# A First Look at Mobile Hand-held Device Traffic

Gregor Maier, Fabian Schneider, and Anja Feldmann

TU Berlin / Deutsche Telekom Laboratories
Ernst-Reuter-Platz 7, 10589 Berlin, Germany
{gregor|fabian|anja}@net.t-labs.tu-berlin.de

**Abstract.** Although mobile hand-held devices (MHDs) are ubiquitous today, little is know about how they are used—especially at home. In this paper, we cast a first look on mobile hand-held device usage from a network perspective. We base our study on anonymized packet level data representing more than 20,000 residential DSL customers. Our characterization of the traffic shows that MHDs are active on up to 3 % of the monitored DSL lines. Mobile devices from Apple (i. e., iPhones and iPods) are, by a huge margin, the most commonly used MHDs and account for most of the traffic. We find that MHD traffic is dominated by multimedia content and downloads of mobile applications.

**Key words:** Mobile Devices, iPhone, Traffic Characterization

## 1 Introduction

Today advanced mobile hand-held devices (MHDs, e. g., iPhones and BlackBerrys) are very popular. MHDs have evolved rapidly over the years—from pure offline devices, to cell phones with GSM data connectivity, to 3G devices, and universal devices with both cellular as well as WiFi capabilities. Their increased graphics and processing power makes these devices all-in-one PDAs and media centers. Today's MHDs can be used to surf the Web, check email, access weather forecast and stock quotes, and navigate using GPS based maps—to just name some of the prominent features. This increase in flexibility has caused an increase in network traffic. Indeed, cellular IP traffic volume is growing rapidly and significantly faster than classic broadband volume [15].

We, in this paper, cast a first look at Internet traffic caused by mobile hand-held devices. We use anonymized residential DSL broadband traces, spanning a period of 11 month, to study MHD behavior and their impact on network usage. We are thus able to observe the behavior of MHDs when they are connected via WiFi at home and compare their traffic patterns to the overall residential traffic characteristics. Some devices (most notably iPod touch and iPhone) require WiFi connectivity rather than cellular connectivity for some services. Other services are more likely to be used via cellular connectivity due to user mobility, e. g., looking up directions on Google Maps, while walking around town or driving. Although, we in this paper only focus on residential MHD usage and not MHD usage in cellular networks, our analysis gives first insights into what kind of services users are interested in when they are at home and have access to all services. This information is crucial for 3G cellular providers to anticipate usage patterns and future traffic growths.

**Table 1.** Overview of anonymized packet traces.

| Name | Start date | Duration | Size | # MHDs | MHD HTTP Traffic Volume | MHD HTTP Traffic % of HTTP |
|---|---|---|---|---|---|---|
| SEP08 | Thu 18 Sep'08 4am | 24 h | >4 TB | >200 | >2 GB | 0.1 % |
| APR09 | Wed 01 Apr'09 2am | 24 h | >4 TB | >400 | >9 GB | 0.4 % |
| AUG09a | Fri 21 Aug'09 2am | 24 h | >6 TB | >500 | >15 GB | 0.6 % |
| AUG09b | Sat 22 Aug'09 2am | 24 h | >5 TB | >500 | >15 GB | 0.7 % |

The remainder of this paper is structured as follows. In Sec. 2 we present our data sets and methodology, Sec. 3 presents our results. In Sec. 4 we discuss related work before we conclude our paper in Sec. 5.

## 2 Data and Methodology

In this section we describe the anonymized data sets of residential DSL connections and our methodology for analyzing them.

### 2.1 Data Sets

We base our study on multiple sets of anonymized packet-level observations of residential DSL connections collected at aggregation points within a large European ISP. The monitor, using Endace monitoring cards, operates at the broadband access router connecting customers to the ISP's backbone. Our vantage point allows us to observe more than 20,000 DSL lines. The anonymized packet-level traces are annotated with the anonymized DSL line card port id. This enables us to uniquely distinguish DSL lines since IP addresses are subject to churn and as such cannot be used to identify DSL lines [7]. While we typically do not experience any packet loss, there are several multi-second periods (less than 5 minutes overall per trace) with no packets due to OS/file-system interactions.

