MULTI - SERVICE LEVEL AGREEMENT IN WIRED AND WIRELESS ACCESS NETWORKS

Arianna Rufini

Dottorato di Ricerca in Ingegneria dell'Informazione e della Comunicazione
Doctor of Philosophy in Information and Communication Engineering

Sapienza Università di Roma
Sapienza University of Rome

2014
XXV Ciclo
25th Course

Autore
Author……
Arianna Rufini

Dipartimento di Ingegneria dell'Informazione, Elettronica e Telecomunicazioni
Department of Information Engineering, Electronics and Telecommunications

Visto
Certified by……
Prof. Aldo Roveri
Tutore di Dottorato
Thesis Supervisor

Table of Contents

Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor Prof. Aldo Roveri, who has supported and encouraged me throughout my thesis with his immense knowledge, experience and helpfulness.

I would like to express my special appreciation and thanks to my FUB tutor Francesco Matera for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm.

I would like to acknowledge the academic and technical support of the whole Infocom Department in particular I also thank the Prof. Marco Listanti for his support and his advices since the start of my postgraduate .

I would also like to thank the Fondazione Ugo Bordoni and ISCOM Department which enabled me to carry out my PhD research activities, providing input, support and an ideal comparison environment, in particular I would like to express my gratefulness to Guido Riva and Andrea Neri.

I am deeply grateful to my colleagues Alessandro, Sergio and Luca, for their feedbacks, tolerance and encouragements. During this path I have really appreciate the helpful and meaningful exchange with Albenzio, Elena and Edion.
Furthermore I am sincerely grateful to Marco, his support and friendship since the days of the university has been a strength to me.

Un grazie particolare va alla mia famiglia, ai miei genitori e a mia sorella per avermi sostenuto anche durante questo percorso.

Infine il grazie di cuore va a Paolo il mio grande amore. Grazie per esserci stato sempre, per aver creduto in me e per il costante e prezioso sostegno. Grazie per l'affetto e la complicità dimostratami ogni giorno e grazie per far così parte della mia vita.

Introduction

The deployment of Next Generation Networks (NGN) provides new opportunities to increase consumer choice regarding general applications and services based on IP connectivity. The convergence of fixed and mobile networks and the wide availability of new services and applications (as video, data, broadcast TV, voice) bring out new challenges for Quality of Service (QoS) and for consumer protection.

The rapid market evolution of new broadband services and applications brings to the necessity of a deep renovation in telecommunication networks, both in terms of functionality, performance control and management that in terms of actual capacity provided to the customer.

In this context, the access network is heavily involved and it must evolve towards innovative architectures able to exploit the availability of new technologies, making it possible to offer high capacities required by the market, and at the same time realizing a network access infrastructure capable of evolving towards solutions with performance always increasing.

To satisfy these requirements both fixed and wireless mobile accesses are providing ever increasing data rate to network customers. The former with the widespread use of optical fiber, the latter with ever growing radio spectrum efficiency and utilization.

Regarding fixed access we are witnessing the introduction of architectures based on the use of fiber in increasing amounts in the access network (FTTCab, FTTB, FTTH), reducing the copper line length, maximizing performance obtainable from the xDSL technologies (in particular VDSL2 and vectoring).

However, architectures based completely on optical fiber require relevant investment, therefore it is necessary a careful planning with the aim of maximizing the performance reducing cost, so as to allow, in the future, development of these infrastructures in function the new market demands.

In this context, Quality of Service offered to end-user plays a fundamental role, especially with regard to the available bandwidth in the access section.

Focusing on mobile networks in the last years wireless mobile access has become more pervasive and effective in terms of transmission rate.

Mobile Internet access has experienced an explosion and this has triggered interest in the quality with which these services are offered. Standardization bodies of the mobile world (3GPP and IEEE) have developed several radio interfaces, tracing the evolution towards higher throughput.

Although technologies are tagged with theoretical peak throughput, in practice their performances depend on radio quality and sharing of access resources.

In this scenario a necessary step is to allow users to verify the actual performance offered by Internet Service Providers, especially if who use the service supports costs of new technologies.

In recent years Quality of Service (QoS) has become a major issue, not only in the communications field. On the one hand, QoS and its verification are the building blocks of a good relationship between users and operators, and secondly, QoS determines the positions of different actors on the market, motivating competitiveness.

The complexity of these issues instead collides with what should be the guiding principles in the evaluation of QoS, such as:

- providing QoS parameters easy to be understand and meaningful to customers;
- verified by independent organizations;
- have an accuracy of the QoS values at a level consistent with the measurement methods and at the same time as simple as possible and as low as possible with costs;
- get statistics such that the QoS values of different operators can be easily compared to each their users and customers.

These conflicting requirements require proceeding with a method, distinguishing as much as possible three different components that determine the overall quality perception from the user equipment, networks and services point of view [1].

These requirements have led to study and develop systems for assessing network performance that will necessarily involve the end users. For the development of such

systems, multiple implications are opening especially with uniformity and comparability standards.

One of the key elements to provide a customer service is the definition of a Service Level Agreement (SLA), i.e. contractual elements that define properly requirements that a service should satisfy. A SLA should define constraints through parameters defined at the same level of the service abstraction and should be directly interpretable. For example, quality parameters such as bandwidth or latency should be used to define supply contracts of infrastructural services while parameters such as throughput or response time should be adopted to define supply contracts of application services.

This work fits exactly in this area, in particular in the analysis and identification of techniques to measure quality of the service of Internet access from a fixed and mobile location by directly measurable metrics, such as throughput, delay or packets loss.

The purpose of this dissertation is to define a platform for QoS measurements and for Service Level Agreements verifications.

With the aim to establish a reliable measurement, it is started an experimental activity of analysis and performance evaluation of Internet connection. At first was verified the reliability and efficiency of evaluation techniques of network performance based on the most widely used protocols on the Internet.

Currently, most of the Internet services are based on the protocol stack TCP/IP and HTTP, therefore, all regulated by the TCP protocol. While the TCP protocol allows a reliable delivery of data through control and errors recovery mechanisms, on the other hand this mechanism becomes the major bottleneck especially in case of networks with high bandwidth-delay product.

To point out the state of the art on this issue, the ETSI standardization body does not define a measurement mode but suggests performing a measurement using the HTTP protocol (ETSI TS 102 250-5 V2.2.1). As a consequence available standards on this topic are lacking of a strict QoS measurement mode definition and do not address challenges related to high bandwidth-delay product networks.

Therefore considering on the one hand Internet network structure and on the other hand the rapid increase of network data rates, the implementation of comparable and meaningful QoS measures is non-trivial.

This PhD dissertation has exactly this prospective, i.e. develop a measurement framework for QoS verification considering different level of the OSI (Open Systems Interconnection) stack. In particular measures will cover the verification of line capacity, throughput and goodput of a single connection through measures performed respectively at physical (layer 1), transport (layer 4) and application (layer 7) layer.

Starting from previous studies, first step was to analyze TCP performance and subsequently determine the impact of different network parameters on quality evaluation both for fixed and mobile Internet access.

Simultaneously studies were carried out to obtain experimental evidence that quality of service depends not only from the network or from measurement type but also depends from the user terminal that is used.

To this aim, in order to consider mainly the aspect related to the user terminal, the voice service that is the most simple and consolidated service has been investigated.

Chapter 1
SLA Evaluation

1.1 Verification and Certification of Service Level Agreements

The ubiquity of Internet access, and the wide variety of Internet-enabled devices, has made the Internet a principal pillar of the Information Society. As the importance of the Internet to everyday life grows, reliability of the characteristics of Internet service (availability, throughput, delay, etc.) grows important as well.

The networks complexity itself leads to the difficulty of correlating different measurements that are performed by network equipment, test equipment and end user.

This is because telecommunications networks are comprised of several different layers of services, each with its own protocols and measurements. Measurements are realized to verify that network performance meets specific requirements for particular protocols and in each customer's service level agreement.

Service Level Agreements (SLAs) between providers and customers of Internet services regulate the minimum level of service provided in terms of one or more measurable parameters.

Currently SLAs are tested in terms of some network performance parameters as "bandwidth'' (generally expressed in terms of raw throughput). However, with the evolution of applications, SLAs will regard aspects more and more related to user perception.

There is need to define new metrics that take into account Quality of Experience (QoE), and to investigate on the introduction of SLAs based thereon.

In this scenario, each user may have SLAs with different providers (e.g. ISP, VoIP, IPTV). Therefore it is important to consider correlation among different SLAs. This will permit both verification of SLAs between providers and customers from either

the provider or customer end, as well as the customer-end verification of advertised throughput for comparison and regulatory purposes.

1.2 SLA Metrics and Measurements

With service evolution, the trend is to shift SLA verification towards applications.
This implies that metrics and measurements have to be applied at the application layer; therefore suitable SLA parameters must be defined for Web services such as YouTube, specified in terms of application-level measurable.
These Application SLAs (ASLAs) will necessarily depend on lower-layer SLAs: clearly, for high-quality HD video it is necessary to have a very good wide broadband physical connection that has to be guaranteed by a suitable SLA between ISP and user, even though that is not sufficient to guarantee high quality of service at application layer.
In measuring service levels in terms of throughput on a broadband connection, the SLA is often defined in terms of the physical characteristics of the connection from the user gateway to the network. This implies measurement method that is able to exploit all the capacity of the line.
It is well known that the amount of useful bits for seconds (defined as either goodput if considered at application layer, or throughput if at transport layer) can differ very much with respect to the capacity [2-10] offered at the physical layer, or in line capacity; for instance, in case of Transfer Control Protocol (TCP) [11, 12] the measured throughput can be much lower than the line capacity. This results particularly evident in case of network conditions with high values of bandwidth-delay product [6].
This difference between line capacity and throughput is a cause of several problems, especially in terms of SLA verification between customers and ISPs. Therefore, from the SLA point of view, novel tools are needed to simultaneously evaluate throughput (or goodput) and line capacity.
The most reliable method to estimate the line capacity is to locate a specific measurement device at the gateway of the user (modem, CpE,…). This method is clearly not scalable and it is very expensive, since it requires installing monitoring boxes very close to the customers' access links. Conversely, cheap methods to

evaluate the user connection quality are based on “speed test” tools which measure the time to download a specific data file from a dedicated server. Therefore such speed tests measure the application layer goodput (or the transport throughput) and they are not reliable to evaluate the line capacity.

The aim of this chapter is to explain how actual throughput measurements are derived, how they relate to other common network measurements, and how specific factors affect throughput measurements.

1.3 Network Performance Evaluation and OSI Model Relationship

With regard to the verification of SLA between operator and customer end users are interested in the performance experienced by applications of interest.

Unlike service level agreements relate exclusively to the network parameters.

The user applications performance depends not only on the network parameters also by the implementation of the protocols on which they are based. In order to accurately represent the user's perception throughput verification should be close to the applications. The inherent problem with this approach is that Application bandwidth is measured in the upper layers of the network stack (see Figure 1), while the bandwidth requirements stated in most SLAs are typically at layer 1, 2, or 3 measurements. As previously discussed throughput at layer 7 will always be less than that measured at layer 1, 2 or 3.

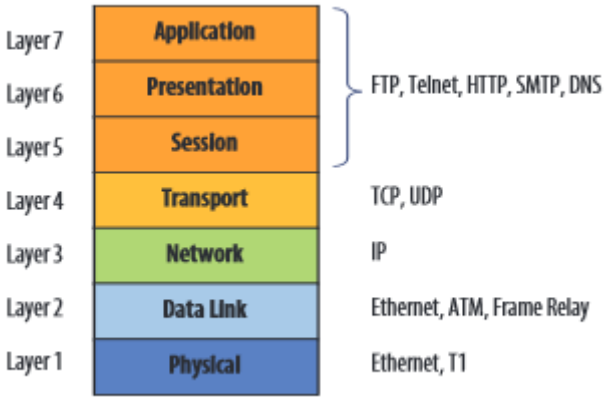


Figure 1: OSI Model

It is important to note that throughput can be measured at each layer of the network stack. Layer 1 capacity is the maximum rate of traffic that can traverse the link independent of the actual traffic type, and it is simplest measurement of throughput. The application throughput is not confrontable with throughput at lower layers because each layer of the network stack introduces overhead and more complex protocols will affect the throughput measurement. Therefore, directly comparing SLA throughput to application throughput is not trivial in order to verify the service provider's SLA.

For such reasons a measurement method to verify multilayer service level agreements is proposed. In particular this method measures simultaneously the line capacity to verify SLA between the user and the ISP (layer 2 measurement), the throughput as available bandwidth for user (layer 4 measurement) and the goodput as available bandwidth for user at application layer (layer 7 measurement). The proposed method provides both QoS parameters and QoE evaluation, in fact on the one hand the layer 2 and layer 4 measurements characterize the internet access infrastructure and protocol implementation respectively from a technical point of view, on the other hand layer 7 performance evaluation points out the actual end user experience. The overall opportunity is to integrate and correlate such key elements of the service provided to the user.

1.4 The TCP/IP Performance

The TCP [2] is a reliable transport layer and the most adopted protocol in the Internet. The protocol supports reliable data transfer by establishing a connection between the transmitting and receiving ends. It is a protocol that adapts to the network requirements regulating the number of packets it sends by inflating and deflating a window.

In fact TCP is also defined a "rate-adaptive protocol", in that the rate of data transfer is intended to adapt to the prevailing load conditions within the network and adapt to the processing capacity of the receiver. There is no predetermined TCP data-transfer rate; if the network and the receiver both have additional available capacity, a TCP sender will attempt to inject more data into the network to take up this

available space. Conversely, if there is congestion, a TCP sender will reduce its sending rate to allow the network to recover. To do that the TCP sender uses the acknowledgment feedback (ACKs) sent by the receiver to regulate the transfer rate. This adaptation function attempts to achieve the highest possible data-transfer rate without triggering consistent data loss.

This mechanism is regulated by the TCP congestion control algorithms. The congestion control scheme in current TCP implementations has two main parts: Slow start and Fast Retransmit and Fast Recovery algorithms.

During slow start phase, a TCP sender starts with a congestion window of one segment and exponentially increases the congestion window. When the congestion window (CWND) reaches a slow start threshold (ssthresh), the sender goes to the congestion avoidance phase increasing linearly the congestion window (see Figure 2).

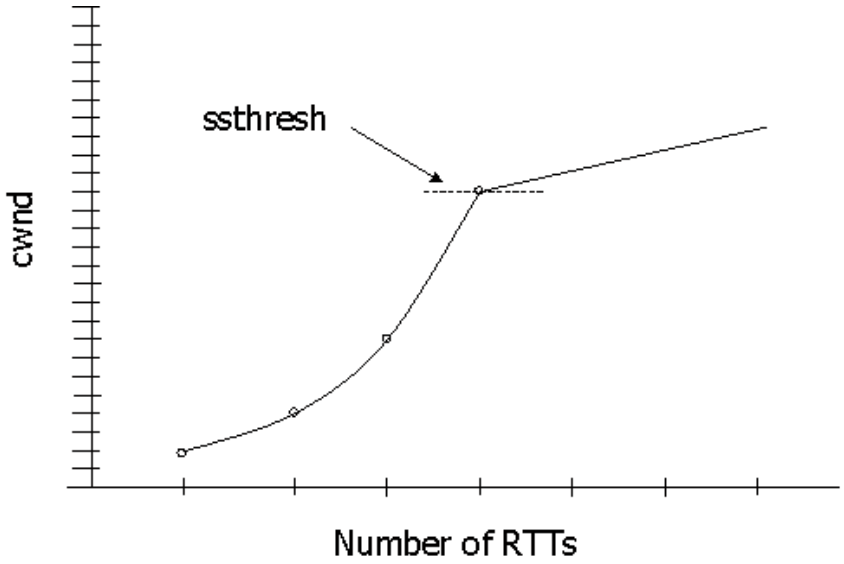


Figure 2: TCP Slow Start and Congestion Avoidance

Clearly the threshold choice is key to the algorithm performance. In the literature there are several studies that investigate on the choice of the threshold or the window size at the beginning of slow start phase [3].

The Fast Retransmit algorithm deduces that a segment has been lost when a sender receives three duplicate ACK's. This is due to receiver mechanism that acknowledges the highest in-order sequence number it has seen; so when it receives out-of-order

packets, it generates acknowledgments for the same highest in order sequence number (i.e. duplicate ACK's).

In addition, slow start threshold (ssthresh) and the congestion window are lowered to approximately half of the congestion window size prior to the Fast Retransmit to slow down the sending rate.

