Traffic Measurement and Analysis of a Broadband Wireless Internet Access

Rastin Pries*, Florian Wamser*, Dirk Staehle*, Klaus Heck†, Phuoc Tran-Gia*
*University of Würzburg, Institute of Computer Science, Würzburg, Germany
Email: {pries,wamser,staehle,trangia}@informatik.uni-wuerzburg.de
†Hotzone GmbH, Berlin, Germany, Email: heck@hotzone.de

***Abstract*—The increasing broadband wireless Internet usage and the limited wireless resources require a careful network management and optimization of the wireless Internet Service Providers (ISPs). Unfortunately, those providers often just have statistics about the overall usage and limited knowledge about the detailed application distribution as well as the traffic characteristics.**

In this paper we present user and traffic characteristics measured at a broadband wireless Internet access. The results show that the applications change quickly but the general characteristics like packet size and TCP/UDP percentage have not changed during the last years.

***Index Terms*—traffic measurements, traffic classification, home networks, BWA, WLAN**

I. Introduction

WIRELESS LAN based on the IEEE 802.11 standard is becoming more and more popular as a broadband wireless Internet access. Laptops and other mobile devices provide convenient wireless access to the Internet. Due to the simple installation and integration into existing networks, Wireless LAN is gaining popularity at home, in public facilities, and in more and more areas of our everyday life. Along with the increasing number of wireless Internet users, the number of wireless ISPs rises.

These service providers have to continuously optimize their networks in order to react to the increasing number of wireless Internet users as well as changing application demands of the mobile devices. The optimization highly depends on the user behavior and the expected network traffic. According to Cisco [1] there is a growth in volume of all applications as well as a considerable variation in the traffic mix. Particularly, the percentage of P2P file sharing traffic is decreasing in favor of streaming and video traffic. Cisco expects about 40 % P2P file sharing traffic in 2009 after 60 % in 2006. However, the total volume of P2P file sharing traffic doubles between 2006 and 2009. Video traffic is estimated with approximately one quarter of all consumer Internet traffic whereas over two third or 7700 petabytes of the entire monthly Internet traffic is estimated as consumer traffic.

To see the changing application structure, we performed measurements at an ISP for home users who provides broadband wireless Internet access. The goal is to react to the changing traffic characteristics caused by new applications, to optimize the network performance, and to reveal the user behavior in a wireless broadband access network.

In this paper, we present the results of these measurements like application distributions as well as changing traffic characteristics caused by user demands and new services. The results are used by a network service provider to optimize its network performance in order to give QoS guarantees for home users in its fixed wireless network.

The remainder of the paper is organized as follows. Section II provides a background of traffic measurements and shows the related work. This is followed by Section III, introducing our measurement scenario and methodology. Section IV shows the results of the measurements. Finally, conclusions are drawn in Section V.

II. Background & Related Work

It has been a challenge for years to structure a reliable and feasible measurement architecture. First, a measurement has to generate detailed traffic characteristics, including global and special statistics, like application-based or user-based ones. Second, a measurement always affects the measured data.

A. Traffic Measurements

Commonly, there are two different approaches to measure a network: active polling and passive monitoring. The measuring process of the active measurements generate new traffic and inject it into the network, while passive measurements monitor and capture the network traffic. Latter systems use the recorded traffic to produce several statistics with the help of analysis software. A well-known passive measurement architecture is the OC3MON by MCI [2] that was used by Thompson et al. [3] and McCreary et al. [4] to monitor optical ATM OC-3 links. The CoralReef suite [5] developed by CAIDA is based on the OC3MON but supports GPS timing and allows only link speeds of up to OC-12 (622 Mbps). One drawback is that the use of passive measurements raises serious privacy and security problems. Furthermore, such systems generate a lot of traffic with large traces depending on the link rate.

