Traffic modeling in FMC network scenarios

Executive Summary of the Deliverable

The COMBO project will propose and investigate new integrated approaches for Fixed Mobile Convergence (FMC) for broadband access and aggregation networks. COMBO will target on an optimal and seamless quality of experience for the end user together with an optimized network infrastructure ensuring increased performance, reduced cost and reduced energy consumption.

COMBO WP2 will provide the preliminary work for fixed and mobile networks analysing their current status and evolution trends. This deliverable (D2.3) of Task 2.3 is called *Traffic modelling in FMC network scenarios*.

The change of traffic patterns at FMC network scenarios in comparison with fixed and mobile networks without considering FMC architectures can be seen from two different perspectives. On one hand, the content-specific characteristics of the traffic – and on the other hand, the way how traffic flows. The volume and the various traffic-descriptor statistics will change, including the application mixture changes in some links/connections. Beside this, the way how traffic flows will also change. This fact is due to the evolution of FMC architectures, for example, with the introduction of offloading scenarios. In short, some traffic will take different routes instead of current ones.

This document summarizes the reasons of these changes, and makes an effort of predicting how the traffic-related changes will take effect in the future - especially in the relation of specific FMC scenarios. It starts with a methodology definition and a brief description of the current traffic situation in the different segments of the network.

The chapter on the analysis of current traffic demands and forecast provides insight and comparative data for WP5 (Techno-economic assessment), as well as for Task 4.2 (Traffic and Performance Management).

The chapter on traffic analysis and modelling provides a general overview of analysis and modelling methodology, and publicly available reference data. The available modelling methodologies have an impact on architectural issues, hence it affects the work in WP3 (Fixed Mobile Convergent architectures). This chapter provides input on some of the traffic-related discussions at Task 2.4 (Performance Management), as well.

Beside this, the analysis results published here - as well as the procedures that invoked data capture and analysis activities at various partners' networks - can be used as reference points at some WP6 activities (Functional Development & Experimental Research Activities).

The discussion on offloading scenarios and the FMC traffic model for aggregation networks support the work of WP3, WP4 and WP5 activities by providing traffic models for some of the scenarios referenced at these work packages.

Table of Contents

1 Introduction

This document provides the basis of traffic scenarios and modelling for fixed and mobile convergent networks starting with the study of the state of the art and the current evolution trends of both networks from a non FMC point of view. It also includes an analysis of the future evolutions paths over time considering a 2020 time horizon in order to establish one roadmap for fixed and one roadmap for mobile networks.

In the first technical chapter after the introduction, the document describes the overall methodology of gathering traffic-related data and attempting to set up models based on them. Furthermore, it aims to find some common understanding on what to analyse at, and how to model traffic scenarios.

Traffic evolution should be analysed based on earlier experiences and the overview of current status. After discussing current traffic demands, and various drivers for traffic growth - such as the increasing number of devices, the introduction of new services and applications, and the effect of these factors on bandwidth demands - a forecast on traffic trends can be provided. The conclusions of these predictions include a 3.5x busy-hour Internet traffic increase from 2012 to 2017 (865 Tbps), and a 1000x traffic increase in a 20 years span. A brief overview of Content Delivery Networks - as an important optimization answer of some traffic routing issues -, and signalling changes are also part of this chapter of the document.

In order to provide traffic models of current and future scenarios - including FMC -, traffic analysis of current patterns should be covered. Although the methods - and hence, the toolset - of modelling and analysis differ, they should be handled together, since they provide input requirements and feedback to each other. The traffic analysis and modelling section of this document covers methodology discussion and actual analysis results as well. These include analysis of publicly available traffic archives, as well as analysis of aggregated traffic patterns and bit-by-bit captures. In the modelling part the document covers packet-level arrival processes and long-range dependency, as well as modelling on macro-level (traffic mixtures, user behavioural patterns, etc.).

The last chapter of the document describes how traffic patterns and models are affected by some of the FMC scenarios. One of the models here is described for offloading scenarios with WLAN femtocell, and the other model describes traffic composition, and the expected evolution within aggregation networks.

2 DEFINITION OF TRAFFIC SCENARIOS AND MODELS

2.1 Motivation and Methodology

This section briefly provides the main motivation behind the need of defining traffic scenarios, and modelling traffic. Furthermore, since partners have different understanding of what a traffic scenario or model may mean, we describe the agreed, simple method of consolidating these terms.

Since partners came from various organizations, there is a need for understanding what they – especially operators – mean by “traffic scenario”, and to identify what is the common understanding of differentiating traffic types. This differentiation can be based on many factors, i.e. what kind of applications provide such a traffic; statistical properties; QoS demands; network type and network segment that carries such a traffic, etc.

Partners agreed on following a certain methodology for reaching an agreement on the above questions. This included steps such as:

- operators providing input, answering with their view of these definitions,
- consolidation of different naming conventions,
- definition of various traffic scenarios.

The first two steps were carried out within task 2.3, whereas the third step was a combined result of discussions within task 2.1 and two use cases of deliverable D2.1 “Framework reference for fixed and mobile convergence”. The use cases UC06 “Common fixed and mobile access termination in hybrid connectivity for FMI customer services” – and UC08 “Universal Access Gateway (UAG) for fixed and mobile aggregation network” were chosen to be part of the analysis within D2.3, due to their effect on changing traffic mix or changing traffic matrix.

2.2 Summary of views on traffic scenario and modelling

Based on the operators’ feedback it seems that providing an overall, general description of FMC traffic mixtures is not an expectation. Rather than that, various scenarios are to be considered, analysed and modelled from traffic evolution point of view. The defining parameters of these traffic scenarios can widely vary depending on the network domain, the traversing architecture (node and connection types), user behaviour, and many more.

This propagates that the traffic modelling used for these scenarios can and should consider various levels of describing traffic. Depending on the underlying motivation, the traffic model can be defined on the level of packet-interarrival times, flow dynamics and other statistical distributions of packet- and flow-level parameters – and on the other extreme it can be defined based on user behaviour descriptions, architectural constraints and low-granularity descriptors of traffic volumes traversed across network segments.

The following sections provide a short summary of answers given by the operators to interview questions about sharing views on traffic scenarios and expectations on traffic models.

2.3 Where that traffic flows: differentiation by traffic paths

This section describes a path in different segment of network for popular applications currently used in fixed and mobile network. Firstly we choose the application most used and/or that generates most of the data traffic (like: Internet data traffic, traffic controlled by operator) based on real traffic data. Secondly we will show a qualitative description of traffic volume in different segments of fixed network. Thirdly we describe a qualitative traffic volume in different segments of Mobile network; finally we indicate a qualitative description of traffic volume in the case of a Fixed-Mobile Convergent network based on use cases 6 and 8.

2.3.1 The choice of applications

Internet Traffic

For Internet traffic, we only consider applications that generate most of traffic in upstream and downstream. These applications are selected from section 3.1 coming from real traffic demand provided by FT and TID.

In downstream, we consider applications that generate most of total traffic and that represent minimum of 15% volume in different segments of the network. These applications are usually video as stated in section 3.1: Streaming Video and Download.

In upstream, Internet usage is strongly related to applications for file sharing in peer-to-peer mode (P2P). These files can be a video or other content (e.g., images, music). Peer-to-peer applications contribute majorly in terms of upstream volume in different segments of the network.

Traffic controlled and provided by operator

For this type of traffic we chose the applications that are frequently used but they can contribute little in load of different segments of the network compared to the Internet traffic. The most important is to know the routing of these prioritized applications in different segment of the network. These applications are: TV Multicast, Video streaming (IPTV unicast) and VoIP.

2.3.2 Qualitative description of traffic volume in different segments of fixed Network

Downstream traffic

IPTV services are routed using multicast, the rate of multicast traffic decreases when this traffic approaches to the final customers (e.g. core → aggregation → access → customers).

IPTV multicast is transported from the head of TV (see Figure 1) via the core and the aggregation networks, arriving at the DSLAM or OLT which will distribute the necessary channels to all clients attached to the tree Passive Optical Network (PON) or DSLAM.

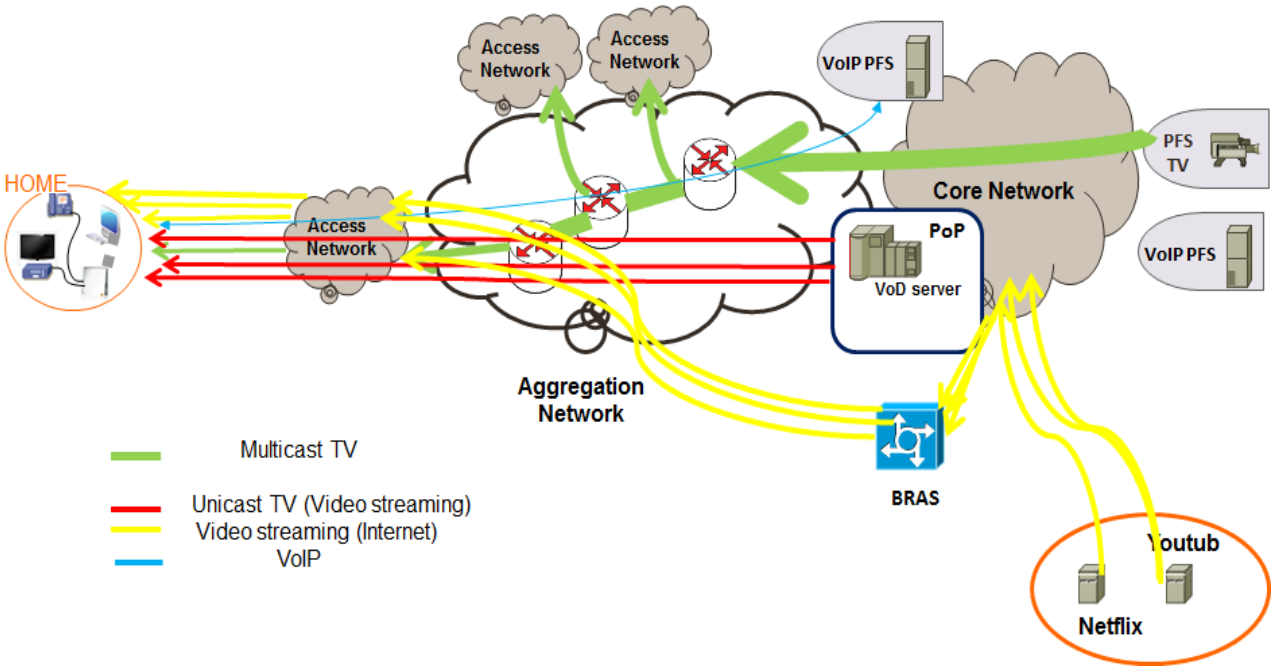


Figure 1: Architecture of Fixed Network.

The traffic generated by the multicast TV is modest and stable in the time. The contribution of multicast video in the network is not important (in terms of volume).

In other video services (unicast traffic) such as streaming video controlled by operator, This traffic is routed from the PoPs (Point of Presence) (platform of video) via the aggregation and access networks (DSLAM/OLT) arriving at the final customers.

Unicast traffic is contrary to multicast traffic. The multicast traffic optimizes bandwidth, by broadcasting one flow for a set of customers but in unicast traffic, the video content required is multiplied by the number of customers requesting this content.

The unicast video managed by the operator represents a little contribution in terms of load in different segment of the network (core / aggregation / access).

The VoIP is a service that currently contributes poorly on the network in terms of load seen its low consumption of network resources.

VoIP is routed from the source to destination client via the aggregation and access Network through a VoIP service platform.

Currently the Internet services (streaming, download) represent the majority of traffic (>50%) and charging increasingly a network in different segments.

The video content (e.g., YouTube, daily motion, Akamai, Netflix) requested by users, they traverse the core Network and BRAS then sent to the end users via the aggregation and access networks.

The following table summarizes the qualitative load in different segments of the fixed network of different applications.

Downstream	Core Network	Aggregation NW	Access Network
TV multicast	high	low	very low
Unicast Video managed by operator	high	high	high
VoIP	very low	very low	very low
Video (Internet Data)	very high	very high	very high

Table 1: Qualitative load in different segment of Fixed Network.

Upstream traffic

In upstream direction, we observe that P2P generates the majority of traffic, therefore we will only consider the routing of P2P in different segments of the network (core/ aggregation/access).

P2P is an exchange of content between two end-users. Datagrams are routed from the access network across subsequently the aggregation, BRAS and finally the core Network.

The following table summarizes the qualitative load in different segments of the network of the application P2P in the case of Fixed Network.

Upstream	Core Network	Aggregation NW	Access Network
P2P	very high	very high	very high

Table 2: Qualitative load in different segment of the network

2.3.3 Qualitative description of traffic volume in different segments of Mobile Networks

Downstream

In this section, we chose the same applications that are used in the fixed network based on data provided by COMBO partners TID and FT (see section 3.1). It is observed that in fixed and mobile network, the customers use the same kind of applications.

For VoIP, we took the example in the case of a call via Skype, but there are other solutions for VoIP whether via 3G (circuit mode) or via IMS.

The path of VoIP starts from the Internet service platform (e.g. Skype, Viber) through the core of the mobile network (PDN-GW and S-GW) then the access network, arriving at the final customer. This path is valid in both directions of transmission (downlink and uplink).

The VoIP is more and more used by the majority of users, but it currently represents a small contribution to the total load of the network (access and core network).

Video streaming in unicast mode (Internet, e.g. YouTube) represents the application that generates most traffic in downstream and contributes strongly in the load of different segments of the mobile network. Video streaming in unicast is sent from video head-end

platform while crossing the mobile core network (PDN-GW and S-GW) and the access network (E-node), then the content is delivered to the end user device. Unicast video managed by the operator currently represents a small contribution in terms of load in mobile network and most users prefer to see the video at home with a classical TV, although this situation is changing lately.

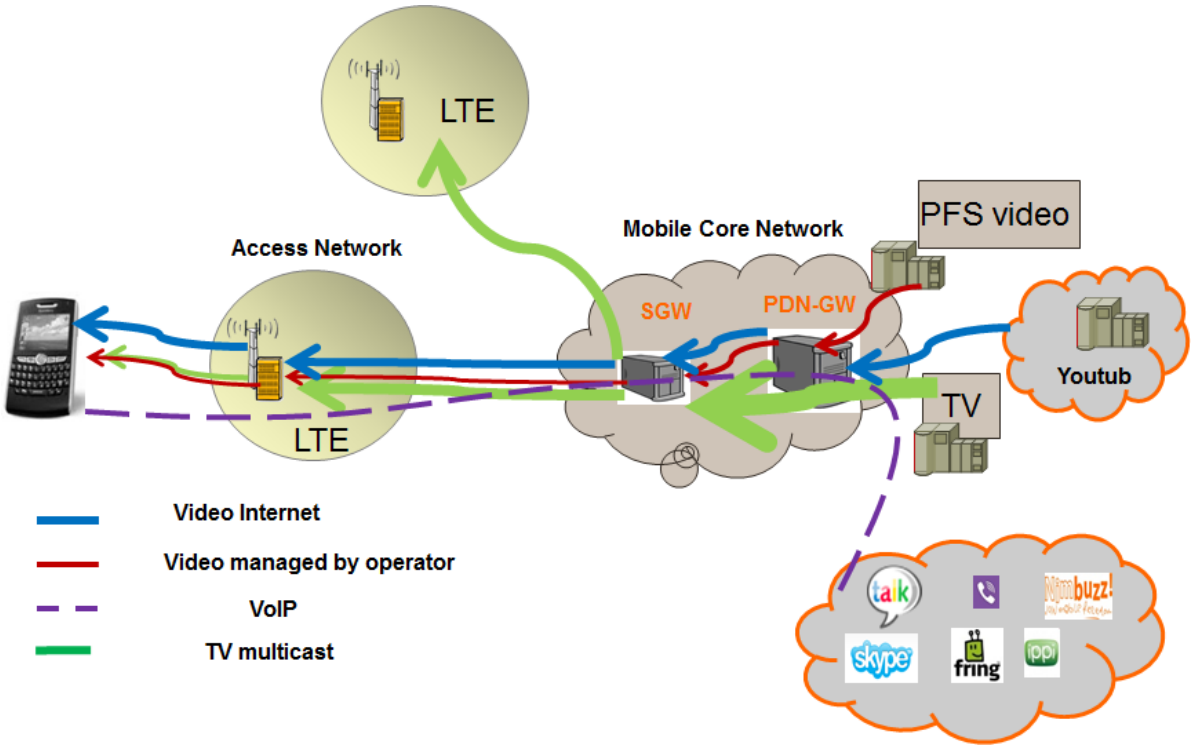


Figure 2: Architecture of Mobile Network (LTE).

Table 3 summarizes the qualitative load in different segments of the Networks of different services / applications in the case of Mobile Network.

Downstream	Core Network	Aggregation Network	Access Network
Unicast Video managed by operator	high	high	high
VoIP	very low	very low	very low
Video (Internet Data)	very high	very high	very high

Table 3: Qualitative load in different segment of Mobile Network.

Upstream

In upstream we find that web browsing is always the dominant application in terms of volume generated by all customers (Web represents 59% of total volume generated by

Orange mobile customers and 47% of total volume generated by Telefónica mobile customers). But the total volume generated in upstream is very little compared to down-stream volume.

2.3.4 Qualitative description of traffic volume in different segment of FMC

Taking the same applications that are used in upstream and downstream of fixed and mobile networks (as discussed in previous sections), we now consider the path of each application in the upstream and downstream and in what extent changes may occur in different segments of an FMC network.

TV multicast is routed from the head of TV to the end customer via the IP backbone then the aggregation and finally the converged access (fixed / mobile) networks.

Generally, IPTV multicast traffic is viewed through a fixed access. The contribution of TV multicast traffic remains modest and stable over time.

VoIP is an application that requires little resource (flow) and is sent from the IP backbone through the aggregation and access network converged (in the use case 8) and/or to fixed (e.g. Wi-Fi for Skype) or mobile access networks (in the use case 6).

Streaming video managed by the operator is routed from the head-end, via the mobile IP edge and then the aggregation converged network of the mobile access network.

Video streaming in unicast (Internet) represents the application that generates most of the traffic in downstream and contributes strongly in the charge (%)

Video streaming in unicast (Internet) represents the application that generates most of the traffic in downstream and contributes strongly in the charge of different segments of FMC networks. Video streaming unicast (Internet) is sent from video servers (or CDNs) while crossing the core and access/aggregation converged network.

2.3.5 What is expected to change in Fixed Mobile Convergence situation?

The load in the different segments of the network is completely different. We can imagine that the load must be important in the converged aggregation network (use-case 6) and reduced in fixed and mobile access networks which are separated. Whereas in the case of use case 8, we can have a very significant traffic load in the converged access and aggregation network which represents a single entity.

At busy hours we can imagine that even home users who prefer to connect their devices in Wi-Fi [1] or fixed access can also use their mobile access to download streaming movies or file sharing P2P, whenever the Wi-Fi connection and fixed access become overloaded.

The following table summarizes the total load in different segments of FMC networks.

Total traffic Load	Fixed Access Network	Mobile Access Network	Converged aggregation Network	IP Backbone
Use case 6	Very high	Very high	Very high	Very high

Table 4: Qualitative load in different segment of FMC Network (use case 6)

Traffic Load	Converged Access & aggregation Network	IP Backbone
Use case 8	Very high	Very high

Table 5: Qualitative load in different segment of FMC Network (use case 8)

3 ANALYSIS OF CURRENT TRAFFIC DEMANDS AND FORECAST

3.1 Current Traffic Demands

FT carried out a traffic analysis in their network in the past years. The measurements were carried out inside the core of dense mobile and dense fixed ADSL and FTTH (Fibre To The Home) networks. Some highlights of the report are selected in the following. Applications generating the most traffic in 2012 in FTTH and ADSL networks:

- Upstream FTTH:
  - 64% of average volume is generated by P2P
  - 6% of applications generate 64% of total traffic
- Upstream ADSL:
  - 50% of total volume is generated by P2P
  - 6% of applications generate 50% of traffic
- Downstream FTTH:
  - Out of the total volume –
    - 36% is generated by streaming
    - 17% is generated by download
    - 16% is generated by web browsing
    - 16% is generated by P2P
  - 27% of applications generate 86% of traffic
- Downstream ADSL:
  - Out of the total volume –
    - 44% is generated by streaming video
    - 18% is generated by download
    - 18% is generated by web browsing
    - 12% is generated by P2P

- 27% of applications generate 92% of traffic

Observations:

- For both FTTH and ADSL, downstream traffic was dominated by streaming, upstream traffic by P2P.
  - Streaming among others may be video, audio or online gaming
  - Online gaming generates symmetric traffic on the up- and downlink.
  - Users on average *consume* more data than they *produce*
  - Upstream traffic is mainly *redistributing* consumed content rather than creating new content
- Comparing FTTH and ADSL, the traffic volume ratios among the top applications were approximately the same (streaming : download : web : P2P).
  - Bandwidth does not seem to affect the top applications used by the customers?
  - Even ADSL seems to be sufficient for the current top used

Profiling FTTH and ADSL traffic distributed by hours of the day, for 5 000 customers each:

- For FTTH, both upstream and downstream show the same profile: Saw tooth profile, with minimum at around 6:00h, maximum at around 23:00h, transition is linear.
- For ADSL, both upstream and downstream show approximately the same minimum at maximums at 6:00h and 23:00 hours, respectively. However, the profile is less linear – peak hours already staring at around 12:00h and going into the evening. Traffic is more differentiated as according to the day of the week.
  - Ratio between the minimum and maximum traffic volume is around 1:6 for upstream, 1:3.5 for upstream – for both FTTH and ADSL.
  - Traffic volume varies half an order of magnitude depending on the hour of the day. If this is an issue for a network operator, then using a smart combination of hour-dependent pricing and appropriate software support, e.g. P2P downloads could be scheduled for less busy hours of the day.

Comparing FTTH and ADSL average traffic in the past year:

- Upstream (2012): The average volume generated by FTTH customers is 4.6 times the average volume generated by ADSL customers
- Upstream (2013): The average volume generated by FTTH customers is 3.6 times the average volume generated by ADSL customers
- The use of symmetric traffic in optical access networks

Distribution of users by volume of traffic –values are based on a one-day measurement

- Heavy users FTTH (1 day 2013):
  - Upstream: 20% of Customers generate 98% of traffic
  - Downstream: 20% of Customers generate 86% of traffic

- Heavy users ADSL (1 day 2013):
    - Upstream: 20% of customers generate 93% of traffic
    - Downstream: 20% of customers generate 84% of traffic
- Also notable that the bottom 20% of the customers generate negligible amount of traffic – both upstream and downstream.

Observations:

- 20% of the customers generate an order of magnitude more traffic than the rest of the customers. They matter more than the others.
- 20% of the customers generate virtually no traffic – if the operator offers a flat price plan, then the bottom 20% percent of the customers would pay for the traffic of the top 20%.

Volume of TCP and UDP traffic (FTTH):

- Downstream: 90% of volume generated by TCP flows
- Upstream:, volume generated by TCP flows is similar to that generated by the UDP flows:
    - This is highly related to the use of the application P2P.
    - P2P uses UDP for signalling and for some file sharing phase.

Devices generating most traffic (FTTH)

- Downstream:
    - During the daytime hours, *computers* generate 8-10 times more traffic than mobile devices.
    - During night-time, traffic generated by computers and mobile devices are equally low, being in the same order of magnitude.
- Upstream:
    - Traffic profile (with the daily minima and maxima) depicts the same curve for *computers* and *mobile devices*. There is no significant difference between the two curves.
- The terms *computers* and *mobile devices* were not precisely defined in this context.
- Observation: For sharing and producing content, *mobile devices* are on par with *computers* as far as data volumes are concerned.

Evolution of data usage: 2009–2012:

- Fixed network, traffic volume ratio on a particular day:
    - 2009 downstream: 36% P2P, 35% streaming
    - 2012 downstream: 40% streaming, 20% web, 20% download
- Mobile network, traffic volume ratio on a particular day:
    - 2005: 55% mail, 40% web, 4% streaming

  - 2009 downstream: 29% web, 27% streaming, 12% download, 11% Adaptive Video Streaming Protocol (AVSP)
  - 2012 downstream: 40% streaming, 31% web, 14% download, 7% stream AVSP
- Observation: On a 7-year scale steaming gained 10 times traffic volume ratio compared to other applications.

Evolution of application use for fixed and mobile customers (2009–2012):

- Fixed network, percent of each application:
  - Streaming video went up from 35% to 40%
  - Web browsing went up from 14% to 20%
  - Download went up from 8% to 20%
  - P2P went down from 35% to 8%
- Mobile network, percent of each application:
  - Streaming video went up from 27% to 40%
  - Web browsing stayed at 30%
  - Download went up from 12% to 15%
  - Adaptive video streaming went down from 12% to 7%

FT Conclusions:

- Streaming Video has become the killer application in Fixed and Mobile Network, representing about 40% of the volume of traffic exchanged in Downstream
- P2P is the application that generates most of traffic in Upstream
- Pareto Law:
  - 20% of applications generate 80% of traffic in fixed and mobile network
  - 20% of customers generate 80% or 90% of total traffic

3.2 Forecast of traffic trends

3.2.1 General trends

For future years, all reports claim that mobile data traffic will grow faster than fixed data traffic. Indeed, during 2013 it is forecasted that mobile data traffic will continue to double each year. Anyway, fixed data traffic will remain higher than mobile data traffic in absolute terms in all the foreseen period (see Figure 3, Figure 4) by at least one order of magnitude.

The rapid growth of the mobile traffic is mainly due to the rise of the smartphone traffic which is a consequence of the increased number of smartphone sales and subscriptions. The total number of mobile subscriptions is expected to reach about 9 billion in 2017 [2]. In addition, even the traffic generated by single subscribers is forecasted to become 5 times larger in the next 6 years (Figure 5).

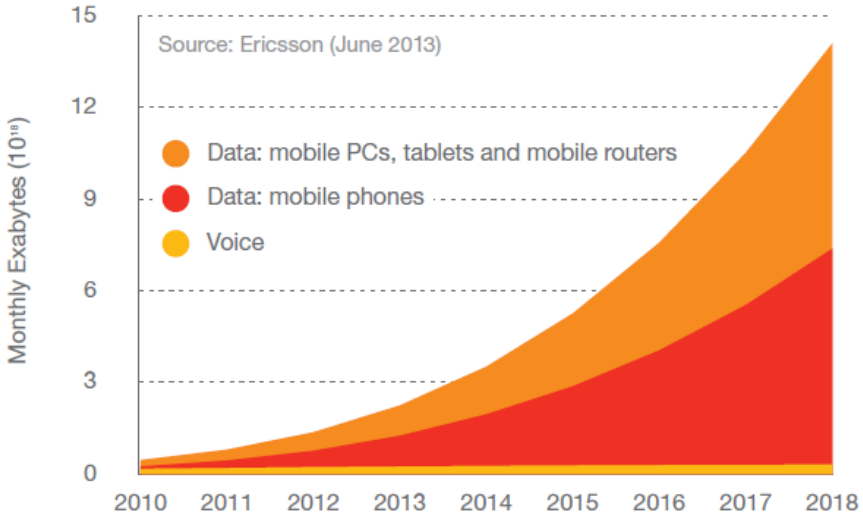


Figure 3: Global Mobile Traffic 2012 – 2018. [2]

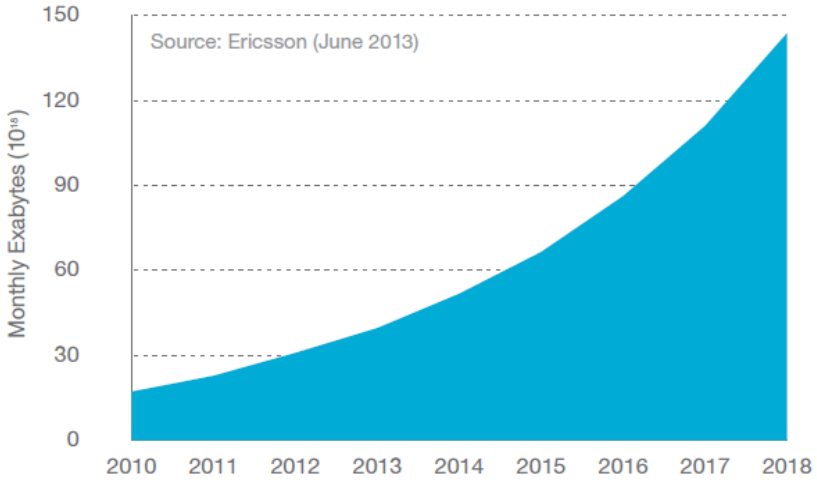


Figure 4: Global Fixed Traffic 2012 – 2018. [2]

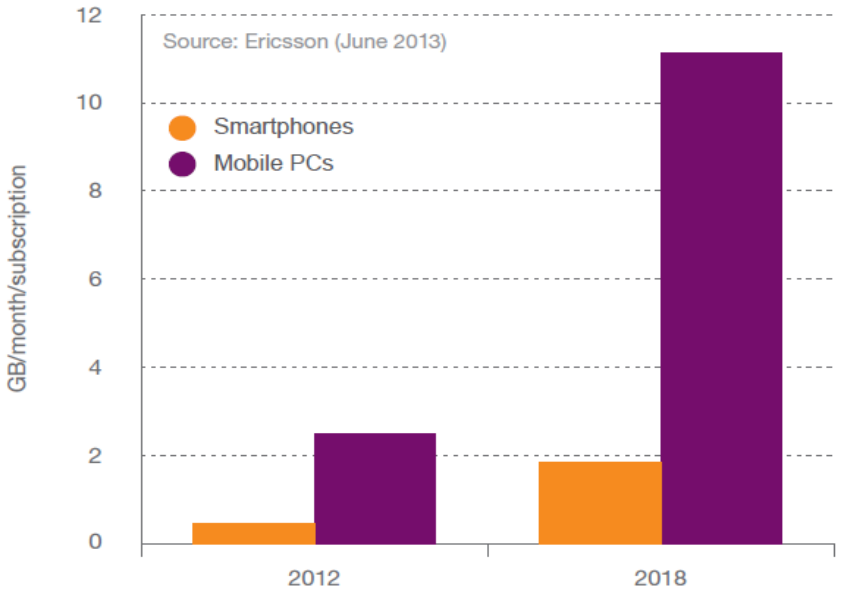


Figure 5: Mobile Device Traffic per month and per subscription forecasted in 2012–2018. [2]

According to the Cisco VNI Global Mobile Data Forecast 2012–2017 [3], the future outlook for mobile data traffic forecasts that by 2017, global mobile data traffic will reach 11.2 exabytes per month, or a run rate of 134 exabytes annually. Smartphones will be 68 percent of total mobile data traffic in 2017, compared to 44 percent in 2012; 4G connections will be 10 percent of total mobile connections in 2017, and 45 percent of mobile data traffic.

It is expected that data traffic will be always more fairly split between mobile phones on the one hand, and tablets, mobile routers and mobile PCs on the other. Moreover, the increase of the utilization of the mobile data network is driven by the growth in the amount of content (i.e., applications, services) available and to the improved network speed which follows the development of HSPA and LTE.

A forecast related to the population covered by different mobile systems in the future years state that it is expected that more than 90% of the global population will be covered at least by GSM/EDGE technology while 85% will be reached also by WCDMA/HSPA transmissions. Finally, more than 60% of the world population will be able to connect to the mobile data network through LTE systems. Such measures (Figure 6) witness that in 2018 the overall mobile data traffic will grow also due to a higher number of users which use broadband mobile technologies.

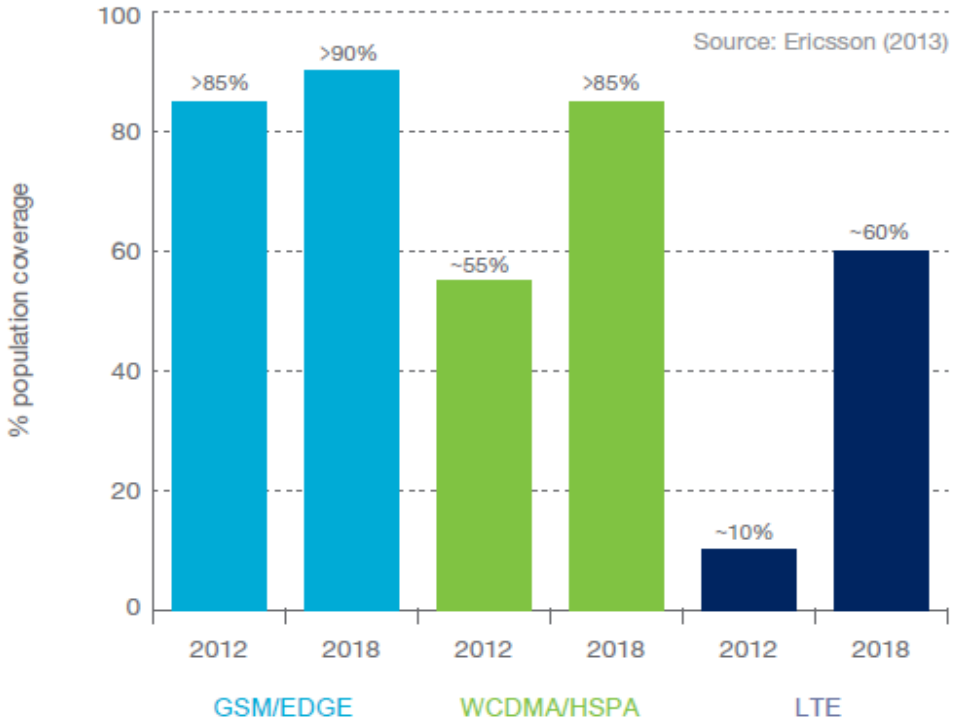


Figure 6: Forecast of population coverage divided by mobile access technology.[2]

Considering the traffic growth region by region, Figure 7 shows that in Asia-Pacific and Middle East the growth of the number of mobile subscriptions will be the strongest among all the regions [2]. On the other hand, Figure 7 shows also that market in Europe is saturated. Therefore, more data traffic will not come from more users, but from other end devices which allow more/different use and thus require more bandwidth.

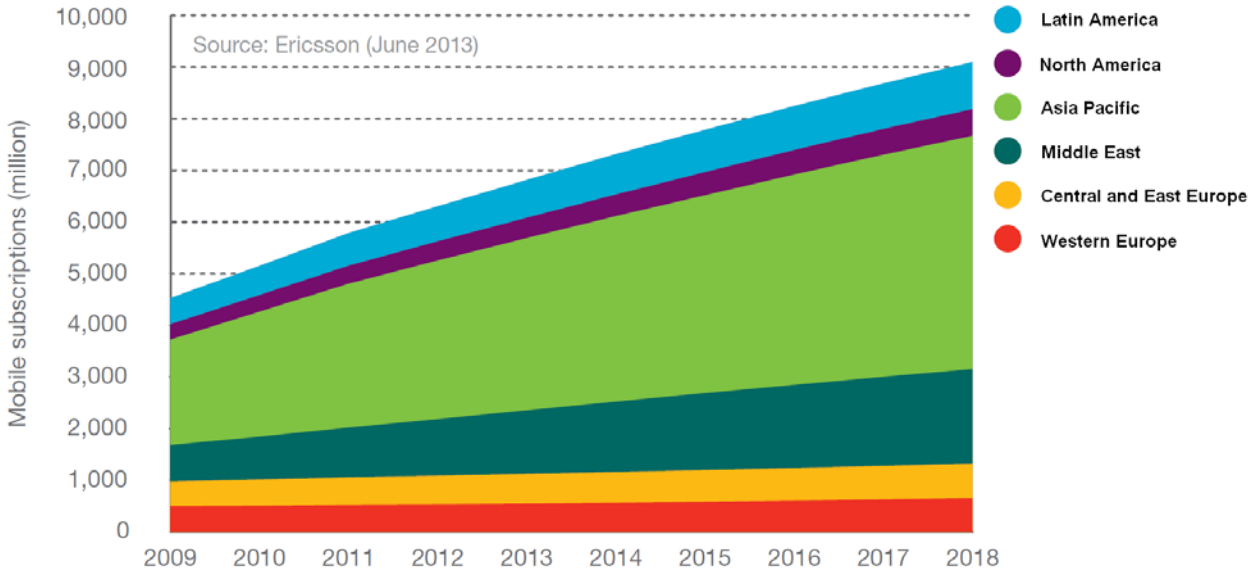


Figure 7: Mobile subscriptions by region. [2]

Now, the main question that is beginning to be discussed in fora is: what will happen in a longer timeframe, say 10-20 years? A reasonable projection made by some authors is the following.

Today:

- about 4 billion people generate voice traffic on the order of about 20 million Bytes/month/user;
- about 20 million people generate data traffic on the order of about 4 billion Bytes/month/user.

Tomorrow:

- most probably 4 billion people will generate a global mix of data traffic, including any type of service, on the order of about 20 billion Bytes/month/user.

This could be seen as a huge chance, but also as a huge threat. Therefore, the next question that could be asked is: how could we achieve a capacity increase by a factor ×1000? A possible answer, taking into consideration the mobile broadband evolution, could be given by the following considerations:

- High Speed Packet Access (HSPA) and LTE technologies can bring a capacity increase factor ×10 with respect to Wideband CDMA (WCDMA).
- Spectral efficiency improvements from 1 b/s/Hz to 2 b/s/Hz give a capacity increase factor ×2.
- More allocated spectrum gives a capacity increase factor ×5.
- A smaller cell radius (about 1/3 of that currently in use) gives a capacity increase factor ×10.

Overall, these factors multiplied together give a capacity increase factor around ×1000.

3.2.2 Broadband service trends to 2020

The past decade has witnessed innovations which changed global communication opportunities and people's way of life. As the bandwidths, contents and services grew, the volume of the broadband communication increased significantly and will continue to increase due to the constant information communication and entertainment flowing around the globe [4]. An important question is therefore: How much traffic is there on the Internet, and how fast is it growing? Network technologies, architectures, and business models all depend upon and influence the answer. However, there is very little solid data about what is happening on the network, and many conflicting estimates.

Internet traffic is the flow of data across the Internet. What stimulates its growth is a complex set of feedback loops operating on different time scales, involving human adoption of new services and improvements in processing, storage, and transmission technologies [5]. In spite of the widespread claims of continuing and even accelerating growth rates, Internet traffic growth appears to be decelerating [6]. The most, prominent source of IP traffic forecasts is the Cisco Visual Networking Index. Since 2006, Cisco provides an annual prediction of the expected IP traffic for 5 years into the future. Ever

since it started publishing results in 2006, the Compound annual growth rate (CAGR) has been dropping constantly with each update of the report.

In its current version, Cisco forecasts indicate that Internet traffic is to grow 27% annually until 2017, reaching 3 times the Internet traffic in 2012 by that time [7]. However, in order to extend those figures further into the future, one has to take into account that the IP traffic is growing at a sub-exponential rate [8]. This is illustrated in Figure 8, where the blue curve shows the annual growth rate of the total IP traffic reported in the Cisco's VNI from 2006 to 2012. The orange curve shows the regression of those values and the projection into the future. Blue line shows the historical CAGR between years 2006-2012 the green one the resultant prediction of traffic growth until the year 2020 based on fitted curve.

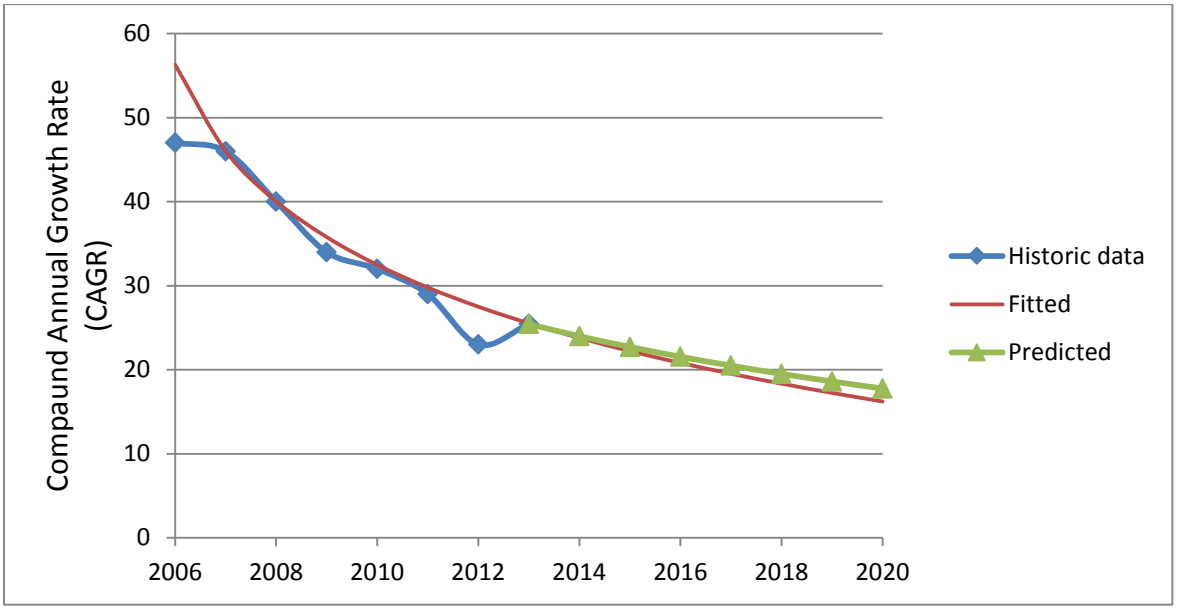


Figure 8: Prediction of the compound annual growth rate for the total IP traffic

The IP traffic can further be classified by the connection type (i.e. it can be divided into "Fixed Traffic" and "Mobile Traffic"), by the data type (e.g. Video, P2P, HTTP, Voice) and if it is generated by consumers or businesses.

The term "Fixed Traffic" denotes IP traffic generated by households, university populations, Internet cafés, corporate IP WAN traffic and traffic generated by traditional commercial TV services. Mobile data includes data and Internet traffic generated by handsets, notebook cards, and mobile broadband gateways.

Table 6 gives an overview of the predicted traffic volume. The table lists the results from the seven previous Cisco reports denoted as "Historic Data". For each of the five categories, the volume and the CAGR are given. One can see that in each category, the CAGR is decreasing. A logarithmic regression of the CAGR based on the historic data is calculated in order to take the sub-exponential growth into account. Then, based on the updated CAGR, the expected traffic is predicted. Cisco report (May 2013), which are also given for the year 2013 and 2017 to enable comparison of both approach.

By the end of 2017, global IP traffic will be around 1.4 zettabytes per year. This corresponds to around 120 exabytes (EB) per month, 4.0 EB per day [7]. If that growth continues, by 2020 the Internet traffic can be 5 times what it was in 2012.

	Total Traffic		Fixed		Mobile		Consumer		Business	
	Volume	CAGR	Volume	CAGR	Volume	CAGR	Volume	CAGR	Volume	CAGR
Historic Data										
2006	4,2	47	4,2	47	0,03	130	2,6	53	1,6	29
2007	6,6	46	6,6	46	0,03	125	4,4	49	2,2	35
2008	10,1	40	10,1	40	0,03	131	7	42	3,1	32
2009	14,7	34	14,6	33	0,09	108	11,6	37	3,1	21
2010	20,2	32	19,9	30	0,24	92	16,2	34	3,9	22
2011	30,7	29	30,1	26	0,6	78	25,8	30	4,9	22
2012	43,5	23	42,6	21	0,9	66	35	23	8,5	21
Forecast										
2013	54,0	25,4	52,4	22,9	1,6	76,1	43,8	25,5	10,2	20,8
Cisco	55,5	23	53,9	20,5	1,6	66	35	23	8,5	21
2017	118,3	20,5	105,9	17,4	12,4	62,8	97,1	19,5	21,2	18,3
Cisco	120,6	17	109,4	20,5	11,1	50	98,9	19,7	21,7	18,5
2020	211	17,8	162,2	14,3	48,8	55,5	172,2	16,2	38,8	17,0

Table 6: IP Traffic Growth

Over the same period, Busy-hour Internet traffic will increase by a factor of 3.5 between 2012 and 2017, while average Internet traffic will increase 2.9-fold. Busy-hour Internet traffic will reach 865 Tbps in 2017, the equivalent of 720 million people streaming a high-definition video continuously [7]. It is expected that in the USA the traffic per subscriber will continue to increase with 30% CAGR [9].

Metro traffic will grow nearly twice as fast as long-haul traffic from 2012 to 2017. The higher growth in metro networks is due in part to the increasingly significant role of content

delivery networks, which bypass long-haul links and deliver traffic to metro and regional backbones. Long-haul traffic is also deposited onto metro networks, so that total metro traffic already exceeds long-haul traffic. In 2012, total metro traffic was 1.8 times higher than long-haul traffic, and by 2017, metro traffic will be 2.4 times higher than long-haul as shown in Figure 3-1 [7]. To project this trend to year 2020 where amount of access and metro traffic is expected as 202 EB/month 5 folding year 2012, long-haul traffic will be 80 EB/month.

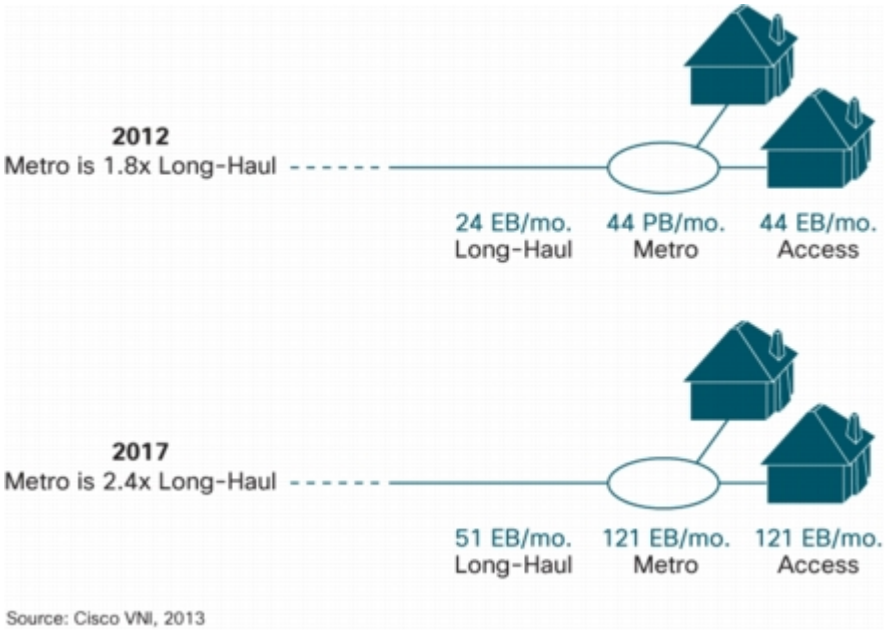


Figure 9. Metro versus Long-Haul Traffic Topology, 2012 and 2017 [7]

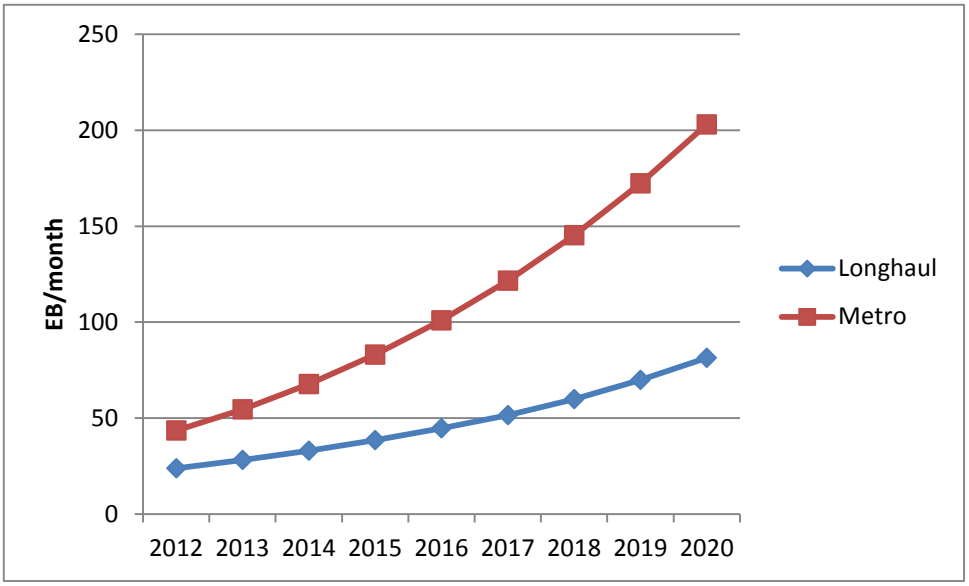


Figure 10. Metro versus Long-Haul Traffic Forecast until 2020

In the following section, the traffic predictions are broken down based on the connection type, the data type and the geographical region.

3.2.2.1 Internet Traffic Based on Connection Type

Internet traffic can be categorized by the connection type. Traffic can originate from fixed devices (i.e. from PCs, TVs, tablets and other devices connected through digital subscriber line, cable, fibre etc.) or mobile devices connected through 3G/4G networks. Based on the sub-exponential growth function, traffic predictions are made until 2020 in Figure 11. Thinking of connection types, fixed access network can be more divided into wired and wireless connection types. Figure 12 shows traffic forecast based on the connection type. Here in 2020 it is expected that the amount of traffic generated by wired access fixed networks will be nearly same as mobile access networks, which corresponds to half of the amount of Wi-Fi accessed fixed networks.

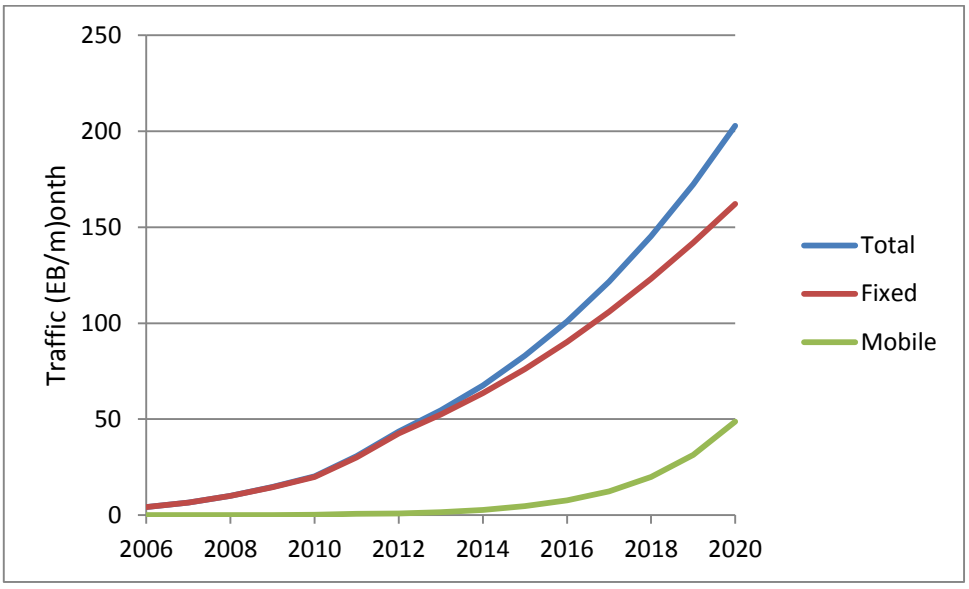


Figure 11. Predictions for the Fixed and Mobile Traffic

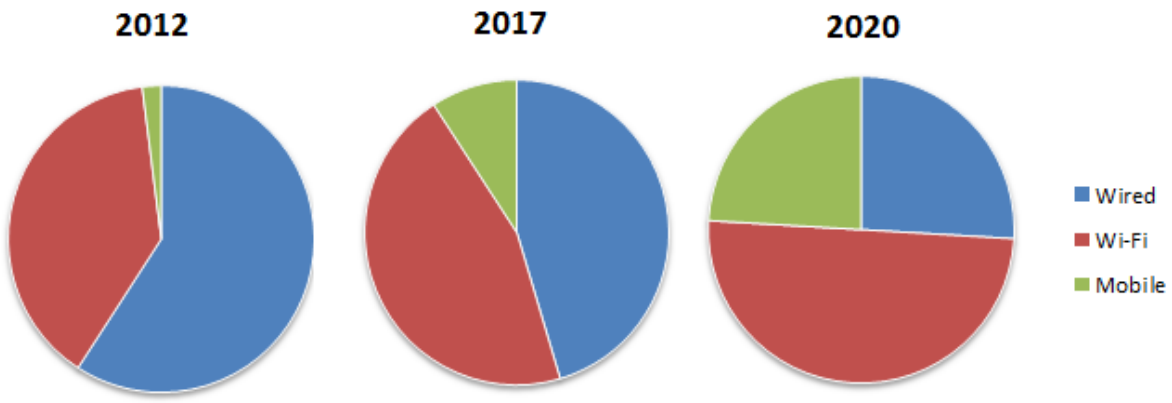


Figure 12. Traffic based on connection type

3.2.2.1.1 Fixed Internet Traffic

Fixed Internet traffic includes Wi-Fi access traffic which is generated by mobile devices (e.g. smart phones, tablets, notebooks etc.) which offload their traffic through a fixed connection when available. IP traffic generated by wired connected end-devices comprises approximately 59% of the total IP traffic in 2012. Studies show that by 2014, IP traffic generated by end-devices connected through Wi-Fi access network is expected to surpass the wired access IP traffic. The same study implies that by 2017, wired connection type portion of Internet traffic will constitute only 39% of the total IP traffic [7]. If the forecasted growth rates continue until 2020, wired connection type portion of fixed IP traffic will constitute only 28% of the total IP traffic.

3.2.2.1.2 Mobile Internet Traffic

Global mobile data traffic is expected to grow three times faster than fixed IP traffic – having an estimated compound annual growth rate (CAGR) of 76% and reaching 12 EB per month by 2017. Hence, the estimated 120 EB per month will include 105 EB of fixed traffic [7]. This indicates that although mobile Internet traffic is increasing rapidly, fixed Internet will continue to comprise most of the consumer Internet traffic. By 2020, mobile traffic will be head to head with wired fixed traffic growing to 49 EB per month which is 25% of global traffic.