The Fast Recovery algorithm refers to the way the congestion window and ssthresh are adjusted so that after a Fast Retransmit, the sender slows down and enters a mode that linearly instead of exponentially increases the congestion window.

The main problem of this algorithm is the delay caused by packet losses due to congestion.

When TCP is employed for data transport, highly RTTs (Round Trip Time) can induce spurious timeouts, even though the involved packet actually is not lost but simply delayed. RTT is the time it takes for a signal to be sent plus the time it takes for an acknowledgement of that signal to be received. In addiction relevant network RTT can delay consistently the achievement of the connection steady state, making it unreachable for short-lived flows.

In this regard, it is important to study and investigate intrinsic limitations of TCP performance related to inner protocol algorithms, in order to detect key parameters and mechanism to further improve TCP performance in large bandwidth delay networks.

1.5 The Performance of TCP/IP for Networks with High Bandwidth-Delay Product

Some observations based on TCP performance investigation in WAN realized in [6] are reported in this paragraph to analyze the Performance of TCP/IP for networks with high bandwidth-delay products. In particular the study reported in [6] examines two cases: one that the bandwidth-delay product of the network is high compared to the buffering in the network and the second packets may incur random loss (due to congestion mechanism).

The interest is in networks with large round-trip delays, so that the buffering on the bottleneck link is typically of the same order of magnitude as, or smaller than, the

bandwidth-delay product. The bottleneck link may be shared by several TCP connections. In addition, this study assumes that each packet may be lost randomly even after obtaining service at the bottleneck link.

It is important to clarify that random loss causes performance deterioration in TCP because it does not allow the TCP window to reach high enough levels to permit good link utilization. On the other hand, when the TCP window is already large and is causing congestion, random early drops of packets when the link buffer gets too full can actually enhance performance and alleviate phase effects.

In this regard there are many simulation studies [2-5]; here I report the most significant ones.

In TCP packets size may be variable, so in this study the analyzed scenario is based on infinite data sources which always have packets to send, so that the units of data are maximum sized packets. A single bottleneck link with capacity packets per second and a FIFO buffer of size *B* packets is considered. Any packet arriving when the buffer is full is lost. It is important to note that random loss may cause additional losses. The number of connections sharing the link is assumed to be constant. The propagation delay of each connection includes:

1) the time between the release of a packet from the source and its arrival into the link buffer;
2) the time between the transmission of the packet on the bottleneck link and its arrival at its destination;
3) the time between the arrival of the packet at the destination and the arrival of the corresponding acknowledgment at the source.

You can see how the propagation delay includes all delays except for service time and queuing at the bottleneck link.

The propagation delay for a packet is taken to be deterministic, which implicitly assumes that deterministic propagation and processing delays are more significant than random queuing delays at all nodes and links other than the bottleneck link.

Each connection is assumed to use a window flow control protocol. At time *t*, the window size for connection is equal to the maximum allowed number of unacknowledged packets. The window size varies dynamically in response to

acknowledgment and detection of packet loss. Upon receiving a packet, the destination is assumed to send an acknowledgment back immediately. These acknowledgments are cumulative and indicate the next byte expected by the receiver.

In [6] the authors study two versions of TCP; here I report only the TCP Reno that is one of the more popular TCP implementations in the internet today and includes the fast retransmit algorithm.

The algorithm description of TCP Reno is reported in the following:

- After every non repeated acknowledgment the algorithm works as:
  - If $W<W_t$, set $W=W+1$; Slow start phase;
  - Else set $W=1+1/[W]$; Congestion avoidance phase.
- When the duplicate ack exceeds a threshold,
  - retransmit “next expected” packet;
  - set $W_t=W/2$, then set $W=W_{t;}$
  - resume congestion avoidance using new window once retransmission is acknowledgement.
- Upon time expiry, the algorithm goes into slow start as before
  - set $W_t=W/2$;
  - set $W=1$.

Reno after the number of duplicate acknowledgments exceeds a threshold retransmits the packets. After an initial slow start transient, it does not return to slow start after Fast Recovery (which ends on receipt of the retransmitted packet), instead it reduces the congestion window to half the current window size.

Each cycle begins when a loss is detected via duplicate ack. In particular if three duplicate ACKs are received (i.e., four ACKs acknowledging the same packet, which are not piggybacked on data, and do not change the receiver's advertised window), Reno will halve the congestion window, perform a Fast Retransmit, and enter a phase called Fast Recovery.

In this state, TCP transmits again the packet, assumed as missing, that was signaled by three duplicate ACKs, and waits for an acknowledgement of the entire transmit window before returning to congestion avoidance. TCP Reno experiences a time out

in case there is no acknowledgement; as a consequence of the timeout TCP enters the slow start state.

Assuming that loss occurs at windows size W_{max}, the windows size at the beginning of each cycle is $W_{max}/2$. In TCP Reno in each cycle the algorithm starts from $W=W_t=W_{max}/2$. The algorithm resumes congestion avoidance mode using new window until the window size reaches W_{max} again. At this point reached the value W_{max} a loss due to buffer overflow occurs and a new cycle with windows size $W_{max}/2$ begins.

The authors use W_{max} as a generic notation for the windows size at which congestion avoidance ends. So the W_{max} value could change from cycle to cycle if loss occurs randomly, or could be the same for all cycles if loss occurs periodically.

In this work the authors describe the evolution (see Table 1) of a single connection to derive expressions to calculate the throughput of the connection.

Define the normalized buffer size $\beta=B/(\mu\tau+1)=B/\mu T$, where τ denotes the propagation delay for each packet of the connection and $T=\tau+1/\mu$ denotes the propagation delay plus the service time. So the maximum window size that can be accommodating in steady state in the bit pipe is:

$$W_{max} = \mu T+B = \mu\tau+B+1 \quad (1)$$

The buffer is always fully occupied and there are μT packets in flight.

To conclude if the number of packets successfully transmitted during a cycle is N_c, and the duration of a cycle is T_c, then the periodic evolution implies that the average throughput is given by $\lambda=N_c / T_c$.

Time	Packet Acked	Window Size	Packet(s) Released	Queue Length
0		1	1	1
T	1	2	2,3	2
$2T$	2	3	4,5	2
$2T+1/\mu$	3	4	6,7	$2-1+2=3$
$3T$	4	5	8,9	2
$3T+1/\mu$	5	6	10,11	$2-1+2=3$
$3T+2/\mu$	6	7	12,13	$2-1+2-1+2=4$
$3T+3/\mu$	7	8	14,15	$2-1+2-1+2-1+2=5$
$4T$	8	9	16,17	2

Table 1: Evolution during slow start phase

The slow start evolution described in Table 1 consists in mini cycles of duration equal to the Round Trip Time (T), where the i_{th} mini cycles refers to the time interval [iT, (i+1)T]. The ack for a packet transmitted in a mini cycles i arrives in mini cycles (i+1) increasing the window size. The consequence is a doubling of the window in each mini cycles. The 1/μ value represents the service time for acknowledgment for consecutive packets served in mini cycles i.

The preceding evolution assumes implicitly that the normalized buffer size β<1, so that the window size during the slow start phase is smaller than μT and the queue empties out by the end of each mini-cycle. Denoting the window size at time t by W(t), and the queue length at time t by Q(t), the maximum queue length during the (n+1)th mini cycles is 2^n+1, which is approximately half the maximum window size during that mini cycle.

During a slow start phase with threshold W_t buffer overflow occurs only if $W_b \leq W_t$; where $W_b = 2^{nb}+B$ and $n_b = [\log_2(B-1)]$.

Case 1) When $W_b > W_t$ during the slow start phase (one slow start for cycle) the window size reaches $W_t = W_{max}/2$. The window size W(t) at time t, is equal to $W(t) = 2^{t/T}$, so to approximate the duration of this phase is $t_{ss} = [RTT^* \log_2(W_t)]$. In this phase $n_{ss} = W_t$ is the number of packets transmitted in the slow start phase. It's underline that during slow start the window size grows by one for every acknowledgment, so starting from an initial value of one, the number of packets successfully transmitted during slow start is equal to the windows size at the end of this period.

Case 2) When $W_b \leq W_t$ and buffer overflow occurs there are two slow start phases in a cycle. So the average throughput is given by the calculation of the number of packets transmitted during the two phases. The total spent in the two slow start phases is given by $t_{ss}=t_{ss1}+t_{ss2}$, where t_{ss2} is equal to the t_{ss} (case 1), instead $t_{ss1\ is}$ calculated by $[RTT^* \log_2(W_b+1)]$, which is the time to reach W_b and the time taken to detect the loss after the buffer overflow. Time to detect the loss is approximated to one Round Trip Time (RTT).

The congestion avoidance phase starts from an arbitrary window size W_0 (in Reno $W_0=W_{max}/2$ since the window size is halved after losing a packet) and terminates when the window size reaches $W_{max.}$

Theory and experimental tests [6-10] and [13, 14] showed that TCP flow and congestion control mechanisms suddenly become the major bottleneck when exploiting high capacity paths with large Round Trip Time (RTT). In particular the download time of a file depends on the Receiver Window (RWND) and the Congestion Window (CWND). The relationship between throughput and TCP parameters is shown below:

$$\mathrm{B}_t \le \frac{\min(CWND, RWND)}{RTT} \le \min\left(\mathrm{B}_c, \frac{RWND}{RTT}\right) \qquad (2)$$

Where CWND is the congestion window, B_t the throughput and B_c the line capacity, and RTT is the Round Trip Time.

In particular the sender is allowed to transmit no more than the minimum amount of data specified by the CWND and the RWND per each RTT. As for any sliding window protocol, TCP throughput is proportional to the window size, and inversely proportional to the sender-receiver RTT.

So the estimation of the average TCP throughput, calculated in (2), requires the knowledge of following parameters:

- window size (segment);
- segment size (byte);
- link propagation delay (ms);
- packet transmission time (ms).

$$\text{Rate} = W_{nd} / \text{RTT [byte/sec]} \qquad (3)$$

where:

$$W_{nd} = \text{window size} * \text{segment size [byte]}$$

To conclude this stack TCP/IP implementation provides congestion control algorithms based on slow start and Fast Retransmit and Fast Recovery algorithms. This implementation cut the window to half when it detects a loss. While this does provide better throughput under ideal conditions, the basic problem is that there

can be multiple window cutbacks due to a single congestion episode, and that multiple losses can lead to a timeout (which in practice can lead to significant throughput reduction if coarse timeouts are used). After the timeout expiration it is forced to restart with the slow start.

TCP may experience poor performance when multiple packets from a window of data are lost. One solution in order to handling isolated losses in a window without changing the window size is by using some form of selective acknowledgment (SACK) [15].

Basically considering only cumulative acknowledgments, a TCP sender can only learn about a single lost packet per round trip time. As mentioned above, an aggressive sender could choose to retransmit packets early, even if retransmitted segment have already been successfully received.

Selective Acknowledgment (SACK) is a strategy which combined with a selective repeat retransmission policy, can help to corrects this behavior in the face of multiple dropped segments.

The receiving TCP sends back SACK packets to the sender informing the sender of data that has been received. With selective acknowledgments, the receiver can inform the sender properly about all segments that have arrived successfully, so the sender can retransmit only the segments that have actually been lost.

It is important to note that this mechanism allows recovering multiple packets per RTT.

Chapter 2
Bandwidth Evaluation Techniques: Analysis and Characterization

Today's bandwidth estimation plays a key role in telecommunication evolution and market regulation in order to correctly define service level agreements between customers and Internet Service Providers (ISP). On the other hand customers are interested in network performance verification to better exploit ISP market competition.

Next Generation Networks provide opportunities and challenges for data communication, in particular the ever increasing bandwidth pose a major challenge to transport layer protocols such as TCP.

This chapter is focused on limits and opportunities related to measurements implementation based on TCP in wired and wireless broadband access networks.

The key role of TCP protocol in performance evaluation and verification is due to the wide diffusion of such protocol as a building block of the most Internet Services. In fact, most commonly used application-protocols, most notably HTTP, deeply rely on TCP so that download throughput for most applications is regulated by TCP at the transport layer.

In this section an experimental investigation about TCP protocol performance in a broadband access network is reported, considering current bandwidth evaluation best practice and looking forward at the Next Generation Access Networks (NGAN).

In particular an accredited standard technique for end-user bandwidth evaluation in a wired and wireless access scenario showing limitations in the bandwidth exploitation of user due to the transport protocol behavior has been applied.

This investigation analyzes how the end-to-end bandwidth measurements depend strongly on the TCP implementation in a broadband access network.

2.1 Overview on Bandwidth Definition and Measurement: the Quality Contest

QoS is the term that characterizes the performance from the network point of view and it is usually measured by means of some parameters as bandwidth, jitter, delay and packet loss. According to measurements carried out in field trials [13], the only parameter that is much critical to be measured is the "bandwidth", since it depends on too many factors that are independently by the physical channel and, in particular, to the OSI Layer we refer to. Currently, the Internet services and applications are based on IP protocol at the network layer, and most of them rely on TCP at the transport layer most notably HTTP. This implies that data transfer throughput is regulated by TCP. As mentioned in the previous chapter (see paragraph 5) TCP flow and congestion control mechanisms suddenly become the major bottleneck when exploiting high capacity paths with large Round Trip Time (RTT). The download time of a file depends on the Receiver Window (RWND) and the Congestion Window (CWND). The former depends on the Operating System (OS) installed on end user equipments [6, 7], while the latter depends on the congestion control implemented algorithms.

As describe in equation 2 the mechanism of the TCP congestion control drives the CWND values to the actual path capacity that cannot be larger than the capacity of the bottleneck link.

For example, in [9], it was shown that Windows XP (which is still today very popular) suffered a strong throughput reduction with respect to the channel capacity in case of 20Mb/s access as the one offered by ADSL2+ technology.

Important improvements were showed up by OSes like Windows 7 and Linux, even though limitations were well observed in exploiting higher speed access capacities in the presence of moderate RTT, cases that verify for accesses obtained by means of GPON.

The difference between goodput and line capacity could be deeply limited by avoiding flow control algorithms, as for an instance using UDP. Notably, UDP

implements no retransmission in case of packet loss, and does not perform congestion control.

These simple considerations point out that the QoS should be defined for each Layer of OSI stack, and as a consequence it is possible to define corresponding SLAs related to different OSI Layers, from the physical one to the application Layer.

From user point of view the most important metric is the Quality of Experience (QoE) that is just related to the user perception and depends on several human factors.

QoE measures user satisfaction related to the use of a service, in other words is a subjective measure of the user experience. Taking into consideration a video streaming service QoE should consider all degrading effects such as blur, jerkiness, blocking, freezes and so on [17]. These effects depend on the network performance but the relationship with network parameters is usually rather complex. QoE is used to be measured in terms of Mean Opinion Score (MOS) [17-24] and it can be evaluated with either subjective or objective tests. MOS based on subjective tests is generally evaluated by a pool of reviewers that look at the video services and manifest a score either following the quality evolution in time or at the end of the service [18, 19]. Conversely QoE based on objective tests uses software and algorithms tools to quantify the service degradation, for an instance also by using a reference service [19].

In case of HTTP video streaming applications like YouTube, previous studies [23, 24] have shown that the number of stalling events and their duration are the most important features influencing the QoE undergone by the end-user. A stalling event corresponds to the interruption of the video playback due to the depletion of the playback buffer at the user's terminal. It occurs when the available bandwidth is lower than the required video encoding bitrate. A deep investigation on the relationship between MOS and stalling events has been reported in [23] and such results have been used for this QoE evaluation.

2.2 The QoS and QoE Evaluation Method

The user Quality of Service (QoS) is measured by means of a specific QoS tool evaluation technique that was implemented in the framework of an Italian

resolution, decided by AGCOM Del 244 2008, that permits to each Internet user to verify his home wireline bandwidth [25].

Such a tool was implemented according to ETSI EG 202 057, and in particular it is based on File Transfer Protocol (FTP) file transfer [26].

The interest is focused only in the achievable end-to-end TCP throughput so an experimental evaluation is carried out analyzing performances and characteristics of a single TCP connection through conditions typical of broadband access networks.

Concerning Quality of Experience (QoE) (or subjective) measurement, the Mean Opinion Score (MOS) described in [27] and based on an average of values expressed by a group of reviewers that looked at the video services, is adopted. However such a method is difficult to be implemented, especially in field environment.

This approach does not permit to distinguish different kinds of video degradations as blur, jerkiness, blocking, freezes and so on, but it is enough to judge if the network configuration is either suitable to deliver a service with a quality required by user or if a reconfiguration is necessary.

