LUND UNIVERSITY

TRAMMS: Monitoring the evolution of residential broadband Internet traffic

Aurelius, Andreas; Lagerstedt, Christina; Sedano, Inigo; Molnar, Sandor; Kihl, Maria; Mata, Felipe

Published in:
[Host publication title missing]

2010

Link to publication

Citation for published version (APA):
Aurelius, A., Lagerstedt, C., Sedano, I., Molnar, S., Kihl, M., & Mata, F. (2010). TRAMMS: Monitoring the evolution of residential broadband Internet traffic. In *[Host publication title missing]* IEEE - Institute of Electrical and Electronics Engineers Inc..

Total number of authors:
6

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

TRAMMS: Monitoring the evolution of residential broadband Internet traffic

Andreas AURELIUS[1], Christina LAGERSTEDT[1], Iñigo SEDANO[2], Sándor MOLNÁR[3], Maria KIHL[4] and Felipe MATA[5]

[1]*Acreo AB, Electrum 236, 16440 Kista, Sweden*
Tel: +48 8 6327802, Fax: + 46 8 6327710, Email: andreas.aurelius@acreo.se
[2]*TECNALIA-Telecom, Parque Tecnológico, Edificio 202, 48170 Zamudio, Spain*
Tel: +34 946002266, Fax: +34 946002299, Email: isedano@robotiker.es
[3]*High Speed Networks Lab., Dept. of Telecommunications and Media Informatics, Budapest University of Technology and Economics,*
H-1117, Magyar tudósok körútja 2, Budapest, Hungary
Tel: +36 1 463 3889, Fax: +36 1 463 3107, Email: molnar@tmit.bme.hu
[4]*Dept. of Electrical and Information Technology, Lund University, Box 118, 221 00 Lund, Sweden*
Tel: +46 46 2229010, Fax: +46 46 129948, Email: maria.kihl@eit.lth.se
[5]*Networking Research Group, Universidad Autónoma de Madrid, 28049 Madrid, Spain*
Tel: +34 91 4972268 , Fax: +34 91 492235, Email: felipe.mata@uam.es

Abstract: Traffic measurements in broadband access networks are crucial from several points of view. A better understanding of traffic patterns can lead to more efficient network design, which leads to energy and cost savings for the operator and improved end user services. This paper reports on selected traffic measurement results from the Celtic TRAMMS project. The measurements were performed in broadband access networks in Sweden and Spain. The vast majority of the traffic volume is video based and from peer-to-peer applications. This suggests that future access networks should be symmetrical in order to properly cope with the traffic patterns of the future. Geographic locality of end-to-end flows has been identified for incoming and outgoing traffic in Spain, which may be used to analyze peering between ISP's. Finally, we show that stricter legislation in Sweden targeting illegal file sharing led to a dramatic decrease of the traffic. This means that factors not controlled by the networking community may seriously impact traffic patterns and user behaviour thereby indicating the need for closer collaboration between researchers and network designers on one side and politicians and regulators on the other side.

Keywords: Traffic measurements, broadband access networks, content locality, Internet user behavior.

1. Introduction

The Internet as we know it today is a diverse and heterogeneous network that has continuously evolved over the years and which has been shaped by many different actors with fundamentally different agendas. These agendas could be for instance technical, ideological, social, commercial, political, developmental, or even criminal, and all have in combination shaped what we see today. Due to the complexity of the driving forces and in the underlying technologies that constitute the Internet, it is increasingly important to monitor and analyze the traffic in the networks to identify user trends and traffic patterns, to

recognise paradigm shifts, and to basically better understand the mechanisms of the Internet.

