Computer Sciences Department

Handheld vs. Non-Handheld Traffic: Implications for Campus WiFi Networks

Aaron Gember
Ashok Anand
Aditya Akella

Technical Report #1679

October 2010

UNIVERSITY OF WISCONSIN MADISON

Handheld vs. Non-Handheld Traffic: Implications for Campus WiFi Networks

Aaron Gember, Ashok Anand, Aditya Akella
University of Wisconsin, Madison
{agember,ashok,akella}@cs.wisc.edu

ABSTRACT

Smartphones, portable music players, and other handheld devices have become a major computing platform. Wherever users go, they utilize 3G and WiFi connectivity to access a wide array of Internet services. The small, mobile nature of these devices results in a unique mix of application and network usage. As more handheld devices connect to campus, enterprise, and community WiFi networks, administrators need to adjust their network's configuration to cope with the unique traffic characteristics of handhelds. Other studies have used WiFi and 3G wireless traces to analyze session, mobility, and performance characteristics for handheld devices. We complement these studies by examining how the content and flow characteristics of handheld traffic elicit network management changes. We analyze packet traces from the University of Wisconsin-Madison campus and computer science wireless networks, with 3 days of traffic for over 32,000 unique devices. Trends for handheld devices include a lower usage of UDP, a high volume of HTTP traffic, and a greater proportion of video traffic. We summarize key implications for network management and suggest configuration changes to maintain suitable network performance as handheld devices become the primary users of wireless networks.

1. INTRODUCTION

Handheld devices are quickly augmenting, and in some cases even replacing, laptops as the computing and Internet perusal platform of choice for many users on the go. Individuals are using smartphones, portable music players, ebook readers, and other handhelds to access a variety of Internet applications and data.

In this paper, we seek to understand what would happen as the number of mobile handheld devices devices becomes comparable to, or exceeds, the number of laptops on campus wireless networks. In particular, we wish to understand *what new approaches must be developed to manage campus networks in order to improve the experience of mobile handheld users, assuming current handheld access patterns hold?*

We focus on issues pertaining to the management of the wired backends, and of network-based services deployed across the backend. The characteristics of handheld traffic inform the management of the network for performance and security, and guide administrators in supporting and deploying new middleboxes and network services. To the best of our knowledge, these issues have not been considered in prior work. We ignore low-level wireless transmission, connectivity, and mobility issues as these have already been well studied [4, 10, 14].

Understanding these issues is both timely and important: According to a 2009 EDUCAUSE study of technology on college campuses, 51% of undergraduates own an Internet-capable handheld device and 12% plan to purchase a device within the next 12 months [19]. In addition, a recent PEW study comparing 2007 and 2009 wireless Internet usage among Americans found a 73% increase in the rate users went online with their handhelds [12]. While the number of non-handheld portable devices, such as laptops, is also growing, the number of handheld devices is growing at a much faster pace.

Although many handheld users have 3G data plans from a cellular carrier, 802.11 WiFi access is still a preferred Internet access mechanism, when available, because of its higher bandwidth, lower latency, and lower energy usage. WiFi networks in campuses, enterprises, and communities are seeing increasing numbers of handhelds as the popularity of these devices continues to grow.

The traffic characteristics of WiFi networks change as handheld devices represent a larger proportion of clients. Compared to wireless desktop and laptop users, handheld users access a different mix of Internet services and content. The small screen size and limited resources of handheld devices make some applications infeasible to run, while other applications are more appropriate for the "take anywhere" nature of these devices. Applications like web browsers and email clients are used on both handheld and non-handheld devices, but service providers tailor the content differently for the two classes of devices. Furthermore, the browser interface on handheld devices in itself places limitations on the range of both Internet-based and local network-based services that users can access. Thus, the network traffic of handheld devices is likely to differ in several crucial respects from non-handheld devices, e.g., in terms of flow lengths, protocol usage, access to services, the prevalence and nature of multimedia content, and temporal and spatial locality of content access.

We use traces of network traffic gathered at the University of Wisconsin-Madison from two independently-managed multi-AP wireless networks over a 3 day period to identify how handheld network traffic differs from the traffic characteristics of non-handheld devices. The traces contain over 32,000 unique clients, with about 15% of these being handhelds. We examine the traffic patterns, protocols and content of the transport and application layer traffic associated with handhelds. Our goal is to identify network management changes stemming from our observations, assuming these patterns remain the same but the number of handheld users grows.

Our key findings can be summarized as follows. The majority of handheld traffic (97%) is web, with small amounts of email traffic. In contrast, 82% of non-handheld traffic is web, with miscellaneous UDP traffic (14%) and internal services accounting for the remaining share. As handheld populations increase, administrators should focus less on providing internal services and instead focus on the quality of Internet access. In terms of TCP flow characteristics, handhelds tend to have smaller flows and a narrower range of flows durations. However, both types of devices have similar flow rates, with a median rate of 100 KBps. The similarity in flow rates for handhelds and non-handhelds implies that the aggregate network throughput is unlikely to change as handheld usage increases.

We look in-depth at HTTP traffic because it accounts for such a large share of traffic for both types of devices. Handheld devices access content from a narrower range of hosts than non-handheld devices; the top ten handheld hosts account for 74% of handheld HTTP content, while the top hosts for non-handhelds only account for 42% of non-handheld content. Less than 100 different types of content are accessed by handhelds, compared to over 275 types for non-handhelds. Administrators can deploy in-network security scanners for handheld devices with relatively few HTTP content signatures but scan a large majority of handheld traffic without imposing an energy burden on handhelds. The top content type for handhelds is video, accounting for 42% of HTTP traffic, compared to 23% of non-handheld HTTP content. The streaming video flows represent the largest, fastest, and highest throughput flows of all handheld TCP flows. We recommend administrators use traffic shaping, instead of resource reservation, to ensure the network meets the bandwidth demands and delay intolerance of streaming video flows.

The remainder of this paper is organized as follows. Section 2 presents background information on handheld device properties impacting traffic characteristics. Details on the data set and the methods used to isolate handheld and non-handheld traffic are presented in Section 3, followed by an overview of the client population in Section 4. We look at the protocols and services used by devices and the associated flow characteristics in Section 5. Section 6 looks in depth at HTTP and streaming media usage and the impact on network management. Section 7 considers a "chunk-based" content similarity system as a mechanism for reducing network transfers. Lastly, we present related work in Section 8 and conclude in Section 9.

Vendor	Model	CPU Speed	Memory	O/S	Connectivity
Apple	iPhone 3GS	600 Mhz	256 MB	iPhone OS	UMTS, EDGE, 802.11b/g
Apple	iPod Touch	532 Mhz	256 MB	iPhone OS	802.11b/g
BlackBerry	Storm	528 MHz	128 MB	BlackBerry OS	UMTS, EDGE
HTC	Nexus One	1 GHz	512 MB	Android	UMTS, EDGE, 802.11b/g
Palm	Pixi	600 MHz	8 GB	Palm webOS	UMTS, 802.11b/g

Table 1: Specifications for widely-used mobile devices

2. HANDHELD DEVICE PROPERTIES

Handheld devices have unique hardware and software properties that differentiate them from non-handhelds. Table 1 lists important hardware specifications of some well-known and widely-used handheld devices. Below, we highlight some properties of handhelds to provide context for our work and identify the potential impact of these properties on handheld traffic characteristics.