However the interest for MOS evaluation was mainly to analyze network conditions that could induce some form of video services degradation. Therefore the adopted approach is based on detection of events that induced also a minimum MOS degradation (i.e. from 5 to 4) and we considered less relevant a different interpretation of an event in terms of MOS between 2 and 3. Therefore, according to the MOS investigation reported in [23, 24] it is adopted a MOS evaluation based on an only reviewer that had to decide a score according to the following guidelines:

- 5: no stalling,
- 4: only one stalling with a duration shorter than 0.5 s.,
- 3: only one stalling event with a total duration between 0.5 and 3s,
- 2: either for 2 stalling events with a duration lower than 4s or for one stalling with a duration between 3 and 8s,
- 1: for all the worse cases.

The evaluation was carried out on different video services according to the available throughput under test, and in particular from YouTube standard (360p) to HD

(1080p), up to H.264 video (1920 x 1080 pixel requiring an video encoding bitrate equal to 10 Mb/s), and MPEG-2, (1280 x 720 pixel, 18Mb/s).

2.3 Analysis and Bandwidth Measurement

This study aims for investigating bandwidth estimation techniques based on active probes designed considering protocols commonly used by customers: in this way bandwidth estimation matches the experienced performance closer to real user experience.

In this field some studies are already carried out in [8, 9], regarding measurements on ADSL2+ that is the most adopted technology by Telecommunication Operators in Europe.

The goal of this study is to understand the protocol dependencies of bandwidth measures both in wireless and wired environments.

Wireless environments pose formidable challenges when attempting to provide reliable, end-to-end data transmission for transport protocols such as TCP. In particular wireless networks have peculiar characteristics which can have a negative impact on the transmission performance:

- random packet loss;
- high and variable RTT;
- low channel capacity compared to wired access networks;
- asymmetric channels;
- frequent disconnections.

For these reasons to carry out this experimental investigation a fixed access network, such as ADSL2+, has been adopted. This choice comes from the necessity to directly control parameters and consequently network performances, excluding all random phenomena typical of wireless environment. In order to emulate mobile access networks, in this investigation typical parameters of the radio access, such as high RTT, have been taken into account.

In the assumption above mentioned, the Transmission Control Protocol (TCP) plays a key role in evaluating network performance, since it directly regulates the data flow.

In particular the TCP receiver sends back an acknowledgement for every received data segment, ensuring the proper execution of communication. On the other end of the connection if the transmitter does not receive an acknowledgement for a given packet when the corresponding timeout period expires, the packet is considered lost and subject to retransmission.

To take into account network condition, the transmitter starts a timeout mechanism when sending a packet to the receiver and constantly tracks the round-trip times (RTTs) for its packets as a means to determine the appropriate timeout period.

Moreover TCP implements some proactive mechanisms to prevent packet loss trying to avoid exceeding both the network and the receiver capacity.

These aims are prosecuted by means of two key algorithms: Flow Control [rfc 793] and Congestion Control [rfc 2581] and [12] [16] respectively, which cooperate data transfer tuning.

The choice of a TCP dependent technique for bandwidth estimation, tries to keep QoS evaluation as closer as possible to the end-user effective experience of broadband access services.

To determine the throughput and analyze the TCP window mechanism, network topology illustrated in paragraph 2.4 has been adopted; it is a bidirectional link between a source and a destination node characterized by a fixed capacity and delay. In particular the maximum throughput of a TCP flow depends directly on the sliding window size; if the sliding window is too large, there is a high probability of packet loss because the network and the receiver have resource limitations.

As known, at the start of a new TCP connection, the sender does not know the proper congestion window for the path, so in the slow start phase the congestion window grows exponentially, doubling every Round Trip Time until reach the threshold value. This value is an arbitrary default value depending on the operating system implementation; this setting is critical because it influences the TCP performance. In fact if the threshold value is set too high relative to the network Bandwidth Delay Product (BDP), the exponential increase of congestion window generates many packets, causing multiple losses with significant reduction of the connection throughput. On the other hand if it is too low, TCP may need a very long

time to reach the proper window size resulting in poor start-up utilization especially when Bandwidth Delay Product is large such as in wireless communications networks.

Thus, packet losses reduction requires minimizing the sliding window. Therefore, the open issue is finding an optimal value for the sliding window that provides good throughput, yet does not overcome the network and the receiver. This value is usually referred to as the congestion window. Furthermore, TCP should be able to recover from packet losses timely. This means that the shorter the interval between packet transmission and loss detection, the faster TCP can recover. However, this value cannot be too short, since the sender could early detect a loss and often results in an unnecessarily retransmission.

Starting from equation 2 and 3, the layer 2 Bandwidth - Delay product, determines the amount of data that can be in transit in the network just as Receive Window (RWND). It is the product of the available bandwidth and the latency. BDP is a very important concept in a window-based protocol such as TCP, since the throughput is bounded by the BDP value. As shown below:

$$\text{BDP (bytes)} = \text{bandwidth (Kbyte/sec)} * \text{RTT (ms)} \quad (4)$$

where

$$\text{RTT} = (\text{packet transmission time} + \text{link propagation delay})*2$$

To get full TCP performance, the ideal TCP window size needs to be large enough to accommodate the Bandwidth Delay Product and it represents the amount of information that uses completely the link between the transmitter and receiver.

To conclude it is necessary a good window size to maximize the transmission capacity of each link.

Then the optimal receive window size (W_{id}) for TCP flow control is equal to the maximum number of bits that the connection can hold at the same time divided by the size of each segment:

$$W_{id} = (\text{bandwidth} * \text{RTT}) / \text{segment size [byte]} \quad (5)$$

2.4 Experimental Setup

In this section a briefly description of adopted experimental and simulation setup to analyze network performance is reported. The investigation scenario is typical of wireless environment where networks are characterized by high delay. The network test bed is composed by ADSL2+ systems for the access part of the network; the ADSL2+ system consists of an Alcatel DSLAM (ISAM 7324) that allows us to use downstream links with different bit rates. According to reported tests, in Figure 3 a detail of the reference topology is illustrated; it represents a segment of Metropolitan Area Network with a core and an access part: a client accesses the network through ADSL2+ technology and communicates with a server that is linked to router Cisco 3.

Furthermore there is a delay generator (Shunra Wan Emulator) between server and client that provides different and high RTTs, reproducing one of the key characteristic of the mobile environment.

The aim is to outline the impact of TCP implementation in QoS evaluation and bandwidth customer exploitation.

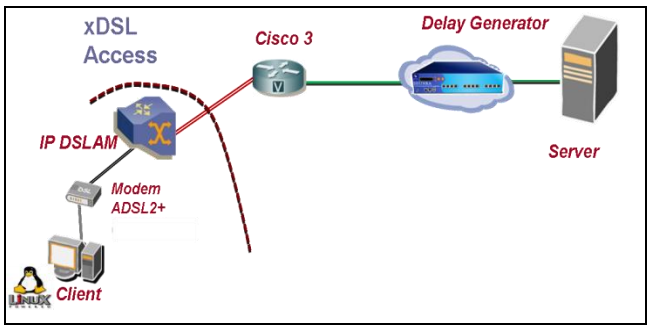


Figure 3: Experimental Setup

2.5 Experimental Results

The current paragraph shows results obtained from a set of experimental tests. This section analyzes the current TCP Slow start mechanisms, and evaluates their startup performance in high delay networks.

All results are obtained using iperf, a network testing tool [28] which is used to measure the end-to-end achievable bandwidth, using TCP data streams, allowing variations in parameters like TCP window size and number of parallel streams. In particular it has a client and server functionality, and can determine the throughput between the two ends, either unidirectional or bi-directionally. Iperf allows the user to set different parameters that can be used for testing a network, either alternately for optimizing or tuning a network.

Results collected in this experimental analysis are referred to different access network scenarios in terms of bit rate, delay and test duration that characterize access technology and protocols performances. To explain the obtained results, following index has been defined:

- Goodput: expressed in kbps represents the application level throughput, i.e., the number of useful bits per unit of time received;
- Line Throughput: expressed in kbps is defined as the average layer 2 data rate offered by xDSL connection;
- Measurement Ratio: is the parameter used to determine bandwidth evaluation accuracy and it represents the ratio between iperf-measured goodput and the offered line throughput.

Throughput tests are carried out establishing an end-to-end iperf TCP connection considering different test duration to download a file and introducing different network delays.

In Figure 4 and Figure 5 the obtained results expressed in terms of Measurement Ratio are reported, respectively for a 7 Mb/s and a 2 Mb/s link. This parameter represents an efficiency index that quantify the difference between estimated throughput and layer 1 maximum bandwidth (Line Throughput) provided by access network. Figures detail the trend of Measurement Ratio, in correspondence of different increasing BDP values (2). Each curve represents a test of different duration.

This analysis highlights the role of connection startup (TCP slow start) on the overall connection throughput; the goal is achieved considering different long-lived connections and their performance.

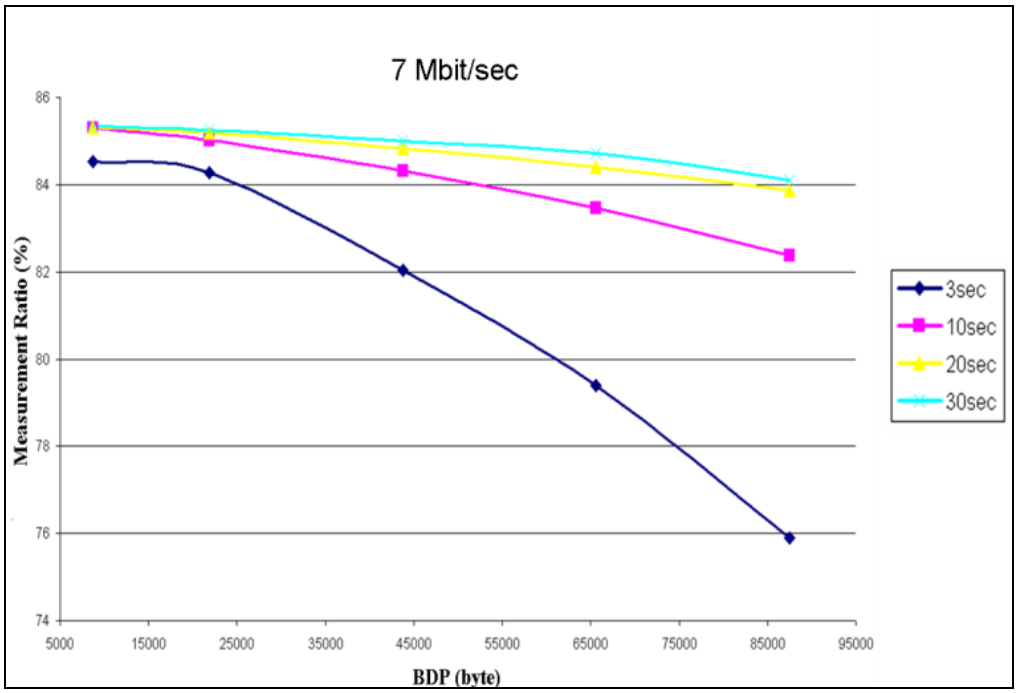


Figure 4: BDP impact over Measurement Efficiency – 7 Mb/s

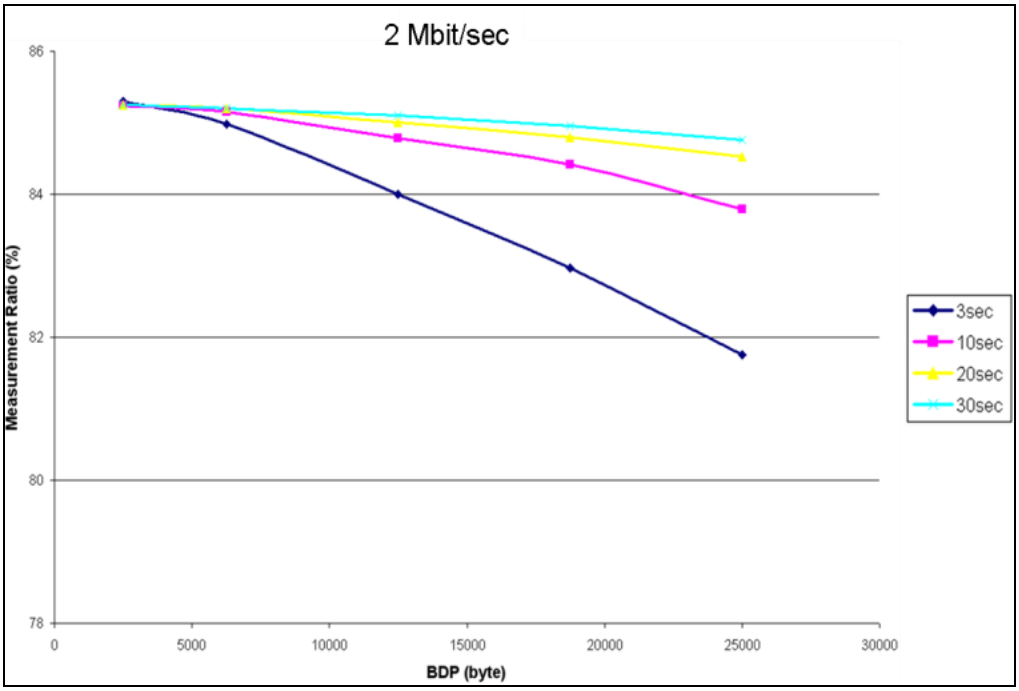


Figure 5: BDP impact over Measurement Efficiency – 2 Mb/s

It is possible to see how the growth of test duration (e.g. 30sec) increases the measured throughput percentage values especially in presence of high delay. This effect is due to TCP startup algorithm that grows progressively the data rate in order to avoid network overloading.

For completeness, in this section, TCP startup performance in large BDP networks, focused on Reno Congestion Window (CWND) dynamics, are evaluated.

As mentioned in the previous chapter, Reno is one of the more popular implementations in the internet today and it includes Fast Recovery mechanism [29]. It does not return to slow start after Fast Recovery (which ends on receipt of the retransmitted packet), instead it reduces the congestion window to half the current window size. In particular if three duplicate ACKs are received Reno will halve the congestion window, perform a Fast Retransmit, and enter a phase called Fast Recovery.

Figure 6, Figure 7 and Figure 8 show the CWND trend of the Reno protocol in the startup stage and the relative throughput to this configuration. In particular the receiver issues an Acknowledgement (ACK) for every data packet received.

Let's assume the receiver's advertised window is always large so that the actual sending window is always equal to CWND. For convenience, the window size is measured in number of packets, and the packet size is 1460 bytes.

In TCP Reno, a sender starts in the slow start phase, the congestion window is below than a threshold value *(CWND < SSTHRESH)* and every ACK received results in an increase of CWND by 1 packet. Thus, the sender exponentially increases CWND. When CWND reaches the slow start Threshold value (SSTHRESH), the sender switches to congestion avoidance phase, increasing CWND linearly. It is important to note that this last phase is considerably slower than in slow start.

To counter the effects of TCP slow start iperf should be run for fairly long periods of time.

In fact by definition the slow start duration (as described in Chapter 1), is roughly:

$$[\log_2 (\text{ideal\_window\_size\_in\_MSS})] * RTT \qquad (6)$$

In practice, it is a little less than twice this value because of delayed ACKs. In this scenario, the bandwidth-delay product for 7 Mb/s network with an RTT of 150 ms is about 90 segments, 1460-byte. The slow start duration, will be approximately [log2 (90) * 0.15], which is 1 second.

Assuming a stable congestion window after slow start, the time for the cumulative bandwidth to reach 90% of the achievable bandwidth will be about:

$$10*slow\_start\_duration - 10*RTT \quad (7)$$

This means for the 7Mb/s-150ms network path, it will take over 8.5 seconds for the cumulative bandwidth to reach 90% of the achievable bandwidth.

In addition, to confirm this model, all packet transmissions are recorded for a detailed analysis using tcpdump. The analysis is carried out by using tcptrace, a tool [30] written by Shawn Ostermann at Ohio University. It analyzes TCP dump files taking as input the files produced by several popular packet-capture programs.

Thanks to this tool it is possible to produce several different types of output containing information on each connection under test, such as elapsed time, bytes and segments sent and received retransmissions, round trip times, window advertisements, throughput, and more.

Results represented in Figure 6, Figure 7 and Figure 8 report on the Y-axis the throughput and the in-flight data respectively and the X-axis the connection time. At the top of figures, red line tracks the throughput seen from the last few samples, calculated as the average of N previous dots (these points are not shown for intelligibility of the figure). The green line represents the average throughput of the connection up to that point in the life time of the connection (total bytes seen/total seconds so far).

