A five year perspective of traffic pattern evolution in a residential broadband access network

Li, Jie; Aurelius, Andreas; Nordell, Viktor; Du, Manxing; Arvidsson, Åke; Kihl, Maria

Published in:
[Host publication title missing]

2012

A five year perspective of traffic pattern evolution in a residential broadband access network

Jie Li[1], Andreas Aurelius[1], Viktor Nordell[1], Manxing Du[1], Åke Arvidsson[2], and Maria Kihl[3]
[1]*Acreo AB, Networking and Transmission Lab, Electrum 236, SE-164 40 Kista, Sweden*
Tel: +46 8 632 7746, Fax: +46 8 632 7710, Email: jie.li@acreo.se
[2]*Ericsson AB, Ericsson Research, Färögatan 6, SE-164 80 Stockholm, Sweden*
Tel: +46 10 7151897, Fax: + 46 10 7146525, Email: Ake.Arvidsson@ericsson.com
[3]*Lund University, Dept. of Electrical and Information Tech., John Ericssons väg 4, Lund, Sweden*
Tel: +46 46 2229010, Email: Maria.Kihl@eit.lth.se

Abstract: In this paper we describe a systematic study on long-term evolution of residential broadband Internet traffic covering 5 calendar years from June 2007 to May 2011. The traffic evolution is characterized both in the term of the total traffic volume, as well as the traffic volumes and shares for different application categories (file sharing, video streaming etc.), with the focus on comparing the traffic on the per IP user basis and among different broadband subscription groups. The results show that the average daily total traffic generated by each private end user increased only by about 33 % during the past 5 years. Further, the results show that the P2P file-sharing has been dominating the network total traffic, but the daily file-sharing traffic volume per end user largely remains the same. Also, the daily streaming-media traffic volume per end user has increased dramatically by over 500% during the studied period of time. In the meantime, the daily web-browsing traffic volume per end user has increased by about 300%. Finally, a further investigation among 4 different FTTH broadband subscription groups with 1, 10 , 30, and 100 Mbit/s symmetric access speeds shows that the lower the access speed, the more diversified the end user traffic tend to be.

Keywords: Traffic monitoring, residential Internet traffic pattern, end user behavior, long term evolution, file sharing, streaming media

1. Introduction

Internet has now become the global information and communication base for every aspect of people's daily work and life. Even so, global Internet traffic keeps on growing steadily with the ever-increasing mainly video-oriented content distributions and services, as well as with the ever-increasing new devices that are connected to the network [1]. This puts challenges for network operators and ISPs to deliver broadband IP services that meet the quality-of-service (QoS) requirements of multimedia services needed to provide sufficient quality-of-experience (QoE) to the end users [2].

One important part in meeting this challenge is to understand the Internet traffic characteristics, especially the Internet traffic patterns of residential end users. These users have generated according to [1] about 80% of the global IP traffic in the years 2010-2011, and they are expected to have even larger total traffic share in the next few years. Indeed, many research groups have so far published a large number of internet traffic measurement and analysis results. For example, in an early study [3] the traffic statistics in the Sprint IP backbone network were investigated regarding traffic workload, traffic applications, packet delay, TCP flow round trip times etc. The analyses were based on short-term (one week)

traffic measurement results and did not distinguish business and private consumer traffic. In the work presented in [4, 5] the residential traffic of 7 major Japanese ISPs were studied based on up to 3 month long traffic data. However, in this investigation only total traffic volume analyses were carried out. In [6], user traffic characteristics of 20,000 residential DSL customers in an urban area were reported, based on short-term (up to 10 consecutive days) traffic measurements. They found that HTTP including video-over-HTTP traffic dominated the network (57.6%). However, the traffic analyzing tool adopted in this work was not well capable of identifying peer-to-peer (P2P) traffic, hence a large portion of traffic were unclassified. In [7], Fiber-to-the-home (FTTH) customer activities were characterized, based on up to 6 month traffic data for about 2500 end users. In [8, 9] traffic characteristics of wireless networks were reported, which showed that even for wireless networks P2P file sharing constitutes more than 50% of the total network traffic.

