach is that it is easy to deploy without requiring any software installation on the device, or device registration. It can be deployed by a network administrator, and it will learn BYOH behavior on-the-fly, and label and manage devices according to a desired strategy, as we explain below. However, the administrator has the ability to label specific devices (e.g. the tablet of the president) and treat them differently.

Note that we do not claim that our approaches are the best (or the only) approaches. Rather, we showcase the benefit of having a deep and intuitive understanding of the BYOH traffic, which our approach provides.

### *5.1. Informed DHCP address allocation*

In the previous sections, based on the profiling results from the perspective of the protocol and control plane in Brofiler, we found that the operation of DHCP in our network is far from ideal: (a) BYOHs issue a large number of non-conforming IP lease requests, (b) nearly half of BYOHs acquire IP addresses, but do not send any traffic over campus network, and (c) many "vagabond" BYOHs appear in the campus network. To address these issues, we propose a strategy, based on two design principles, for tailoring IP lease allocations to BYOHs.

Principle #1: Device-centric privileges: The on-the-fly learning of our approach could lead to the creation of device profiles, which can be stored and then used to provide different privileges and permissions to each device.

One reasonable implementation of device-centric management could be as follows: (a) all never-before seen devices are given IP addresses, (b) verified zombie devices may be entered in a blacklist that will preclude them from procuring an IP address, unless the user makes an explicit request to be removed from that list, (c) there could be a privileged-list of devices that receive preferential treatment, with respect to DHCP but also bandwidth as we will see later, and (d) periodically, we could flush old entries in the database and the lists. This way, we allow visitors to use the network as guests, but we eliminate inefficiencies for devices that do not use the network anyway, and allow administrators to hard-code schemes for particular devices.

Principle #2: Compliance validation: To cope with non-conforming IP lease requests, we mandate that the DHCP server ignore non-conforming IP lease requests by maintaining a BYOH's current state of IP lease allocation. If a BYOH currently holds one active IP lease, the DHCP server will ignore any subsequent non-conforming or unnecessary IP lease requests.

Our intention is not to find the optimal protocol. Rather, our goal is to reveal that some simple strategies can be used to improve DHCP operation. All our principles require the system to maintain client state. The features necessary to implement such a scheme are supported, though not required, by the DHCP RFC [20]. NAT (Network Address Translation) is an alternative, but NAT has its own issues that sysadmins often want to avoid [21]. Managing address allocation properly is an issue of efficient resource usage which management approaches should strive for, as resource waste can be particularly problematic in large networks.

### 5.2. Data usage quotas

The large number of BYOHs and the rapid growth in data demands can potentially complicate the task of managing the campus network and fill up the available capacity. We found that in our case, this increase is driven by a small fraction of BYOHs: 6% of BYOHs generate 82.1% of total BYOHs' traffic, as noted in Section 4.2. The lopsided data usage patterns of these "heavy users" could affect other users. Therefore, we set out to recommend to place limits on the data usage per device. In a manner similar to *data plan* limits imposed by commercial wireless cellular carriers or ISPs, we propose to impose a maximum volume of data that each BYOH consumes per month and per day, namely monthly quota and daily quota, which may be tailored by the device profiles obtained from the Brofiler. We envision that it could ensure the fair share of bandwidth.

In practice, bandwidth management is a concern: the consulted network administrators said that they impose traffic shaping near their egress point of the network. They also expressed the desire to ensure that users get a "fair" share of bandwidth, and they were interested in the idea of grouping users based on their profiles (e.g., users who are streaming) and managing the usage per groups. limiting heavy-hitters on a daily or monthly basis can ensure that light users are not dominated by heavy-hitters. The specifics of the solution have to do with the network architecture, the user base behavior, and the optimization goals (fairness, differentiated services, resource utilization, etc.), and extends beyond the scope of this work.

### 5.3. Towards enhancing the access control

Security is one of the most pressing BYOHs concerns inside the network perimeter of a campus or an enterprise. Management of security risks introduced by BYOHs is still in its infancy [4]. The concern stems from: (a) lack of standard access control schemes for BYOHs in contrast to the well established schemes and tools to ensure the safety of laptops and desktops, (b) large variations among devices OS or device OS versions, and (c) the vulnerabilities introduced by apps, easily bought for as little as $0.99, whose behavior and potential risks are not well-understood.