Hardware Constraints: Relative to desktops and laptops, handhelds have limited processing power and memory, translating to longer computation times and potentially longer data transfer times. Energy conservation techniques can also affect transfer duration and throughput. Using WiFi power save mode (PSM) is a well known technique for saving energy and has been widely studied [6, 13]. Screen size and keyboard interface is also a limitation that impacts the applications and services used on handhelds.

Specialized Content: Many web services offer specialized content for handheld devices, e.g., text-based pages with little or no graphical components. The typically goal is to decrease transfer sizes. Multimedia content served to handheld devices may also be provided at a lower resolution in an effort to decrease the volume of traffic sent to the device.

Multiple Connectivity Methods: Handheld devices often offer multiple forms of network connectivity: 802.11a/b/g WiFi, 3G, GSM, and other cellular protocols. In contrast, non-handheld devices typically only use 802.11 WiFi for network access. Since we only look at WiFi traffic from handhelds, it is possible that our analysis misses some network traffic originating from these devices. We know Apple iPods only feature WiFi connectivity, so we isolate traffic from these devices to determine if WiFi-only handhelds exhibit different trends from handhelds that also feature other forms of connectivity. At the same time, missing traffic traversing over 3G and other cellular protocols is not a major concern because our focus is on managing WiFi networks.

3. METHODOLOGY

We collect and analyze data for the University of Wisconsin-Madison campus wireless network (UW) and the University of Wisconsin-Madison computer science department wireless network (CS). From the UW network, full bi- directional packet traces were captured from six wireless aggregation points, covering about 80% of the approximately 2,400 APs on campus. Traces were captured over a period of 3 days during April 2010, yielding 8 TB worth of data. From the CS network, full bi-directional packet traces were captured from all APs for a period of 3 days in June 2010, yielding 50 GB worth of data. Traces from both networks only include traffic destined for hosts external to the wireless subnets; neither trace includes traffic sent between wireless clients.

3.1 Isolating Handheld Traffic

The packet traces contain data from all wireless clients connected to the network—laptops, smartphones, and other devices. Since we focus on the differences between handheld and non-handheld devices, we need to differentiate traffic based on device type. We rely on user-agent strings in HTTP packets for differentiation.

We consider handheld user-agents to have at least one of the following keywords: Android, ARCHOS, BlackBerry, CUPCAKE, FacebookTouch, iPad, iPhone, iPod, Kindle, LG, Links, Linux armv6l, Linux armv7l, Maemo, Minimo, Mobile Safari, Nokia, Opera Mini, Opera Mobi, PalmSource, PlayStation, SAMSUNG, Symbian, SymbOS, webOS, Windows CE, Windows Mobile, Zaurus.[1] This keyword list is based on common knowledge and published lists [22].

Note that it is possible for a particular device (i.e., a MAC address) to be identified as both handheld and non-handheld. This can occur if multiple types of devices exist behind a router which is connected to the wireless network. Alternatively, a user may "spoof" the user-agent in some browsers, causing conflicting identifications. We exclude such devices from both sets of traffic we analyze.

Other studies [4] have used Organizationally Unique Identifiers (OUIs) in MAC addresses to differentiate device types. However, not all manufacturers use different OUIs for different device types. Apple, for example, groups all types of devices (MacBooks, iPods, iPhones) into the same OUIs [4]. Also, OUIs can only provide hardware manufacturer identification, while user-agent strings often also contain operating system and application information.

As confirmation of our identification approach, we verify our categorization of mobile and non-mobile devices using OUIs. We found that the MAC addresses for all devices identified as mobile are registered to manufacturers known to make mobile devices (or components for them).

3.2 Unidentifiable Devices

Not all devices can be identified as mobile or non-mobile using user-agent strings in HTTP packets. About 17% of devices in the UW traces do not send any HTTP packets. To classify these devices, we resort to using OUIs.

We use the manufacturers of devices classified by HTTP user-agent to guide our classification of the remaining devices. Two lists of manufacturers are generated (for handheld and non-handheld devices) based on the OUIs of already classified devices. If a manufacturer appears in both lists, then all OUIs registered to the manufacturer are excluded from our list of classifiable OUIs. Any device whose OUI is registered to a manufacturer in our handheld or non-handheld list is classified accordingly. A small percentage of devices (14% for UW and 11% for CS) remain as unknown and are excluded from our analysis.

[1] The keywords for non-handheld devices are: Windows 7, Windows Vista, Windows XP, Windows Server, Windows NT, Intel Mac OS X, PPC Mac OS X, MacBook, iMac, Fedora, Ubuntu, Gentoo, SUSE, Linux x86_64, Linux i686, WiiConnect.

Device Type	UW	CS
Handheld	5060	9
Non-handheld	22485	90
Dual-identify	113	–
Unknown	4508	13
Total	32166	112

Table 2: Client counts by device type

Handheld Vendor	UW	CS
Apple	4337	6
HTC	134	–
Research in Motion (BlackBerry)	173	–
Motorola	118	–
Palm	113	1
Nokia	88	1
Samsung	20	–
Other	77	1

Table 3: Handheld counts by vendor

4. USER POPULATION

Over the 3 day capture periods, 32,166 unique clients connect to the UW network and 112 unique clients connect to the CS network. Table 2 shows the total number of clients of each type present in the trace data. Non-handheld devices account for the majority of clients in both networks. However, administrators from both networks provided anecdotal evidence that handheld devices are much more prevalent than in prior years, and industry and campus studies have shown the number of handheld devices is expected to continue increasing [19]. The number of laptop users on college campuses is also increasing (with desktop usage decreasing), but at a less rapid rate than the growth in handheld clients. The unique handheld characteristics we identify will become pronounced and have a greater impact on network management as the number of handheld clients and their network usage increases.

Table 3 lists the number of handheld devices by manufacturer. We see devices from 7 primary vendors, with Apple iPods and iPhones accounting for over two-thirds of all handhelds (85% in UW network and 66% in CS network).

4.1 Client Authentication

In both the UW and CS networks, wireless clients are required to authenticate with the network by providing user credentials. The CS network consists of only one subnet and SSID, but the UW network is broken into 134 subnets each with a separate SSID. When clients move within the

UW network, they need to re-connect and re-authenticate to a new SSID. Other studies have already shown handheld clients move more frequently than non-handheld clients [10, 4], and this re-authentication requirement places an additional burden on handheld clients. As the number of handheld devices grows, administrators should consider implementing authentication mechanisms that require users to only authenticate with the network once. Using certificate authentication, instead of requiring users to enter a username and password could further improve the handheld user experience.

Scaling of DHCP services has already occurred in most networks as handheld clients have increased. Network administrators from Marquette University provided anecdotal evidence of needing to increase the IP address pool in popular areas like the campus library due to rapid growth of the handheld client base. As handheld usage continues to grow, administrators need to further scale authentication and client addressing mechanisms. Other studies have also shown that handheld devices have shorter session durations [**?**, 7], so we recommend administrators shorten address assignment and authentication validity durations as handheld usage increases.