Despite the numerous publications so far in the context of Internet traffic measurement and analyses, there are few studies focusing on the long-term traffic pattern evolutions. In [10] a 7 year long traffic investigation was reported. However, in this effort, only 15 minute traffic each day from 14:00 to 14:15 were recorded, and the traffic data were extracted from the backbone of a Japanese academic network connecting only universities and research institutes. In another effort [11], focus was put on comparing specially P2P traffic during years 2002-2004 on two OC48 (2.5 Gbps) links of Tier 1 ISPs. In another paper [12] week-long residential traffic data sets from 2004 to 2008 (one week in each year) of 6 major Japanese ISPs were used to study the traffic volume evolution trend, however no further results of traffic application breakdown were given in this work. Furthermore, apart from research papers, there are also commercial reports published yearly on the global Internet traffic statistics [e.g. 1, 13, 14]. These reports usually focus on the macro traffic application shares and total traffic volumes over large geographic areas (West Europe, North America etc), but reveal little information on e.g. the actual end user number and broadband access speed distributions, thus lack the information of the traffic volume and applications on the per end user and per access speed bases.

In this report, we describe, to the best of our knowledge, the first time systematic study on long-term evolution of residential broadband Internet IP traffic covering 5 calendar years from June 2007 to May 2011. The traffic data were collected from a medium-sized Swedish municipal network that has about 2600 end users, mainly using FTTH broadband. The traffic evolution is characterized both in the term of the total traffic volume, as well as the traffic volumes and shares for different application categories of the network traffic (file sharing, video streaming etc.), with the focus on comparing the traffic on the per IP end user and per access speed bases.

2. Target network, traffic data collection and storage

The network, in which the measurements were performed, is a medium-sized municipal network in Sweden [7]. There are approximately 2500 FTTH households connected to the network, and a small number (~ 100) of DSL lines and some enterprise and campus end users. Further, the network is an open network, hence there are several ISPs to choose from, and each ISP offers a set of subscription types with the maximum symmetric access speed at 100 Mbit/s.

The traffic data collection tool used in the study was PacketLogic (PL) [15], a commercial traffic management device. In PL, traffic is identified based on packet content (deep packet inspection and deep flow inspection) instead of port definitions. The device can identify more than 1000 Internet application protocols, and the signature database is continuously updated. The identification process is connection-oriented, which means that each established connection between two hosts is matched to a certain application protocol. Using PL, for 90-99% of the traffic, a match was found during the identification period. PL

uses the traffic in both directions in the identification process, and records all the traffic (both the inbound and the outbound) that pass through it every five-minute in the form of the traffic volume, the traffic application, the actual timing, and the IP address for each traffic record during the 5 minute period of time.

For collecting the traffic data, the PL was connected to the network via optical 50/50 splitters. The measurement point is the Internet Edge (IE) aggregation point, where the service providers are connected to the network [7]. In order to enable long-term study of the target network traffic statistics, a MySQL data base was established to store the original traffic records collected by PL. When transferring data to the MySQL server, the IPs are hashed, ensuring that no violation of integrity or law is done. In parallel, the log data of the DHCP server of the network are also stored in the same data base. The DHCP server log contains information as timing, (hashed) IP address, broadband service subscription type, access switch and access port. Thus, in linking the two data set tables (PL record data and DHCP log) in the same data base, each traffic record collected by PL can be traced to e.g. the end user's broadband subscription type, the access switch etc., and by matching the (hashed) IP addresses of the two data sets, the traffic that were not generated by private end users can be excluded in the data analyses, as is the case in this study.

3. Traffic evolution over 5 years

In this study, all the traffic analyses are based on the 'daily' traffic statistics. Here, 'daily traffic' means that the stored raw traffic data of every 5 minutes (with the time stamping from kl 00:00 to kl 23:55) were summed up over every 24 hour period of time. In the meantime, the summed up daily traffic are also further grouped into different application categories according to the following classifications:

- *File Sharing*, including peer-to-peer (P2P) file sharing traffic and client-server file sharing traffic, but not including HTTP-based direct download file sharing traffic offered by e.g. RapidShare and Megaupload [16];
- *Streaming Media*, including both video and audio streaming traffic, such as flash video over HTTP (e.g. YouTube), HTTP media stream, real-time streaming protocol (RTSP), and real-time messaging protocol (RTMP) family. P2P media stream is also included in this category;
- *Web Browsing*, traffic generated by HTTP including its plugins. Note that the traffic generated by file hosting direct download services (RapidShare, Megaupload etc.) as mentioned above is classified into this category (as HTTP download traffic);
- *Unknown*, traffic that are not recognizable by PL;
- *Others*, including traffic categories of *File Transfer*, *Messaging and Collaboration*, *Entertainment*, *Network Infrastructure*, *Business Systems*, *Remote Access*, *Information*, and *Malware*. This is because that, as can be seen in the following sections of this paper, the total traffic belonging to these application categories amounted only up to a maximum of 6% of the total network traffic during one calendar year period of time.

3.1 End-user number and broadband subscription distributions

As important background information, Fig. 1 shows the counted daily active end user number and their broadband subscription distributions for 3 periods at the beginning, middle, and the end of the studied period of time. One can see that, apart from the end of 2007 to the early 2009 where the measurement was targeted specifically to the small amount of DSL end users, the measurement covered broadband subscribers with 1 Mbit/s up to 100 Mbit/s access speeds, among which the 10 Mbit/s subscription has been the most popular choice. In the meantime, there have been in average about 140 (symmetric) 30 Mbit/s broadband subscribers and particularly about 75 (symmetric) 100 Mbit/s broadband subscribers that already from year 2007 have been connected to the network and

under the measurement coverage. Further, Fig. 1 shows that there are unfortunately a few discontinuity periods of traffic data collection during the studied period of time. This is partly due to administration reasons of the studied network, e.g., the reconfiguration of the network, and partly due to the replacement of the measurement equipment.

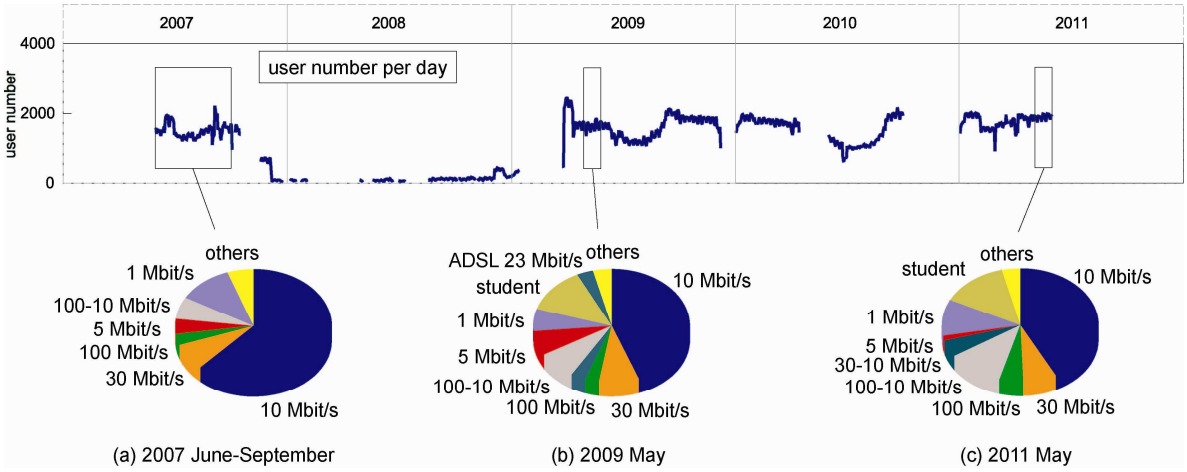

Fig. 1. Daily active end user number and their broadband subscription type distributions during the selected 3 periods of time.