The *Traffic Measurements and Models in Multi-Service Networks* (TRAMMS) [1] project is part of the Celtic framework, a EUREKA cluster focusing on telecommunications. It is a three year project with 11 partners from Sweden, Spain and Hungary. The main objective of TRAMMS is to model traffic in multi-service IP networks, and to use the models as input for capacity planning of tomorrow's networks. The models are built upon data acquired with advanced traffic measurements on the application level with deep packet/deep flow inspection in different parts of Europe, combined with bottleneck analysis and inter-domain routing analysis. In the project, Internet traffic is analyzed and characterized, in order to find and establish patterns rather than to find new mathematical modules for simulation use. Brownlee and Claffy argue for the former [2], while Hlavacs et al give an overview of the state of the art models for the latter [3]. Patterns are established by user and/or application behaviour, and can be used to identify services currently exploiting the available network resources and thereby controlling flows or streams on individual basis. Arbor Networks, the University of Michigan and Merit Network have studied global Internet traffic from 110 large cable operators, international transit backbones, regional networks and content providers for a period of two years [4]. Content is moving from a multitude of smaller or medium sized providers to a few, very large host, cloud and content providers. These so called hyper giants generate and consume 30% of all Internet traffic. Facebook, Google, YouTube and Microsoft are found among the hyper giants. Another finding is that P2P traffic carrying video and application distributions is decreasing while a smaller number of web and video protocols are carrying more and more of this type of content. The data sources in the TRAMMS project are not nearly as vast as in the above mentioned study, but we have something that is unique in the research area. We have the *detailed* view of the traffic, which means that we can distinguish from which type of access technology the traffic originates as well as the access speed, and number of households generating the traffic. We will in this paper show that the P2P filesharing traffic still has a significant share of the traffic for most access technologies.

The “Future Internet” is presently subject to major attention globally and regionally. For instance, work is going on in the Future Internet Assembly [5] and several national future Internet initiatives. We believe that in order to discuss and shape the Internet of *tomorrow* it is of paramount importance to understand the traffic patterns and mechanisms of the Internet *today*. This belief drives our research, and this paper reports on some of the findings in this endeavour.

2. Methodology

The authors of [6] showed that experiment durations of approximately 30 days are necessary for the traffic processes to show stationarity. Hence, in order to obtain accurate statistics on traffic characteristics of large internetworks using state-of-the-art measurement techniques, long and spatially diverse experiments may be necessary. The TRAMMS measurements have followed these guidelines, performing the measurements in different networks in Spain and Sweden using different measurement equipments, which are described in the following subsections. The measurement intervals for the TRAMMS results are in the order of months (except for some specific cases), as described in Sections 3 and 4, thus stationarity of the corresponding statistics can be assumed.

Networks

A major focus of TRAMMS and a particular strength is the direct access to measurements in several live broadband access networks. In this paper, we present results from

measurements performed in three networks. In the networks, the measurement equipment is installed near the end-users in order to ensure a high level of detail for the analysis. The first network is a *Swedish commercial network*. This network is a municipal network based on the open access model with approximately 2600 FTTH and 200 DSL residential users. The FTTH customers represent many social and ethnic groups, while the DSL customers constitute a more homogeneous group of Swedish middle class living in single family houses. The second network is a *Spanish commercial network*, which contains both fixed and wireless access with residential and enterprise users. The wireline part consists of a fibre to the cabinet (FTTC) network and the last mile consists of Cable Modem Termination System (CMTS). Approximately 10.000 users are connected to the wireline part. The wireless access is a combined GPRS and UMTS system with 250.000 potential users. The wireline network is mainly residential while the wireless network carries mainly enterprise traffic. The third network is a *Spanish university network*, which provides Internet access to more than 300 institutions with more than 1 million users. The network is SDH-based with link speeds from 2.5 Gbps up to 10 Gbps.

Measurements

The measurements in the Swedish and Spanish commercial networks were performed using a commercial traffic management device, PacketLogic (PL) [7]. Traffic is identified based on packet content and flow behaviour using deep packet and deep flow inspection. PL uses the proprietary Datastream Recognition Definition Language (DRDL) [8] to identify application protocols. The PL is able to identify more than 1000 application protocols. In other words, PL accurately determines the applications used over the Internet.

The measurements in the Spanish university network were performed with Cisco NetFlow [9], a network protocol developed by Cisco Systems to run on their routers, implemented by other vendors as well. It monitors the traffic that traverses a router and to keep statistics of the performance by sampling some of the packets. Cisco defines a flow as a unidirectional sequence of packets sharing the flow descriptors commonly referred to as 7-tuple: Source and Destination IP addresses, IP protocol, Source and Destination ports (when the IP protocol is TCP or UDP), Ingress interface and IP Type of Service.