Touch-screen smartphones (launched from around 2007) have been a key driver of mobile data traffic growth [10] – leading to an increase of the mobile Internet usage. By 2009 there were 95 million mobile Internet users in Europe [10]. At the same time, mobile video traffic was growing dramatically. By 2010, YouTube and other Internet services with Flash video have generated the majority of the mobile video traffic. It is predicted that video will account for 66% of mobile data traffic by 2014 [10], [2]. Media rich social network users with mobile Facebook are twice as active as the average user. In 2010, more than 75% of smart phone users accessed social network sites [10],[2].

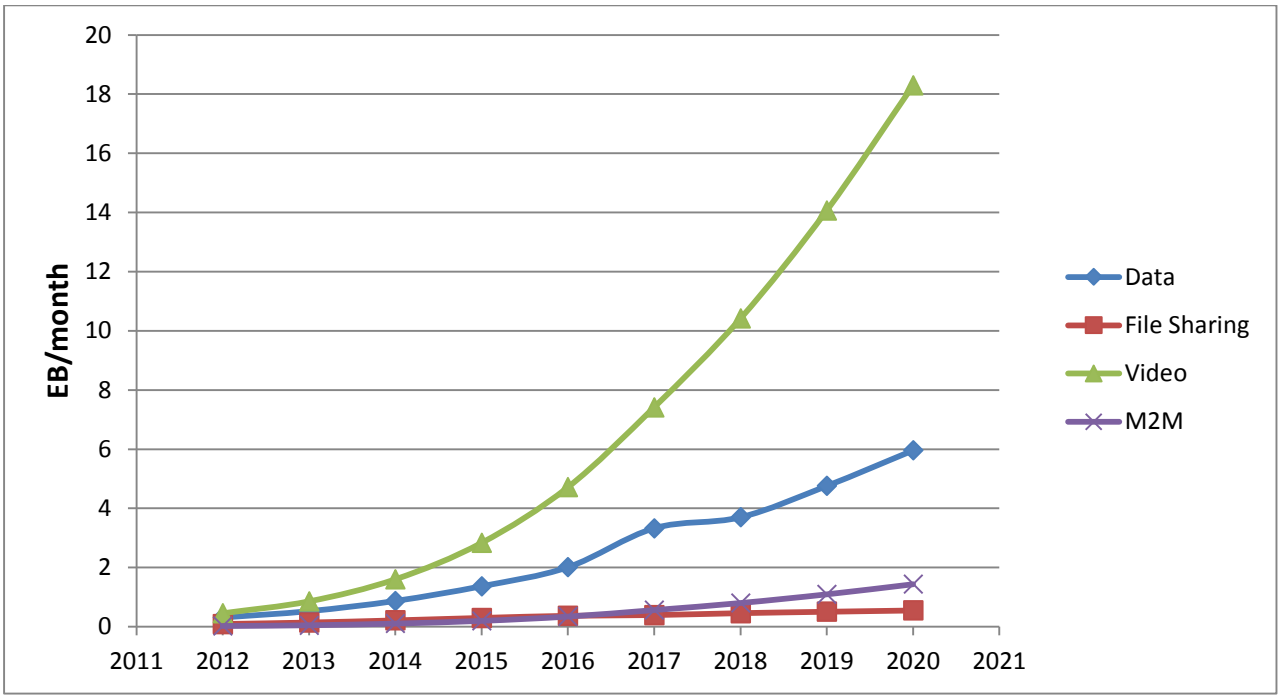


Figure 13. Mobile Internet traffic forecast based on application type

The evolution of the devices and the expected traffic per device has been evaluated by the UMTS broadband forum in [2]. In 2020, 60% of the devices will consist of smartphones. Connected PCs and laptops (through Dongles which may also be included in the laptop) and tablets (denoted as connected devices in [2]) will make up 15% of the mix.

By 2020, the traffic generated by tablets and 3G/4G connected PCs and Notebooks is expected to be similar to the traffic generated today (2012) by a fixed connection on Digital Subscriber Lines or CATV/Cable TV networks [2]. In Europe, in year 2020, it is expected that there will be 1.7 mobile subscriptions per inhabitant [2]. With a population of about 500 million in EU-27, that will lead to 850 million subscriptions.

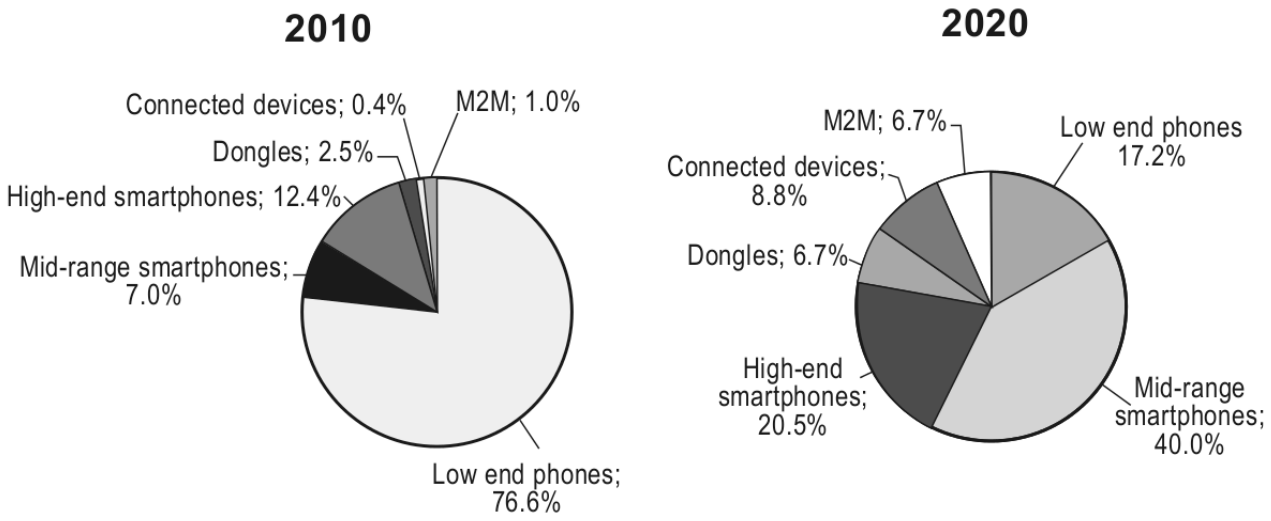


Figure 14. Worldwide mobile device mix in 2010 and 2020 [2]

By 2017, over 30% of the world's population are expected to live in metro and urban areas. These areas represent less than 1% of the Earth's total land area, yet are set to generate around 60% of mobile traffic by 2017 [11]. Broadband mobile population coverage (WCDMA, HSPA and LTE) will increase from 50% (2012) to about 85% in 2017 [11]. Figure 15 shows projection of mobile traffic generation per access technology per area to 2017.

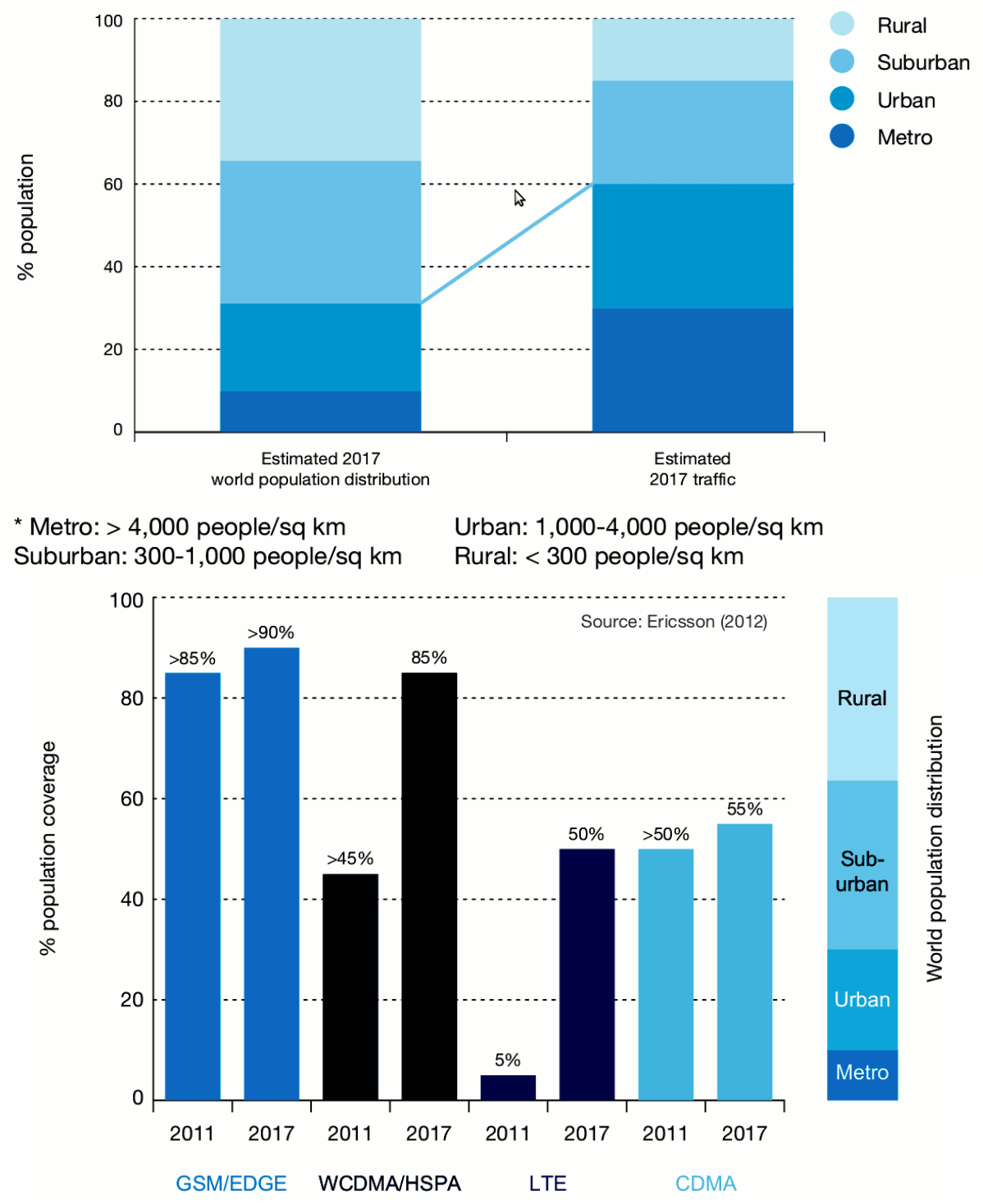


Figure 15. Mobile traffic generation per area in 2017 [11]

3.2.2.2 Consumer and Business Traffic

Traffic predictions for the consumer- and business IP traffic based on the Cisco VNI with sub-exponential growth are given in Figure 16Figure 16. Predictions for the Consumer and Business Traffic

In 2012, the business IP traffic volume was about 20% of the total traffic. This will drop to about 15% in 2020 since the consumer traffic is growing at a slightly higher rate.

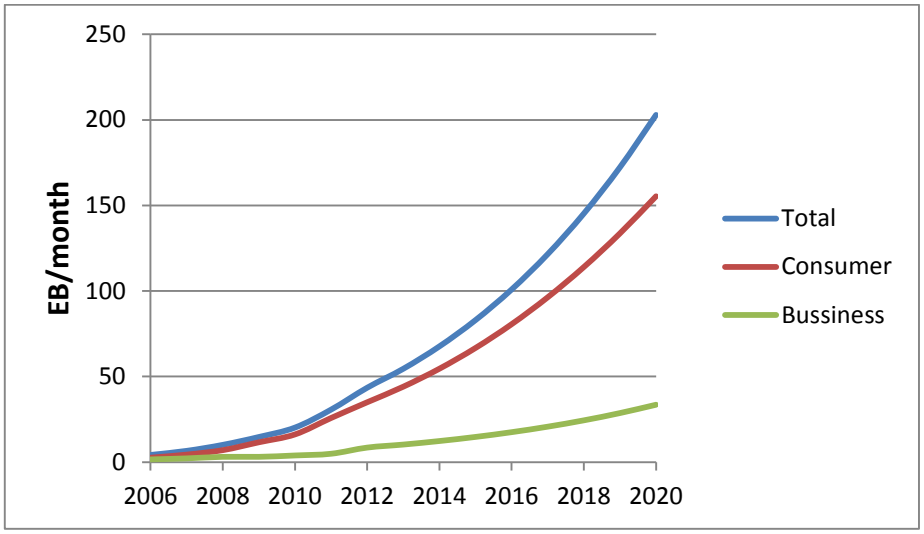


Figure 16. Predictions for the Consumer and Business Traffic

3.2.2.2.1 Consumer Internet Traffic Based on Data Type

The sum of all forms of IP video (Internet video, IP VoD, video files exchanged through file sharing, video-streamed gaming, and videoconferencing) will continue to be in the range of 80 to 90 percent of total IP traffic. Globally, IP video traffic will account for 73 percent of traffic in 2017. Taking a more focused definition of Internet video that excludes file sharing and gaming, Internet video will account for 52 percent of consumer IP traffic in 2017 [7].

Detailed information about the categories of consumer Internet traffic is given below.

Internet Video Traffic

Internet video traffic constituted 38% of the global IP traffic by 2012. It is expected to grow to 52% by 2017 [7]. This statistic does not include video shared via P2P file sharing. Internet video traffic is expected to increase 29% annually. The overall video content from all sources such as Internet, P2P, TV and video on demand, broadly named IP video, will continue to comprise 73% of the global consumer traffic by 2017 [7]. By 2020, it is predicted that Internet Video traffic will reach to 90 EB per month which will be 60% of consumer traffic and %45 of global traffic.

File Sharing

File sharing comprised 23% of the consumer Internet traffic in 2012. It is expected to comprise 23% of the total consumer Internet traffic by 2017 [7]. From 2011 to 2016, file sharing traffic is expected to grow with 26% compound annual growth rate (CAGR) [7]. The trend of file sharing is included in Figure 17. In 2013, estimated file sharing consumer Internet traffic was around 7.1 EB/month. It is expected to reach 9.0 EB/month by 2020 [7].

Web, Email and Data

E-Mail, data and web traffic was 19% of the total consumer Internet traffic in 2012.. The traffic is expected to have a CAGR of 35% [7]. The trend of web, e-mail and data is included in Figure 17. The monthly web, e-Mail and data traffic is approximated to be 14.4 EB/month by 2017 and 24 EB EB/month by 2020 while it was 5.1 EB/month in 2012 [7].

Online Gaming

In online gaming, there is no need to have a powerful graphical processor but only a low delay and a high bandwidth Internet connection, it is becoming more and more popular each day. Most of the online (or cloud) gaming platforms have been in serve since 2011 or 2012 [7]. Therefore they contribute very little to the consumer traffic for now. Specifically, for now cloud gaming traffic constitutes only 0.04% of the total online and offline gaming traffic. However, it has a great potential to become one of the biggest contributors to the Internet traffic. Forecasts imply that online gaming traffic in 2020 will be 5 times higher than it is in 2012 [7].

Internet traffic based on Online Gaming is expected to have one of the most rapid increases for the next five years ahead, with a CAGR of 52% [7]. The projection of online gaming traffic amount towards the 2020 is plotted in Figure 17.

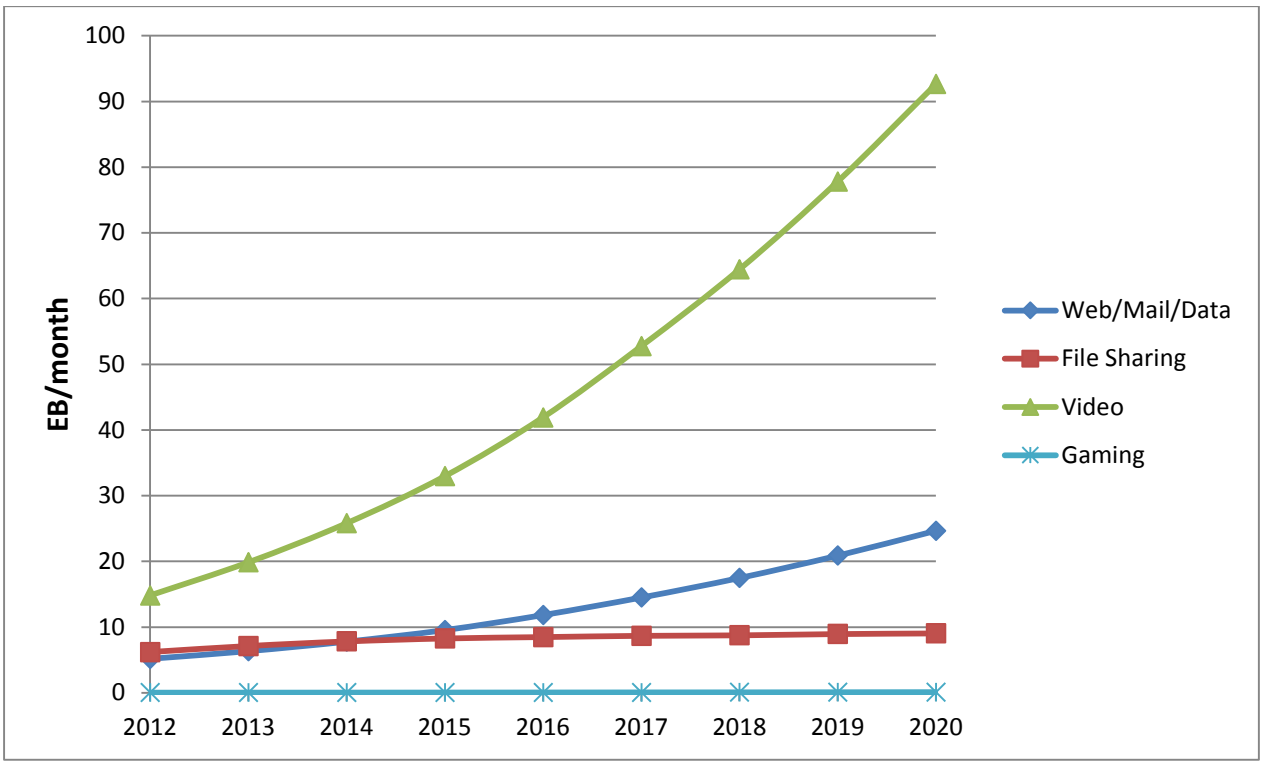


Figure 17. Consumer Internet IP traffic forecast based on data type

As can be seen in Figure 17, the consumer Internet traffic is expected to increase in all data types. Among these data types video will be the dominant data type in consumer Internet traffic.

3.3 Drivers for traffic growth

The growth of the mobile traffic is huge throughout the globe and constantly changing our behaviours and our society. Nowadays, almost half of the population of the earth uses mobile communications. In the last 4 years a billion of mobile subscribers were added providing a total amount of current mobile users of 3,2 billions [12]. Anyway, there are many people who have not the possibility to access the Internet yet. Therefore, it is reasonable to think that mobile traffic will continue to grow in the future years.

According to a report of Cisco [13], in 2012 global mobile data traffic grew more than 70% year over year, 855 petabytes a month. This percentage of growth was different among the regions. The lowest traffic growth was experienced in Western Europe (44% of traffic growth). There are various reasons which explain the low traffic growth rate measured in such area. One of these reasons is the elimination of many unlimited data plans and the consequent introduction of tiered mobile data service plans to price the mobile services, e.g., according to the needs of the users. Moreover, in the Western Europe a slowdown in the increase of the laptops connected to the mobile network was registered. Finally, operators have promoted the offload of the mobile traffic to the fixed network, i.e., onto the Wi-Fi networks.

Conversely, the highest traffic growth rates were recorded in Asia Pacific (+95%), and in Middle East and Africa (+101%). Therefore, such emerging markets are the major contributors to the increase of the mobile traffic. This is mainly due to a decrease of the prices of devices and to the consequent higher possibility to own a mobile technology. Moreover, the development of new technologies which can be used to provide connectivity to rural areas is another cause of the traffic rise in such emerging markets.

It is important to consider that the nature of traffic growth in developed countries and in developing countries is very different [12] [2]. In fact, as it is possible to notice in Figure 18 and as also noticed in previous sections, Africa and Europe have a similar number of total subscribers, but this number is due to different reasons in the two regions. In Africa the growth of connections (8%) is caused by subscribers addition. Instead, in Europe this growth rate is mainly due to the increase in the number of SIM cards per subscriber. This means that many consumers own several different types of devices and use multiple SIM cards.

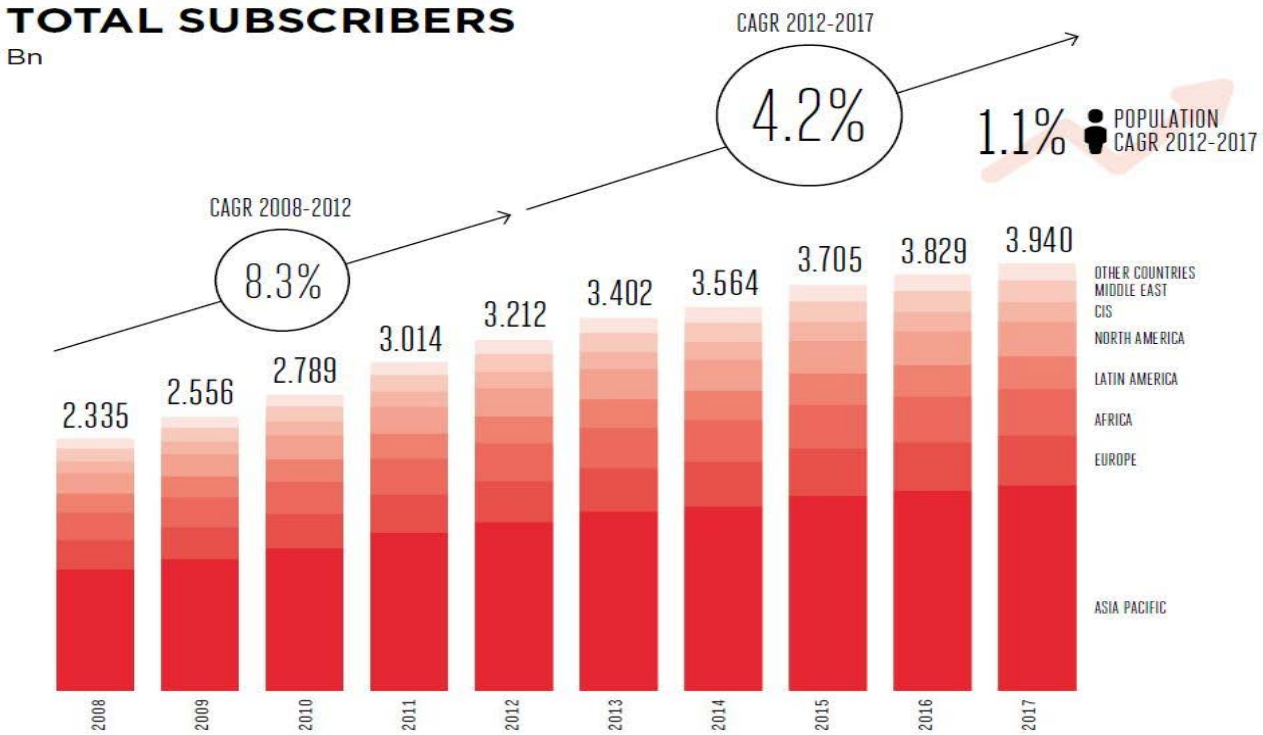


Figure 18: Total subscribers [12]

In general, the current traffic growth is due to many factors. The spread of wireless devices, the increase of network-attached, the diffusion of video services, the introduction in the market of both public and private cloud delivery models, the faster speed of mobile connections, and finally the decreasing costs of smart-phones are the most important drivers of the traffic growth. We can classify such drivers of the traffic growth in main categories:

- Increase of the number of connected devices;
- Increase of the available bandwidth provided to end-users;
- Increase of the number of bandwidth hungry applications and services.

In the following we will discuss in details the main categories listed above which cause the increase of the overall traffic in the networks.

3.3.1 The increasing number of devices

The rise of sales of mobile devices is one of the main drivers of the traffic growth. Recently various devices that incorporate mobile network connectivity such as smartphones, tablets, e-book readers, TV sets and gaming consoles appeared in the market.

These new devices offer improved resolution and screen size. Such improvements caused an increase of data consumption per image and foster people to use bandwidth hungry applications such as video streaming or video calls. Indeed, according to Ericsson [2], the mobile subscriptions are increasing for tablets, smartphones and PCs.

The strongest growth in the last years concerns smartphones sales and subscriptions. At the end of 2012, the total smartphone subscriptions reached 1.2 billion. As a matter of fact, 3.2 billion people out of 7 billion on earth currently own a mobile phone [12].

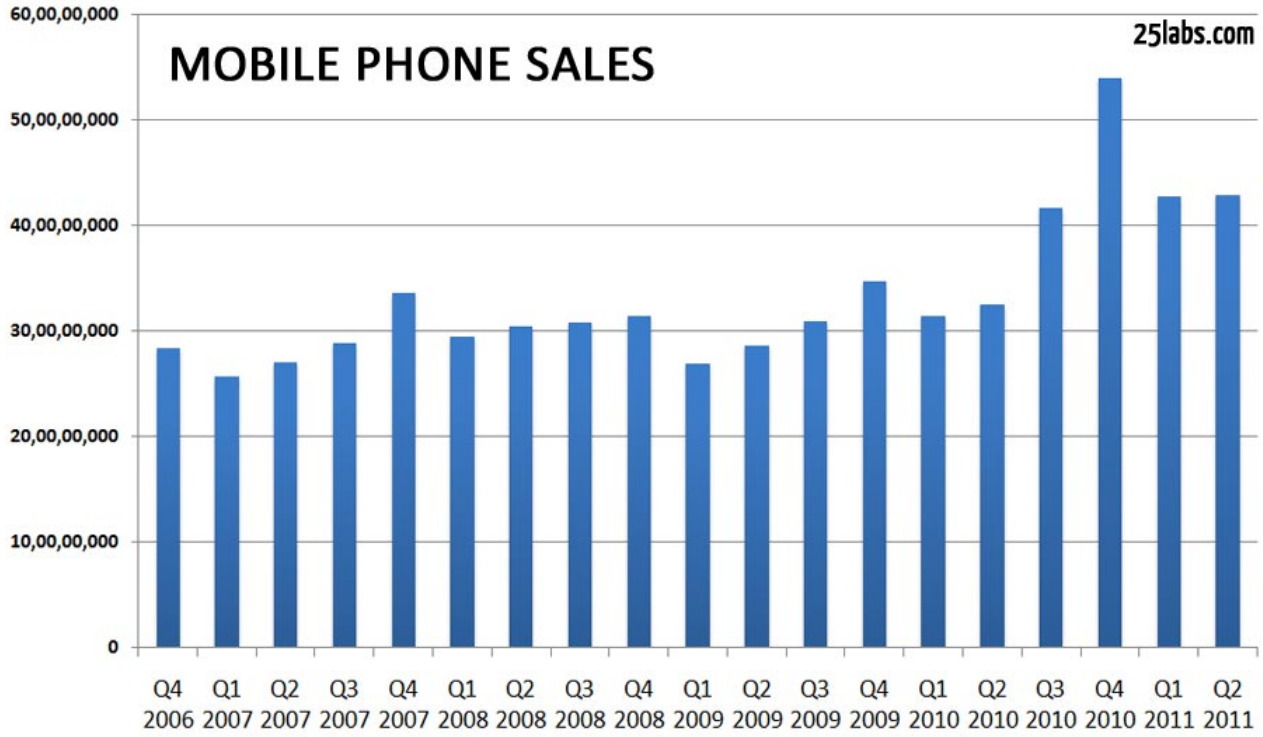


Figure 19: The sales of mobile phones worldwide [14].

In the future, the number of smartphones will continue to increase while the number of non-smartphones will stop to grow. Also the number of mobile devices such as tablets, laptops, and machine-to-machine (M2M) modules will continue to grow in the next years, as shown in details in Figure 20.

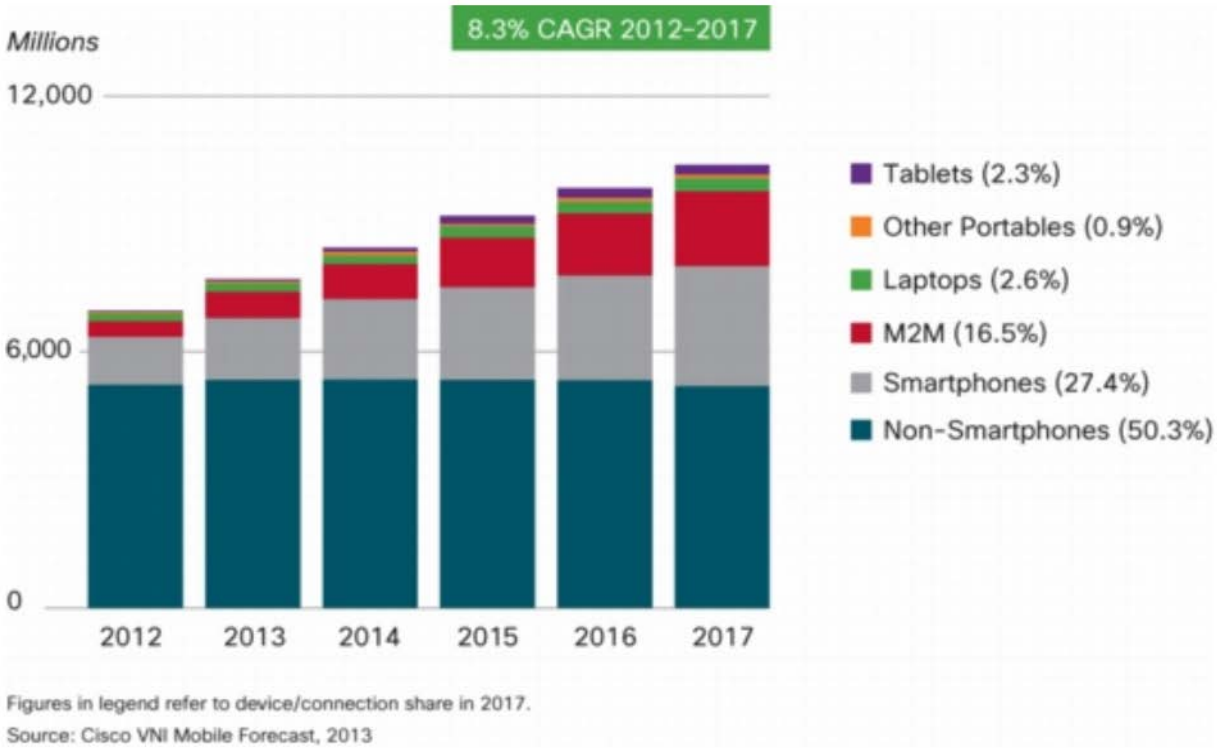


Figure 20: Growth of the number of mobile devices [13].

If we consider that in developed markets a smartphone generates much more traffic than basic phones, we realize that such rise in the smartphones sales have a big impact in the traffic growth in recent and future years.

Figure 21 shows a Cisco forecast where are reported the contributions to the bandwidth consumption due to different devices, in present and future years. In Figure 21, we can notice that, nowadays smartphones and laptops are the main contributors to the bandwidth consumption. In the next years smartphones will become the devices which will provide the highest amount of bandwidth. A very important outcome of the forecast shown in this figure is that the amount of bytes consumed will continue to grow in the future.

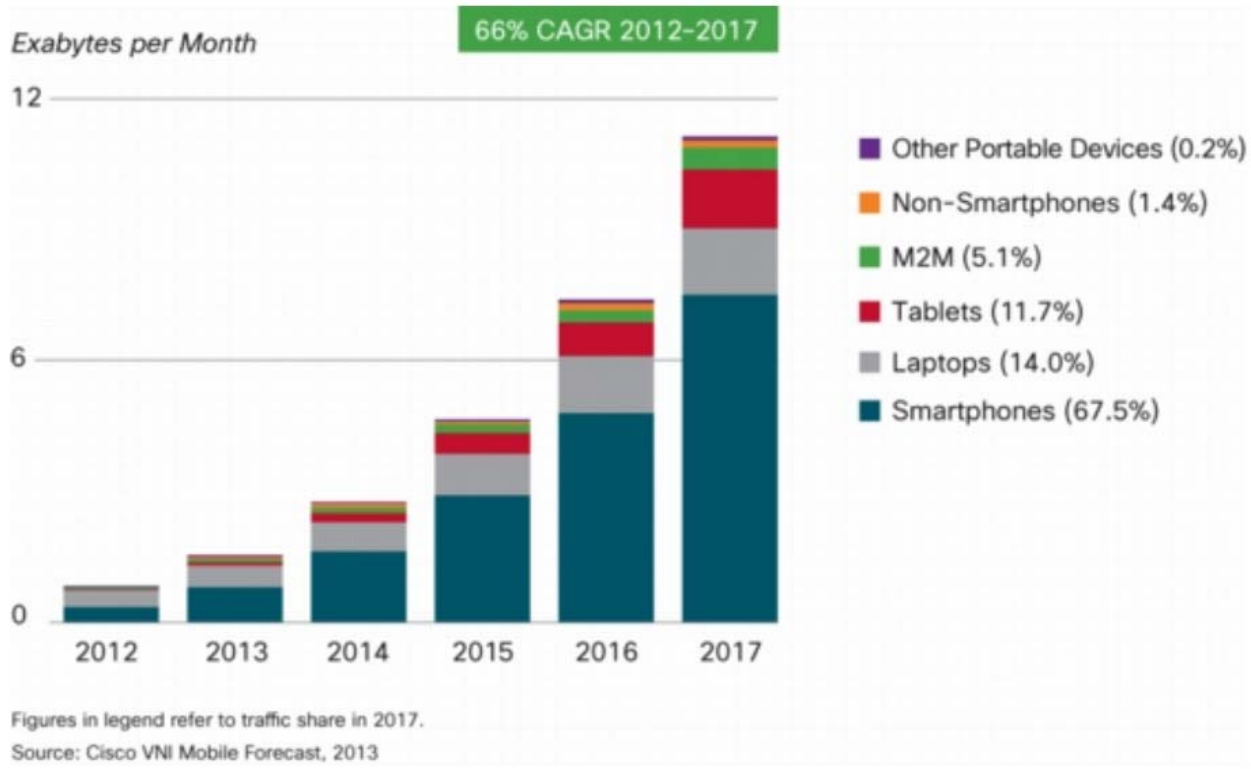


Figure 21: Devices responsible for mobile data traffic growth [13].

3.3.2 The bandwidth domain

Another driver (and enabler) for the traffic growth is the increase of connections numbers and speeds.

Indeed, the augmented speed of the connections allows each users to consume more data, encouraging to use applications which requires higher bit rates, e.g., games, video streaming, video calls, and social network applications. In Figure 22 the current and forecasted trends of the connection speed growth are shown.

	2012	2013	2014	2015	2016	2017	CAGR 2012-2017
Global							
Global speed: All Handsets	526	817	1,233	1,857	2,725	3,898	49%
Global speed: Smartphones	2,064	2,664	3,358	4,263	5,284	6,528	26%
Global speed: Tablets	3,683	4,811	6,082	7,624	9,438	11,660	26%

Figure 22: Projected average mobile network connection speeds in kbps. [13]

The enhancement of the connections speed is mainly due to the increasing proportion of 4G mobile connections, as shown in Figure 23. The 4G connections, which include Long-Term Evolution (LTE) and mobile WiMAX, providing higher bandwidth, lower latency, and

an increased level of security, will promote the generation of a huge amount of mobile traffic.

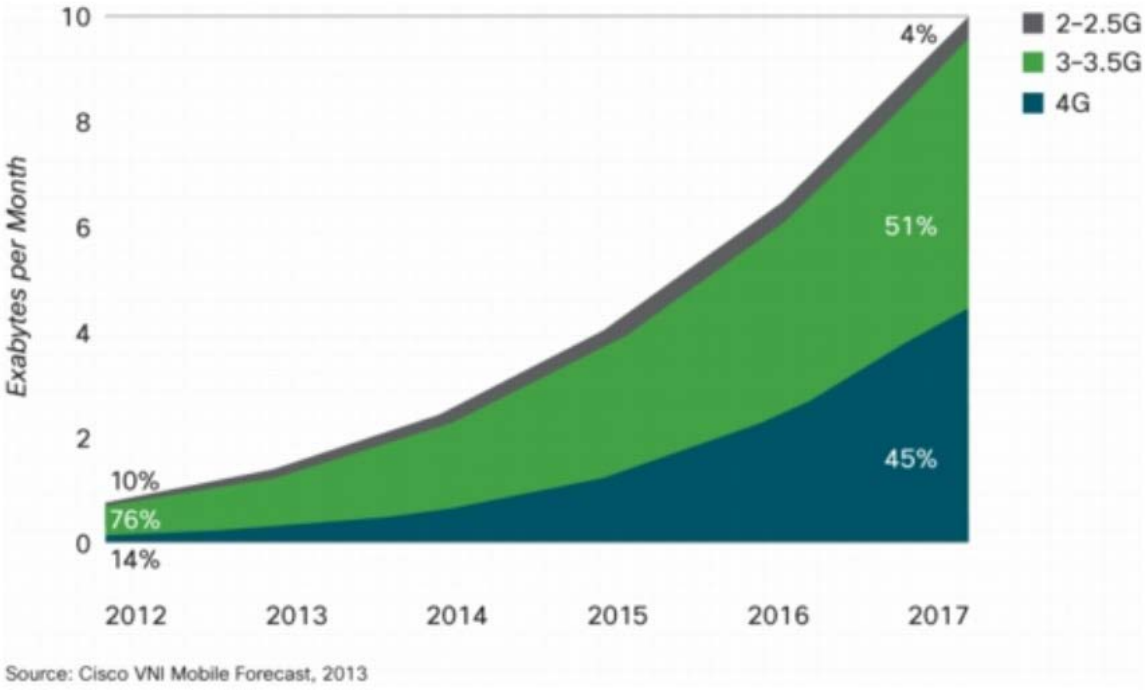


Figure 23: 4G proportion growth forecast. [13]

Today a single 4G connection generates 19 times more traffic than another non-4G connection. This is mainly due to the fact that many of the current 4G connections are used for residential broadband access, which has in general higher average usage. Moreover, since the 4G system provides high connection speeds, it encourages the usage of bandwidth hungry applications.

3.3.3 Services and Applications

Several new mobile services and applications have been recently introduced on the market and are finding large consensus among both consumers and enterprise customers. Some of these services and applications generate a huge amount of data traffic. Examples of these applications are Cloud services, content sharing, music and video streaming, video on demand and video telephony. Among these video streaming is one of the major contributors of bandwidth. Moreover, as shown in Figure 24, the number of mobile requests for video content is expected to increase [15] due to the future enhanced technologies which will be deployed (e.g., LTE and LTE-Advanced) and also to further improvements to the devices (e.g., size, screen resolution, processing power).

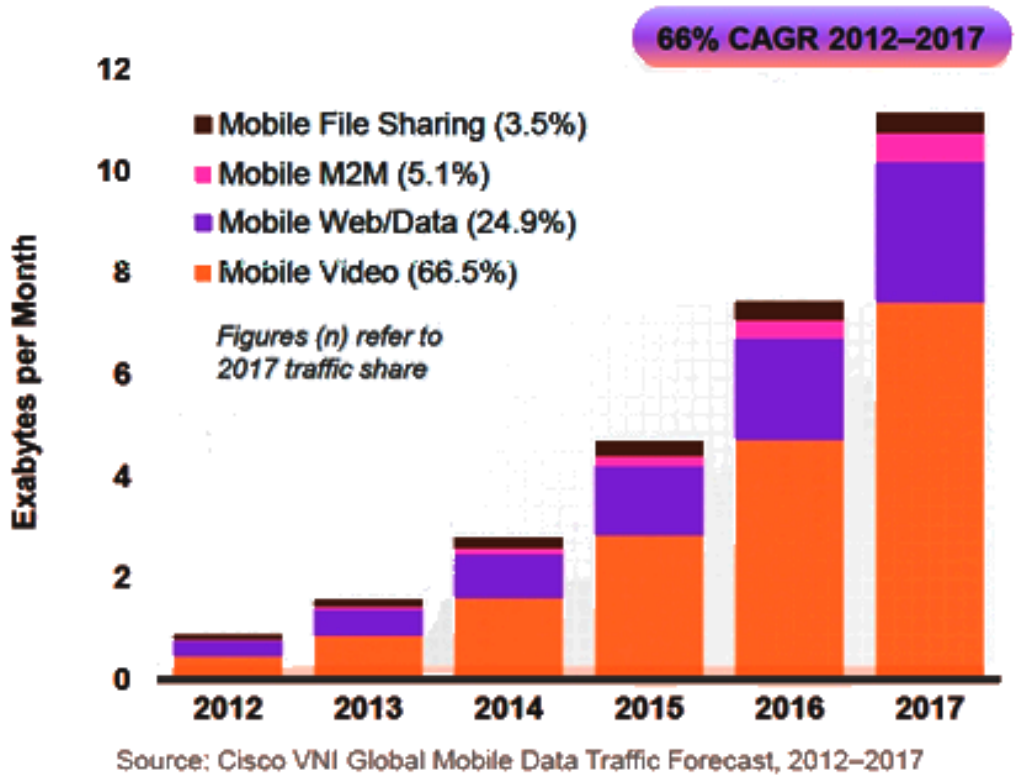


Figure 24: Contribution of the different applications to global mobile data traffic growth. [13]

Mobile voice has been overtaken by mobile data at the end of 2009 and it is expected that the growth of mobile voice traffic will remain limited compared to the rapid and huge growth of the data traffic [2]. This trend can be noted in Figure 25, where the growth of mobile voice traffic is compared with the growth of mobile data traffic during the period from 2007 to 2013. The numbers in the figure are derived from measurements over a large base of commercial networks that cover all regions of the world.

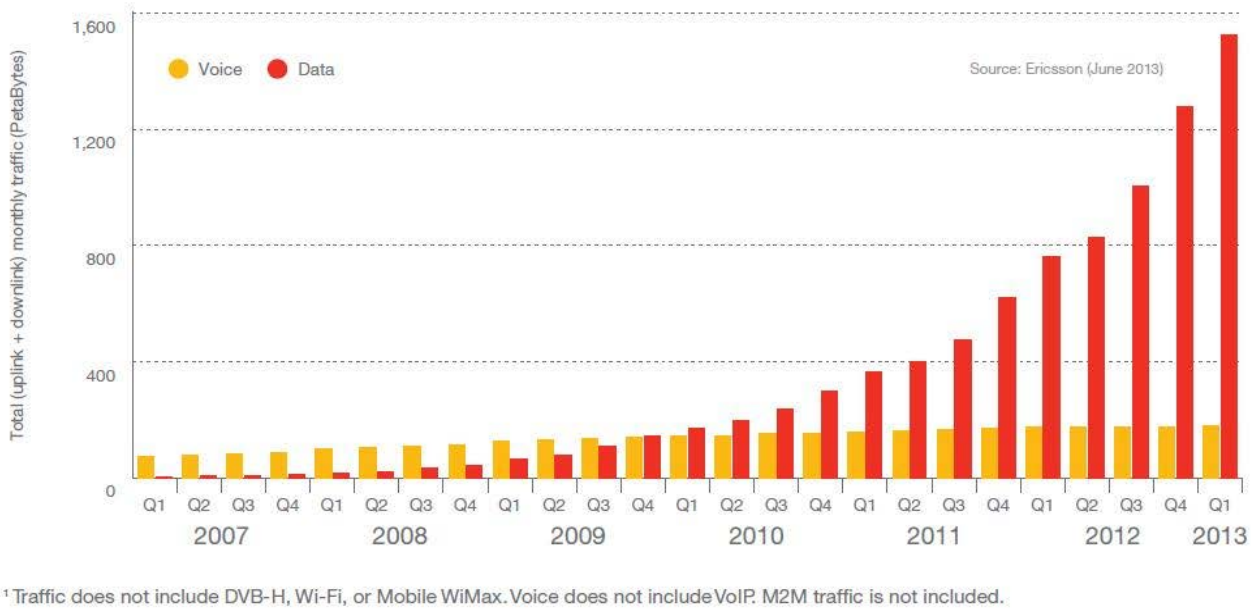


Figure 25: Global total data traffic in mobile networks, 2007-2013. Source [2]

Social networks applications generate significant traffic since they are widely used. The social networks applications themselves are not exactly bandwidth consuming applications, but many times their usage is associated with video on demand utilization, thus generating a big amount of additional traffic. In Figure 26 it is shown how the number of Facebook users through mobile devices has grown during the years.

Figure 26: Faceboook's mobile monthly active users (Q4 2010 – Q1 2013). [16]

3.4 Traffic of new perspectives: CDN

Thanks to the development of multimedia technologies and high-speed networks, multimedia entertainment applications like video on demand, online songs and movies, IP telephony, Internet radio and television, and interactive games have now become popular

networking applications over the Internet. To meet this demand, Content Delivery Networks (CDNs) have recently been widely deployed to deliver the contents from content providers to a large community of geographically distributed users. CDN service may be provided either by CDN service providers such as Akamai that partner with multiple Internet service providers (ISPs) or by a big ISP itself. A CDN can achieve scalable content delivery by distributing load among its servers, serving user requests via servers that are close to the users.

In a typical CDN system, a user's request is redirected to a nearby CDN server through a certain redirection mechanism so that content delivery takes place at the edge of the network where bandwidth is abundant. This server performs admission control to either accept or block the request. If the request is accepted, the local CDN server serves the user if it has the content; otherwise, it performs content routing to locate and then deliver the requested content to the user. In the existing CDN architectures, content delivery is achieved via managed and controlled servers. Resource usage depends on the popularity of the content provided and optimization can be performed on this basis. When considering an FMC scenario (described in a use case of COMBO deliverable D2.1), this architecture somewhat need to change, since all data has to go through the PGW in the mobile core, which is independent to the CDN location.

Cooperative caching and application-level multicast (ALM) are two important technologies in a multimedia CDN for delivering on-demand and live multimedia contents respectively. In cooperative caching, the CDN servers cache all the content cooperatively when single CDN is not capable of caching the whole content. In application level multicast, the content is served from the CDN server having the requested content through a multicast group formed of CDN servers and clients. With its overlay structure, CDNs shorten the paths and delays as well as stabilize the throughput.

With the emergence of popular video-streaming services that deliver Internet video to the TV and other device endpoints, CDNs have prevailed as a dominant method to deliver such content. Globally, 51% of all Internet traffic will cross content delivery networks in 2017, up from 34% in 2012. Globally, 65 percent of all Internet video traffic will cross content delivery networks in 2017, up from 53 percent in 2012 [7]. At 2020, traffic generated by content delivery networks will be reached to 90 EB/month catching the amount of video constituting 45% of global traffic.

3.5 Signalling changes

It is supposed that the always-on data-centric nature of devices in the LTE/EPC networks can results in an explosion not only of data traffic, but signalling traffic as well [17]. Even more, it is expected that signalling traffic growth will significantly outpace mobile data traffic growth [18].

A load on network entities due to signalling traffic can be caused by, e.g.

- Frequent loss of broadband coverage that may potentially generate extremely frequent intersystem change activities [19];
- Flood of registrations caused by 1) mass of mobile users attempting simultaneously to perform registration procedures such as attach or location updating and 2) restart of

RAN and Core Network (CN) nodes and CN nodes which handle mobility management (MSC/VLR, SGSN, MME and HSS/HLR) [19];

- Data traffic offloading can also generate very frequent intersystem change activities.

Moreover, control data signalling tends to scale with the number of users, while user data volumes may scale more dependent on new services and applications [20].Thus, as the number of the LTE/EPS users is being increased the signalling load towards network entities is also growing.

One more trend related to signalling issues is to introduce a full separation between control signalling and user plane operation. The decoupling is motivated by the fact that both the control signalling operation and the user data functionality have to be implemented in optimized ways for separating these functions in the logical architecture [20]. One more fact is that the full separation of control signalling and data operation allows more flexibility handling user data functions in a more distributed way in the networks, while at the same time allowing for a centralised deployment of the equipment handling the control signalling [20]. While most control layer decisions have been decoupled from user plane elements, the responsibility to disseminate them still relies on specific interfaces between the P-GW and S-GW where the S-GW shares both control and user plane responsibilities.

Thus, the signalling traffic trends can be considered as open research issues and it is interesting to analyse both LTE signalling traffic that is expected to be generated towards Evolved Packet Core (EPC) and signalling changes that are required on the architectural design level to support true decoupling of control signalling from user data traffic. The former is considered in the next section, the latter is highlighted in sections 3.5.2 and 3.5.3. In particular, section 3.5.2 deals with required design enhancements towards a full signalling and control separation and section 3.5.3 analyses how feasible it is to add these enhancements to the EPS architectural framework.

3.5.1 Traffic signalling forecast

The always-on data-centric nature of devices in the LTE/EPC networks can results in an explosion not only of data traffic, but signalling traffic as well [17]. Some typical examples why signalling traffic can be generated are given below.

The UEs like smartphones are “always-on” devices that automatically attach to the network [17]. Each attached device generates signalling traffic. Besides, to conserve battery life, the smartphones can detach from the wireless connection after a short period of inactivity. Then, the UE should be reattached when data service is used again. It is obviously that the continual attach/reattach from a large number of smart devices generates considerable signalling traffic.

Besides, the applications installed in the UEs create the signalling traffic. As application usage grows, signalling traffic increases [17]. Some applications, like instant messaging and social networking, are always-on services that require regular message exchange between the client and server. This continuous attach/detach/reattach process further increases the load in the control plane.

Moreover, if there should be a network problem where service is interrupted, once service is restored there will be a large amount of devices all trying to reestablish a connection automatically and simultaneously [17].

It is supposed that Diameter signalling traffic will be dominant quite essential in LTE/EPC networks.

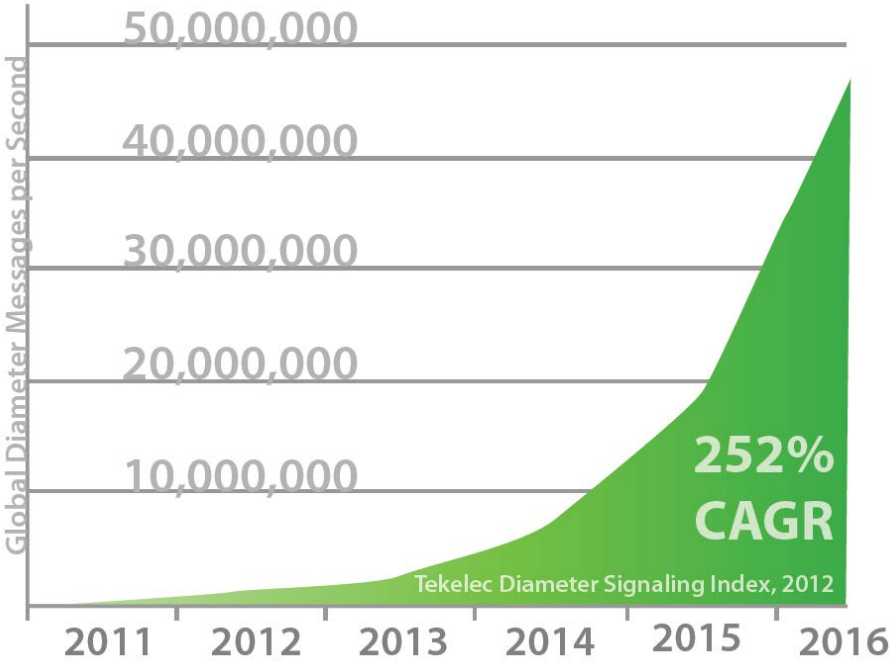


Figure 27. Forecast of global Diameter signalling traffic

Figure 27 shows the growth of compound annual growth rate (CAGR) of signalling messages required to support different services [18]. As one can see from the figure the number of Diameter messages per second (MPS) is increased exponentially and expected to reach 50M messages per second by 2016 in accordance with the forecast [18].

Control plane elements, which are mostly decoupled from the user plane path, handle Policy, Authentication, Privacy, QoS, Charging and Mobility functions [20]. Diameter control plane entities involved providing data and voice services are the HSS, EIR, OCS, and PCRF. The network entities of LTE/EPC architecture for supporting these services are illustrated in Figure 28 [18].

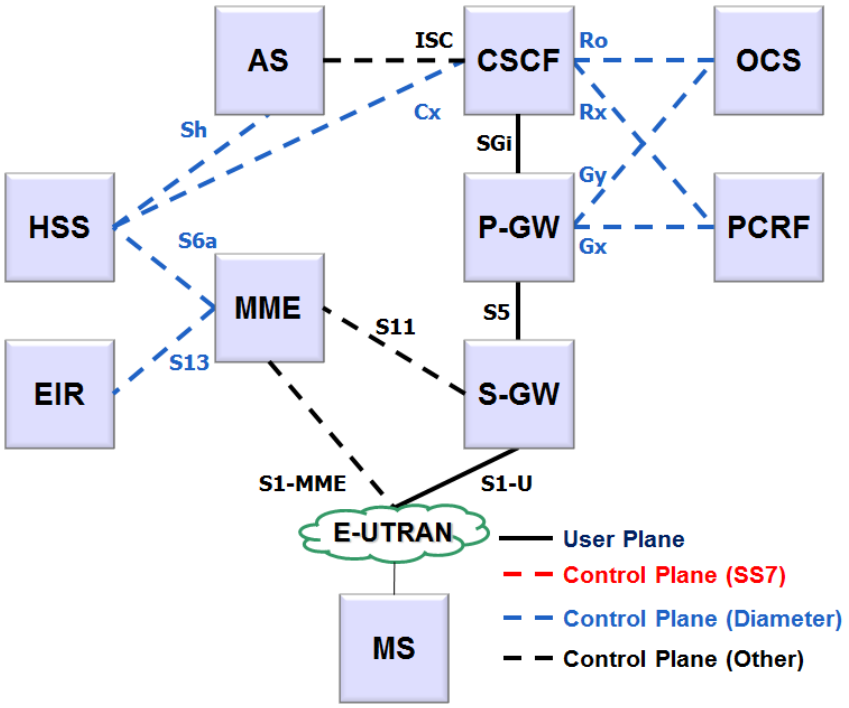


Figure 28. LTE/EPC architecture to support voice and data services

As seen from Figure 29 [18], **policy** has the largest impact on signalling traffic growth since the PCRF interacts very frequent with the charging system (via the P-GW), enforcement points, etc. [18]. There can be several instances of the RCRF in the network because of scalability, redundancy and other reasons [17]. Online charging represents also essential contribution to the total Diameter signalling traffic because of the interactions between the OCS, PGW and PCRF [18]. As one can see mobility generates not so much signalling, but the relative values do not take into account LTE roaming procedures [18].