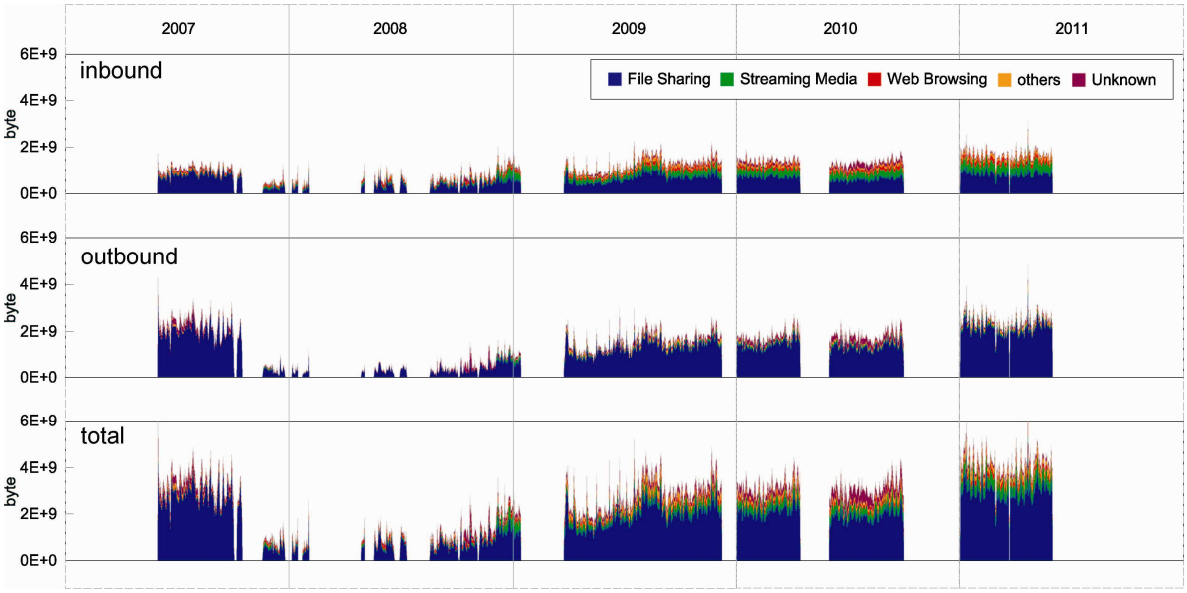

Fig. 2. Daily network traffic evolution over the studied period per end user.

3.2 Daily traffic evolutions on the per end user basis

Fig. 2 shows the end user averaged daily traffic evolutions for the inbound, outbound, and the total traffic, respectively, in the form of stacked traffic volumes belonging to different traffic categories. Note that in order for easy comparison, the Y-axis (traffic volume) scales are the same. We can see that the average daily total traffic per end user increased only by about 33 % over the past five years, from about 3 Gbyte to about 4 Gbyte per day. In the meantime, even though the P2P file-sharing remains the dominant network bandwidth consumer, the daily file-sharing traffic per end user largely remains the same and already reached its peak at the beginning of the studied period of time, and (apart from year 2008 when only a small amount of DSL end users were under investigation) even suffered a major drop from the first half year of 2009. This is attributed to the Swedish enforcement of the European Union's Intellectual Property Rights Enforcement Directive (IPRED) effective from April 1 2009 [7], which has seen almost 50% drop of file sharing traffic on

April 1 2009 compared to the day before. On the other hand, the daily streaming-media traffic per end user has increased dramatically by over 500% since the end of year 2008, from about 0.06 Gbyte per day to about 0.5 Gbyte per day. As shown in Fig. 2, the streaming-media traffic is inbound dominant. Even so, it can be noticed that the outbound streaming media traffic has also gained significant growth from the end of year 2008, and this is due to the dramatic growth of P2P enabled streaming media applications [17]. Also, the daily web browsing traffic volume per user has increased by about 300% over the past 5 years, from about 0.075 Gbyte per day in 2007 to more than 0.2 Gbyte per day in 2011.