3. Popular Applications

The future networks are expected to have a higher degree of content awareness and to better accommodate to user needs in terms of e.g. quality of experience. A first step in this endeavour is to analyze the applications used on the current Internet and characterize the user behaviour.

In Fig. 1, the ratio of the total traffic volume for the most popular application groups are shown for the Spanish Cable modem (CMTS) and mobile (GGSN) customers during peak hours. The reason for only showing the traffic mix for the peak traffic is that this is the dimensioning traffic for the network operators, and thus it is a much more vital metric than the overall traffic mix. Also, we have found that the traffic mix during the peak is similar to that of the total traffic. It can be seen that, in the CMTS case, the P2P file sharing traffic is dominating during the peak hours. The change from 2008 to 2009 is that the relative share of P2P traffic is lower, yielding for increased web and streaming traffic. One explanation for the drop in the P2P share of total traffic volume could be the large increase in Direct Download Link (DDL) based services, like Megaupload or RapidShare, that have made users move away from P2P based services. Further, the traffic volume during peak hours has not significantly increased from 2008 to 2009 (it is the same customer base). This is, however, not true for the mobile broadband users, where the traffic volume has roughly doubled. This is due to more users actively using mobile broadband and also more traffic

per user. In this case, the application group generating most traffic during peak is web traffic, even if P2P traffic has increased significantly from 2008 to 2009.

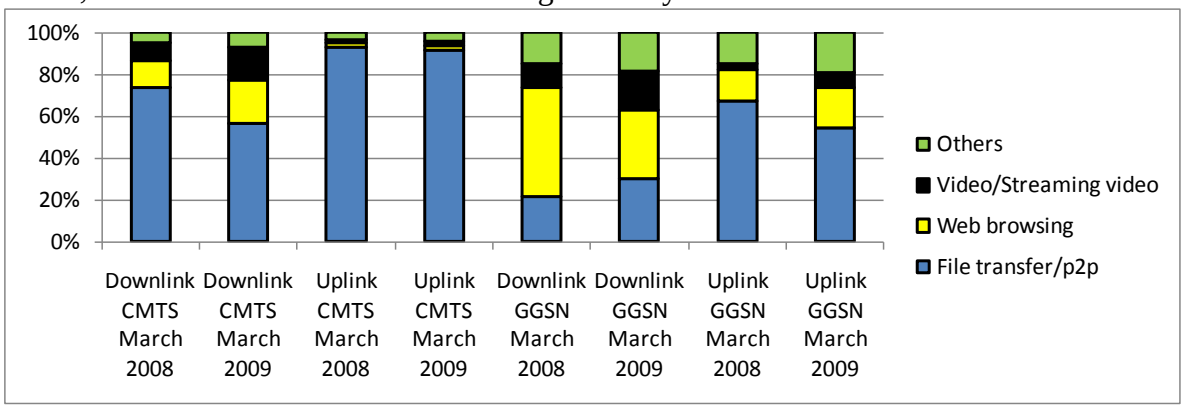

Fig. 1 Relative volume share of application groups in the Cable Modem (CMTS) and mobile (GGSN) Spanish network for March 2008 and March 2009 (23 days, "others" include the rest of the traffic excluding the unknown) considering only the uplink and downlink peak hours

There are many factors that affect the traffic pattern, and one example is policy decisions and law making. In Sweden, the national implementation of the European Union's Intellectual Property Rights Enforcement Directive (IPRED) [10] took effect on the 1st of April 2009. This law is aimed at reducing illegal sharing of content. Another law that may have effects on Internet user behaviour is the Swedish FRA law, a Swedish legislative package that authorizes the state to warrantlessly wiretap all telephone and Internet traffic that crosses Sweden's borders (Swedish proposition 2006/07:63). This law took effect on the 1st of January 2009. Fig. 2 shows the daily traffic volume in the Swedish commercial network from end of March 2009 until end of September 2009. As can be seen, the traffic volumes decreased dramatically after the enforcement of IPRED, with an immediate decrease in traffic volume of about 50 %. This phenomenon has also been reported in [11] [12]. The long term effects of the law are still unknown. One possible effect is that the use of encryption of data will be even more wide spread.