B. Traffic Classification

After collecting the data, the services have to be classified. Service classification has its own research group and with the emergence of new services like P2P, it is getting more and more difficult to identify packets. At the network link an unordered mix of packets are collected that should be first

978-1-4244-2517-4/09/$20.00 ©2009 IEEE

grouped in connections and afterwards classified connection-wise. There are several techniques to classify packets. We used two recent methods which are described in the following:

Payload-based classification: It is also known as content-based method and is a syntactic analysis of the applicative layers of a packet. The classification entity is seeking deterministic character strings in the IP packet payload with fast regular expressions. The problem is that a detailed knowledge of the application as well as the format of its packets are needed. Several disadvantages are known: Character strings are not always available or the payload may be encrypted. However, this method only depends on a few characteristic packets. Karagiannis et al. [6] developed a heuristic for transport layer identification of P2P traffic which includes payload based methods. Baset and Schulzrinne [7] tried to detect among other techniques Skype up to version 1.4 with some characteristic bytes. Ehlert et al. [8] proposed further byte sequences for Skype version 2.0 in 2006. A Wiki devoted to the identification of network protocols is used by the Application Layer Packet Classifier for Linux (L7-filter) [9] to allow a real-time classification.

Host behavior classification: Due to the limitations of the payload-based classification, Karagiannis et al. [10] propose another approach for traffic classification called BLINC. They try to classify the popularity and the transport layer interactions with the help of inherent host behavior. The focus is shifted from classifying flows to associating hosts with applications. The flows are then classified accordingly. They specify three different levels. The social level includes the popularity of a host. It describes whether a host is part of a large group of clients that communicate with each other. A group of clients is called a community. Second, the functional level distinguishes service providers from service consumers. This is done with the IP/port ratio of the connections of a host. A service provider normally have a large amount of different clients connected to a few local ports. Whereas a consumer only opens a few ports per outgoing connection. Finally, at the application level the interactions between particular hosts on specific ports and IPs are examined to characterize a host. With this method, Karagiannis was able to present some heuristics to detect malware, P2P, web, chat, ftp, game, and streaming traffic.

In the next section, we describe our measurement setup and how we used the payload-based and host behavior classification in combination.

III. Measurement Scenario and Methodology

In this paper, we focus on traffic characteristics of home users in a wireless network. The measurements have been performed at a Germany-wide wireless access provider who provides, along with business network access, private Internet access in large housing estates.

A. Measurement Setup

The measurements were taken at an ISP switching center providing access for 250 households. The customers got access over Wireless LAN at several access points before the traffic was multiplexed at an IEEE 802.11a radio link. The dimensioning of the radio link is done by the provider according to the upcoming traffic of the users. Measurements of the provider confirmed that the link almost never operates at full capacity.

The measuring unit was deployed right after the access points in the hard-wired network. The monitoring point for the measurement is shown in Fig. 1. We measured both directions with the help of a receive-only network tap which ensures that the productive network was not interfered by our measurement. Our meter runs on a Linux system. It observes packet headers using two commodity 100BaseT Ethernet cards via libpcap.

The measurement process basically consists of five steps. First, raw traces are captured in pcap packet capture files. Additionally, the real-time classification entity described in the next paragraph stores detection data in log files. Second, the traffic traces are filtered to suppress or to make sensitive information anonymous. The anonymization module scrambles data in order to raise effort needed to obtain sensitive information about the internals of an operational network. Afterwards, the filtered traces are checked for errors and submitted in a database-driven repository. The last step was to analyze the traces by running analysis scripts. The analysis is performed offline at external computers. All further work is done either within the database itself with the help of database languages or by querying the database.

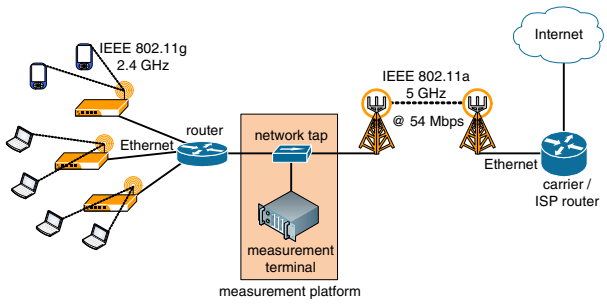


Fig. 1. Measurement setup

B. Service Classification

Our classification involves two levels of detection. On the one hand we used a payload-based detection with the Application Layer Packet Classifier for Linux. However, this method requires the payload of the packets which we were not allowed to store in capture files because of privacy concerns.