5. PROTOCOLS AND SERVICES

The protocols and services used by devices dictate the performance of an enterprise wireless network. Different protocols and services respond differently to bandwidth limitations and network congestion. They also contribute flows of varying sizes, durations, and frequencies to the overall traffic mix. As more handheld devices connect to the network, the varying protocol and service usage of these devices changes the mix and behavior of aggregate wireless traffic.

5.1 Network and Transport Protocols

At the highest level, we categorize traffic based on network and transport layer protocols. Table 4 shows the percentage of traffic in packets and bytes for each type of protocol in the UW and CS traces. As expected, the majority of traffic is TCP or UDP. The remaining traffic is IPSec—IP traffic tunneled over a secure connection—or network control traffic (ICMP, ARP, etc.).

A major difference in protocol usage between handheld and non-handheld devices is the amount of UDP traffic. Non-handheld devices use UDP for 20% of their traffic, while handheld devices use UDP for only 1.5%. In the face of congestion, this difference can impact network performance. As the number of handheld devices increases, a smaller percentage of traffic will be UDP and more traffic will be TCP. More TCP traffic results in devices using a fairer-share of network bandwidth when there are large numbers of competing flows. However, this increase in TCP flows also increases the amount of flow state that needs to be stored on NAT boxes, intrusion detection systems, or other network middleboxes which rely on knowledge of TCP flow state. The flow state increase will be most prevalent in wireless-only networks with a large percentage of handheld clients.

Protocol	**UW Handhelds**		**UW Non-handhelds**	
	% of Packets	**% of Bytes**	**% of Packets**	**% of Bytes**
UDP	5.9%	1.5%	25.7%	19.9%
TCP	92.0%	98.3%	74.0%	80.0%
IPsec	0.3%	0.05%	0.05%	0.03%
ICMP	0.1%	0.01%	0.2%	0.04%
Other (ARP, etc.)	1.5%	0.1%	0.2%	0.04%

(a) UW

Protocol	**CS Handhelds**		**CS Non-handhelds**	
	% of Packets	**% of Bytes**	**% of Packets**	**% of Bytes**
UDP	4.5%	1.7%	18.4%	14.4%
TCP	93.0%	98.0%	81.4%	85.6%
IPsec	–	–	0.05%	0.05%
ICMP	0.1%	0.02%	0.05%	0.01%
Other (ARP, etc.)	2.3%	0.3%	0.1%	0.01%

(b) CS

Table 4: Network/Transport protocol usage

Category	**Transport Protocol**	**Applic. Protocol**	**UW Handheld**	**UW Non-Handheld**	**CS Handheld**	**CS Non-Handheld**
Web	TCP	HTTP	94.0%	74.2%	64.7%	71.8%
	TCP	HTTPS	3.0%	8.3%	26.4%	0.9%
Email	TCP	IMAP4	0.1%	$\prec$	–	–
	TCP	SIMAP	1.0%	0.1%	–	0.03%
	TCP	POP3	0.01%	0.06%	–	$\prec$
	TCP	SPOP	0.4%	0.3%	–	0.01%
	TCP	SMTP	$\prec$	0.04%	–	$\prec$
Chat	TCP	IRC	$\prec$	$\prec$	–	–
Remote	TCP	FTP	$\prec$	$\prec$	–	$\prec$
	TCP	SSH	–	$\prec$	–	0.05%
Enterprise Services	TCP/UDP	NFS	$\prec$	$\prec$	–	$\prec$
	TCP	SMB	–	$\prec$	–	$\prec$
	TCP	IPP	$\prec$	0.01%	–	0.3%
	TCP	LPD	–	0.04%	–	–
	TCP	LDAP	$\prec$	$\prec$	–	$\prec$
	TCP	SQL	–	$\prec$	–	–
Management	TCP/UDP	DNS	0.2%	0.3%	1.5%	0.1%
	TCP/UDP	NetBIOS	$\prec$	0.03%	0.02%	0.02%
	UDP	NTP	$\prec$	$\prec$	–	$\prec$
	UDP	SNMP	$\prec$	0.01%	–	$\prec$
Other	TCP	Other	0.2%	2.9%	5.7%	8.7%
	UDP	Other	1.0%	13.7%	1.7%	18.1%

Table 5: Application protocol usage by percent of bytes ($\prec$ *less than 0.01%*, *– none*)

5.2 Application Protocols

We further categorize the mix of wireless traffic based on application protocol using Bro Intrusion Detection System [16]. Table 5 shows the percentage of traffic in bytes for each type of application protocol in the UW and traces. Web protocols (HTTP and HTTPS) account for the largest percentage of traffic for both handheld (97%) and non-handheld (82%) devices. However, web usage is lower for both types of devices in the CS network. Email protocols are the second most popular application but account for less than 2% of traffic for both device types. (We believe clients actually generate more email traffic than this and attribute the low percentages of email protocols to the common usage of web based email.) These protocol usage observations are consistent with other network measurement studies [10].

The majority of UDP traffic for non-handheld devices can not be identified by Bro's dynamic protocol detection. Manual categorization using port numbers and IP addresses reveals some of the traffic is from VPN, Symantec Systems Center, DropBox, and Microsoft Simple Service Discovery Protocol. However, more than 90% of the unidentified UDP traffic is large flows, ranging from 1 MB to 20 MB in size, whose application protocol remains unidentifiable. This traffic is likely from streaming media or peer-to-peer file sharing, as no traffic of this type is explicitly identified but we expect should be present.

Handhelds do not exchange any peer-to-peer (P2P) file sharing traffic, and non-handhelds only exchange this traffic in small amounts (potentially identified as TCP or UDP *other*), if at all. This low volume of P2P traffic differs from a 2003 study [10] which attributed over 35% of traffic to P2P applications. Low P2P usage is a good sign for network management and performance, because traffic of this type has been known to be difficult to control and prevent from overtaking network capacity. Further decreases in P2P traffic as handhelds become the most prevalent devices, translates to more bandwidth for streaming media and other bandwidth-intensive Internet services.

Based on the application protocols that are identified, we observe a larger percentage of non-handheld devices using internal enterprise services—filesystem, printing, database, etc. These services represent a small percentage of total application traffic (about 0.07% in total) but bring to the forefront an important observation: non-handheld users care more about internal services. This observation is reinforced by the presence of an internal website (`library.wisc.edu`) in the top HTTP hosts by request for non-handheld devices (Table 10). From a management perspective, administrators need to balance their focus on maintaining network services and network performance to satisfy the needs of non-handheld users. An increase in the number of handhelds requires administrators to place less emphasis on internal services and more emphasis on increasing network performance.

5.3 TCP Flow Characteristics

As mentioned earlier, TCP traffic accounts for 98% of handheld traffic. We compare the TCP flow characteristics of handheld and non-handheld traffic to determine *if* and *how* the network dynamics will change as the network client base becomes primarily handhelds. We look at the flow size, duration, and rate for the downlink half of TCP connections— data flowing from remote host to the wireless client—since the majority of data flows in this direction. Incomplete flows (flows which do not end with a *FIN* or *RESET*) are excluded from the analysis. In all cases, the distributions for the both the UW and CS networks are equivalent; we omit inclusion of the CS distributions for brevity.