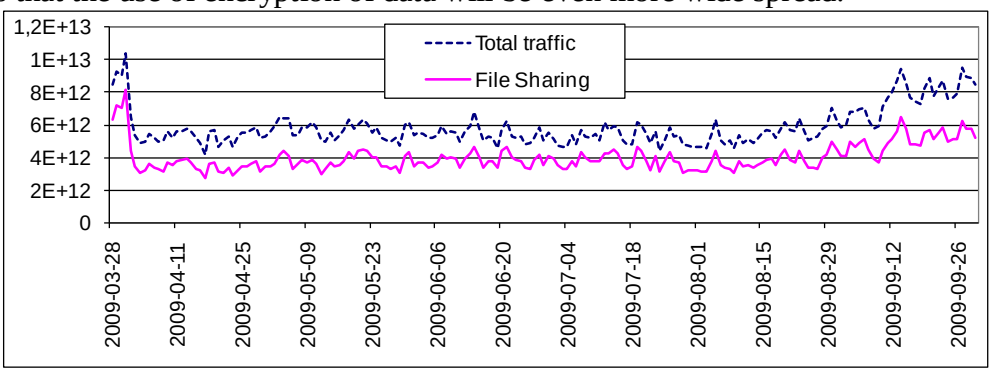

Fig. 2 Measurements of total traffic and P2P file sharing traffic in a segment of the Swedish municipal network (FTTH). The number of households in the measurement is approximately 1500, of which roughly 50% are using Bit Torrent.

In Fig. 3, the relative share of encrypted data from some P2P file sharing protocols is shown. It is obvious that the relative share of encrypted traffic has increased in all 4 cases, except for the DSL (the customer base is quite small in the DSL case, so this number may not be representative). Another effect that may have been caused by enforcement of the IPRED directive and the FRA law is the increased number of users that are using anonymization services. Measurements in the Swedish commercial network show that the number of IP addresses using the tunnelling protocol PPTP has increased, from virtually none in the beginning of 2009 to around 10 % in the end of 2009. If this trend continues, there may be dramatic changes in traffic patterns on the Internet, since traffic is forced to

the service provider that delivers the anonymization service, causing extra router hops, changing paths in the network, possibly causing new bottlenecks and increasing complexity.

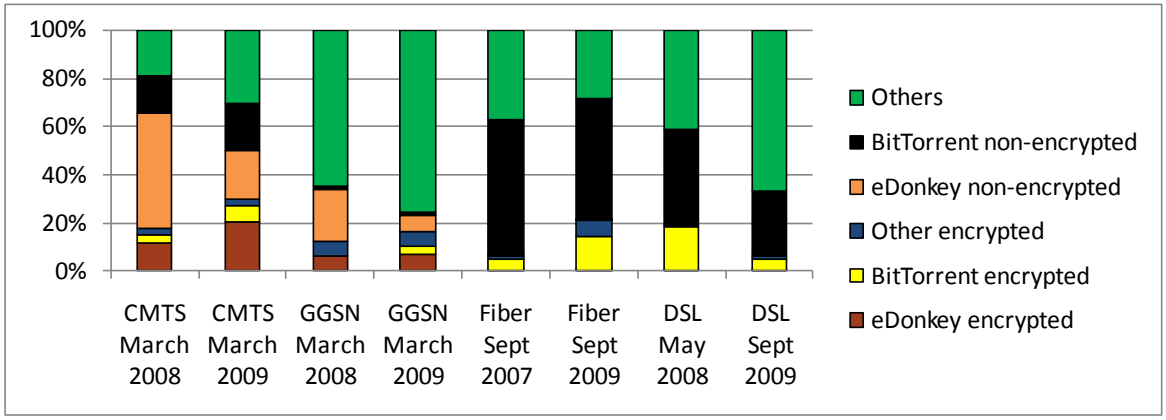

Fig. 3 Relative share of encrypted data in the CMTS and GGSN Spanish networks March 2008 and 2009 (23 days), the Swedish Fiber network Sept 2007 and 2009 and finally the Swedish DSL network May 2008 and Sept 2009. "Others" include the rest of the traffic excluding the unknown.