3.3 Traffic application shares

Fig. 3 shows the traffic application shares of the total recorded traffic data during each calendar year period. Note that apart from the period of 2008 during which the traffic shares are based on the traffic of about 100 DSL broadband subscribers, all the traffic share statistics are dominantly based on traffic of FTTH end users. From this figure we can see that, firstly, file sharing has been dominating both the inbound and outbound network traffic, even though the total file sharing traffic share has decreased from 86% in 2007 to 71% in the first five months of 2011. For the studied network, the file sharing traffic has been overwhelmingly P2P based with just a very little amount of client-server traffic (< 0.1%), and BitTorrent has been the mostly used P2P file sharing protocol (accounting for over 90% of the P2P file sharing traffic). This is in contrast to what was stated in [14] and particularly in [1] that file sharing amounted to only 40% and 34% of global consumer Internet traffic in year 2010 and 2011, respectively. One possible explanation of this observation difference is that P2P file sharing, e.g. BitTorrent, may tend to find local peers within the same access network, hence a large amount of P2P traffic are kept flowing among them and not visible at a higher level Internet exchange point. This suggests that in order to have a full picture of the residential Internet user traffic characteristics, it is necessary to do the traffic measurement as close to the end user as possible. The other feature of file sharing traffic for the studied network is that the outbound traffic share has been significantly more than that of the inbound traffic. This behaviour is well known for FTTH based networks where a P2P user with high uplink bandwidth is tended to be chosen as a good peer acting as a virtual server for other P2P users.

Secondly, streaming media has emerged to be the second largest contributor to the traffic of the studied network, from amounting to only 2% of the total network traffic in 2007 to up to 13% of the total network traffic in the first 5 months of 2011. As stated earlier, the streaming media traffic consists of video, audio, as well as P2P streaming media traffic. The dominant contributors to the streaming media traffic in the network are not surprisingly video applications consisting mainly of flash video over HTTP (YouTube is classified into this sub-category), HTTP media stream, and RTMP family that, for example, amounted to 89% of the total streaming media traffic in May 2011, while for the same period of time P2P and audio amounted to 4% and 7% of the total streaming media traffic, respectively. Hence, unlike file sharing that the uplink outbound traffic dominates, for streaming media the downlink inbound traffic has a dominantly larger share, amounting up to 26% of the total inbound traffic in the first five months of 2011.

Thirsdly, for web browsing, the traffic also features a much larger inbound traffic share than that of the outbound traffic. Apart from year 2007, the web browsing traffic share remains largely the same at the level between 11%-14% for the inbound traffic, and 6%-8% for the total traffic. Worth to note here also is that the HTTP download traffic within this category has increased significantly from less than 30% in May 2009 to over 40 % in January 2011 of the total web browsing traffic. This is in accordance with the observations in [16] and attributed to the ever-growing popularities of one-click file hosting direct download services offered by e.g. RapidShare and Megaupload.