The payload-based classification is done in the following way: The first classification method is performed in real-time. A connection tracking assigns the packets to flows. If a new flow is detected, the classification scans the first N packets of this flow and the first M bytes within these packets for payload signatures. This is done online before the capturing. First, the traffic is checked for P2P data cause it may use arbitrary ports. Then, it is scanned for well-known common applications. At the very end we try to detect Skype traffic. If the payload does not match at all, the packet is classified as “unknown”,

especially all encrypted and new protocols are classified as unknown in the payload-based detection.

Our second classification method is a host behavior analysis similar to the proposed one by Karagiannis [10]. The connections of a host were investigated as in the functional level approach. We recorded the usage of ports and IP addresses per host and compared the results of unknown hosts to already classified hosts. Thus, we were able to distinguish P2P traffic from web and streaming traffic. The host behavior classification is done after capturing at the data repository. The major advantage is that it is also capable to detect encrypted traffic. However, a detection of certain applications is in turn not possible. Therefore, a traffic class called "unclassified P2P" is shown in Section IV which is P2P file sharing traffic of an unrecognized application.

C. Limitations

The monitoring and classifying of unknown traffic has always some difficulties and limitations which have to be taken into account. Several issues occurred during the measurement which are enumerated below for completeness.

Classification payload patterns: The traffic patterns tend to underestimate or overestimate the traffic. It is difficult to find reliable packet signatures that match only the intended protocol. In all cases, a random encrypted stream may fit to several patterns. The other way round, some patterns are only able to match a part of the whole desired traffic. Namely, in our case the Skype pattern was one of the pattern that tend to overestimate. Furthermore, some badly designed unimportant application patterns are simply left out.

Anonymization, packet capture length: During the capturing of packets the capture length is set to 96 bytes to make sure that the whole header is included in the traces. Due to privacy issues, the IP and payload anonymization cleared the rest of the payload in such a way that only the packet headers remained in the trace files. Consequently, we had no usable information about the payload during the offline analysis.

Measuring times: The traces are not collected equally distributed over the week. They differ in duration and time of capturing.

D. Trace Description

The measurements were made from July 13th, 2007, till July 24th, 2007. The whole measurement last about 7 days and about 150 GB measurement data was collected. Further on, the Internet service provider gave us Cisco Netflow statistics of routers, which proved our measurements in data volume and packet count. The billing system was flat rate. Moreover, the packet loss during the capturing of packets in trace files was negligible and sums up to 0.18 % in downlink direction and 0.09 % in uplink direction.

IV. Measurement Results

This section presents the results of the traffic measurements at the broadband wireless Internet access. The general traffic statistics are included in the first part and the second part deals with the traffic classification.

A. General Traffic Statistics

The general traffic statistics involve daily traffic fluctuations as well as packet size distributions. The traffic fluctuations of one day are shown in the upper figure of Fig. 2. The traffic statistics were gathered on July 23th, 2007, and the curve presents a 5 min average. Although the results are shown for one day only, we have seen a similar characteristic during all our days of measurement.

We can observe that the downlink bandwidth varies during the day. Comparing the figure to the results presented by Perenyi et al. [11] where the throughput triples between 6 and 12 o'clock, we can see a decrease of the ratio caused by the constant nighttime P2P file sharing traffic. In our result, the traffic increases in the late afternoon up to midnight. This is obvious because the measurements have been performed in a home network in contrast to a business network. During the day, the users are not at home and use the Internet only in the evening. This becomes even more obvious when classifying the traffic. We observed almost no YouTube streaming traffic during the day, but in the evening hours with a peak at 10 pm. A similar daily behavior was seen for web browsing and email traffic. P2P file sharing was the only protocol that was used constantly all day long.

A complete week with focus on the weekend can be seen in the lower figure of Fig. 2. Looking at the statistics both, the daily variations and the weekend can be identified. Due to the fact that a lot of users spend their weekends not in front of a PC, the total average throughput is lower compared to the weekdays. Concluding the results of the daily fluctuations, we have seen a difference between our measurements and the measurement results taken in the backbone. The differences come from the fact that our results just include the home users and the backbone measurements include both, home and business Internet usage. P2P file sharing was used all day and streaming applications are mainly used in the evening.