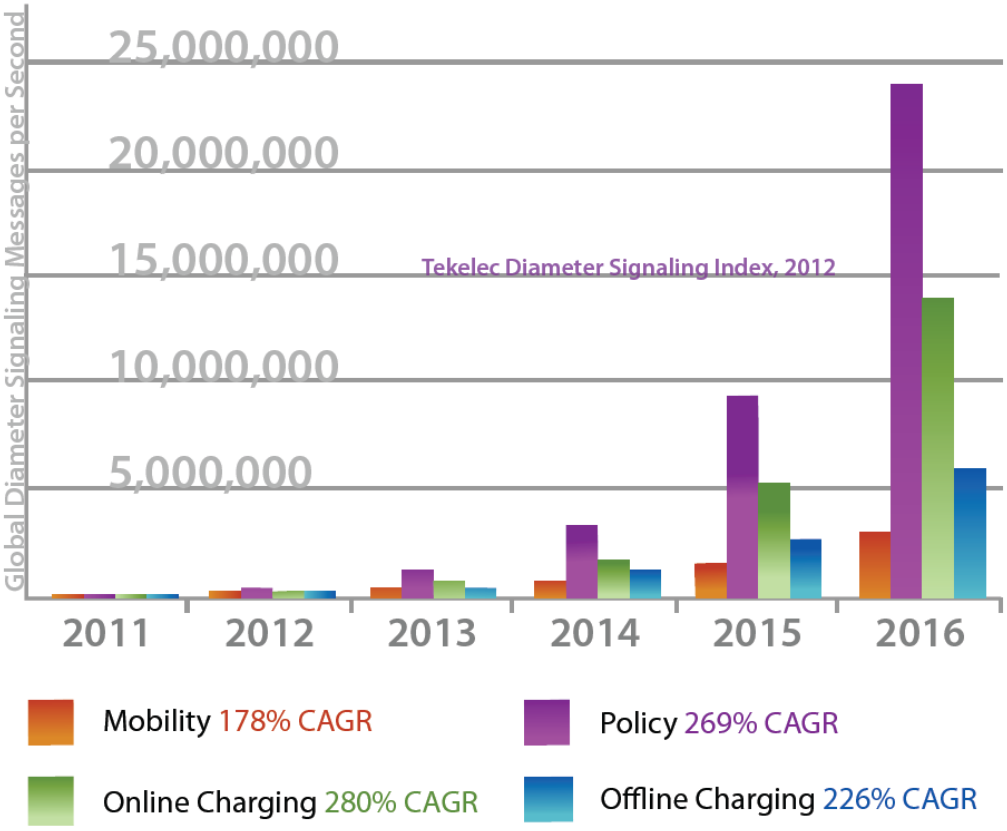


Figure 29. Forecast of global Diameter signalling traffic by message type

With regard service type contributions to Diameter signalling growth the report [18] states that voice and video over LTE produce the largest amount of signalling traffic.

As was mentioned above the attach/detach frequency of the UEs can essentially impact on amount of the signalling traffic, in particular, Diameter signalling traffic. We have used the Diameter traffic calculator [22] to illustrate dynamics of changes of signalling messages if attach/detach procedure is initiated each 15, 30, 45, and 60 minutes correspondingly. The methodology for calculation is based on the following events that can generate Diameter messages [21]:

- Attach of the UE to the E-UTRAN. It requires authentication of the terminal at the HSS and the EIR using S6a/S13. The HSS updates the MME with the subscriber data, the default data session is created at the S-GW and P-GW requiring policy using Gx from the PCRF and reports the initial charging event using Gy to the OCS.
- Tracking Area Update. When the device attaches, it is assigned a list of tracking areas as the registration area. When the UE moves outside the registration area a procedure is required with the MME, S-GW and P-GW to update the location of the UE and possibly change the S-GW and/or MME that involves the use of S6a to the HSS and Gx to the PCRF.
- Service Request: The UE or the network may decide that the data session needs to be modified, e.g. to increase the QoS for a video download or reduce the QoS when the

S-GW is overloaded. This requires modification of the data session at the S-GW and P-GW that involves the use of S6a to the HSS and Gx to the PCRF.

- Detach of the UE from the E-UTRAN: When the UE is powered off or the UE is not used for a prolonged period it detaches from the E-UTRAN requiring the use of the S6a, Gx and Gy.

  Initial data that we use for Diameter signalling traffic evaluation is summarized in Table 7.

Number of LTE Devices (Million)	1
Annual Growth Rate of LTE Devices (%)	150%
Number of Simultaneous always-on Apps per Device	1
Annual Growth Rate of Simultaneous Apps per Device (%)	20%
Frequency of device attach/detach (min)	15, 30, 45, 60
Frequency of tracking area updates (min)	60
VoLTE Enabled (%)	0
Annual Growth Rate of VoLTE Enabled (%)	25
VoLTE BHCA per Subscriber	2
Annual Growth Rate of VoLTE BHCA per Subscriber(%)	25
Policy Enabled (%)	100

Table 7. Initial data for signalling traffic forecast

The calculation is done for 1 Million of LTE devices. It is assumed that initially there are no LTE subscribers that use VoLTE. However, there will be a 25% annual growth of subscribers using LTE infrastructure to make voice calls. The number of attempts of the VoLTE calls in the busy hour is also supposed to be increased by 25% annually.

Diameter signalling traffic forecast based on these input data depending on UE attach/detach frequency is shown in Figure 30. Diameter signalling traffic forecast..

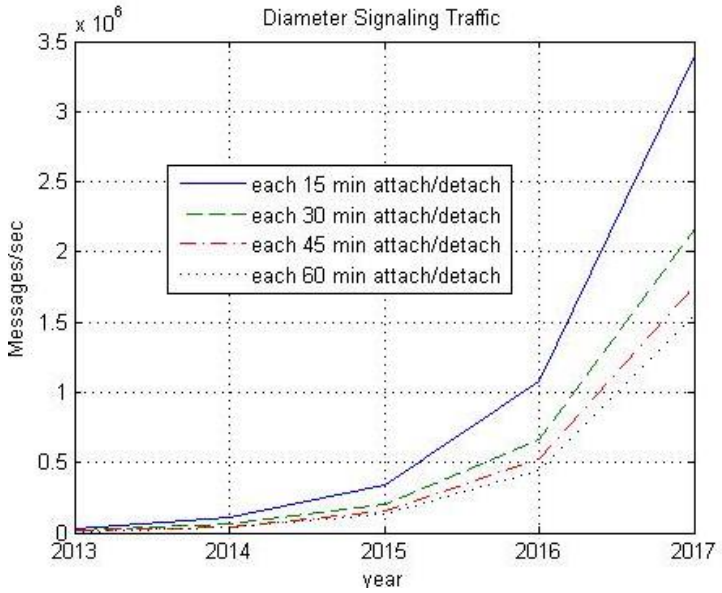


Figure 30. Diameter signalling traffic forecast

As seen from Figure 28, total amount of Diameter signalling messages is increased dramatically when frequency of device attach/detach to/from network is changed. In particular, if the device attach/detach frequency is changed from once per hour to 45 min, the increment of signalling traffic is 13%. If the attach/detach is completed each 30 min the traffic growth is around 40%. If it is each 15 min the amount of signalling traffic is increased more than twice in compare with 60 min attach/detach. Therefore, the mobile network operators (MNOs) should be careful with this parameter. For instance, if a MNO introduces offloading strategies in its network on the one hand it can reduce essentially the amount of data traffic but on the other hand it can lead to explosion of signalling traffic. Very frequent attach/detach process can happen, e.g. if users move by cars or trains from home to office and back and along roads there are many offloading areas.

3.5.2 Towards a full decoupling of control signalling from user data traffic

Since it is supposed that in the proposed FMC reference architecture the commonly distributed elements of mobile core network (e.g. S-GW, PDN-GW) will be separated from the commonly centralised elements of the mobile core (e.g., MME, PCRF), the idea is to consider in this Section some design aspects of the mobile network architecture to decouple *completely* signalling operation from user plane functionality and estimate how many signalling changes it requires on the control plane level to manage connectivity and mobility procedures.

One of the EPS architecture aims is to introduce a clear separation between control and user plane operation. Control plane elements, which are mostly decoupled from the user plane path, handle authentication, privacy, QoS and mobility functions. The decoupling is motivated by several factors [17].

One of these factors is that control data signalling tends to scale with the number of users, while user data volumes may scale more dependent on new services and applications. The second factor is that the full separation of control signalling and data operation allows

more flexibility in handling user data functions in a more distributed way in the networks, while at the same time allowing for a centralised deployment of the equipment handling the control signalling [17].

Moreover, without full decoupling of control and user planes, multihoming support that is required for deployment of advanced 3GPP concepts where multiple interfaces are used simultaneously (e.g., IFOM, MAPCON, S1-flex [17], [18], [19]) is an issue. It requires the duplication of user plane operation that should be coordinated appropriately with control plane functionality. If the decoupling between two planes is not designed in its full extension, the complexity of the process is significantly increased.

The decoupling between the two planes is not realized completely up to now even though there is significant progress in this direction. In particular, when observing the Control Plane structure of the EPS network one can notice that the architecture builds around the tandem MME/S-GW that is located at the centre of the whole system. While MME only has control plane functionality, the S-GW shares both control and user plane responsibilities. In particular, the S-GW acts as a transit point for the signalling exchange between the MME and the PDN-GW. Interestingly, while most control layer decisions have been decoupled from user plane elements, the responsibility to disseminate them relies on specific interfaces between the PDN-GW and S-GW. In this sense, the future evolution path should be done towards a full separation of control signalling from user data traffic.

Thus, the basic proposal here is to move all control signalling responsibilities of the S-GW to the control plane (the MME in particular) and, as a consequence, establish additional interfaces between control and user plane elements. Specifically, a design principle that it is worthwhile to propose is to establish an interface between the MME and the PDN-GW that substitutes the role of S5/S8 interface for control plane information (see Figure 31). As a consequence the MME acquires the role of being the sole responsible for the establishment and maintenance of the user plane path during the attach process and during mobility events. To achieve this goal there is the need to define an additional interface between the MME and the PDN-GW that will be used to exchange all those signalling messages that are currently exchanged through a combination of S11 and S5/S8 interfaces. Even more, the S11 interface needs to be extended to accommodate those signalling messages that are directly exchanged between the PDN-GW and the S-GW.

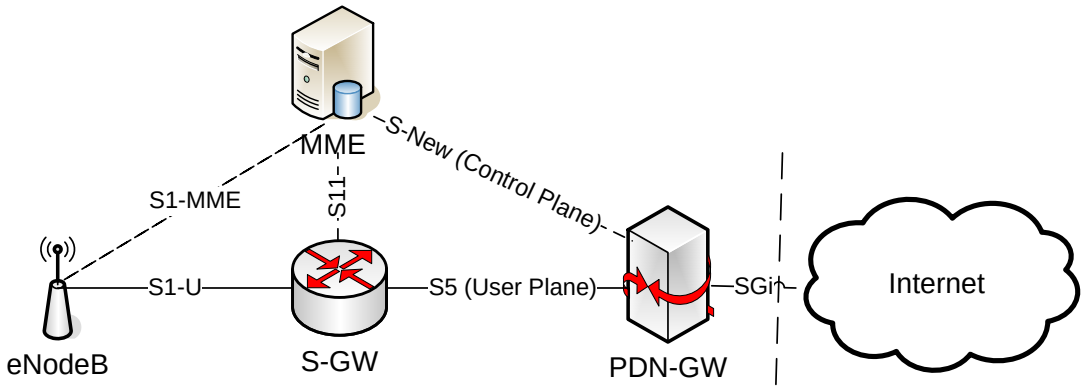


Figure 31. New interface for full decoupling of control signalling from the user data traffic

With a complete decoupling, the formation of multiple user plane paths becomes simpler (among other advantages it can bring), as it is reduced to informing the certain user plane nodes (e.g., the eNodeBs, the PDN-GW) about the multiple options that they have to forward data by the special entity (MME) that handles all control plane functionality as illustrated in Figure 32.

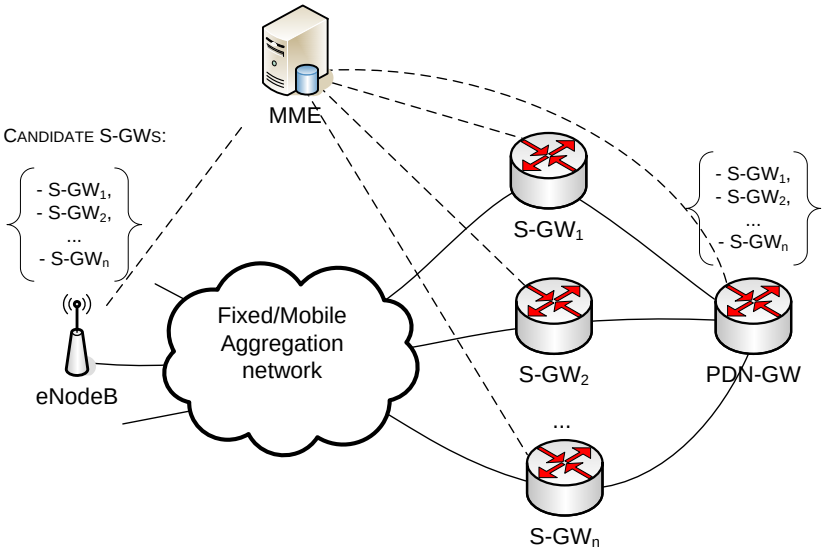


Figure 32: Multiple data path formation with full decoupling of control signalling from the user data traffic

It could be realized by two ways. As the first way, the MME can send to the eNodeB a signalling message (e.g., Initial Context Setup Request/Attach Accept [20]) including addresses of several candidate S-GWs for user plane instead of one as it is currently realized. As the second way, the MME can send to the eNodeB the same signalling message, but several times, with a different S-GW address. Similarly, the PDN-GW can receive the same information about candidate S-GWs, directly from the MME responsible of this particular User-Plane path.

The feasibility to move the messages exchanged between the PDN-GW and the MME in the current EPS architecture to the new proposed interface is analysed in next Section.

3.5.3 Analysis of signalling changes related to a full decoupling of control and user planes

This section analyses the feasibility of full separation of control signalling operation from the user plane functionality when managing connectivity and mobility procedures. In particular, we review how the singling messages currently exchanged between the MME and the PDN-GW over the S-GW (using S11 and S5 interfaces) can be redistributed taking into account the new interface called conditionally S-New to release the S-GW from control responsibilities.

Within the scope of this analysis, we review two basic EPS procedures [20], namely, the attach procedure and the X2-based handover (without S-GW relocation).

The attach procedure with the S-New interface

Figure 33 presents a flow diagram with control messages related to default bearer establishment within the attach procedure when there is the S-New interface between the MME and the PDN GW.

As one can see from the diagram, the Create Session Request message (initiated by the MME) that is currently released between the S-GW and the PDN-GW (crossed-out lines in Figure 33) can be sent directly between the MME and the PDN-GW (dash-and-dot line). Among other parameters the message can inform the PDN-GW about MME's TEID for sending control messages on the S-New interface towards the MME.

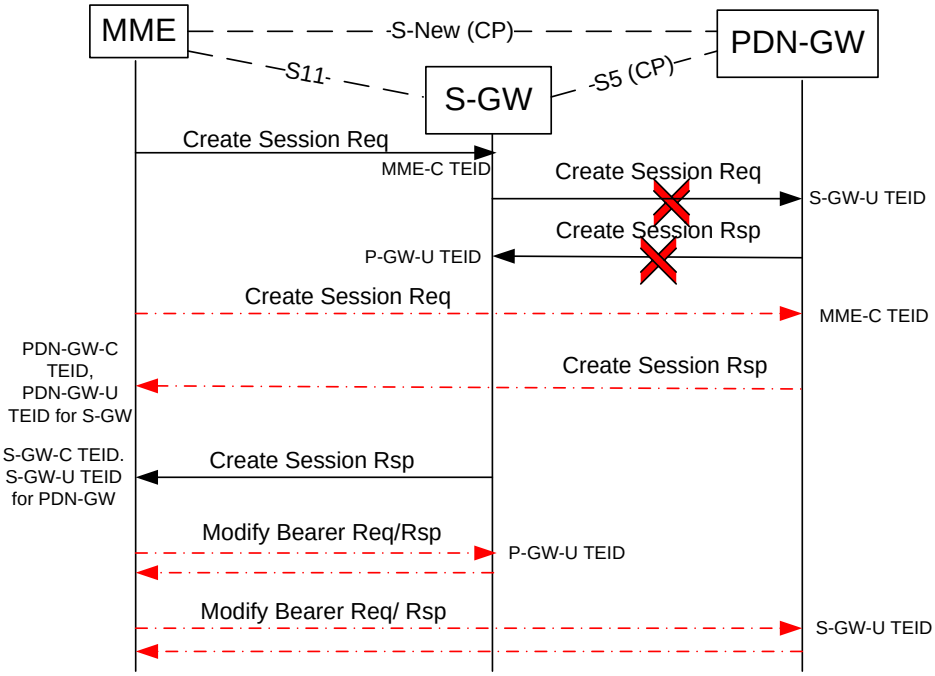


Figure 33: Default bearer establishment with the S-New interface

Once the PDN-GW receives this request it sends the Create Session Response with some parameters including the PDN GW's TEID for sending control messages on the S-New interface from the MME to the PDN GW. Besides, the message contains the PDN GW's TEID for user plane. In turn, with the Create Session Response the S-GW informs MME about the S-GW's TEID for user plane.

Now the MME is aware of the PDN GW's TEID and S-GW's TEID for user plane and it should forward this information to the S-GW and the PDN-GW for sending PDUs on S5. It can be realized, for instance, by means of sending the Modify Bearer Request (or Create Session Request as an option) from the MME to the S-GW and from the MME to the PDN-GW with the PDN GW's TEID and S-GW's TEID, respectively. The S-GW and the PDN-GW should acknowledge by sending the corresponding response messages to the MME. Then, the S-GW and the PDN-GW can exchange data with each other in both directions.

Therefore, for default bearer establishment with the S-New interface we can release two signalling messages from S5 interface, but we need to add four messages to the S-New interface and two messages to the S11 interface.

Figure 34 presents a flow diagram with control messages related to dedicated bearer activation within the attach procedure when there is the S-New interface between the MME and the PDN-GW.

With the S-New interface, the PDN-GW can send the Create Bearer Request message directly to the MME without the S-GW involvement. The message should contain the PDN-GW's TEID for user plane. The MME forwards this information to the S-GW.

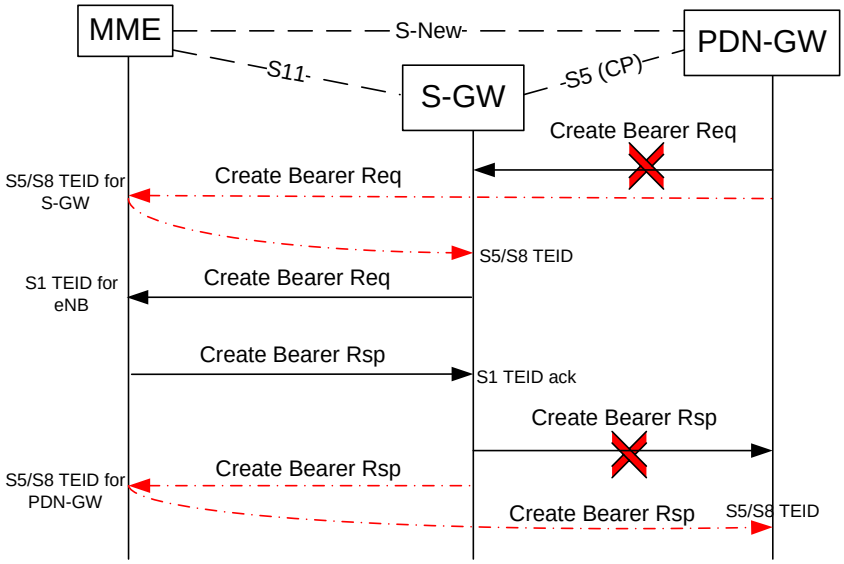


Figure 34: Dedicated bearer establishment with the S-New interface

Once the S-GW receives this message, it sends the Create Bearer Response to the MME with the S-GW's TEID for user plane. The MME should propagate this message towards the PDN GW.

As a result, for the dedicated bearer activation process with the S-New interface we can release two signalling messages from S5 interface, but instead, we need to add two messages to the S-New interface and extend the S11 interface by the two messages.

Figure 35 presents a flow diagram with control messages related to the bearer modification procedure when there is the S-New interface between the MME and the PDN GW.

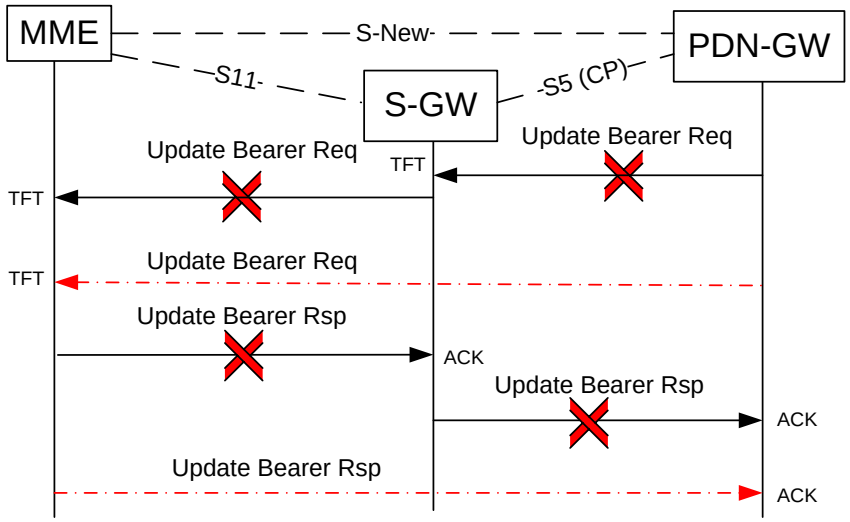


Figure 35: Bearer modification procedure with the S-New interface

The PDN GW can send the Update Bearer Request with the generated TFT and updated QoS parameters directly to the MME without the use of the S-GW as a middle point. Upon reception of the message the MME issues the Update Bearer Response acknowledging the bearer modification.

Thus, for the bearer modification process with the S-New interface we can release two signalling messages from S5 interface and two signalling messages from the S11 interface adding two messages to the S-New interface.

Handover with the S-New interface

Figure 36 presents a flow diagram with control messages related to X2-based handover for the proposed architecture with the S-New interface between the MME and the PDN-GW.

With the S-New interface, the MME can send the Modify Bearer Request message simultaneously to the S-GW and the PDN-GW with different context information. The message sent to the S-GW contains the eNodeB address and TEID for downlink user plane. The Modify Bearer Request sent to the PDN-GW contains the information required by the PDN-GW, for instance, user location IE, UE Time Zone IE, serving network IE, etc. [20].

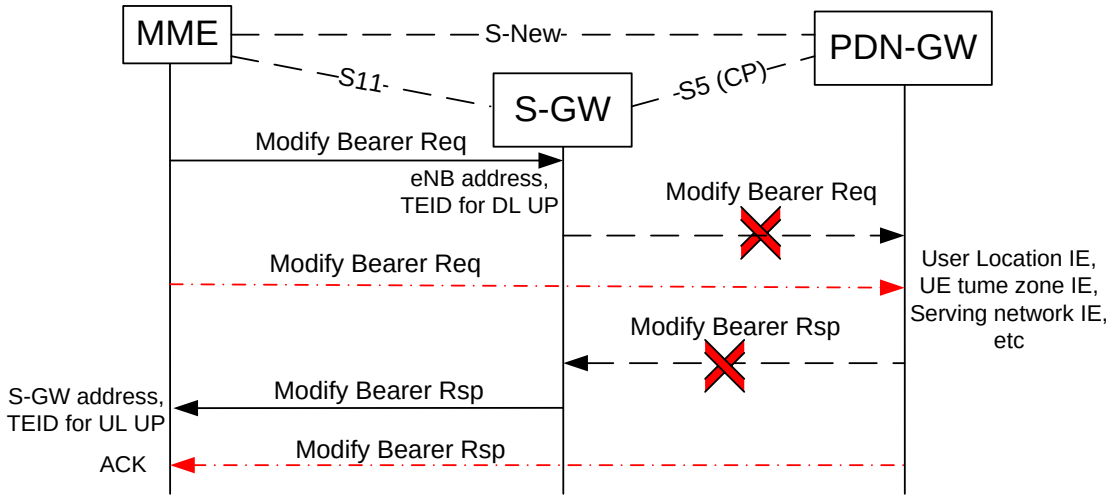


Figure 36: X2-based handover with the S-New interface

The S-GW returns a Modify Bearer Response to the MME with the S-GW address and TEID for uplink user plane. The PDN-GW acknowledges the MME with the Modify Bearer Response. Thus, for the handover process we can release two signalling messages from S5 interface moving them to the S-New one.

Control messages reallocation in the new architecture

A summary of control messages reallocation related to the attach and mobile procedures considered above is presented in Table 8.

Procedure	**Signalling messages**		
	S-New interface	*S11 (extension)*	*S5 (Released)*
Default bearer establishment	Create Session Req/Rsp, Modify Bearer Req/Rsp	Modify Bearer Req/Rsp	Create Session Req/Rsp
Dedicated bearer activation	Create Session Req/Rsp	Create Session Req/Rsp	Create Bearer Req/Rsp
Bearer modification	Update Bearer Req/ Rsp	-	Update Bearer Req/Rsp, Update Bearer Req/Rsp (S11 released)
X2-based handover	Modify Bearer Req/ Rsp	-	Modify Bearer Req/Rsp

Table 8: Signalling reallocation to release S5 (CP)

The procedures can be supported by means of regular signalling message exchange through the new proposed interface with some extension of S11 interface functionality to release the S5 interface from control plane (CP) operation. For instance, in the considered procedures (see Table 8), ten signalling messages are sent through the new interface and four additional messages are sent through the S11 interface (from the MME to the S-GW) to release ten control messages that are currently exchanged through a combination of S5 and S11 interfaces.

Thus, we have considered some high-level aspects related to a network architecture design enhancement towards a full decoupling of control signalling from user data traffic and signalling changes it can require. In particular, we revealed how signalling exchange related to both the attach and mobility procedures can be supported by means of the new interface between the MME and the PDN-GW with some extension of the functionality provided through S11 interface. As a result of the analysis, we conclude that much less changes than one could initially have expected are needed to apply the proposed design principle for network architecture evolution. It enables a basis for effective deployment of advanced 3GPP concepts related to multihoming and offloading scenarios (e.g. S1-flex, IFOM, MAPCON [17], [18], [19], [23]).

3.5.4 Summary on decoupling control and user plane signalling in the mobile architecture

The goal of the section, as stated at the beginning, is to consider design aspects of the mobile network architecture to decouple control from user plane and estimate how much of signalling changes it requires on the control plane level to manage connectivity and mobility procedures. This approach affects the standard signalling and some modifications have been proposed in the section. The number of new signalling messages has been counted and the number of removed signalling messages has been computed as well. However, a complete analysis to understand whether the total signalling is reduced had not been done in this section – this question is out of the scope of the section itself and it is subject to further studies.

4 TRAFFIC ANALYSIS AND MODELLING

4.1 Survey of traffic archives

This section provides a summary of some of the most important online archives of Internet traffic measurements. In particular, archives are listed with information about accessibility, data type, link capacity, etc. For further information, links to websites are provided too.

4.1.1 CAIDA

Web site: http://www.caida.org/home/

These datasets contain anonymized traffic traces coming from high-speed Internet backbone links located between San Jose and Los Angeles and between Chicago and Seattle; in both cases, measures are divided in two directions. The anonymized traces are available from 2008 to 2013 and they are stored in pcap format with timestamps truncated to microseconds. The original nanosecond timestamps are provided as separate ASCII files alongside the pcap files (times files).

4.1.2 CRAWDAD

Web site: http://crawdad.cs.dartmouth.edu/data.php

CRAWDAD (Community Resource for Archiving Wireless Data At Dartmouth) is a wireless network data (including cellular and Bluetooth) available for the research community. CRAWDAD requests registration to enable the user downloading data. Available data are composed of data traces, a lot of tools for traffic management and some publications related to the data. Traces are in tcpdump format.

4.1.3 WITS

Web site: http://wand.net.nz/wits/catalogue.php

WITS (Waikato Internet Traffic Storage) brings together a multitude of different data recorded in New Zealand. Traffic traces were provided by some university (Auckland and Waikato) and some unnamed ISP. The IP addresses were anonymized using Crypto-Pan AES encryption. Archives are available for free but you can find some problems if the download is not done with an IPv6 address, that is why trying downloading with common IPv4 address may be useless. Still, there is a part of WITS database available on the site http://www.ripe.net that can be downloaded through ipv4 address.

4.1.4 MAWI Working Group Traffic Archive

Web site: http://mawi.wide.ad.jp/mawi/

This web page contains a large amount of traffic traces, most coming from trans-Pacific line (connecting US with Japan) and also some traces about IPv6. Data are daily updated and it is possible to download (without registration) full years of measures divided day-to-day.

4.1.5 UMass Trace Repository

Web site: http://traces.cs.umass.edu/index.php/Network/Network

This is a big free archive with lots of trace files measured on many different links. These traces are not sniffed from real user connections, so this is a simulation made for testing network performance. You have to pay attention because files are sorted by associated publication year, not by trace collection year. Each file has its own documentation text with date and hour of registration and explanation of the kind of link involved.

TCP packet and Ethernet frame (2013)

The network topology for these experiments is shown in Figure 35 below:

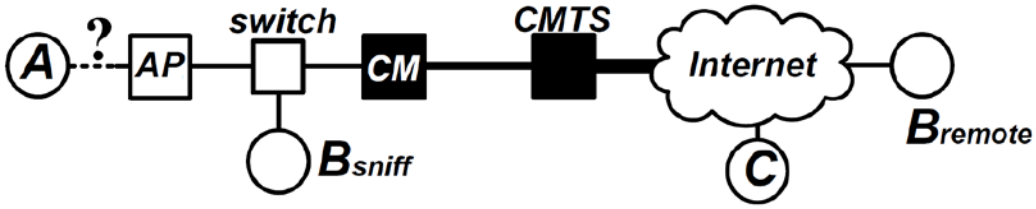


Figure 37: Network topology for the monitoring experiments

There are two monitoring points: B_{sniff}, and B_{remote}. Node A generates TCP traffic to B_{remote} using iPerf to transmit TCP data at the maximum rate possible; hence the TCP transmit queue is never starved for data. At B_{sniff} there was a sniffer that captured frames before they were transmitted via the CM. These experiments were using only traces gathered at B_{remote}. The number of hops between A and B_{remote} varied, as determined via the *traceroute* path reporting software tool. Node C is used as the sink for TCP flows originating at A that compete with traffic to B. Each above experiment setting was performed ten times for 10 sec, 1 min, or 10 minutes. Target flow is captured via tcpdump at B_{sniff} and B_{remote}.

Cellular Phone GPS, Signal Strength, and TCP Data (2012)

This contains a collection of TCP (pcap) and GPS/signal strength (gpx) traces. The files were generated by streaming music to mobile phones in the Amherst area between April and November 2012. The traces contain only the TCP headers, and not the payload.

Measurements were based on four Android cell phones instrumented to record traces of GPS location and signal strength. A server streamed music continuously to the phones during measurement trials, TCP traces were logged at the server at the same time.

Two sets of traces are available:

- Mobile 3G Measurement Set: Samsung Nexus S, Samsung Galaxy S, Motorola Atrix, and HTC Inspire phones, all connected to the AT&T UMTS (3G) network, were used to record traces. The 802.11 radio on the phone remained powered off during the experiments. Data were collected during a one-month period under varying traffic and weather conditions.
- Stationary 3G Measurement Set: traces from stationary phones, connected to the UMTS (3G) network, located in different locations

UMass DieselNet

Many traces were sniffed in this testbed network, that is also open for public experiments.

It currently consists of 35 buses each with a Diesel Brick, which is based on a HaCom Open Brick computer. The brick is connected to three radios: an 802.11b Access Point (AP) to provide DHCP access to passengers and passers-by, a second USB-based 802.11b interface that constantly scans the surrounding area for DHCP offers and other buses, and a longer-range MaxStream XTend 900 MHz radio to connect to throwboxes (throw-boxes are small, inexpensive devices equipped with wireless interfaces and deployed to relay data between mobile nodes). Additionally, a GPS device records times and locations. You can run experiments on DieselNet on http://geni.cs.umass.edu/dome .

4.1.6 Universita' degli Studi di Napoli "Federico II"

Web site: http://traffic.comics.unina.it/Traces/ttraces.php

Little free archive created for academic purpose. The link observed is a link at 200 Mbps connecting the University of Napoli "Federico II" network to the rest of the Internet. These traces are in tcpdump format. Packet lengths are variable because, for each packet, full TCP headers are stored, including optional headers (e.g. MSS). Moreover, because of privacy concerns, IP addresses have been replaced preserving network membership.

4.1.7 Simple Web - Dropbox Traces

Web site: http://www.simpleweb.org/wiki/Traces

As described in the paper, the data was captured at 4 vantage points in 2 European countries. The first 4 files were collected from March 24, 2012 to May 5, 2012. A second dataset was collected in Campus 1 in June and July 2012.The data was captured using Tstat, an open source monitoring tool developed at Politecnico di Torino. All IP addresses in the datasets are anonymized. Tstat was installed at 4 vantage points in 2 European countries and collected data from March 24, 2012 to May 5, 2012. This setup provided a pool of unique datasets, allowing to analyse the use of cloud storage in different environments, which vary in both the access technology and the typical user habits. Further information available at: http://eprints.eemcs.utwente.nl/22286/01/imc140-drago.pdf

4.2 Traffic Models

In practical networks the evaluation of the performance can be very complicated. In particular, in order to predict the performance of a network, it is important to be able to properly model the input traffic.

Moreover, the proliferation of wireless technologies has introduced a new era of convergence providing aggregation of contrasting devices and access technologies. These changes have also led to the integration of different services such as voice, video and broadband data services through seamless integration of wireless and fixed networks [23]. Technological development, market demands and deregulations made it possible to introduce the changes required to develop the aforementioned convergent environment (FMC). Such convergence yields the need of traffic models capable of accurately characterizing the behaviour of complex network traffic.

Several models of traffic have been proposed for all kinds of network environment. Each model has a different impact on network performance evaluation. Within such models, two major focuses are the packet length distribution and the packet inter-arrival time

distribution (that is, the *time correlation* of traffic). To accurately model the traffic, in particular IP traffic, both the packet length and the packet inter-arrival distributions are required to be properly characterized.

Moreover, modelling heterogeneous network traffic remains an open issue, where the main challenges are due to the different structures and features in the traffic behaviour for interconnected (IP based) communication networks.

If we focus on time correlation of traffic, two main categories of traffic models can be identified: *non-self-similar traffic models* (both memoryless, or Poisson-based, and with memory, or time-correlated) and *self-similar traffic models*.

Memoryless traffic models based on Poisson processes, e.g. packet arrivals or phone calls modelled as Poisson events, are the oldest models. They are based on the well-known result from stochastic processes theory that large aggregations of (even non-Poisson) traffic from independent sources tend to form a Poisson process.

Anyway, it has been observed that real network traffic may exhibit bursts, contrary to Poisson-type processes. For this reason, new models were proposed to describe bursty traffic, including appropriate time correlation.

More recently, the intriguing property of self-similarity (fractal behaviour) has been widely detected in network traffic measurements. This property seems ubiquitous and independent of how large the aggregation of traffic sources is. Its origin has been debated lengthy, but the empirical evidence is that network traffic exhibits self-similarity at all network segments, although not in all cases.

Next sections outline the characteristics of traffic models considering self-similarity and not. Moreover, a separate section is devoted to the mobile (or wireless) traffic modelling, due to some specific features of this type of networks.

4.2.1 Non-Self-Similar Traffic Models

Non-self-similar traffic models generally operate under the assumption that the traffic is aggregated from a large number of independent traffic sources. Thus, aggregation tends to smooth out bursts. Moreover, as the number of sources increases, the burstiness and the overall time correlation decrease.

4.2.1.1 Poisson Model

The Poisson model [24] is one of the most widely used traffic models in particular to study the traditional telephony networks. Moreover, with a proper selection of the parameters, a Poisson traffic model can fit many network traffic traces reasonably well for short periods.

The traffic Poisson model is a truly memoryless traffic model and is characterized as a renewal process, where the packets inter-arrival times are exponentially distributed with a rate parameter λ and a mean value of $1/\lambda$. Packet arrivals over a certain interval $[t_0; t_1]$ with $t = t_1 – t_0$ have a Poisson distribution with mean $t\lambda$.

The two main assumptions made by the Poisson model are that the number of sources that generate the traffic are infinite and independent. Indeed, the Poisson traffic distribution can be efficiently used in scenarios where the traffic is generated by a very large number of independent sources.

Among the mathematical properties of the Poisson processes, one of the most important is that the superposition of independent Poisson processes results in a new Poisson process with a rate which is the sum of the rates of the independent Poisson processes. Another fundamental property, also called *independent increments*, is that the number of arrivals in disjoint intervals is statistically independent.

4.2.1.2 Compound Poisson Traffic Models

The compound Poisson model is based on the Poisson model where batches of packets are delivered at once. In this case, the inter-batch arrival times are exponentially distributed and the size of batches is geometric.

This model is characterized by two parameters: λ which is the arrival rate, and ρ which is the batch parameter and is defined in $(0,1)$. In such traffic model, the mean number of packets in a batch is $1/\rho$, and the mean inter-batch arrival is $1/\lambda$. The mean packet arrivals over a certain interval $[t_0; t_1]$ with $t = t_1 – t_0$ are $t\lambda/\rho$.

Like the Poisson traffic model, also the compound Poisson model is memoryless, at least from batch to batch, and the aggregation of different streams is still a compound Poisson traffic distribution.

When this model was proposed, it was not convincing since true back-to-back packet arrivals were rare. However, modern CPUs are usually fast enough to saturate the outgoing links and to create real batches of packets. Therefore the compound Poisson traffic models should be efficiently used.

4.2.1.3 Markov Modulated Poisson Traffic Model

In general, the Markov process models the activities of a traffic source on a network using a finite number of different states. In Markov Modulated Poisson Processes (MMPPs) [24], a continuous-time Markov chain determines the arrival rate in a Poisson model.

For example, the Markov chain is a two-state chain, each with an associated rate λ_i, where i defines the state of the Markov chain. Each state is characterized by a distinct mean sojourn time r_i. Therefore, a MMPP is defined by the set of parameters $(\lambda_1,\lambda_2,r_1,r_2)$. Such model can also be extended to more than two states in the Markov Chain, and the algebraic solution to compute the mean delay of a packet arrival which is used in a two-states MMPP can still be used.

One of the characteristics of the Markov-Modulated Poisson Traffic models is that, like for basic Poisson processes, if another stream with rate λ’ is added to a MMPP defined as $(\lambda_1,\lambda_2,r_1,r_2)$, the resultant traffic is still a MMPP characterized by $(\lambda_1+\lambda',\lambda_2+\lambda',r_1,r_2)$.

4.2.2 Self-Similar Traffic Models

Since the non-self-similar traffic models suffer from a lack of large time-scale analysis, self-similar traffic models have been studied in literature. A self-similar traffic is characterized by distributions that have the same characteristics at all time-scales. Therefore, a self-similar traffic will continue to show bursts at all time-scales. This behaviour is opposite to the behaviour of the non-self-similar traffic, e.g., the Poisson traffic model.

4.2.2.1 Self-Similarity and Long-Range Dependence

Internet traffic exhibits *self-similarity* (SS) and *long-range dependence* (LRD) [28][30]. These properties emphasize long-range time-correlation between packet arrivals. Fractional noise and fractional Brownian motion models are often used to describe such behaviour of Internet traffic series, e.g. cumulative or incremental bit count transmitted over time.

In an SS random process, a dilated portion of a realization, by the scaling Hurst parameter *H*, has same statistical characterization as the whole. On the other hand, LRD is a long-memory property, usually equated to an asymptotic power-law decrease of the power spectral density (PSD) as $\sim f^{-\alpha}$ (for $f \to 0$) or, equivalently, of the autocovariance function. Under some hypotheses [28], the integral of a LRD process is SS with *H* related to α (e.g., fractional Brownian motion, integral of fractional Gaussian noise).

It has been well recognized [31]−[34] that traffic LRD yields long queues in network buffers. In the case of fractional Gaussian traffic, for example, it has been found [31][32] that the queue tail is Weibull-distributed, i.e. the buffer occupancy *Q* exceeds a given threshold *q* with asymptotic probability $P\{Q > q\} \sim exp(-\beta q^{1-\alpha})$, where β is a positive function of α and of other network parameters. The Weibull queue length distribution departs significantly from the plain exponential result in the case of Poisson input ($\alpha = 0$). In particular, when α tends to 1, the Weibull distribution flattens and average and variance of the queuing delay even tend to infinite.

Self-similarity (or LRD) is a property of network traffic that is almost ubiquitous and very hard to remove or control. The presence of LRD traffic causes the network engineer to plan for larger buffers or higher-capacity links, compared to figures designed according to traditional network design algorithms based on simple Poisson traffic models.

In particular [47], a random process $X(t)$ (where $X(t)$ is the cumulative packet arrivals in time interval $[0, t]$), is said to be *self-similar*, with scaling parameter of self-similarity or Hurst parameter $H>0$, $H \in \Re$, if

$$X(t) =^{d} a^{-H} X(at) \qquad (1)$$

for all $a>0$, where $=^{d}$ denotes equality for all finite-dimensional distributions [27][28][29]. In other terms, the statistical description of $X(t)$ does not change by *scaling* simultaneously its amplitude by a^{-H} and the time axis by a.

In practice, the class of self-similar (H-SS) processes is usually restricted to that of *self-similar processes with stationary increments* (or H-SSSI processes), which are "integral" of some stationary process. For example, consider the δ-increment process of $X(t)$, defined as $Y_\delta(t) = X(t)-X(t-\delta)$ (say, packet arrivals in the last δ time units). For a *H*-SSSI process $X(t)$, $Y_\delta(t)$ is stationary and $0 < H < 1$ [28][29][58].

Long-range dependence (LRD) of a process is defined by an asymptotic power-law decrease of its autocovariance or equivalently PSD functions [27][28][29]. Let $Y(t)$ be a second-order stationary stochastic process. The process $Y(t)$ exhibits LRD if its autocovariance function follows asymptotically

$$R_Y(\delta) \sim c_1 |\delta|^{\gamma-1} \quad \text{for } \delta \to +\infty, 0 < \gamma < 1 \qquad (2)$$

or, equivalently, its power spectral density (PSD) follows asymptotically

$$S_Y(f) \sim c_2 |f|^{-\gamma} \quad \text{for } f \to 0, 0 < \gamma < 1 \qquad (3).$$

In general, a random process with non-integer power-law PSD is also known as fractional (not necessarily Gaussian) noise. It can be proven [28][29] that H-SSSI processes $X(t)$ with $1/2 < H < 1$ have long-range dependent increments $Y(t)$, with

$$\gamma = 2H - 1 \qquad (4).$$

The Hurst parameter of a LRD process (characterized by parameter γ) denotes the Hurst parameter $H = (\gamma+1)/2$ of its integral H-SSSI parent process.

By definition, LRD consists in a power-law behaviour of certain second-order statistics versus the duration τ of the observation interval and several techniques exist to estimate H and γ of series of data supposed to be self-similar or LRD.

4.2.2.2 Fractional Brownian Motion

Fractional Brownian Motion (FBM) [25][59], also known as the “Random Walk Process”, has been widely used for its simplicity. It consists of steps in a random direction with a step-length with particular characteristic values. FBM processes have power-law PSD as $\sim f^{-\alpha}$ (for $f \to 0$) with $\alpha = 2$.

4.2.2.3 SWING

A problem with self-similar traffic models is that, to estimate the self-similarity parameters from real network measurements, a very large amount of data and extended computation are required. SWING [26] is a self-similar traffic generator that accurately captures the packet interactions of a range of applications using a simple structural model.

The model starts from the observation of the traffic in a single point of the network and it automatically extracts the distributions for application, user, and network behaviour. Then it generates a traffic corresponding to the considered model and it is able to reproduce the burstiness of the traffic across a range of different timescales.

4.2.3 Traffic Models for Wireless Access Networks

With the evolution of cellular mobile system in the last decade, the problem of traffic modelling and analysis of mobile traffic over macro, micro and picocells has become a new research challenge [79].

Within mobile network systems, classical analytical models may not apply [84]. Indeed, interactions among multiple interfaces [79], picocells and femtocells environment [77][78] traffic modelling [80]-[82] and user mobility modelling [81]-[83] (including handoff processes) pose various challenges in designing integrated networks. Therefore new traffic models especially designed for mobile traffic need to be investigated.

4.2.3.1 An Introduction to Traffic Modelling and Analysis for Wireless Networks

There is an extensive body of literature on traffic modelling and analysis for cellular networks. Because of handoffs between cells, standard models applied for decades in telephony had to be adapted. Furthermore, with different traffic types, networks with

different types of cells (e.g., mixed microcells and macrocells) and the introduction of features like guard channels, and so on, the problem can be quite challenging.

In the great majority of the papers and books, Markov models are used because of the ease in applying standard queuing theory analysis and results to the problem. Even if the actual traffic characteristics may not be exactly Poisson, it is hoped that they are close enough so that the analytical or simulation results obtained by using a Poisson process model are still meaningful.

The Poisson process assumption has been challenged by some researchers (especially the handoff traffic, rather than new call traffic [90]), as a result of which newer models have been proposed that are supposedly more accurate. One example is the proposed use of a Hyper-Erlang distribution for the inter-handoff time [53], rather than an exponential distribution. Closed-form analytical expressions (like the Erlang B and Erlang C formulas) are often not achievable (even when using Poisson assumptions) because modelling of wireless networks is far more complex than modelling standard telephone networks. Thus, in general, computer simulations are used to obtain teletraffic performance of wireless systems.

One of the main thrusts of research is how to differentiate the handling of handoff traffic and new call traffic. Handoff calls are often given higher priority than new calls, since subscribers would find it more annoying if existing calls are dropped when handoff fails, than for new calls to be blocked [98]. Several methods for such prioritization [91] have been proposed. One method uses guard channels [95][96]. Another method to reduce dropped-call probability is known as a channel borrowing [93]. Channels are borrowed from neighbouring cells to accommodate handoff calls.

The complexity of the standard queuing system model grows exponentially as the number of states increases, i.e., as the number of channels and cells increases. Approaches to simplifying the teletraffic analysis problem to something more computable have generally focused in simplifications to the model itself, or on ways to simplify the computations for a given model.

Of the attempts to simplify the model itself, making many assumptions related to homogeneity and loose coupling, some researchers have attempted to obtain useful information about a whole network by using a one-cell model. The idea is that if the coupling between cells is “loose enough”, it suffices to pick a single cell in the network and examine the teletraffic dynamics of that cell alone to have a good idea what is happening in the network. The cell can be arbitrarily chosen, based on homogeneity arguments. The incoming/outgoing traffic from/to neighbouring cells may be approximated as Poisson processes [92]. It may be shown that such an approach allows simplification to a well-known M/M/m/m Markov model [92][97], whose closed-form solution is easily available from standard queuing theory. However, the accuracy of single-cell models is very questionable for sophisticated wireless networks that employ channel borrowing, dynamic guard channels, etc. Furthermore, it is not useful for multi-layer networks.

4.2.3.2 Telephone Traffic Modelling in Mobile Cellular Networks

In classic telephone traffic theory, developed for wired networks, call arrivals to a local exchange are usually modelled as a Poisson process, at least over short observation intervals to assume stationary arrival rate, since the user population served by the

exchange is very large and with negligible correlation among users. This assumption of memoryless traffic has been often retained also in presence of mobile users: in literature, incoming calls in cellular networks are mostly modelled as a Poisson process, with both call holding time and interarrival time assumed with negative exponential distribution.

Nevertheless, as discussed in [49], it has been argued that this Poisson assumption might not be valid in wireless cellular networks for a number of reasons:

- cells partition the user population in small sets, each served by a small number of channels and with possible correlation between users;
- congestion and repeated call attempts, more likely with radio access impairments, are major causes of peaks and bursts in offered traffic and of levelling off the carried traffic;
- user mobility during calls (handover) adds further complexity to the problem.

Therefore, not surprisingly, traffic characterization in wireless cellular networks has been attracting a lot of attention in literature since few years.

In most cases, researchers focused on characterizing the channel holding time or the call holding time, sometimes based on empirical data. In many studies, the channel holding time has been modelled by negative exponential distribution. Nevertheless, several other works contradicted this simple assumption. In papers [50][51], for example, the probability distribution that better fits empirical data, by the Kolmogorov-Smirnov test, was found to be a sum of lognormal distributions.

On the other hand, the channel holding time is affected by user mobility, characterized by the cell residence time. With exponentially-distributed call holding time, the merged traffic from new and handoff calls is Poisson if and only if the cell residence time is exponentially distributed too [52]. For cell residence time having general distribution, the channel holding time distribution was derived analytically in [53][54]. The channel holding time distribution was also studied in [55], when the cell residence time has Erlang or Hyper-Erlang distribution.

As for the correlation between call arrival times, the distributions of the channel idle time (time between the end of an answered call and the beginning of the next one on the same channel) and of the call interarrival time in a Public Access Mobile Radio (PAMR) cellular system were investigated in [56]. In that work, the former distribution (channel idle time) was approximated by the Erlang-*j*,*k* function and the latter distribution (call interarrival time) resulted different from the Poisson negative exponential distribution.

More recently, a further empirical study on real GSM telephone traffic data was reported in [57]. Answered call holding and interarrival times were found to be best modelled by the lognormal-3 function, rather than by the Poisson negative exponential distribution.

In summary, *several studies contradicted the ubiquitous likelihood of the classic Poisson model for telephone traffic in cellular networks and suggested that call arrivals may be significantly time-correlated, due for example to access congestion, user mobility and possible correlation between nearby users.*

However, we note that the Poisson traffic model is still assumed in almost all works, mainly for the sake of simplicity, when cellular network performance is evaluated. Questions may arise, therefore, on the practical relevance of this simplifying assumption.

In paper [49], authors investigated possible time-correlation of both originated and terminated answered call arrivals in sets of real GSM telephone traffic data, provided by an Italian operator. After examination of such empirical data, the authors found that call arrivals proved excellent consistency, by MAVAR analysis and χ^2-test evaluation, with a *non-homogeneous Poisson model with diurnal variable rate*.

Results presented in [49] confirm, at least to the limited extent of those empirical data, that the Poisson model may be still adequate to describe realistically telephone traffic in cellular networks, unless focusing specifically on particular issues like small user population, access congestion and very frequent handovers.

4.2.3.3 Models based on Poisson distributed call arrival with handoff processes

In [80], a traffic model for cellular mobile telephone systems, where handoff procedures are considered, is presented.

In the system model, the new call origination rate is uniformly distributed over the mobile service area. It is also assumed that a very large population of mobiles is served, and the average call origination rate is independent of the number of calls in progress.

Moreover, in the proposed model handoff attempts are made. The channel holding time in a cell is defined as the time duration between the instant that a channel is occupied by a call and the instant it is released alternatively by: i) the completion of a call, or ii) a cell boundary crossing by the mobile (i.e., handoff).

The time that an assigned channel would be held if no handoff is required (referred as the unencumbered message duration) is denoted as a random variable and it is assumed to be exponentially distributed. Because of the presence of handoffs in the studied system, the distribution of such random variable will differ, in general, from that of the channel holding time.

Also the speed of a mobile is assumed to be a random variable uniformly distributed in [0, V_{max}], constant during the travel of a mobile in a cell. Moreover, the model assumes that the speed and the direction of the motion of a mobile change when a mobile crosses a cell boundary, and the direction of travel is independent of speed and uniformly distributed.

When a call is originated in a cell and gets a channel, the call holds the channel until the call is completed in the cell or until the mobile moves out of the cell. Given this consideration, in the proposed model, the channel holding time (T_{Hn}) is either the unencumbered message duration (T_M) or the time for which the mobile resides in the cell (T_n).

When a call has been handed off successfully, the channel in the new cell is held until the call is completed in the cell or the mobile moves out of the cell again before call completion. Because the exponential distribution has the property to be memoryless, the remaining message duration of a call after handoff has the same distribution as the unencumbered message duration.

In this case, the channel holding time T_{Hn} is either the remaining message duration T_M or mobile residing time T_h in the cell, whichever is less. Precise mathematical model and expressions are derived in a complete manner in the Appendix A of [80].

4.2.3.4 Poisson-arrival-location model

In the Poisson-arrival-location model (PALM) [85] customer arrivals are modelled by a non-homogeneous Poisson process and move independently through a general location state space according to a location stochastic process.

PALM assumes that different users do not interact and that the network has no capacity constraints in term of the number of radio channels available for calls. Such model can be used to represent both a time-dependent behaviour and users mobility in wireless communication networks.

4.2.3.5 Models based on general distributed Poisson call arrival for general distributed handoff processes

One of the main differences between telephone fixed public switched transport networks (PSTN) and mobile cellular networks is the presence of the handoff traffic [79]. Handoff calls, in several works e.g. [80][82][86], are assumed to be memoryless Poisson processes. However, in microcellular systems, a mobile user can request handoff multiple times at various cell boundaries altering the traffic behaviour.

Authors of [79] and [87] consider that the Poisson assumption may not apply for multiple handoffs of a call, and suggest a two moment approach suitable for both traffic streams of micro- and picocells. Each of the two moments are represented by their mean and variance. Their studies assume fixed channel allocation scheme.

The two models in [88], called Probability Generating Function (PGF) and Binomial Moment Generating Function (BMGF), are more complex. They are based on semi-Markov analysis with no closed form solution and have complex algebraic solutions.