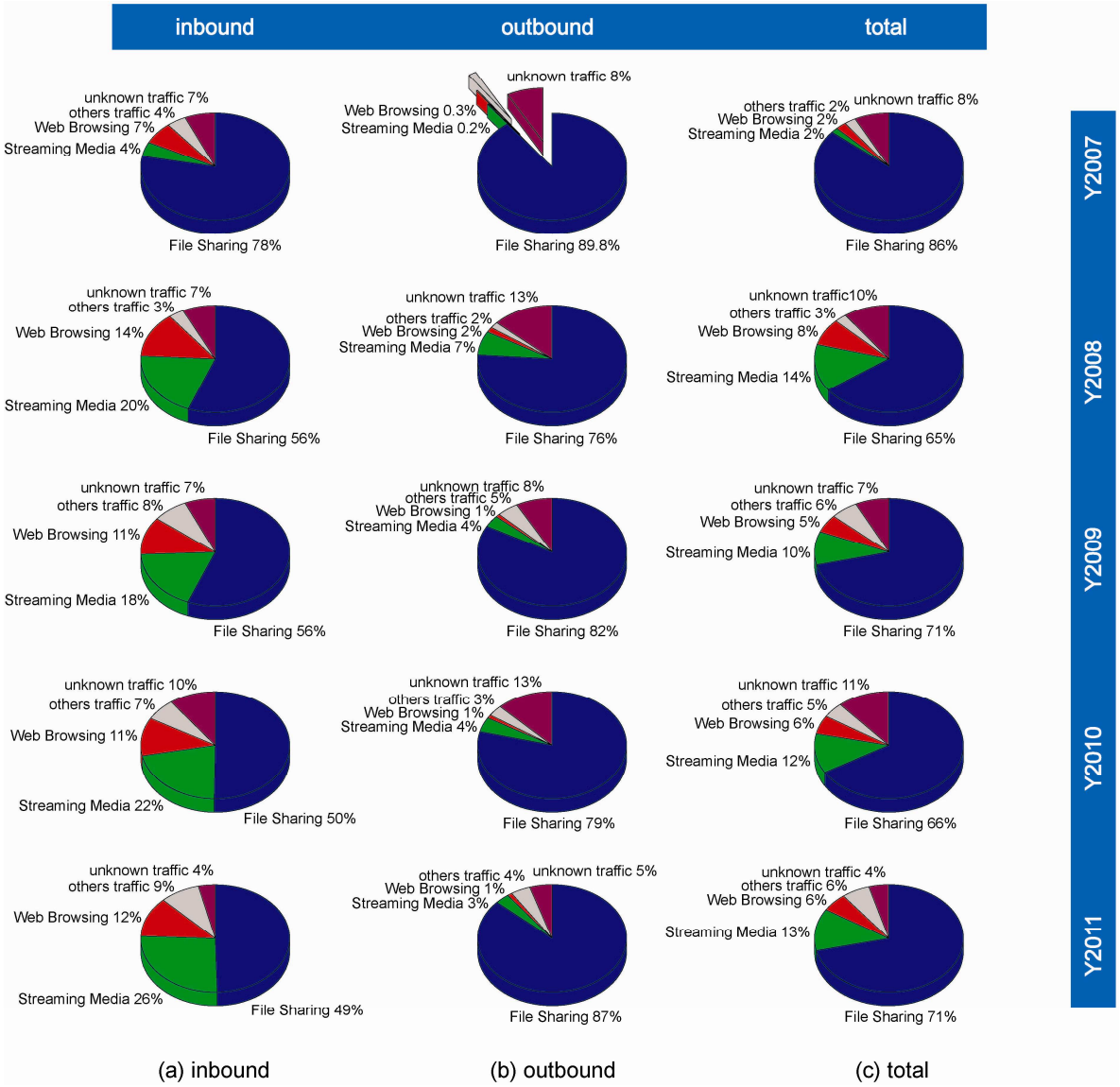

Fig. 3. Total traffic application share evolution over the studied 5 year period of time.

3.4 Traffic evolutions of different FTTH broadband end users

As a further step, we investigated the daily traffic pattern evolutions for different end user broadband subscription groups. In this step, we have chosen the ISP of the network that has the largest group of broadband subscribers. 4 groups of end users of this ISP were selected, with the symmetric broadband access speeds of 1, 10, 30, and 100 Mbit/s, respectively. For the 10 Mbit/s and 30 Mbit/s broadband connections, the covered end user numbers under the study period have remained at the level of 500 and 110, respectively, while for the 1 Mbit/s and 100 Mbit/s access speeds, the end user numbers were at the level of 100 and 50 at the beginning of the studied period in year 2007, and gradually increased to the level of 200 and 100 since year 2011, respectively.

Fig. 4 summarizes the total (of both inbound and outbound) daily traffic evolutions per end user for the 4 different end user groups. Note here that the Y-axes denoting the traffic volumes have different scales due to the large (traffic volume) differences among them. From this figure it is clearly shown that the traffic evolution differs significantly among the 4 end user groups. More specifically, first of all, the lower the access speed, the more diversified traffic evolution. For the first year of 2007, the blue coloured file sharing traffic exclusively dominated the total daily traffic for all the 4 different subscription groups. However, since the year 2009, the 10 Mbit/s and especially the 1 Mbit/s end users' traffic

have tended to be more and more diversified, with the streaming media traffic shares reached to 15% and even up to 36% of the total traffic during the first five months in 2011, for the two subscription groups respectively. In the mean time the file sharing traffic ratio has dropped to 71% and even down to 32%, for the 10 Mbit/s and 1 Mbit/s end users, respectively. On the other hand, for the 30 Mbit/s and 100 Mbit/s end users, the file sharing has always been dominating their daily internet usage with over 80% of the total traffic share all the time throughout the studied period.

Secondly, apart from the 1 Mbit/s end users, the daily total traffic volume per end user remains largely the same over the studied period of time. The daily traffic volumes per end user during the first five months of 2011 are 0.5, 3.2, 6.4, and 13.2 Gbyte, for the 1, 10, 30, and 100 Mbit/s end users, respectively

Thirdly, for the 1 Mbit/s end users, apart from streaming media, the web browsing and messaging traffic also have significantly larger traffic shares than those with higher access speeds, suggesting these 'lower end' users may be more socially oriented. On the other end, for those 'higher end' 100 Mbit/s users, apart from file sharing, they also generate a significantly larger amount of remote access and file transfer traffic, suggesting that at least some of those 100 Mbit/s access lines are used for some kind of home enterprise activities.