The idea behind the graphs at the bottom of figures is to estimate the congestion window (CWND) at the sender side. Since this cannot be determined accurately, in-flight unacknowledged data as estimation are adopted.

- Green line tracks the window advertised by the opposite end-point i.e., the receive window (RWND);
- Blue line represents instantaneous in-flight data samples at various points in the lifetime of the connection (CWND);
- Yellow line tracks the average in-flight data up to that point;
- Red line tracks the weighted average of in-flight data up to that point.

These results are carried out for a bottleneck bandwidth of 7 Mb/s, for RTT values respectively of 150 and 750 ms (typical of the mobile access networks) and for different test duration. The buffer size is set equal to BDP in each test.

Figure 6 shows that a 3 second measurement is not sufficient for a 7Mb/s-150ms bandwidth-delay network path, since the slow start duration is comparable with the session duration. As a consequence, the achieved throughput is only 4.5 Mb/s, much lower than the actual 7 Mb/s.

In Figure 7, instead, when measurement test increases (30sec), the achieved throughput for the same operative scenario is equal to 5.9 Mb/s.

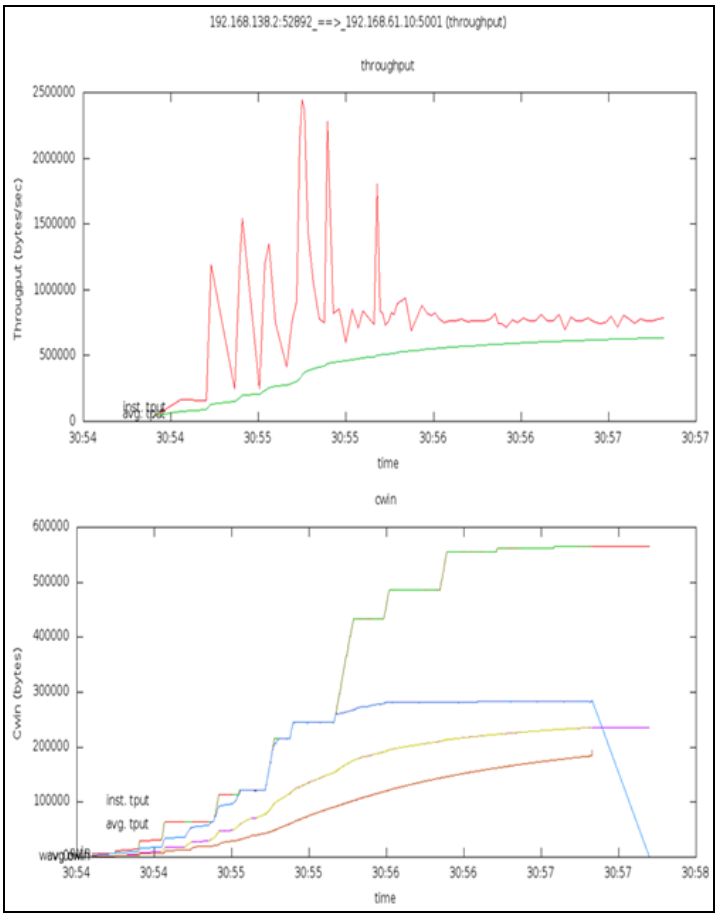


Figure 6: Throughput and CWND vs time
(line capacity=7Mb/s, RTT=150ms, duration test=3sec)

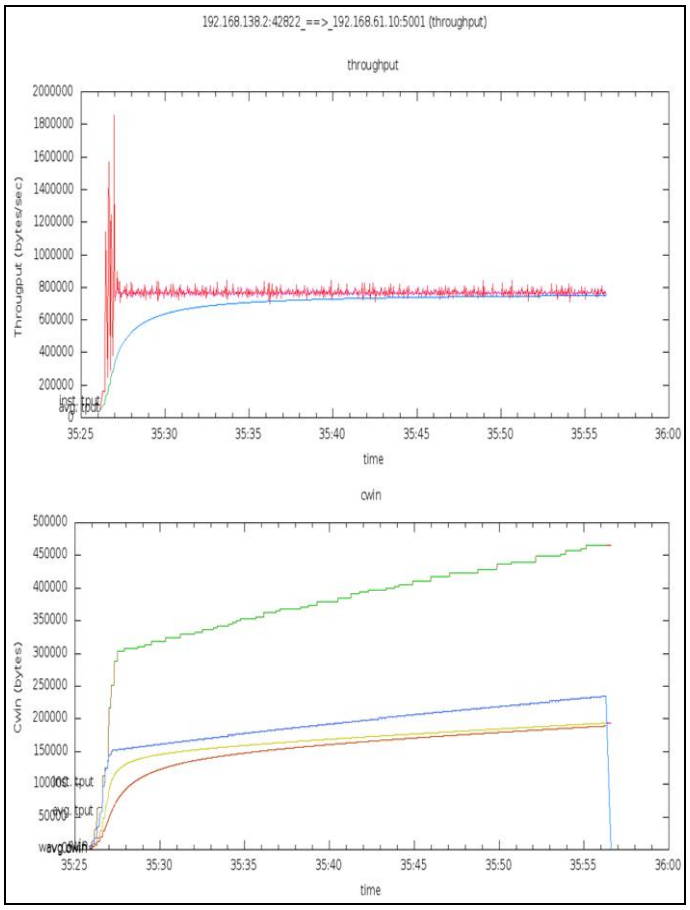


Figure 7: Throughput and CWND vs time
(line capacity=7Mb/s, RTT=150ms, duration test=30sec)

Considering a 7Mb/s-750ms network path, results for a 3 second measurement have not been reported due to the heavy network delay impact on the slow start algorithm.

In particular the duration of the slow start phase is about 7sec. For a 30 second measurement (see Figure 8), Reno stops exponentially growing CWND long before it reaches the ideal value (BDP=656Kbyte). After that, CWND increases slowly, and has not reached 450 packets by 30sec. As a result, the achieved throughput is only 3.8 Mb/s, much lower than the desired 7 Mb/s.

To conclude like other window-based protocols, TCP performances depend on RTT value. In particular when RTT increases (e.g. 750ms), the ideal window grows too.

On the other hand, because CWND increases 1 packet per RTT during congestion-avoidance, longer RTT means slower CWND growth, resulting in even lower utilization.

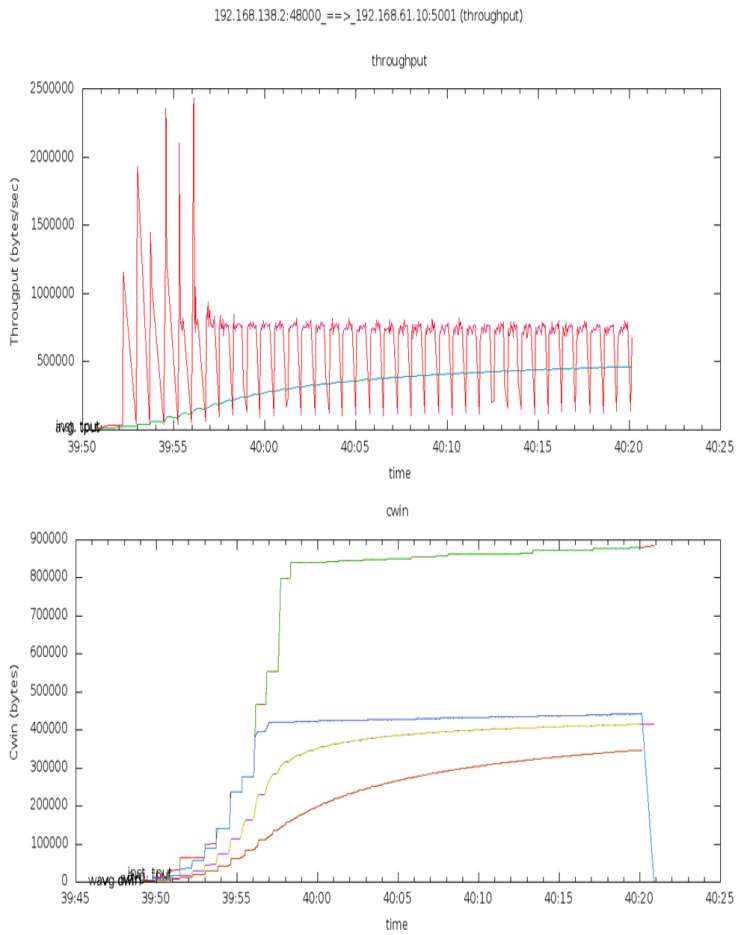


Figure 8: Throughput and CWND vs time
(line capacity=7Mb/s, RTT=750ms, duration test=30sec)

To conclude the obtained results show how the TCP session duration is an important parameter in order to fully exploit the available bandwidth. This consideration heavily impacts in performance measurements based on TCP, as a consequence in an accurate tuning of the session duration is mandatory to perform an effective bandwidth evaluation. In particular session duration dimensioning must to take into consideration the slow start period of the TCP session; as illustrated in Figure 8, in a 7Mb/s-750ms network path the duration of the slow start phase is about 7sec so the duration test must be greater than 30sec otherwise TCP does not reach the ideal value (BDP=656Kbyte).

2.6 Simulation of TCP Behavior

To confirm these results and analyze the performance of a 100 Mb/s connection, software which estimates the TCP performance reproducing the TCP protocol behavior, has been realized. In this software implementation the tcp protocol phases are emulated considering the sender/receiver interaction and the link delay, in each protocol phase from slow start to the congestion avoidance steady state. In this way the software is able to reproduce the protocol behavior accordingly with equations 1 and 2.

This software measures the connection throughput according to the congestion control mechanism of TCP and taking as input following parameters:

- Bandwidth value of the connection to be analyzed;
- RTT value (both fixed and variable);
- Packet size;
- Initial congestion window size;
- Threshold value;
- Receive buffer value;
- Test duration.

Mechanism of congestion window evolution has been implemented with the aim to emulate TCP behavior. Figure 9 shows a screenshot of the parameters set to calculate the throughput of a 100Mb/s connection with a RTT equal to 50msec.

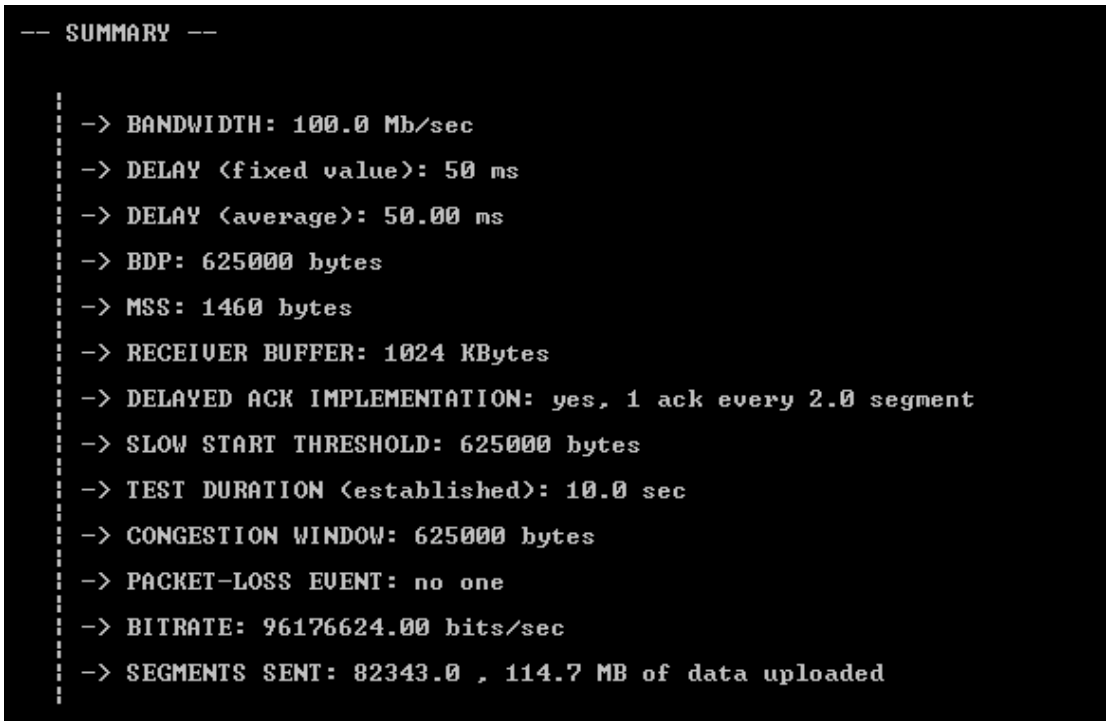


Figure 9: TCP Simulator implementation

The software is implemented in C considering the TCP algorithm implemented in Linux operating system (kernel 2.6) which enables the “window scale option”. This field allows to increase the TCP transmission window [31], therefore the maximum value reached by the window is equal to 65535 * 2^14=1,073,725,440 bytes [32].

Also in the Linux kernel 2.6, TCP adopts dynamic adaptation mechanisms of the receive window in order to have the window size large enough to accommodate the Bandwidth Delay Product [33].

In Figure 10 a detail of the packages exchange implemented in the simulator is shown. Figure reports the slow start (SS) phase where there is an exponential growth in the window size and the transition to the Constant Phase (CP) where the congestion window is totally opened. This transition occurs as soon as the CWND reaches the BDP value.

For the test implementation, approximately the congestion window size is around to the BDP value.

Each RTT value the simulator generates a number of segments according to the TCP protocol mechanism.

```
..OK            STARTING SIMULATION.....

TIME>    50 ms    CWIN>       1460 B    SEG>    1   RTT>    50 ms    MODE> SS
TIME>   100 ms    CWIN>       2920 B    SEG>    2   RTT>    50 ms    MODE> SS
TIME>   150 ms    CWIN>       5840 B    SEG>    4   RTT>    50 ms    MODE> SS
TIME>   200 ms    CWIN>      11680 B    SEG>    8   RTT>    50 ms    MODE> SS
TIME>   250 ms    CWIN>      23360 B    SEG>   16   RTT>    50 ms    MODE> SS
TIME>   300 ms    CWIN>      46720 B    SEG>   32   RTT>    50 ms    MODE> SS
TIME>   350 ms    CWIN>      93440 B    SEG>   64   RTT>    50 ms    MODE> SS
TIME>   400 ms    CWIN>     186880 B    SEG>  128   RTT>    50 ms    MODE> SS
TIME>   450 ms    CWIN>     373760 B    SEG>  256   RTT>    50 ms    MODE> SS
TIME>   500 ms    CWIN>     747520 B    SEG>  512   RTT>    50 ms    MODE> SS
TIME>   550 ms    CWIN>     625000 B    SEG>  428   RTT>    50 ms    MODE> CP
TIME>   600 ms    CWIN>     625000 B    SEG>  428   RTT>    50 ms    MODE> CP
TIME>   650 ms    CWIN>     625000 B    SEG>  428   RTT>    50 ms    MODE> CP
TIME>   700 ms    CWIN>     625000 B    SEG>  428   RTT>    50 ms    MODE> CP
TIME>   750 ms    CWIN>     625000 B    SEG>  428   RTT>    50 ms    MODE> CP
TIME>   800 ms    CWIN>     625000 B    SEG>  428   RTT>    50 ms    MODE> CP
TIME>   850 ms    CWIN>     625000 B    SEG>  428   RTT>    50 ms    MODE> CP
```

Figure 10: Transition from the Slow Start to Constant Phase

This result is confirmed by laboratory tests performed by the same experimental setup described in Paragraph 2.4 and reported in Figure 3.

In figure 10 the achieved throughput (upper part) and the congestion window (lower part) are shown, both in function of time.

As previously described, at the top of Figure 11, green line represents the average throughput of the connection up to that point in the life time of the connection (total bytes seen/total seconds so far).

In the lower part of the figure, instead, the congestion window (CWND) is represented by the blue line while the receive window (RWND) by the red line.