Brofiler's profiling can provide a basis to *start the discussion* on providing security enhancement, and enable administrators to *begin reasoning* about the access control scheme. As a showcase, we discuss how an administrator can think about enhancing the access control. This case-study is more tentative than the two previous studies, where we demonstrated tangible benefits. Nevertheless, we show that having intuitive groups can provide a good starting point for developing a better access control for BYOHs.

We now state our assumptions; while reasonable to us, in practice the specific needs for the access control might vary across organizations or administrators. For the sake of the case-study, we focus on non-zero traffic BYOHs. Specifically, we could assess the security threat of each such device considering three aspects of its behavior: (a) is the device talking to Internal Servers (IS)? (b) does it belong to the group of non-regular devices (NRE)?, (c) does the device use unencrypted HTTP flows (HTTP)? Clearly, the three aspects capture indications that the BYOH may pose a risk. The first aspect captures whether the device accesses sensitive information of the institution. The second aspect represents the regularity of the BYOH with the consideration that a regular BYOH could be treated differently from an ephemeral BYOH. The third aspect provides an indication of how security-sensitive is the device owner or the apps that the device runs. For example, using unencrypted traffic could allow an eavesdropper to gain access to sensitive personal information, and subsequently, allow the hacker to impersonate that user, including potentially the student account, and thus putting the network at risk.

**Table 8**
The number of affected devices after enforcing blocking strategies at a group level.

| Blocking Strategy | Affected users (%) | Affected flows (%) |
|---|---|---|
| Block-All-IS | 42 | 19 |
| Block-NRE-IS | 14 | 2.7 |
| Block-All-IS-HTTP | 40 | 11 |
| Block-NRE-IS-HTTP | 12 | 1.8 |

While this is not a comprehensive list of security aspects—depending on the scenario and the network configuration, other issues may also be of interest in the context of device security—note how Brofiler-supplied information allows administrators to quantify risk and design approaches to mitigate this risk.

Given this profile-driven approach, an administrator could restrict access to or completely block traffic to potentially sensitive resources for certain groups of BYOHs.

For example, let us assume that the goal is to protect the internal servers from being accessed by BYOHs that raise concerns, e.g., vagabonds.

Based on the groups obtained above, we consider enabling different blocking strategies per group. We also provide an assessment of how many BYOHs will be affected by such a restriction in Table 8. Block-All-IS means we block all the connections that talk to internal resources from BYOHs. Block-NRE-IS means we only block the NRE BYOHs that talk to internal servers. Block-All-IS-HTTP or Block-NRE-IS-HTTP means we block BYOHs with unencrypted HTTP flows from all or NRE BYOHs correspondingly. To sum up, we have illustrated how behavioral profiles and groups facilitate an intuitive discussion on how to develop the access control for BYOHs. Evaluating the effectiveness of the aforementioned strategies in a real deployment is beyond the scope of this work; nevertheless, Brofiler has allowed us to draw up management strategies that have intuitive appeal.

## 6. 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 [22,23], which is related to the DHPC part of our work. However, these are earlier studies, around 2007, when smartphones and tablets were not widely used. In their work, a fingerprinting technique was first proposed to classify devices by type and to manage IP lease time according to device type [16]. In our paper, we use a DHCP point of view to capture the operational behaviors of BYOHs in both the protocol and control plane, which could offer a fine-grained behavior profiling of BYOHs with a new dimension. 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 [12,24], network performance [25] to mobility [26]. However, smartphones were only widely adopted recently, which makes these work not suitable for current networks that accommodate BYOHs. Later, Afanasyev et al. [27] indicated that the number of smartphone users significantly increased in WiFi networks, which again proves the importance to manage these BYOHs. Deshpande et al. [28] compared the performance between 3G and WiFi networks and found that significant benefits could be obtained through the hybrid network design. Gember et al. [10] 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. However, they studied only at the aggregated network traffic from these devices. Chen et al. [19] have studied the network performance of smartphones in campus networks, focusing on delay and congestion, which is different with our focus. In contrast, we focus on BYOH management from the point of view of the network administrator and focus on *individual* BYOH behavior, and study and group their 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 only focused on the aggregated traffic, and also do not look at the BYOH management problem from the point of view of a network administrator. Falaki et. al.[8], as the first step, have analyzed network traffic from 43 smartphones and focused on TCP transfer performance, network congestion, and delay issues. The same group [9] 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, which is not our focus in this paper. Later, Maier et al. [11] have analyzed smartphone traffic from the home, by analyzing DSL line traces. From the large-scale infrastructure perspective, Shafiq et al. [29] have studied the traffic of smartphones as aggregated over backbone Internet links. Sommers et al. [30] compare the performance of cellular and WiFi in metropolitan areas. Gember et al. [31] developed guidelines to accurately assess smartphone performance from the perspective of in-context. Nikravesh et al. [32] observed significant performance differences of mobile devices across different carriers, different access technologies, different geographic regions and over time. Erman et al. [33] investigated the impact of the large event on the resource provision of wireless networks. Fukuda et al. [34] studied the effectiveness of mobile traffic offloading in the wild. Ashkan et al. performed a large-scale measurement study to improve the visibility into mobile network performance [35]. Moreover, mobile apps are the important element in the BYOH devices. Huang et al. [36] have studied smartphones on 3G networks, and focused on application performance issues. PROTEUS was developed to passively collect network information and forecast future network performance for mobile apps [37]. Qian et al. [38] investigated Redundancy Elimination techniques to achieve the reduction of smartphone traffic. Huang et al. [39] studied the impact of protocol and application behaviors on the network performance based on a large-scale LTE measurement. A tool QoE Doctor was proposed to better understand Quality of Experience problems across multiple layers [40]. A large scale traffic study on thousands of cellular towers was conducted to derive a powerful model on the traffic pattern in urban environment [41],and this model could enable various applications. Again, all these work focused on general network performance and application issues from smartphones, e.g., delay, congestion, application usage and optimization. In addition, MAPPER was developed to enforce management policies on diverse smartphones apps [3], which showed unique challenges to manage BYOH devices in the enterprise networks.