We use several 24 h traces collected over a period of 11 months which gives us the the opportunity to track changes in mobile device usage over time. Table 1 summarizes characteristics of the traces, including their start, duration, size, and number of observed MHDs. We note that while the number of observed DSL lines remains about the same in each trace, the number of observed MHDs has increased significantly.

The data anonymization, classification, as well as application protocol specific header extraction is performed immediately on the secured measurement infrastructure using the Bro NIDS with dynamic protocol detection [3].

### 2.2 Identifying MHDs

To understand how MHDs are utilized we need to identify not only their presence in our traces but also their contributions. This is non-trivial as MHD users commonly do not just operate the MHD over their DSL-line but also/mainly computers or set-top boxes. Note, that all devices active via one DSL-line usually share a single IP address. Therefore, we rely on network signatures which we gather by observing and recording MHD behavior in a controlled environment.

Among the currently popular MHD devices are Symbian based phones, BlackBerrys, iPhones and iPods, Windows Mobile based phones, and Google Android phones [12]. We collected manual traces using tcpdump for all device types except BlackBerrys[1]. With each device we performed the following set of actions using a wireless access-point for data collection: connecting to the access-point, accessing several Web sites, watching videos on YouTube, using other mobile applications like Weather and Stocks, checking and sending emails, using Facebook, and updating/installing mobile applications on the MHD.

Analyzing these manual traces reveals that HTTP dominates the protocol mix and that most mobile applications, including Weather, Stock quotes, AppStore, and YouTube, use HTTP. From our manual traces we extract a list of HTTP user-agent strings for each device and OS combination.[2] We further augment this list by well-known strings from other mobile devices, e.g., BlackBerrys. This captures the strings of the standard applications. However, it is not possible to compile a list of all user-agent strings that MHD application writers may use. However, since most rely on standard libraries, we can add patterns for these. For example, most applications for Apple devices use the Apple CFNetwork library for communication and CFNetwork usually adds its name and version number to the end of user-agent strings. While Mac OS X also uses CFNetwork, the version numbers used by the iPhone and Mac OS X are disjoint and we can distinguish them. Based on this collection of user-agent strings we create patterns for *(i)* identifying DSL lines that “host” MHDs and *(ii)* identifying and classifying MHD usage of HTTP.

### 2.3 Application Protocol Mix

Finding signatures for identifying non-HTTP traffic caused by MHDs is more difficult since most other application protocols, e. g., POP, do not add device related information to their user-agent strings. Furthermore, they may use encryption.

One obvious approach for overcoming this limitation is to assume that MHDs and regular computers are used consecutively, i. e., not used at the same time at the same DSL line. Based upon this assumption one can classify all traffic after a HTTP request from a MHD on a DSL line as MHD traffic (relying on a timeout). However, we show in Sec. 3.1 that the underlying assumption is incorrect. A majority of the lines shows contemporaneous activity from MHDs and regular computers.

Therefore, we take advantage of another characteristic of network devices—their IP TTLs. The default IP TTLs of popular MHDs differ from those of the most commonly used home OSs. The default TTL of iPhones/iPods and Macs is 64, Symbian uses 69, while Windows uses 128. This enables us to separate MHD usage from regular PC

[1] Manual trace collection was performed with Google’s G1 (Android 1.5), Apple’s iPod touch (iPhone OS 2 & iPhone OS 3), HP’s iPaq (Windows Mobile), HTC Touch 3G (Windows Mobile), Nokia 810 (Maemo Linux), and Nokia E61 (Symbian). Thanks to all device owners.

[2] We note that these MHD user-agent strings differ from user-agent strings used by PCs/Macs.

