©2009 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The definitive version of this paper has been published in Next Generation Internet Networks 2009 (NGI 2009) 2009 10.1109\/ngi.2009.5175772.

On Traffic Characteristics 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*—Internet traffic measurements and traffic characterization are essential for managing and optimizing network infrastructures. The increasing number of wireless Internet users and the changing application demands require consecutive traffic measurements. Therefore, we have performed measurements of home users at a broadband wireless access service provider in order to reflect the current traffic characteristics. 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 Quality of Service (QoS) guarantees for home users in its fixed wireless network.**

***Index Terms*—traffic measurements, traffic classification, broadband wireless access**

I. Introduction

DURING the last years the Internet has emerged as the key component for commercial and personal communication which is reflected by the exponential traffic increase. The German Commercial Internet Exchange (DE-CIX) point has had a peak traffic rate of 500 Gbps on October 20th 2008 whereas the peak traffic rate was approximately 200 Gbps 12 months before. According to Cisco Systems [1] this trend will continue. Over two third or 7700 petabytes of the monthly traffic is generated by consumers. The large bandwidth demands are caused by the fast changing application requirements. The applications range from low bandwidth email traffic over web browsing and Peer-to-Peer (P2P) traffic to high bandwidth multimedia streaming. YouTube, as an example for video streaming, generates 5 % to 8 % of the complete Internet traffic. Cisco Systems [1] claims that 38 % of the consumer Internet traffic is generated by Internet video in 2009 and expects that it increases to 48 % in 2012.

The total traffic increase and especially the increase of real-time applications require a careful network planning and optimization. This applies for fixed-line as well as for wireless providers. Traffic measurements are one essential part for the Internet Service Provider (ISP) to optimize their network. According to the measurement results, the ISP can adapt its prioritization strategies in order to guarantee a good perceived quality for the end user. However, most public available measurement data was gathered in the backbone and show the global traffic characteristics but do not reveal the user and application demands. According to Fukuda [2] there is a significant different traffic usage pattern in residential broadband traffic. Therefore, we performed the measurements close to the end user, namely at an ISP for home users who provides a shaped broadband wireless Internet access.

The measurements were performed in 2008 and reflect the Internet usage of 250 households. Afterwards, the measurement data was classified using a combination of payload-based classification and host behavior. This paper shows the results of these measurements like daily traffic fluctuations, session statistics as well as application distributions. In contrast to our previous publication which is based on measurement data from 2007 [3], we have seen an immense growth of streaming traffic which is also underlined by Cisco Systems [4].

The remainder of the paper is organized as follows. Section II gives an overview of traffic measurements and its classification together with the related work. This is followed by Section III introducing our measurement scenario and methodology. Section IV shows the results of the measurements and 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. The following measurement systems meet these main issues.

A. Traffic Measurements

Commonly, there are two different approaches to measure a network: active polling and passive monitoring [5], [6]. 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. The following monitoring systems use the passive approach.

Brownlee et al. [7] use RTFM [8], an Internet standard real-time flow measurement system with its open source implementation NeTraMet. It is a versatile and very general system for collecting flow data and includes a high level language for filtering, managing, and aggregating observed packets into flows. However, due to the fact that it needs to see headers for every packet through a device, it is not easy to implement in a switch or a router.

Fraleigh et al. [9] designed a passive monitoring system to capture packet level traffic measurements on various ATM and SONET links. It is called IPMON and is inspired by the well-known OC3MON architecture by MCI [10] that is used by Thompson et al. [11] and McCreary et al. [12] to monitor optical ATM OC-3 links. IPMON has the capability to collect packet traces of up to OC-48 link speeds (2.4 Gbps) for a period of at least several hours. In addition, it uses GPS for synchronization. The CoralReef suite [13], developed by CAIDA, is originally based on the OC3MON, too. It is similar to IPMON, but does not support GPS timing and allows only link speeds of up to OC-12 (622 Mbps). Tools like CoralReef provide network card drivers, various programming APIs, and applications for capturing and analysis. A popular application programming interface for capturing network traffic is libpcap. Compared to the solutions above, it is only a computer library on top of network drivers and not a whole architecture. Shannon et al. [14] used libpcap to capture network traffic for further analysis.