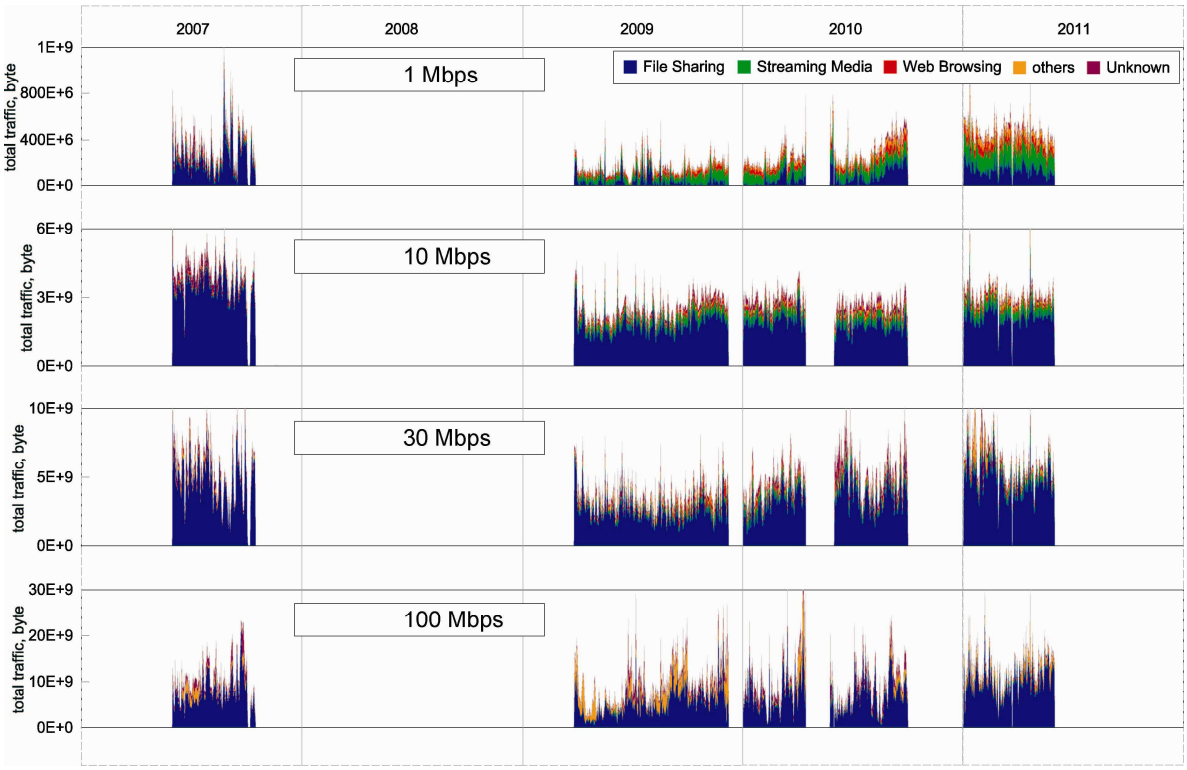

Fig. 4. Total daily network traffic evolution per end-user over the studied period for 1, 10, 30, and 100 Mbit/s end users, respectively. Note the difference of the Y-axis scales in the figure.