4. Content Locality

For Internet service providers (ISPs), traffic exchange has a direct impact on business. One option to reduce the transit costs for an ISP is by means of peering agreements. In a peering agreement, two ISPs reciprocally provide connectivity between each other's users [13]. The reduction in transit costs is obtained by reducing transit traffic, routing it directly to the destination ISP. Peering agreements also offer a reduction in the path inflation, i.e. routing paths longer than needed, thus lowering the latency in the traffic [14][15].

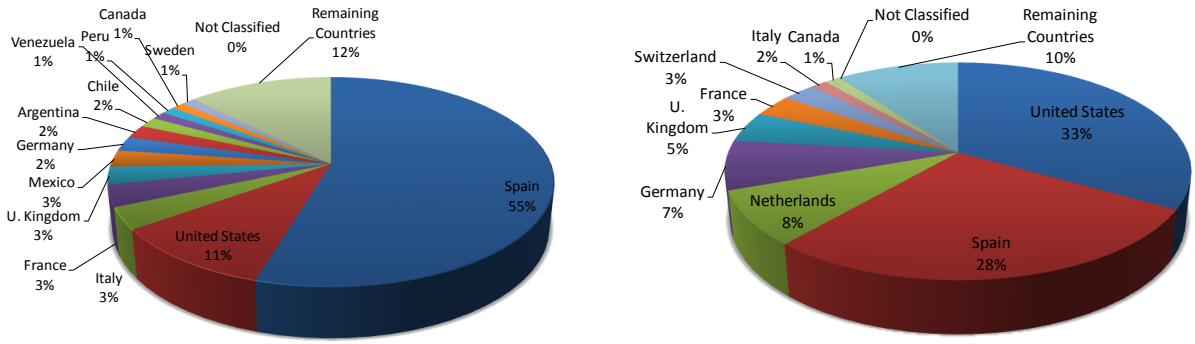

Fig. 4 Most contributing countries in the Spanish university network when measuring bytes. ***Left****: outgoing direction;* ***Right****: Incoming direction.*

A better understanding of the geographical location of end-to-end flows can justify the establishment of peering agreements between ISPs [16]. As ISPs tend to cover certain geographic regions (usually countries), a geographic study of end-to-end flow connections is crucial when analyzing the opportunities of establishing a peering agreement ([17], [18]). Fig. 4 shows a content locality analysis from the Spanish university network. IP addresses were mapped to their corresponding countries using MaxMind [19], a public database giving the geographic localization of IP addresses with an accuracy of 99.5% (that improves IP block based geolocation [20]). The analysis was performed for both incoming traffic, i.e. traffic from the Internet to campus, and outgoing traffic, i.e. the traffic from campus to the Internet. The results show that the majority of the traffic is sent to and received from Spain and the United States. The same analysis has been performed for flows and packets, but the results are quite similar as for bytes and are not included in this paper.

Also, there is an asymmetry in the traffic directions. The asymmetry to different countries may be due to different applications that are used in each location. One plausible explanation can be that the traffic from the US is dominated by streaming traffic and downloads. Therefore, one interesting study would be to analyze in greater detail the level of asymmetry between ISPs and possible ways to mitigate this asymmetry, by i.e. local caching of content. The asymmetry exhibited for some countries in Fig. 4 has to be taken into account when analyzing peering agreements. However, there are countries (such as the United Kingdom or France) whose traffic exhibits a higher level of symmetry. Although the percentage of traffic that is destined to and from these countries is smaller, it could be more profitable to establish peering agreements with ISPs in those countries than with ISPs in the former ones. Further analysis is needed in order to quantify the benefits of peering.