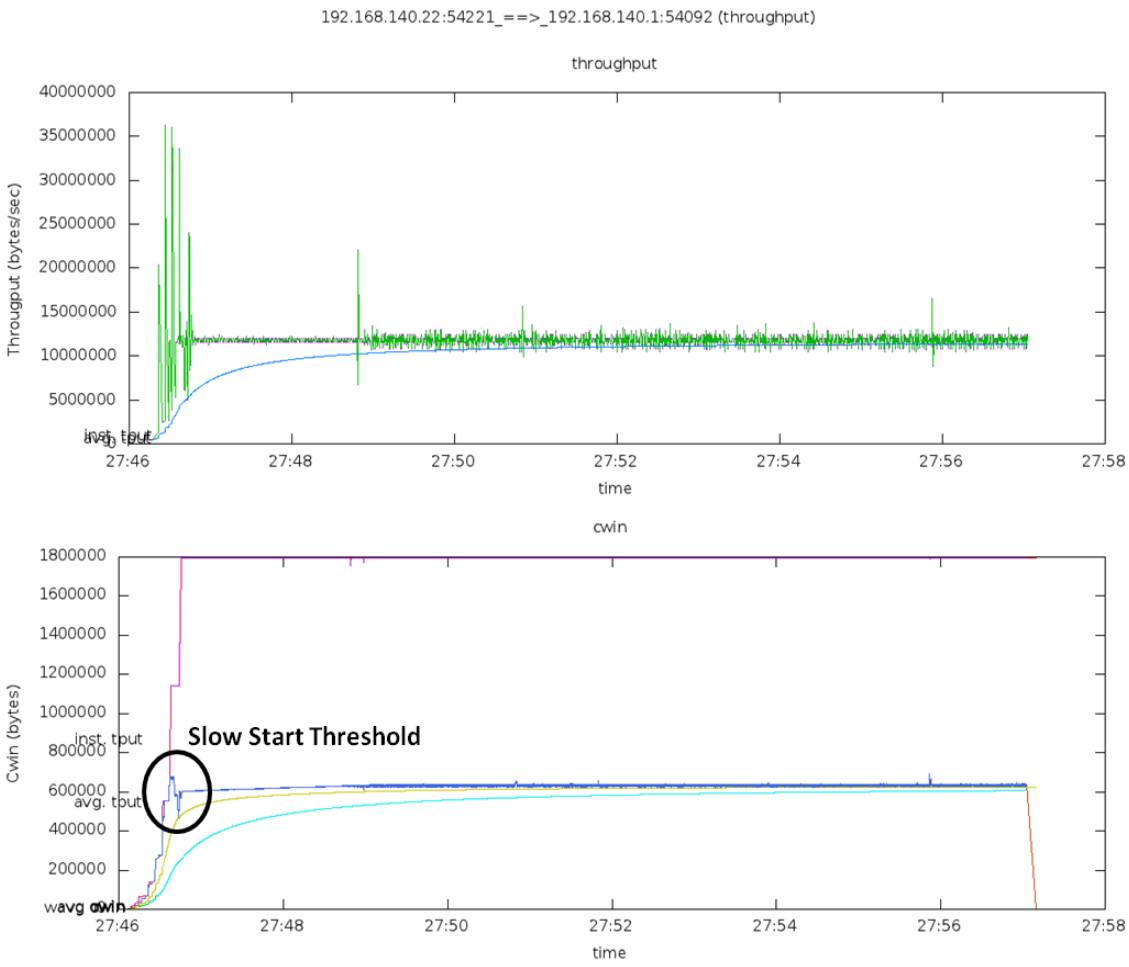


Figure 11: Throughput and CWND vs time (line capacity=100Mb/s, RTT=50ms)

In order to optimize network performance, the slow start threshold parameter for the transition from a slow start to congestion avoidance phase is not set a priori but through the "window scale option" mechanism implemented in Linux operating system.

In particular, this parameter varies on the basis of information exchanged by client and server during the handshaking phase (SYN messages).

Figure 12 reports a screenshot of the packages exchange, implemented in the simulator (when RTT is equal to 100 ms), in the transition from Slow Start (SS) to Congestion Avoidance (CA) phase.

During slow start an exponential growth of the window size is observed. After reaching the receive buffer size (1024 kbyte), enters the congestion avoidance phase where congestion window has a more conservative growth through the Delayed ACK mechanism [34]. The transition between these two phases, observable both from

tests performed by simulator (see Figure 12) both by laboratory tests (see Figure 13), requires a transient state period.

As mentioned above, after this stage, the growth of the congestion window continues in congestion avoidance mode until the end of the test or until the receive buffer value [35].

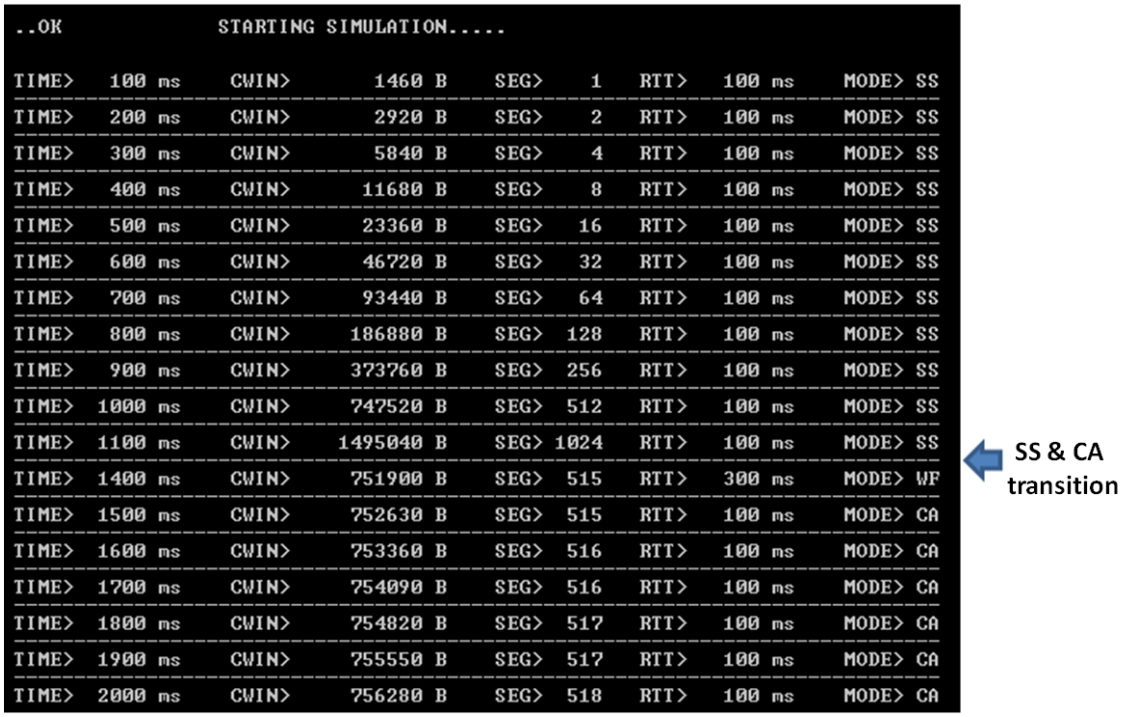


Figure 12: Transition from Slow Start to Congestion Avoidance phase

As you can see from the screenshot reported in Figure 12 and from Figure 13 CWND grows more slowly (one segment every two RTT) because it follows the Delayed ACK mechanism. As explained in RFC 3465 [34], when receiver implements this mechanism, half of all ACK is observed; this implementation requires sender to open the congestion window more conservatively.

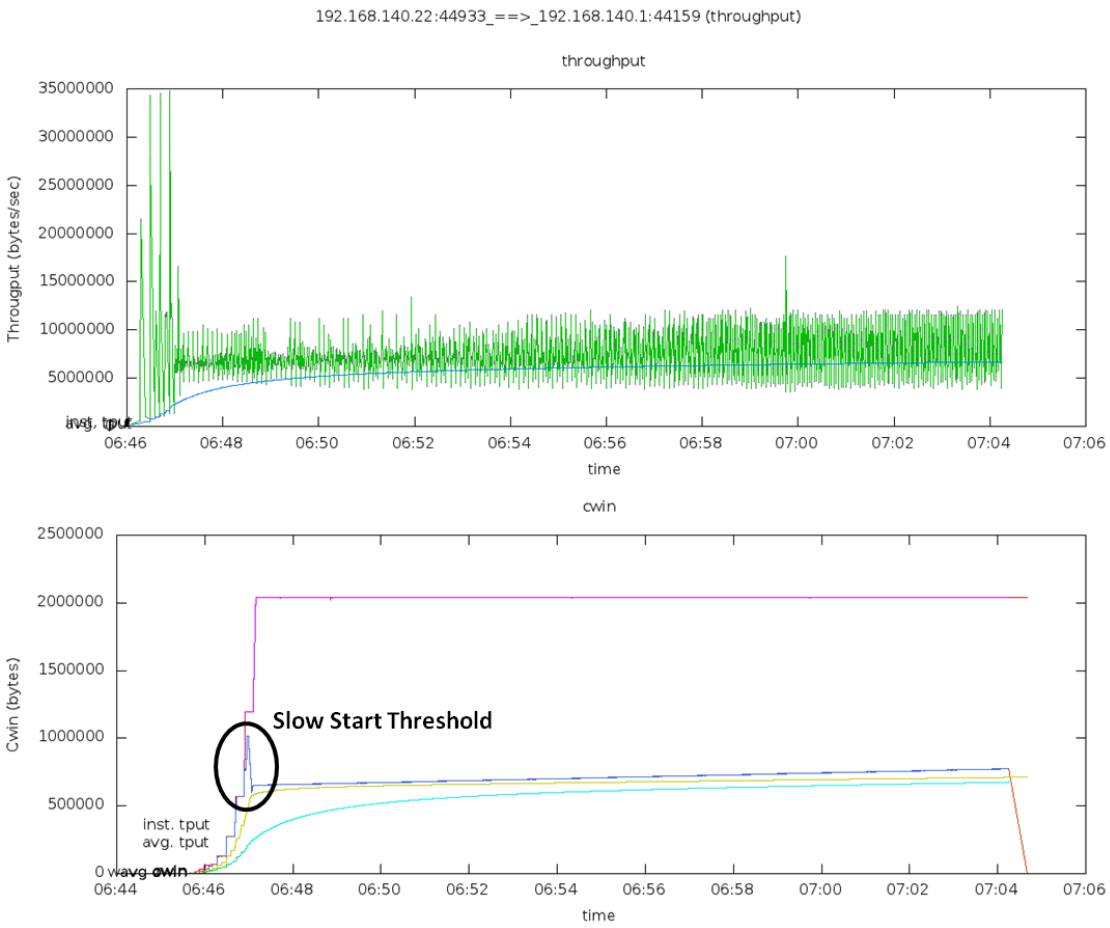


Figure 13: Throughput and CWND vs time (line capacity=100Mb/s, RTT=100ms)

2.7 Multi-Level QoS Measure Definition: the Proposed Solution

The results shown in this chapter suggest the need to find a method to simultaneously investigate QoS according to different aspects related to the OSI layer [36].

In particular measures should include the verification of:

- the line capacity to verify SLA between the user and the ISP;
- the throughput (as available bandwidth for user at physical layer);
- the goodput (as available bandwidth for user at application layer).

Starting from this study, a software agent performing this set of measurements is implemented [37]. In particular, the building blocks of the algorithm implemented by the agent are the following:

i. First, the agent performs throughput measurement by adopting a classic TCP file download method to evaluate B_t. The throughput measurement was implemented according to ETSI EG 202 057, and in particular it is based on File Transfer Protocol (FTP) file transfer.

ii. In the meanwhile it measures other QoS parameters as delay t_d (RTT measured with PING), congestion window and threshold value.

iii. After RTT and B_t measurements, an estimation of the channel capacity, B_c, can be obtained by means of equation (2).

iv. Secondly, to measure the line capacity an UDP test to obtaining other QoS parameters such as jitter and packet loss is performed. UDP test is divided in two steps. First a UDP stream is sent from the server with a line capacity of B_c', measuring the lost datagrams with respect to the transmitted datagrams. In such a way the useful transmitted bits BT_U are obtained and from the total transmitted bits BT_T, the estimated line capacity as $B_c = (BT_U/BT_T) * B_c'$ is achieved. In the second step the effective line capacity is measured by means of an UDP test based on a downloading of a file and forcing the transmission rate $B_c^* = B_c - \varepsilon$ (i.e. $\varepsilon = 0.001 * B_c$), verifying that a packet loss is lower than 0.1%. The value of ε depends on the required reliability of the measurement and in particular on the SLA between ISP and user. If the required packet loss is verified B_c^* is the line capacity.

This approach permits to simultaneously measure throughput and line capacity offering a method to verify multilayer service level agreements. Results are illustrated in the next chapter.

Chapter 3
Multi-Service Level Agreement Verification

3.1 Bandwidth Measurements in Gigabit Passive Optical Networks

The ever-increasing bandwidth-hungry applications bring Internet Service Providers (ISP) and industries to adopt new architectures based on Fiber To The Building (FTTB) and Fiber To The Home (FTTH) solutions and in particular by adopting Gigabit Passive Optical Network (GPON) systems [38]. These new architectures provide wider and wider bandwidth but how and if such a bandwidth will be really exploitable by end-user is still an open question, as described in Chapter 1, especially since the relationship among network quality (Quality of Service, QoS) and the user quality (Quality of Experience, QoE) can depend on different factors [18].

This work has a dual aim:

- To show methods to simultaneously measure throughput and physical channel capacity in very wide broadband GPON access,
- To illustrate how to overcome the throughput limitations due to TCP implementation analyzed in Chapter 2.

This investigation has been realized in a Gigabit Passive Optical Networks (GPON) Access. Preliminary results point out that the huge capacity offered by the GPON highlights the enormous differences that can be showed among the available and actually exploitable bandwidth. In fact, while the physical layer capacity can reach value of 100 Mb/s and more, the bandwidth at disposal of the user (i.e. either throughput or goodput) can be much lower when applications and services are based on TCP protocol (as illustrated in Chapter 1). As discussed in the previous chapters this is more evident in the case of network conditions with high values of bandwidth-delay product which are especially common in high capacity accesses with line rate higher than 20 Mb/s.

Furthermore in [16] it was shown that the goodput is quite close to the channel capacity in case of conventional ADSL2+ connections having a channel bit rate lower than 10 Mb/s. However, for line rate higher than that, the obtained throughput can be very different from the actual line capacity.

This is due to TCP protocol mechanism as expressed in equation 2 (Chapter 1).

In particular the download time of a file depends on RTT values but especially on the receiver window (RWND) adopted by the Operating System (OS) that are installed on a computer by an end user.

For example, in Figure 14, it was shown that Windows XP (which is still today very popular) suffered a strong throughput reduction with respect to line capacity in case of 18 Mb/s access as the one offered by ADSL2+ technology.

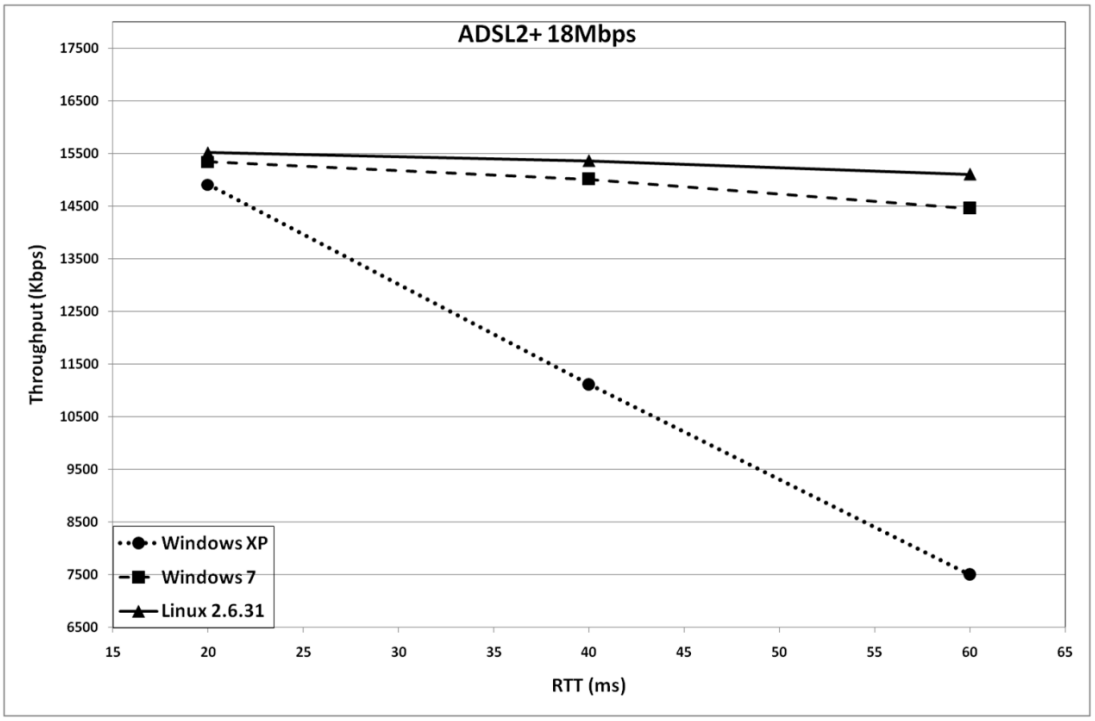


Figure 14: Throughput vs RTT for different OSes

The difference between line capacity and throughput is a cause of several problems, especially in terms of Service Level Agreement (SLA) verification between user and ISP.