Commercial solutions are available from Endace like DAG cards. Endace sells DAG cards for Ethernet and optical networks which allow to collect packet traces of up to OC-192 or Gigabit Ethernet link speeds. Karagiannis et al. [15], [16] and John et al. [17] used DAG cards and software for their measurements.

Finally, some routers have the ability to export global per-flow summaries including start time, flow duration, byte and packet volume, IP addresses, and port numbers. In Cisco routers the tool for this purpose is called Netflow [18], [19]. It is embedded within the Cisco IOS software and is widely used to collect IP traffic information. Even though initially implemented by Cisco, Netflow will be standardized by the IETF. Juniper Networks, Nortel Networks, and Huawei Technology provide similar features within their routers.

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 [16]. At the network link an unordered mix of packets is collected that should be first grouped in connections and afterwards classified connection-wise. Along with Port-based classification, several techniques and methods exist to classify packets:

Port-based classification: The correlation between port number and application type defined by the Internet Assigned Numbers Authority (IANA) is used. It is the simplest and most traditional method, but has some drawbacks. The port numbers are not defined for all applications. Especially some applications use port ranges or they even assign the ports dynamically so that the mapping of the ports and the applications can not be trusted. Hence, a detection with this method is not possible. Thompson et al. [11], McCreary et al. [12], and Shannon et al. [14] used port-based classification and mapped each IP packet to a named application by choosing the first matching rule from an ordered collection of protocol/port patterns.

Payload-based classification: It is also known as content-based method. Payload-based classification 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. [15], [20], [21] developed a heuristic for transport layer identification of P2P traffic which includes payload based methods. A Wiki devoted to the identification of network protocols is used by the Application Layer Packet Classifier for Linux (L7-filter) [22] to allow a real-time classification.

Host behavior classification: Due to the limitation above, Karagiannis et al. [23] proposed another approach for traffic classification. 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. With this method, Karagiannis was able to present some heuristics to detect malware, P2P, web, chat, ftp, game, and streaming traffic.

Statistical classification: This is a recent method that uses statistical descriptions of the traffic with supervised learners. A statistical parameter can be the packet size or the inter-arrival time. First order Markov chains or k-Nearest Neighbors, Linear or Quadratic Discriminant Analysis are proposed by [24], [25] to calculate the probability of a packet to the statistical data model of an application. The statistical method has some performance issues but can also detect tunneled or encrypted traffic.

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 offers, along with business network access, private Internet access in large housing estates. The measurement and the classification is done according to proposals and papers introduced in the related work section.

A. Measurement Setup

The measurements were performed at an ISP switching center which provides access for 250 households. The customers have access over Wireless LAN at several access points before the traffic is 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 is set up right after the access points in the 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

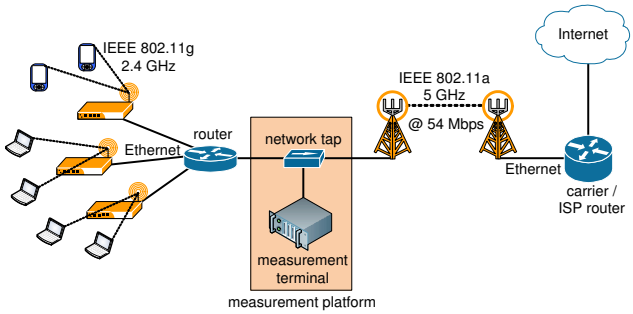


Fig. 1. Measurement setup

productive network is 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 is the analysis of the traces which 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.