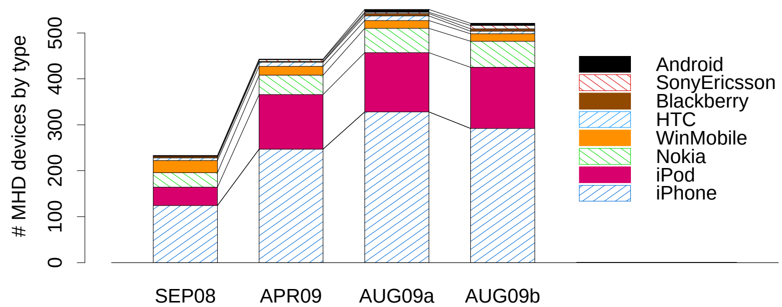



**Fig. 1.** Popularity of MHD device types

usage for some combinations of OSs. While we cannot distinguish iPhones/iPods from Macs or Windows Mobile from Windows we can use IP TTLs to separate the other combinations. Our observations show that the majority of home OSs is Windows while the majority of MHDs are iPods or iPhones. In order separate those, we first select all DSL lines for which *every* HTTP request with a TTL[3] of 64 or 69 is originated by a MHD (as identified via the user-agent). The assumption is that all traffic on these lines with TTL 64/69 is then caused by a MHD. Thus, we can then use Bro's DPD [3] on this traffic to get a first impression of the application protocol mix of MHDs. Since this approach excludes lines with certain combinations of MHDs and regular computers we are left with 54–59 % of the lines with MHDs. In addition, if the activity of the regular computer does not include HTTP we might misclassify its traffic. We note that we use this heuristic only for analyzing the application protocol mix, we use user-agent strings for all other analyses.

## 3 Results

After reporting on the pervasiveness of MHDs we focus on their protocol mix. Then we characterize MHDs' HTTP traffic, analyze mobile application usage, and present results on iTunes and AppStore usage.

### 3.1 MHD Pervasiveness

On a significant number of the DSL lines we observe traffic from MHDs (see Table 1). Indeed, in the most recent trace, AUG09, 3 % of active lines have MHD activity. Moreover, the contribution of MHDs to the observed HTTP traffic is also substantial (up to 0.7 % of HTTP bytes). This indicates that some MHD users may find it more convenient to use their mobile devices at home even if they have a regular computer as well. Note, HTTP's share of overall traffic volume is 50–60 % [4, 7].

There is a strong temporal trend underlined by the rapid growth in the number of lines with MHDs' activity and in the MHDs' HTTP traffic volume. The number of lines

[3] We take NAT devices and our hop distance to the end system into account.

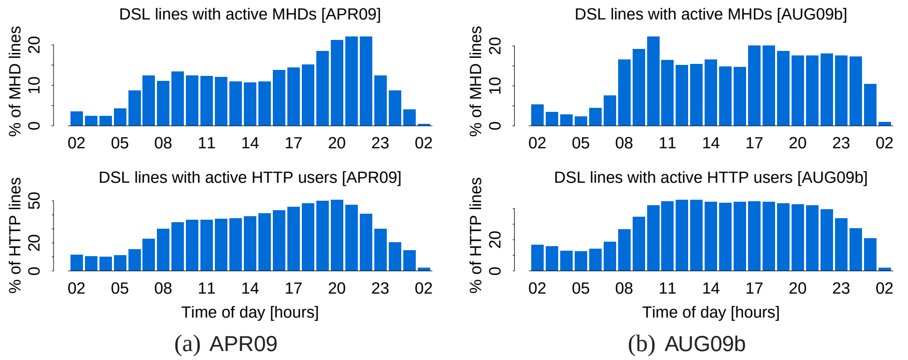



**Fig. 2.** Number of lines with MHD activity (top) vs. Number of lines with HTTP activity (bottom)

with MHDs almost doubled between SEP08 and AUG09. The HTTP traffic volume from MHD grew sixfold while the overall traffic volume increased only slightly and the overall HTTP volume increased by 22 % at our vantage point.

Fig. 1 shows the distribution of active devices types for all traces. We observe that Apple devices (iPhone and iPod touch) clearly dominate, both in terms of number of lines and traffic volume (not shown). They account for 86–97 % of MHDs' HTTP traffic and 71–87 % of the devices. This is in contrast to the market shares of the devices [12]. Possible explanations are that Apple users *(i)* find their device very convenient even for home use and/or *(ii)* are looking for a multimedia device that "also works as a phone". Indeed, the iPod Touch is an iPhone without phone capability. We note that starting from APR09 the number of lines with iPods outnumber the number of lines with all non-Apple MHDs combined.