This difference could be deeply limited by avoiding mechanisms of flow control as for an instance in the UDP case, even though such a protocol is not able to control and manage packet loss processes and does not perform congestion control.

These simple considerations point out that the QoS should be defined for each Layer of OSI stack, and as a consequence it is possible to define corresponding SLAs related to different OSI Layers, from the physical to the application layer.
With service evolution, trend is to shift the SLA verification towards applications.
In this work it is realized a tool that simultaneously evaluates throughput and line capacity, and it is defined and developed a method to suitably exploit the huge capacities offered by the optical fiber accesses.

3.2 The Experimental Test-Bed

The experimental tests were carried out in a multivendor integrated network test bed. Figure 15 reports the experimental test bed consisting of a core and an access segment. The core consists of four IP routers Juniper M10i (Ji, i=1,…,4) and three Alcatel SR7750 routers (ALCi, i=1,…,3)) interconnected by the optical fibers contained in the Roma-Pomezia cable (50 km) [39]. Three Cisco 3845 routers (Ci, i=1,…,3) are deployed at the access part of the network by means of Gigabit Ethernet optical links. The access network consists of copper and fiber sections. GPON access network is composed of an OLT (Optical Line Terminal) and up to eight Optical Network Terminations (ONT), offering a shared bandwidth equal to 1.244 Gb/s. A desktop PC is connected to the user ONT. A high-end server is then connected using Gigabit Ethernet links.
To simulate different network conditions that can arise in the core/backbone networks, traffic by means a delay generator in the test bed, is included.

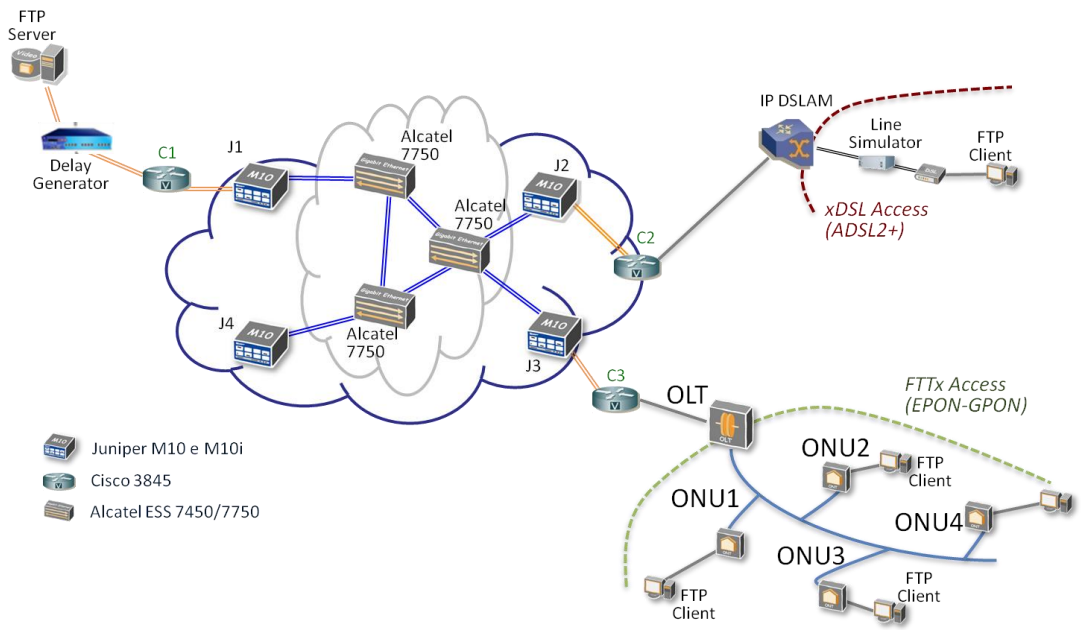


Figure 15: Experimental Test-Bed

This network can deliver audio/video services based on IP, in unicast and multicast modality, and on DVB-T over fiber as broadcast modality. IP services rely for transport on application layer protocols such as HTTP (Hypertext Transfer Protocol) or RTSP (Real Time Streaming Protocol), enabling all the functionalities of a common video player (Play, Pause, Stop, Fast Forward, etc.) [40]. Video server has VLC (Video Lan Client) configured by VLM (Video Lan Management) console. Both Standard Definition (SD) and High Definition (HD) Video are available and in particular Full HD (1920x1080p) is encoded in H.264 video and AAC Multichannel audio with a total bit-rate of about 9 Mb/s.

DVB-T over fiber broadcast modality is achieved by optically converting radio frequency signal coming out from a commercial terrestrial TV antenna.

TV quality perceived by the user is investigated by looking at four HD TV devices, each one located at the ONU output. Each HD TV is connected to the ONU both by means of a PC, that is used as a Set-Top-Box for IP services and by means of a commercial DVB-T Set-Top-Box, connected to the RF ONU output.

It has be pointed out that the PC and the Set-Top-Box could be either a unique device or included directly in a HD TV.

One of the multicast forwarding advantages is the fact that the replica of the packets is carried out only at the network edge avoiding traffic jam. Multicast can be

performed both at layer 3 and layer 2. On test-bed network, to guarantee an End-to-End minimum bandwidth in the backbone path, a technique described in [41] is implemented.

This technique allows us to assign a guaranteed bandwidth between two end-points of the network by means of different tagging techniques, i.e. Virtual LAN and Virtual Private LAN Service (VPLS). This technique is setup to avoid that background traffic, that may be present on the test bed, affects ongoing tests.

QoS analysis is carried out by a specific evaluation technique, based on the ETSI recommendation [26], using FTP tests and iperf tool [28]. Each experiment has been repeated 50 times; average results are described in the following paragraph.

3.3 Experimental Gigabit Passive Optical Networks Verification

In order to carry out a performance analysis in a common reference framework is referred to a bandwidth estimation method proposed by European Telecommunications Institute (ETSI) in [26]. In particular each run in the experiment consists in downloading of a file sized 10 times the line speed between client and server.

In other words means that the data transfer duration is at least 10sec.

The choice of using a long transfer size is in order to limit the impact of transient behavior at the start of a TCP connection.

QoS tests are carried out by means the three most popular OSes as client, namely e.g. Windows XP SP3, Windows 7, and Linux Ubuntu.

Figure 16 illustrates the throughput versus RTT behavior for different OSes and for a line capacity of 100 Mb/s. The analytical behavior from eq. (2) in case of Windows XP and Linux, assuming as RWND 512 and 64 Kbyte for Linux and Window XP respectively, is also reported.

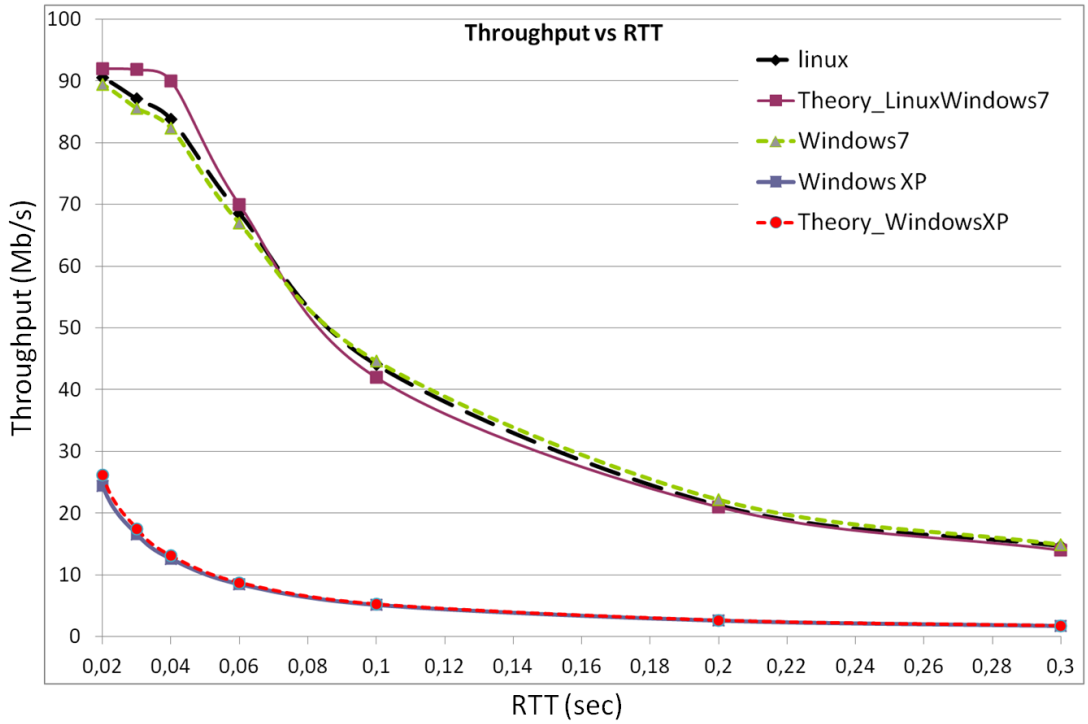


Figure 16: Throughput vs RTT for different OSes

Even in presence of a very wide bandwidth access, as in case of GPON (100 Mb/s), both network degradations and intrinsic limit of TPC can induce severe limitations for end user services, furthermore, another important limiting effect is the delay due to RTT, that in conjunction with the behavior of the operating system can limit the bandwidth exploitation. Evident substantial differences between Windows XP and advanced OSes (Windows 7 and Linux) are due to enhanced TCP algorithm implementation, related to adaptive parameters in the algorithm (i.e. Auto-Tuning of the RWND which can grow much larger than 64kB offered by Windows XP).

It is important to note that performance degradation can be observed also considering advanced OSes for high value of network delay. In particular, for 100 Mb/s profile, only 65 Mb/s are measured when RTT is equal to 60 ms.

The effects of the limitations due to the OS on the HD video quality perception are shown in Figure 17 where the MOS vs RTT behavior is illustrated. For all the OS, but XP, the MOS is always equal to 5; conversely XP drops when RTT is higher than 50 ms.

In case of a XP operating system, results show that also for a small RTT as 60 ms a significant degradation for HD video is revealed [42].

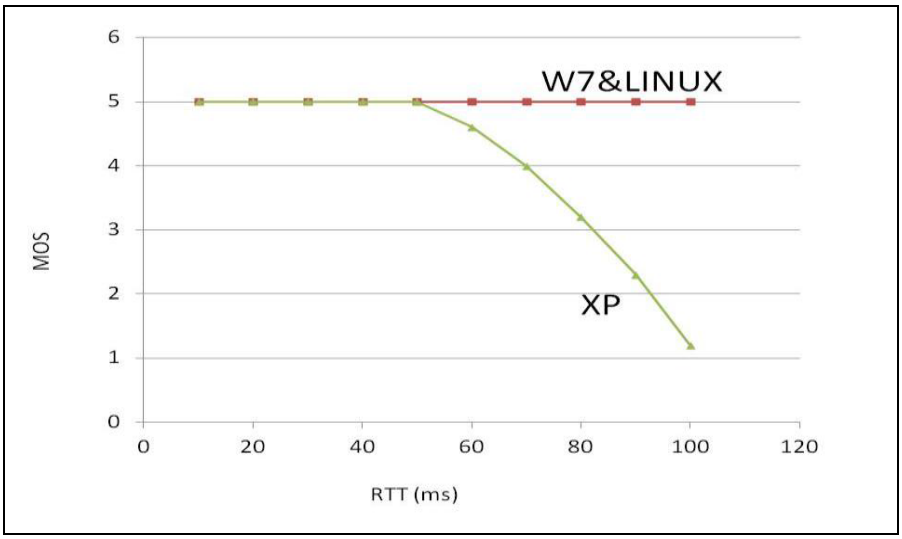


Figure 17: MOS vs RTT

To outline the differences between throughput and line capacity, QoS parameters measurements in the GPON access are carried out. Results are obtained profiling different access capacities at the ONU output ranging between 20 and 100 Mb/s always considering different Operating Systems (OSes).

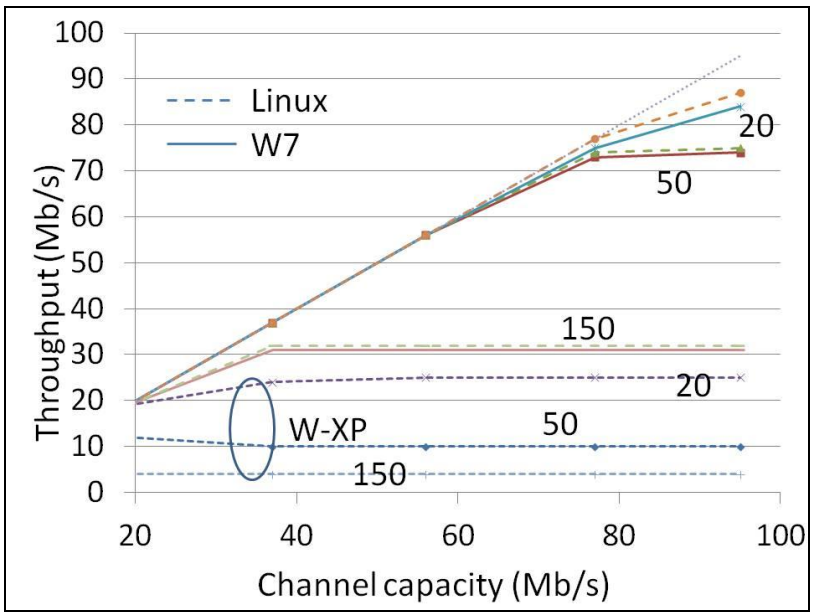


Figure 18: Throughput vs line capacity for Linux, W7, Windows XP for different RTT (20, 50, 150 ms).

In Figure 18 throughput versus line capacity for different RTT is reported. The difference is relevant for high RTT values and overall by considering Windows XP.

As already shown in [18] the TCP limitation could strongly degrade video services based HD streaming.

For example Figure 19 reports MOS versus throughput in Windows XP case (referred to results in Figure 18) for H264 video and for MPEG2 video. The RTT variation impacts the throughput with respect to the line capacity; in this case the line capacity is equal to 60 Mb/s. It is possible to see as RTT has a fundamental role in terms of QoE; let's further observe a weak better immunity of MPEG2 with respect to H264.

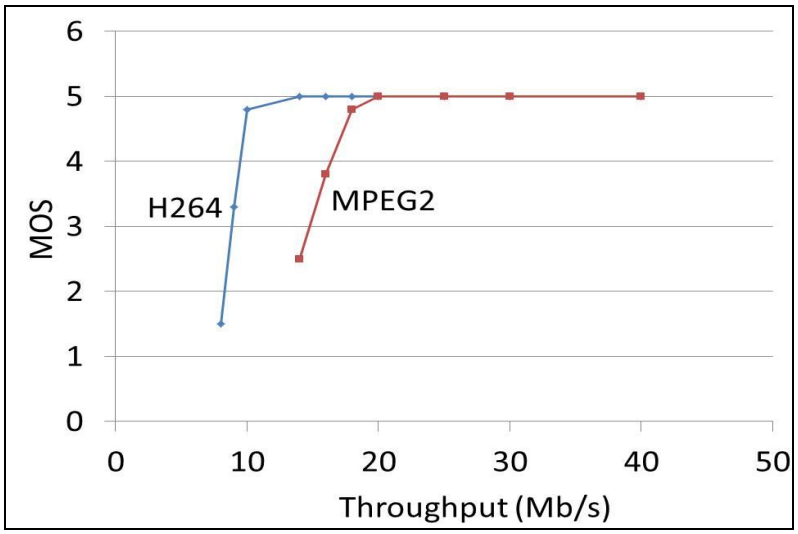


Figure 19: MOS vs throughput for W-XP in case of 60Mb/s line capacity for two different RTT; 50ms for H264 and 20ms for MPEG

Figure 18 also shows that XP can show strong throughput reduction for high RTT, which could induce degradation also for YouTube videos. For an example for RTT=150ms the XP throughput was equal to 4.7 Mb/s.