Network management related work. Network management has been studied extensively by researchers on different topics. For example, some studies focused on the network management's middleware and policy implementation. Rendon et al. [42] jointly used the situation management and mashup technologies for the network management in order to facilitate the daily work of network administrators. Han et al. [17] surveyed a variety of policy languages that are used to express the intentions of network administrators and explained how the languages are used to implement the policies. Due to the blooming of Internet of Things(IoT) devices, Sicari et al. [43] presented a framework to enforce the policies for the security and data quality issues from IoT devices and services. Our work focused on profiling the behavior of BYOHs, which could later help administrators to develop the corresponding policies and integrate them into the policy enforcement. Recently, software-defined networking has been introduced to network management that provides the flexible management of the networks. For example, a software-defined networking based policy enforcement framework was proposed to efficiently manage the traffic though the middle-boxes [44]. In Cellular networks, Sou et al. [45] presented a reference model to perform the application awareness with the data flows in order to realize the application-based charging management. Furthermore, an advanced mobility management approach was proposed to manage the two-tier LET-Advanced network by jointly considering the impact of user mobility, interference, and power consumption [46]. In addition, Sanchez et al. [47] presented an experimental study on the connectivity management on three mobile operating systems, Android, iOS and Windows. This work focused on the connectivity management in handover technologies, which is different with our work. A scalable solution was proposed to automatically balance the traffic from the heterogeneous network operation environments [48] in order to optimize the target network. All these works focused on either traffic optimization, or middlewares, or policy language and enforcement. However, our paper focused on how to manage the BYOHs, which are not controlled by installing any software, effectively and systematically via profiling the individual BYOH behavior from a multi-dimensional perspective.

## 7. Conclusions

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. Finally, we show that using profiles, a network administrator can develop effective strategies for managing BYOHs.

## Acknowledgments

This work was supported in part by funds from the University of Cincinnati CECH, as well as 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] Gartner, 2015, http://www.gartner.com/newsroom/id/3010017.
[2] NIST (National Institute of Standards and Technology), Guidelines for managing and securing mobile devices in the enterprise, 2012, http://csrc.nist.gov/publications/drafts/800-124r1/draft_sp800-124-rev1.pdf.
[3] A. Sapio, Y. Liao, M. Baldi, G. Ranjan, F. Risso, A. Tongaonkar, Per-user policy enforcement on mobile apps through network functions virtualization, ACM MobiArch, 2014.
[4] Enterasys, Trends in BYOD:network management and security are leading concerns, 2013, http://blogs.enterasys.com/trends-in-byod-network-security-and-management-are-leading-concerns/.
[5] X. Wei, L. Gomez, I. Neamtiu, M. Faloutsos, ProfileDroid: multi-layer profiling of Android applications, ACM Mobicom, 2012.
[6] Y. Zhou, X. Jiang, Dissecting Android malware: characterization and evolution, IEEE S&P, 2012.