We already pointed out that we have a substantial number of DSL lines "hosting" MHDs. Now we want to illustrate how the use of MHDs is distributed over the course of a day. To determine how the use of MHDs is distributed across time we plot the relative number of lines with active MHDs per hour (top) and the percentage of lines with HTTP traffic per hour for APR09 and AUG09b in Fig. 2. We see that MHDs are used throughout the day. While we see a similar behavior when looking at overall HTTP traffic, we see that MHD usage has a stronger pick-up in the morning (AUG09b even shows a peak). Overall HTTP traffic on the other hand slowly ramps up during the day. Again the convenience of using the mobile device may be a possible explanation. Users can use them to check their emails or the weather when "starting their day". The low byte contribution of mobile devices in the morning hours supports this claim (figure not shown).

Next, we examine if MHDs and regular computers are used consecutively or whether they are used contemporaneously. To asses this, we compute for each DSL line and for any two subsequent HTTP requests their inter-request-times (IRTs) and label them as *(i)* both from MHDs, *(ii)* both from non-MHDs, or *(iii)* from MHD and non-MHD. Using this information and timeouts of one second, one minute, and five minutes we compute the number of DSL lines with mixed activity (MHD and non-MHD). We find

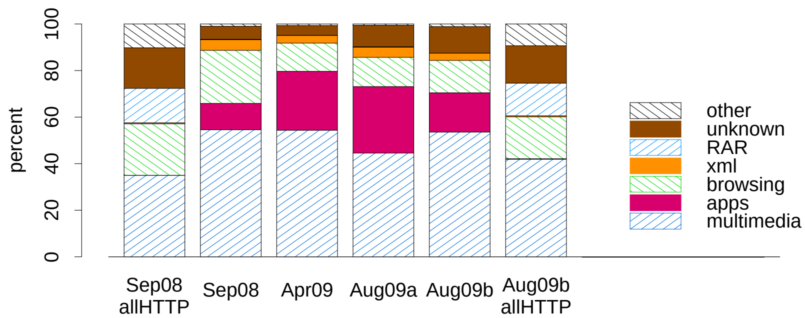



**Fig. 3.** HTTP content type categories by volume. Comparing MHD traffic all HTTP traffic.

that 33–39 % of MHD lines exhibit mixed MHD/non-MHD activity with IRTs of less than one second. For IRTs of less than one minute (five minutes) up to 62 % (72 %) of the lines have mixed activity.

### 3.2 Application Protocol Mix

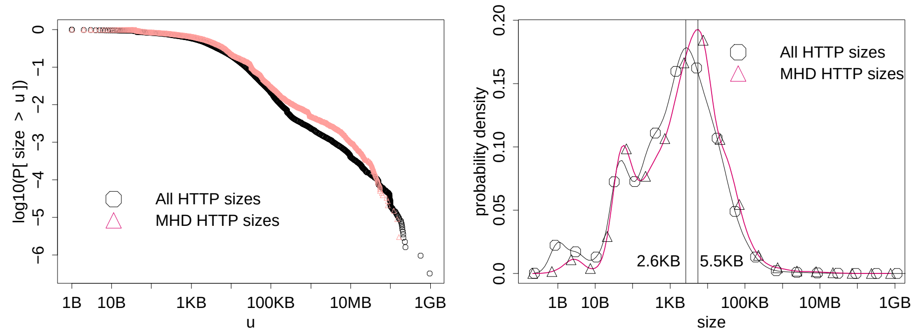
While our approach for analyzing the application protocol mix of MHDs is limited (see Sec. 2.3), it still gives us a first impression of MHDs' traffic composition. We find that HTTP clearly dominates across all of our traces. HTTP contributes 80–97 % of all MHD bytes. Email related protocols account for more than 9 % of the bytes in SEP08, 2.3–2.5 % in APR09 and AUG09a. However, it drops to 0.2 % in AUG09b most likely due to a different usage patterns on weekends. In general, no other protocol has a traffic share of more than 1.5 % with the exception of 13 % unclassified traffic in APR09, and 15 % RTMP streaming in AUG09a, caused by only a handful of MHDs.