In the next paragraph, the results obtained by the proposed method based on multilevel measurement are reported.

In particular following results report the throughput behavior in GPON accesses distinguishing between TCP (a) and UDP (b) cases.

3.3.1 TCP case

Figure 20 reports the throughput measurement at time instant during a file transfer test that last more than 30s. Line capacity is set to 100 Mb/s. No additional delay (d=0) has been introduced in the network path between server and client, so RTT is equal to network delay. Linux is used as OS.

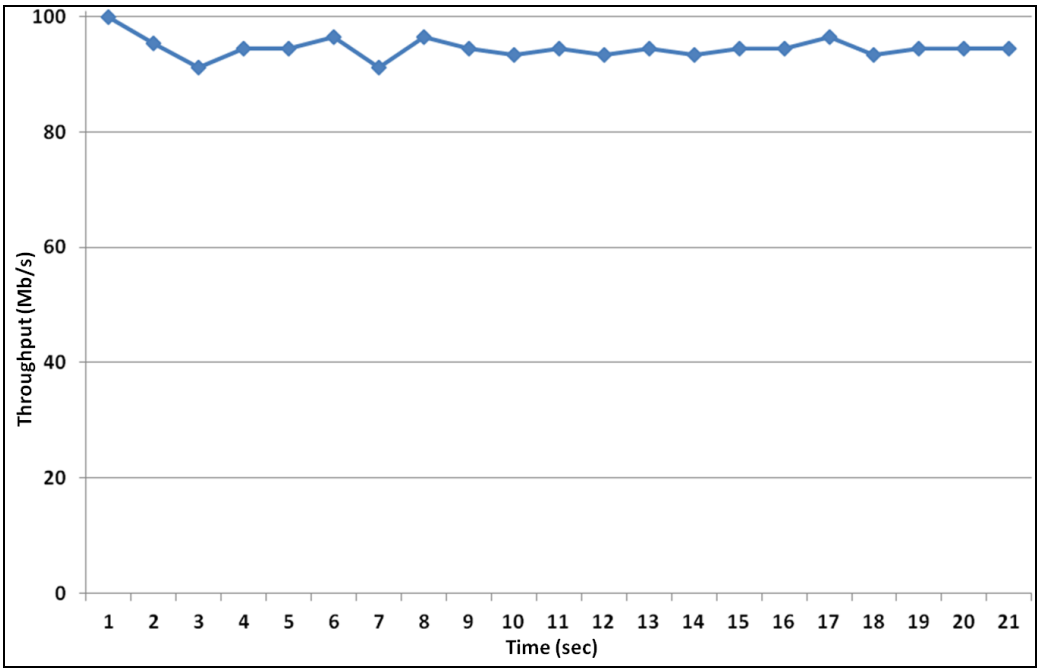


Figure 20: Throughput measurements over time; line capacity of 100 Mb/s and delay *d*=0

Figure 20 shows a modest fluctuating behavior due to the TCP congestion control algorithm: the TCP connection tries to exploit the maximum line capacity increasing the CWND; when the transmission bit rate overcomes the line capacity, packets may be lost, and TCP reacts by reducing CWND with a consequent rate decrease. This continuous increasing-decreasing behavior manifests with the small bandwidth fluctuations, with an average throughput equal to 94.5 Mb/s and a standard deviation of 2.3 Mb/s.

In Figure 21, time series of TCP throughput measurements in the case of RTT equal to 100 ms is reported.

This RTT can be considered as a typical delay between server and client in a continental environment. With respect to Figure 20 the behavior is totally different and in particular let's observes a typical transit behavior characterized by an initial fast growth of the throughput in time. This is due to the TCP slow start phase, when the congestion window grows exponentially, doubling every Round Trip Time until reach the maximum value given by Equation 1 as RWND/RTT.

This value is an arbitrary default value depending on the operating system implementation; its setting is critical because it influences the performance.

In fact if the threshold value is set too high relative to the network Bandwidth Delay Product, the exponential increase of congestion window generates many packets losses with subsequent significant reduction of the connection throughput.

On the other hand, if it is too small, TCP results in poor utilization especially when Bandwidth Delay Product is large. In Figure 21, the Operating Systems sets RWND=512 Kbyte that corresponds to a throughput of about 42 Mb/s for RTT=100ms.

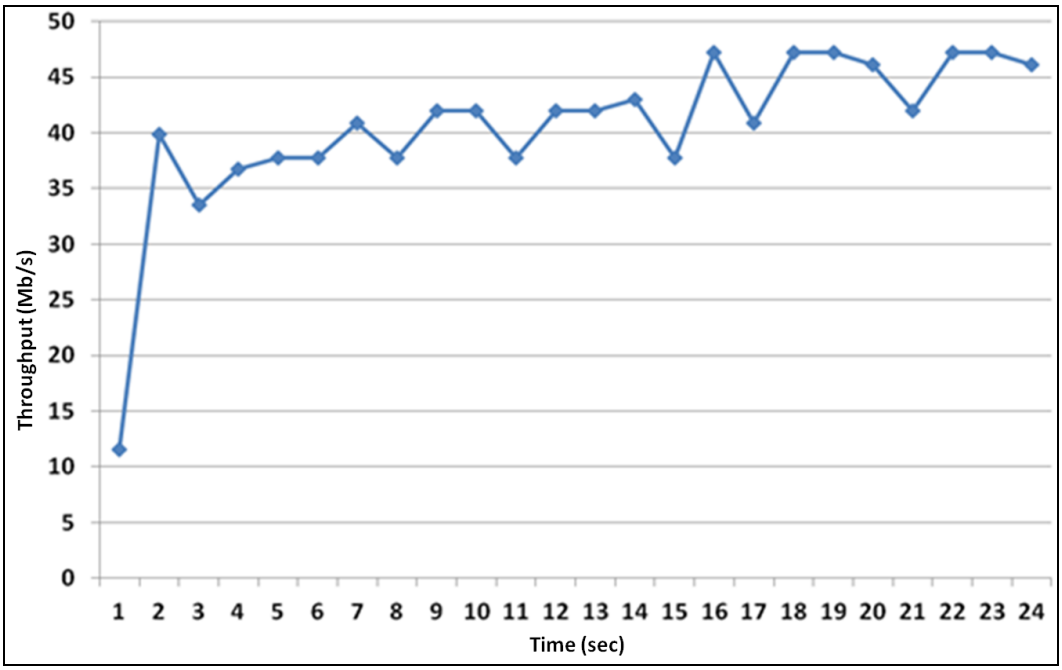


Figure 21: Throughput measurements over time; line capacity of 100 Mb/s; delay *d*=100 ms

After this transition period an almost flat behavior, as shown in Figure 19, is observed. Figure 21 clearly shows the difficulties of TCP in exploiting huge capacity as in case of fiber accesses in presence of large RTT.

This behavior is better summarized in Figure 16 where the throughput versus RTT for different OSes and for a line capacity of 100 Mb/s is reported.

This strong bandwidth reduction could be avoided by adopting UDP protocol. It is important to remember that UDP protocol does not have retransmission mechanisms so it performs very well in absence of network losses.

3.3.2 UDP case

UDP protocol does not implement any sliding window algorithm that limit the throughput; therefore, if the server-client transmission is the same as the line capacity, the UDP throughput is maximum and equal to the line capacity apart the overheads of the IP and lower layers encapsulation. This is shown in Figure 22 where the throughput versus time in the UDP case is reported. The server is imposed to transmit data at 95.6 Mb/s. No packet loss was thus observed at the receiver. These tests were performed with an additional delay between client and server equal to 100 ms.

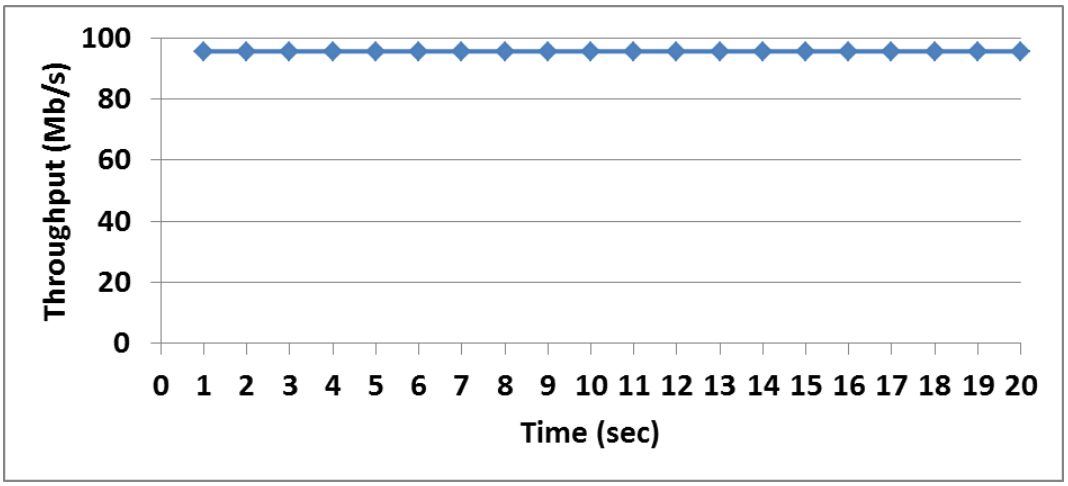


Figure 22: Time evolution of the UDP Throughput for a line capacity of 100 Mb/s and RTT=100 ms.

Since UDP does not implement congestion control mechanism, it is expected packet losses when the server transmission rate exceeds the bottleneck capacity.

This behavior is clearly shown in Figure 23 where the Packet loss (in percentage with respect to the total packets) versus server data transmission (Tx Rate) is reported. As expected, loss occurs as soon as the data transmission rate from the server exceeds 95.5 Mb/s. Figure 23 clearly shows that no packet loss is present if the server-client bit rate remains below 95.5 Mb/s.

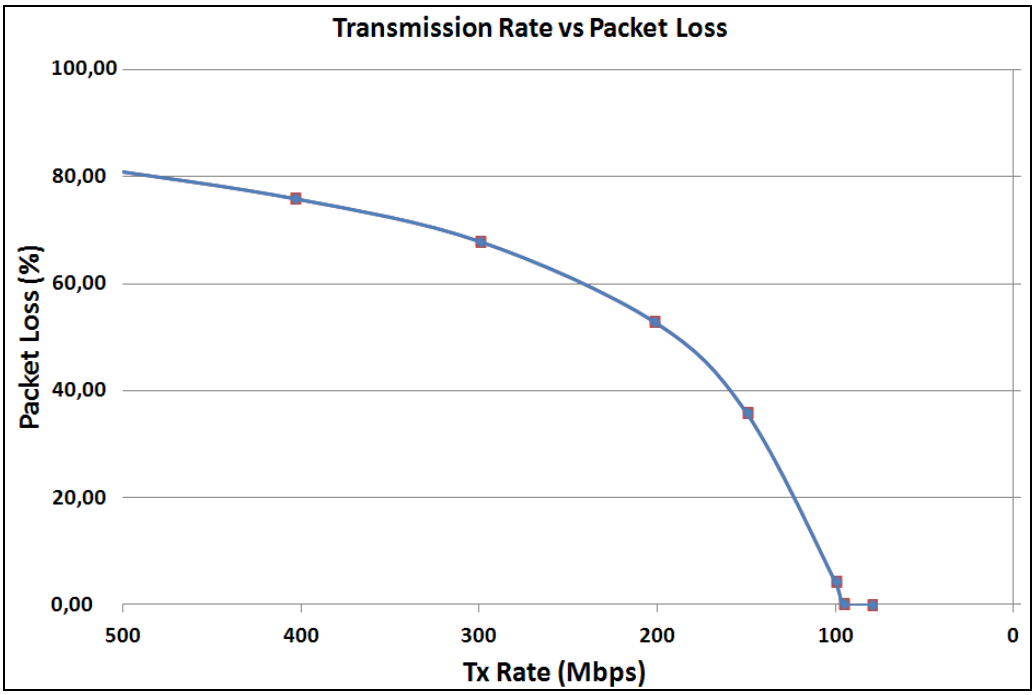


Figure 23: Packet loss (%) versus server data transmission (Tx Rate) in case UDP with RTT=100 ms.

Conversely when the bit rate is higher than the line capacity we observe packet losses corresponding to difference between line bandwidth and server transmission rate. No difference was observed among Windows XP, W7 and Linux as far as the data loss behavior was concerned.

The obtained results show that a multilevel measurement based on TCP and UDP protocol allows to calculate the goodput (bandwidth available to the user) and the line capacity (bandwidth provided by the ISP).

As described in the previous chapter the software agent performs a goodput measurement by adopting a TCP downloading. After, an estimation of the line capacity by means B_t and RTT measurements and equation (2), is calculated.

Subsequently a UDP stream is sent from the server with a line capacity of B_c', measuring the lost datagrams with respect to the transmitted datagrams. In such a way the useful transmitted bits BT_U are obtained and from the total transmitted bits BT_T, the estimated line capacity as $B_c = (BT_U/BT_T) * B_c'$ is achieved.

Finally the effective line capacity is measured by means of an UDP test based on a downloading of a file and forcing the transmission rate $B_c^* = B_c - \varepsilon$ (i.e. $\varepsilon = 0.001 * B_c$), verifying that a packet loss is lower than or equal to a predetermined value. It is important to note that the value of ε depends on the required reliability of the

measurement and in particular on the SLA between ISP and user. So, once the required packet loss is verified B_c* is the line capacity.

The transmission rate is a function of the line capacity and the value ε.

3.4 Multisession Transmission

The previous paragraph shows that when the RTT is large UDP could be more appealing than TCP, with the conditions to have a transmission bit rate lower than the line capacity. However, how to perform congestion control would require a much deep analysis, especially considering a noisy channel characterized by a packet loss that cannot be neglected.

However, no congestion control would be available, thus making the usage of UDP impractical, especially if we consider the case of a shared channel where packet losses can be caused by concurrent flows competing for the same capacity.

On the other hand, most services are based on TCP. Those would be strongly limited in case of large RTT, and some solutions have to be proposed to overcome the TCP limitations shown in the previous section to exploit all the physical layer capacity.

The proposal based on multisession transmission is here experimentally illustrated.

A multisession data connections means that a file is split in N files and each file is transmitted at the same time with a bit rate $\mathrm{B_m} = \mathrm{B_c} / \mathrm{N}$.

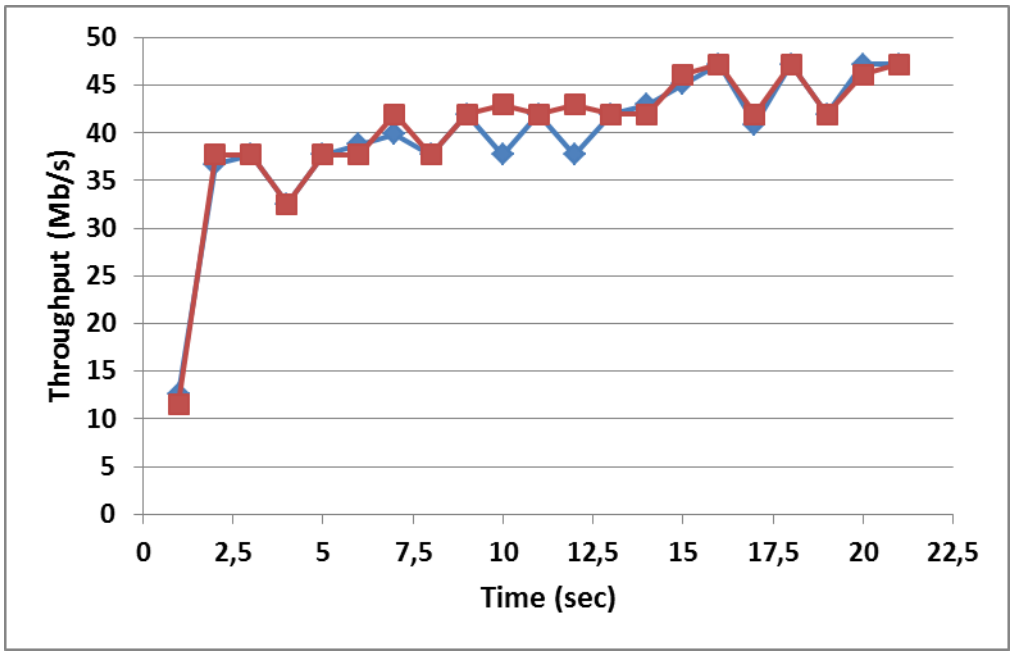


Figure 24: Time evolution of the TCP Throughput for a line capacity of 100 Mb/s and RTT=100 ms in case of 2 flows.