4.2.3.6 Traffic Model for calls in integrated femto-macro cellular network

A tractable traffic model for the offload from macrocell to femtocell is a research area that received little attention in the research community. The large and dense scale deployment of femtocells makes the handover issues more challenging. A well designed femtocell-macrocell integrated network can divert heavy traffic from congested macro cellular network to femtocell network.

Chowdhury and Jang [89] studied a traffic model based on macrocell-femtocell integrated network. Call inter-arrival process is assumed to be Poisson. Femtocells are randomly deployed within the macrocell coverage area.

In such model, during the handover of a call both macrocell-to-macrocell and femtocell-to-femtocell handover calls can occur. Moreover, also direct calls can arrives to the macrocell during the considered handover. A femtocell may not accept the call in case of poor signal to noise interference ratio (SNIR).

4.3 Traffic Analysis Methodology

In this study, we aimed primarily at detecting the degree of self-similarity or long-range dependency (LRD) of network traffic, because of its large impact on network resource planning. This is true in both fixed and mobile access networks, whenever buffers or link capacities need to be designed to meet given performance requirements. However, it

would be interesting to understand if traffic in new FMC networks will exhibit peculiar behaviours in terms of self-similarity or not.

We recall that, by definition, LRD of a process is defined by an asymptotic power-law decrease of its autocovariance or equivalently PSD functions [27][28][29]. For example, in the Fourier frequency domain, the PSD of a LRD process *Y*(*t*) follows asymptotically

$$S_Y(f) \sim c_2 |f|^{-\gamma} \quad \text{for } f \to 0, 0 < \gamma < 1$$

In other words, LRD consists in a power-law behaviour of certain second-order statistics versus the duration τ of the observation interval in the time domain, or the Fourier frequency *f* in the frequency domain.

In short, detecting the degree of self-similarity or long-range dependency in measured traffic data means to estimate model parameters *H* and γ, assuming a LRD model based on a power-law PSD of traffic data, with

$$\gamma = 2H - 1.$$

Several techniques exist to estimate *H* and γ of data series, supposed self-similar or LRD, both in the time domain (e.g., variance-time plot) and in the frequency domain (e.g., periodogram) [27][29][36], which are all based on measuring the slope of a linear fit in a log-log plot.

A class of more advanced techniques is based on wavelet analysis [29][35][37][30][39]. Among wavelet-based techniques, the so-called *log scale diagram* (LD) is one of the most important [29][38]. It analyses data over an interval of time scales (octaves), ranging from 1 (finest detail) to a longest scale given by data finite length. Also in this case, by observing the diagram slope, *H* and γ are estimated.

To estimate LRD parameters, we use instead the method based on the Modified Allan Variance (MAVAR), which was proposed only recently and was demonstrated to be more accurate, with finer resolution and simpler than such wavelet methods [46][47].

4.3.1 The Modified Allan Variance

This section introduces MAVAR and briefly summarizes basic properties most relevant to our aim.

In stability characterization of precision oscillators, PSD power-law models are common for phase and frequency noise, with integer values $0 \le \gamma \le 4$ found in experimental results (e.g., the phase deviation follows a random walk, when the instantaneous frequency is affected by white noise). Although values $\gamma \ge 1$ yield model pathologies, such as infinite power (variance) and even nonstationarity, this model is widely used, considering also that real-world constraints imply measurement finite bandwidth and duration. For an excellent survey on characterization of frequency stability in precision oscillators, read Rutman [60].

To circumvent such pathologies, in particular the variance increasing indefinitely with data length if $\gamma \ge 1$, a useful approach is evaluating the variance of the *M*-th derivative (supposed stationary) of the process (in wavelet analysis, this is equivalent to increasing the number of vanishing moments). In particular, the Allan Variance (AVAR), recommended by IEEE in 1971 [61] for characterization of frequency stability after D. W.

Allan [62], is a kind of variance of the second difference of phase samples. The structure function theory, developed by Lindsey and Chie [63], gives a unifying view of time-domain quantities evaluated more generally on the *M*-th difference (supposed stationary) of data. To probe further, see also [44] and [64]-[67].

In the different context of statistical inference, the estimation of parameters of fractional Brownian motions by the *K*-th moments of their discrete variations was studied in [68][69], covering theoretical aspects as robustness of estimators and convergence theorems. Although not mentioned explicitly, Allan and related variances may be set also in this more general framework.

Abry and Veitch, in their fundamental paper [38] proposing the wavelet LD, did mention the AVAR and noticed that its definition can be rephrased in terms of the Haar wavelet. Under this perspective, the second difference in AVAR definition corresponds to the two vanishing moments of this wavelet. Several properties of Allan and related variances may be derived also via the wavelet formalism. However, while these Authors recognized fine qualities of AVAR for *H* estimation, they did not investigate it further, since their method based on Daubechies wavelets outperforms it.

A substantial improvement of AVAR is the Modified Allan Variance, which has played a prominent role in stability characterization of precision oscillators since 1981 [40]-[45]. Being based on data second difference, as AVAR, it converges to finite values for all power-law noise types with $\gamma < 5$ and is insensitive to data linear drift. Nevertheless, it allows more accurate estimation of the parameter γ, over the full range $0 \le \gamma < 5$ and in particular for $0 \le \gamma \le 1$, where AVAR exhibits poor discrimination capability. These properties suggest its fruitful application also to LRD and self-similar traffic analysis and call for a thorough investigation on its usefulness in this field.

4.3.1.1 Definition and Estimator in the Time Domain

Given an infinite sequence $\{x_k\}$ of samples of $x(t)$ with sampling period τ_0, MAVAR is defined as

$$\begin{aligned} \mathrm{Mod}\,\sigma_y^2(\tau) &= \frac{1}{2n^2\tau_0^2}\left\langle\left[\frac{1}{n}\sum_{j=1}^{n}\left(x_{j+2n}-2x_{j+n}+x_j\right)\right]^2\right\rangle = \\ &= \frac{1}{2}\left\langle\left[\frac{1}{n}\sum_{j=1}^{n}\left(y_{j+n}-y_j\right)\right]^2\right\rangle \end{aligned} \qquad (5)$$

where <·> denotes infinite-time averaging, $\tau = n\tau_0$ is the observation interval and y_k is the average value of $y(t) = x'(t)$ over interval τ beginning at t_k, i.e. the k-th sample of the first difference of $\{x_k\}$ with lag τ:

$$y_k(\tau) = \frac{1}{\tau}\int_{t_k}^{t_k+\tau} y(t)dt = \frac{x_{k+n}-x_k}{n\tau_0} \qquad (6).$$

Thus, MAVAR is a kind of variance of the first difference of $\{y_k\}$ or of the second difference of $\{x_k\}$ (note: differences multiplied by τ). In very brief, it differs from the unmodified Allan

variance in the additional internal average over n adjacent samples: for $n = 1$ ($\tau = \tau_0$), the two variances coincide.

In practice, given a finite set of N samples $\{x_k\}$, spaced by τ_0, an estimate of MAVAR can be computed using the ITU-T standard estimator [44][45]

$$\operatorname{Mod}\sigma_y^2(n\tau_0) = \frac{1}{2n^4\tau_0^2(N-3n+1)} \sum_{j=1}^{N-3n+1} \left[\sum_{i=j}^{n+j-1} \left(x_{i+2n} - 2x_{i+n} + x_i \right) \right]^2 \qquad (7)$$

with $n = 1, 2, ..., \lfloor N/3 \rfloor$. A recursive algorithm for fast computation of this estimator exists [58], which cuts down the complexity of evaluating MAVAR for *all* $\lfloor N/3 \rfloor$ values of n to $O(N^2)$ instead of $O(N^3)$.

The point estimate (7), computed by averaging N-3n+1 terms, is a random variable itself. Exact computation of confidence intervals is not immediate and, annoyingly enough, depends on the spectrum of the underlying noise [70]-[76]. However, in general, confidence intervals are negligible at short τ and widen for longer τ, where fewer terms are averaged. Interval width is approximately proportional to $m^{-1/2}$, with m equal to the number of averaged terms. In practice, being N usually in the order of 10^4 and above, $\operatorname{Mod}\sigma_y^2(\tau)$ exhibits random ripple due to poor confidence only at the right end of the curve.

4.3.1.2 **Equivalent Definition in the Frequency Domain**

As for other variances [67], the MAVAR time-domain definition can be translated to an equivalent expression in the frequency domain, allowing a more profound understanding of the behaviour of this quantity [43][44]. In fact, (5) can be rewritten as the mean square value of the output of a linear filter with impulse response $h_{MA}(n, t)$ properly shaped, i.e.

$$\operatorname{Mod}\sigma_y^2(\tau) = \left\langle \left[\int_{-\infty}^{\infty} y(\xi)\, h_{\mathrm{MA}}(n, t-\xi)\, d\xi \right]^2 \right\rangle \qquad (8).$$

Hence, MAVAR can be equivalently defined in the frequency domain as

$$\operatorname{Mod}\sigma_y^2(\tau) = \int_0^{\infty} S_y(f) \left| H_{\mathrm{MA}}(n, f) \right|^2 df = \int_0^{\infty} S_y(f) \frac{2\sin^6 \pi\tau f}{(n\pi\tau f)^2 \sin^2 \pi \frac{\tau}{n} f} df \qquad (9)$$

where $S_y(f) = S_x(f)\cdot(2\pi f)^2$ is the one-sided PSD of $y(t) = x'(t)$ and $H_{MA}(n, f)$ is the filter transfer function.

The square magnitude $|H_{MA}(n, f)|^2$ takes the asymptotic expression, for $n \to \infty$ and keeping constant $n\tau_0 = \tau$:

$$\lim_{\substack{n\to\infty \\ n\tau_0=\tau}} \left| H_{\mathrm{MA}}(n, f) \right|^2 = 2\frac{\sin^6 \pi\tau f}{(\pi\tau f)^4} \qquad (10).$$

This transfer function is pass-band, with a narrow main lobe at $f \cong 0.4/\tau$. Hence, MAVAR allows high-resolution spectral analysis by computation over τ. Analogously to the wavelet

LD, *H* and γ of LRD data can be estimated by observing the $\mathrm{Mod}\,\sigma_y^2(\tau)$ curve slope in a log-log plot.

4.3.2 Behaviour of MAVAR

In this section, the behaviour of MAVAR is studied with power-law random signals, drifts, periodic components and steps.

4.3.2.1 Power-Law Random Signals

It is convenient to generalize the LRD power-law model of spectral density. As customary in characterization of phase and frequency noise of precision oscillators [43][44][60][73], we deal with random processes $x(t)$ with one-sided PSD modelled as

$$S_x(f) = \begin{cases} \sum_{i=1}^{P} h_{\alpha_i} f^{\alpha_i} & 0 < f \le f_h \\ 0 & f > f_h \end{cases} \qquad (11)$$

where P is the number of noise types considered in the model, α_i and h_{α_i} are parameters (α_i, $h_{\alpha_i} \in \Re$) and f_h is the upper cut-off frequency. Such random processes are commonly referred to as *power-law noise* or *fractional noise*. Note that $x(t)$ is not necessarily assumed Gaussian in this model.

Power-law noise with $-4 \le \alpha_i \le 0$ has been revealed in practical measurements of various physical phenomena, including phase noise of precision oscillators [44][60][61][73] and Internet traffic [27][46][48], whereas P should be not greater than few units for the model being meaningful. If the process $x(t)$ is simple LRD (3), then P=1 and $-1 < \alpha_i < 0$. Finally, the case $\alpha_i > 0$ is less interesting and will be not considered in this work [29].

By considering separately each term of the sum in (11) and letting P=1, $\alpha=\alpha_i$, evaluation of frequency-domain definitions with (10) yields corresponding time-domain expressions $\mathrm{Mod}\,\sigma_{\mathrm{H},M}^2(\tau)$ and $\mathrm{Mod}\,\sigma_y^2(\tau)$. Complete formulas for MAVAR are available in [43][44]. Moreover, Rutman [60] presents a detailed overview about recognizing power-law random noise and polynomial drifts in time-domain measures, including unmodified Allan and Hadamard variances.

In summary, under the power-law PSD model (11) (P=1) and in the whole range of convergence $-1-2M < \alpha \le 0$, MAVAR is found to obey the simple power law (ideally asymptotically for $n\to\infty$, $n\tau_0=\tau$, but in practice for n>4)

$$\mathrm{Mod}\,\sigma_{\mathrm{H},M}^2(\tau) \sim A_\mu \tau^\mu \quad , \quad \mu = -3-\alpha \qquad (12)$$

If P>1, it is immediate to generalize (12) to summation of powers $\sum_i A_{\mu_i}\tau^{\mu_i}$.

This is a fundamental result. If $x(t)$ obeys (11) and assuming sufficient separation between components, a log-log plot of $\mathrm{Mod}\,\sigma_{\mathrm{H},M}^2(\tau)$ looks ideally as a broken line made of P straight segments (corresponding to the piecewise linear trend of the $S_x(f)$ log-log plot), whose

slopes μ_i yield the estimates $\alpha_i = -3-\mu_i$ of the fractional noise components dominant in different ranges of observation interval τ.

The log scale diagram [29][38] exhibits analogous behaviour. Actually, plots of log(MAVAR) versus log(τ) can be also seen as particular log scale diagrams, since modified Allan and Hadamard variances can be redefined in terms of appropriate wavelets too. In [29], log scale diagrams displaying two regions with different average slopes are referred to as *biscaling* phenomenon.

4.3.3 Using MAVAR for Estimating the Hurst Parameter

Let us consider a LRD process *x(t)* with PSD (3) and Hurst parameter $1/2 < H < 1$. Then, from (4) and (12), MAVAR follows $\sim\tau^{\mu}$ (ideally for $n\to\infty$) with $\mu = 2H-4$. In brief, the following procedure is suggested to estimate *H*:

- 1) compute MAVAR by (7), based on $\{x_k\}$, for integer values $1 \le n < N/(M+1)$ (we use a geometric progression of ratio 1.1, i.e. 24 values/decade, for finest rendering of trend);
- 2) by least-square linear regression, estimate the average slope μ of MAVAR in a log-log plot for $n>4$ and excluding highest values of *n*, where confidence is lowest;
- 3) if $-3 < \mu < -2$ (i.e., $-1 < \alpha < 0$, $0 < \gamma < 1$), get the estimate of the Hurst parameter as

$$H = \mu/2 + 2 \qquad (13).$$

Under the more general hypothesis of power-law PSD (11), then up to *P* slopes μ_i can be identified ($-3 \le \mu_i < 2M-2$) to yield the estimates $\alpha_i = -3-\mu_i$ ($-1-2M < \alpha_i \le 0$) of the *P* components of f^{α_i} noise.

Some care should be exercised against non-stationary terms in data analysed (e.g., big steps, slow trends), which cause slope changes that may be erroneously ascribed to random power-law noise. On the other hand, polynomial drifts are cancelled, unless their order is greater than *M*. Thus, the order *M* can be conveniently adjusted. In [38] (Sec. III.B.4), similarly, it is suggested to increase the number of vanishing moments until the *H* estimate converges to a stable value, thus indicating that all smooth trends have been cancelled.

A key issue is to determine the confidence of these estimates and whether they are unbiased or not. In [38] (Sec. III.C), this problem is studied for the *H* estimator based on wavelet decomposition. Provided that the number of the vanishing moments is chosen appropriately, the estimator is proven to be unbiased (or with low bias on finite data sets). Closed forms of the variance and confidence intervals of this estimator are derived as well, although under a number of simplifying assumptions. Since MAVAR can be rephrased in terms of appropriate wavelets, it can be argued that similar results may be valid also for the estimator proposed herein.

Nevertheless, deriving exact expressions for the confidence intervals of *H* and α_i estimates is not immediate. Exact computation of confidence intervals of MAVAR estimates is tedious and even depends on the spectrum of the underlying noise. Therefore, the evaluation of confidence intervals of estimates of *H* and α_i results even more complex, depending also on the algorithm used to estimate the average slope of curves and on the interval on which this is carried out.

4.4 Traffic analysis and complex modelling methodology at a macro-level

4.4.1 The cycle of model-based traffic generation

The following methodology aims at providing a model for complex control plane as well as data plane traffic at the edge of the mobile access network.

This methodology will allow to model the traffic behaviour of hundreds of thousands of users as they attach to the network, initiate user-related communication, and move within the network while producing and receiving traffic.

A major task of the methodology is traffic **analysis**, carried out in order to find proper **models** and parameters that fit this behaviour. The models are implemented in the form of a traffic generator, which produces synthetic traffic whose statistical parameters match the **observed** real-life patterns. After **verification** and fine-tuning of the model, traffic generators can be **deployed**, and operated based on the verified traffic models. This can help revealing the performance limitations of the core of FMC.

Figure 38 illustrates this methodology through a general cycle of model-based traffic generation: observation, analysis, model creation, implementation and finally verification and deployment.

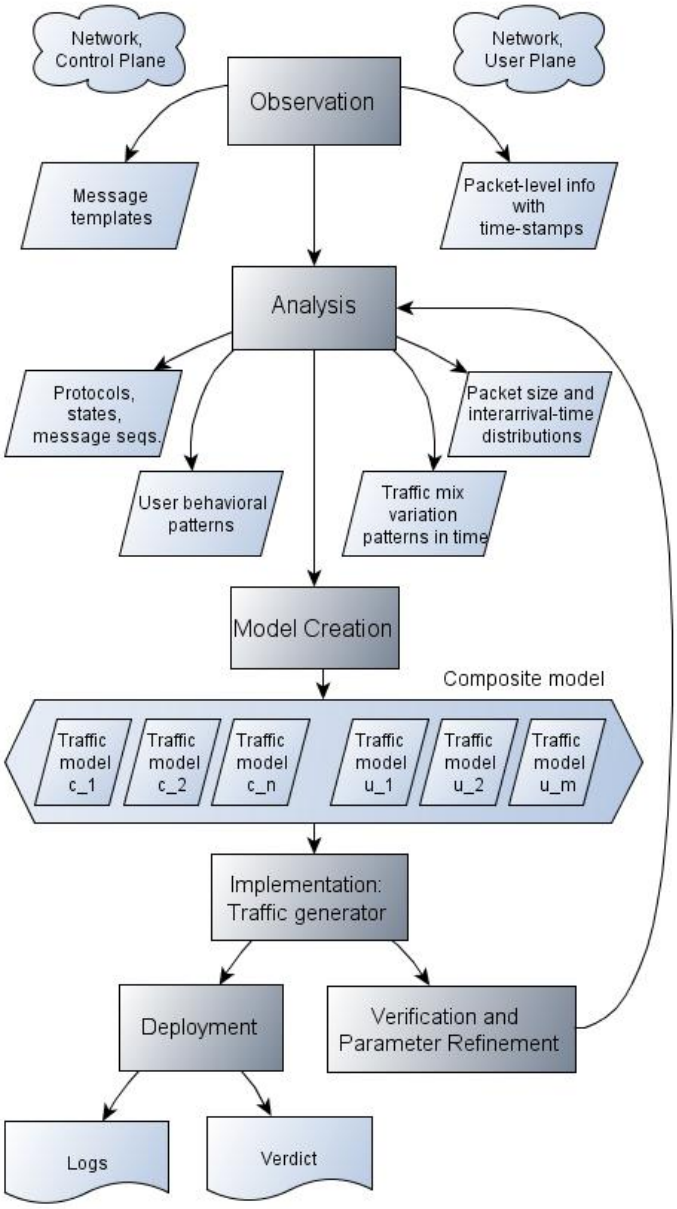


Figure 38: Tasks and their results in the methodology. Shaded grey rectangles illustrate procedures; parallelograms represent intermediate data.

Observation: Capturing of real-life traffic data in an operational network. Traces are collected from both the control plane and the user plane.

Analysis is performed on the collected data. Types of various network activities are identified. For each activity, relevant features and statistical parameters are identified, and their values are extracted from the data.

Model creation: Based on the relevant parameters and message samples models are built, which account for the various activities and traffic patterns observed in the network.

Implementation: The models are realized as a traffic generator. The device simulates the operation of a specific network segment through the parameters of the implemented models.

Verification and refinement: The statistical properties of the synthetic traffic are matched against those observed in real patterns. The models' parameters are refined in an iterative process in order to improve the prediction accuracy of the models.

Deployment: Once the models are considered sufficiently accurate, the traffic generator can be deployed in a pilot network (or live network) to carry out complex load testing tasks. At this point the models may be operated outside the previously observed realistic parameter range. Thus the device can be used to simulate extreme network activities or boundary conditions, which would otherwise be difficult or impossible to produce in a real-life setup.

Due to their distinct characteristics, control and user planes need to be addressed separately throughout the model creation process.

4.4.2 Modelling of the Control plane for traffic generation

Control plane messages are captured bit-by-bit. In the analysis phase, control message sequences and protocol state transitions are identified and stored. Subscriber mobility patterns are analysed and their relevant parameters are identified.

The models created for the control plane are mainly based on protocol specifications: message sequences and state transitions need to conform to the standards and vendor-specific extensions. Subscriber mobility models, on the other hand, can as well employ statistical parameters.

4.4.3 Modelling of the User plane traffic

User plane traffic capture is a process of collecting data packets sent over the network.

The analysis needs to identify categories of user activities (such as voice calls, video streaming, web browsing or email traffic) and define relevant parameters which characterize the activities (e.g. packet delay and jitter for VoIP and video; expected value and variance of packet sizes for email download).

Typically different sets of parameters are identified for different types of observed activities.

4.4.4 The Composite traffic model

In order to model the traffic in the actual network, a collection of different traffic models need to be elaborated, each one accounting for a particular type of user activity. That is, different models need to be used for characterizing e.g. voice calls, video streaming, web browsing or email traffic. Similarly, on the control plane different models are needed for e.g. describing subscriber attach demand or mobility within the network.

The specific models are then combined into a composite model; its structure is depicted in Figure 39). The composite model characterizes the different types of subscriber activities on the control plane and tells us how particular subscriber activities contribute to the

overall observed traffic on the user plane. In this sense the composite model is a super-model, whose parameters are the constituting models themselves.

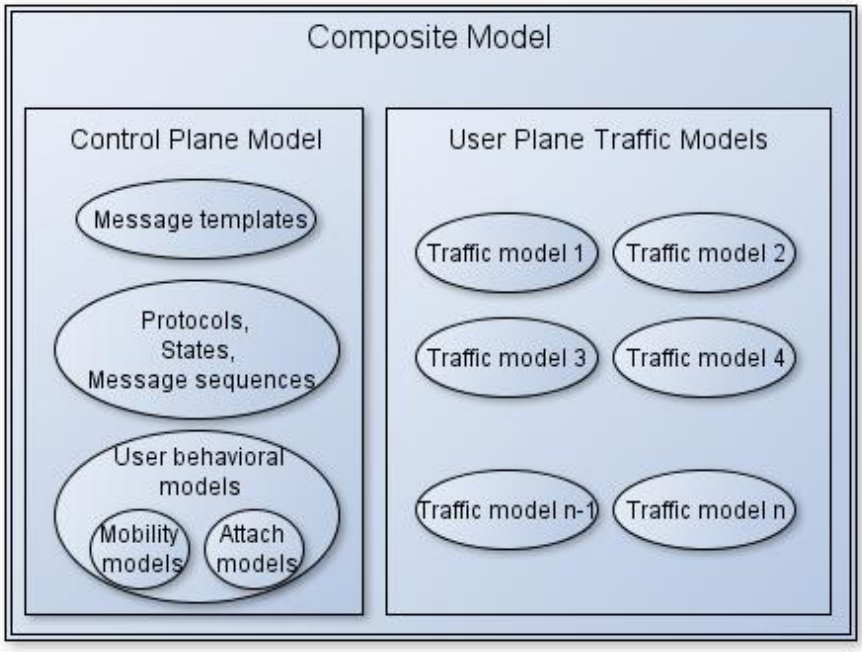


Figure 39: Elements of the composite activity model.
Different types of models describe different activities on the control and user plane.

In the following subsections the process of data capture, analysis, model creation, implementation, verification and deployment are detailed.

4.4.5 Data capture

The motivation behind traffic capture in the described method is to provide data for analysis and model creation. Two key features of the capture process are *losslessness* and *accurate time-stamping*, which are addressed below. Capture is performed at multiple network interfaces of the control and the user plane.

Control plane data capture

The collected data are practically bit-by-bit captures of control plane traffic. The capture contains actual protocol messages along with mobility information. These control messages can be collected into a set of templates, used by the implemented traffic generator in the network testing phase.

Losslessness is a key requirement here. Missing even one control message may lead to missing e.g. a temporary identifier update, which in turn may lead to loss of a whole communication session.

In general, accurate time stamping is not of paramount importance for the control messages, as long as the original order of the messages is preserved. Time stamping of the **mobility** control messages, on the other hand, carry valuable information for model

creation. Accurate time stamping is necessary in order to build a mobility model from the statistical properties of how location changes appear within the live network.

Note that control plane data account for a relatively small portion of the total traffic - the counter example being e.g. machine-to-machine communication.

User plane data capture

The actual contents of the user packets are not relevant from the model creation point of view. However, in-band signalling in the user plane needs to be captured and stored accurately, just like in the case of control plane messages.

In the case of user plane packet capture, losslessness is a less strict requirement. Actually it may be satisfactory to capture packet headers only. Whether or not headers only would suffice depends on the planned depth of the analysis, and the granularity of the modelling. As an example: traffic generation with random payload does not require payload fingerprint analysis. It is completely satisfactory to chop such packets during their capture, and merely analyse their time-stamp and the 5-tuple of “from-IP”, “to-IP”, “from-port”, “to-port”, and “protocol”.

The requirement for the accurate time-stamping is also determined by the purpose of the modelling. When the model includes packet inter-arrival time modelling (or even taking long-range dependence into account), the theoretically smallest packet inter-arrival time determines the required time-stamping accuracy for proper capture. (It is 672ns for 1Gbps Ethernet, and 6.72ns for 100Gbps Ethernet.)

4.4.6 Data analysis

The purpose of data analysis is to categorize activities in the network and identify relevant parameters for the various activities in order to build models from them. We would also like to find typical and non-typical traffic patterns so as to use them for generating varying, real-life-like synthetic traffic. The user and the control plane data should be analysed separately.

Protocol analysis in the control plane

The control plane of the mobile core is responsible for call and session management, mobility management, policy control and charging. (When referencing mobile-related behaviour, we describe them within the LTE terminology.)

Signalling on the control plane uses the message-oriented, reliable SCTP transport protocol. This ensures that the communicating devices always receive complete signalling messages in correct order.

The actual task of protocol analysis is to

- identify the dialogs to be simulated (examples include the Initial Attach, Periodic Location Update, Normal Location Update and PDP Context Activation procedures),
- collect message templates for the dialogs, and
- identify relevant message parameters.

Some subscriber activities can be characterized by statistical parameters. Such are the probability of changing location within the network, activating a PDP context or detaching from the network.

Traffic analysis in the user plane

Categorization of network activities is important for successful model creation. Data traffic is a result of parallel user activities: users are simultaneously engaged in voice calls, view video streams, browse the web, read emails, and so on.

Different user activities produce different traffic patterns. Each traffic pattern can be characterized by various parameters. These parameters are chosen so that they help building an effective traffic model for the particular traffic pattern.

From this point of view traffic analysis is a complex deconvolution problem. In the aggregate traffic flow of the user plane one needs to identify the various user activities and extract the relevant parameters, which characterize the traffic pattern produced by the particular type of activity.

Deep packet inspection at the user plane alone may not be sufficient to differentiate types of user traffic. In this case analysis of the control plane messages can give us further clue for correctly identifying user plane traffic.

User plane packets account for a significant portion of the overall traffic volume. Relevant parameters (such as packet type, size, source and destination addresses) are extracted from the data flow. The actual contents of the user plane packets can be discarded.

4.4.7 Model creation

The composite model described in this section incorporates building blocks from earlier published modelling methods, briefly summarized in the following references.

Chandrasekaran [24] gives an overview of the various traffic models. The Cisco paper [99] on VoIP traffic patterns compares various traffic models for voice calls, including Erlang B and C, Extended Erlang B, Engset, Poisson, EART/EARC, Neal-Wilkerson, Crommelin, Binomial and Delay.

The UMTS Forum Report44 [2] is built on a traffic model which distinguishes service categories (video/audio streaming, mobile gaming, etc.), device categories (smartphones, tablets, connected embedded devices, etc.) and various subscriber activity patterns. These components contribute to the overall traffic models through parameters such as traffic per service/device, device mix, upload/download direction, period of the day/week/year, and others. The traffic model selection criteria are based on the number of sources (finite, infinite), call arrival patterns (smoothed, peaked, random), holding times (exponential, constant) and what happens to blocked calls (cleared, held, delayed, retied).

Aalto et al. [100] compared various scheduling algorithms from link delay and fairness aspects, and found that scheduling algorithms in the access network have their impact on the observed traffic in the core network.

The composite model for traffic generation

This section describes the creation methodology for the composite traffic model, which consists of independent, classical traffic models. The parameters of the composite model include:

- the total number of simulated subscribers in the system,
- number of subscribers taking part in each activity (i.e. number of subscribers for each traffic model),
- triggering events/probabilities to change any parameter in any of the traffic models, and
- the rate at which subscribers start/stop particular activities (i.e. the change rate in the number of subscribers simulated by a particular traffic model, such as attach/detach rate).

In the composite model we assume that one subscriber is engaged in one activity at a time. That is, a simulated subscriber is used in only one traffic model at a time. In the pool of simulated subscribers, each subscriber is assigned to a given traffic model for a set period of time, in an on-off manner. Figure 40 depicts an example of how the user traffic models contribute to the composite model in time.

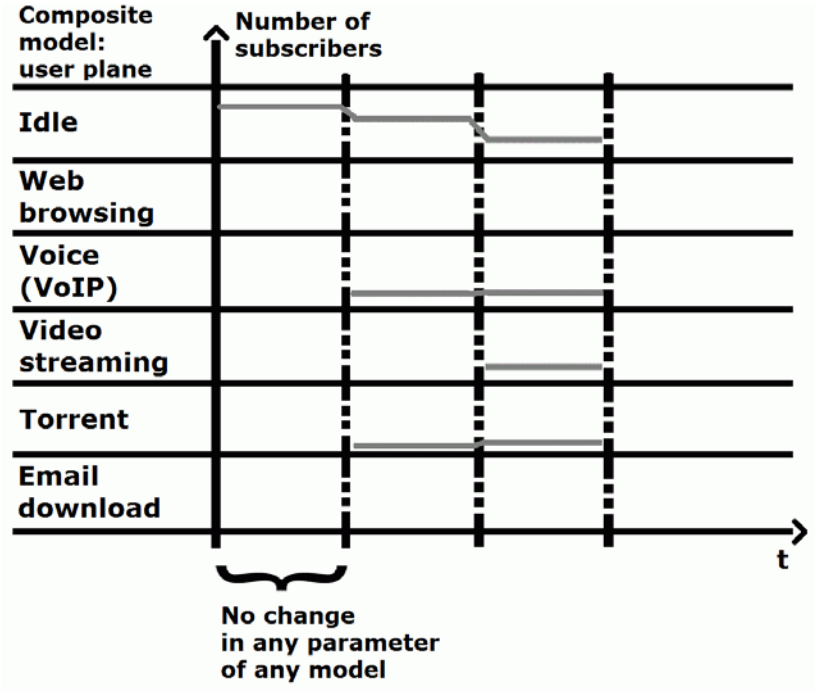


Figure 40: Superposition of user traffic models in time.

In order to simulate a limited number of subscribers, we found it practical to introduce an *Idle* traffic model. This model acts as the subscriber pool, which holds available users for the various activities simulated by the models.

The traffic model parameters are considered as target values only, in the context of other model parameters. The target may not be met if fewer subscribers are simulated than

required by a model or the traffic generator's hardware and software resource limitations are reached, or the resources of the Network Under Test are exhausted. In a load testing scenario the latter may actually be a desired event -- provided that the aim is to discover the limitations of the system.

4.4.8 Implementation

In the control plane messages, the network-, service- and user-specific parameters are filled out as required. In practice, a relatively huge part of these messages never change: many parameters in call-setup, session or mobility management are set to exactly the same value. Network- and service- specific parameters (e.g. point codes, service capabilities, expected QoS parameters) have no or limited variance; user-related parameters (e.g. endpoint identifiers, temporary codes), on the other hand, vary a lot.

From this viewpoint it is possible to build a protocol message pool in which each kind of dialog is represented by a set of message templates. In these templates there are

- parameters filled out with *fixed values*,
- variable parameters, whose values are *chosen from a range* (based on certain rules), and
- variable parameters *matching the protocol logic* (i.e. temporary identifiers, sequences, etc.).

In the user plane, individual dummy packets are generated as defined by the composite model. Actual packet lengths, and inter-arrival times are also derived from the composite model in use.

4.4.9 Verification and refinement

The general purpose of making a model is to give predictions. In our case, after the composite model is implemented, the traffic generator device is deployed and its output is matched against the earlier-captured real-life traffic.

In the control plane, the generated traffic needs to match the captured traffic as set out by the protocol standards. The statistical properties of subscriber mobility patterns need to match that of the captured patterns.

In the user plane, the statistical properties of the generated traffic need to match those of the captured data -- provided that the same subscriber activity patterns (same traffic mix) are used. Until these requirements are matching, the model parameters need refinement through further traffic analysis and fine-tuning of the parameters.

4.4.10 Deployment

Network testing in an active way is carried out by simulating nodes that

- send control-plane protocol messages to the Network Under Test (NUT),
- keep track of the status of dialogs, transactions and contexts, and
- handle user-plane traffic sent to/received from the NUT.

Depending on the type of network (segment) we actually deploy such a model-based traffic generator, the control plane protocols – including their internal logic – as well as the user plane protocols has to be made properly available in the traffic generator and tuned to the NUT.

4.5 Traffic Analysis Results

In this section, some results of MAVAR analysis over three different data sets of real IP traffic sequences, based on the procedure described in section 4.3.3, are presented.

4.5.1 Traffic Measured by AITIA (10 Gb/s link SGSN - RNC pool)

The first sequence $\{x_k\}$ (bytes per time unit, [bytes/t.u.]) was derived by processing the traffic trace provided by AITIA, measured on a 10 Gb/s link towards the SGSN from the RNC pool. This traffic is highly aggregated indeed, including traffic of all RNCs connected to the SGSN.

Figure 41 presents the sequence $\{x_k\}$ of $N \cong$ 30000 samples, acquired with time unit $\tau_0 =$ 50 µs over a measurement interval $T \cong$ 1,5 s. Figure 42 plots the histogram of the sequence. Figure 43 depicts the $\mathrm{Mod}\,\sigma_y^2(\tau)$ of $\{x_k\}$ in a log-log scale. Here, MAVAR exhibits a regular slope $\mu \cong$ -2,7115 estimated by least-square linear regression (excluding small and large values of τ). Therefore, the traffic trace seems to obey a simple power law (11) with $\alpha \cong$0,28, which is a kind of LRD process with $H \cong$ 0,64.

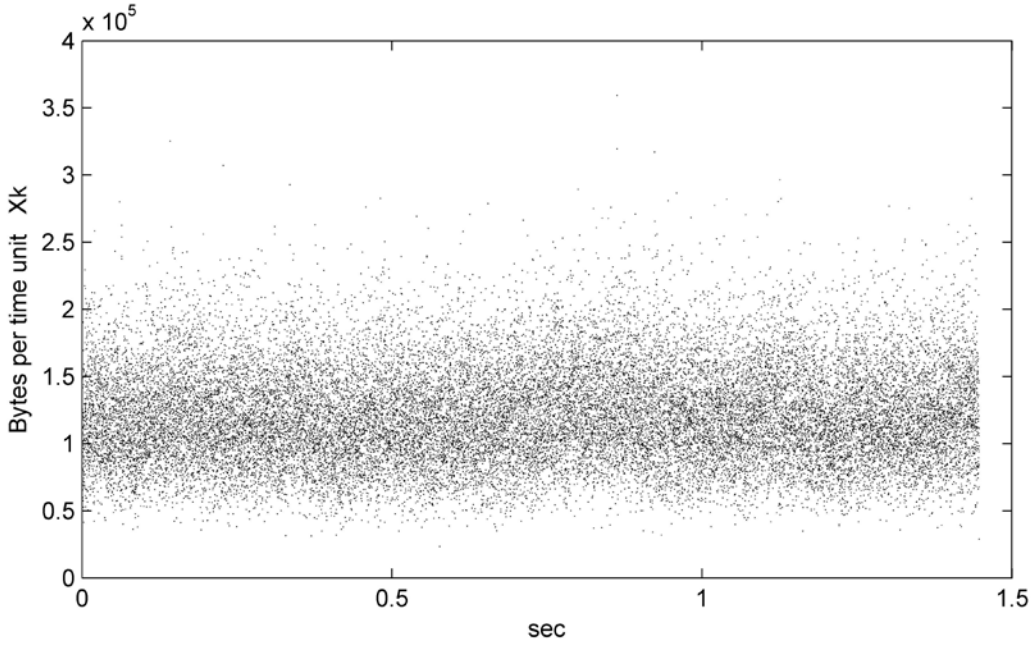


Figure 41: Traffic sequence $\{x_k\}$ AITIA 10 Gb/s link (N=30000, τ_0=50 µs, T=1,5 s).

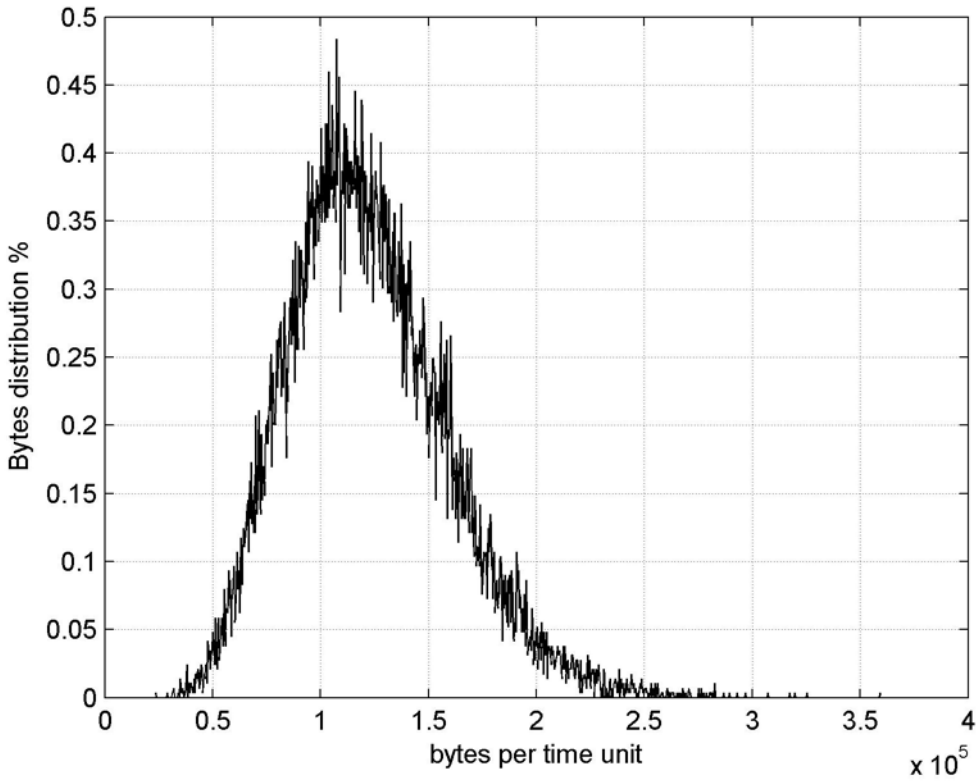


Figure 42:
Normalized histogram of {x_k} AITIA 10 Gb/s link (N=30000, τ_0=50 µs, T=1,5 s)

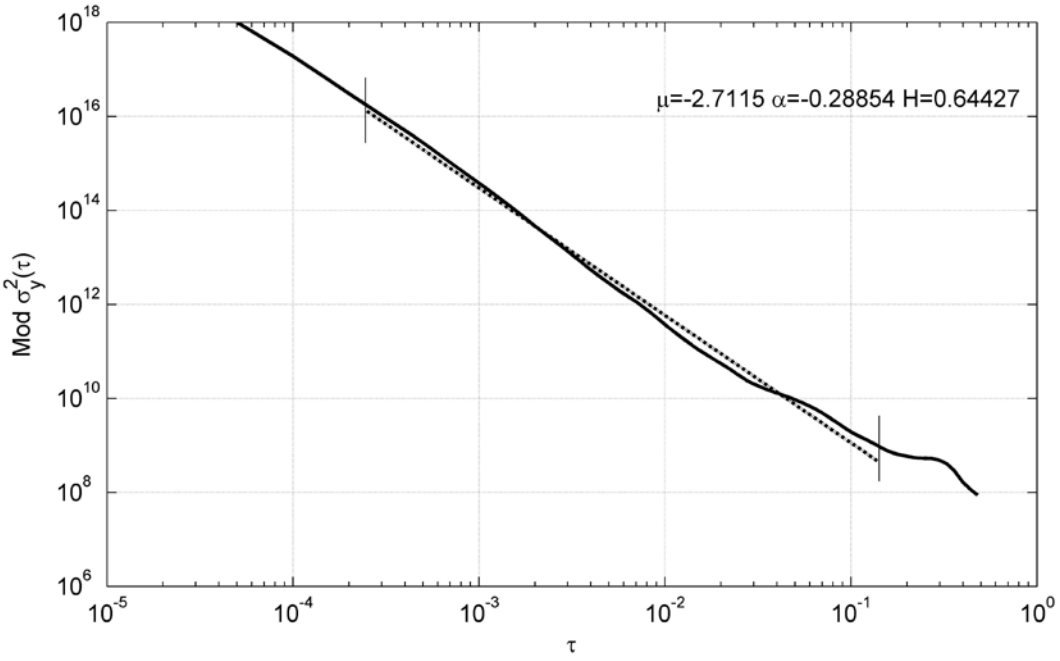


Figure 43:
$\mathrm{Mod}\,\sigma_y^2(\tau)$ of traffic sequence {x_k} AITIA 10 Gb/s link (N=30000, τ_0=50 µs, T=1,5 s).

4.5.2 Traces from the CAIDA Project Repository (10 Gb/s link Seattle-Chicago)

The second set of real IP traffic series $\{x_k\}$ [bytes/t.u.] was captured on a 10 Gb/s link connecting Seattle with Chicago in the CAIDA UCSD Project [101]. Both series are made of N=360000 samples, acquired with time unit τ_0=10 ms over a measurement interval T=3600 s, between 11:59-12:58 in two different days (day 1: 29 May 2013 and day 2: 20 June 2013).

Figures 44 and 46 depict the traffic sequences $\{x_k\}$ [bytes/t.u.] of day 1 and day 2 respectively. In figure 46 at $t \cong 500$ s a step-like behaviour can be appreciated, however MAVAR is not significantly affected by such non-stationary trend [47] in estimating the Hurst parameter of random components. Figure 45 and 47 present the histogram of the sequences $\{x_k\}$ corresponding to day 1 and day 2 respectively.

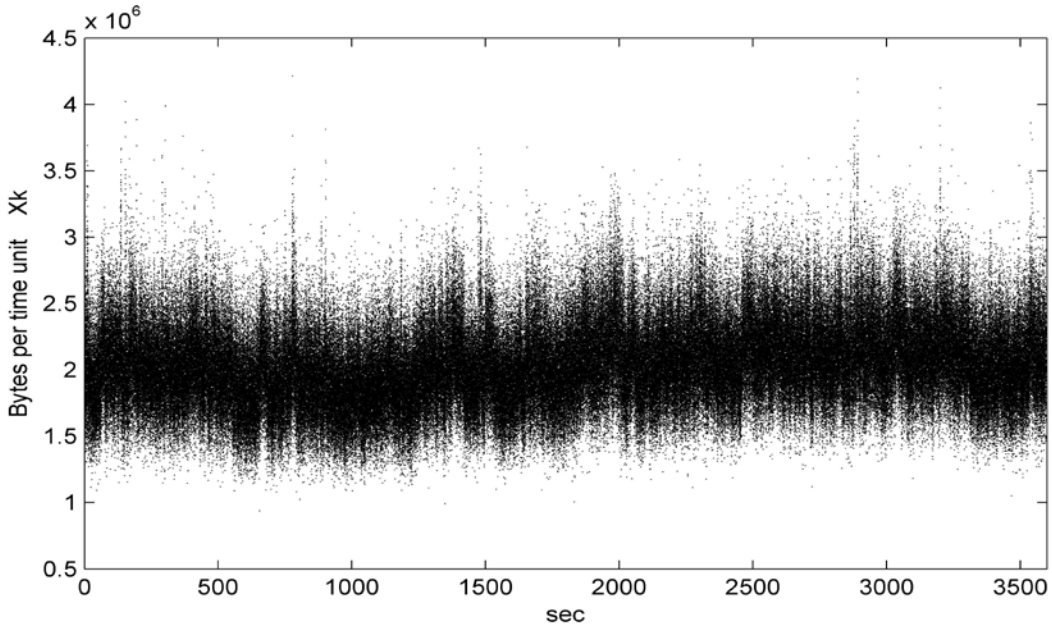


Figure 44: Traffic sequence $\{x_k\}$, CAIDA, 10 Gb/s link Seattle-Chicago, day 1 (N=360000, τ_0=10 ms, T=3600 s).

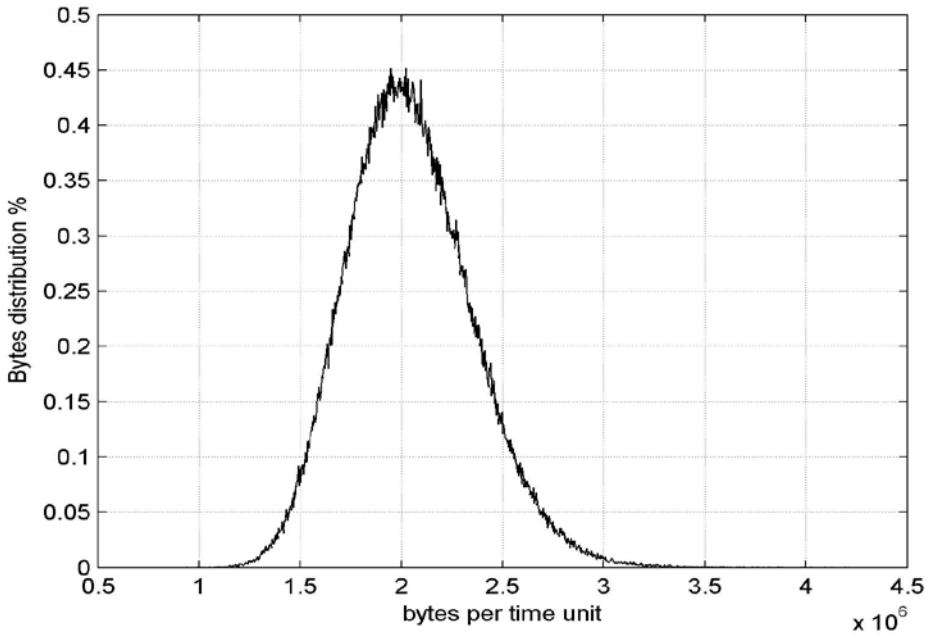


Figure 45: Normalized histogram of {x_k}, CAIDA, 10 Gb/s link Seattle-Chicago, day 1 (N=360000, τ_0=10 ms, T=3600 s).

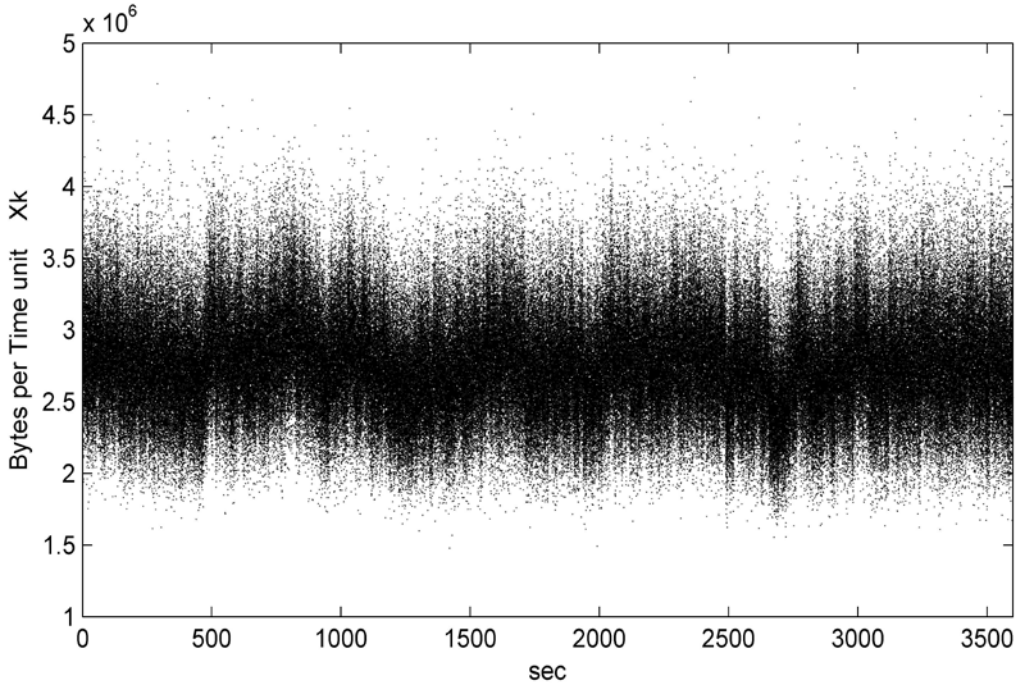


Figure 46: Traffic sequence {x_k}, CAIDA, 10 Gb/s link Seattle-Chicago, day 2 (N=360000, τ_0=10 ms, T=3600 s).

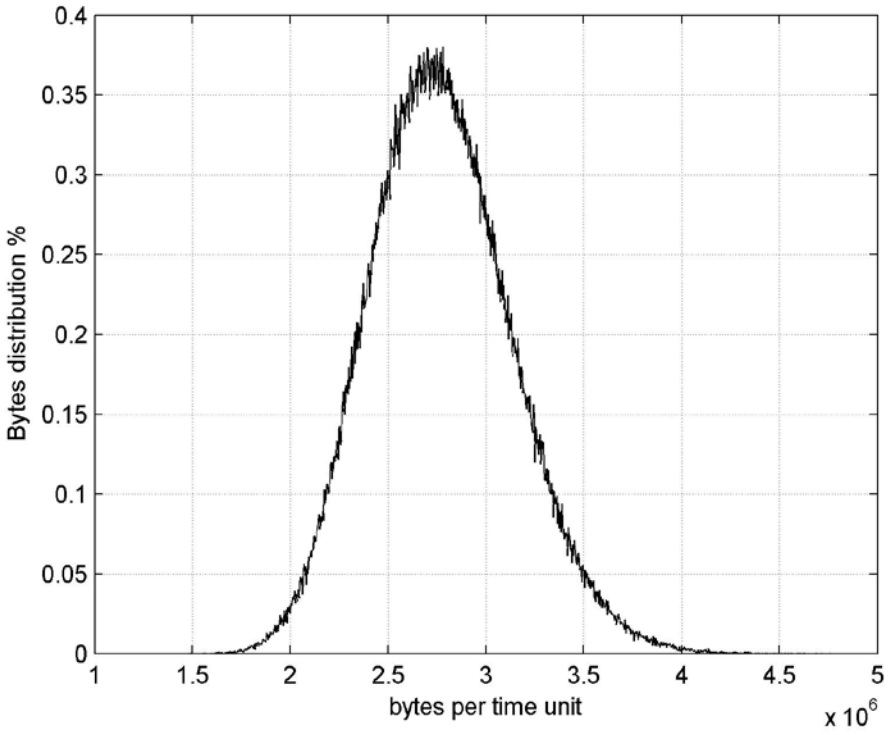


Figure 47: Normalized histogram of {x_k}, CAIDA, 10 Gb/s link Seattle-Chicago, day 2 (N=360000, τ_0=10 ms, T=3600 s).

In Figure 48 the MAVAR diagrams in log-log scale of the sequences {x_k} of day 1 and day 2 are presented. Curves are almost identical in the two days, only with a slightly difference for the largest values of τ, were MAVAR loses confidence. Excluding the lowest and highest values of τ, a nearly straight line behaviour is observed in both sequences with slope μ_1=-2,12 and μ_2=-2,11 (obtained by least square regression) of day 1 and day 2 respectively. Therefore, the traffic traces seem to obey simple power laws (11) with $\alpha_1 \cong$ -0,88 for day 1 and $\alpha_2 \cong$ -0,89 for day 2, corresponding to LRD processes with $H_1 \cong$ 0,94 and $H_2 \cong$ 0,95.

Then, the 1h overall time span of the traffic trace CAIDA 10 Gb/s link day 2 was divided in distinct subsequent time intervals T = {600 s, 100 s, 60 s, 10 s}. For each sweep of the overall time span, MAVAR was computed for each single segment of data (Figures 49 through 52). A summary of the values of μ, α and H estimated by MAVAR diagrams are presented in Table 9.

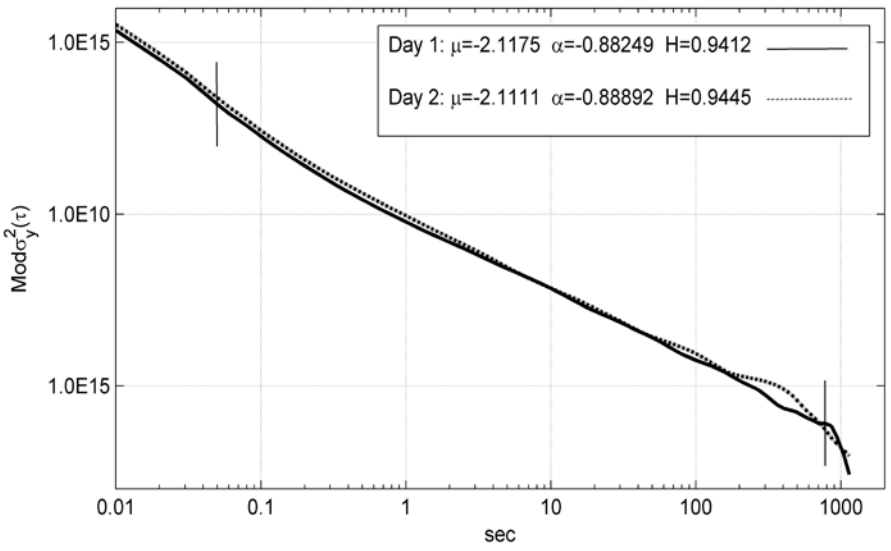


Figure 48: $\mathrm{Mod}\,\sigma_y^2(\tau)$ of sequences $\{x_k\}$, CAIDA, 10 Gb/s link Seattle-Chicago, from day 1 and day 2 (N=360000, τ_0=10 ms, T=3600 s).