A CDF of flow size in bytes is shown in Figure 1. Handheld devices tend to have smaller flows than non-handheld devices: the median handheld flow size is 50 KB, compared to a median non-handheld flow size of 100 KB. The middle 80% of handheld flows range in size from 10 KB to 1 MB, while the middle 80% of non-handheld flows range in size from 25 KB to 1 MB. At the lower tail, there are fewer small non-handheld flows than handheld flows. At the upper tail, maximum non-handheld flow size is larger (2 GB) than the maximum handheld flow size (630 MB). The distribution of flow sizes for WiFi-only handhelds is identical to the distribution of flow sizes for all handhelds.

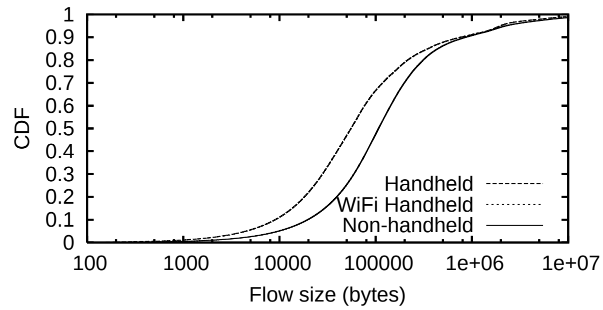

Figure 1: CDF of TCP flow size (UW)

Handhelds and non-handhelds also differ in their distribution of flow duration. A CDF of flow duration in seconds is shown in Figure 2. Handhelds have a narrower range of flow durations than non-handhelds. The middle 80% of handheld flows range in duration from 250 ms to 15 sec, compared to a range of 100 ms to 75 sec for non-handhelds. The median flow duration is approximately the same for both types of devices. The lack of long flows for handhelds can be attributed to the typically short usage sessions of handhelds reported by Falaki [7]. Again, the distribution of flow durations is similar for WiFi-only handhelds and all handhelds.

The flow durations for a subset of specific applications are shown in Table 6. On average, TCP flows for web traffic are five times shorter for handheld devices than non-handhelds. Handhelds are served simplified versions of many web pages, which we suspect is the cause of generally shorter TCP flows. For email traffic, receiving protocols (IMAP, POP) have shorter

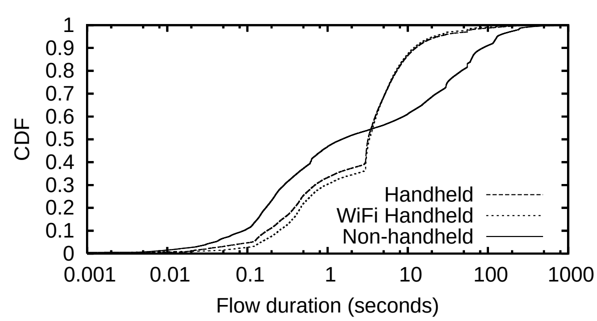

Figure 2: CDF of TCP flow duration (UW)

Category	Transport Protocol	Application Protocol	Handheld Avg	Handheld StdDev	Non-handheld Avg	Non-handheld StdDev
Web	TCP	HTTP	4.93	15.09	24.06	48.53
	TCP	HTTPS	2.42	14.62	11.60	38.14
Email	TCP	IMAP4	59.33	67.85	0.69	1.91
	TCP	SIMAP	36.30	64.23	5.64	24.03
	TCP	POP3	33.06	62.41	1.46	7.69
	TCP	SPOP	36.50	51.98	1.95	4.82
	TCP	SMTP	3.63	11.45	26.56	40.63
Other	TCP	Other	16.95	50.68	3.10	18.18
	UDP	Other	13.87	201.96	21.72	2106.98

Table 6: Application Connection Duration (UW)

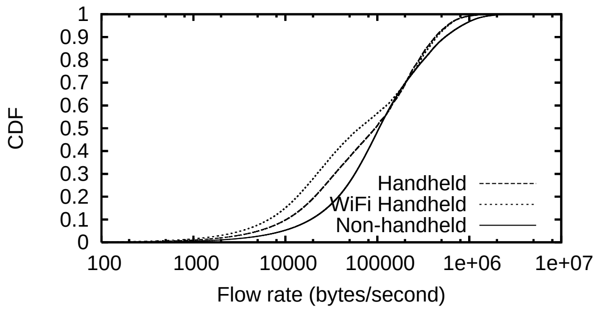

Figure 3: CDF of TCP flow rate (UW)

average TCP flows on non-handhelds, while the sending protocol (SMTP) has shorter flows on handhelds. We hypothesize the discrepancy in SMTP is caused by a higher likelihood of non-handheld users including attachments in emails, thus increasing the about of data that must be transferred.

The distribution of flow rate is similar for both types of devices. Figure 3 shows a CDF of flow rate in bytes/second. The middle 80% of flows have a rate ranging from 10 KBps to 500 KBps, with a median rate of 100 KBps. WiFi-only handhelds have slightly faster flow rates than all handhelds for the lower 60% of flows. The factors associated with flow rate are also consistent across handheld and non-handheld devices: The average round trip time for 90% of TCP flows is less than 100 ms. Only 4% of flows have one or more retransmissions due to retransmission time out, and 1% of flows have one or more retransmissions due to fast retransmit. Figure 4 shows a CDF of the average receive window. Handheld devices have a more linear distribution of window sizes than non-handheld devices, but the distribution is relatively similar. The similarity between the handheld and non-handheld flow rates implies both types of devices have similar TCP stacks.

Our analysis shows that the aggregate TCP flow characteristics of network traffic will remain relatively unchanged as the handheld client population increases. Handhelds have smaller flows and a narrower range of flow durations, but throughput will remain consistent. From a management perspective, no major network configuration changes are required to maintain similar performance.

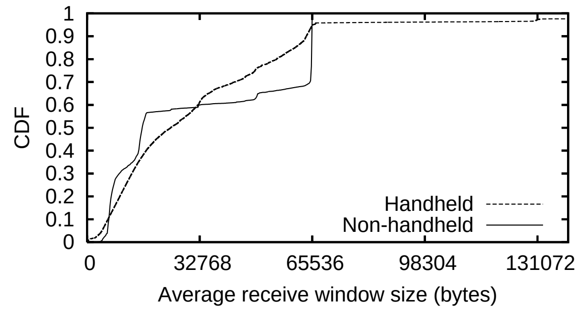

Figure 4: Average receive window (UW)

5.4 Management Implications

Based on our analysis of the protocols and services used by handheld devices, we make the following management recommendations for networks with primarily handheld clients:

- **Focus less on internal services** Handheld clients do not use most internal services like network file systems, printers, and authentication services. Most traffic is external HTTP traffic, so clients care more about network performance and the quality of Internet access.

- **No concerns with peer-to-peer traffic** The bane of many network administrators, peer-to-peer traffic is not present in handheld communications (and has also become only a small fraction of non-handheld traffic). Instead of shaping peer-to-peer traffic, administrators should focus on providing strong performance for the new source of high traffic volumes: streaming media.