4. Discussions and conclusion

In this work we carried out a systematic study on long-term evolutions of residential broadband Internet traffic covering 5 calendar years from June 2007 to May 2011. The novelty of this work is that the traffic volume and traffic application evolutions are analyzed on the per end user and per access speed bases, which is important and inevitable in understanding and predicting a network total traffic growth trend based on the maturity of the studied network (the increase rate of new end users, the broadband subscription distributions etc). In this study, we find that the average daily total traffic generated by each private Internet end user has an annual growth rate of only 6%. One conclusion drawn from

this observation is that the major Internet traffic growth driving forces are from the new end users and new devices that are connected to the network. Another conclusion may be that the customers in FTTH networks have reached a maturity in Internet consumption much before other users, and hence the growth rate has slowed down in these networks. It should be noted here, that much of the traffic volume has been attributed to file sharing traffic, and that major changes in usage are seen due to legislative decisions. This affects the end results heavily. The other observation of this study is that even though the P2P file sharing has been dominating the network traffic, the daily file-sharing traffic volume per end user largely remains the same. On the other hand, the daily streaming-media traffic volume per end user has experienced a dramatic annual growth rate of about 40%, and has now become the second largest traffic application category. From this observation, one may conclude that streaming-based new video and audio services such as YouTube, on-line TV etc. are indeed the driving forces for the increase of the total network traffic generated by the end users. A further investigation among 4 different FTTH broadband subscription groups with 1, 10, 30, and 100 Mbit/s symmetric access speeds shows that the lower the access speed, the more diversified the end user traffic tend to be. This may suggest that online streaming media services, though have enjoyed a dramatic growth in the past few years, are still not mature enough to topple the P2P file sharing oriented online habits especially for heavy Internet end users.

Acknowledgement

This work was supported by European CELTIC initiative in the project of IPNQSIS. Christina Lagerstedt, Qu Shuai and Hemamali Pathirana are acknowledged for their continuous effort in traffic data collections.

References

[1] White Paper, “Cisco Visual Networking Index: Forecast and Methodology, 2010–2015”, June 2011, Cisco

[2] IP Network Monitoring for Quality of Service Intelligent Support (IPNQSIS), http://projects.celtic-initiative.org/ipnqsis/

[3] 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, 2003.

[4] K. Fukuda, K. Cho, and H. Esaki, “The impact of residential broadband traffic on japanese isp backbones”, *ACM SIGCOMM Computer Communications Review*, vol. 35, no. 1, 2005.

[5] K. Cho, K. Fukuda, H. Esaki, and A. Kato, “The impact and implications of the growth in residential user-to-user traffic”, in *Proc. of ACM SIGCOMM’06*, 2006.

[6] G. Maier, A. Feldmann, V. Paxson, and M. Allman, “On dominant characteristics of residential broadband internet traffic”, in *Proc. of ACM SIGCOMM’09* 2009, pp. 90-102

[7] M. Kihl, P. Odling, C. Lagerstedt, and A. Aurelius, “Traffic analysis and characterization of Internet user behavior”, *Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)* 2010, pp. 224 -231

[8] *A. Riikonen*, “Mobile Internet Usage - Network Traffic Measurements”, Master's Thesis, Department of Communications and Networking, TKK Helsinki University of Technology

[9] R. Pries, F. Wamser, D. Staehle, K. Heck, and P. Tran-Gia, “Traffic measurement and analysis of a broadband wireless internet access”, *IEEE 69th Vehicular Technology Conference*, 2009.

[10] P.Borgnat, G.Dewaele, K.Fukuda, P.Abry, and K.Cho, “Seven Years and One Day: Sketching the Evolution of Internet Traffic”, *IEEE Infocom 2009* proceedings, April 2009, pp.711-719

[11] T. Karagiannis, A. Broido, N. Brownlee, K. Claffy, and M. Faloutsos, “Is P2P dying or just hiding?”, in *Proc. of IEEE Globecom*, 2004.

[12] K. Cho, K. Fukuda, H. Esaki, A. Kato “Observing Slow Crustal Movement in Residential User Traffic”, *ACM CoNEXT*, 2008

[13] H. Schulze, K. Mochalski, “Internet Study 2008/2009”, http://www.ipoque.com

[14] “Global Internet Phenomena Report, Spring 2011”, http://www.sandvine.com

[15] Procera Networks, http://www.proceranetworks.com

[16] J. Sanjuàs-Cuxart, P. Barlet-Ros, J. Solé-Pareta, "Measurement Based Analysis of One-Click File Hosting Services," Journal of Network and Systems Management, 12 May 2011, pp. 1-26

[17] A. Aurelius, C. Lagerstedt, and M. Kihl, “Streaming media over the Internet: Flow based analysis in live access networks”, *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)* 2011, pp.1-6