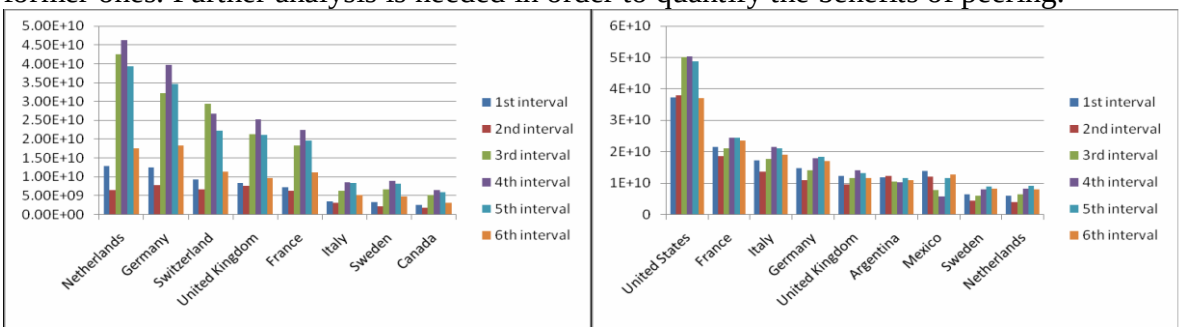

*Fig. 5 Day-night traffic pattern for the ten most contributing countries (**Left**: sources; **Right**: destinations). **Left**: after removing Spain and the United States; **Right**: after removing Spain*

Fig. 5 shows the same content analysis when splitting the day into 6 non-overlapping intervals of 4 hours length starting from midnight, and accounting the same parameters only for the traffic generated during the corresponding interval. The results show a clear day-night pattern for the outgoing traffic. We observe similar traffic shares for those countries in the different time frames when compared to those of the whole day. These results suggest that the network traffic locality has a fractal nature (i.e. the distribution is nearly the same independently of the time instant and of the hours of aggregation). However, for the outgoing direction we do not find a clear pattern. During night hours, Spain is the main source of traffic followed closely by the United States. During day hours the United States is the main source of traffic. These results suggest an interaction between the day-night patterns of the countries that send traffic to Spain that superimposes with time zones of the countries and results in a not clearly distinguishable pattern.

5. RTT measurements

Internet delays have significant impact on user perceived quality of numerous applications. For example, the web experience of users is mainly determined by the webpage download times. Also, online gaming applications have strict requirements on delays. Furthermore, popular P2P applications are also sensitive to the underlying delay characteristics of the Internet. In case of these P2P applications, geographical locations can also significantly influence the delay-sensitive system performance [21]. In the area of mobile Internet the delay is an even more crucial characteristic of the perceived quality of the applications [22]. We have carried out a Round-Trip- Time (RTT) analysis study based on measurements in the Spanish commercial network in order to reveal these characteristics. Here, RTT is defined as the time required for a packet to travel from a source to a destination and back. In this analysis, the method of measuring RTT is via the TCP SYN and SYN-ACK packets. Every packet has a timestamp and the RTT is calculated by the time difference between a certain SYN packet and the corresponding SYN-ACK packet.

Fig. 6 shows the average RTT values with standard deviations from Spain to different countries in the wired part of the network (CMTS). The RTT values are between 30ms and

100ms measured between Spain and most of the European countries. We can see that the RTT is not higher than 100ms by reaching USA due to the high speed transoceanic links between Europe and USA. The highest RTTs are found in between Spain and Asian or South American countries having values around 400ms.