6. WEB TRAFFIC

Web traffic accounts for almost all handheld data (97%) and a large fraction of non-handheld data (82%). HTTP is used so commonly because of its wide interoperability and ability to distribute all types of content. Web usage differs between handheld and non-handheld devices because of differences in the way individuals use these devices. We see differences in the range and type of hosts being accessed and variability in the type and length of content. We also observe that 82% of handheld HTTP traffic is consumed by non-browser applications, while only 10% of non-handheld traffic is destined for other applications. Most notably, we see a higher usage of HTTP-based streaming media services on handhelds, where video accounts for 42% of handheld HTTP content, compared to only 23% for non-handhelds.

Extracting information from HTTP traffic requires re-assembling packet payloads into streams. We built a custom tool using the PCAP and NIDS libraries to read traces, reassemble streams, and extract the necessary fields from HTTP headers. Libnids is part of Network Intrusion Detection System [1]; we take advantage of its TCP stream reassembly capabilities to reconstruct HTTP headers and payloads from

% of Bytes	Host	Top Content Types
35.48%	googlevideo.com	video/mp4
18.12%	pandora.com	application/octet-streaming, image/jpeg
10.57%	phobos.apple.com	text/plain, image/jpeg, video/m4v, audio/mp4
2.45%	fbcdn.net	image/jpeg, text/javascript, image/png
2.43%	vo.llnwd.net	video/m4v, video/mp4, audo/mpeg
1.23%	m.nbc.com	video/mp4, image/jpeg, text/javascript
1.17%	espn.go.com	text/plain, text/html, image/jpeg
1.16%	video.ted.com	video/mp4
0.82%	gdata.youtube.com	application/atom+xml
0.64%	s3.amazonaws.com	audio/3gpp, image/jpeg, image/png

Table 7: Top handheld HTTP hosts by response size

sets of TCP packets. Incomplete streams are not reassembled by NIDS so we miss the data from some HTTP streams, but this proportion is small. As a stream is reassembled, we extract the values of relevant HTTP fields (method, URI, host, content-type, content-length, etc.) and keep byte and packet statistics. Due to anonymity concerns, we only look at HTTP traffic from the UW traces.

6.1 Hosts

HTTP hosts provide a rough understanding of the types of services accessed by clients. We group hosts by subdomain to cope with websites which load-balance amongst multiple servers—i.e. we consider traffic for `f500.mail.yahoo.com` and `f504.mail.yahoo.com` to all be associated with the same host `mail.yahoo.com`. However, using subdomains still keeps key service information intact—i.e. we can differentiate between `yahoo.com` and `mail.yahoo.com`. Also, some web services provide special mobile versions of content to handheld devices, typically identified by a `m` or `mobile` subdomain. (In some cases we group by domain instead of submdomain—for example `fbcdn.net`—because load-balancing across lots of servers occurs at the domain level.)

Table 7 lists the top HTTP hosts for handheld devices based on the size (content-length) of the data served to the devices. Over 35% of handheld HTTP content originates from `googlevideo.com`, followed by 18% originating from `pandora.com`. For each host, we also list the most frequent content type (based on on the total content-lengths for each type). Multimedia content is the most frequent for eight of the top ten hosts.

The top hosts for non-handheld devices, based on content size, are listed in Table 8. The most popular host, `c.youtube.com`, accounts for 11% of the data, followed by `pandora.com` which accounts for 7% of the data. In total, the top 10 non-handheld hosts account for 42% of non-handheld data, while the top 10 handheld hosts account for 74% of handheld data. These percentages indicate a much greater diversity in hosts for non-handheld devices. In addition, non-handheld devices are more likely to receive content from hosts providing more than text or multimedia content. A Microsoft site hosting application downloads, `dlservice.microsoft.com`, appears in the top non-

% Bytes	Host	Top Content Types
11.45%	c.youtube.com	video/flv, video/mp4
7.00%	pandora.com	application/octet-stream, image/jpeg, audio/mpeg
6.63%	fbcdn.net	image/jpeg, image/png, text/javascript
4.63%	dlservice.microsoft.com	application/octet-stream
2.89%	vo.llnwd.net	video/wmv, audio/mp4
2.80%	stileproject.com	application/octet-stream, image/jpeg, video/mp4
2.53%	com.edgesuite.net	video/wmv, audio/wma, application/octet-stream
1.69%	phobos.apple.com	text/plain, audio/mp4, image/png
1.51%	www.facebook.com	text/html, text/javascript
0.94%	cdn.turner.com	text/javascript, image/jpeg, video/flv

Table 8: Top non-handheld HTTP hosts by response size

% of Requests	Host	Top Content Types
10.58%	fbcdn.net	image/jpeg, text.javascript, image/png
4.26%	phobos.apple.com	image/png, image/jpeg, text/plain
3.10%	espn.go.com	text/html, image/gif, image/jpeg, image/png
2.75%	facebook.com	text/plain, text/html
2.74%	googlevideo.com	video/mp4
2.56%	www.apple.com	text/html
1.30%	ad.doubleclick.net	text/javascript, text/html, video/ms-asf, image/gif
1.26%	i.ytimg.com	image/jpeg
1.23%	www.google.com	text/html, text/javascript
1.18%	itunes.apple.com	text/xml

Table 9: Top handheld HTTP hosts by request volume

handheld hosts with *application/octet-stream* as the primary content type.

We also look at the top hosts based on number of HTTP requests. Table 9 lists the top hosts for handheld devices and Table 10 for non-handheld devices. The top 10 handheld hosts account for 30% of handheld HTTP requests and non-handheld hosts account for 32% of non-handheld requests. There is a greater diversity of content in the top hosts by number of requests: social networking, streaming media, advertising, search, and news. Only one streaming media host is in the top hosts for handheld devices and none are in the top hosts for non-handheld devices. Also, the non-handheld top hosts includes an internal UW website, `library.wisc.edu`, re-enforcing our earlier observation (Section 5.2) that non-handhelds use internal services more than handhelds.

6.2 Content Type and Length

The type of HTTP content access by devices further characterizes the services used. Table 11 lists the top HTTP con-

% of Requests	Host	Top Content Types
15.36%	fbcdn.net	image/jpeg, image/png, text/javascript
5.70%	www.facebook.com	text/plain, text/javascript, text/html
2.92%	www.google.com	text/html, text/javascript, application/json
1.84%	ad.doubleclick.net	text/javascript, text/html, image/gif
1.24%	cdn.turner.com	image/jpeg, image/gif, text/javascript
1.22%	library.wisc.edu	text/html, image/gif, text/javascript
1.11%	g.doubleclick.net	text/html, text/javascript, image/gif
1.08%	graphics8.nytimes.com	image/jpeg, image/gif, text/javascript
1.05%	www.google-analytics.com	image/gif, text/javscript
0.90%	espn.go.com	image/gif, text/html, image/jpeg

Table 10: Top non-handheld HTTP hosts by request volume

Protocol	Handheld	Non-handheld
Application	*19.73%*	*22.1%*
atom+xml, rss	0.97%	0.26%
binary	0.35%	0.02%
compress, gzip, tar, zip	0.64%	0.74%
json	0.23%	0.18%
octet-stream	16.93%	11.41%
pdf	0.17%	0.74%
shockwave-flash	0%	6.71%
Audio	*3.43%*	*5.59%*
3gpp	0.54%	0%
mp4	0.23%	0.80%
mpeg	2.56%	4.45%
wma	0.05%	0.29%
Image	*17.41%*	*29.39%*
gif	1.64%	2.85%
jpeg	12.81%	22.75%
png	2.95%	3.69%
Text	*16.85%*	*18.92%*
css	0.65%	1.36%
html	2.39%	7.38%
javascript	2.81%	6.47%
plain	10.11%	2.86%
xml	0.87%	0.83%
Video	*42.55%*	*23.35%*
3gpp	0.79%	0%
flv	0.01%	18.62%
mp4	41.40%	2.30%
quicktime	0.28%	0.41%
wmv	0%	1.99%

Table 11: HTTP content-types by content-length (UW)

tent types, based on content length, for the UW traces. We group the types based on the top-level category—application, audio, image, text, and video—and list the top types for each category. The largest volume of handheld content is video (42%), while images are the top content type for non-handheld devices (29%).