B. Service Classification

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

The payload-based classification is done in the following way: First, 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. The traffic is checked for P2P file sharing data because it may use arbitrary ports. Afterwards, it is scanned for well-known common applications. 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 [23]. The connections of a host are investigated as in the functional level approach. We record the usage of ports and IP addresses per host and compare the results of unknown hosts to already classified hosts. Thus, we are able to distinguish between P2P file sharing, web, and streaming traffic. The host behavior classification is done at the data repository after the packet capturing. 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 is one of the pattern that tend to overestimate and therefore added to the unknown traffic. Furthermore, some badly designed unimportant application patterns are simply left out in our analysis.

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 have no usable information about the payload during the offline analysis.

Diverse HTTP usage: During the measurements, we noticed that the HTTP usage statistics are varying. Some customers use extreme HTTP downloads from large file-hosting sites. Although these downloads do not represent the typical web browsing behavior, they are included in the web traffic statistics.

Traffic Shaping: The wireless access provider uses traffic shaping to control the Internet traffic. Due to the fact that the Cisco routers are configured to prefer web and real-time traffic, P2P traffic might be underestimated in the following results.

D. Trace Description

The measurements were made from July 11th, 2008, until July 29th, 2008. The whole measurement last 19 days and about 400 GB measurement data was collected. Further on, the Internet service provider gave us Cisco Netflow statistics of routers, which prove our measurements in data volume and packet count. The billing system of the ISP is flat rate. Moreover, the packet loss during the capturing of packets in trace files is negligible and sums up to 0.18 % in downlink direction and 0.09 % in uplink direction.

IV. Measurement Results

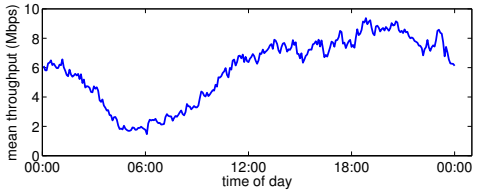
This section presents the results of the traffic measurements at the broadband wireless Internet access. The general daily traffic fluctuations are included in the first part, the second part deals with session statistics of users, and the last part shows a detailed traffic classification.

A. Daily Traffic Fluctuations

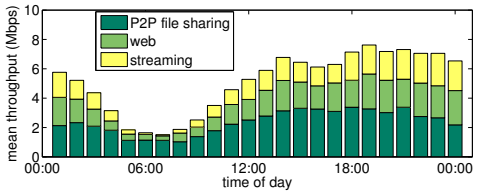
First of all, we take a look at the mean throughput variations during a day. The mean throughput is calculated by first dividing all measurement data into days, then splitting each day into 5 min samples, and finally calculating the mean of all nineteen 5 min samples. The throughput fluctuations during the day are shown in Fig. 2(a). The x-axis shows the time of the day while the y-axis shows the mean throughput. It is obvious that the throughput decreases after 1:00 o'clock down to a minimum at 6:00 o'clock. Afterwards, the throughput increases with a maximum throughput at 19:00 o'clock in the evening. Similar daily traffic fluctuations can be found in [2], [26].

Since the traffic is varying over the day, we distinguish between constant and fluctuating traffic in Fig. 2(b). The figure shows the daily traffic statistics according to the three main applications P2P file sharing, web, and streaming traffic. Comparing the three different application categories we can see that P2P file sharing traffic is still the dominating application during the whole day with the largest percentile at night. These are users running their computers 24/7.

Web and streaming traffic consume almost the same amount of traffic but the total throughput varies during the day. At night, almost no web and streaming traffic is present in contrast to P2P file sharing traffic. The high streaming throughput is rather surprising compared to Ploumidis et al. [27] with only 0.177 % of streaming traffic measured in 2005. The reason for this streaming traffic increase is the popularity of new video streaming services. YouTube for example generates 5 % to 8 % [1] of the complete Internet traffic and the platform was set up at the end of 2005. Summarizing, we see that 1.6 Mbps is constantly used by P2P over the day. Web and streaming traffic are varying over the day because they need user interaction. Further investigations on the traffic classes are shown in Section IV-C.