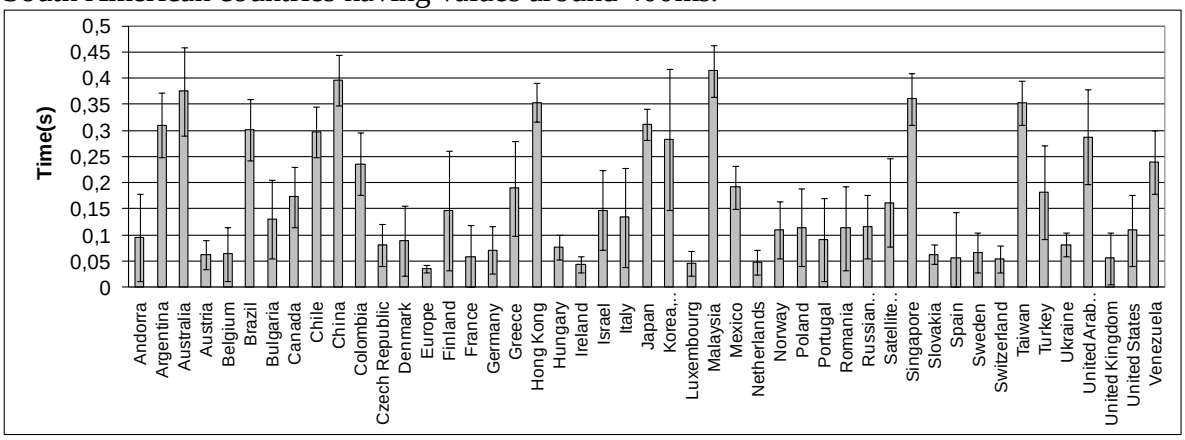

Fig. 6 Average RTT and standard deviation to different countries from Spain (CMTS).

Performing the same analysis in the wireless part of the network, we note that the values are 4-6 times higher compared to the CMTS RTT values. In some extreme cases, RTT can even be larger than 1sec. However, it is interesting to note that the average RTTs have smaller deviations regarding geographical locations and almost all RTTs are between 400ms and 600ms. This may suggest that if this range is acceptable for mobile applications, then geographical locations have only minor influence on delay-related performance.

6. Conclusions and discussion

In this paper, we have presented some results from the CELTIC TRAMMS project. The results come from measurements in both Spanish and Swedish access networks with real users. We have shown the relative usage of popular applications, as well as analysis of content locality and round-trip times.

We observe that the P2P paradigm is being used more and more as a content delivery mechanism also for legal content. There are several examples of this, for example, the Norwegian broadcasting company NRK that has set up its own Bit Torrent tracker [23], and the P2P based streaming services Spotify, PPlive, and TVUPlayer. Thus, we envisage that there will be an increasing traffic volume due to *legal* P2P applications. In particular, we believe that more content providers will follow the footsteps of NRK and give access to their material through P2P applications in order to gain control of their contents, while at the same time considerably offloading their servers. If the majority of video content (which is dominating the network traffic) moves to P2P distribution, it will lead to a higher demand for symmetrical access connections, which has also been discussed in [24]. Other services that benefit from symmetric connections are e.g. health care in the home, where information needs to be transmitted from the home to the hospital; or video conferencing.

Also, it is clear that laws like IPRED and other policy decisions or regulations may have tremendous effects on the behaviour and traffic patterns and consequently also on the networks in the future. In this paper, we have shown two effects from the IPRED, namely the decrease in P2P filesharing traffic volume and the increased usage of PPTP, which is commonly used by anonymization services. We therefore believe that there is a strong need for researchers, network designers and policy makers to have a closer cooperation in analyzing the consequences of decisions that are under consideration. Further, the business models should be taken into account when analyzing the user behaviour regarding

applications used. This will have a great impact on the uptake and usage of services, and is thus an important piece of the complete picture.

One of the major outcomes of the measurement campaign in the TRAMMS project may be the framework and the methods for measurements and analysis that have been developed. Long term measurements are performed in access networks, for benefit of research and development. Since the measurements are long term and close to the end users, it is possible to identify changes in user and protocol behaviour. It is important to note here that application layer data is available, per household, which is a prerequisite for drawing conclusions from the data. Thus the major impact of the reported measurements is still to be explored, and apart from traffic modelling, we envisage this type of methodology to be beneficial also for analysis of techno economics, business models and detailed user behaviour. In this endeavour we also plan to perform measurements of user behaviour in relation to the broadband subscription type.

7. Acknowledgements

The authors would like to acknowledge the Swedish Agency for Innovation Systems, VINNOVA for funding the Swedish part of the TRAMMS project, and the Spanish Ministry of Industry, Tourism and Trade for funding the Spanish part of the project. Moreover, parts of this work were performed within the VINNOVA project Broadband Behaviour, as well as the VINNOVA VINNMER program. We would also like to acknowledge the support of the DIOR project that provided the RedIRIS measurements.