Application HTTP content is data associated with specific applications, for example documents, compressed files, or streaming media control information. For both types of devices, *octet-stream*—a generic binary stream of data which an application can interpret as desired—is the most common, accounting for 86% of handheld and 51% of non-handheld application type data. Some streaming media sites, for example Pandora, use the *octet-stream* content type. The average *octet-stream* is 713 KB for handhelds (std dev 882 KB) and 189 KB for non-handhelds (std dev 658 KB) The second most common for handhelds is RSS feeds (average length of 29 KB), while Shockwave Flash is the second most common for non-handhelds (average length of 38 KB). No handhelds access Shockwave Flash content because there was no Flash support on these devices until very recently.

The diversity of application content types is greater on non-handheld devices than handhelds. Over 185 different types of application content are accessed by non-handhelds compared to only 58 different types for handhelds. This variety in types results from the greater diversity of applications running on non-handheld devices.

Stemming from this observation, we envision in-network security scanners as a viable service to deploy in networks with large proportions of handhelds. The energy and performance cost of running a malware scanner on a handheld is prohibitive. In contrast, a middlebox could scan a large majority of handheld traffic for malware with relatively few HTTP content signatures.

The content for regular web browsing falls mostly into the image and text content types. Three image types—gif, jpeg, and png—make up the majority of image content. JPEG images are the largest with an average length of 13 KB on handhelds and 11 KB on non-handhelds. HTML, CSS, JavaScript and XML are used for the web page itself. For both types of devices these text types average 3-7 KB in length. Over two-thirds of the text content received by handheld devices is identified as *plain* text. This content is access by non-browser-based applications retrieving data from a web service, for example a news or sports application.

The remaining two top-level categories of content are multimedia traffic, namely audio and video. Multimedia accounts for 46% of handheld content and 29% of non-handheld content. In particular, video accounts for 93% of multimedia traffic in the handheld case and 80% in the non-handheld case. We examine video traffic in greater detail next.

6.3 Streaming Video

Streaming video is a major source of traffic for handheld devices. Video content accounts for 40% of all handheld traffic, compared to only 17% of all non-handheld traffic. Large volumes of multimedia streams bring forth interesting issues because of their size and sensitivity to delay.

6.3.1 Video Flows

We compare the flow characteristics of handheld streaming video flows to all handheld flows to better understand streaming media's impact on the network. Figure 5 shows the flow size in bytes for handheld TCP flows (excluding incomplete flows). As expected, handheld video flows are large compared to overall handheld traffic: Eighty percent of video flows are greater than 50 KB in size, whereas 50KB is the median flow size among all handheld flows. Nearly 20% of video flows are larger than 1 MB in size. The median flow size is 400KB and the average flow size is 305 KB for handheld devices. Also, we note that video flows in non-handhelds are even larger: The median flow size is nearly an order or magnitude higher (3MB).

The flow duration for handheld TCP flows is shown in Figure 6. Video flows for handhelds appear to be of a short duration. Eighty percent of video flows are less than 1 second in duration, with a median video flow duration of 0.5 seconds. The median durations for all handheld flows and for non-handheld video flows are significantly higher, at 5 and 50s, respectively. We expect the short flow durations result from handhelds' goal of conserving energy by receiving data over wireless in short periods of time.

Based on the short duration of video flows, we expect high throughput rates for handheld video flows. Figure 7 shows the rate of handheld TCP flows. Eighty percent of video

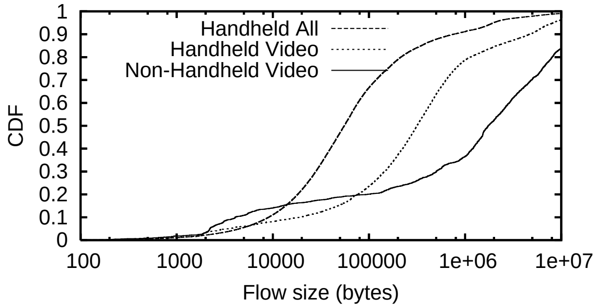

Figure 5: Flow size in bytes

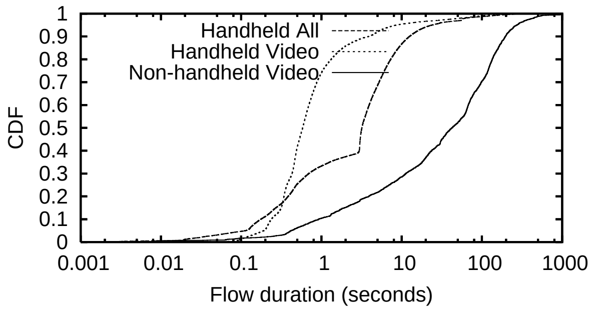

Figure 6: Flow duration in seconds

flows have a rate faster than 100KBps (0.8 Mbps), with a median flow rate of 250 KBps (2 Mbps). In contrast, the media flow rate for all handheld flows and for non-handheld video flows is roughly 75 KBps (0.6Mbps).

On the whole, handheld video flows are long in size (although not as long as non-handheld video flows), significantly short in duration, and achieve high end-to-end throughputs which are comparable, if not slightly higher than non-handheld video flows.

6.3.2 Video Format

Video content streamed to handheld devices differs from video streamed to non-handheld devices because of differences in decoding capabilities. Most streaming video services use Adobe Flash, but Flash support did not exist on

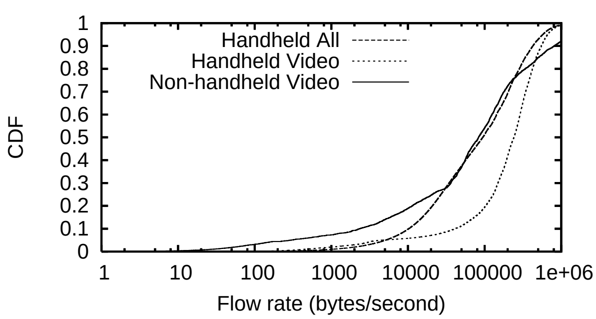

Figure 7: Flow rate in bytes/second

handheld devices until very recently [3]. Instead, handheld devices receive video content that is encoded using MPEG 4. Table 11 shows that *mp4* (MPEG 4) is the top video type for handhelds and *flv* (Flash video) is the most common for non-handhelds. Video streaming sites like YouTube serve two versions of videos: one encoded as mp4 and the other encoded as flv.