(a) Mean daily traffic fluctuations



(b) Mean daily application distribution

Fig. 2. Mean throughput and application distribution

B. Session Statistics

After having seen the daily throughput statistics we investigate the user behavior in more detail. We evaluate the session statistics of the users. A session is thereby defined as follows. Several flows of one user regardless of the application with an inter-flow-time lower than 5 min (timeout: 5 min) belong to one session. Furthermore, a session has to last longer than 10 s and needs a minimum session volume of 10 KB in order to distinguish between periodic signaling and normal traffic.

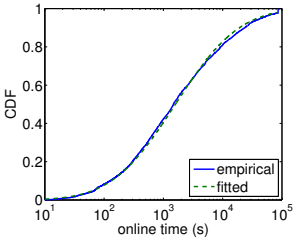
4590 sessions are identified during the whole measurement and the maximum online time is 24 h due to the fact that each measurement run lasts 1 day. The mean number of sessions per user is 2.1 and the maximum number of sessions 74. Further session statistics are shown in Table I.

TABLE I
ONLINE SESSION DURATION AND SESSION VOLUME

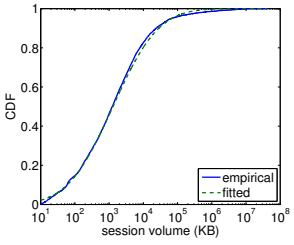
	Mean	Max.	Percentile 25th	Percentile 50th	Percentile 90th
Session duration					
All-day	136 min	24 h	6.7 min	24.6 min	395.2 min
Weekend	167 min	24 h	8.3 min	33.4 min	465.0 min
Weekday	129 min	24 h	6.5 min	23.7 min	362.9 min
Session volume					
All-day	80 MB	42 GB	246 KB	1.24 MB	23.52 MB
Weekend	98 MB	22 GB	294 KB	1.74 MB	33.55 MB
Weekday	76 MB	42 GB	238 KB	1.15 MB	22.29 MB

The weekend data is gathered at 2 weekends and the mean session duration during the weekend is 167 min compared to 129 min during the week. This is not surprising as most home users spend more time in front of their computers during the weekend. However, what is surprising is the median of the session volume. 1.24 MB in 24.6 min seems to be a very low amount of data. We think that the reason for this lies in instant messaging services and periodic email checking. Although only 3 % of the complete data volume belong to messaging services, 10 % of all traffic flows belong to this class.

Finally, we can see a large gap between the 50 % quantile and the maximum session volume (1.24 MB to 42 GB). The few large sessions belong to P2P and web file downloads. In order to further analyze the session duration and session volume, the Cumulative Distribution Functions (CDFs) of both statistics are plotted in Fig. 3.



(a) Session duration



(b) Session volume

Fig. 3. Empirical cumulative distributions and lognormal distributions fitting the empirical functions

The x-axis in Fig. 3(a) shows the logarithmic scale of the session duration. Unfortunately, we gathered the measurements on a daily basis and therefore it is not possible to identify session longer than one day. However, we can see that 3 % of the sessions last at least one day and these sessions belong to P2P file sharing traffic. The curve can be well fitted by a lognormal distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad (1)$$

with $\mu = 2.8879$ and $\sigma = 2.0577$.

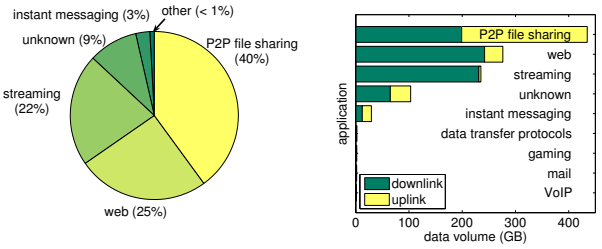
Looking at the CDF of the session volume in Fig. 3(b) we can see a larger heterogeneity compared to the session duration. About 3 % of the sessions have a volume larger than 1 GB whereas 90 % of the sessions have a volume smaller than 23.52 MB. This curve can also be well fitted by a lognormal distribution with $\mu = 7.1650$ and $\sigma = 2.4066$.