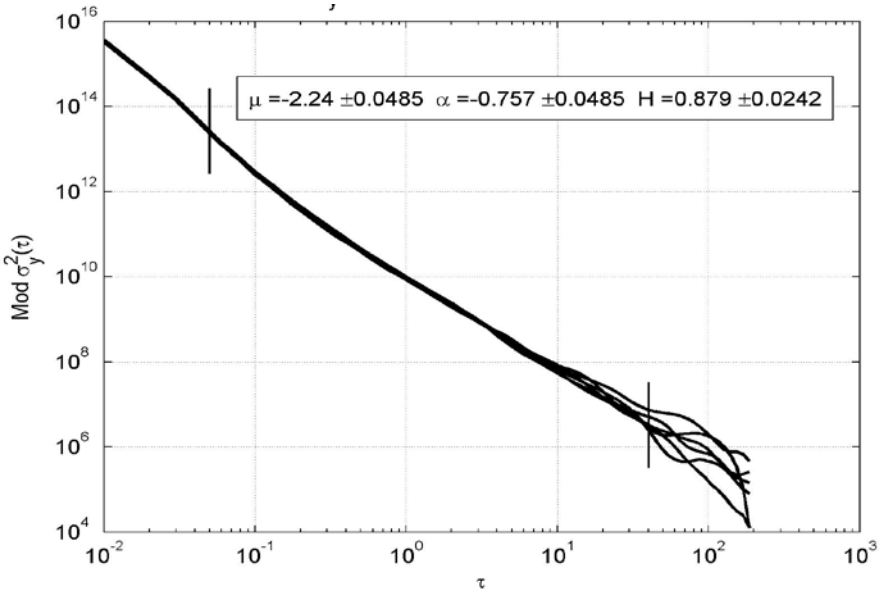


Figure 49: $\mathrm{Mod}\,\sigma_y^2(\tau)$ of 6 sequences $\{x_k\}$, CAIDA, 10 Gb/s link Seattle-Chicago, day 2 (sweep of 1 hour 6 x T=600 s, N=60000, τ_0=10 ms).

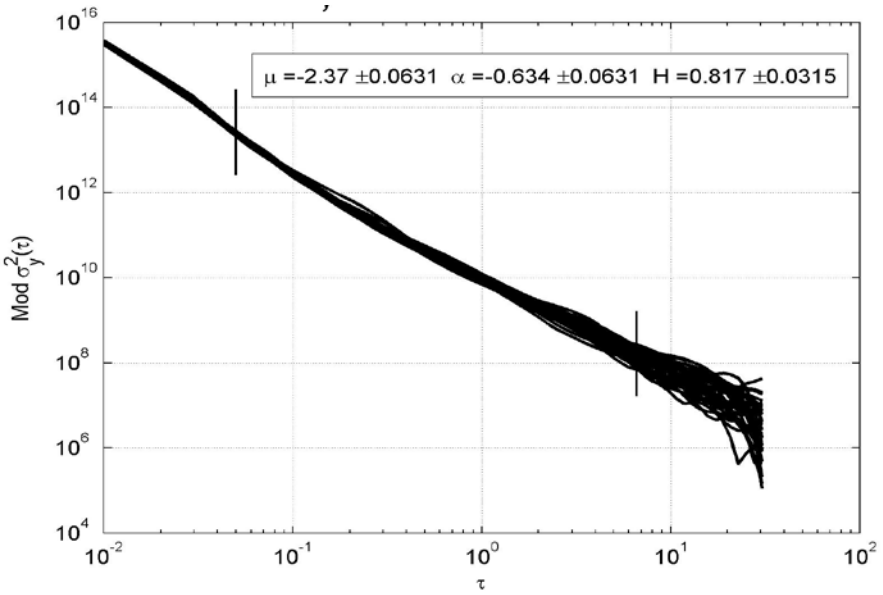


Figure 50: $\mathrm{Mod}\,\sigma_y^2(\tau)$ of 36 sequences {x_k} , CAIDA, 10 Gb/s link Seattle-Chicago, day 2 (sweep of 1 hour 36 x T=100 s, N=10000, τ_0=10 ms).

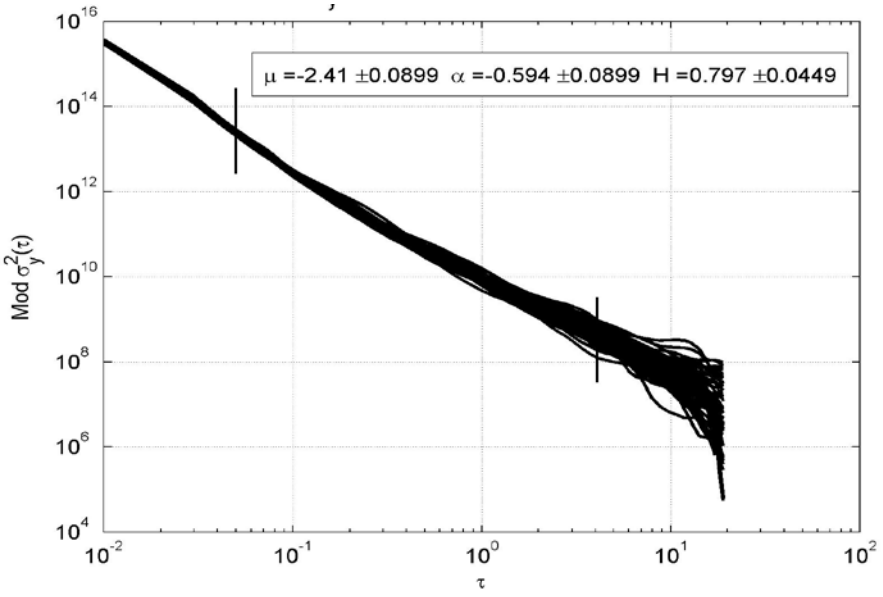


Figure 51: $\mathrm{Mod}\,\sigma_y^2(\tau)$ of 60 sequences {x_k} , CAIDA, 10 Gb/s link Seattle-Chicago, day 2 (sweep of 1 hour 60 x T=60 s, N=6000, τ_0=10 ms).

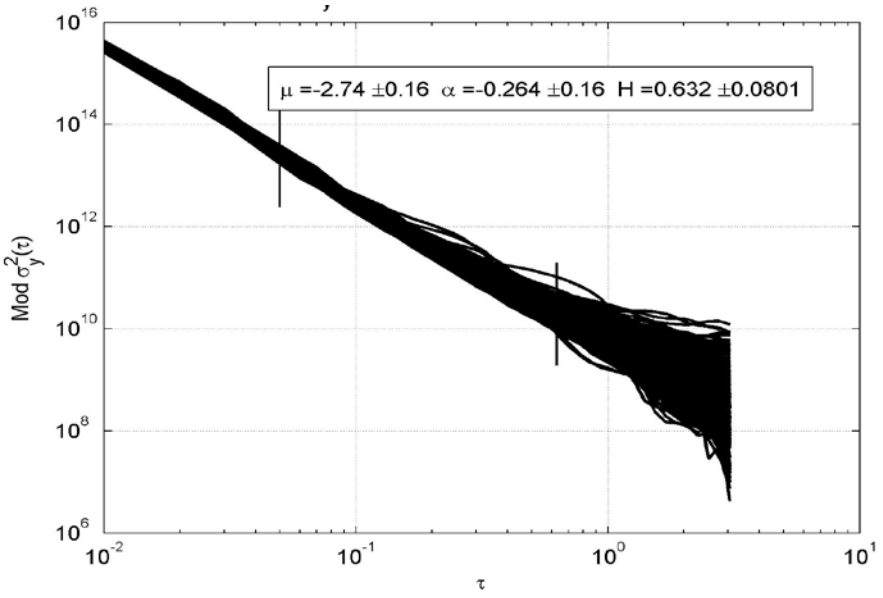


Figure 52: $\mathrm{Mod}\,\sigma_y^2(\tau)$ of 360 sequences $\{x_k\}$, CAIDA, 10 Gb/s link Seattle-Chicago, day 2 (sweep of 1 hour 360 x T=10 s, N=1000, τ_0=10 ms).

T	*μ*	*α*	*H*
3600 s	-2,11	-0,889	0,941
600 s*	-2,24±0,0485	-0,757±0,0485	0,879±0,0242
100 s*	2,37±0,0631	-0,634±0,0631	0,817±0,0315
60 s*	-2,41±0,0899	-0,594±0,0899	0,797±0,0449
10 s*	-2,47±0,16	-0,264±0,16	0,632±0,0801

Table 9: Summary of values of μ, α and H estimated by MAVAR diagrams presented in Figures 48-52. *Values presented with mean and standard deviation.

4.5.3 Traces from the CAIDA Project Repository (OC48 Peering Link, *T* = 1h)

The third set of real IP traffic series $\{x_k\}$ [bytes/t.u.] was captured on a west coast OC48 peering link for a large ISP in the CAIDA Project 2002 [102]. The dataset is made of N=360000 samples, acquired with time unit τ_0=10 ms over a measurement interval T=3600 s, at time 09:00-09:59 on the 14 August 2002. Figure 53 depicts the traffic sequence $\{x_k\}$ [bytes/t.u.]. Figure 54 presents its normalized histogram.

In Figure 55, the MAVAR diagram of the sequence $\{x_k\}$ of the CAIDA, 2,5 Gb/s ISP´s peering link is presented. Excluding the lowest and highest values of τ, a nearly straight line behaviour is observed with slope μ = -2,3 . Therefore, the traffic trace seems to obey a simple power law (11) with $\alpha \cong$ -0,69, presenting LRD with $H \cong$ 0,85.

Then, the 1h overall time span of the traffic trace CAIDA 2,5 Gb/s peering link was divided in distinct subsequent time intervals T = {600 s, 100 s, 60 s, 10 s}. For each sweep of the overall time span, MAVAR was computed for each single segment of data (Figures 56 through 59). A summary of the values of μ, α and H estimated by MAVAR diagrams are presented in Table 10.

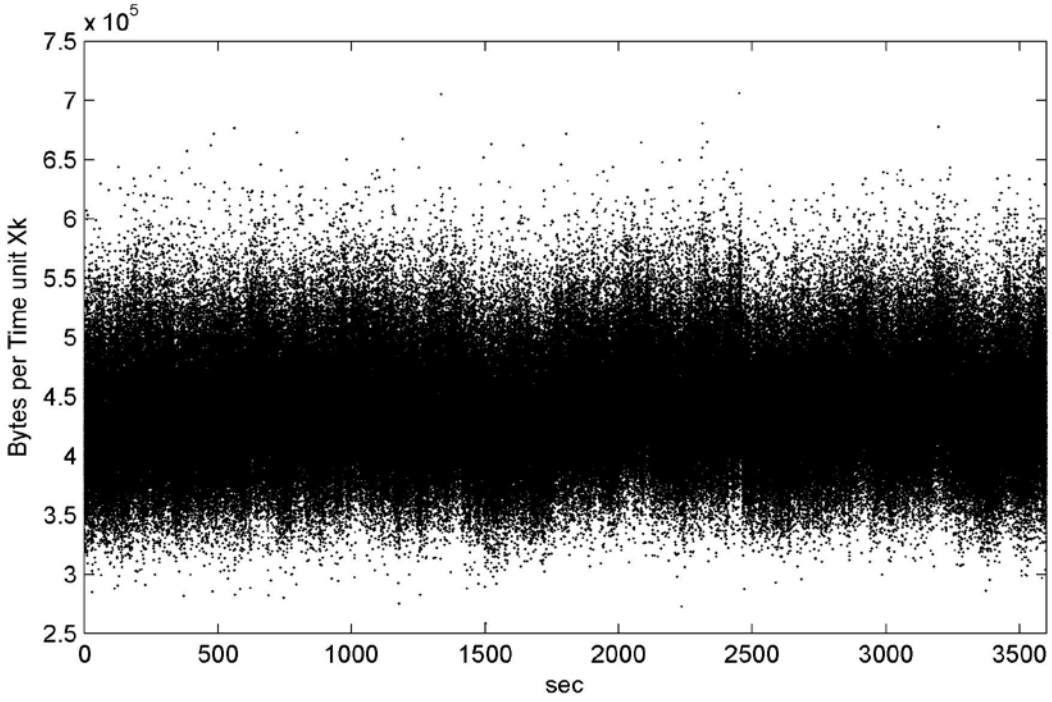


Figure 53: Traffic sequence {x_k} [bytes/t.u.], CAIDA, 2,5 Gb/s ISP´s peering link (N=360000, τ_0=10 ms, T=3600 s).

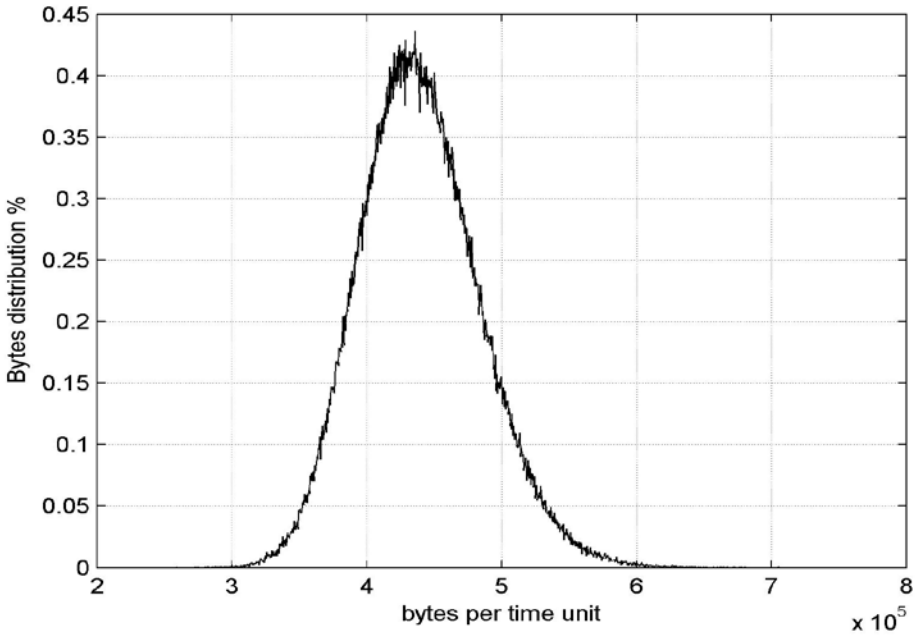


Figure 54: Normalized histogram of sequence {x_k},
CAIDA, 2,5 Gb/s ISP´s peering link (N=360000, τ_0=10 ms, T=3600 s).

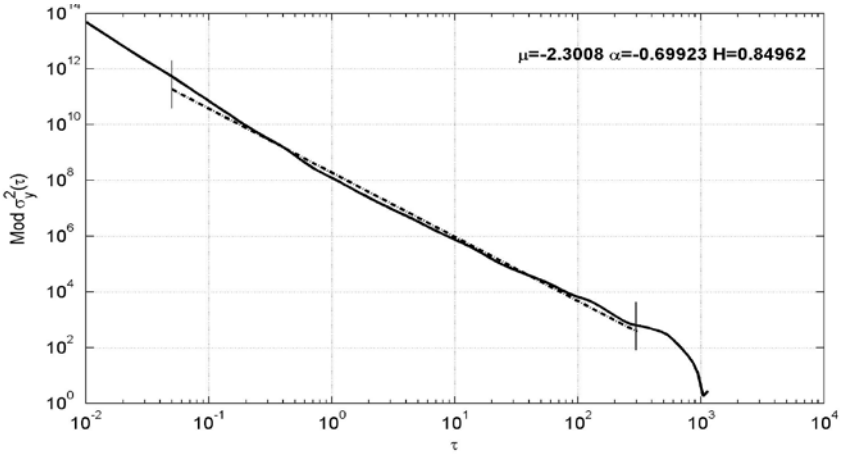


Figure 55: $\mathrm{Mod}\,\sigma_y^2(\tau)$ of sequence {x_k}, CAIDA, 2,5 Gb/s ISP´s peering link
(N=360000, τ_0=10 ms, T=3600 s).

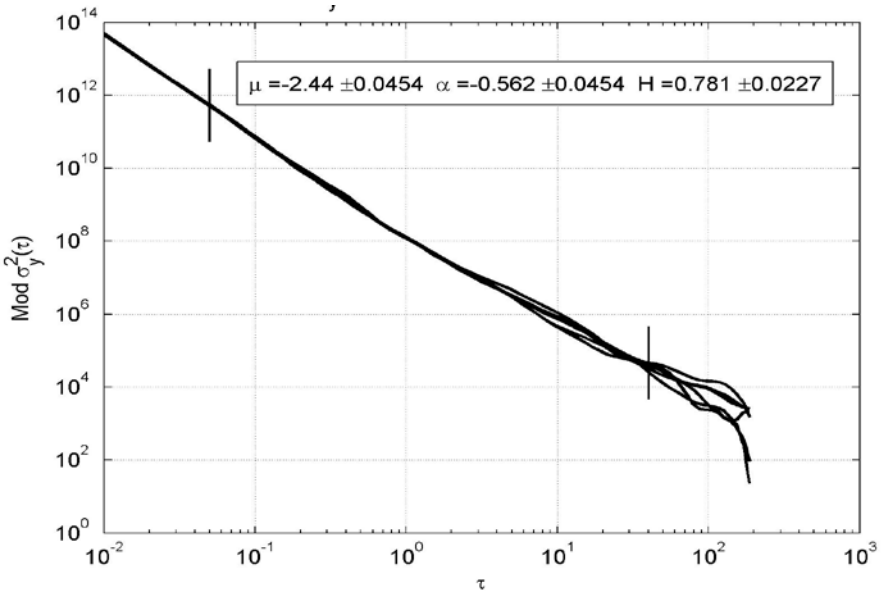


Figure 56: $\mathrm{Mod}\,\sigma_y^2(\tau)$ of 6 sequences {x_k}, CAIDA, 2,5 Gb/s ISP´s peering link (sweep of 1 hour 6 x T=600 s, N=60000, τ_0=10 ms).

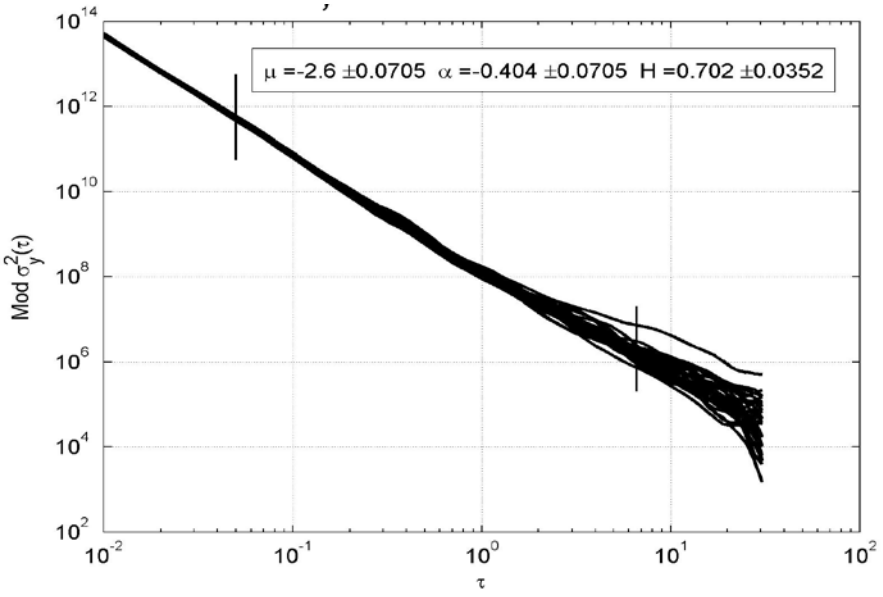


Figure 57: $\mathrm{Mod}\,\sigma_y^2(\tau)$ of 36 sequences {x_k}, CAIDA, 2,5 Gb/s ISP´s peering link (sweep of 1 hour 36 x T=100 s, N=10000, τ_0=10 ms).

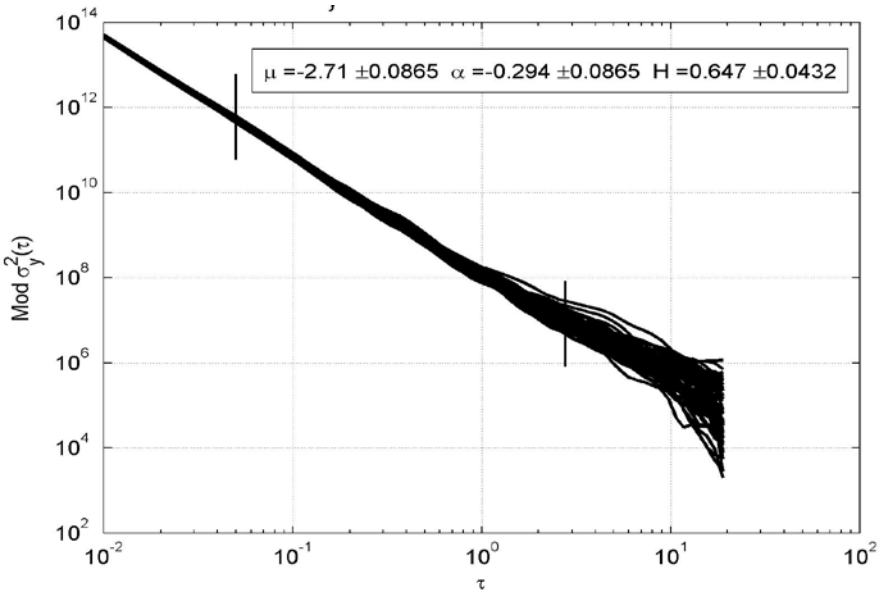


Figure 58: $\mathrm{Mod}\,\sigma_y^2(\tau)$ of 60 sequences {x_k}, CAIDA, 2,5 Gb/s ISP´s peering link (sweep of 1 hour 60 x T=60 s, N=6000, τ_0=10 ms).

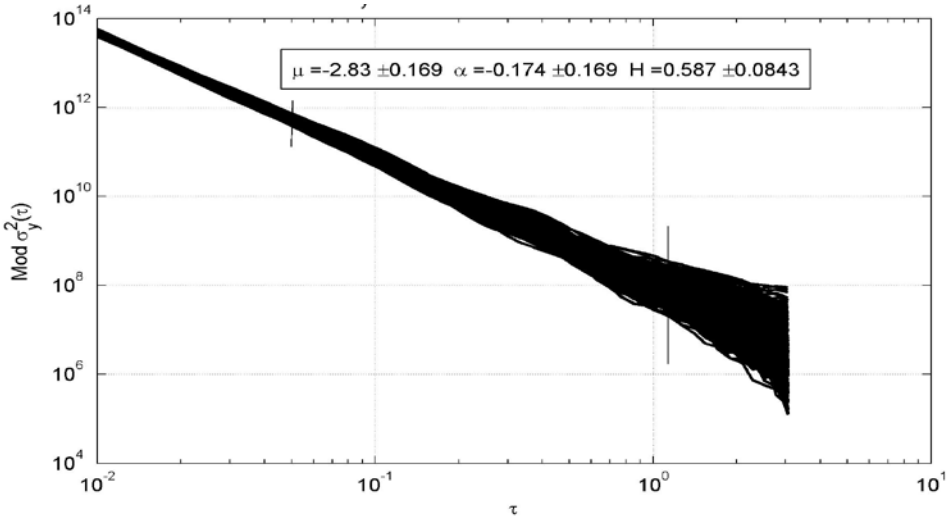


Figure 59: $\mathrm{Mod}\,\sigma_y^2(\tau)$ of 360 sequences {x_k}, CAIDA, 2,5 Gb/s ISP´s peering link (sweep of 1 hour 360 x T=10 s, N=1000, τ_0=10 ms).

T	*μ*	*α*	*H*
3600 s	-2,301	-0,699	0,849
600 s*	-2,44±0,045	-0,562±0,045	0,781±0,023

100 s*	-2,6±0,071	-0,404±0,071	0,702±0,035
60 s*	-2,71±0,086	-0,294±0,086	0,647±0,0432
10 s*	-2,83±0,169	-0,174±0,169	0,587±0,084

Table 10: Summary of values of *μ, α* and *H* estimated by MAVAR diagrams presented in Figures 55 through 59. *Values presented with mean and standard deviation.

4.5.4 Traces from the CAIDA Project Repository (OC48 Peering Link, *T* = 3h)

The fourth set of real IP traffic series $\{x_k\}$ [bytes/t.u.] was captured on a west coast OC48 peering link for a large ISP in the CAIDA Project 2002 [102] (same as third dataset). The dataset is made of N = 1.080.000 samples, acquired with time unit τ_0=10 ms over a measurement interval T=10800 s (3 hours), at time 09:00-11:59 on the 14 August 2002. Figure 60 depicts the traffic sequence $\{x_k\}$ [bytes/t.u.]. Figure 61 presents its normalized histogram.

In Figure 60, a step of the data sequence is clearly visible, as noted already before for Figure 46 but for a significantly smaller step. However, as demonstrated in [47], MAVAR is not significantly affected by such steps or non-stationary trends in estimating the Hurst parameter of random components of the data sequence analysed.

In Figure 62, the MAVAR diagram of the sequence $\{x_k\}$ is presented. Excluding the lowest and highest values of τ, by least-square linear regression the straight line slope results $\mu \cong$ -2.01 . Therefore, the corresponding power law (11) would be with $\alpha \cong$ -0,99, that is LRD with $H \cong$ 0,99.

Then, the 3h overall time span of the traffic trace was divided in distinct subsequent time intervals T = {3600 s, 1800 s, 600 s, 100 s, 60 s, 10 s}. For each sweep of the overall time span, MAVAR was computed for each single segment of data (Figures 63 through 68). A summary of the values of *μ*, *α* and *H* estimated by MAVAR diagrams are presented in Table 11.

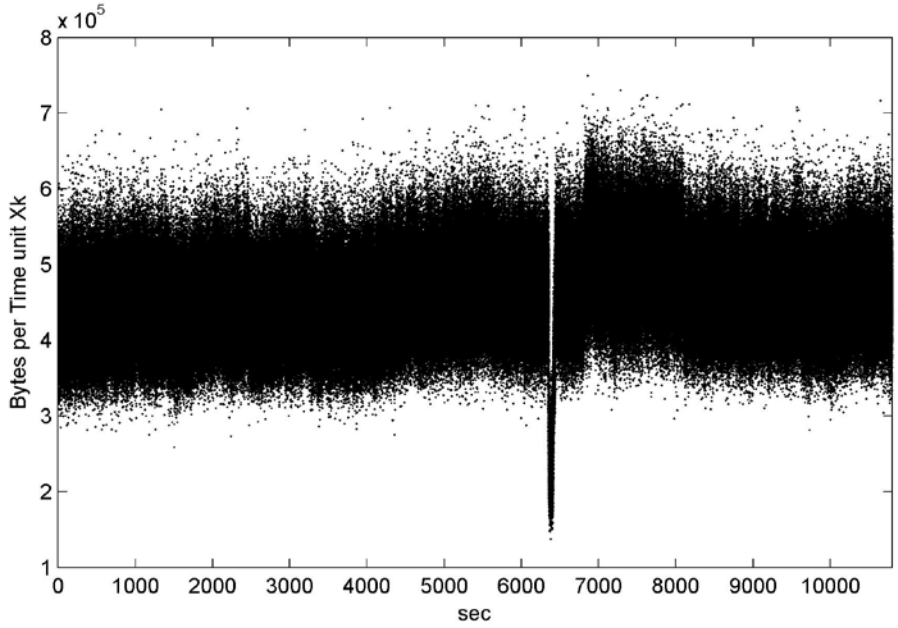


Figure 60: Traffic sequence $\{x_k\}$ [bytes/t.u.], CAIDA, 2,5 Gb/s ISP´s peering link (N=1080000, τ_0=10 ms, T=10800 s).

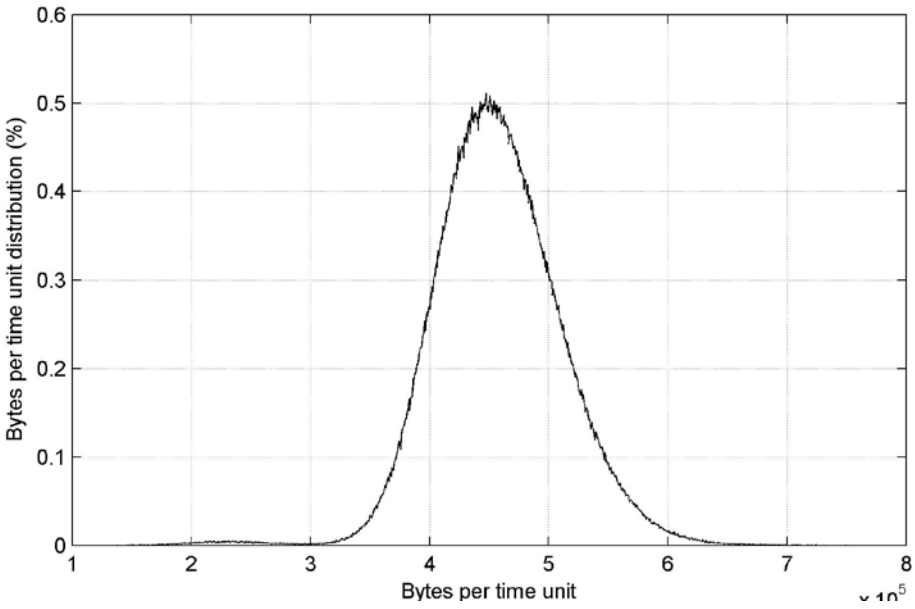


Figure 61: Normalized histogram of sequence $\{x_k\}$, CAIDA, 2,5 Gb/s ISP´s peering link (N=1080000, τ_0=10 ms, T=10800 s).

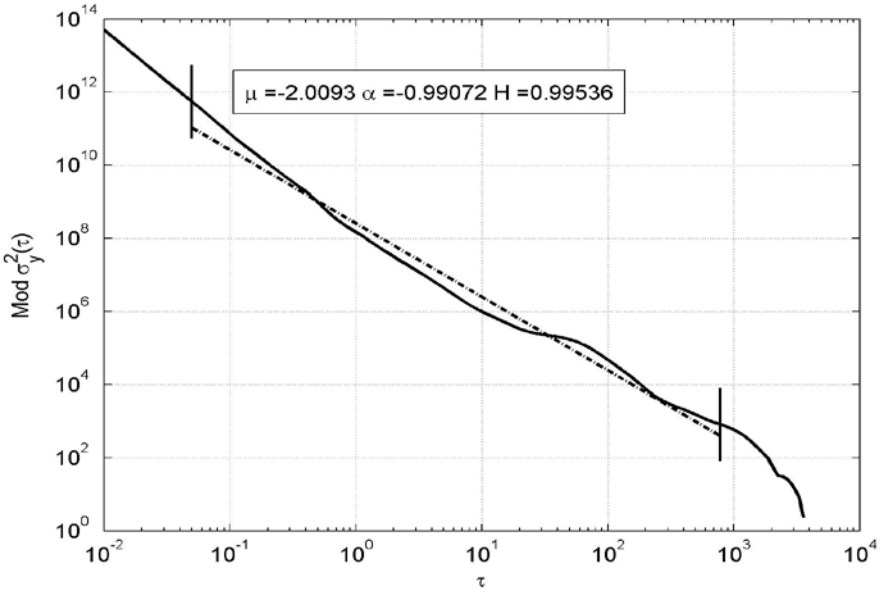


Figure 62: $\mathrm{Mod}\,\sigma_y^2(\tau)$ of sequence {x_k}, CAIDA, 2,5 Gb/s ISP´s peering link (N=1080000, τ_0=10 ms, T=10800 s).

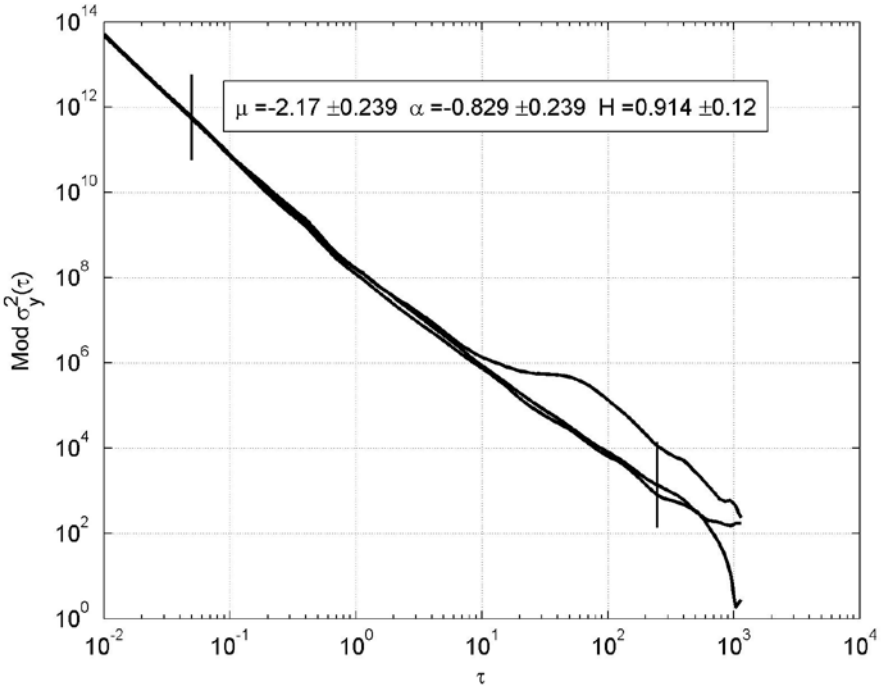


Figure 63: $\mathrm{Mod}\,\sigma_y^2(\tau)$ of 3 sequences {x_k}, CAIDA, 2,5 Gb/s ISP´s peering link (sweep of 3 hours: 3 x T=3600 s, N=360000, τ_0=10 ms).

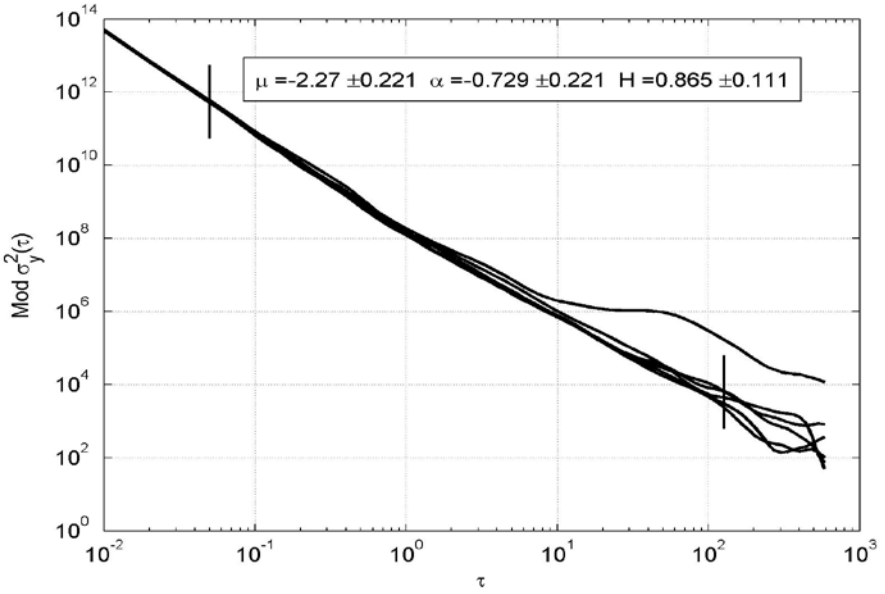


Figure 64: $\mathrm{Mod}\,\sigma_y^2(\tau)$ of 6 sequences {x_k}, CAIDA, 2,5 Gb/s ISP´s peering link (sweep of 3 hours: 6 x T=18000 s, N=180000, τ_0=10 ms).

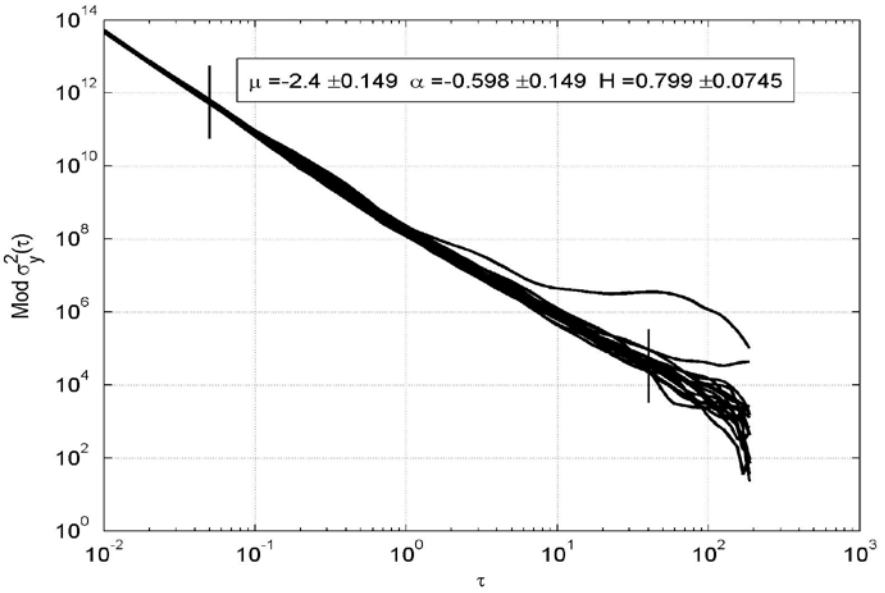


Figure 65: $\mathrm{Mod}\,\sigma_y^2(\tau)$ of 18 sequences {x_k}, CAIDA, 2,5 Gb/s ISP´s peering link (sweep of 3 hours: 18 x T=600 s, N=60000, τ_0=10 ms).

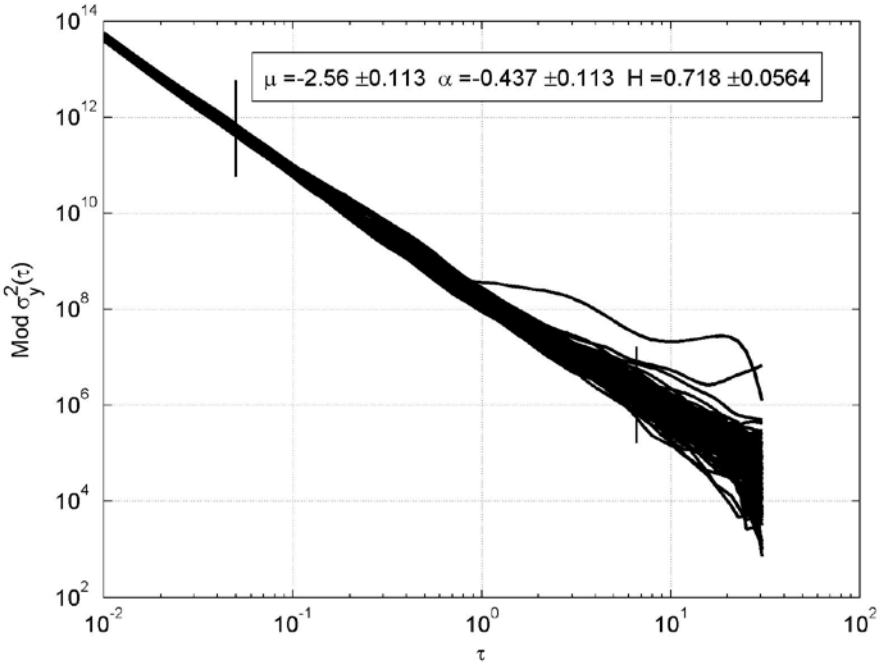


Figure 66: $\mathrm{Mod}\,\sigma_y^2(\tau)$ of 108 sequences {x_k}, CAIDA, 2,5 Gb/s ISP´s peering link (sweep of 3 hours: 108 x T=100 s, N=10000, τ_0=10 ms).

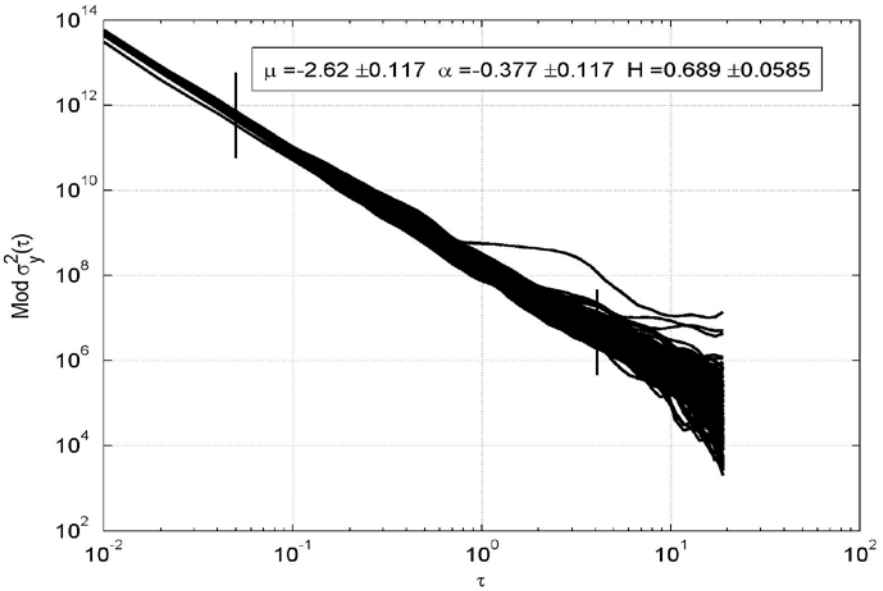


Figure 67: $\mathrm{Mod}\,\sigma_y^2(\tau)$ of 180 sequences {x_k}, CAIDA, 2,5 Gb/s ISP´s peering link (sweep of 3 hours: 180 x T=60 s, N=6000, τ_0=10 ms).

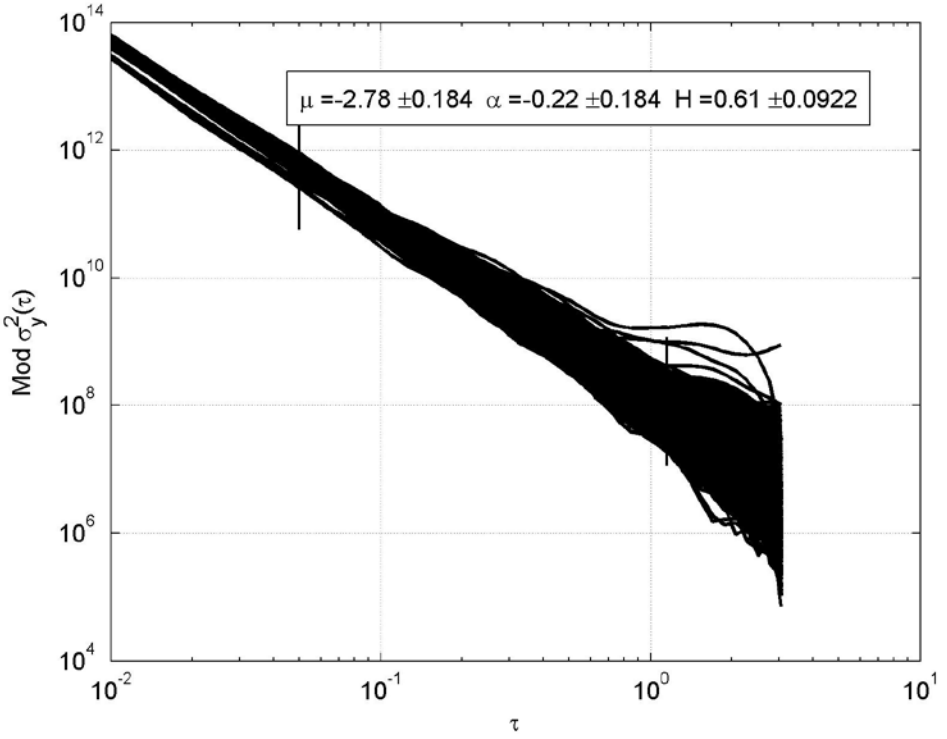


Figure 68: $\mathrm{Mod}\,\sigma_y^2(\tau)$ of 1080 sequences $\{x_k\}$, CAIDA, 2,5 Gb/s ISP´s peering link (sweep of 3 hours: 1080 x T=10 s, N=1000, τ_0=10 ms).

T	***μ***	***α***	***H***
10800 s	-2,009	-0,991	0,995
3600 s*	-2,17 ± 0,239	-0,829 ± 0,239	0,849 ± 0,12
1800 s*	-2,27 ± 0,221	-0,729 ± 0,221	0,865 ± 0,111
600 s*	-2,4 ± 0,149	-0,598 ± 0,149	0,799 ± 0,075
100 s*	-2,56 ± 0,113	-0,437 ± 0,113	0,718 ± 0,056
60 s*	-2,62 ± 0,117	-0,377 ± 0,117	0,689 ± 0,058
10 s*	-2,78 ± 0,184	-0,22 ± 0,184	0,61 ± 0,092

Table 11: Summary of values of *μ*, *α* and *H* estimated by MAVAR diagrams presented in Figures 62-68. *Values presented as average ± standard deviation.

4.5.5 A Comparison of α Values Estimated over Time

In order to present how the estimated values of α depend on the time span of the traffic traces, Figure 69 plots the mean and standard deviation of the values of α estimated on the same traffic trace of length 3600 s, but on data subsegments for increasing values of *T*, as estimated by MAVAR diagrams presented in Figures 48-52. Similarly, Figures 70 and 71 plot the mean and standard deviation of the values of α on data subsegments for increasing values of *T*, as estimated by MAVAR diagrams presented in Figures 55-59 and 62-68, respectively.

We see that, the longer is the time span of the data segment, the higher is the LRD that is detected (higher values of *H*, or, equivalently, lower values of α).

Moreover, in order to detect possible trends of α over time, plots of values of $\alpha(t)$ estimated by MAVAR on data subsegments of 60 seconds over three different traffic sequences $\{x_k\}$ are presented in Figures 72, 73, 74.

In all cases, no peculiar trend is evident. Only some random variation, as expected. It would be interesting to investigate further on this direction, by estimating α (or equivalently *H*) over longer measurement time intervals, for example along days. As traffic average values may exhibit diurnal trends, also the LRD might exhibit slow periodic trends.

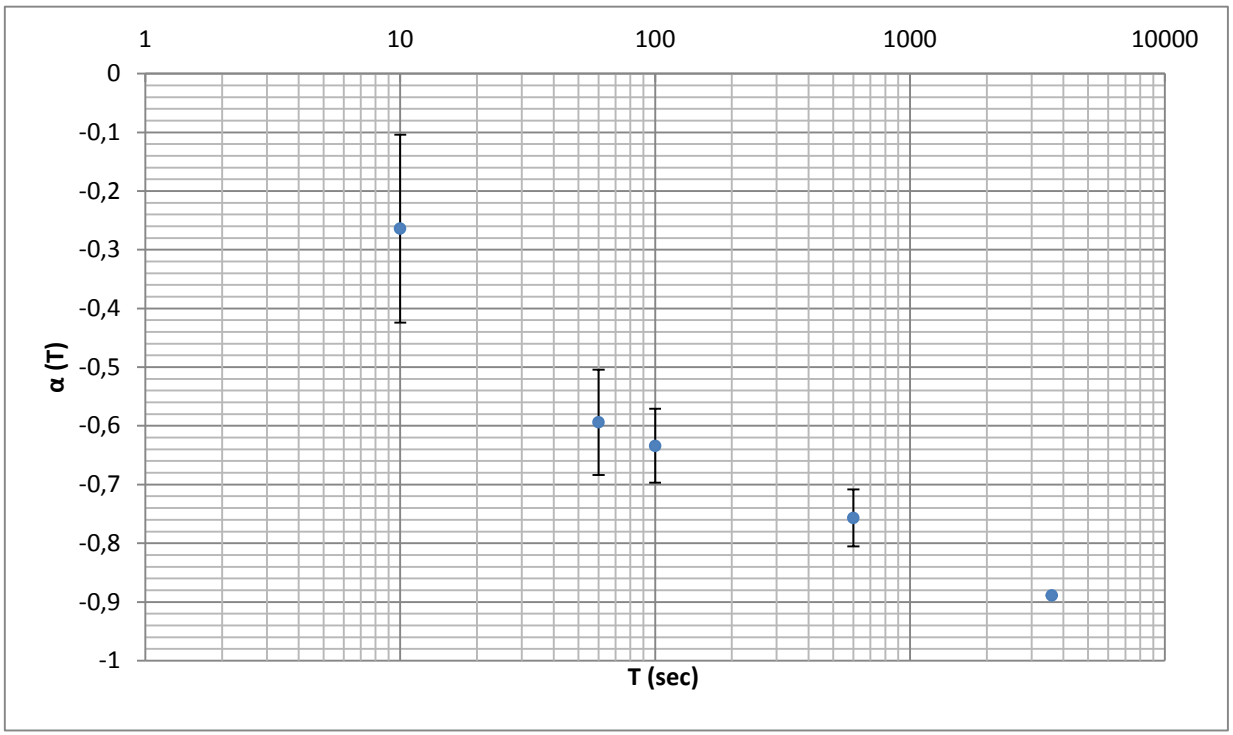


Figure 69: Plot of α estimated by MAVAR diagrams presented in Figures 48-52.
(Mean and standard deviation for T<3600). T in log scale.

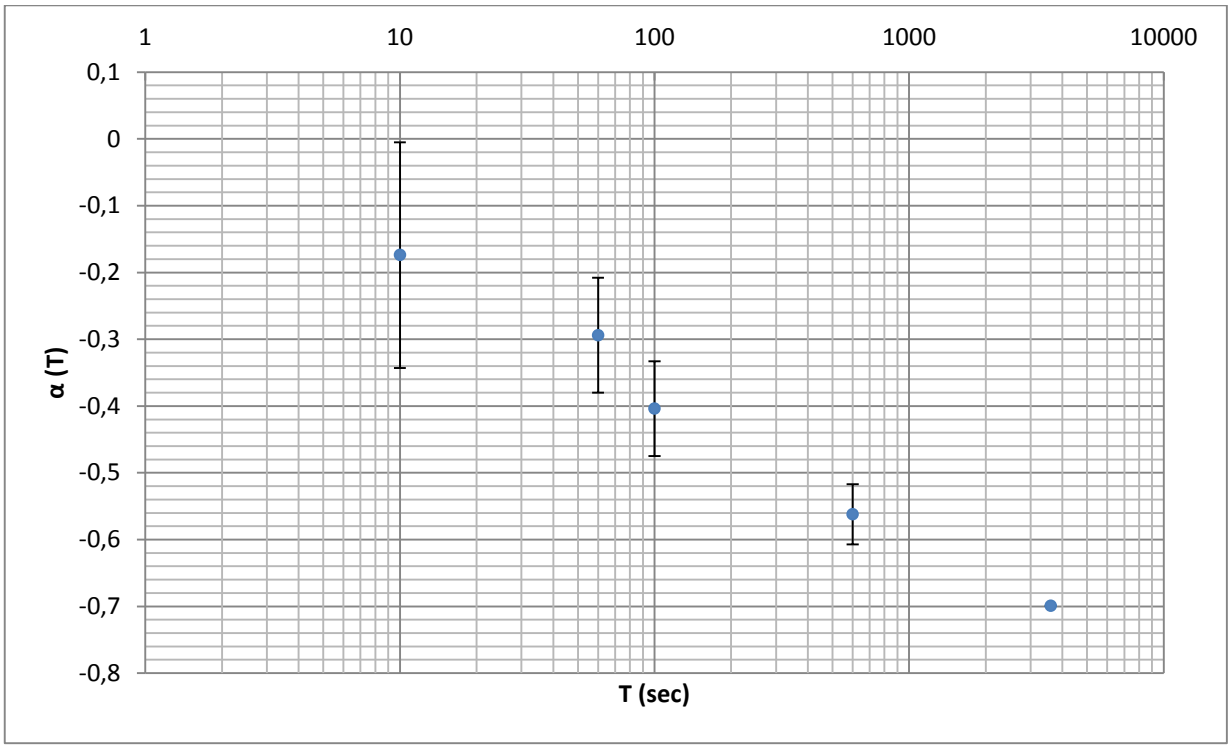


Figure 70: Plot of α estimated by MAVAR diagrams presented in Figures 55-59.
(Mean and standard deviation for T<3600). T in log scale.