To better understand the differences in the video content served to the two different types of devices, we watch the same 3 minute video [2] from YouTube on both a Google Android HTC Dream smartphone and a Lenovo X201 laptop. On the phone, we use the standalone YouTube application and on the laptop we use Mozilla Firefox. The handheld device receives 7362 KB `video/mp4`; the non-handheld device receives 11792 KB `video/flv`. Both videos have the same resolution of 320 x 240, but the mp4 version is encoded at 30 fps (frames per second) and 200 kbps, while the flash version is encoded at 25 fps and 231 kbps. The flash video is of slightly higher quality, with more bits per second, but both versions are comparable. The audio is encoded in stereo at 44100 Hz and 128 kbps for the mp4 version and mono at 22050 Hz and 64 kbps for the flv version. The mp4 audio is higher quality than the Flash audio, but both versions are closely comparable. We conclude the video served to handheld devices is of approximately equal quality, but the content is smaller in size—the handheld version is about 62% of the size of the non-handheld version.

The size of both handheld and non-handheld video flows are relatively small compared to the size of the sample YouTube video. The average handheld video flow size is 305 KB, much less than the 7.4 MB size of the 3 minute video. This size gap implies individuals watch only a small fraction (e.g., the initial few seconds) of most videos on their handhelds.

6.4 Management Implications

Our analysis of HTTP traffic yields the following management implications for wireless networks with a primarily handheld client base:

- **Deploy in-network malware scanners** The majority of handheld traffic (93%) is HTTP and consists of less than 100 different content types. Deploying a malware scanner as a middlebox requires knowledge of a limited number of content types, yet can provide relatively high coverage of handheld data. It also avoids the energy and performance penalties of running an anti-malware application directly on handhelds.

- **Use traffic shaping for suitable streaming video performance** Almost half of all handheld traffic is streaming video. This content is sensitive to delay and requires sufficient bandwidth to transfer data in a reasonable period of time. Quality of service and other forms of resource reservation are not scalable in networks where a large percentage of traffic is from multimedia streams. Instead, administrators need to use traffic shaping or admission control to ensure sufficient resources are available to meet the streaming media demands of handheld devices. However, designing appropriate traffic shapers and admission controllers may not be easy because of the ephemeral nature (i.e., short duration) of the video flows.

7. CONTENT SIMILARITY

In this section, we examine the similarity in the content perused by mobile smartphone users and compare it against non-handheld users. We evaluate a “chunk-based” content similarity system, akin to DOT [20], SET [17], EndRE [5] and LBFS [15]: we first divide the payloads of packets exchanged by users into chunks using value sampling [20]; the sizes of the chunks range between 32B and 64B. We then identify if the chunks have appeared in an earlier accesses. Unless otherwise specified, we assume that a total of 2GB worth of chunks are stored across all users, as done in prior systems [5].

This analysis helps us estimate the benefit of employing content similarity suppression schemes which eliminate duplicate chunks from network transfers by serving them from a local cache [8, 5, 20, 17]. Such schemes help improve both the end-to-end latency as well as transfer throughput experienced by users, and they can also help save mobile battery life by conserving network transmissions. Chunk-based schemes in particular are more effective than object caching schemes such as Web caches as they are known to identify more duplicates, e.g., sub-object duplicates, uncacheable content etc. Thus, our analysis places an upper bound on the effectiveness of using caching and similarity suppression.

In performing this study, we identify two types of similarity: that found in content accessed by the same user, which we refer to as “intra-user” similarity, and that found in content accessed by a different user in the network, which we refer to as “inter-user” similarity. The former can be exploited much more easily, by deploying caches close to the users, such as per-user or per-AP caches. In order to exploit the latter, the network admin may have to employ schemes for issuing cross-cache queries across per-user or per-AP caches, or maintaining a single larger cache that aggregates content accessed by all users. Furthermore, if intra-user redundancy dominates then partial deployment of per-user or per-AP caches will result in partial benefits.

In Figure 8, we show the extent of intra- and inter-user content similarity observed over every 1 million packets worth of handheld and non-handheld traffic. The traces are 0.8-2GB in size. We measure content similarity as the ratio of similar bytes to all bytes in the 1 million-packet trace subset. The graph plots a CDF of average redundancy of different types observed per trace.

The graph shows that content similarity varies across different traces (and, thus, over time) and across similarity types. First, we observe a greater amount of similarity in handheld traces than in non-handheld traces. Second, similarity due to inter-user matches is quite small: less than 2% for more than 95% of both handheld and non-handheld trace subsets. Third, we observe that in more than 40% of the non-handheld trace subsets, and more than 70% of the handheld device trace subsets, $\geq$ 8% of the similar bytes are due to intra-user matches. In some cases, we observed upto 20-25% intra-user similarity for both kind of traces. Finally, the extent of intra-user similarity is greater in the case of handhelds than in the case of non-handhelds.

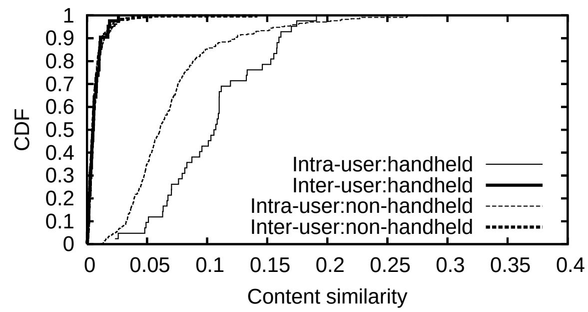

Figure 8: Average intrauser and interuser redundancy across multiple traces corresponding to handheld devices

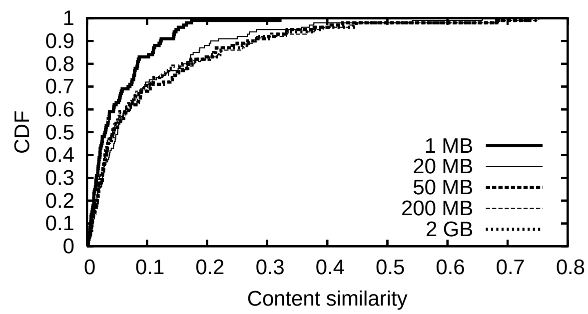

Figure 9: CDF of intrauser redundancy across top 100 (by volume) handheld device traces for different dictionary sizes

Given that the dominant fraction of similar bytes belonged to intra-user traffic, we further delve intro intra-user similarity. In particular, we explore the efficacy of deploying per device caches and the cache size configuration issues therein. We split the handheld traffic on a per device basis and study the effect of different dictionary sizes on amount of similarity identified per device.

Figure 9 shows the CDF of similarity across top 100 devices by traffic volume for different dictionary sizes. First, we observe that almost 80% of users have less than 20% similarity with their own traffic. However, for certain users, the similarity proportion was much higher (more than 50%). Second, we observe that most of the similarities can be identified by using only 50 MB cache; Larger cache sizes exhibit diminishing returns in the amount of similarity they can identify. These two observations imply that partial deployment of small per-device caches can result in significant benefits.

Since our chunk-based analysis upper bounds the similarity that can be identified by object caching approaches, it is safe to conclude from our analysis that even if object level caches were employed:

- It is better to employ per device, or per-AP, caches as most matches are intra-user, and per-device caches care simpler and can help provide partial deployment benefits.
- Small caches, of roughly 50MB per device, suffice in practice.