### 3.3 MHD Web Traffic

Given that HTTP traffic accounts for the vast majority of MHD traffic we now examine it more closely to characterize its usage and how it differs from overall HTTP usage. We use anonymized HTTP headers and identify HTTP requests from MHDs using user-agents strings as discussed in Sec. 2.2.

To identify the content-type of each transfered HTTP object we join information from the Content-Type HTTP header field and an analysis of the initial part of the HTTP body using libmagic, see [7]. We then group these into a handful of categories. We classify downloads of mobile applications as apps, video and audio content as multimedia, and images as web-browsing since the latter are usually an integral part of Web pages.

Fig. 3 shows the HTTP content type categories for MHDs and compares them with all HTTP traffic. We find that multimedia content is the most voluminous MHD content-type across all traces followed by application downloads. Interestingly, XML objects are also common. They account for 2–5 % of the transfered HTTP bytes. XML is used by many applications for status and data updates, e.g., weather forecasts, stock quotes, and sport results. Surprisingly, Web surfing itself (text based content-types and images) is




(a) CCDF (b) PDF (logarithmic scale)

**Fig. 4.** Size of HTTP objects for all traffic and MHD traffic for trace APR09.

only the third largest category contributing less than 14 % in the 2009 traces (23 % in SEP08).

Comparing these results to all HTTP traffic [7] we find that downloads of mobile applications and XML contribute a significantly smaller fraction to the content type mix. In contrast the volume contributed by RAR archives to all HTTP traffic is significantly larger. Browsing is a bit more prevalent in all HTTP traffic (18–22 %). Multimedia content is the biggest contributor for both. However, for all HTTP traffic flash-video is the most popular video codec, while MHDs use MPEG coding.

The volume share per DNS domain reflects the distribution of MHD content-types. Apple's `apple.com` is responsible for most of the traffic due to application downloads. Note, only the AUG09a trace shows a significant number of iPhone application downloads from third-party sites rather than the Apple's AppStore. YouTube and Stream.fm are the next most popular domains. For overall HTTP traffic One-Click-Hosters and video portals are among the top domains by volume.

To answer the question if MHD HTTP traffic characteristics differ from overall HTTP traffic we compare the distribution of HTTP object sizes. See Fig. 4 for a plot of the Cumulative Complementary Distribution Function (CCDF) and Probability Density Function (PDF) for APR09[4]. The results for the other traces are similar. We find that both distributions are consistent with a heavy-tailed distribution (see Fig. 4(a)). While the dominating mode of objects sizes downloaded by MHDs is larger (see support lines in Fig. 4(b)) the tail is heavier for all HTTP traffic.

### 3.4 Mobile Applications

Fig. 5 shows the popularity of the top MHDs' applications. The most popular application is Apple's browser Safari. Up to 62 % of all devices are using it. This is followed by iTunes (up to 37 %) and Weather (up to 32 %). For non-Apple MHDs we observe

[4] Coupled with a logarithmic scale on the *x*-axis, plotting the density of the logarithm of the data facilitates direct comparisons between different parts of the graphs based on the area under the curve.