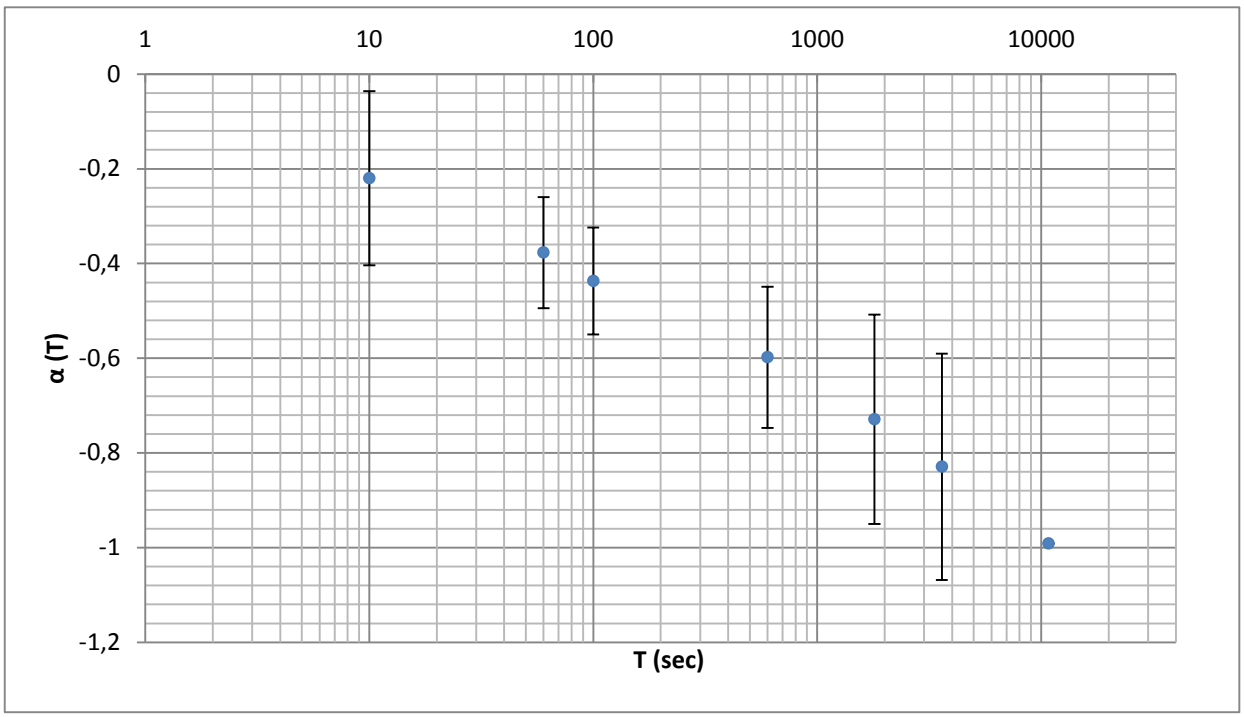


Figure 71: Plot of α estimated by MAVAR diagrams presented in Figures 62-68.
(Mean and standard deviation for T<108000 s). T in log scale.

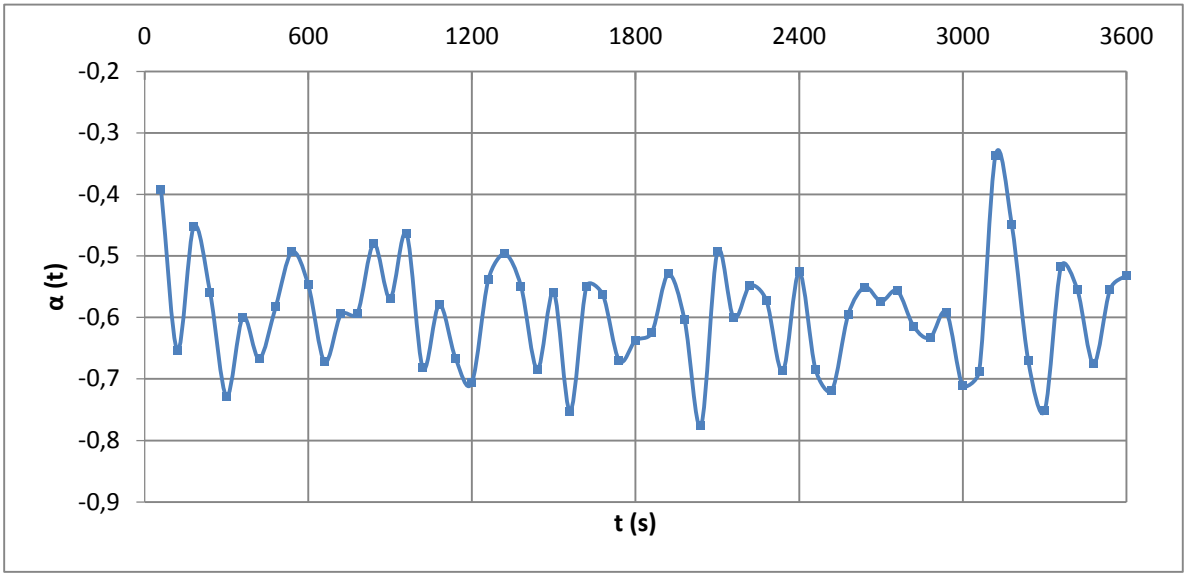


Figure 72. Plot of $\alpha(t)$ estimated from 60 computations of MAVAR every 60 seconds (T=60) over the 1 hour traffic sequence $\{x_k\}$ corresponding to the Seattle-Chicago 10 Gb/s link, on day 2.

Such MAVAR diagrams are presented in Figure 51.

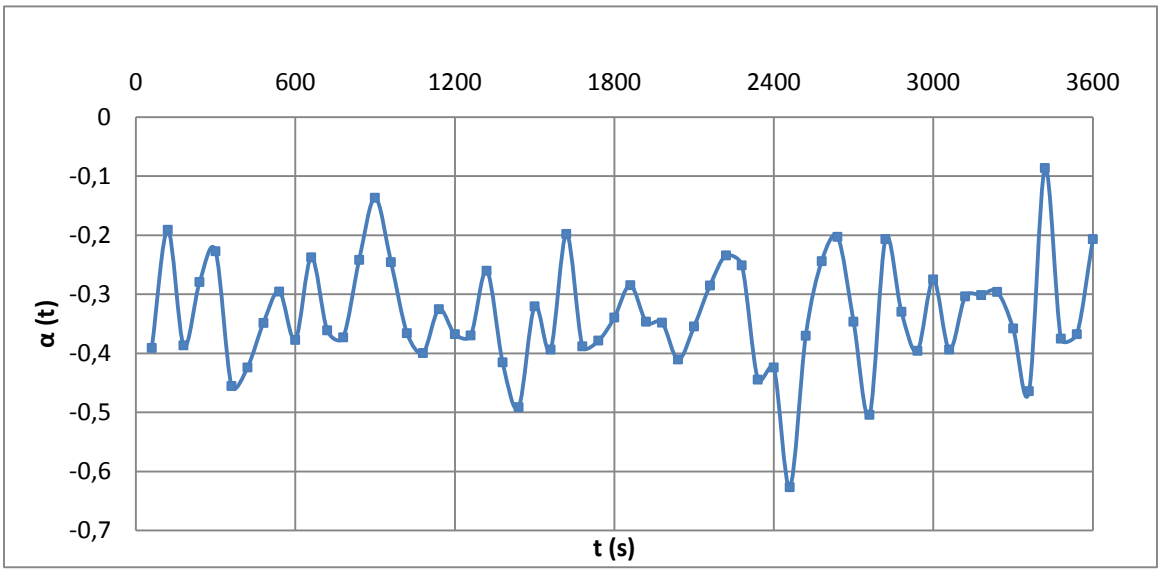


Figure 73. Plot of $\alpha(t)$ estimated from 60 computations of MAVAR every 60 seconds over the 1 hour traffic sequence $\{x_k\}$ corresponding to the west coast 2,5 Gb/s peering link in 2002. Such MAVAR diagrams are presented in Figure 58.

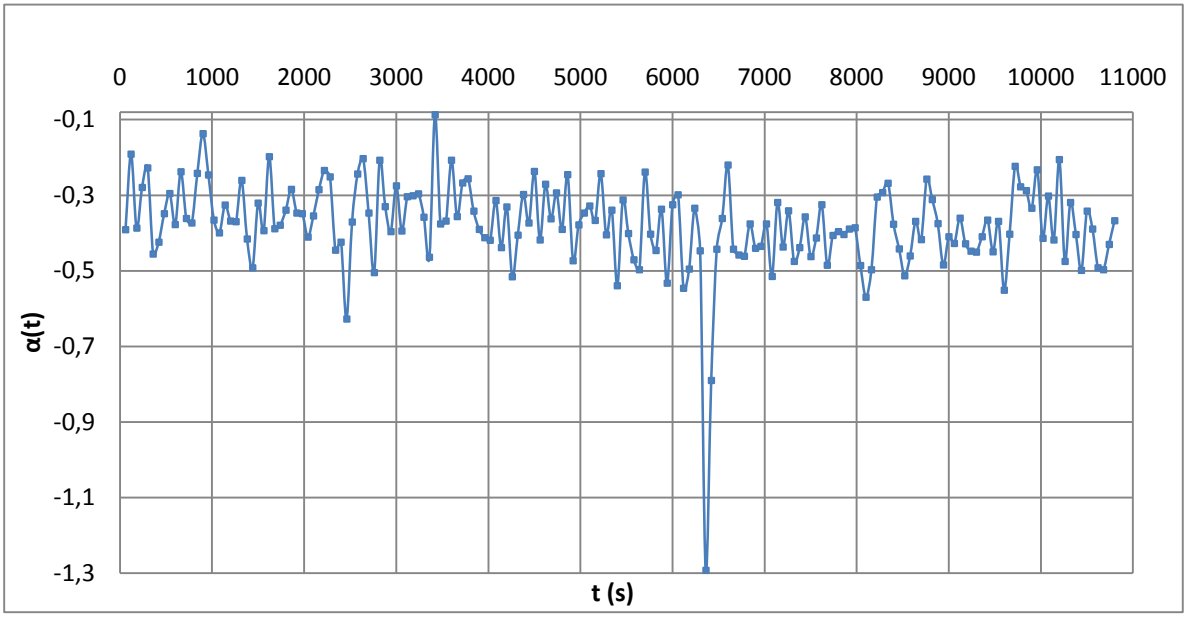


Figure 74. Plot of $\alpha(t)$ estimated from 180 computations of MAVAR every 60 seconds over the 3 hours traffic sequence $\{x_k\}$ corresponding to the west coast 2,5 Gb/s peering link in 2002. Such MAVAR diagrams are presented in Figure 67.

4.5.6 Some Conclusions about Self-Similarity Analysis

The analysis results presented were computed on data measured

- on a 10 Gb/s link towards the SGSN from the RNC pool;
- on a 10 Gb/s link connecting Seattle with Chicago in the CAIDA UCSD Project;
- on a west coast OC48 (2.5 Gbps) peering link for a large ISP in the CAIDA Project 2002.

In all cases, traffic measured is highly aggregated from various sources, including mobile cellular networks and fixed networks. As a matter of fact, today there are no FMC networks yet available for traffic measurements. However, the measurement results provided represent various cases where fixed and mobile traffic are aggregated.

In all cases presented here, as well as in all others we studied but did not report here for the sake of brevity, significant values of self-similarity were measured, except on shortest time scales.

Even in the AITIA traffic data, which span only 1.5 s of measurement interval, the traffic proves to be not purely Poisson (with zero time-correlation). Here, the parameters measured have been $\alpha \cong 0.28$ and $H \cong 0.64$, rather far from the pure Poisson traffic where $\alpha = 0$.

In other traffic traces, which span longer time intervals up to a few hour, the traffic exhibits much higher self-similarity, with values of α ranging from -0.25 to -0.9 (and *H* from 0.6 to almost 1). Such values denote a strong self-similarity of traffic, under all cases of traffic aggregation under consideration.

Such high values of self-similarity call for significant over-provisioning of network resources (buffers, link capacity), compared to the ideal case where all traffic is supposed to be Poisson.

Interestingly, no significant slow trends in the values of α and H were detected. As further development, we believe that it would be interesting to analyse if such trends can be identified over periods spanning at least some days.

4.5.7 Probes in FT network

In operational networks, traffic measurements are mainly performed according to two methods.

The first one is an active method based on the capture and analysis of the live transmitted IP traffic, and injected tests frames.

The second one is a passive method in which all transmitted IP traffic is captured and analysed in real time.

In the FT network, per example, traffic probes allow a detailed capture and analysis of the traffic every 6 minutes. Figure 75 and Figure 76 illustrate the location of traffic probes in the FT (fixed and mobile) network. They are installed between aggregation and core networks within the fixed network. For the mobile network, the probes are installed between SGSN and GGSN; these probes allow analysing traffic of a few thousands of customers. The probe contains a card that performs Data Acquisition and Generation (DAG, which realizes in real time an extraction of all packets passing on a link). The card precisely stamps the packets (with a precision of the 1 ns), and all results of the measurement are stored on a disk drive. The probe used for the measurement of FT network traffic was located in a highly dense area in France with thousands of xDSL and FTTH or 3G customers in the fixed network and the mobile network, respectively. The probes measure traffic generated by 13 categories of applications each category containing a set of applications (see Figure 77).

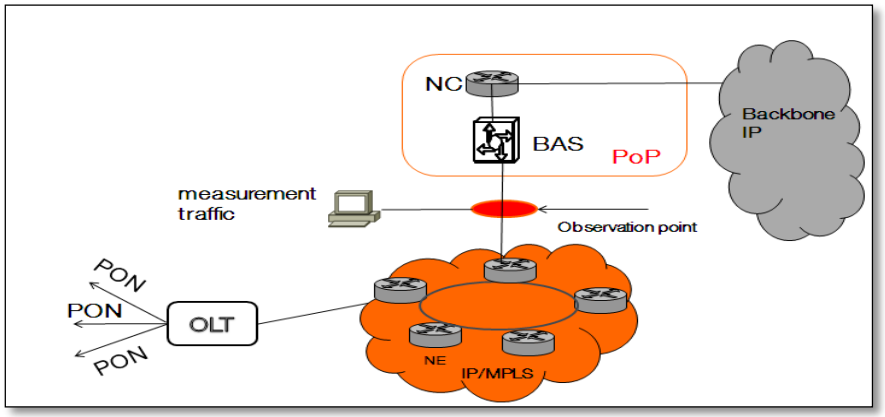


Figure 75. Probes location in the fixed-network

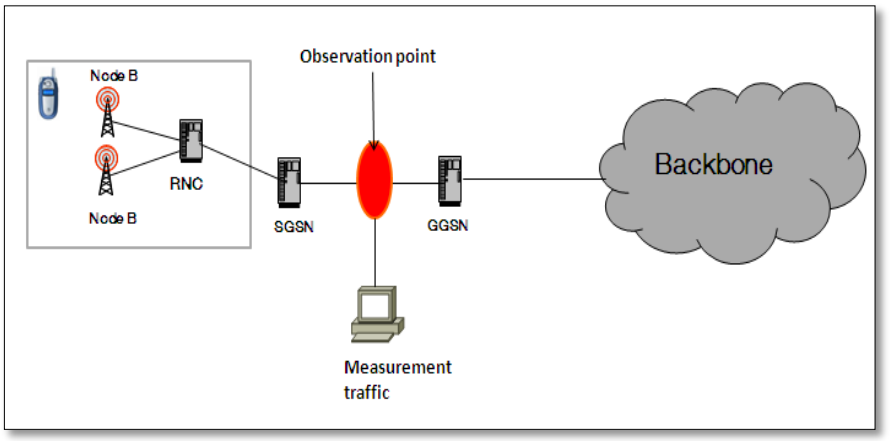


Figure 76. Probes location in the Mobile-network

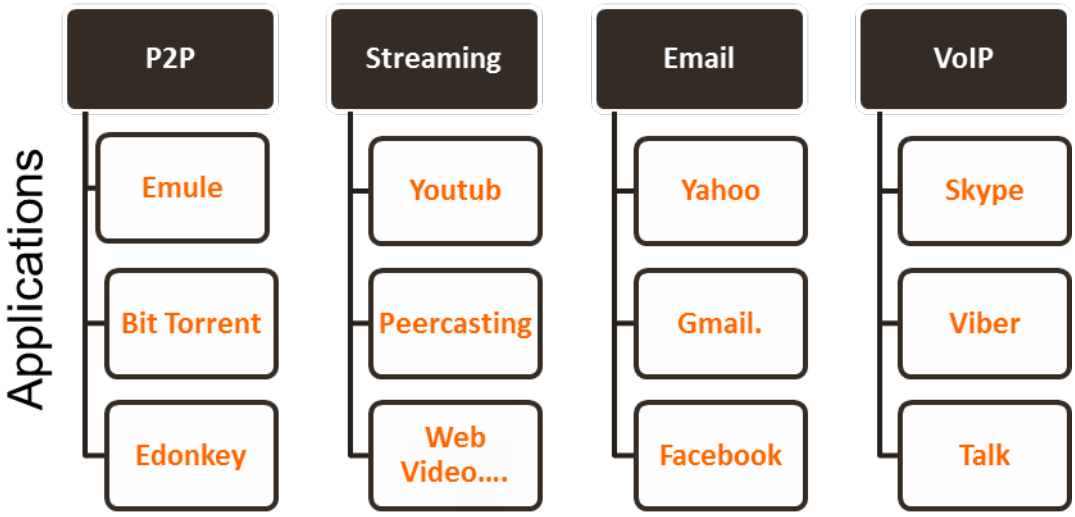


Figure 77. Classification of applications

4.5.8 FT traffic analysis report

4.5.8.1 Analysis of Real traffic Data (FT Network)

Introduction

Internet bandwidth consumed by residential customers continues to increase over the time and is more and more tied to video contents (e.g., video streaming, download...). In the United States, Netflix video service generates about 32.7% of total downstream Internet traffic during the busy hours [103]. In Korea, Korea Telecom (KT) decided in 2012 to block

Internet access to connected TV as they were responsible of a drastic increase of Internet volume. Indeed, it was 15 times higher than the volume generated by managed service IPTV [104]. Internet video also represents a major contribution of the data traffic on mobile network [105][106] (YouTube videos represent about 22% and 52% of videos are consulted in a streaming mode).

In France, there is great competition between Internet service providers in the FTTH access network, as the maximum offered bandwidth is 1 Gbps. In order to guarantee an optimal user experience, operators need to know the amounts of traffic in the different network segments to identify any possible bottleneck and to avoid QoS degradation. This knowledge will allow to correctly dimensioning the network in order to avoid congestion points (like peering links).

Several papers already presented traffic measurements analysis about Internet applications. The work detailed in this document shows the tendencies evolution like the way residential customers use the network, and the impact of FTTH customers on the upstream Internet traffic and the importance of heavy users who follow Pareto law (few customers generate a most of the traffic). This report is based on an analysis of real Internet traffic measured in the field in the FT fixed networks in an ADSL/FTTH context in October 2013.

The second section details the major applications generating traffic in upstream and downstream either for ADSL or for FTTH customers. This section details the customers' traffic profile day by day during a week and presents the notion of busy hours defined by [121].

The third section presents the downstream and upstream traffic evolution for FTTH customers with respect to ADSL ones. It also addresses the notion of heavy users in the case of ADSL and FTTH customers. It also classifies the applications generating most of Internet traffic according to the OSI Layer 4.The last section proposes some conclusions.

4.5.8.2 Internet profile observation of residential customers

A. Major tendencies of downstream and upstream Internet traffic

In this sub-section we will focus on Internet applications generating the most of traffic in upstream and downstream and that in the case of FTTH and ADSL customers. This kind of analysis is performed to improve the network operator knowledge of Internet applications and help in anticipating network dimensioning according to customers' use cases evolution.

- *Upstream Internet traffic for ADSL and FTTH customers*

Figure 78 illustrates the total volume evolution over the time for upstream Internet traffic in the case of FTTH access network. Moreover this figure depicts volume for some Internet applications.

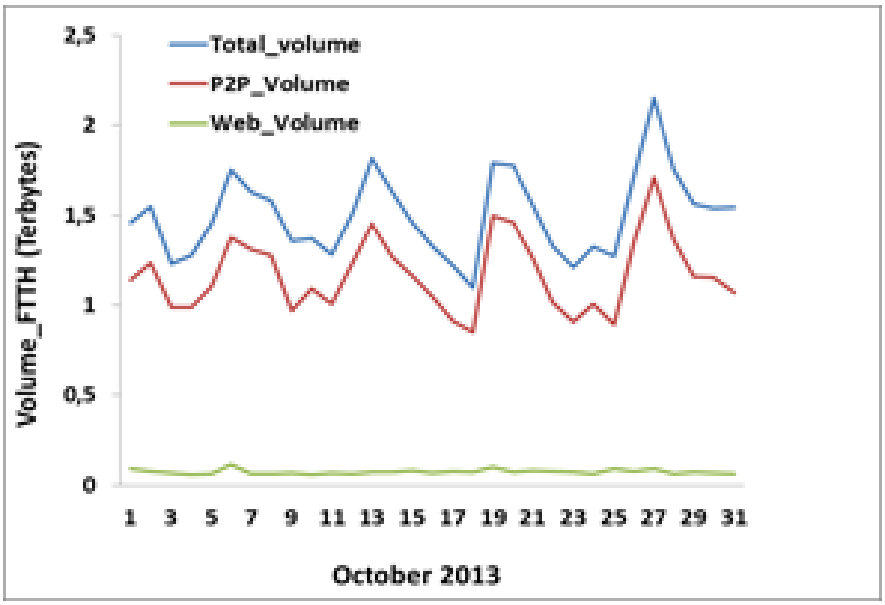


Figure 78. Upstream Internet traffic for FTTH customers, measurements performed in October 2013.

The total volume generated by FTTH customers is 46Tbytes. P2P is the application generating 78% of the total volume and the other applications (such as Web browsing, mailing…) generate the rest of the volume. During the same observation period (October 2013), ADSL customers generated a total upstream Internet traffic of 22Tbytes and P2P remains the application generating the main proportion of the total volume of traffic (about 48%) and Web browsing represents 18% of the total volume, while in 2008, P2P volume was mainly generated by eDonkey application (10% of P2P volume was only generated by BitTorrent [123]), now, there is an inversion of this situation Bit Torrent represents the application generate a most P2P traffic.

- Downstream Internet traffic for ADSL and FTTH customers

Figure 79 illustrates the total volume evolution over the time for downstream Internet traffic in the case of FTTH access network. Moreover these figures depict volume for some Internet applications. Figure 77 shows that the total downstream Internet traffic generated by FTTH customers in represents 89 Tbytes. Video streaming is the application generating alone 36% of the total volume generated by FTTH customers. Download application generates 26% of the global traffic. Web browsing and P2P represent each a volume of 16%, and the rest of the application generates a very low volume. The average downstream Internet traffic generated by ADSL customers in also reaches 126Tbytes. Here again, video streaming represents 25% of the total volume, followed by file downloading (21%), P2P (12%) and Web browsing (18%).

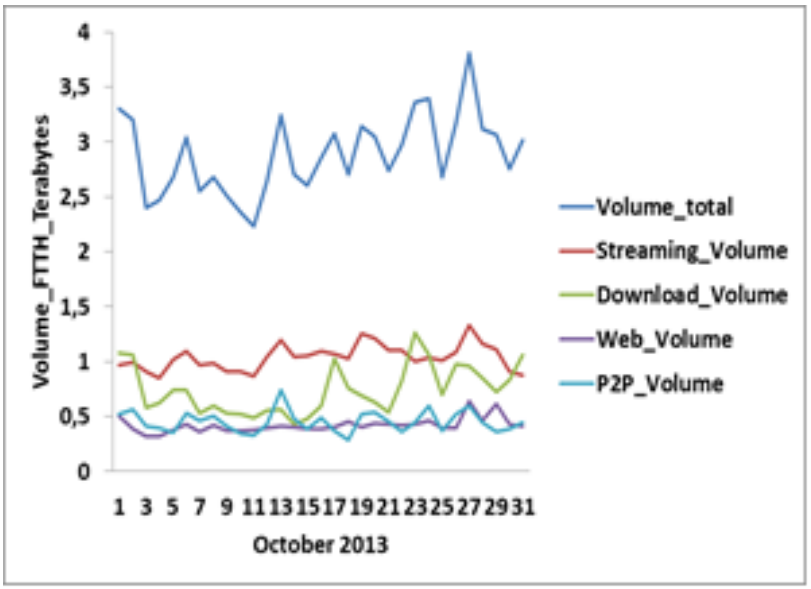


Figure 79. Downstream Internet traffic for FTTH customers, measurements performed in October 2013.

B. Upstream and downstream Internet traffic profiles for FTTH and ADSL customers

- *Upstream profile for ADSL and FTTH customers*

This sub-section explains daily downstream and upstream Internet profiles for ADSL and FTTH customers thanks an analysis during 24 hours over a week. These observations allowed us in identifying the busy hours period as defined in [124]. Moreover, for a network operator, it is important to understand how the traffic evolves during the day, in order to perform maintenance, firmware updates during periods with low traffic activity [105].

Figure 80 illustrates the daily upstream Internet traffic profile for FTTH customers over a period between the 07th of October 2013 and the 13rd of October 2013.

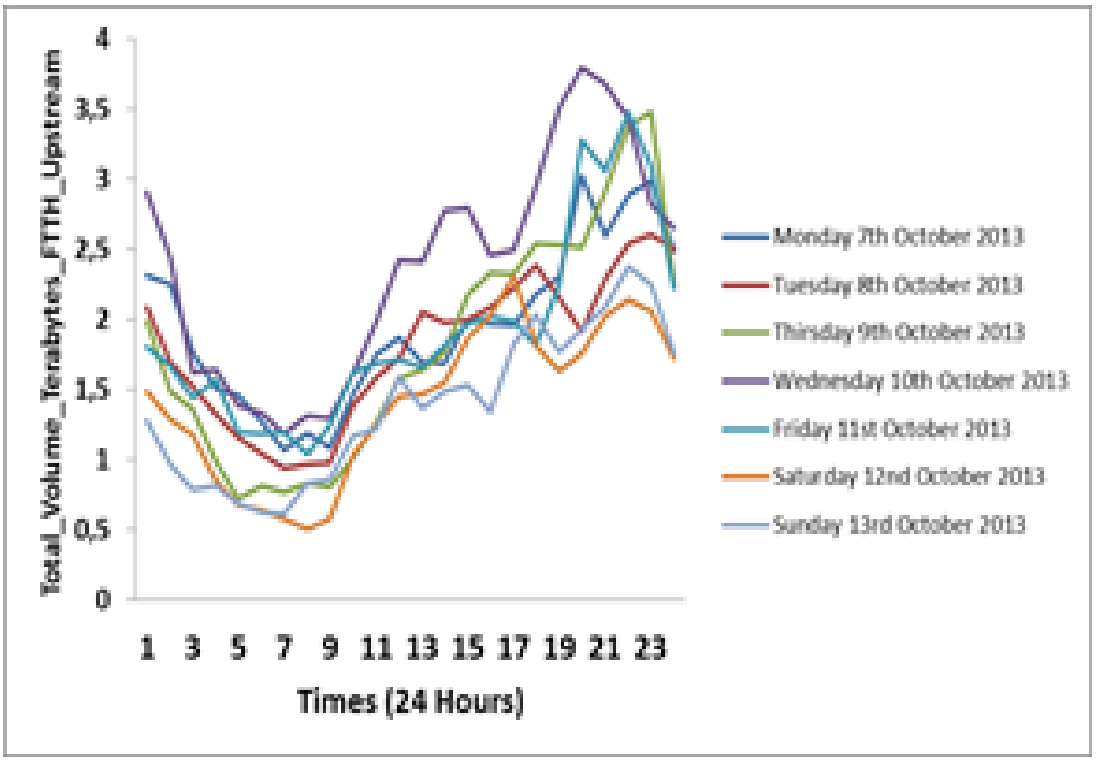


Figure 80. Daily upstream Internet profile for FTTH customers, observation performed over a week of October 2013

The volume generated by all customers (FTTH or ADSL ones) achieve a maximum during the time period [7pm - 10pm]. This time slot represents a busy hours period. We don't observe a difference in the customers' behaviour between the days of the week and the week-end. Indeed, the customers' traffic generally increases after 7pm. Nevertheless, the busy hour period is the same the week-end than the days of the week. The daily upstream Internet profile is similar for ADSL and FTTH customers.

- *Downstream profile for ADSL and FTTH customers*

Figure 81 describes the daily downstream Internet traffic profile for ADSL customers over a period between the 07th of October 2013 and the 13rd of October 2013. These figures demonstrate that there is a similarity in terms of use of the network between the days of the week and the week-end. The downstream volume reaches a maximum in the time slot [7pm – 10pm] for FTTH and ADSL customers.

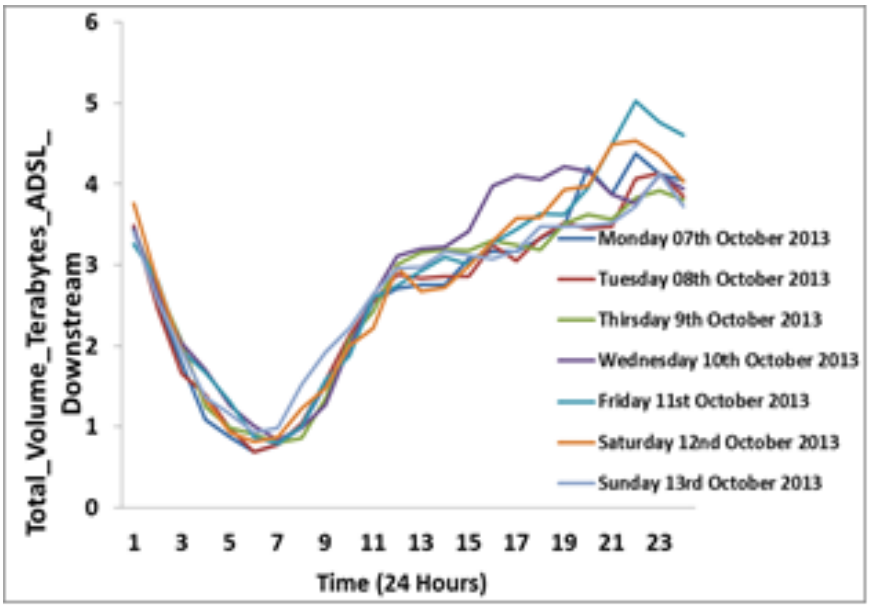


Figure 81. Daily downstream Internet profile for ADSL customers, observation performed over a week of October 2013

4.5.8.3 Comparison between ADSL and FTTH

Analysing the symmetry of upstream and downstream traffic

Table 12 summarizes the average Internet volume generated by FTTH and ADSL customers in upstream (US) and downstream (DS) in October 2013 and gives the ratio between downstream and upstream Internet volume and an average volume generated by the most bandwidth-consuming applications.

	FTTH customers	ADSL customers
Average Volume US (Gbytes) per customer per month	19,183	4,3
Average Volume US P2P (Gbytes) per customer per month	14,94	2,04
Average Volume DS (Gbytes) per customer per month	36,95	24,5
Average Volume DS streaming (Gbytes) per customer per month	13,29	10,54
Average Volume DS Download (Gbytes) per customer per month	9,51	5,023
Average Volume DS P2P (Gbytes) per customer per month	5,8	2,94
Average Volume DS Web (Gbytes) per customer per month	5,36	4,36
Ratio (DS/US) per customer per month	1,92	5,69

Table 12. Ratio for FTTH and ADSL in October 2013

Table 12 shows that the volume generated in the upstream by FTTH customers is 4.5 times higher than the average volume generated by ADSL ones. In downstream, the average Internet traffic volume generated by FTTH customers is 1.5 times higher than the average volume generated by ADSL customers.

In ADSL access networks, the traffic is unbalanced due to the limitation of the physical capacities of the upstream channel. Also, it is quite logic to observe a ratio of 5.69 between the downstream Internet volume and the upstream one. Because of increased performance and capacity of the upstream channel in FTTH access networks, the downstream and upstream Internet traffics are more symmetric in FTTH (the ratio is 1.92). In the optical access networks, file sharing protocols (P2P applications) have a high appetence for such symmetrical network behaviour, and it is observed that the ratio between upstream and downstream P2P volume in FTTH (resp. ADSL) is 2.57 (resp.0.69). This shows that FTTH customers send out more data than they receive in P2P. On the contrary, ADSL customers receive more data than they send in P2P.

This allows us to conclude that some FTTH customers are becoming P2P servers. This is particularly true in a context of coexistence of different access technologies (ADSL and FTTH).

4.5.8.4 Heavy users in FTTH and ADSL access networks

Figure 82 shows that a small proportion of customers generate the most of the traffic in upstream and downstream. In the case of FTTH access networks, we can observe that 20% of the customers generate 98% of the traffic in upstream and 86% in downstream. In the case of ADSL access network, 20% of the customers generate 93% of the traffic in upstream and 84% in downstream. These customers are named “heavy users”.

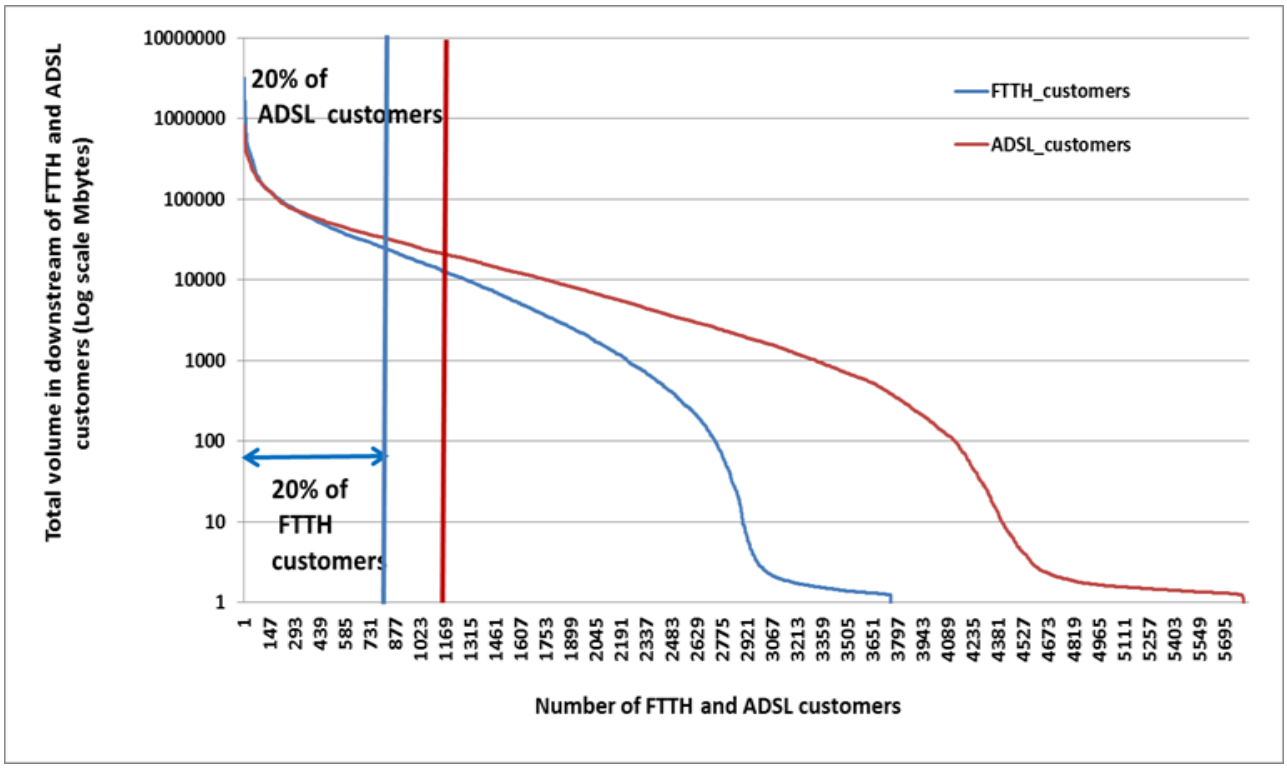


Figure 82. Total upstream Internet traffic volume generated in a day of October 2013 by ADSL and FTTH customers.

Identification of Heavy users

During our studies, we identified Internet heavy users in the upstream and downstream during a day of October 2013. We observed that the upstream Internet heavy users are not the same heavy users than in downstream. This is true for FTTH and ADSL residential customers, but we can have a few heavy users belonging to both upstream and downstream heavy users categories. Some heavy users have a tendency to load the network in uplink with application P2P and other heavy users load the network in downstream using applications like video streaming, downloads, etc.

FTTH VS ADSL

The majority of the volume of traffic generated by heavy users connected by FTTH and ADSL means in upstream is strongly related to the use of applications based on P2P, which generate 70% or more of the total traffic volume.

The volume generated by the heavy users FTTH in upstream is higher than the volume generated by the heavy users in the downstream which indicates that the heavy users FTTH are becoming servers of file sharing protocol P2P for FTTH and ADSL customers. But the volume generated by the heavy users ADSL is very low than the volume generated by heavy users in downstream. This difference is due to the limitation capacity in upstream of ADSL technology. In downstream, the heavy ADSL and FTTH users generate almost the same volume.

The majority of traffic generated by heavy users in downstream is a mixture of applications, some heavy users use a download and others use the application P2P or two applications like web and P2P or download and streaming.

Upstream and downstream TCP and UDP traffic

Figure 83 (resp. Figure 84) first show the upstream (resp. downstream) total volume generated by FTTH customers, and secondly, the associated volume per type of transport protocol. Also, in downstream, 90% of total volume is generated by applications using TCP, mainly involved in video streaming applications, Web browsing, files downloading and P2P applications (for signalling processes). UDP is used for other applications like streaming or VoIP. In upstream, Internet traffic volume generated by TCP is close to the volume generated by UDP. The use of applications like Bit Torrent (P2P) can either be based on TCP or UDP.

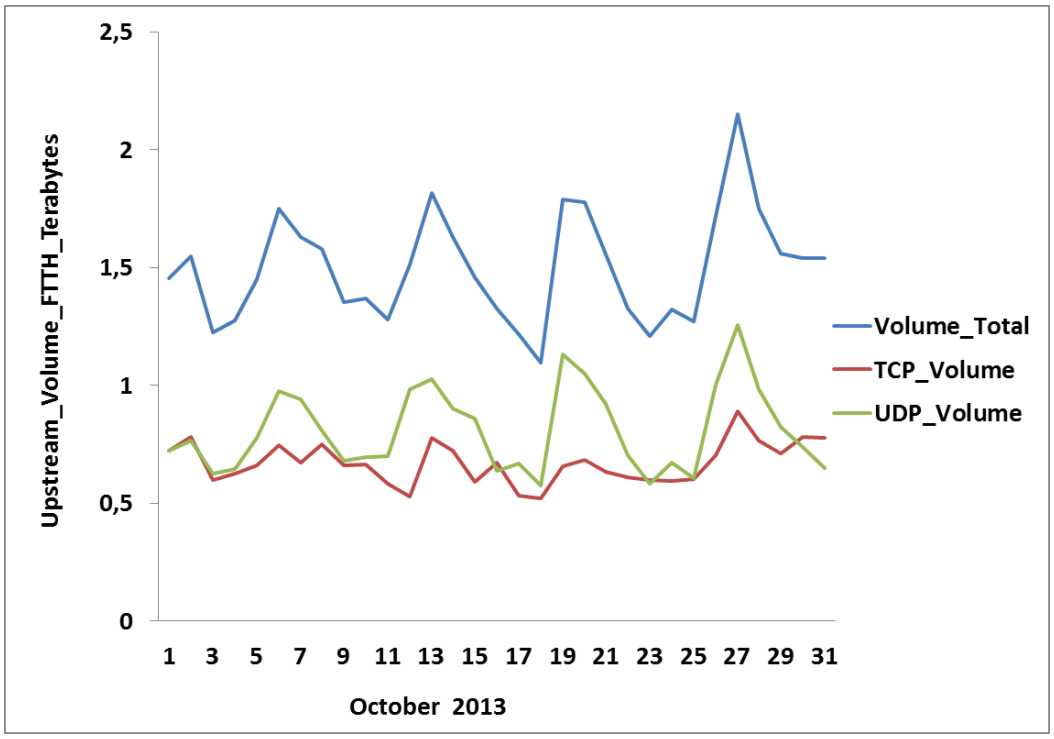


Figure 83. Upstream TCP and UDP traffic volume in FTTH acces networks.

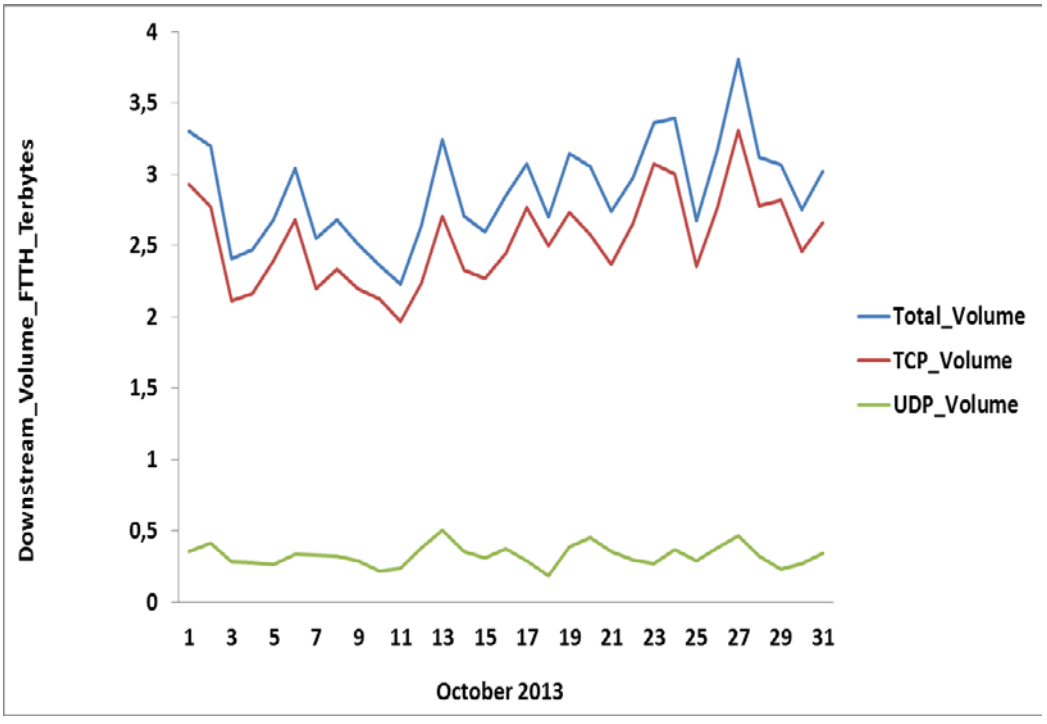


Figure 84. Downstream TCP and UDP traffic volume in FTTH acces networks.

For ADSL customers in downstream, traffic is mainly based on TCP protocol and the UDP traffic is quite low but is also used for some applications such as: streaming, VoIP, P2P.By cons, upstream traffic volume is mainly UDP because the dominant application P2P (Bit Torrent) based on UDP protocol.

4.5.8.5 Conclusion

In upstream, P2P application generates the most of the traffic (in terms of volume).It generates more than 50% of traffic in the case of ADSL and more than 60% in the case of FTTH. In downstream, streaming, downloading, web browsing and P2P applications represent the applications that generate more than 80% of total volume in FTTH and ADSL.

Whatever the day (weekend, weekday), the type of customer (ADSL / FTTH) and the direction of traffic (upstream or downstream), customer profiles are similar, i.e., the volume is very low in the morning and starts to rise in the afternoon and usually reached its maximum between 7 pm and 11 pm.

We can observe that 20% of customers (ADSL or FTTH) generate 80% or 90% of total traffic.

Not all customers use their Internet access in the same manner, some usage tends to load the network in downstream (streaming video, web, download,), and others also request upstream resources (file sharing with P2P protocol).

TCP is the dominant transport protocol in downstream regardless of the type of customer (FTTH or ADSL). In upstream, P2P applications (generally Bit Torrent) use either UDP or TCP. UDP is also used for signalling phases of file sharing protocol P2P and for a few applications like streaming or VoIP.

5 EVOLUTION OF TRAFFIC GROWTH CONSIDERING FMC SCENARIOS

5.1 OFFLOADING SCENARIOS WITH WLAN FEMTOCELLS

5.1.1 INTRODUCTION

Nowadays there is enormous growth of the number of a new generation of mobile devices like various smart phones (iPhone, Android-based), laptops, netbooks, etc. in the market. At the same time, mobile networks operators are incorporating actively Internet applications and services for the mobile devices. There are thousands of web data applications and services available now (e.g. YouTube, Facebook, Spotify, IM, mobile TV, etc.) that are becoming extremely popular in the mobile user environment. According to the Cisco VNI Global Mobile Data Traffic Forecast [105], overall mobile traffic is expected double every year from 2013 onwards.

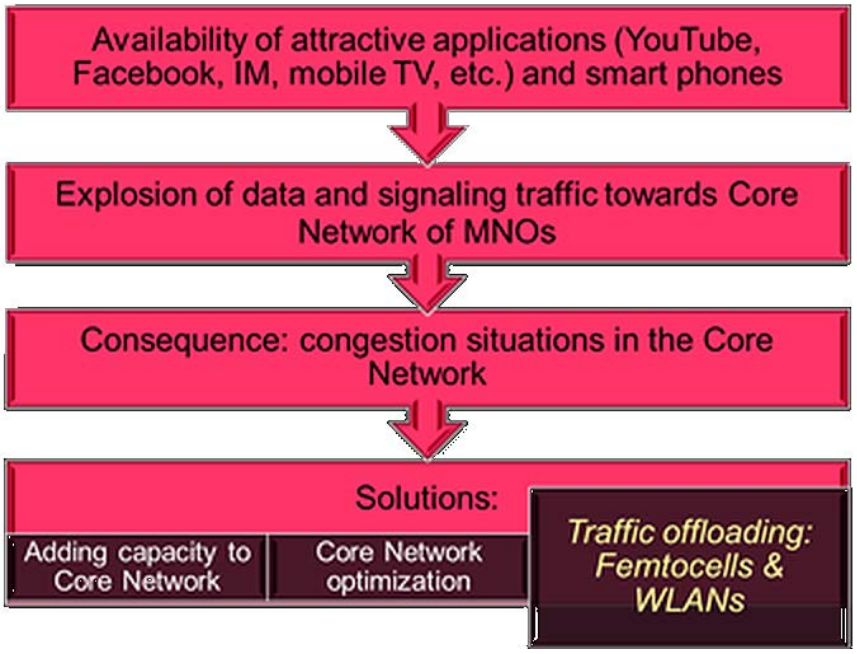


Figure 85: Factors motivating offloading

As a result of these two factors, there is an explosion of both data and signalling traffic towards the core network of Mobile Network Operators (MNOs). As a consequence, congestion situations can arise in the network. Thus, solutions to avoid unnecessary traffic load at network nodes are needed. One such solution is to apply a traffic offloading mechanism by means of femtocells or WLAN. It can solve macro core network capacity crunch avoiding future upgrades of the network infrastructure. The factors motivating the offloading process in the network of the MNOs are mentioned in Figure 85.

It is interesting to consider the traffic offloading process from the viewpoint of a MNO perspective. In this context it is important for the MNO to care about a traffic load that goes to its core network (e.g. Evolved Packet Core (EPC)) after implementing offloading technique. What is needed for the MNO is to know what was before offloading and what happens after in the context of traffic parameters to adjust with them dimensioning characteristics of its network, e.g. to evaluate required system capacity.

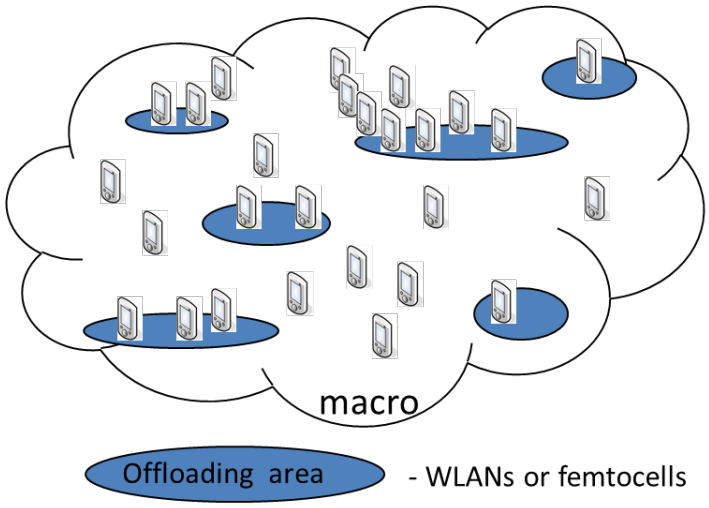


Figure 86: A view of a network implementing offloading

A view of a network implementing offloading is presented in Figure 86. It contains some offloading areas within the MNO network coverage. Thus, with the advent of femtocells, WLANs or Integrated Femto-WLAN solutions [108], the network of the operator could be offloaded when the user is at home or at an enterprise, as her/his traffic would be routed through the corresponding alternative ways. When a user is out of offloading areas traffic goes by traditional way through the macro network of the operator. However, definitely there will be also users who prefer for some reasons to have connections to network services over macro network even if they are inside offloading coverage.

To describe performance of mobile networks implementing offloading an appropriate analytical model is needed. In the next subsection we propose a model for traffic of a single source that eventually reaches the network of the operator after offloading.

5.1.2 Traffic model for offloading

In the same way as in previous literature, the activity of a single source can be modelled as a strictly alternating ON/OFF process [109]. In the model, we introduce the effect of offloading over such source by means of an additional strictly alternating ON/OFF process. The resulting traffic sent to the network of the MNO in a regular way (i.e., the non-offloaded traffic) from a single source is then modelled as the product of the two above processes. Therefore, the aggregation of several such resulting processes models the traffic still needing to be handled by the operator in the conventional way, which is the one to be used for the purposes of network dimensioning. Below we introduce the details and notation of all above processes.

Model of user activity

Thus, the activity of a single source/user is modelled as a strictly alternating ON/OFF process, where ON periods are independent and identically distributed (i.i.d.), OFF periods are i.i.d., and ON and OFF periods are independent. Furthermore, previous measurements have also shown that such periods follow heavy-tailed distributions (e.g., due to file size distributions, web pages) [110] (and references therein). This is also assumed in our model.

Therefore, the process the process Y (t) is a stationary binary time series {Y(t), t ≥ 0} such that [109]:

$$Y(t) = \begin{cases} 1 & \text{activity period} \\ 0 & \text{idle period} \end{cases}$$

A representation of such a process is presented in Figure 87. During the activity period, the user is transmitting and receiving packets. On the other hand, no traffic is exchanged with any network during idle periods, e.g., user reading time of downloaded content. Notice that Y(t) only describes source/user behaviour and is independent from the network through which the traffic is sent, i.e., no offloading considerations have been made in this process.

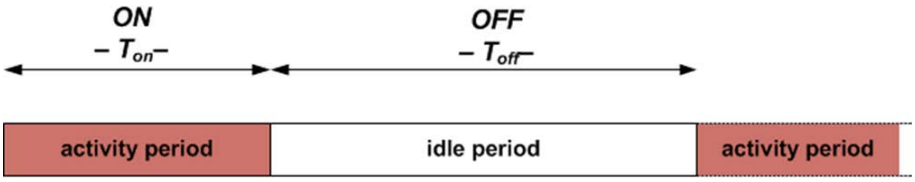


Figure 87: User activity is modelled as a strictly alternating ON/OFF process, Y(t)

Model of offloading periods

Previous measurements carried out in real networks have shown that smartphone connection and disconnection periods to offloading areas follow heavy-tailed distributions [114]. This observation led us to also model offloading periods for a single source as strictly alternating ON/OFF process. In the same way as before, we also assume that ON periods are i.i.d., OFF periods are i.i.d., and ON and OFF periods are independent. Therefore, the process X(t) is a stationary binary time series {X(t), t ≥ 0} such that

$$X(t) = \begin{cases} 1 & \text{user flow sent through MNO network} \\ 0 & \text{user flow sent via offloading network} \end{cases}$$

A representation of such a process is presented in Figure 88. During ON periods, the traffic generated by the source (if any) would be sent towards the network of the MNO in a regular way (i.e., traffic is not offloaded). On the other hand, during OFF periods all traffic generated by the source would be routed through the offloading network (e.g., smartphone under the coverage of a femtocell or Wi-Fi).

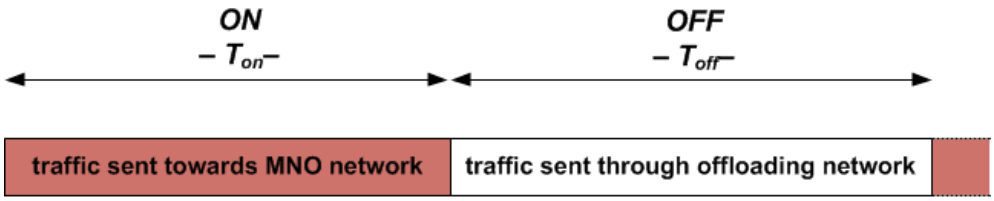


Figure 88: Offloading periods are modelled as a strictly alternating ON/OFF process, X(t)

Model of non-offloaded traffic from a single source

The traffic generated by a single source that is treated in a conventional way by the MNO (i.e., non-offloaded traffic) can be modelled as the product of the previously defined processes. That is,

$$Z(t)=X(t)Y(t).$$

Figure 89 represents such a process, which is also a strictly alternating ON/OFF process, with ON and OFF periods following heavy-tailed distributions and whose characteristic parameters can be derived from those of the original ones. During ON periods, the traffic being generated by the source (i.e., user activity in ON state) is forwarded to the network of the MNO as usual. On the other hand, during OFF periods, either there is no activity from the source or traffic is being sent through the offloading network (e.g., through Wi-Fi, femtocells).

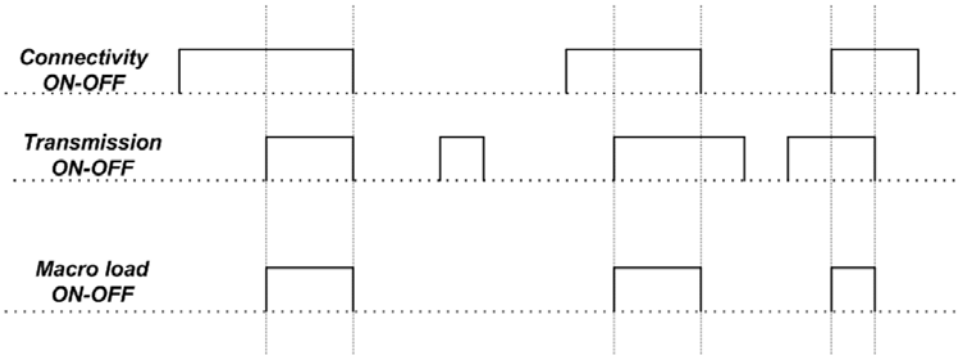


Figure 89: Non-offloaded traffic from a single source is modelled as the product of two strictly alternating ON/OFF processes, Z(t)=X(t)Y(t)

High-level view of the modelling procedure

A high level view of steps for the modelling procedure is as follows. Since, as explained above, Z(t) has ON and OFF periods that follow heavy tailed durations, this process is long-range dependent. Therefore, the aggregation of several such processes is self-similar and can be characterized by means of the Hurst parameter (H). This can be done by means of the same techniques presented in [109]. In turn, parameter H can be derived from the parameters characterizing the heavy-tailed behaviour of the ON and OFF periods of Z(t). Besides, such parameters can be obtained from those of processes of the original processes X(t) and Y(t). Therefore, once the parameters of the original processes are known, by following the above steps, we can characterize the behaviour of the aggregated non-offloaded traffic, and hence, the resources needed in the network of the MNO to serve it.