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

[8] H. Falaki, D. Lymberopoulos, R. Mahajan, S. Kandula, D. Estrin, A first look at traffic on smartphones, ACM IMC, 2010.

[9] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, D. Estrin, Diversity in smartphone usage, ACM MobiSys, 2010.

[10] A. Gember, A. Anand, A. Akella, A comparative study of handheld and non-handheld traffic in campus Wi-Fi networks, PAM, 2011.

[11] G. Maier, F. Schneider, A. Feldmann, A first look at mobile hand-held device traffic, PAM, 2010.

[12] T. Henderson, D. Kotz, I. Abyzov, The changing usage of a mature campus-wide wireless network, in: Computer Networks 52(14), 2008, pp. 2690–2712.

[13] X. Wei, N. Valler, H.V. Madhyastha, I. Neamtiu, M. Faloutsos, A behavior-aware profiling of handheld devices, IEEE INFOCOM, 2015.

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

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

[16] I. Papapanagiotou, E.M. Nahum, V. Pappas, Configuring DHCP leases in the smartphone era, ACM IMC, 2012.

[17] W. Han, C. Lei, A survey on policy languages in network and security management, Elsevier Comput. Netw. 56 (2012) 477–489.

[18] D. Pelleg, A.W. Moore, X-means: extending K-means with efficient estimation of the number of clusters, ICML, 2000.

[19] X. Chen, R. Jin, K. Suh, B. Wang, W. Wei, Network performance of smart mobile handhelds in a university campus WiFi network, ACM IMC, 2012.

[20] RFC, Dynamic host configuration protocol, 1997, http://www.ietf.org/rfc/rfc2131.txt.

[21] RFC, Architectural implications of NAT, 2000, ftp://ftp.ripe.net/rfc/rfc2993.txt.

[22] V. Birk, J. Stroik, S. Banerjee, Debugging DHCP performance, IMC, 2004.

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

[24] D. Tang, M. Baker, Analysis of a local-area wireless network, ACM MobiCom, 2000.

[25] A. Balachandran, G. Voelker, P. Bahl, V. Rangan, Characterizing user behavior and network performance in a public wireless LAN, ACM Sigmetrics, 2002.

[26] M. Balazinska, P. Castro, Characterizing mobility and network usage in a corporate wireless local-area network, ACM MobiSys, 2003.

[27] M. Afanasyev, T. Chen, G.M. Voelker, A.C. Snoeren, Analysis of a mixed-use urban WiFi network: when metropolitan becomes neapolitan, ACM IMC, 2008.

[28] P. Deshpande, X. Hou, S.R. Das, Performance comparison of 3G and metro-scale WiFi for vehicular network access, ACM IMC, 2010.

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

[30] J. Sommers, P. Barford, Cell vs. WiFi: on the performance of metro area mobile connections, ACM IMC, 2012.

[31] A. Gember, A. Akella, J. Pang, A. Varshavsky, R. Caceres, Obtaining in-context measurements of cellular network performance, ACM IMC, 2012.

[32] A. Nikravesh, D.R. Choffnes, E. Katz-Bassett, Z.M. Mao, M. Welsh, Mobile network performance from user devices: alongitudinal, multidimensional analysis, PAM, 2014.

[33] J. Erman, K.K. Ramakrishnan, Understanding the super-sized traffic of the Super Bowl, ACM IMC, 2013.

[34] K. Fukuda, K. Nagami, A measurement of mobile traffic offloading, PAM, 2013.

[35] A. Nikravesh, DR. Choffnes, E. Katz-Bassett, Z.M. Mao, M. Welsh, Mobile network performance from user devices: alongitudinal, multidimensional analysis, PAM, 2014.

[36] J. Huang, Q. Xu, B. Tiwana, Z.M. Mao, M. Zhang, P. Bahl, Anatomizing app performance differences on smartphones, ACM MobiSys, 2010.

[37] Q. Xu, S. Mehrotra, Z.M. Mao, J. Li, PROTEUS: network performance forecast for real-time, interactive mobile applications, ACM Mobisys, 2013.

[38] F. Qian, J. Huang, J. Erman, Z.M. Mao, S. Sen, O. Spatscheck, How to reduce smartphone traffic volume by 30%? PAM, 2013.

[39] J. Huang, F. Qian, Y. Guo, Y. Zhou, Q. Xu, Z.M. Mao, S. Sen, O. Spatscheck, An in-depth study of LTE: effect of network protocol and application behavior on performance, Sigcomm, 2013.