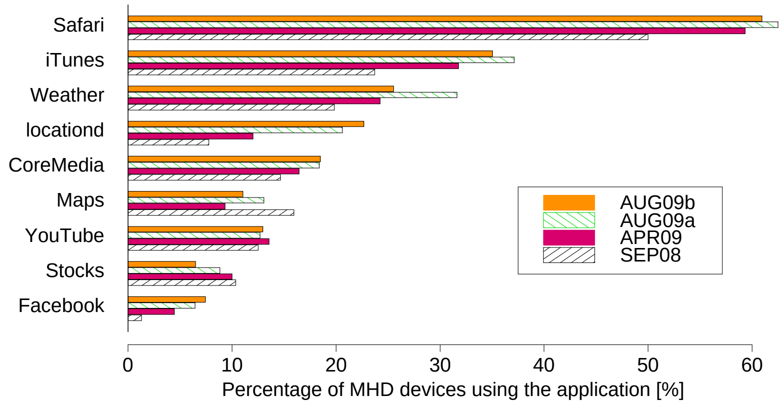



**Fig. 5.** Application popularity by number of MHD devices using this application

that the browser is also the most popular application. Overall we find that Apple's default applications clearly dominate. Surprisingly, given our own usage, the popularity of Maps is relatively low. One possible explanation is that one rarely needs directions while at home. CoreMedia, the media player of iPhones and iPods, is also quite prevalent. This application is e. g., responsible for playing videos accessed via the YouTube application or the browser. The YouTube application itself is only used for searching videos, tagging, and navigating within YouTube. Locationd is the wireless positioning system used on Apple devices.

To understand if users take advantage of specialized applications available for popular Web services we select two Online Social Networks that are popular in our user base: Facebook and StudiVZ. For both OSNs there are specialized applications available for the iPhone/iPod MHDs. We find that roughly half of the users (50 % $\pm$ 10 %) use the specialized applications while the other half continues to use the built-in browser. This relationship is stable throughout our 11 month observation period.

### 3.5 Application and Media Downloads

Given that we are observing traffic from residential DSL lines we have the ability to evaluate if users use their mobile devices or their regular computer to download mobile applications and/or multimedia content. Due to the prevalence of Apple devices in our dataset we now focus on Apple iTunes store and Apple AppStore.

We find that applications are predominantly downloaded directly to the MHD (see Table 2), e.g., more than 70 % of downloads for the 2009 traces. Surprisingly, we see that for AUG09a and AUG09b the volume of application downloads in terms of bytes is almost the same for regular computers and MHD, i. e., the mean application size is larger for applications downloaded by PC/Macs. A detailed analysis reveals that this is caused by outliers; the median application size is the same for both.

We see a vastly different behavior for media downloads or purchases from Apple's iTunes store. Downloads are almost exclusively done via the regular computers. We see

**Table 2.** Downloads from AppStore

| Trace | # Apps available | by PC/Mac Volume | by PC/Mac # Req | by MHD Volume | by MHD # Req |
|---|---|---|---|---|---|
| SEP08 | 3,000 | <1 GB | <100 | <1 GB | <100 |
| APR09 | 7,500 | <1 GB | >100 | >2 GB | >250 |
| AUG09a | 70,000 | >2 GB | >150 | >3 GB | >450 |
| AUG09b | 70,000 | >3 GB | >150 | >3 GB | >400 |

several thousand media files being accessed in the 2009 traces. However, only a handful of downloads are via MHDs which results in a small byte contribution.

## 4 Related Work

Only a small number of studies have focused on Internet traffic in 3G mobile or cellular networks. Svoboda et al. [8] analyze various aspects of GPRS and UMTS traffic using anonymized header traces from 2004 and 2005. They study traffic volume per user and protocol mix. In terms of protocol mix, they find that HTTP is the dominant protocol with 40–60 % of traffic. Heikkinen et al. [5] analyze P2P usage from passive UMTS header traces in Finland from 2005–2007. Web traffic accounts for 57–79 % of bytes from mobile hand-held devices, email for 10–24 %, and P2P is not noticeable. Williamson et al. [13] analyze packet/data call event traces from a CDMA2000 network from 2004. They focus their analysis on link-layer behavior, session properties, and user mobility.

Several studies have analyzed TCP performance and low-level traffic characteristics in GPRS and CDMA data networks [2, 6, 14]. Other studies analyze the content requested or available for mobile devices. Using data from 2000, Adya et al. [1] analyze the Web server logs of a major commercial site and study the requests of mobile clients. They find that stock quotes, news, and yellow pages were the most commonly accessed content in their traces. Timmins et al. [9] use active measurements to crawl the Web for sites offering specialized content for mobile devices. Verkasalo [11] studies how Symbian phone features are used by instrumenting the handset. He finds that the camera feature and games are the most common multimedia applications.