Chlebus and Divgi presented session statistic of a Wi-Fi hotspot network in [28], [29]. Their definition of a session slightly differs from ours. It is created when a user logs into the network and ends when the user logs out or is timed out of the network. Unfortunately, the length of the timeout is not defined. According to their statistics, a user has on average 2.16 sessions per day, consuming a mean of 12.24 MB in about one hour. The maximum session duration was 34 hours consuming 1.5 GB of data and the maximum number of sessions per user was measured 37. They fitted the curves with a truncated Pareto distribution. Although their results differ from ours, the general distribution of the session duration and the session volume are similar.

C. Traffic Classification

After we evaluated the daily traffic fluctuations and the session statistics and compared them to the related work, we want to evaluate if the application distribution differs compared to fixed-line networks. Towards the end of 2005, P2P file-exchange applications overtook web traffic as the major contributor of traffic on the Internet. P2P traffic was measured at 60 % to 80 % of the total broadband traffic [26]. Cisco Systems states in their annual report that 60 % or 1358 PB per month belong to P2P traffic at the end of 2006 [1], [4]. However, this percentage decreases and Cisco estimates a P2P traffic percentage of 40 % (3075 PB per month) at the end of 2009 and an increase of streaming traffic to 40 % (3073 PB per month).

Looking at Fig. 4(a), our measurements underline these statements. 40 % of the complete measured traffic belong to P2P file sharing applications. However, we have to point out that this value is only achieved with the traffic shaping of the ISP which is essential in order to perceive an acceptable streaming and web quality. On the other hand the percentage of P2P file sharing traffic without traffic shaping is estimated based on former measurements at around 60 %, which is a higher value as in backbone measurements. A higher percentage of P2P file sharing traffic clearly results from the measurements in a home network. Mainly, this is especially



(a) Relative application distribution (b) Application distribution

Fig. 4. Application distribution

interesting for home network service providers to optimize their services.

Web traffic was measured with up to 50 % in the core [15]. In our environment only 25 % web traffic was detected. However, our web traffic fraction includes browsing and file downloads with HTTP but not streaming over HTTP which belong to a separate streaming traffic category. Surprisingly, we notice 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 is detected as a problem. HTTP proxies may help here to limit the outbound traffic. Fig. 4(b) shows the exact data volume of the traffic categories and further distinguishes between downlink and uplink volume.

Although VoIP and FTP (data transfer protocol) are prioritized 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 wired interfaces. Besides the low usage of VoIP and FTP, we have also seen only a few 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 low usage of VoIP and Internet games is seen as characteristic for a wireless broadband access network at the moment.

In contrast, streaming traffic with about 22 % of the whole traffic is now besides web and P2P file sharing traffic one of the main traffic categories used in home environments. On the one hand this is surprising when comparing it with previous publications from Ploumidis et al. [27] with 0.177 % and Pries et al. [3] with 4 % of streaming traffic. On the other hand, this is nearly the predicted value of Cisco Systems [1]. 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, HTTP video are used by Quicktime, Real Player, and the Windows Mediaplayer. However, all players support RTP/RTSP streaming as well. The only difference between these two groups is the way the connection is established.