As the daily fluctuations differ from the backbone measurements, we want to evaluate if this is also the case for the packet size distributions. Thompson et al. [3] show a trimodal packet size distribution where nearly half of the packets are 40 to 44 bytes, 20 % are 576 bytes, and 10% are 1500 bytes in length. Sean McCreary and kc claffy [4] show that about 80 % of the

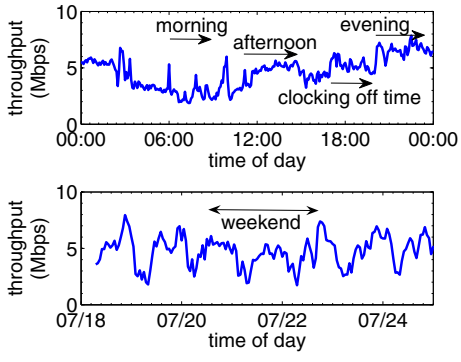
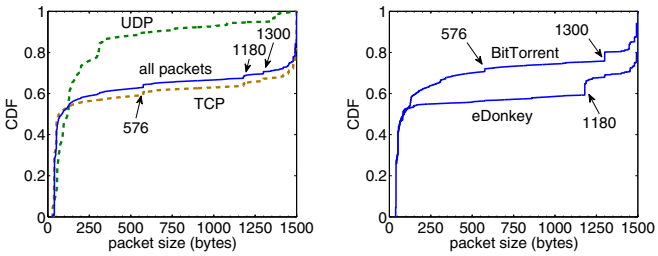
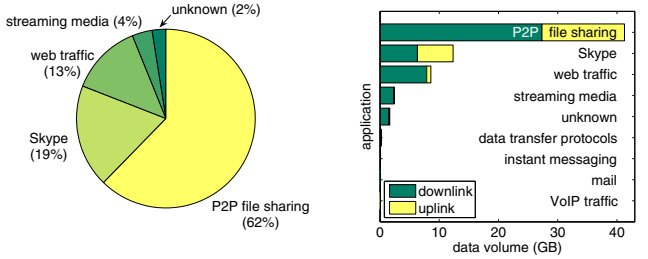


Fig. 2. Daily and weekly downlink throughput



(a) TCP and UDP packet size distribution (b) eDonkey and BitTorrent packet size distribution

Fig. 3. Cumulative IP packet size distribution



(a) Relative application distribution (b) Daily application distribution

Fig. 4. Application distribution

packets are smaller than 600 bytes but have observed the same trimodal packet size distribution as Thompson. The newest backbone traffic packet distribution we found is presented by John and Tafvelin [12] in 2007. In contrast to the previous two papers, they show a bimodal traffic distribution where 40 % are of size smaller than 44 bytes and another 40 % of the packets are between 1400 bytes and 1500 bytes. Their results are similar to our measurements at the broadband wireless Internet access, shown in Fig. 3(a).

Looking at the figure, we can observe several things. First, 90 % of the UDP packets are smaller than 500 bytes. This might be P2P control or real-time streaming traffic. Second, looking at the curve for all packets, we observe a bimodal packet size distribution. The first peak occurs at around 40 bytes and the second step at 1500 bytes. This shows that most packets are transmitted via TCP, with the 40 bytes Acknowledgments and the 1500 bytes Ethernet Maximum Transfer Unit (MTU), which is also underlined with the TCP packet size distribution curve. However, we can also observe three small steps at 576 bytes, 1180 bytes, and 1300 bytes. These packet sizes are used by P2P file sharing protocols as shown in Fig. 3(b).

Packets of size 1180 bytes are only used by the eDonkey protocol which was also observed by Karagiannis et al. [6]. Furthermore, they have shown a similar packet size distribution for BitTorrent. 1300 bytes is the MTU recommended by some ISPs for DSL connections. Therefore, we think that these packet sizes result from downloads from clients of such ISPs.

Finally, we take a look at the protocol distribution on the transport layer. Almost 88 % of all measured packets are transmitted via TCP, only 11.6 % via UDP, and less than one percent is used for ICMP control traffic. Considering the total throughput in bytes, 95 % of the complete data is transmitted via TCP.