5.1.3 Applicability of the model

There are various ways in which offloading could be deployed in a network. Currently, the most popular one is the use of Wi-Fi (instead of 3G), when available, for all the traffic exchanged by the user [111]. With the advent of femtocells, some parts of the network of the operator could also be offloaded when the user is at home or at an enterprise, as

her/his traffic would be routed through the femtocell. The proposed model is focused on the traffic that eventually reaches the network of the operator in a regular way (i.e., the non-offloaded traffic). That is, the MNO can compare data traffic (its volume and characteristics) that went through the network before implementing offloading and after. Then, it is possible to evaluate the benefits that offloading generates in the network of the MNO in terms of resource consumption (network capacity). In this sense, the traffic pattern generated towards the network of the MNO in the two above scenarios (i.e, Wi-Fi or femtocells) is the same. That is, all traffic generated by the user when outside coverage of a Wi-Fi access point (or a femtocell) is sent to the network of the MNO as usual. On the other hand, traffic is offloaded when under the coverage of the access point. If a user within the offloading coverage still prefers to use macro connectivity the traffic is treated as non-offloaded. Therefore, the model captures this traffic pattern disregarding the technology in use for offloading, and in this sense, it is agnostic from the specific technology in use.

There are many offloading techniques (e.g., LIPA, SIPTO, I-WLAN, MAPCON, IFOM [111],[112],[113]) that are being defined by 3GPP nowadays. In some techniques offloading (e.g. SIPTO) is done on a PDN connectivity basis, other techniques (e.g. IFOM) realise offloading on a per-flow basis, i.e., some flows are offloaded and some other flows are not when under the coverage of a Wi-Fi access point or a femtocell. In a more generic case, there may be intermediate nodes in which offloading decisions are taken (e.g., the gateway of a company towards the network of the MNO). However, in terms of traffic pattern generated towards the network of the MNO, the resulting flow would behave in the same way as in previous cases. Therefore, the proposed model might still be used for those flows that are offloaded, no matter where the offloading point is in the network (e.g., terminal, femtocell/access point, intermediate node). Hence, the total aggregated traffic toward the network of the MNO would result from the aggregation of offloaded flows (output of the model) with regular flows that are not offloaded at all.

5.2 FMC traffic model for aggregation networks

5.2.1 Introduction

Increased convergence of fixed and mobile networks necessitates a combined network modelling approach to identify optimum locations for core network functions and potentially leverage synergies. The FMC traffic modelling approach documented in this chapter focuses on, but is not restricted to, Germany and can be easily generalised. Figure 90 depicts our modelling approach in terms of collected data which has been used in the modelling process, and can serve as guideline in future FMC modelling attempts.

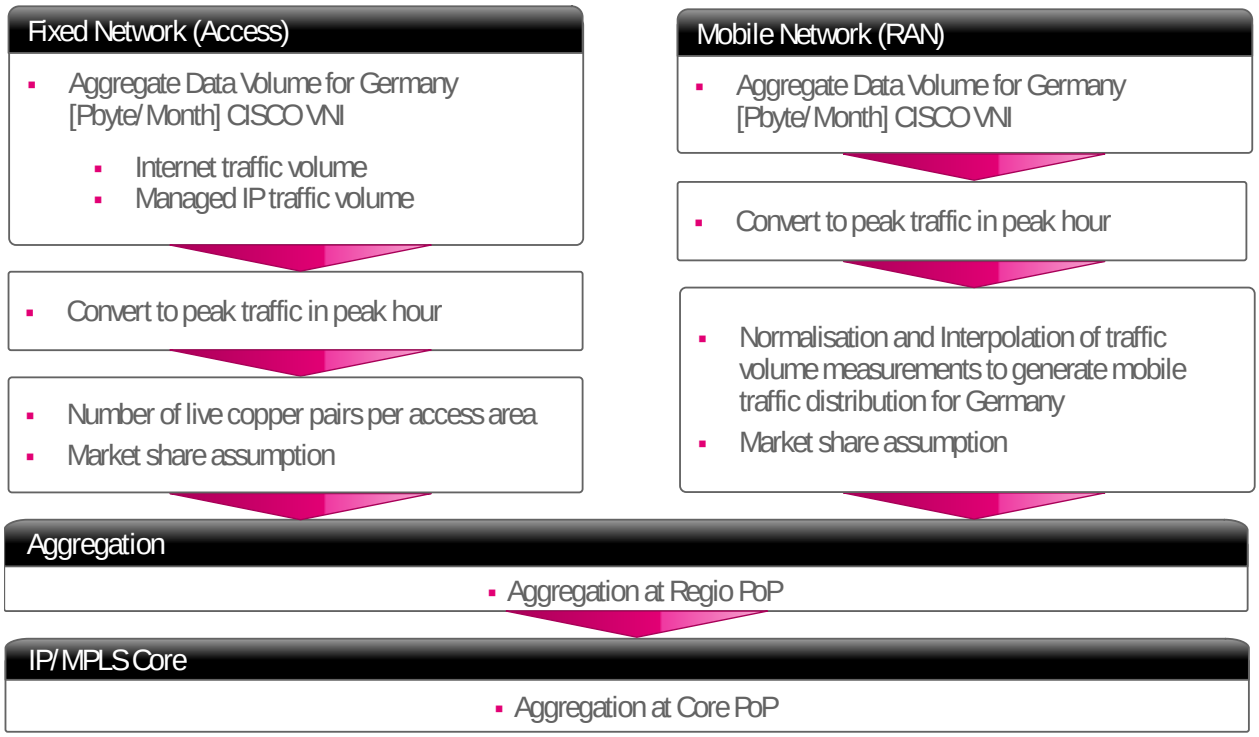


Figure 90: FMC modelling methodology

To build on a common and openly accessible foundation, the CISCO visual networking index has been chosen to provide the forecast data as input for the traffic model. To identify region or core PoP locations, struggling to keep up with corresponding throughput requirements, we will use the *peak traffic in peak hour* as metric. This metric enables us to dimension aggregation and core networks, because traffic contributions of a high number of independent customers are multiplexed together rendering traffic fluctuations on short time scales irrelevant. However, at the edge of the network significantly less customers are aggregated, necessitating a more refined statistical model operating on shorter time scales to cope with inherent traffic burstiness. Since the data provided by the CISCO VNI is available as *traffic volumes per month* it has to be converted to peak traffic in peak hour as indicated in Figure 90.

Aggregated peak traffic in peak hour has to be distributed across all access areas according to certain metric. In case of the fixed network, this metric is the number of copper pairs available for xDSL connections per fixed network access area, which gives a reasonable indication of the regionally distributed traffic. Unfortunately, in case of the

mobile traffic, this distribution is harder to come by, which is due to the space varying nature of mobile traffic. Here we used traffic volume measurements as well as estimates for cell coverage areas to interpolate a Germany-wide mobile traffic distribution. To operate on a common geographical metric, this distribution is tied to fixed network access areas. Fixed and mobile traffic components are scaled by a market share assumption. Finally the fixed and mobile traffic is aggregated together via a common fixed/mobile aggregation network (Figure 91 shows such an example) consisting of region and core PoPs.

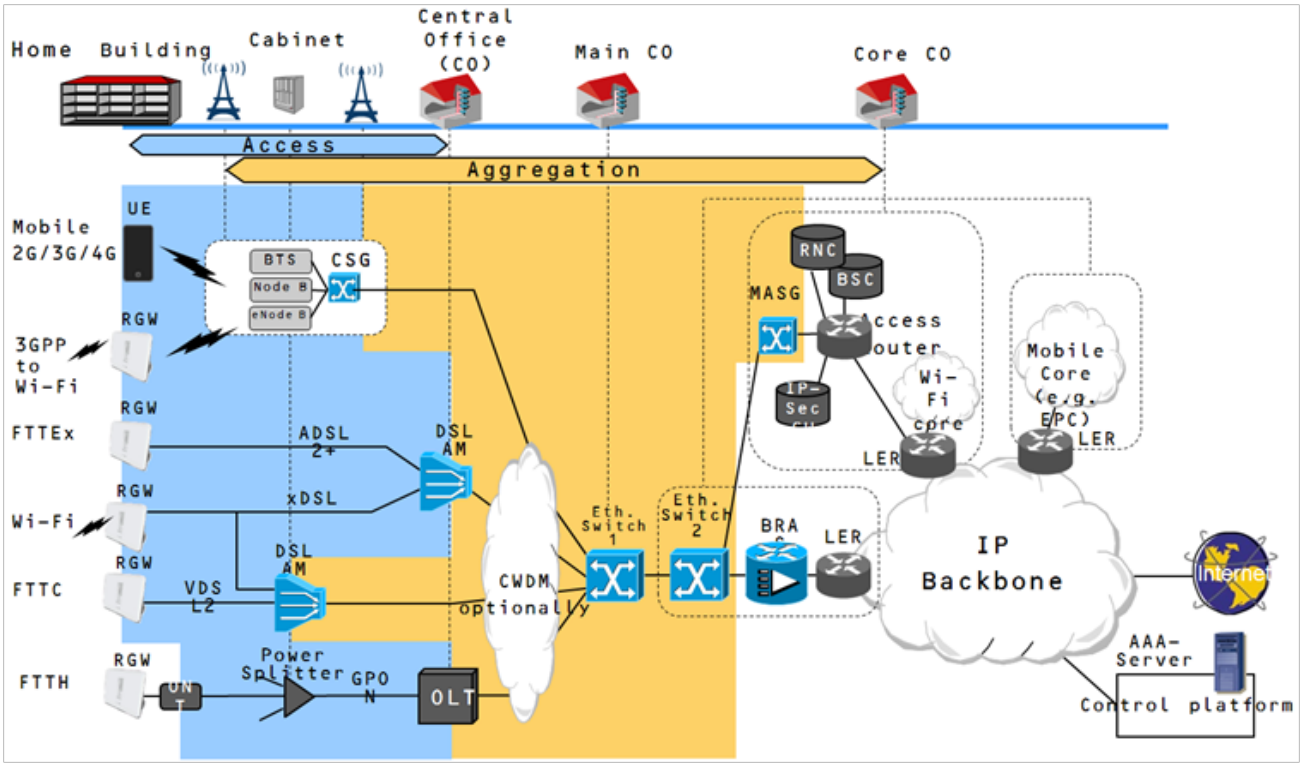
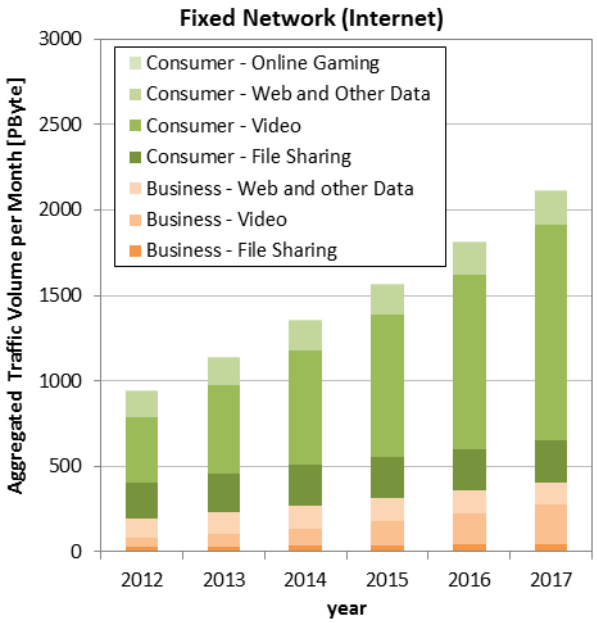


Figure 91 Example of FMC in current Networks

5.2.2 Traffic Volume Forecast

The traffic model introduced here is based on traffic forecast data from CISCO visual network index 2012 [105], which forecasts the IP traffic evolution up until the year 2017. The study distinguishes between fixed and mobile traffic volume per month, where the fixed traffic share is again subdivided into Internet and managed IP (largely IPTV). Furthermore a basket of applications is introduced covering online gaming, web and other data, and video and file sharing for consumer and business customers. Figure 92 and Figure 93 show fixed network traffic volume predictions for Germany from 2012 to 2017. The dominant fixed network traffic share (Internet as well as managed IP) is traffic attributed to video streams, which is expected to become even more important in the future. In terms of fixed network Internet traffic, web and other data as well as file sharing traffic constitute important proportions of the overall traffic, but do not fuel the growth. Business traffic accounts for roughly 20% of the overall IP Internet traffic and for 43% of managed IP traffic. Note that the share of business traffic declines to 28% in 2017, even though the absolute traffic is increasing slightly.

(a)



(b)

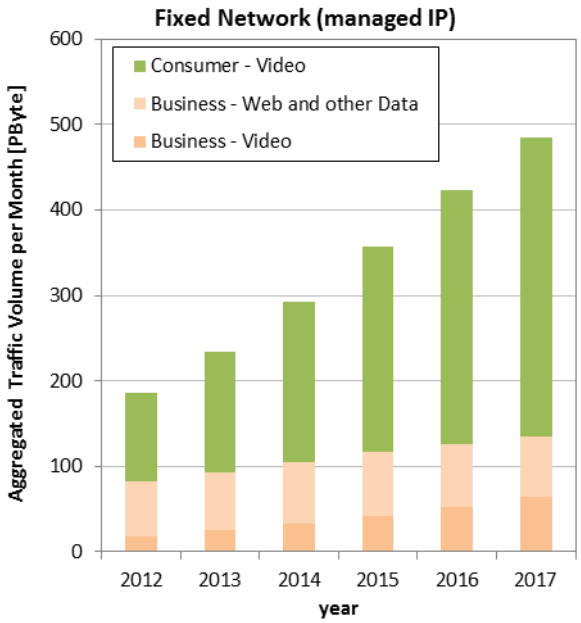


Figure 92: fixed network Internet traffic volume (a) and managed IP traffic volume for Germany projected until 2017

Figure 93 depicts mobile network traffic volume in Germany from 2012 to 2017. Here, video and web and other data services are the most important drivers for traffic growth, while file sharing accounts for 6% in 2012 (declining to 2% in 2017). Similarly to the fixed network Internet traffic, business traffic accounts for roughly 20% of the overall traffic in mobile networks as well.

Note that shares of each application and customer classes are given for completeness only. For the sake of reduced complexity, we will restrict ourselves in the following to the categories *Internet* and *managed IP* for *fixed network traffic* as well as *mobile network traffic* without any subdivisions.

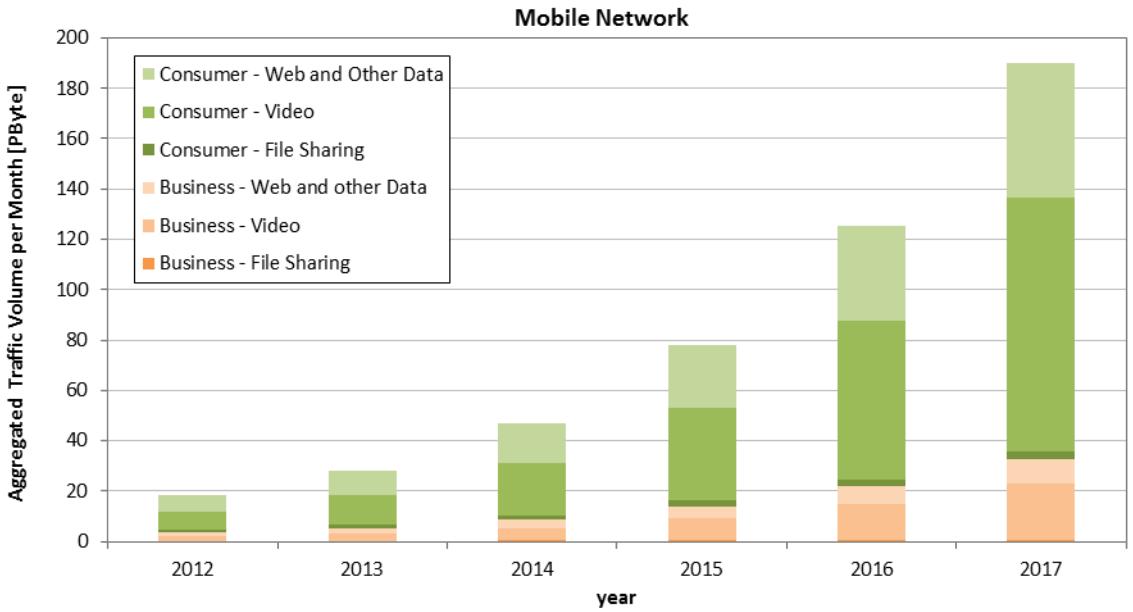


Figure 93: mobile traffic volume for Germany projected until 2017

Figure 94 directly compares the growth of fixed and mobile network IP traffic volume according to CISCO VNI for Germany. It can be seen, that mobile traffic is expected to grow much faster than fixed network traffic with a compound annual growth rate (CAGR) of more than 50%, whereas fixed network Internet and managed IP traffic grow only with around 20-25% over the next years. Consequently, mobile traffic which corresponded in 2012 to only 1.6% of the fixed network traffic volume is expected to rise to 7.3% in 2017. This development, which is certainly due to the expected proliferation of smartphones exploiting 4G and future 5G mobile networks, underlines the growing importance of FMC traffic models, since the growing mobile traffic has to be backhauled and aggregated via fixed network infrastructure.

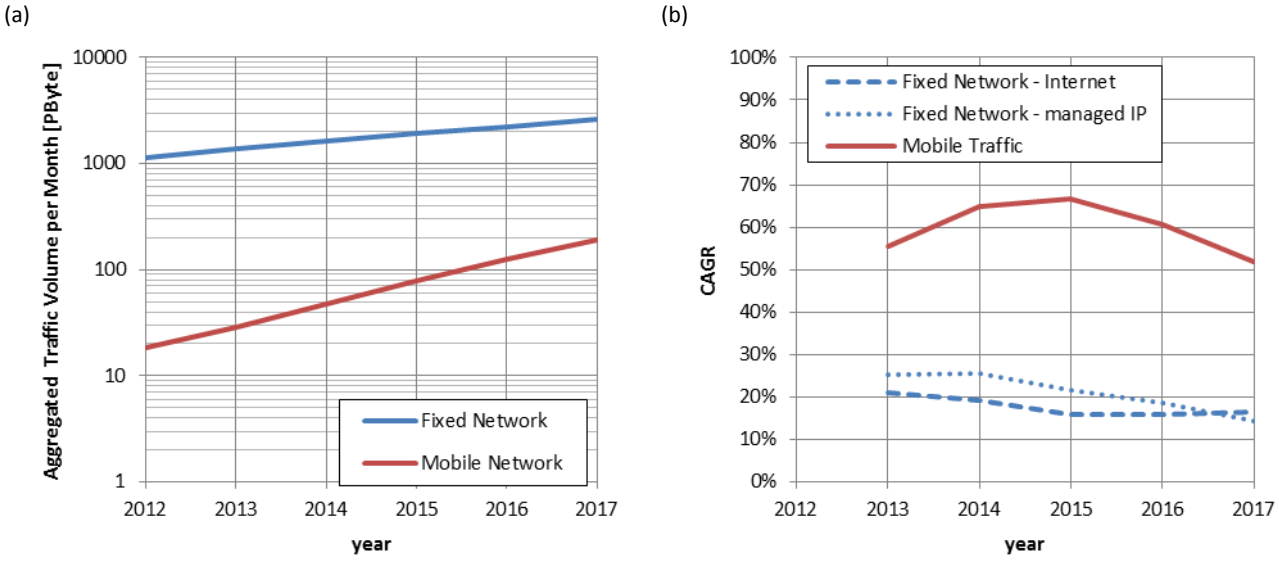


Figure 94: (a) Traffic Volume per month for fixed and mobile traffic, as well as (b) compound annual growth rate (CAGR) for Internet and managed IP traffic in the fixed network and mobile traffic according to CISCO VNI for Germany

The peak traffic in peak hour is generally defined as the average traffic during the busiest hour of the day. To convert aggregate traffic volume per month to peak traffic in peak hour the following formula has been used:

$$\text{Peak Traffic in Peak Hour} = \frac{\text{Traffic Volume per month} \cdot 8\,{}^{\text{bit}}/_{\text{Byte}}}{17 \cdot 30.44\text{d} \cdot 3600\text{s}}$$

assuming that 1/17 of the overall diurnal traffic is accumulated during the peak hour as depicted in Figure 95.

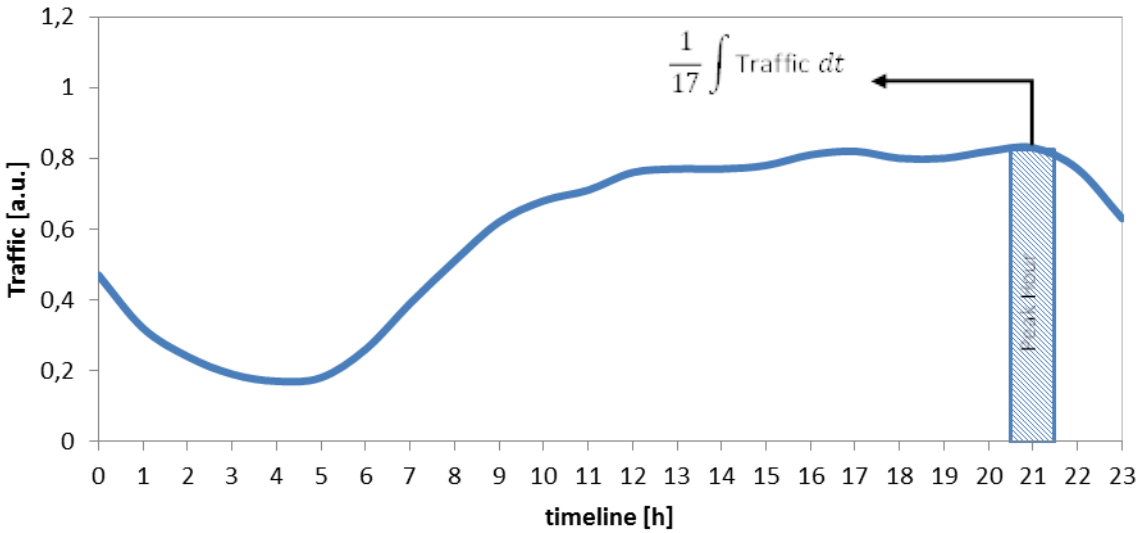


Figure 95: Typical load curve for a remote access router, with 1/17 of the diurnal traffic being accumulated in the peak hour.

5.2.3 Regionalisation of aggregated traffic

Aggregated mobile traffic described in the previous section has been regionalized by using per-cell traffic volume measurements, estimations of coverage areas for particular cells as well as geo-referenced fixed network service area boundaries. In the following the methodology is explained based on an exemplary access area.

Initially, the spatial overlap between cell coverage and each fixed network access area is determined as shown in Figure 96 (a). Subsequently, the traffic volume measured for each cell is distributed according to the overlap between cell coverage and access area (see Figure 96 (b)).

(a) (b)

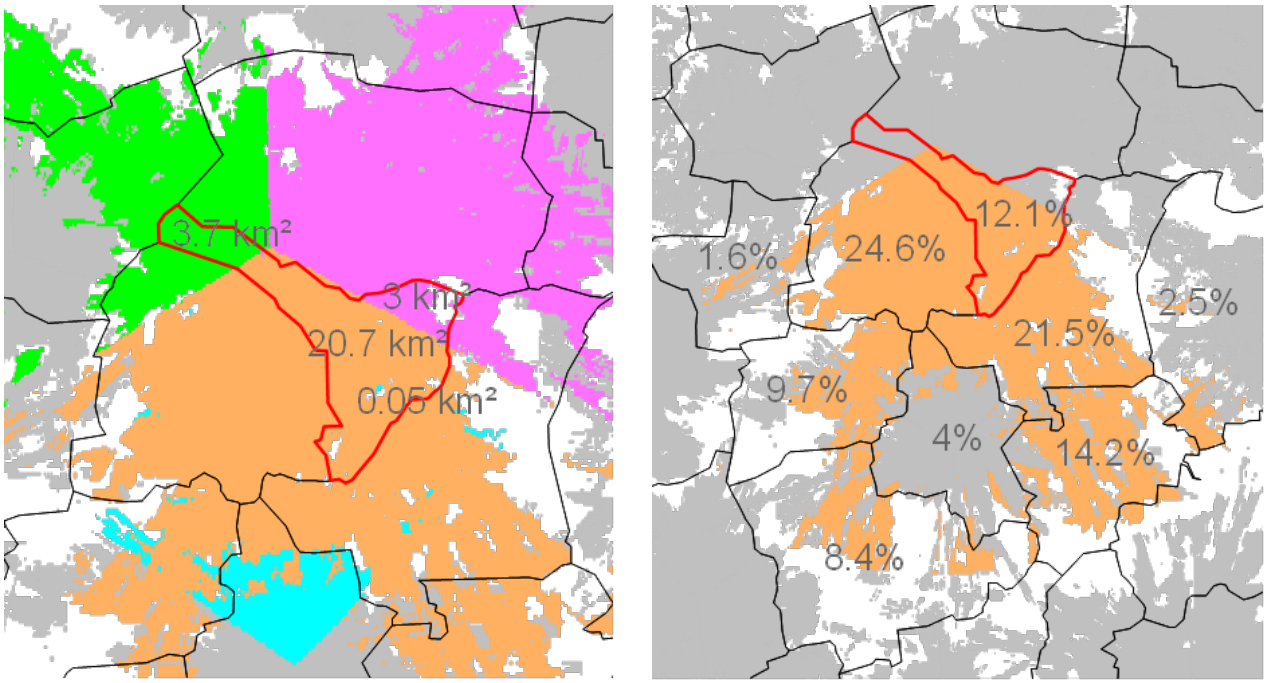


Figure 96: (a) spatial overlap between multiple mobile cells and an exemplary fixed network access area, (b) traffic of each cell is distributed to access areas, covered by the corresponding cell

As a next step, it is possible to sum up data volumes from multiple cells and determine the overall volume originating in the corresponding access areas as depicted in Figure 96. However, there are white spots which are not covered by any mobile cell (Figure 96 upper right hand corner of access area). In this case, we still assign a certain data volume to this currently unserved area under the assumption that 100% of each access area will be covered in the future to improve customer service. If more than 60% of the corresponding access area is currently covered by existing cells, data volume contributions from unserved areas are interpolated according to the data volume density in the served area. Otherwise (less than 60% of the access area is covered), uncovered white spots are interpolated according to the mean data volume density of access areas of the same cluster (dense urban, urban, rural). To achieve this, all fixed network service areas have been clustered according to their population density (population/area):

$$\textbf{Traffic density} = \frac{\text{Sum of UMTS } \textbf{data volume} \text{ of all access areas within cluster}}{\text{Sum of UMTS } \textbf{coverage area} \text{ of all access areas within cluster}}$$

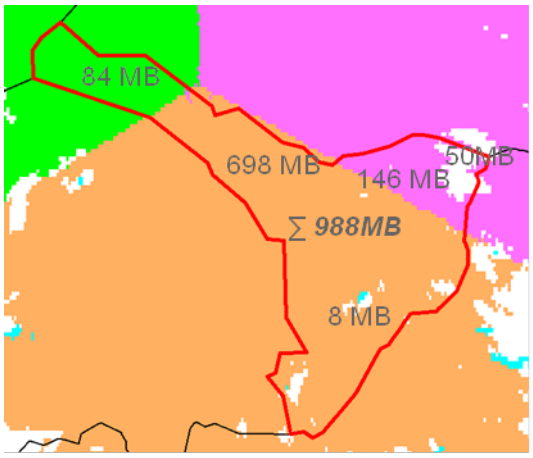


Figure 97: data volumes originating in exemplary access area

Figure 97 (a) shows the result of the regionalization in terms of the varying traffic density across Germany. Due to the lack of customer mobility, fixed network traffic per access area can be regionalized by using the number of live copper pairs per access area in Germany as shown in Figure 97 (b).

(a) (b)

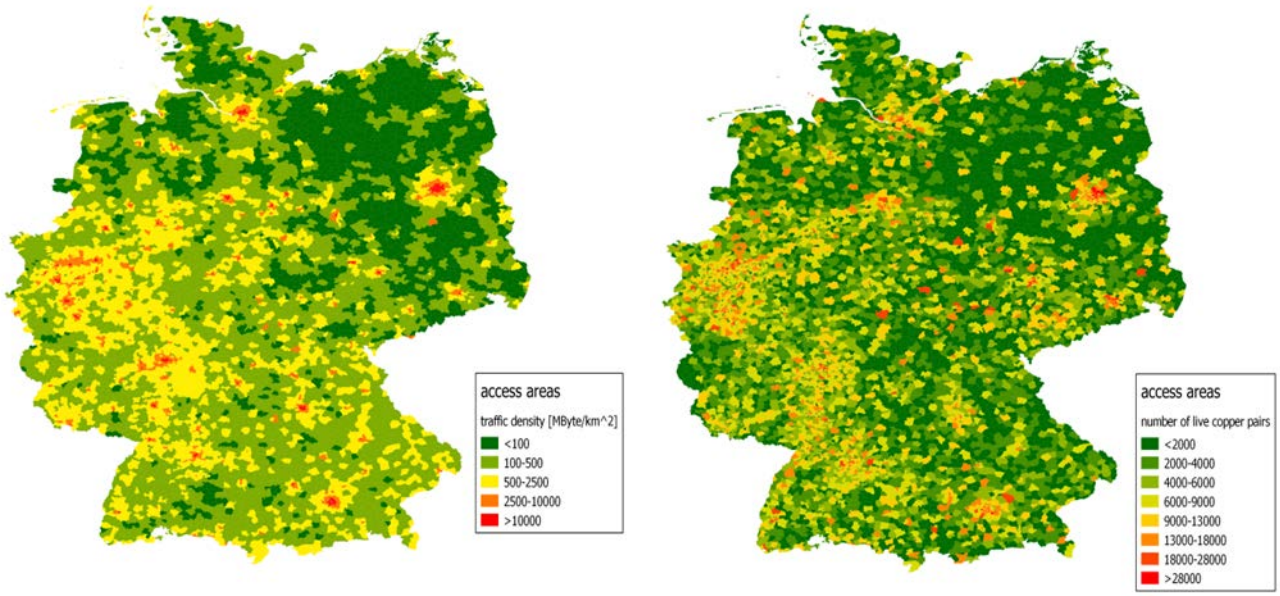


Figure 98: Base to regionalize forecasted traffic for (a) mobile network [traffic density per access area] and (b) fixed network [number of live copper pairs]

After assigning each access area its corresponding peak traffic in peak hour share, the value is scaled by the Deutsche Telekom market share which is 33.15% for fixed network and 31.2% for the mobile network as published by the German regulating authority responsible for the telecommunication sector [115].

Fixed network traffic originating in the access areas is initially concentrated in local exchanges and than aggregated at region PoPs as well as core PoPs as depicted in Figure 91. On the mobile network side traffic is originating from the antennas,

concentrated via a common fixed/mobile aggregation network (region and core PoPs) and transported to the mobile core network. Figure 99 features a potential fixed/mobile aggregation network topology with more than 800 region PoPs as well as 18 core PoPs at 9 locations. The present topology is designed to aggregate large traffic volumes which are assumed to be related to the population density.

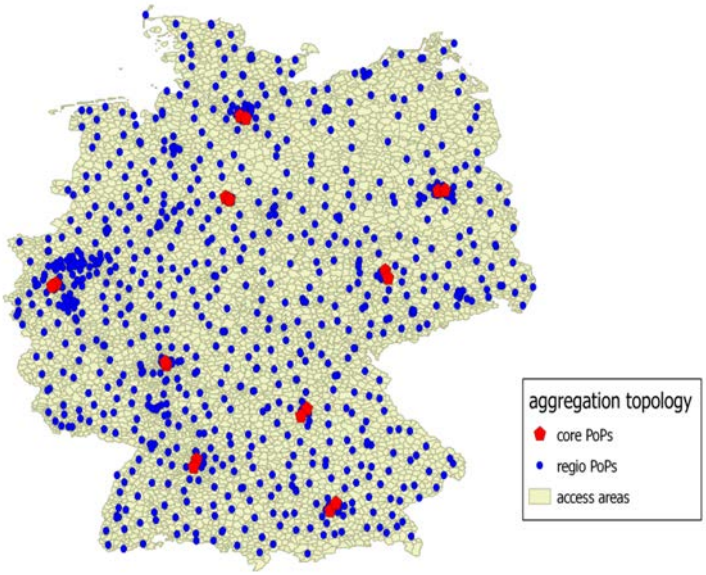


Figure 99: Potential Germany-wide aggregation topology with region and core PoPs

5.2.4 Traffic Evolution at Region and Core PoPs

Figure 100 shows the projected rise in IP traffic of the busiest region and core PoPs for fixed and mobile network traffic between 2012 and 2017. The busiest region PoP does not exceed a peak traffic of 15.1 Gbps fixed network traffic and 1.3 Gbps mobile network traffic by the end of 2017. In 2017 the fixed network traffic at the busiest region PoP will not even have doubled compared to 2013, whereas mobile traffic at the busiest PoP will increase by more than a factor of 6 during the same time span.

Considering the core network: the busiest core PoP will experience fixed network traffic of 390 Gbps in 2017, while the busiest mobile network core PoP increases to 21.3 Gbps (the entire mobile network traffic in 2012 amounts to 24 Gbps). Again, mobile traffic at the busiest core PoP increases by more than a factor of 6 between 2013 and 2017, while fixed network traffic not even doubles. Note that, the fixed network traffic will probably be lower in reality, since IPTV traffic (the dominant part of managed IP traffic – see Figure 92) is multicast traffic leading to a multicast gain in core and aggregation networks.

(a) (b)

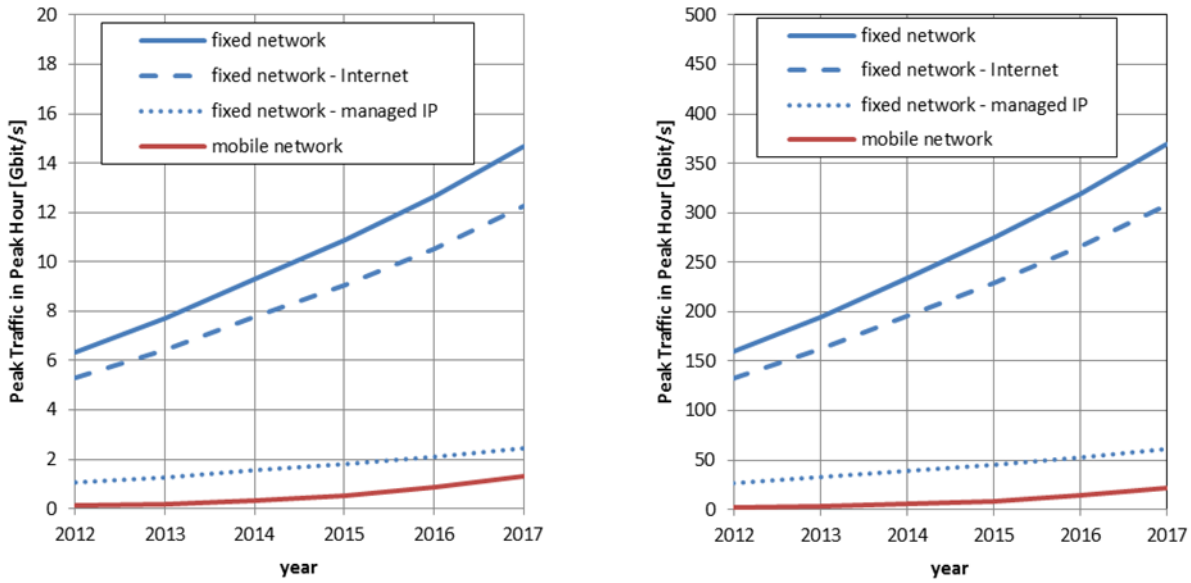
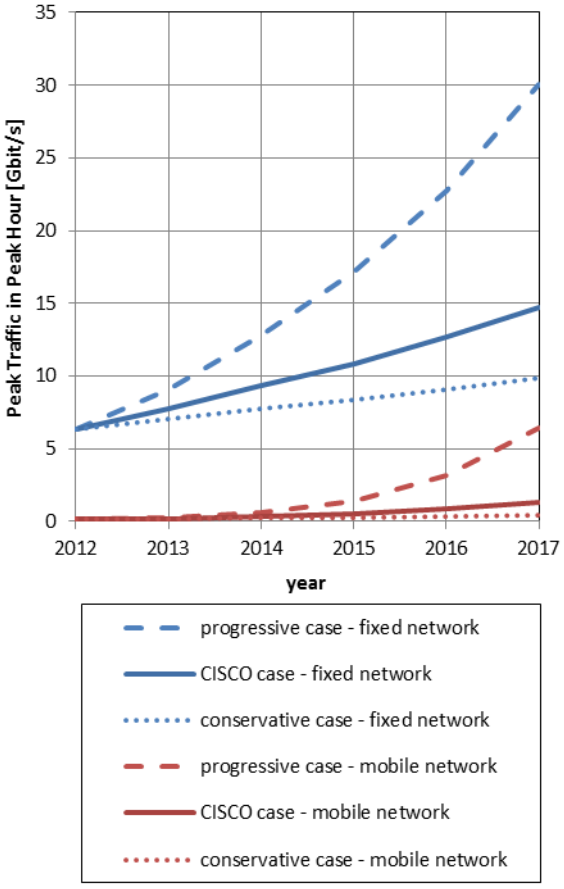


Figure 100: Peak Traffic in Peak hour of the busiest region (a) and core PoPs (b) for fixed and mobile network traffic

To cover potential variations in traffic growth a progressive and conservative case have been introduced as shown in Figure 101. In the progressive case, the traffic is assumed to grow twice as fast compared to the CISCO predictions, while in the conservative case the traffic grows at half the CISCO rate. As a consequence the fixed network traffic doubles in 2017 when comparing the progressive case to the CISCO predictions, while the conservative case leads to 35% less fixed network traffic. In the mobile network the development is more extreme due to a generally higher growth rate (see Figure 94 b): the progressive case leads to an increase by a factor of 5 in 2017 compared to the CISCO predictions, while the conservative traffic assumption leads to a 65% reduction in mobile network traffic, again compared to the CISCO case.

(a)



(b)

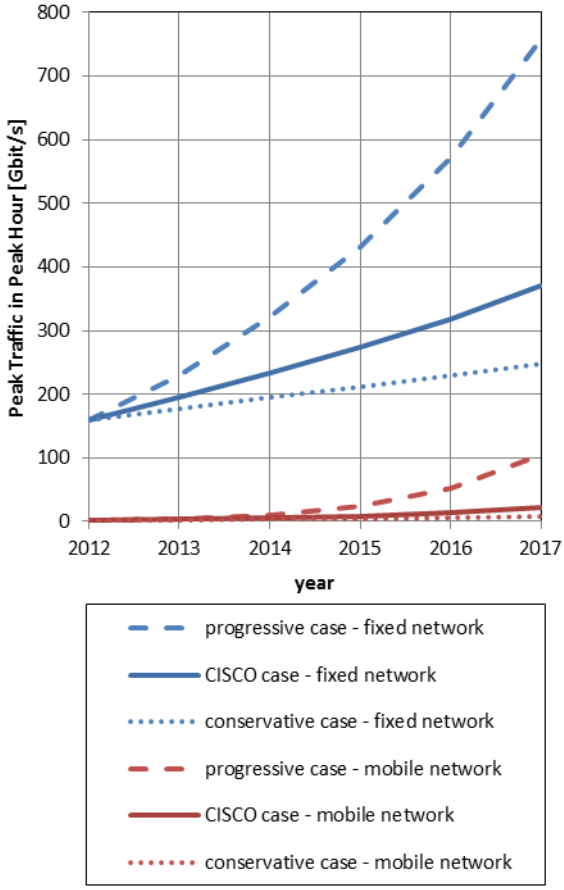


Figure 101: Peak Traffic in Peak hour of the busiest region (a) and core PoPs (b) for progressive, CISCO and conservative cases

Figure 102 depicts the cumulative distribution of peak traffic for region and core PoP in the year 2017 of the CISCO case. The average region PoP deals with 4.1 Gbps of peak traffic in the peak hour, compared to the average core PoP, which has to deal with 205 Gbps. While less than 3% of the region PoPs exhibit a traffic of more than 10 Gbps, the upper 10% of the core PoPs process 390 Gbps. The median for region PoPs lies at 3.4 Gbps and for core PoPs at 194 Gbps.

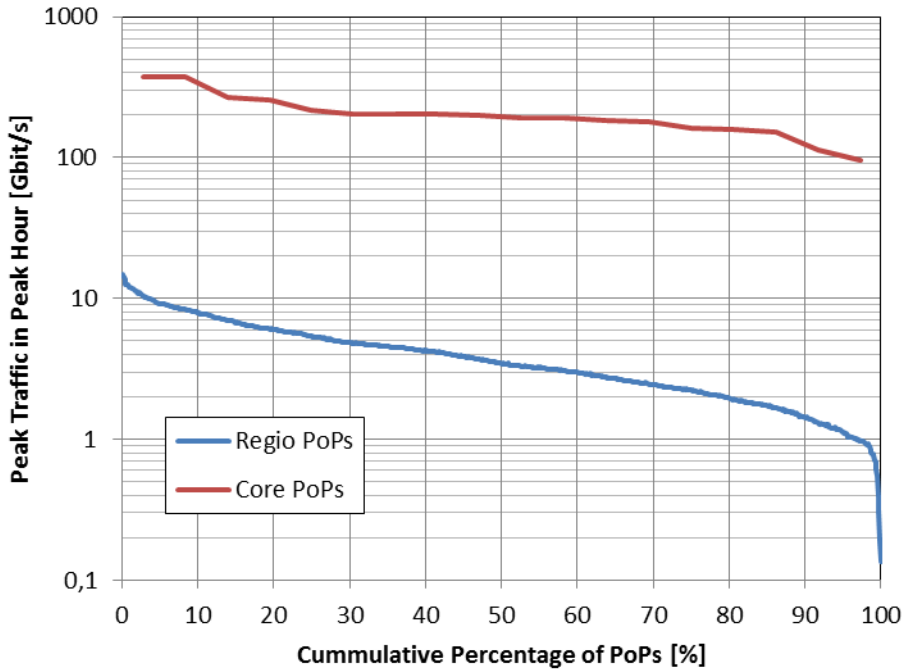


Figure 102: Cumulative distribution of peak traffic in peak hour at region and core PoPs in 2017 (CISCO case)

6 CONCLUSIONS

This deliverable serves as a reference point of traffic-related considerations within COMBO. It provides a snapshot of currently visible trends for the ways traffic changes, and provides modelling instruments for revealing the complex patterns of traffic. These trends – and their timing – affect architectural plans, capacity plans, decisions for offloading scenarios, performance management considerations, and techno-economic studies.

The main drivers for traffic growth are well-known: the increasing number of devices, the need for high definition, streaming media, the introduction of new services and applications, and the effect of these factors on bandwidth demands. Various sources predict a 1000-times traffic increase within a 20 years span – if not earlier.

This document contains traffic analysis results from various parts of the network and service domains of various operators. It does not attempt to cover the countless combination of analysis possibilities at all domains. The aim was less than that: to show traffic analysis instrumentation at work, answering questions at various domains of the architecture (both fixed and mobile, aggregation and core), and providing methodologies for modelling the traffic – based on the analysis results. Since the actual FMC traffic scenarios are not live yet, at this stage it is not possible to provide models based on analysis results. Nevertheless, it is possible to provide the building blocks of such models and to define the methodology on building such models. During WP6 tasks at COMBO it is worth considering building the actual traffic models for live (or laboratory) scenarios, based on the results of the current document.

Although analysing data of “FMC traffic” was not possible at this stage, this deliverable describes how traffic patterns and models are expected to be affected by some of the FMC scenarios. There are two models detailed in this document: one for offloading scenarios with WLAN femtocell, and another for evolution of traffic composition within aggregation networks.

7 REFERENCES

[1] M. Scott, “Study on Impact of traffic off-loading and related technological trends on the demand for wireless broadband spectrum”, 2013.

[2] UMTS Forum Report 44 “Mobile traffic forecasts 2010-2020© UMTS Forum January 2011”, June 2013.

http://www.umts-forum.org/component/option,com_docman/task,doc_download/gid,2537/Itemid,213/

[3] Cisco. Cisco visual networking index: Forecast and methodology, 2011-2016. White paper, 2012.

http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf

[4] Solon. Broadband on demand: Cable’s 2020 vision. White paper, 2011.

[5] Minnesota Internet Traffic Studies (MINTS), Online: http://www.dtc.umn.edu/mints/

[6] Kilper, D.; Atkinson, G.; Korotky, S.; Goyal, S.; Vetter, P.; Suvakovic, D. & Blume, O.; Power Trends in Communication Networks; IEEE Journal of Selected Topics in Quantum Electronics, 2011, 17, 275-284

[7] Cisco. The zettabyte era. White paper, 2013.

[8] Korotky, S. K.; Traffic Trends: Drivers and Measures of Cost-Effective and Energy-Efficient Technologies and Architectures for Backbone Optical Networks; Optical Society of America, 2012

[9] AT&T Labs. Traffic types and growth in backbone networks. Technical report, 2011.

[10] ITU-R M.2243. Assessment of the global mobile broadband deployments and forecasts for international mobile telecommunications. Technical report, 2012.

[11] Ericsson. Traffic and market report. White paper, 2012.

[12] A.T. Kearney, GSMA, “The Mobile Economy 2013". http://www.atkearney.com

[13] “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast 2012–2017”, Cisco Systems.

[14] “25 labs”, Technology blog, 8 September 2011. 25labs.com

[15] Analysys Mason “Global mobile network traffic” ,29 June 2011.

[16] “Facebook’s Mobile Users, Q4 2010 - ”, May 2013, available at http://trends.e-strategyblog.com/2013/05/08/facebooks-mobile-monthly-active-users/10979

[17] D. Wonak, “A storm is brewing – an LTE signalling storm”, Diametriq, white paper, September, 2012.

[18] Tekelec LTE Diameter signalling index: forecast report and analysis 2011-2016, 2012

[19] 3GPP TR 23.843, “Study on Core Network Overload Solutions”, Rel’12, V0.8.0 (2013-04)

[20] M. Olsson, et al., “SAE and the Evolved Packet Core: Driving the mobile broadband revolution”, Elsevier, 2009.

[21] C. Knight “Measuring the explosion of LTE signalling traffic – a Diameter traffic model”, Diametriq, white paper, September 2012.

[22] Diametriq Diameter Traffic Calculator, http://www.diametriq.com/lte-signalling-traffic-model/

[23] K. Iniewski, et al. “Convergence of mobile and stationary next-generation networks”, Wiley, 2011.

[24] Balakrishnan Chandrasekaran, “Survey of Traffic Models”, available at http://www.cse.wustl.edu/~jain/cse567-06/ftp/traffic_models3/index.html.

[25] Jacques Lévy Véhel, Rudolf H. Riedi, “Fractional Brownian motion and data traffic modelling: The other end of the spectrum”, INRIA Rapport de Recherche August 1997.

[26] K. Vishwanath, A. Vahdat, “Swing: Realistic and Responsive Network Traffic Generation”, *IEEE/ACM Transactions on Networking 2009.*

[27] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, D. Veitch, “The Multiscale Nature of Network Traffic”, *IEEE Signal Processing Mag.*, vol. 19, no. 3, pp. 28-46, May 2002.

[28] K. Park, W. Willinger, "Self-Similar Network Traffic: An Overview", in *Self-Similar Network Traffic and Performance Evaluation*, K. Park, W. Willinger, Eds. Chichester, UK: John Wiley & Sons, 2000, pp. 1-38.

[29] P. Abry, P. Flandrin, M. S. Taqqu, D. Veitch, "Wavelets for the Analysis, Estimation, and Synthesis of Scaling Data", in *Self-Similar Network Traffic and Performance Evaluation*, K. Park, W. Willinger, Eds. Chichester, UK: John Wiley & Sons, 2000, pp. 39-88.

[30] V. Paxson, S. Floyd, “Wide-Area Traffic: the Failure of Poisson Modelling”, *IEEE/ACM Trans. Networking,* vol. 3, no. 6, pp. 226-244, June 1995.

[31] N. G. Duffield, Neil O'Connell, "Large Deviations and Overflow Probabilities for The General Single Server Queue, with Applications", *Math. Proc. Cam. Phil. Soc.*, vol. 118, pp. 363-374, 1995.

[32] I. Norros, "On The Use of Fractional Brownian Motion in the Theory of Connectionless Networks", *IEEE J. Select. Areas Commun.*, vol. 13, no. 6, pp. 953–962, Aug. 1995.

[33] Y. Cheng, W. Zhuang, "Calculation of Loss Probability in a Partitioned Buffer with Self-Similar Input Traffic", *Proc. IEEE GLOBECOM 2004*, Dallas, TX, USA, Nov. 2004.

[34] J. Choe, N. B. Shroff, "Queueing Analysis of High-Speed Multiplexers Including Long-Range Dependent Arrival Processes", *Proc. IEEE INFOCOM 1999*, New York, NY, USA, March 1999.

[35] G. W. Wornell, A. V. Oppenheim, "Estimation of Fractal Signals from Noisy Measurements Using Wavelets", *IEEE Trans. Signal Processing*, vol. 40, no. 3, pp. 611-623, March 1992.

[36] M. S. Taqqu, V. Teverovsky, W. Willinger, “Estimators for Long-Range Dependence: an Empirical Study”, *Fractals*, vol. 3, no.4, pp. 785-798, 1995.

[37] P. Abry, P. Gonçalvès, P. Flandrin, "Wavelets, Spectrum Analysis and 1/*f* Processes", in *Lecture Notes in Statistics: Wavelets and Statistics*, vol 103, A. Antoniadis, G. Oppenheim, Eds. Springer, 1995, pp. 15-29.

[38] P. Abry, D. Veitch, “Wavelet Analysis of Long-Range Dependent Traffic”, *IEEE Trans. Inform. Theory*, vol. 44, no.1, pp. 2-15, Jan. 1998.

[39] D. Veitch, P. Abry, “A Wavelet-Based Joint Estimator of the Parameters of Long-Range Dependence”, *IEEE Trans. Inform. Theory*, vol. 45, no.3, pp. 878-897, Apr. 1999.

[40] D. W. Allan, J. A. Barnes, "A Modified Allan Variance with Increased Oscillator Characterization Ability", *Proc. 35th Annual Freq. Contr. Symp.*, 1981.

[41] P. Lesage, T. Ayi, "Characterization of Frequency Stability: Analysis of the Modified Allan Variance and Properties of Its Estimate", *IEEE Trans. Instrum. Meas.*, vol. 33, no. 4, pp. 332-336, Dec. 1984.

[42] L. G. Bernier, "Theoretical Analysis of the Modified Allan Variance", *Proc. 41st Annual Freq. Contr. Symp.*, 1987.

[43] D. B. Sullivan, D. W. Allan, D. A. Howe, F. L. Walls, Eds., *"Characterization of Clocks and Oscillators"*, NIST Technical Note 1337, March 1990.

[44] S. Bregni, "Chapter 5 - Characterization and Modelling of Clocks", in *Synchronization of Digital Telecommunications Networks*. Chichester, UK: John Wiley & Sons, 2002, pp. 203-281.

[45] ITU-T Rec. G.810 "Definitions and Terminology for Synchronisation Networks", Rec. G.811 "Timing Characteristics of Primary Reference Clocks", Rec. G.812 "Timing Requirements of Slave Clocks Suitable for Use as Node Clocks in Synchronization Networks", Rec. G.813 "Timing Characteristics of SDH Equipment Slave Clocks (SEC)", Geneva, 1996-2003.

[46] S. Bregni, L. Primerano, “The Modified Allan Variance as Time-Domain Analysis Tool for Estimating the Hurst Parameter of Long-Range Dependent Traffic“, *Proc. IEEE GLOBECOM 2004*, Dallas, TX, USA, 2004.

[47] S. Bregni, L. Jmoda, "Accurate Estimation of the Hurst Parameter of Long-Range Dependent Traffic Using Modified Allan and Hadamard Variances", to appear in *IEEE Trans. Communications*, 56.11 (2008): 1900-1906.. Extended version available: http://home.dei.polimi.it/bregni/public.htm.

[48] S. Bregni, W. Erangoli, “Fractional Noise in Experimental Measurements of IP Traffic in a Metropolitan Area Network“, *Proc. IEEE GLOBECOM 2005*, St. Louis, MO, USA, 2005.

[49] S. Bregni, R. Cioffi, M. Decina, "An Empirical Study on Time-Correlation of GSM Telephone Traffic", *IEEE Transactions on Wireless Communications*, Vol. 7, No. 9, Sept. 2008, pp. 3428-3435.

[50] F. Barcelò, J. Jordan, "Channel Holding Time Distribution in Public Telephony System (PAMR and PCS)", *IEEE Trans. Veh. Technol.*, vol. 49, no. 5, Sep. 2000, pp. 1615-1625.

[51] C. Jedrzycky, V.C.M. Leung, "Probability Distribution of Channel Holding Time in Cellular Telephony System", *Proc. IEEE Veh. Technol. Conf.*, Atlanta, GA, USA, May 1996.

[52] Y. Fang, I. Chlamtac, Y.B. Lin, "Channel Occupancy Times and Handoff Rate for Mobile Computing and PCS Networks", *IEEE Trans. Comput.*, vol. 47, no. 6, June 1998, pp. 679-692.

[53] Y. Fang, "Hyper-Erlang Distributions and Traffic Modelling in Wireless and Mobile Networks", *Proc. Wireless Communications and Networking Conference* (WCNC), New Orleans, LA, USA, Sep. 1999.

[54] Y. Fang, I. Chlamtac, "Teletraffic Analysis and Mobility Modelling of PCS Networks", *IEEE Trans. Commun.*, vol. 47, no. 7, July 1999, pp. 1062-1072.