Trestian et al. [10] analyzes mobility and web-application usage in a 3G network from a metropolitan area. We on the other hand, focus on stationary usage when MHDs are connected at home via WiFi. Trestian et al. characterize web-application usage by counting the number of HTTP request and find that social networking, music, and e-mail are the most common web. They do not asses who many *users* utilize a particular application, which is the approach we use to characterize application usage.

## 5 Conclusion

Our analysis of residential broadband DSL lines of a large European ISP shows that there is a significant and increasing number of active MHDs. We find that iPhones and iPods are by far the most commonly observed MHDs. This has an impact on the most popular mobile applications: Safari (Apple's browser), iTunes, and Weather. The largest fraction by volume of MHD HTTP content is multimedia. Comparing HTTP object sizes of overall and MHD traffic we find that MHD HTTP objects are on average larger. The contribution of MHDs to the overall traffic volume is still small, but rapidly growing, especially compared to the overall traffic growth. In future work we plan a more detailed analysis of non-HTTP protocols and refine our methodology for protocol classification. In addition, we plan to extend our analysis to traces from cellular data networks.

## References

1. ADYA, A., BAHL, P., AND QIU, L. Characterizing alert and browse services of mobile clients. In *Proc. Usenix Annual Technical Conference* (2002).
2. BENKO, P., MALICSKO, G., AND VERES, A. A large-scale, passive analysis of end-to-end TCP performance over GPRS. In *Proc. IEEE INFOCOM* (2004), vol. 3.
3. DREGER, H., FELDMANN, A., MAI, M., PAXSON, V., AND SOMMER, R. Dynamic application-layer protocol analysis for network intrusion detection. In *Proc. Usenix Security Symposium* (2006).
4. ERMAN, J., GERBER, A., HAJIAGHAYI, M. T., PEI, D., AND SPATSCHECK, O. Network-aware forward caching. In *Proc. World Wide Web Conference* (2009).
5. HEIKKINEN, M., KIVI, A., AND VERKASALO, H. Measuring mobile peer-to-peer usage: Case Finland 2007. In *Proc. Passive and Active Measurement Conference* (2009).
6. LEE, Y. Measured TCP performance in CDMA 1x EV-DO network. In *Proc. Passive and Active Measurement Conference* (2006).
7. MAIER, G., FELDMANN, A., PAXSON, V., AND ALLMAN, M. On dominant characteristics of residential broadband internet traffic. In *Proc. Internet Measurement Conference* (2009).
8. SVOBODA, P., RICCIATO, F., PILZ, R., AND HASENLEITHNER, E. Composition of GPRS, UMTS traffic: snapshots from a live network. In *IPS-MOME: Workshop on Internet Performance* (2006), Salzburg Research Forschungsgesellschaft.
9. TIMMINS, P., MCCORMICK, S., AGU, E., AND WILLS, C. Characteristics of mobile web content. *Hot Topics in Web Systems and Technologies* (2006).
10. TRESTIAN, I., RANJAN, S., KUZMANOVIC, A., AND NUCCI, A. Measuring serendipity: connecting people, locations and interests in a mobile 3g network. In *Proc. Internet Measurement Conference* (2009).
11. VERKASALO, H. Empirical observations on the emergence of mobile multimedia services and applications in the U.S. and Europe. In *Proc. Conference on Mobile and ubiquitous multimedia* (2006).
12. WIKIPEDIA. Smartphone: Operating systems. `http://en.wikipedia.org/w/index.php?title=Smartphone&oldid=315621107#Operating_systems`, Sep 2009.
13. WILLIAMSON, C., HALEPOVIC, E., SUN, H., AND WU, Y. Characterization of CDMA2000 cellular data network traffic. In *Proc. IEEE Conference Local Computer Networks* (2005).
14. WON, Y. J., PARK, B.-C., HONG, S.-C., JUNG, K. B., JU, H.-T., AND HONG, J. W. Measurement analysis of mobile data networks. In *Proc. Passive and Active Measurement Conference* (2007).
15. WOOD, N. Mobile data traffic growth 10 times faster than fixed over next five years. In Total Telecom: `http://www.totaltele.com/view.aspx?ID=448681`, Sep 2009.