8. References

[1] Traffic measurements and models in multi-service networks, *http://projects.celtic-initiative.org/tramms/*
[2] Brownlee, Nevil, and KC Claffy. "Internet Measurement." *IEEE Internet Computing* (IEEE Computer Society), 2004.
[3] Hlavacs, Helmut, Gabriele Kotsis, and Christine Steinkellner. *Traffic Source Modeling.* Technical Report, Institute of Applied Computer Science and Information Systems, University of Vienna, Vienna: Institute of Applied Computer Science and Information Systems, University of Vienna, 1999.
[4] *Two-Year Study of Global Internet Traffic , Presented At NANOG47.* http://www.arbornetworks.com/en/arbor-networks-the-university-of-michigan-and-merit-network-to-present-two-year-study-of-global-int-2.html (accessed 10 20, 2009).
[5] *http://www.future-internet.eu/*
[6] J. L. García-Dorado, J. A. Hernández, J. Aracil, J. E. López de Vergara, F. J. Monserrat, E. Robles, and T. P. de Miguel. "On the Duration and Spatial Characteristics of Internet Traffic Measurement Experiments." *IEEE Communications Magazine*, 2008.
[7] Procera networks, http://www.proceranetworks.com.
[8] PacketLogic drdl signatures and properties 10340 (beta), *tech. rep.*, 2007.
[9] Cisco NetFlow. http://www.cisco.com/go/netflow.
[10] European Union directive*: Directive 2004/48/EC*
[11] http://stats.autonomica.se/mrtg/sums/All.html
[12] http://www.dn.se/kultur-noje/illegal-fildelning-minskar-stort-1.865430 (in Swedish)
[13] W. B. Norton, Internet service providers and peering, 2001. *http://www.equinix.com*.
[14] W. B. Norton, A business case for ISP peering, 2001. *http://www.equinix.com*
[15] N. Spring, R. Mahajan, and T. Anderson, *The causes of path inflation.* In Proceedings of the ACM SIGCOMM Conference on Applications, technologies, architectures, and protocols for computer communications, pages 113{124, Karlsruhe (Germany), Aug. 2003
[16] S. Shakkottai and R. Srikant. Economics of network pricing with multiple ISPs. *IEEE/ACM Trans. Netw.*, 14(6):1233-1245, Dec. 2006.
[17] W.-C. Feng, F. Chang, W. chi Feng, and J. Walpole. A traffic characterization of popular on-line games. *IEEE/ACM Trans. Netw.*, 13(3): 488-500, June 2005.
[18] A. Lakhina, J. Byers, M. Crovella, and I. Matta. On the geographic location of internet resources. *IEEE J. Sel. Areas Commun.*, 21(6):934-948, Aug. 2003.
[19] Geolocation and Online Fraud Prevention from MaxMind, *http://www.maxmind.com/app/geolitecountry*.
[20] B. Gueye, S. Uhlig, and S. Fdida. Investigating the imprecision of IP block-based geolocation. In *Proceedings of the ACM SIGCOMM conference on Internet measurement*, pages 237-240, Louvain-la-neuve (Belgium), 2007.
[21] Kaune, S., Pussep, K., Leng, C., Kovacevic, A., Tyson, G., and Steinmetz, R., *Modelling the Internet Delay Space Based on Geographical Locations*. In Proceedings of the 17th Euromicro International Conference on Parallel, Distributed and Network-Based Processing, February 18-20, 2009.
[22] Romirer-Maierhofer, P., Ricciato, F., D'Alconzo, A., Franzan, R., and Karner, W., *Network-Wide Measurements of TCP RTT in 3G*. In Proceedings of the First international Workshop on Traffic Monitoring and Analysis (Aachen, Germany, May 11 - 11, 2009.
[23] Web article posted at 8 March 2009, *http://nrkbeta.no/norwegian-broadcasting-corporation-sets-up-its-own-bittorrent-tracker/*
[24] M. Forzati and C.P. Larsen, "On the symmetry requirements for tomorrow's fibre access networks", *ICTON 2009*, The Azores, Portugal, July 2009