[55] J.A. Barria, B.H. Soong, "A Coxian Model for Channel Holding Time Distribution for Teletraffic Mobility Modelling", *IEEE Commun. Lett.*, vol. 4, no. 12, Dec. 2000, pp. 402–404.

[56] F. Barcelò, S. Bueno, "Idle and Inter-Arrival Time Statistics in Public Access Mobile Radio (PAMR) System", *Proc. IEEE Globecom '97*, Phoenix, AZ, USA, Nov. 1997.

[57] A. Pattavina, A. Parini, "Modelling Voice Call Interarrival and Holding Time Distributions in Mobile Networks", *Proc. 19th International Teletraffic Congress (ITC)*, Bejing, Aug. 2005.

[58] W. Vervaat, "Properties of General Self-Similar Processes", *Bulletin of the Internat. Statistical Inst.*, vol. 52, pp. 199-216, 1987.

[59] P. Flandrin, "On the Spectrum of Fractional Brownian Motions", *IEEE Trans. Inform. Theory*, vol. 35, no. 1, pp. 197-199, Jan. 1989.

[60] J. Rutman, "Characterization of Phase and Frequency Instabilities in Precision Frequency Sources: Fifteen Years of Progress", *Proc. IEEE*, vol. 66, no. 9, pp. 1048-1075, Sept. 1978.

[61] J. A. Barnes, A. R. Chi, L. S. Cutler, D. J. Healey, D. B. Leeson, T. E. McGunigal, J. A. Mullen Jr., W. L. Smith, R. L. Sydnor, R. F. C. Vessot, G. M. R. Winkler, "Characterization of Frequency Stability", *IEEE Trans. Instrum. Meas.*, vol. 20, no. 2, pp. 105-120, May 1971.

[62] D. W. Allan, "Statistics of Atomic Frequency Standards", *Proc. IEEE*, vol. 54, no. 2, pp. 221-230, July 1966.

[63] W. C. Lindsey, C. M. Chie, “Theory of Oscillator Instability Based upon Structure Functions”, *Proc. IEEE*, vol. 64, no. 12, pp. 1652-1666, Dec. 1976.

[64] L. S. Cutler, C. L. Searle, "Some Aspects of the Theory and Measurement of Frequency Fluctuations in Frequency Standards", *Proc. IEEE*, vol. 54, no. 2, pp. 136-154, Feb. 1966.

[65] P. Lesage, C. Audoin, "Characterization and Measurement of Time and Frequency Stability", *Radio Science (American Geophysical Union)*, vol. 14, no. 4, pp. 521-539, 1979.

[66] F. L. Walls, D. W. Allan, "Measurement of Frequency Stability", *Proc. IEEE*, vol. 74, no. 1, pp. 162-168, Jan. 1986.

[67] J. Rutman, F. L. Walls, "Characterization of Frequency Stability in Precision Frequency Sources", *Proc. IEEE*, vol. 79, no. 6, pp. 952-960, June 1991.

[68] J.-F. Coeurjolly, "Simulation and Identification of the Fractional Brownian Motion: a Bibliographical and Comparative Study", *J. of Stat. Software*, vol. 5, no. 7, 2000, pp. 1-53.

[69] J.-F. Coeurjolly, "Estimating the Parameters of a Fractional Brownian Motion by Discrete Variations of Its Sample Paths", *Stat. Inference Stoch. Process*, Kluwer Academic Publishers, vol. 4, no. 2, 2001, pp. 199-227.

[70] P. Lesage, C. Audoin, "Characterization of Frequency Stability: Uncertainty Due to the Finite Number of Measurements", *IEEE Trans. Instrum. Meas.*, vol. 22, no. 2, pp. 157-161, June 1973. "Comments on '——' ", *IEEE Trans. Instrum. Meas.*, vol. 24, no. 1, p. 86, Mar. 1975. "Correction to '——' ", *IEEE Trans. Instrum. Meas.*, vol. 25, no. 3, p. 270, Sept. 1976.

[71] P. Lesage, C. Audoin, "Estimation of The Two-Sample Variance with Limited Number of Data", *Proc. 31st Annual Freq. Contr. Symp.,* 1977.

[72] K. Yoshimura, "Characterization of Frequency Stability: Uncertainty Due to the Autocorrelation Function of Frequency Fluctuations", *IEEE Trans. Instrum. Meas.*, vol. 27, no. 1, pp. 1-7, Mar. 1978.

[73] S. R. Stein, "Frequency and Time - Their Measurement and Characterization", in *Precision Frequency Control*, vol. 2, ch. 2, E. A. Gerber A. Ballato, Eds. New York: Academic Press, 1985, pp. 191-232.

[74] C. A. Greenhall, "Recipes for Degrees of Freedom of Frequency Stability Estimators", *IEEE Trans. Instrum. Meas.*, vol. 40, no. 6, pp. 994-999, Dec. 1991.

[75] C. A. Greenhall, W. J. Riley. "Uncertainty of Stability Variances Based on Finite Differences". Available: http://www.wriley.com.

[76] W. J. Riley, "Confidence Intervals and Bias Corrections for the Stable32 Variance Functions", Hamilton Technical Services, 2000. Available: http://www.wriley.com.

[77] S. Hur, et al. "Performance limits of pico-cell systems using radio-over-fiber techniques with an electro absorption modulator." *Optical and Quantum Electronics* 39.7 (2007): 561-569.

[78] H. Claussen. "Co-channel operation of macro-and femtocells in a hierarchical cell structure." *International Journal of Wireless Information Networks* 15.3-4 (2008): 137-147.

[79] S. Anand, A. Sridharan, and K. N. Sivarajan. "Performance analysis of channelized cellular systems with dynamic channel allocation." *Vehicular Technology, IEEE Transactions on* 52.4 (2003): 847-859.

[80] D. Hong, and S. S. Rappaport. "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures." *Vehicular Technology, IEEE Transactions on* 35.3 (1986): 77-92.

[81] T.S. Kim, et al. "Mobility modelling and traffic analysis in three-dimensional indoor environments." *Vehicular Technology, IEEE Transactions on* 47.2 (1998): 546-557.

[82] Y. C. Kim, et al. "Dynamic channel reservation based on mobility in wireless ATM networks." *Communications Magazine, IEEE* 37.11 (1999): 47-51.

[83] P. A. Dintchev, B. Perez-Quiles, and E. Bonek. "An improved mobility model for 2G and 3G cellular systems." *3G Mobile Communication Technologies, 2004. 3G 2004. Fifth IEE International Conference on*. IET, 2004.

[84] S. Bhattacharya, H. M. Gupta, and S. Kar. "Traffic model and performance analysis of cellular mobile systems for general distributed handoff traffic and dynamic channel allocation." *Vehicular Technology, IEEE Transactions on* 57.6 (2008): 3629-3640.

[85] W.A. Massey, and W. Whitt, “Networks of finite server queues with nonstationary Poisson input”, Queueing Syst. Vol. 13, pp. 183-250.

[86] G. Foschini, B. Gopinath, and Z. Miljanic. "Channel cost of mobility." *Vehicular Technology, IEEE Transactions on* 42.4 (1993): 414-424.

[87] M. Rajaratnam, and F. Takawira. "Hand-off traffic modelling in cellular networks." *Global Telecommunications Conference, 1997. GLOBECOM'97., IEEE*. Vol. 1. IEEE, 1997.

[88] S. Bhattacharya, H. M. Gupta, and S. Kar. "Performance Modelling of Cellular Mobile Systems: A Review of Recent Advances." *IETE Technical Review* 27.1 (2010): 15.

[89] M. Z. Chowdhury, and Y. M. Jang. "Handover management in high-dense femtocellular networks." *EURASIP Journal on Wireless Communications and Networking* 2013.1 (2013): 1-21.

[90] E. Chlebus, and W. Ludwin, “Is Handoff Traffic Really Poisson?”, IEEE International Conference on Universal Personal Communications (ICUPC), 1995.

[91] S. Choi, and K. Sohraby, “Analysis of a mobile Cellular System with Hand-off Priority and Hysteresis Control”, IEEE International Conference on Computer Communications (INFOCOM), 2000.

[92] E. Ekici, and C. Ersoy, “Multi-tier Cellular Network Dimensioning” , Wireless Networks, vol. 7, no. 4, 2001.

[93] I. Katzela, and M. Naghshineh, “Channel Assignment Schemes for Cellular Mobile Telecommunication Systems: A Comprehensive Survey”, IEEE Personal Communications, vol. 3, no. 3, 1996.

[94] L. Kaufman, B. Gopinath, and E. Wunderlich, “Analysis of Packet Network Congestion Control Using Sparse Matrix Algorithms”, IEEE Transactions on Communications, vol. 29, no. 4, 1981.

[95] S. Oh, and D. Tcha, “Prioritized Channel Assignment in a Cellular Radio Network”, IEEE Transactions on Communications, vol.40. no. 7, 1992.

[96] M. Oliver, and J. Borras, “Performance Evaluation of Variable Reservation Policies for Hand-off Prioritization in Mobile Networks”, IEEE International Conference in Computer Communications (INFOCOM), 1999.

[97] L. Ortigoza-Guerrero, and A. Aghvami, “On the Optimum Spectrum Partitioning in a Microcell/Macrocell Cellular Layout with Overflow”, IEEE Global Communications Conference (GLOBECOM), 1997.

[98] M. Sidi, and D. Starobinski, “New Call Blocking Versus Handoff in Cellular Networks”, IEEE International Conference on Computer Communications (INFOCOM), 1996.

[99] Cisco, Traffic Analysis for Voice over IP, White paper, 2001., http://www.cisco.com/en/US/docs/ios/solutions_docs/voip_solutions/TA_ISD.pdf

[100] S. Aalto and P. Lassila, “Impact of size-based scheduling on flow level performance in wireless downlink data channels”, Proceedings of the 20th International Teletraffic Conference, 2007.

[101] The CAIDA UCSD Anonymized Internet Traces 2013. <2013/05/29 and 2013/06/20>. http://www.caida.org/data/passive/passive_2013_dataset.xml

[102] The CAIDA UCSD OC48 Internet Traces Dataset. <2002/08/14> http://www.caida.org/data/passive/passive_oc48_dataset.xml

[103] Traffic flow on broadband Access Networks, Terry D. Shaw, PhD-Cable Labs-2012

[104] V. Flood, “Bandwidth a barrier for (Dis) connected TV in South Korea”, February 2012. http://www.videoadnews.com/2012/02/13/bandwidth-a-barrier-for-disconnected-tv

[105] Cisco Visual Network Index Global Mobile Data Traffic Forecast Update, 2012-2017,February 2013

http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf

[106] J.Ankey, “Google: Mobile devices generate 10% of all YouTube video downloads”, September 2011\ www.fiercemobileit.com

[107] 3GPP TS 22.278 “Service requirements for the Evolved Packet System (EPS)”, V12.2.0 (2013-03).

[108] “Integrated Femto-Wi-Fi (IFM) Networks”, Small Cell Forum, February, 2012.

[109] M. Taqqu, W. Willinger, and R. Sherman, ”Proof of a fundamental result in self-similar traffic modelling”, Computer Communications Review 26:5-23, 1997.

[110] S. Stoev, G. Michailidis, and J. Vaughan, ”On Global Modelling of Network Traffic”, INFOCOM 2010, The 29th Conference on Computer Communications, San Diego, California, March 2010.

[111] White paper ”Analysis of Traffic Offload: Wi-Fi to Rescue”, WirelessE2E LLC, September, 2010.

[112] 3GPP TR 23.829, “Local IP Access and Selected IP Traffic Offload (LIPA-SIPTO)”, V10.0.1 (2011-10).

[113] 3GPP TR 23.861 “Network based IP flow mobility ”, V1.7.0 (2012-11).

[114] K. Lee, Y. Yi, J. Lee, I. Rhee, S. Chong, ”Mobile Data Offloading: How Much Can Wi-Fi Deliver?”, in proc. of ACM CoNEXT 2010, USA, December, 2010.

[115] Bundesnetzagentur: „Wettbewerbsintensität im Mobilfunk nimmt weiter zu“, press report, 2012. [online] http://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Allgemeines/Presse/Pressemitteilungen/2012/120824_WettbewerbMobilfunkpdf.pdf

8 ANNEX I. - Analysis of real traffic data in fixed and mobile networks of Telefónica

8.1 Introduction

The analysis of real traffic data allows the operator to know the needs of these customers, their mode of use, to improve the quality of service (QoS) and a Quality of Experience (QoE).

We construct a network, if we know the traffic to flow into each segment of the network, mainly when we talk about a converged network (Fixed and Mobile Network).In the converged Network we need to know the mode of use of Fixed and Mobile customers, their traffic profiles, the convergence of the mode of use in terms of applications. All these parameters allow us to dimensioning a converged network.

Several papers already presented traffic measurements analysis about Internet applications. This document is based on an analysis of real internet traffic measured in the field on Telephonica Fixed and Mobile networks in March 2013.

8.2 Traffic probes

In Telefónica network, traffic probes measure a traffic aggregated (average bit rate bit/s for 3 hours) every 3 hours. Figure 103 and Figure 104 illustrate the location of traffic probes in the Telefónica networks. In fixed network (Figure 103), the traffic probes are installed between aggregation and core network and in mobile network (Figure 104), they are installed between SGSN and GGSN. The probes measure traffic of approximately 970.000 and 1500.000 customers in Fixed and mobile network respectively. The probe used for the measure is the passive probe using a port mirroring (not injected a test frames) and identify more than 700 types of protocol and several applications.

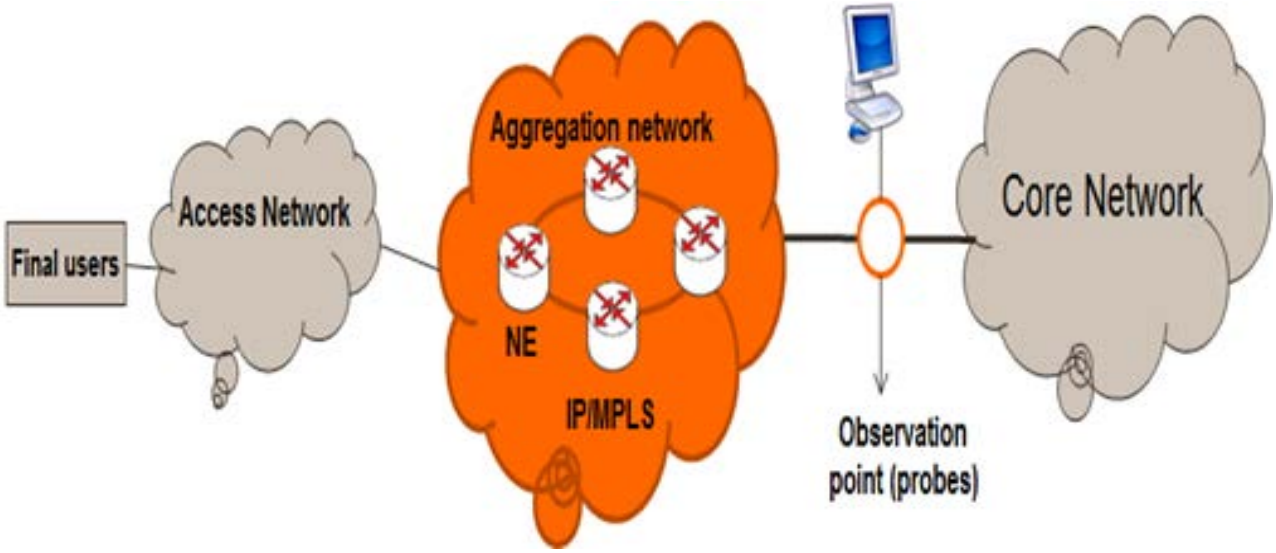


Figure 103. Probes location in the fixed network

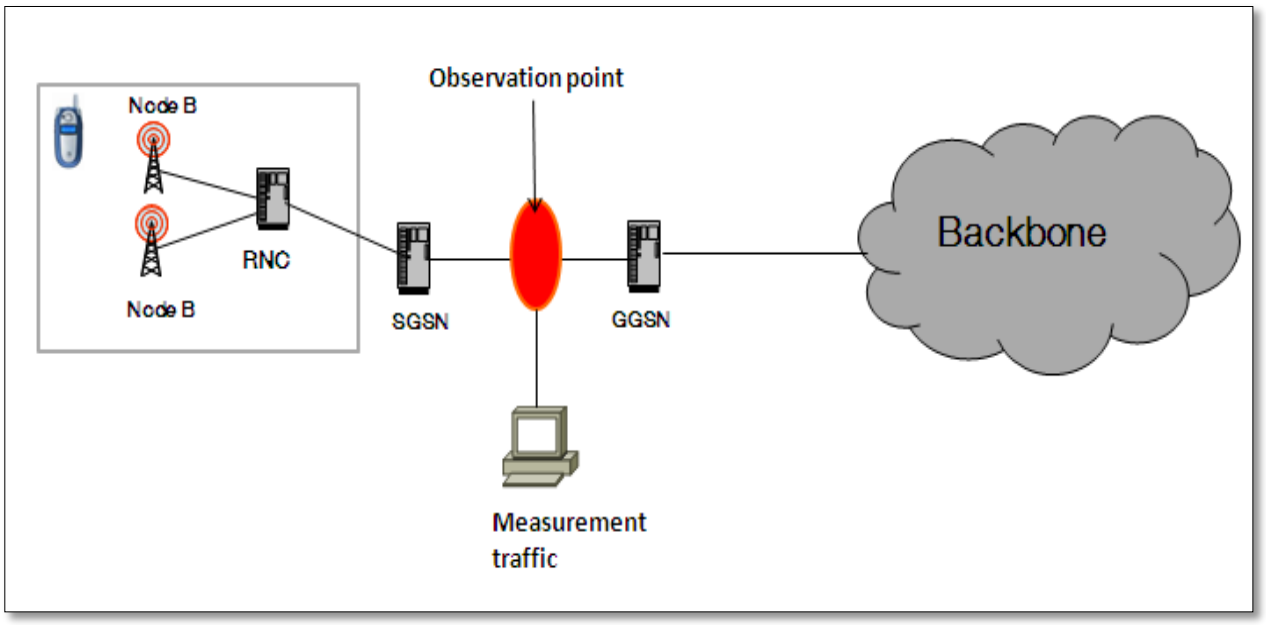


Figure 104. Probes location in the mobile network

8.2.1 Major tendencies of downstream and upstream Internet traffic

In this sub-section we will focus on Internet applications generating the most traffic in upstream and downstream and in the case of fixed and mobile Networks.

8.2.2 Upstream Internet Traffic for fixed and mobile customers

Figure 105 (resp. Figure 106) illustrates the total volume evolution over the time for upstream Internet traffic in the case of fixed (resp. mobile) network. Moreover, these figures depict volume for some Internet applications.

The total volume generated by all fixed network customers is 4374.2Tbytes. P2P is the application generating 30% of the total volume and the other applications (such as Web browsing, mailing, etc.) generate the rest of the volume. During the same observation period (March 2013), mobile customers generated a total upstream Internet traffic of 77Tbytes and Web represents the application generating the main proportion of volume of traffic (about 47%).

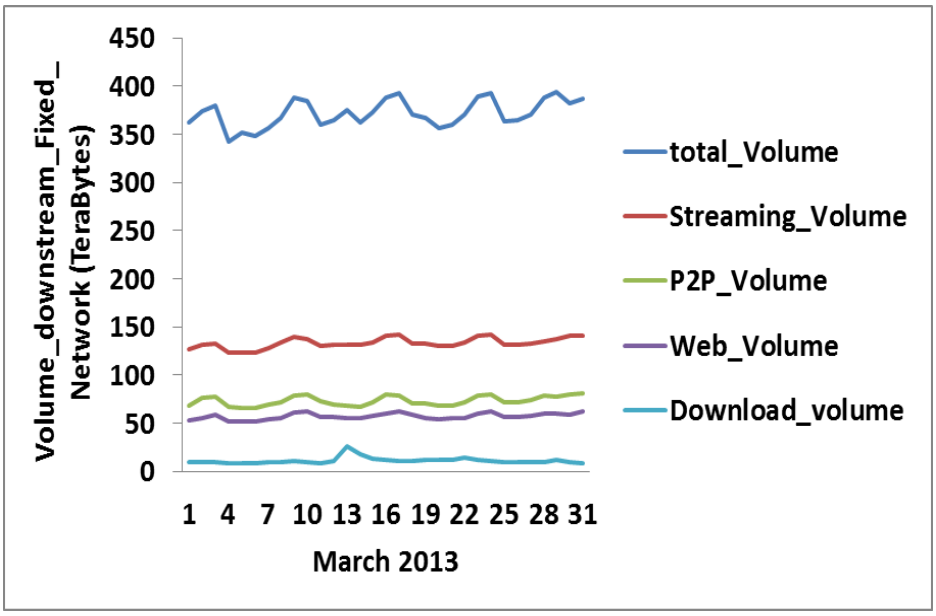


Figure 105. Upstream Internet traffic for fixed network customers, measurements performed in March 2013

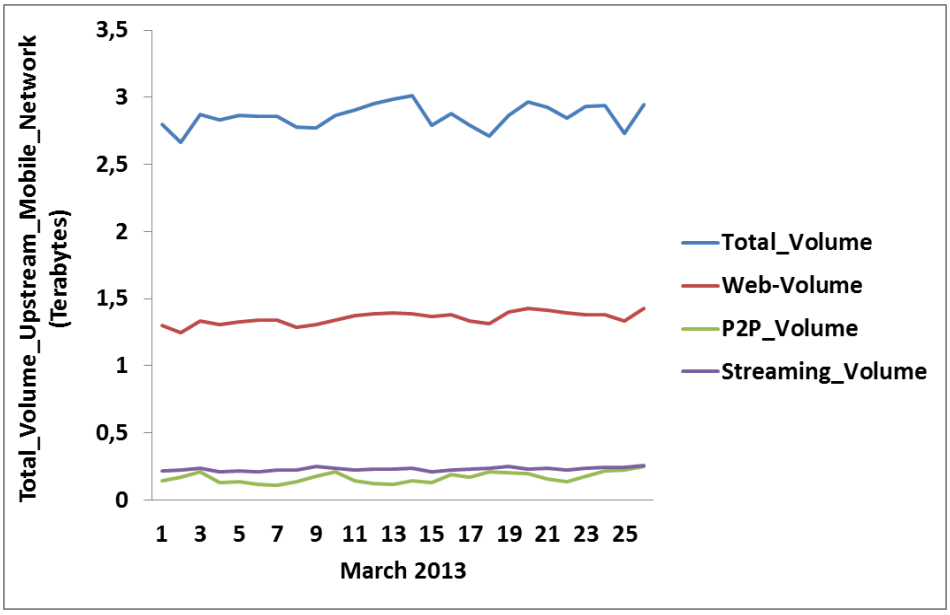


Figure 106. Upstream Internet traffic for mobile network customers, measurements performed in March 2013

8.2.3 Downstream Internet traffic for fixed and mobile customers

Figure 107 (resp. Figure 108) illustrates the total volume evolution over the time for downstream Internet traffic in the case of fixed (resp. mobile) network. Moreover, these figures depict bitrate for some Internet applications.

Figure 107 shows that the total downstream Internet traffic generated by total customers in fixed network represents 11528 Terabytes. Video streaming is the application generating alone 36% of the total volume. P2P application generates 20% of the global traffic. Web browsing generates 15% of the traffic and the rest of the application generates a very low traffic.

Figure 108 indicates that the downstream Internet traffic generated by all customers in mobile network in represents 634 Terabytes. Web represents 47% of the total volume, followed by streaming (33%), P2P (12%) and the rest of the application generates a very low traffic.

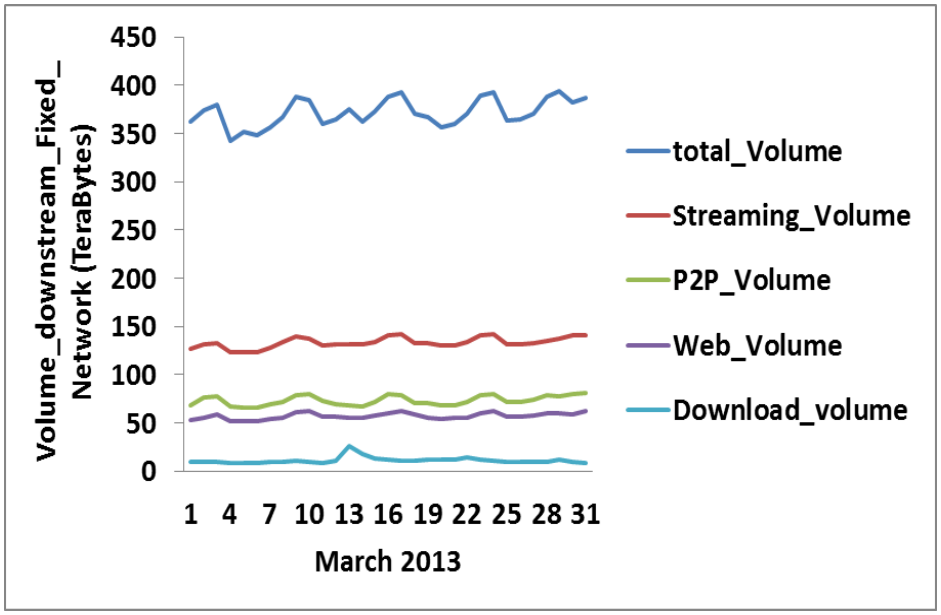


Figure 107. Downstream Internet traffic for fixed network customers, measurements performed in March 2013

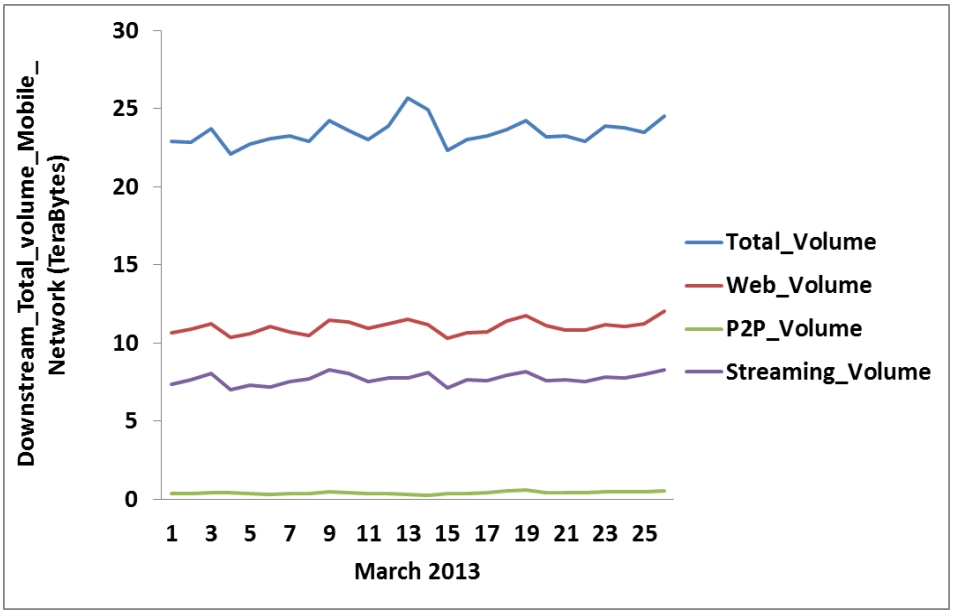


Figure 108. Downstream Internet traffic for mobile network customers, measurements performed in March 2013

8.3 Upstream and downstream Internet traffic profiles in fixed and mobile networks

This sub-section explains daily downstream and upstream Internet profiles for fixed and mobile customers thanks to an analysis during 24 hours over a week. These observations allowed us in identifying the busy hours period. Moreover, for a network operator, it is important to understand how the traffic evolves during the day, in order to perform maintenance, firmware updates during periods with low traffic activity.

8.3.1 Upstream profile for fixed and mobile networks

Figure 109 (resp. Figure 110) illustrates the daily upstream Internet traffic profile for fixed (resp. mobile) customers over a period between the 23rd of March 2013 and the 29th of March 2013.

The traffic generated by all customers in fixed and mobile network achieve a maximum during the time period [7pm-9pm].This time slot represents a busy hours period. We observe a small difference in the customers' behavior between the days of the week and the week-end (generally the Sunday). Indeed, the customer's traffic increases in the afternoon (from 12:00am), while during the day of the week, it increases after 6pm. Figure 109 and Figure 110 show that the daily upstream Internet profile is similar in fixed and mobile networks.

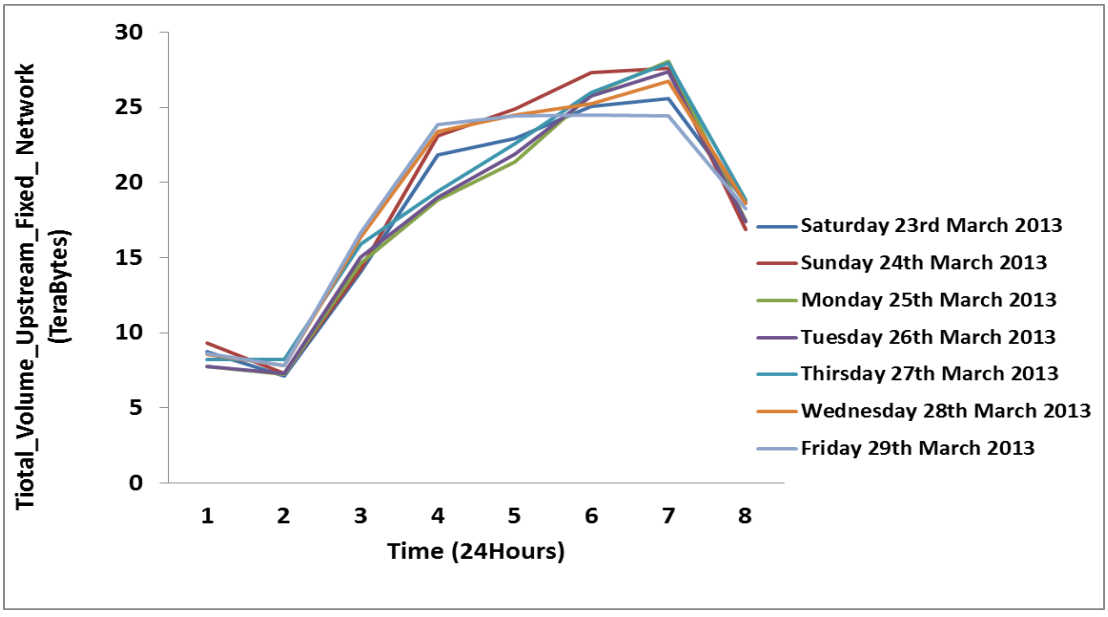


Figure 109. Daily upstream Internet profile for customers in fixed network, observation performed over a week of March 2013.

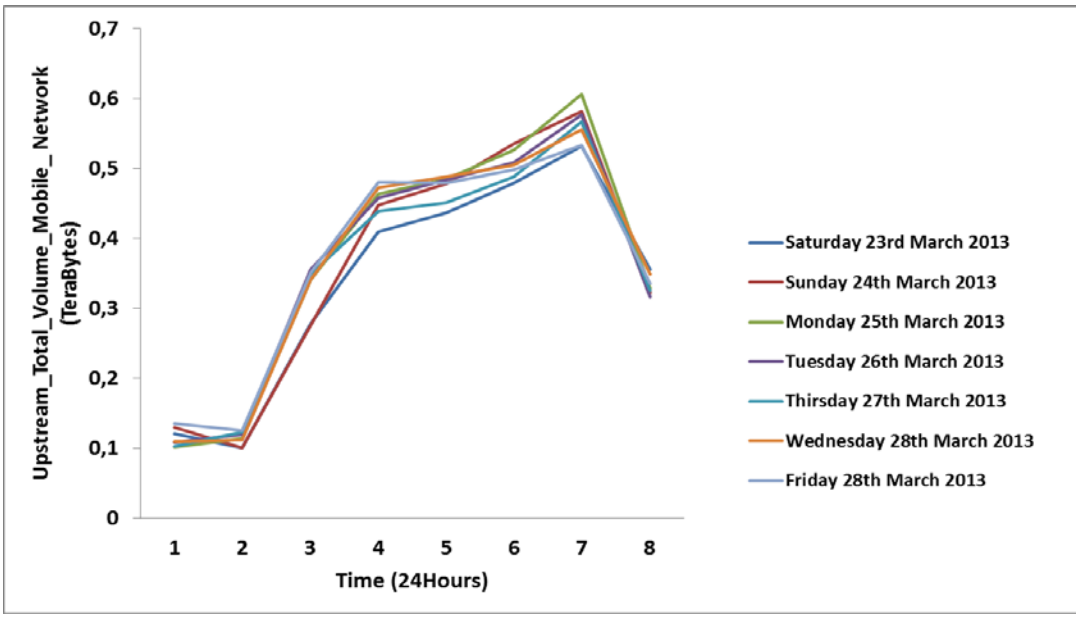


Figure 110. Daily upstream Internet profile for customers in mobile network, observation performed over a week of March 2013

8.3.2 Downstream profile in fixed and mobile networks

Figure 111 (resp. Figure 112) describes the daily downstream Internet traffic profile in Fixed (resp. Mobile) customers over a period between the 23re of March 2013 and the 29th of March 2013. These figures demonstrate that there is a similarity in terms of use of the network between the days of the week and the week-end. The downstream volume reaches a maximum in the time slot [7pm – 9pm] for fixed and Mobile customers.

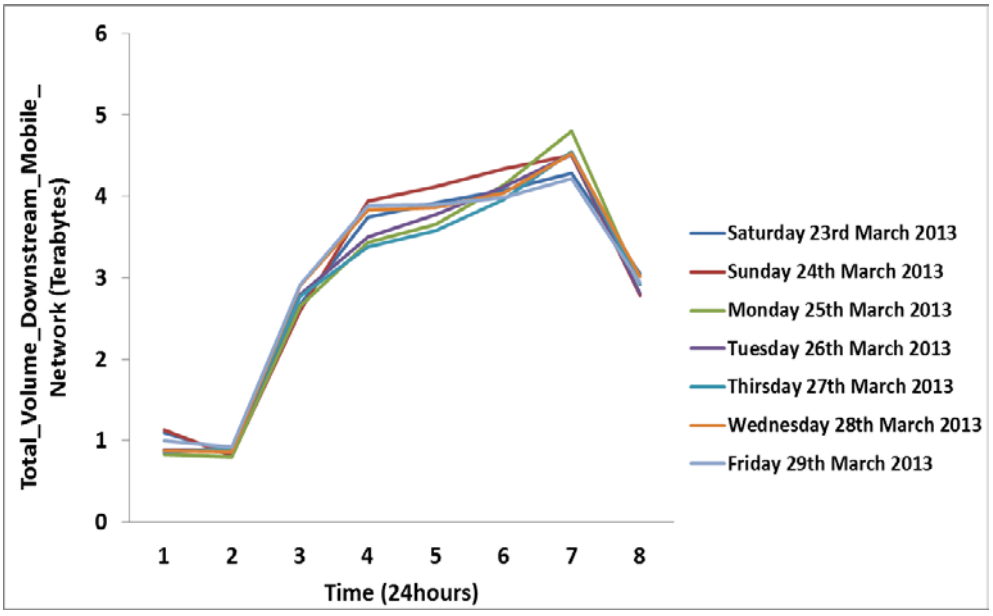


Figure 111. Daily downstream Internet profile for customers in fixed network, observation performed over a week of March 2013

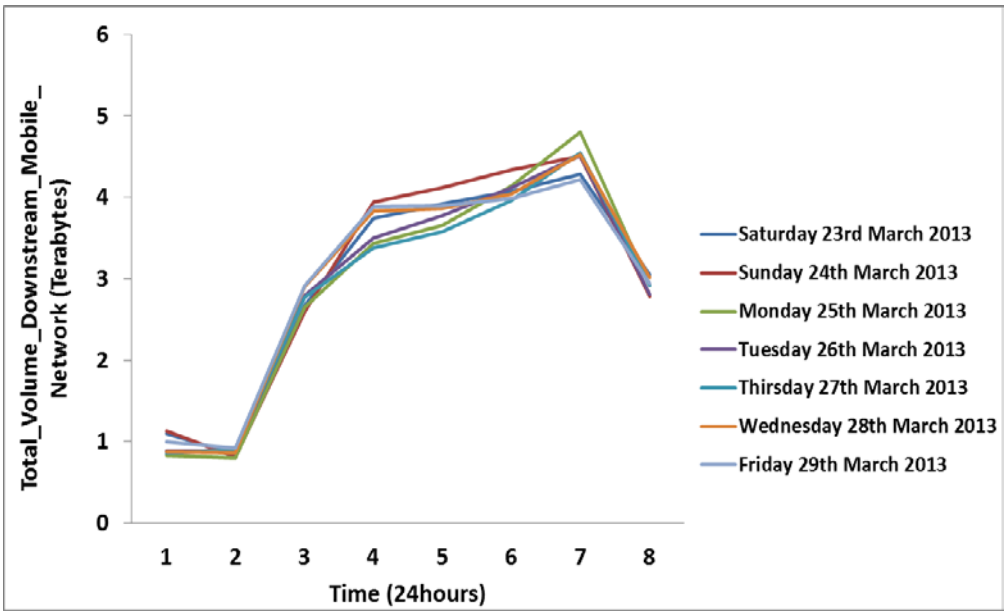


Figure 112. Daily Downstream Internet profile for customers in mobile etwork, observation performed over a week of March 2013

8.4 Fixed versus Mobile customers

The table 1 summarizes the average Internet volume generated by Fixed and Mobile customers in upstream and downstream in March 2013 and gives the ratio between downstream and upstream Internet volume.

	Fixed Network (Gbytes)	Mobile Network (Gbytes)
Average upstream volume	4.5	0.05
Average upstream volume P2P	1.32	3E-3
Average upstream volume Web	0.81	0.024
Average upstream volume Streaming	1.01	4.15E-3
Average Downstream volume	12	0,42
Average Downstream volume Streaming	4.26	0.138
Average Downstream volume Web	1.83	0.197
Average Downstream volume P2P	2.34	7E-3
Ratio DS/US	2.66	8.4

Table 13. Ratio for Mobile and Fixed customers in March 2013

We observed that the mode of use of fixed and mobile customers becomes convergent in terms of applications (web, streaming).

The ratio between downstream and upstream is important in the case of mobile network, the fact that mobile clients do not exploit the upstream direction, but in fixed network the ratio is low because the fixed customers use the uplink direction for applications P2P.

The average volume in Fixed Network is 28 times the average volume in mobile network.

We can imagine very important volume in mobile network (LTE) 20 or 30 times the volume in 3G and we can switch the fixed traffic on mobile network.

8.5 Conclusion

In upstream, P2P based applications generate of the traffic (in terms of volume) of fixed customers. It generates 30% of traffic in the case of fixed network. In mobile network a Web generate a most traffic in Upstream (47%), a P2P is not used in mobile network, the users prefer shared their contained P2P in fixed network with laptop.

In downstream, streaming, Web browsing and P2P applications represent the applications that generate more than 70% of total volume in fixed and mobile network.

Whatever the day (weekend, weekday), the type of customer (Fixed / Mobile) and the direction of traffic (upstream or downstream), customer profile is similar (i.e. the volume is very low in the morning and starts to rise in the afternoon and usually reached its maximum between 7 pm and 9 pm).

We can observe the convergent a mode of use of fixed and mobile customers in terms of applications like web-based streaming.

9 ANNEX II. - Comparison of real traffic data of Orange France and Telefónica

9.1 Upstream Internet traffic for xDSL and FTTH customers

Whatever the technology (ADSL, VDSL, FTTH), P2P remains the application that generates most of the upstream traffic in Fixed Network; it represents a significant proportion (in terms of volume) of data traffic from FTTH and xDSL customers.

Main usage in upstream of fixed Network of Orange and Telefónica customers is similar in terms of applications; generally customers of both operators use P2P and Web applications.

9.2 Upstream Internet traffic for Mobile customers

For Mobile network, the way how Telefónica and Orange customers use mobile Internet service in upstream is similar in terms of applications, the Web represents the application generating most of the traffic while P2P is little used in mobile network. Customers prefer to use a computer to share files with P2P application because computers offer an important storage capacity compared to a mobile device. Moreover, generally mobile users can't install P2P applications on IOS and Android mobile.

9.3 Downstream Internet traffic for xDSL and FTTH customers

In downstream fixed network, we will find a common mode of use between Orange (FTTH and ADSL) and Telefónica (xDSL) customers, like a dominant use of Streaming Video which represents the application that generates most of the traffic. Web browsing represents less than 20% of the total volume generated by Orange and Telefónica customers. We can also find differences in terms of use: Telefónica customers use P2P applications which ranks as the second application that generates data traffic (22%), while Orange customers use Download applications (20% of total volume) and P2P represents a low volume(<8%).

9.4 Downstream Internet traffic for Mobile customers

Orange and Telefónica mobile customers have a very similar mode of use in downstream. Video streaming and Web represents more than 70% of the total volume generated by Orange and Telefónica customers.

9.5 Convergent of applications used by Fixed and Mobile customers

9.5.1 Upstream Fixed and Mobile customers

We find a little similarity in terms of use of fixed and Mobile customers in upstream direction. This little similarity mainly stands in the use of Web application with a very different proportion in terms of contribution of total volume. In Fixed Network, Web browsing traffic generates less than 20% of total volume, while in mobile network Web application represents more than 40% of total volume generated by customers.

P2P-based applications represent the majority of volume generated by all fixed-network customers' while in mobile network, the customers use a little the application P2P (P2P represents a low volume in mobile network). Generally customers use their fixed access to share files P2P and exactly they use their computer as other devices for reasons of storage capacity.

In upstream direction, the mode of use of Fixed and Mobile customers is different for both operators (Telefónica and Orange) with a little similarity which is mainly the use of Web application.

9.5.2 Downstream Fixed and Mobile customers

In downstream direction, there is a strong similarity in terms of use of fixed and mobile customers. Streaming video remains the application generating most traffic and also contributing with a similar volume generated by all customers in Fixed and mobile network for both operators. We observe also that Web browsing is the second common application used by fixed and Mobile customers for both operators and represents less than 20% of total volume in fixed network, while in mobile network a Web represents more than 30% of total volume.

As we observe other common applications used by fixed and mobile customers of the same operator like Download (resp.P2P) which is used by fixed and mobile Orange customers (resp. Telefónica fixed and mobile customers) with a little different proportions of the total volume.

9.6 Upstream profile for xDSL and FTTH customers

The results shows that a common "busy hour" for Orange customers and Telefónica customers can be defined in the range of [7pm-9pm].

Traffic profiles of ADSL Orange and xDSL Telefónica users are flatter than the profile of FTTH customers, because of a higher number of customers in the case xDSL Telefónica (resp. ADSL Orange) of 970 000 customers (resp.5180). These figures explain the flattest profile of xDSL Telefónica customers.

9.7 Upstream profile for mobile customers

A typical profile from a Telefónica customer is different than a profile from an Orange customer during week days and also during the week end.

During a week day, we observe that busy hours are different: profile of Orange customers represents two times slots of busy hours and profiles of Telefónica customers presents only one time slot of busy hours. The results indicate that we cannot find a common time period of busy hours between the two sets of customers from the two operators.

During week-end periods, we have one time slot of busy hours for Orange and Telefónica customers but the time period is different.

Profiles from Orange mobile customers are flatter than profiles from Telefónica mobile customers because of a higher number of Orange customers than Telefónica mobile ones.

9.8 Mobile versus fixed upstream traffic

A profile of fixed customers is flatter than the profile of mobile customers, because application usage of Mobile and fixed customers is different. In Fixed network customers use mainly P2P applications and P2P file sharing often requires a lot of time for full file sharing results. On the other hand, in mobile network, customers use web applications that represent a majority of traffic in upstream direction. These kinds of applications do not need to keep long sessions alive. That is a reason why profile of fixed customers are flatter than profiles in mobile network, even though the number of mobile customers is higher than the fixed customers for this category of measured of traffic.

The busy hours of Telefónica are similar for fixed and mobile customers. In the case of Orange France network, busy hour are very different: in fixed network we observe one time slot of busy hours which represent a period after work and when customers are at home. A second time slot of busy hours of mobile customers can be observed between 5pm and 7pm (after work while commuting for example) and the period of busy hour for fixed customers typically spans between 7 pm and 9 pm. This difference in busy hours period between fixed and mobile customers results in when the customers get home prefer use their fixed access instead their mobile access.

9.9 Downstream profile for xDSL and FTTH customers

The profile of xDSL Telefónica customers is flattest than the profile of FTTH and ADSL Orange customers because a number of Telefónica customers is greater than the Orange customers in this measure.

The number of ADSL customers is higher than the number of FTTH customers, this is a reason why a profile of ADSL customers id flatter than the FTTH.

We can observe a common period of busy hours between profiles of Orange and Telefónica customers, this time is between 7 pm and 9 pm.

9.10 Downstream Internet traffic profile for Mobile customers

A profile of Orange and Telefónica customers represent a common period of busy hours in period of [7pm-9pm]. We observe a difference in profile of Orange customers between a week-end and a week day, this difference is mainly due to lunch break between noon and 2pm on weekdays and it doesn't exist in weekends.

For Telefónica customers we don't observe a difference between a week-end and a week-day for profile customers. A profile of Orange mobile customers is flatter than a profile of Telefónica customers because a number of Orange mobile customers extracted in this measure is higher than the number of Telefónica customers.

9.11 Mobile versus fixed downstream traffic profiles for both operators

We don't observe any difference between fixed and mobile customer's profile because the customers use the same applications (mainly streaming video). A profile of Telefónica customers is similar in fixed and mobile networks in terms of use and a period of busy hours. On the other hand, in Orange France Network, we observe a difference between fixed and mobile customers profile. There is only one period of busy hours in fixed network because the customers return home and two periods of busy hours in mobile network

(period of lunch break and after work in the evening). The second time slot of busy hours of Orange customers in mobile network is similar to a period of busy hours in fixed network.

10 Glossary

Acronym / Abbreviations	Brief description
2G	2nd Generation (mobile service)
3G	3rd Generation (mobile service)
3GPP	3rd Generation Partnership Project
ADSL	Asymmetric Digital Subscriber Line
ALM	Application-Level Multicast
AVAR	Allan Variance
AVSP	Adaptive Video Streaming Protocol
BHCA	Busy-Hour Call Attempts
BMGF	Binomial Moment Generating Function
BRAS	Broadband Remote Access Server
CAGR	Compound Annual Growth Rate
CDN	Content Delivery Network
COMBO	COnvergence of fixed and Mobile BrOadband
DAG	Data Acquisition and Generation
DHCP	Dynamic Host Configuration Protocol
DSLAM	Digital Subscriber Line Access Multiplexer
EDGE	Enhanced Data Rates for GSM Evolution
EIR	Equipment Identity Register
eNodeB	Evolved Node B (base station)
EPC	Evolved Packet Core
E-UTRAN	Evolved UMTS Terrestrial Radio Access network

Acronym / Abbreviations	Brief description
FBM	Fractional Brownian Motion
FMC	Fixed Mobile Convergence (Converged)
FTTH	Fiber to the Home
GGSN	Gateway GPRS Support Node
HLR	Home Location Register
HSPA	High Speed Packet Access
HSS	Home Subscriber Server
HTTP	Hypertext Transfer Protocol
i.i.d.	Independent and Identically Distributed
IP	Internet Protocol
ISP	Internet Service Provider
ITU-T	International Telecommunications Union- Telecommunication Standardisation Sector
LTE	Long Term Evolution (3GPP standard)
LRD	Long-Range Dependence
M2M	Machine-to-Machine
MAVAR	Modified Allan Variance
MSC	Mobile Switching Center
MME	Mobile Management Entity
MNO	Mobile Network Operator
MPS	Messages Per Second
MSS	Mobile Switching Centre Server
NG-PoP	Next Generation Point of Presence

Acronym / Abbreviations	Brief description
NUT	Network Under Test
OCS	Online Charging System
OLT	Optical Line Termination
P2P	Peer-to-Peer
PALM	Poisson-arrival-location model
PAMR	Public Access Mobile Radio
PCRF	Policy and Charging Rules Function
PDN	Packet Data Network
PDN-GW	Packet Data Network Gateway
PDP	Packet Data Protocol
PON	Passive Optical Line
PoP	Point of Presence
PGF	Probability Generating Function
P-GW	Packet Data Network Gateway
PSD	Power Spectral Density
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RAN	Radio Access Network
RNC	Radio Network Controller
SCTP	Stream Control Transmission Protocol
SIM	Subscriber Identity Module
SGSN	Serving GPRS Support Node

Acronym / Abbreviations	Brief description
S-GW	Serving Gateway
SS	Self-Similar
SSSI	Self-Similar Processes with Stationary Increments
TEID	Tunnel Endpoint Identifier
TFT	Traffic Flow Template
TCP	Transmission Control Protocol
UE	User Equipment
UDP	User Datagram Protocol
UMTS	Universal Mobile Telecommunications System
WLAN	Wireless Local Area Network
VoD	Video on Demand
VoIP	Voice over Internet Protocol
VoLTE	Voice over LTE
VLR	Visitor Location Register
WAN	Wide Area Network
WCDMA	Wideband Code Division Multiple Access
Wi-Fi	Wireless Local Area Network – Commercial name
WiMAX	Worldwide Interoperability for Microwave Access

11 List of Tables

12 List of Figures

13 List of authors

Full Name – E-mail	Company – Country Code
C. Behrens (BehrensC@telekom.de) E. Bogenfeld (eckard.bogenfeld@telekom.de)	DTAG – DE
J. De Biasio (joseph.debiasio@orange.com) M. Feknous (moufida.feknous@orange.com)	FT – FR
S. Höst (Stefan.Host@eit.lth.se)	ULUND – SE
A. Krendzel (andrey.krendzel@cttc.es) P. Dini (paolo.dini@cttc.es)	CTTC – ES
S. Bregni (bregni@elet.polimi.it) A. Pattavina (pattavina@elet.polimi.it)	POLIMI – IT
P. Varga (pvarga@tmit.bme.hu) P. Olaszi (polaszi@aitia.ai)	AITIA – HU
O. Eker (onur.eker@turktelekom.com.tr)	TT – TR

13.1 List of reviewers

Full Name – E-mail	Company – Country Code
Jose Alfonso Torrijos Gijón – jgijon@tid.es	TID – ES
Klaus Grobe – kgrobe@advaoptical.com	ADVA-DE
Dirk Breuer – d.breuer@telekom.de	DTAG-DE

13.2 Approval

Approval	Full Name – E-mail	Company – Country Code	Date
Task Leader	Pal Varga – pvarga@tmit.bme.hu	AITIA - HU	
WP Leader	Jose Alfonso Torrijos Gijón – jgijon@tid.es	TID – ES	
Project Coordinator	Jean-Charles Point – pointjc@jcp-consult.com	JCP - FR	
Other (PMC, SC, etc)			

13.3 Document History

Edition	Date	Modifications / Comments	Author
0.1	25/02/13	Initial revision with TOC	Pal Varga
0.2	16/04/13	Second revision with TOC and methodology suggestion	Pal Varga
	16/04/13	TT amendments	Onur Eker
	17/04/13	FT additional views on methodology	Joseph De Biasio
0.3	28/05/13	Editor instructions on section contents	Pal Varga
0.4	07/06/13	Amendments on methodology; partners contributions added	Pal Varga
0.5	18/08/13	Added sections on signalling changes by CTTC	Pal Varga
0.6	14/10/13	Partners contributions added to Chapters 2, 3, 4	Pal Varga
	15/10/13	CTTC amendments added, AITIA results added	Pal Varga
	31/10/13	TT contribution on Traffic trends and on CDN added	Onur Eker
	31/10/13	CTTC contributions corrected after review	Paolo Dini
0.7	31/10/13	Amended with internal review feedback and contributions from POLIMI	Pal Varga
	1/11/13	TT contributions added	Onur Eker
	1/11/13	TT contributions amended	Onur Eker
	1/11/13	TT contributions further amended	Onur Eker
	1/11/13	Included section 5.2	Carsten Behrends
	8/11/13	Included FT analysis results, analysis methodology	Joseph De Biasio
	12/11/13	Amended FT analysis results	Joseph De Biasio
0.8	15/11/13	Included partner review feedback and POLIMI results	Pal Varga
	15/11/13	Added further analysis results	Joseph De Biasio
	22/11/13	Amended results	Joseph De Biasio
	27/11/13	Content tuned after internal review	Joseph De Biasio
	29/11/13	Added analysis results for TID	Joseph De Biasio
0.9	01/12/13	Included partner content after internal review feedback	Pal Varga
	03/12/13	Editorial changes	Pal Varga
	03/12/13	Additional analysis results	Moufida Feknous
	04/12/13	Editorial changes	Pal Varga
	06/12/13	Editorial changes	Peter Olaszi
0.95	09/12/13	Version offered for review	Pal Varga

Edition	Date	Modifications / Comments	Author
0.96	12/12/13	After-review version, Section Editors feedback included, not yet fully edited	Pal Varga
0.97	15/12/12	Ordering references, tables, figures; implementing reviewers comments	Pal Varga
0.98	15/12/13	Implementing reviewer's comments	Peter Olaszi
0.99	16/12/13	Glossary finalized	Peter Olaszi
0.991	16/12/13	Release version	Pal Varga
0.992	16/12/13	Version after reviewer's comment	Pal Varga
0.993	17/12/13	Added Annex	Peter Olaszi
1.00	16/12/13	Release version	Pal Varga

13.4 Distribution List

Full Name or Group	Company	Date
PMC	Public deliverable (will be made available through COMBO website)	
SC		
Other		

14 Further information

Grant Agreement number: 317762

Project acronym: COMBO

Project title: COnvergence of fixed and Mobile BrOadband access/aggregation networks

Funding Scheme: Collaborative Project – Integrated Project

Date of latest version of the Deliverable: 17-12-2013

Delivery Date: **Month 11**

Leader of the Deliverable: Dr. Pal Varga, AITIA

File Name: COMBO_D2.3_WP2_17December2013_AITIA_V1.0.docx

Version: V1.0

Authorisation code: PU = Public

Project coordinator name, title and organisation: Jean-Charles Point, JCP-Consult

Tel: + 33 2 23 27 12 46

E-mail: pointjc@jcp-consult.com

Project website address: www.ict-combo.eu

- - - End of Document - - -