If the player is called using ”rt[s]p://”, the l7-filter assigns the connection to the RTSP class and if the connection

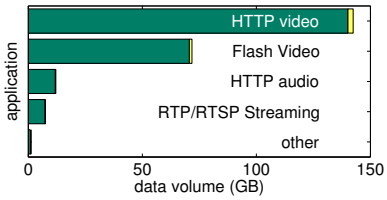
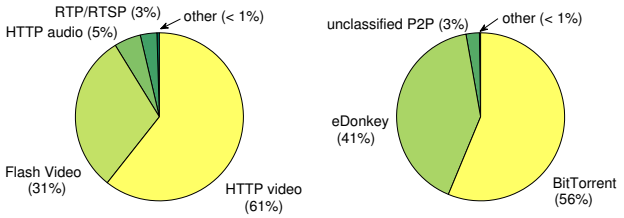


Fig. 5. Application distribution of streaming media (dark part of the stack: downlink, bright one: uplink)



(a) Streaming media (b) P2P traffic

Fig. 6. Subcategory application distribution

is established using "http://" the connection belongs to the HTTP video class. However, the Real Player normally uses RTSP for streaming. Besides these two classes, another similar streaming protocol can be used called Microsoft Media Server Protocol (MMS), which was not detected during our measurements.

Fig. 6(a) shows the percentage of the streaming traffic. Similar to VoIP traffic, real-time streaming traffic has higher QoS requirements. Consequently, it is not surprising that the fraction of non live streaming as Flash Videos is measured with 31 % of the whole streaming traffic.

Finally, the P2P differentiation is shown in Fig. 6(b). Compared to old statistics with the largest portion of eDonkey traffic, this is different in our measurement. About 56 % of the P2P file sharing traffic belongs to BitTorrent. The 3 % 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.

V. Conclusion

This paper presents the results of our Internet traffic measurements in a commercial broadband wireless access network for home users. The results of the daily traffic fluctuations show a similar behavior compared to the statistics from the German Internet point DE-CIX. A breakdown of the application distribution during the day shows that P2P file sharing traffic is used all day long whereas the amount of web and streaming traffic increases in the evening hours with a peak at 19:00 o'clock. This is also reflected in the session statistics with P2P sessions lasting all day long. Although the statistics differ from Chlebus and Divgi [28], [29], the general distribution of the session volume and session duration are similar.

Our traffic classification statistics affirm the predicted trends of P2P, web, and streaming traffic with empirically determined values. The percentage of P2P file sharing traffic is with 40 % lower compared to 62 % measured in 2007 [3]. The decrease is caused by the increase of streaming traffic to 22 %. Within the streaming traffic Flash Video increases to 31 %. Furthermore, a second reason for the decrease might be a change in the download behavior of some customers. They use extensive HTTP downloads as alternative to P2P and FTP file sharing. A breakdown of the P2P file sharing traffic shows that eDonkey is with 41 % not the most popular P2P application anymore. BitTorrent is now responsible for the largest portion of the P2P traffic. This might indicate a slightly change in the P2P protocols in Germany.

The low fraction of VoIP and gaming traffic in our measurements is seen as characteristic for broadband wireless access networks and isolates them from other access technologies. In case of VoIP it is on the one hand caused by the network and on the other hand by the lack of wireless IP phones and wireless capable VoIP devices.

Summarizing we want to point out that the general traffic fluctuations remain similar to previous measurements whereas the application distribution differs. Streaming applications become more and more important and are now responsible for one fourth of the complete traffic. The high QoS requirements of streaming applications necessitates a change in the prioritization scheme of the ISP.

References

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

[2] K. Fukuda, K. Cho, and H. Esaki, "The impact of residential broadband traffic on Japanese ISP backbones," *SIGCOMM Comput. Commun. Rev.*, vol. 35, no. 1, pp. 15–22, 2005.

[3] R. Pries, F. Wamser, D. Staehle, K. Heck, and P. Tran-Gia, "Traffic Measurement and Analysis of a Broadband Wireless Internet Access," in *IEEE VTC Spring 09*, Barcelona, Spain, April 2009.

[4] Cisco Systems Inc., "Approaching the Zettabyte Era," White Paper, June 2008.

[5] V. Paxson, J. Mahdavi, A. Adams, and M. Mathis, "An Architecture for Large-Scale Internet Measurement," *IEEE Communications*, vol. 36, no. 8, pp. 48–54, August 1998.