[40] Q.A. Chen, H. Luo, S. Rosen, Z.M. Mao, K. Iyer, J. Hui, K. Sontineni, K. Lau, QoE doctor: diagnosing mobile app QoE with automated UI control and cross-layer analysis, IMC, 2014.

[41] H. Wang, F. Xu, Y. Li, P. Zhang, D. Jin, Understanding mobile traffic patterns of large scale cellular towers in urban environment, ACM IMC, 2015.

[42] O.MC. Rendon, F. Estrada-Solano, V. Guimarães, L.MR. Tarouco, LZ. Granville, Rich dynamic mashments: an approach for network management based on mashups and situation management, Elsevier Comput. Netw. 94 (2016) 285–306.

[43] S. Sicari, A. Rizzardi, D. Miorandi, C. Cappiello, A. Coen-Porisini, Security policy enforcement for networked smart objects, Elsevier Comput. Netw. 108 (2016) 133–147.

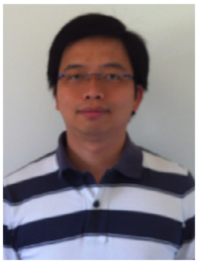
[44] ZA. Qazi, C. Tu, L. Chiang, R. Miao, V. Sekar, M. Yu, SIMPLE-fying middlebox policy enforcement using SDN, SIGCOMM, 2013.

[45] S. Sou, H. Hsieh, Modeling application-based charging management with traffic detection function in 3GPP, Elsevier Comput. Netw. 91 (2015) 625–637.

[46] S.D. Xenakis, N. Passas, L. Merakos, C. Verikoukis, Advanced mobility management for reduced interference and energy consumption in the two-tier LTE-Advanced network, Elsevier Comput. Netw. 76 (2015) 90–111.

[47] M.I. Sanchez, A.D.L Oliva, CJ. Bernardos, Experimental analysis of connectivity management in mobile operating systems, Elsevier Comput. Netw. 94 (2016) 41–61.

[48] J. Moura, C. Edwards, Efficient access of mobile flows to heterogeneous networks under flash crowds, Elsevier Comput. Netw. 107 (2016) 163–177.



**Xuetao Wei** has been a tenure-track faculty member at University of Cincinnati since January 2014, and is affiliated with both School of Information Technology and Department of Electrical Engineering and Computing Systems. He received his Ph.D. in Computer Science from University of California, Riverside in 2013. His research interests span the areas of cybersecurity, mobile computing and networking.

**Nicholas Valler** is a Senior DevOps Engineer for CrowdCompass by Cvent. He holds a Bachelor of Science degree in Computer Science from California State University, Long Beach, and a Doctorate in Philosophy from the University of California, Riverside. His research interests include Internet topology, epidemic spreading behaviors and network measurement techniques. In his role as a DevOps Engineer, he fancies himself a toolsmith and is dedicated to developing simple, elegant testing frameworks.

**Harsha V. Madhyastha** is an Assistant Professor in the EECS Department at University of Michigan. His research interests include distributed systems, networking, and security and privacy, and his work has resulted in award papers at the USENIX NSDI, ACM SIGCOMM IMC, and IEEE CNS conferences. He received Ph.D. and MS degrees from the University of Washington and a B.Tech. degree from the Indian Institute of Technology Madras, all in Computer Science and Engineering.

**Iulian Neamtiu** is an Associate Professor of Computer Science at the New Jersey Institute of Technology. His research interests are in programming languages, software engineering, and the smartphone side of systems and security.

**Michalis Faloutsos** is a faculty member at the Computer Science department in the University of California, Riverside. He got his bachelor's degree at the National Technical University of Athens and his M.Sc. and Ph.D. at the University of Toronto. His interests include, network protocols and measurements, network and systems security, and online social networks. With his two brothers, he co-authored the paper "On powerlaws of the Internet topology" (SIGCOMM'99), which received the "Test of Time" award from ACM SIGCOMM. His work has been supported by several NSF and DAPRA grants, including the prestigious NSF CAREER award with a cumulative of more than $10M in grants. His research has resulted in more than 15K citations, an h-index greater than 50, and an i10-index greater than 100. He is the co-founder of stopthehacker.com, a web-security start-up, which received two awards from the National Science Foundation, and got acquired in November 2013. In Aug 2014, he co-founded programize.com, which provides product development as a service.