B. Traffic Classification

After we evaluated the general traffic statistics and compared them to the related work, we want to evaluate if the application distribution differs compared to fixed-line networks. In 2005, P2P file sharing traffic overtook HTTP traffic in terms of traffic volume. Perenyi et al. [11] measured 60 % to 80 % P2P file sharing traffic of the total broadband traffic in 2006. Cisco Systems also estimates the whole P2P file sharing traffic for the year 2006 to about 60 % or 1358 PB per month [1]. However, according to Cisco's traffic forecast, P2P will fall behind HTTP in the next years. They estimate for 2009 about 40 % P2P traffic and a strong increase of streaming and video traffic.

The measurements presented in Fig. 4 show that still 62 % of the complete traffic is used by P2P file sharing applications. The large percentage of P2P file sharing traffic results from the measurements in a broadband home network. Mainly, this is especially interesting for home network service providers to optimize their services.

According to Karagiannis et al. [6], web traffic consumes with 50 % the largest amount of traffic in the core. In our measurement however, only 13 % web traffic was measured. Fig. 4(b) shows the exact data volume of the traffic categories and further distinguishes between downlink and uplink volume. Web traffic includes browsing and file downloads with HTTP but not streaming over HTTP. Surprisingly, we noticed a new user download behavior. Some customers use extreme HTTP downloads from large file-hosting sites as an alternative to P2P file sharing. Most notably, during the prioritizing and the shaping of the traffic, this was detected as a problem. HTTP proxies may help to limit the outbound traffic.

Although VoIP and FTP are prioritized by the ISP, the usage is very low. In case of VoIP this has several reasons. First, the network can not meet the user expectations and second, IP phones and VoIP devices mainly provide hard-wired interfaces. Surprisingly, we have not seen any gaming traffic. This might result from the fact that gamers normally use a DSL connection with smaller delays compared to the measured multi-hop broadband wireless Internet access.

The high percentage of measured Skype traffic shown in Fig. 4(a) results from the heuristics to detect the traffic. It is difficult to detect Skype traffic as it uses a large variety of ports and the protocol changes with every version. Therefore, we think that the used heuristics can not be trusted and we have to count Skype traffic to the unknown traffic.

In addition to conversational Skype traffic, we measured about 4 % streaming traffic which does not fulfill our expectations and was lower than in the core. Similar to VoIP traffic, real-time streaming traffic needs higher QoS requirements. Consequently, it is not surprising that the fraction of non-live streaming as Flash Videos was measured with 14 % of the

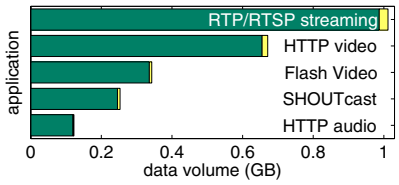


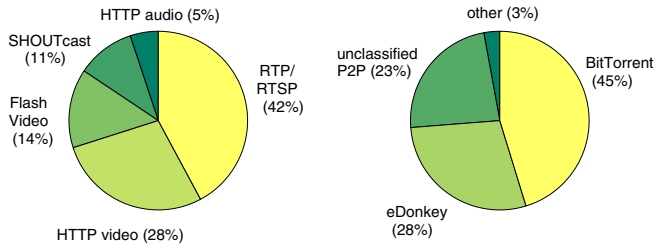
Fig. 5. Application distribution of streaming media (dark part of the stack: downlink, bright one: uplink)

whole streaming traffic. The exact distribution of the streaming traffic is shown in Fig. 5.

It is rather complicated to assign specific media players to the different protocols since most players are able to handle several protocols. The biggest portion, RTSP and RTP are used by Quicktime, Real Player, and the Windows Mediaplayer. However, all players support HTTP video as well. The only difference between these two groups is how the connection is established. Fig. 6(a) shows the percentage of the streaming traffic. The 14 % Flash Video belong to YouTube videos. Surprisingly, SHOUTcast, an audio streaming service, is still frequently used.

Finally, the P2P differentiation is shown in Fig. 6(b). Compared to old statistics with the largest portion of eDonkey traffic, this has changed in our network. Now, about 45 % of the P2P traffic belong to BitTorrent. The 23 % unclassified P2P file sharing traffic has been detected by the P2P host behavior statistics as P2P traffic but the filter was unable to assign the traffic to eDonkey or BitTorrent.

Summarizing, our measurements still underline the traffic characteristics of previous published papers with a high fraction of P2P file sharing traffic. However, we expect an increase of streaming traffic in the near future.