8. RELATED WORK

Our work complements and extends prior studies of campus wireless networks. Some of these focussed on PDA and smartphone usage. We discuss these studies next.

Multiple measurement studies have analyzed traffic patterns in campus wireless network. Hederson et. al identify session and application trends at Dartmouth College and observe how usage evolved four years after the network's initial deployment [10]. Wireless AP workloads at Darthmouth are compared to the University of North Carolina at Chapel Hill by Hernandez-Campos and Papadopouli [11]. Lastly, McNett and Voelker study the wireless access and mobility patterns of students using PDAs at the University of California, San Diego [14]. While all of these studies focus on campus wireless networks none explore in detail the applications used specifically by mobile device users and the traffic characteristics thereof. In addition, mobile device usage is a rapidly changing field and trends observed five years ago are different than today's mobile device usage.

More recent studies have focused on mobile device usage in public Wi-Fi and 3G networks. Application, session, and mobility trends in the Google Wi-Fi network in Mountain View, CA were studied in 2008 [4]. The connections between geo-location and usage of specific types of web services was studied in an urban 3G network in 2009 [21]. The 3G study is most similar to our work, but only considers HTTP traffic. While we observe this is a large portion of mobile device traffic, it leaves out the other applications mobile users utilize.

Application usage on mobile devices has also been studied outside of the context of wireless networks. Usage logs from 255 smartphone users were analyzed by Falaki et. al [7]. Interestingly, communication applications account for 49% of usage and browsing accounts for 12% of usage, resulting in at least 60% of application usage generating network traffic. Mobile device application usage has implications for human computer interaction resulting in multiple log- or diary-based studies [18, 9].

9. CONCLUSION

Handheld devices have become a significant fraction of the client base in campus wireless networks, and their usage is expected to continuing growing. Using traces from two wireless networks at the University of Wisconsin-Madison, we identify differences in the traffic characteristics of handheld and non-handheld devices which have implications for network management. We observe that handheld devices make less use of internal services and are more concerned with the performance and quality of Internet access. Over 98% of handheld traffic is web traffic, and handheld devices do not send any peer-to-peer traffic, avoiding the bane of many network administrators. The HTTP communications of handhelds are spread across a smaller number of hosts and content types than non-handhelds, paving the way for easy implementation of in-network security scanners. Forty percent of all handheld traffic is HTTP-based streaming media, prompting administrators to ensure adequate bandwidth is available by shaping other network traffic. Lastly, administrators may consider employing "chunk-based" content distribution systems to improve caching and decrease handheld network transmissions. As handheld usage continues to increase, administrators will need to change the way they manage wireless networks and deploy new services to meet the unique demands of these devices.

10. REFERENCES

[1] Libnids. `http://libnids.sourceforge.net`.

[2] Christmas lights gone wild. `http://www.youtube.com/watch?v=rmgf60CI_ks`, November 2005.

[3] Adobe announces availability of flash player 10.1 for mobile. `http://www.adobe.com/aboutadobe/pressroom/pressreleases/201006/06222010FlashPlayerAvailability.html`, June 2010.

[4] M. Afanasyev, T. Chen, G. M. Voelker, and A. C. Snoeren. Analysis of a mixed-use urban wifi network: when metropolitan becomes neapolitan. In *IMC '08: Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, pages 85–98, New York, NY, USA, 2008. ACM.

[5] B. Aggarwal, A. Akella, A. Anand, P. Chitnis, C. Muthukrishnan, A. Nair, R. Ramjee, and G. Varghese. EndRE: An End-System Redundancy Elimination Service for Enterprises. In *NSDI*, 2010.

[6] F. Dogar and P. Steenkiste. Catnap: Exploiting High Bandwidth Wireless Interfaces to Save Energy for Mobile Devices. In *MobiSys '10: Proceedings of the 8th international conference on Mobile systems, applications, and services*, 2010.

[7] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin. Diversity in Smartphone Usage. In *MobiSys '10: Proceedings of the 8th international conference on*

Mobile systems, applications, and services, 2010.

[8] L. Fan, P. Cao, J. Almeida, and A. Z. Broder. Summary cache: a scalable wide-area web cache sharing protocol. In *SIGCOMM '98*, 1998.

[9] J. Froehlich, M. Y. Chen, S. Consolvo, B. Harrison, and J. A. Landay. Myexperience: a system for in situ tracing and capturing of user feedback on mobile phones. In *MobiSys '07: Proceedings of the 5th international conference on Mobile systems, applications and services*, pages 57–70, New York, NY, USA, 2007. ACM.

[10] T. Henderson, D. Kotz, and I. Abyzov. The changing usage of a mature campus-wide wireless network. *Comput. Netw.*, 52(14):2690–2712, 2008.

[11] F. Hernandez-Campos and M. Papadopouli. A comparative measurement study the workload of wireless access points in campus networks. In *Personal, Indoor and Mobile Radio Communications, 2005. PIMRC 2005. IEEE 16th International Symposium on*, volume 3, pages 1776 –1780 Vol. 3, sept. 2005.

[12] J. Horrigan. Wireless internet use. *Washington, DC: Pew Internet & American Life Project*, pages 2009–07, 2009.

[13] J. Korhonen and Y. Wang. Power-efficient streaming for mobile terminals. In *NOSSDAV '05: Proceedings of the international workshop on Network and operating systems support for digital audio and video*, pages 39–44, New York, NY, USA, 2005. ACM.

[14] M. McNett and G. M. Voelker. Access and mobility of wireless pda users. *SIGMOBILE Mob. Comput. Commun. Rev.*, 9(2):40–55, 2005.

[15] A. Muthitacharoen, B. Chen, and D. Mazières. A low-bandwidth network file system. *SIGOPS Oper. Syst. Rev.*, 35(5), 2001.

[16] V. Paxson. Bro: a system for detecting network intruders in real-time. In *SSYM'98: Proceedings of the 7th conference on USENIX Security Symposium*, pages 3–3, Berkeley, CA, USA, 1998. USENIX Association.

[17] H. Pucha, D. G. Andersen, and M. Kaminsky. Exploiting similarity for multi-source downloads using file handprints. In *NSDI*, 2007.

[18] A. Rahmati, A. Qian, and L. Zhong. Understanding human-battery interaction on mobile phones. In *MobileHCI '07: Proceedings of the 9th international conference on Human computer interaction with mobile devices and services*, pages 265–272, New York, NY, USA, 2007. ACM.

[19] S. Smith, G. Salaway, and J. Caruso. The ECAR Study of Undergraduate Students and Information Technology, 2009. *EDUCAUSE Center for Applied Research*, 2009.

[20] N. Tolia, M. Kaminsky, D. G. Andersen, and S. Patil. An architecture for Internet data transfer. In *NSDI*, 2006.

[21] I. Trestian, S. Ranjan, A. Kuzmanovic, and A. Nucci. Measuring serendipity: connecting people, locations and interests in a mobile 3g network. In *IMC '09: Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 267–279, New York, NY, USA, 2009. ACM.

[22] ZyTrax. Mobile browser id (user-agent) strings.