[6] C. Fraleigh, S. Moon, B. Lyles, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, and C. Diot, "Packet-Level Traffic Measurements from the Sprint IP Backbone," *IEEE Network*, vol. 17, no. 6, pp. 6–16, November-December 2003.

[7] N. Brownlee and K. C. Claffy, "Understanding Internet traffic streams: Dragonflies and tortoises," *IEEE Communications Magazine*, vol. 40, no. 10, pp. 110–117, October 2002. [Online]. Available: http://www.caida.org/outreach/papers/2002/Dragonflies/cnit.pdf

[8] N. Brownlee, C. Mills, and G. Ruth, "Traffic Flow Measurement: Architecture," October 1999.

[9] C. Fraleigh, C. Diot, B. Lyles, S. B. Moon, P. Owezarski, D. Papagiannaki, and F. A. Tobagi, "Design and Deployment of a Passive Monitoring Infrastructure," in *IWDC '01: Proceedings of the Thyrrhenian International Workshop on Digital Communications*, Taormina, Italy, 2001, pp. 556–575.

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

[11] 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.

[12] 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.

[13] 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.

[14] C. Shannon, D. Moore, and K. C. Claffy, "Beyond folklore: observations on fragmented traffic," *IEEE/ACM Transactions on Networking (TON)*, vol. 10, no. 6, pp. 709–720, December 2002.

[15] 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.

[16] T. Karagiannis, M. Faloutsos, A. Broido, N. Brownlee, and K. C. Claffy, "Is P2P dying or just hiding?" in *IEEE Global Telecommunications Conference, 2004. GLOBECOM '04*, November-December 2004, pp. 1532–1538. [Online]. Available: http://www.caida.org/outreach/papers/2004/p2p-dying/p2p-dying.pdf

[17] 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.

[18] R. Caceres, N. G. Duffield, A. Feldmann, J. Friedmann, A. Greenberg, R. Greer, T. Johnson, C. Kalmanek, B. Krishnamurthy, D. Lavelle, P. P. Mishra, K. K. Ramakrishnan, J. Rexford, F. True, and J. E. van der Merwe, "Measurement and Analysis of IP Network Usage and Behavior," *IEEE Communications Magazine*, vol. 38, no. 5, pp. 144–151, May 2000.

[19] Cisco Systems, Inc., "Cisco IOS NetFlow," White Paper, October 2007.

[20] T. Karagiannis, A. Broido, M. Faloutsos, and K. C. Claffy, "Transport layer identification of P2P traffic," in *IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, New York, NY, USA, 2004, pp. 121–134.

[21] T. Karagiannis, M. Molle, and M. Faloutsos, "Long-Range Dependence: Ten Years of Internet Traffic Modeling," *IEEE Internet Computing*, vol. 8, no. 5, pp. 57–64, 2004.

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

[23] 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.

[24] H. Dahmouni, S. Vaton, and D. Rosse, "A markovian signature-based approach to IP traffic classification," in *MineNet '07: Proceedings of the 3rd annual ACM workshop on Mining network data*, New York, NY, USA, 2007, pp. 29–34.

[25] L. Bernaille, R. Teixeira, and K. Salamatian, "Early application identification," in *CoNEXT '06: Proceedings of the 2006 ACM CoNEXT conference*, New York, NY, USA, 2006.

[26] 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.

[27] M. Ploumidis, M. Papadapouli, and T. Karagiannis, "Multi-level application-based traffic characterization in a large-scale wireless network," in *International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Helsinki, Finland, 2007.

[28] E. Chlebus and G. Divgi, "The Pareto or Truncated Pareto Distribution? Measurement-Based Modeling of Session Traffic for Wi-Fi Wireless Internet Access," in *IEEE Wireless Communications and Networking Conference, 2007.WCNC 2007*, Hong Kong, China, March 2007.

[29] G. Divgi and E. Chlebus, "User and Traffic Characteristics of a Commercial Nationwide Wi-Fi Hotspot Network," in *IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007*, Athens, Greece, September 2007.