(a) Streaming media (b) P2P file sharing

Fig. 6. Subcategory application distribution

V. Conclusion

This paper presents the results of our Internet traffic measurements in a commercial broadband wireless network for home users. The results show that the daily traffic fluctuations differ from measurements in the backbone. As the measurements were taken in a home network, the overall throughput increases during the evening hours in contrast to backbone measurements with a traffic increase between 9 am and 12 pm. The increase is mainly caused by web and streaming traffic which are frequently used in the evenings. In contrast, P2P file sharing traffic is used all day and night.

Similar to the latest backbone measurements [12], we observed a bimodal packet size distribution. 43 % of the packets have a length of 40 bytes and 30 % of the packets contain 1500 bytes of information. This results from the 88 % measured TCP packets, containing 95 % of the complete measured traffic.

Our second part, the traffic classification has shown that over 60 % of the measured traffic belongs to P2P file sharing applications. This percentage underlines other results from backbone measurements with 60 % to 80 % P2P file sharing traffic [11]. Surprisingly, eDonkey with 28 % is not the most popular P2P file sharing application anymore. BitTorrent is now responsible for the largest portion of the P2P traffic.

In contrast to P2P file sharing traffic, VoIP and online games are used very seldom which is seen as characteristic in a broadband wireless network at the moment. Finally, we have to point out that we detected a different download behavior. Some customers use extensive file downloads as alternative to P2P and FTP. To reduce the traffic of such downloads HTTP proxies should be considered.

References

[1] Cisco Systems, “Cisco Visual Networking Index - Forecast and Methodology, 2007-2012,” White Paper, June 2008.

[2] J. Apisdorf, K. C. Claffy, K. Thompson, and R. Wilder, “OC3MON: flexible, affordable, high performance statistics collection,” in *Proc. of INET 97*, June 1997.

[3] K. Thompson, G. J. Miller, and R. Wilder, “Wide-Area Internet Traffic Patterns and Characteristics (Extended Version),” *IEEE Network*, vol. 11, no. 4, pp. 10–23, November/December 1997.

[4] S. McCreary and K. C. Claffy, “Trends in Wide Area IP Traffic Patterns - A View from Ames Internet Exchange,” in *Proceedings of the 13th ITC Specialist Seminar on Internet Traffic Measurement and Modelling*, Monterey, CA, 2000.

[5] K. Keys, D. Moore, R. Koga, E. Lagache, M. Tesch, and K. C. Claffy, “The Architecture of CoralReef: an Internet Traffic Monitoring Software Suite,” in *PAM2001 — A workshop on Passive and Active Measurements*, April 2001.

[6] T. Karagiannis, A. Broido, N. Brownlee, K. C. Claffy, and M. Faloutsos, “File-sharing in the Internet: A characterization of P2P traffic in the backbone,” University of California, Riverside, University of California, Riverside Department of Computer Science, Surge Building, Riverside, CA 92521, Tech. Rep., November 2003.

[7] S. A. Baset and H. Schulzrinne, “An Analysis of the Skype Peer-to-Peer Internel Telephony Protocol,” in *Proceedings of INFOCOM 2006. 25th IEEE International Conference on Computer Communications.*, Barcelona, Spain, April 2006, pp. 1–11.

[8] S. Ehlert and S. Petgang, “Analysis and Signature of Skype VoIP Session Traffic,” Fraunhofer FOKUS, Berlin, Germany, Tech. Rep. NGNI-SKYPE-06b, July 2006.

[9] “Application Layer Packet Classifier for Linux (L7-filter).” [Online]. Available: http://l7-filter.sourceforge.net/

[10] T. Karagiannis, K. Papagiannaki, and M. Faloutsos, “BLINC: multilevel traffic classification in the dark,” in *SIGCOMM ’05: Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*, New York, NY, USA, 2005, pp. 229–240.

[11] M. Perenyi, T. D. Dang, A. Gefferth, and S. Molnar, “Identification and Analysis of Peer-to-Peer Traffic,” *Journal of Communications (JCM)*, vol. 1, no. 7, pp. 36–46, November/December 2006.

[12] W. John and S. Tafvelin, “Analysis of internet backbone traffic and header anomalies observed,” in *IMC ’07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, New York, NY, USA, 2007, pp. 111–116.