In Figure 24 the case of two flows is reported. Comparing with Figure 20 it can see that flows behave independently. After initial transient, the throughput reaches a value around 42 Mb/s for each flow. In this case the multisession permits to much better exploit the line capacity doubling the throughput, even if the sum of flow rates is lower than the line capacity.

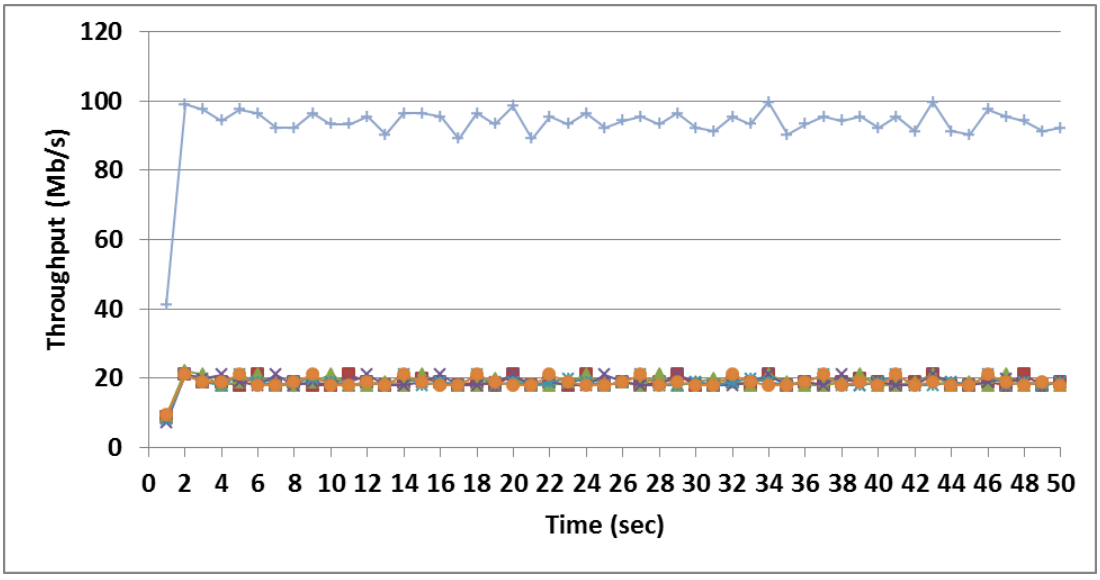


Figure 25: Evolution over time of TCP Throughput for a Multisession with 5 flows. The highest curve is the sum of the five flows. Line capacity of 100 Mb/s and RTT=100 ms.

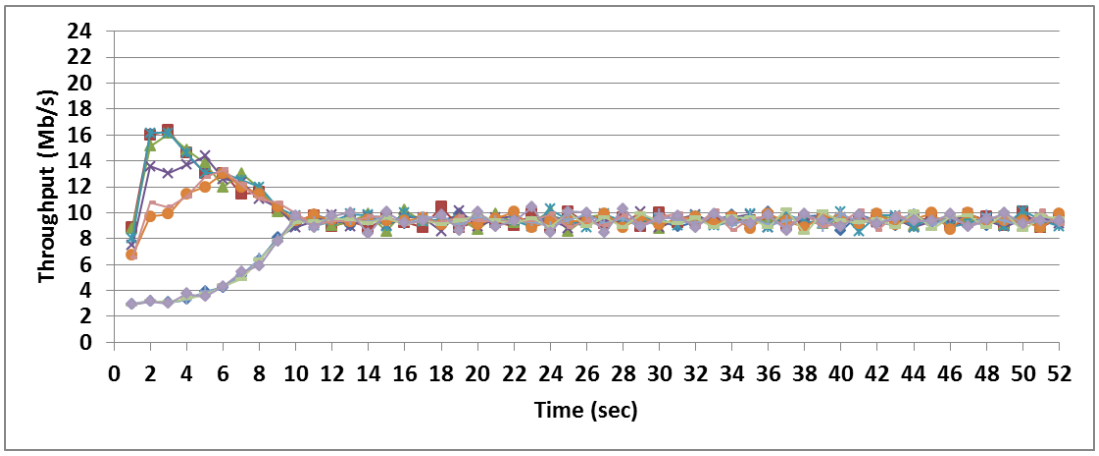


Figure 26: As in Figure 25 but for a Multisession with 10 flows.

In Figure 25 and Figure 26 two cases of multisession download using 5 TCP flows (each one obtaining approximately a transmission rate of 20 Mb/s, up) and 10 TCP flows (each one about 10 Mb/s, bottom) are reported. As shown in these figures, the total throughput coincides with the maximum throughput (95.5 Mb/s) for a line capacity of 100 Mb/s.

Results show how better to exploit the physical layer capacity by adopting multiple TCP connections to avoid the bottleneck of a single connection.

The measured throughput is lower than the line capacity because of the overhead introduced by layers of the network stack. As expected, the multisession download fully exploits the line capacity, therefore this results an important method to measure the line capacity in GPON networks by adopting the TCP protocol.

To conclude the huge capacity offered by the optical fiber accesses gives birth to new aspects regarding the measurement and the exploitation of the bandwidth.

In fact due to the intrinsic behavior of the Internet protocols, the effective bandwidth at disposal of the user can be much lower than the line capacity. Furthermore the evaluation of the Quality of Service and in particular the bandwidth measurement requires further insights, since we have to distinguish to which OSI Layer we refer to.

The question is if we wish to measure either the line capacity or the available bandwidth at disposal of the user, since this has important consequences on the Service Level Agreement verification, especially between customers and their Internet Service Provider (ISP).

This chapter reports an experimental investigation on the bandwidth measurement behavior in a GPON network, connected to a backbone segment, analyzing the impact of the Internet protocols. In particular results showed that the available bandwidth can be much smaller than the line capacity when TCP is adopted.

The proposed measurement method shows how UDP permits to fully exploit the GPON capacity. This approach allows measuring with few steps both the goodput and line capacity.

Finally results show how multisession approach exploits the line capacity in a TCP environment; this method can be also adopted to evaluate the line capacity.

Chapter 4
Mobile QoS Voice Evaluation

Mobile broadband networks support multiplay applications, such as voice, video, and data, on a single IP-based infrastructure.

The increasing popularity of wireless mobile networks indicates that wireless links will play an important role in future internet.

Wireless mobile environment is much more complex than the wireline one since much more information are needed to characterize service quality due to random behavior of the radio channel. It is well known that cellular systems have by nature limited resources: radio spectrum and transport (backhaul) resources are expensive and shared between users and services. To date both scientific community and network providers mainly focus their studies on issues regarding the design and management of radio mobile networks.

Besides the network problem there are some issues related to the user environment and in particular to the user terminals.

In fact with the widespread diffusion of the last generation devices exploiting new mobile network standards and several frequency bands, the QoS of mobile networks perceived by users could also be affected by the devices they are using.

Furthermore, each mobile device is characterized by the adopted frequency band and the installed operating system: these features could have effects on the QoS perceived by users.

The studies and results reported in this chapter aim to obtain experimental evidence of a given known yet vague and elusive in scientific terms: the quality of a mobile service depends not only from the mobile network, even by the user terminal that is used and from any service provider.

In order to consider mainly the aspect related to the user terminal, it was decided to focus on the voice service that is the most simple and consolidated service. The influence of the terminal will be shown up.

Terminals performance are measured in terms of the two most important parameters to characterize a session voice; the Setup Failure Rate (SFR), which is the probability that the establishment of the call is unsuccessful, and the Drop Call Rate (DCR), which is the probability that the call is suffering unnecessary shutdowns.

It is important to note that SFR and DCR are two basic key performance indicators (KPI) usually adopted by network providers to assess the performance of their networks. These KPIs have been chosen also considering the quality of voice service parameters proposed by the European Telecommunications Standards Institute in the ETSI EG 202 057-2 V1.3.1 [43] guide. These parameters allow us to perform objective and comparable measures of the QoS delivered to users.

On the basis of this consideration, this study aims to verify the dependence of the QoS, perceived by end-users of mobile voice services, from the specific cellular device. This goal is achieved testing several models of mobile phone in the same network conditions, time of day and location by “drive-test”. The strength of this activity is the implementation of field measurements instead of laboratory simulation.

4.1 Measurement Methodology

The main goal of this activity is the experimental evaluation of the performance of some models of mobile phones with respect to the voice service through outdoor drive-tests. Thanks to the measurement campaign each tested mobile phone model is classified in terms of quality categories.

As mentioned above the KPIs selected for the measurement activity are the Setup Failure Rate (SFR) and the Drop Call Rate (DCR).

The SFR test call refers to mobile terminated calls considered as unsuccessful if not established within 30 seconds.

Let p be a cellular phone, the SFR is defined:

$$SFR(p) = \frac{|unsuccessfulSfrTestCall(p)|}{|SfrTestCalls(p)|}$$

where at the numerator and at the denominator are reported respectively the number of unsuccessful SFR test calls to p, and the total number of SFR test calls to

p. The SFR(p) value is calculated on the basis of a sequence of SFR test calls started every 90 seconds, as depicted in Figure 27. The duration of a SFR test call (T_{SFR}) is set to 30 seconds.

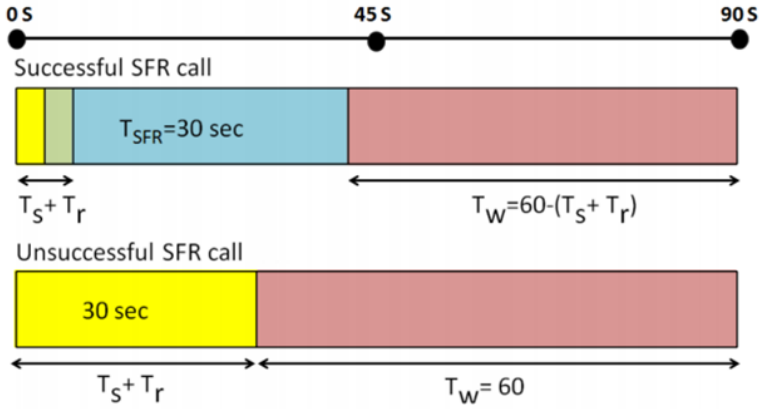


Figure 27: SFR test call diagram. A SFR test call is successful if the sum of Setup Time (T_s) and Response Time (T_r) is less than 30 seconds. T_w is the Waiting Time between two consecutive test calls.

Similarly, DCR test call refers to a mobile-terminated call considered as unsuccessful if, once the call is set-up and started, it lasts less than 120 seconds.

Given a mobile phone p, the DCR is defined:

$$DCR(p) = \frac{|unsuccessfulDcrTestCall(p)|}{|DcrTestCalls(p)|}$$

where at the numerator and at the denominator, the number of unsuccessful DCR test calls to p, and the total number of successfully set-up and started DCR test calls to p are shown.

The DCR(p) value is calculated on the basis of a sequence of DCR test calls started with a period of 180 seconds, as reported in Figure 28.

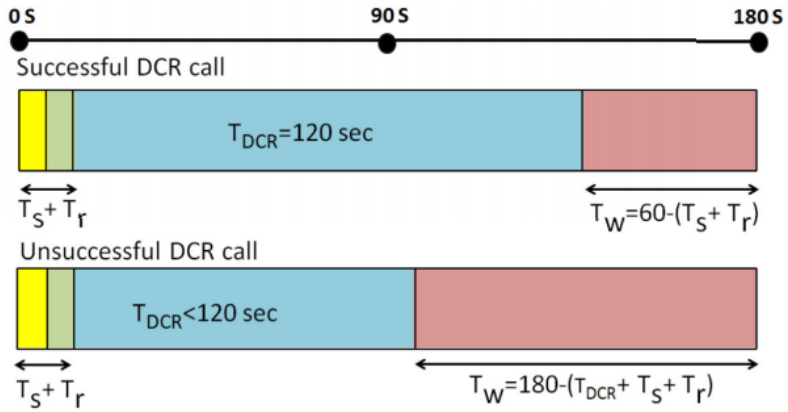


Figure 28: DCR test call diagram. A DCR test call is successful if its duration equals 120 seconds.

It is worth noting that an SFR test call could be derived in principle from a DCR test call.

The CSR parameter quantifies the rate of calls that were started and terminated correctly, that are calls which had neither accessibility nor retainability problems. More precisely, CSR is defined as:

$$CSR = (1 - SFR) * (1 - DCR)$$

In the measurement campaign Test Sessions were implemented.

In particular a Test Session is composed of one Model Test Session for each mobile phone model *m* involved in the test. Given a mobile phone model *m* and *n* network providers, we denote p_j^i the i − th device of the *m* model using the j − th network provider.

The MTS (p_1^1, ..., $p_n^{|P|}$) (i.e. Model Test Session for *m*) is defined as a fixed length sequence of SFR and DCR test calls started at the same time, to p_1^1, ..., $p_n^{|P|}$.

Let M be the set of cellular models and T the set of mobile terminals considered, the function F: T → M such as F(t) = m, if the terminal $t \in T$ is a device of the model $m \in M$ is defined. The Setup Failure Rate for a single model is defined as follow:

$$SFR_{model}(m) = \sum_{P:=\{p:F(p)=m\}} \frac{SFR(p)}{|P|}$$

Model Test Sessions are scheduled to be started separately, with a random delay between 1 and 3 seconds to avoid network overload. The starting order of the Model Test Session is random as well.

4.2 Measurement Campaign Settings

Measurement campaign is conducted adopting the following settings:

- the sequence length of Model Test Sessions is fixed to 10 (|P|=10). At the beginning of the campaign, the Model Test Session to perform 50% of both DCR and SFR test calls is set. After about a week, the percentage of DCR test calls to 70% was implemented on the basis of an initial analysis of statistical significance of collected samples;
- network providers involved in the measurement activity are two (n = 2);

- 7 smartphones and 3 traditional phone models (|M| = 10) were chosen considering the most popular ones on the market at that moment and taking into account different electronics devices. The list of selected phones is reported in Table 2. For each phone the automatic answering function was set and the voicemail disabled.
- 2 vehicles each one carrying 5 phone models (2 devices per model, one for network providers) driving one after the other. The position of each device was periodically shifted inside the vehicle in order not to favor one over the other terminal.

Brand	Model	Operating System	Class
Apple	iPhone 4	iOS	Smartphone
Nokia	N8	Symbian	Smartphone
Nokia	5230	Symbian	Smartphone
RIM	Torch 9800	RIM	Smartphone
HTC	Wildfire	Android	Smartphone
Samsung	Galaxy S2	Android	Smartphone
LG	Optimus 7	Windows Mobile	Smartphone
Samsung	C3050	-	Traditional
Nokia	2720	-	Traditional
Nokia	C1-01	-	Traditional

Table 2: List of the adopted mobile models

The campaign has been conducted in late 2011, from 3 November till 12 December, for a total of 34 calendar measurement days, weekdays and holidays included; each measurement day was made of two separate sessions approximately from 8:00 to 13:00 and from 14:00 to 18:00 respectively. About 7500 mobile-terminated test calls have been performed for each mobile phone. We also took into account traffic variations of mobile networks during the time of the day and the day of the week. In order to obtain an adequate characterization of the performance of the terminals, the tests were carried out in urban (75%) and extra-urban (25%) driving scenarios. As for the urban scenario, 17 cities were selected among the most populated ones in Italy in order to take measurements all over the national territory. The extra-urban scenario includes also major Italian highways, driven while moving from a city to another. For major detail of paths and measurement scheduling see [44].

4.3 The Measurement Platform: a Logical Architecture Overview

The measurement platform consists of all the necessary hardware and software to implement the measurement procedures described in the previous paragraph. It is composed by five logical components, as shown in Figure 29:

- a telephone Private Branch exchange (PBX), with the goal to perform phone calls towards all mobile phones involved in the measurement campaign;
- the Measurement Database, storing results for each phone call;
- the Measuring Session Management (Controller), to manage Test Sessions from remote using a mobile phone. This component acts as a controller capable to configure, start and stop test calls;
- the Measurement Software (MS), having the goal to properly synchronize phone calls requests to be sent to the PBX, catch and manage events generated by the PBX and store measurement results on the Measurement Database;
- a Web Server hosting a Web service used by the MSM to communicate with the MS;

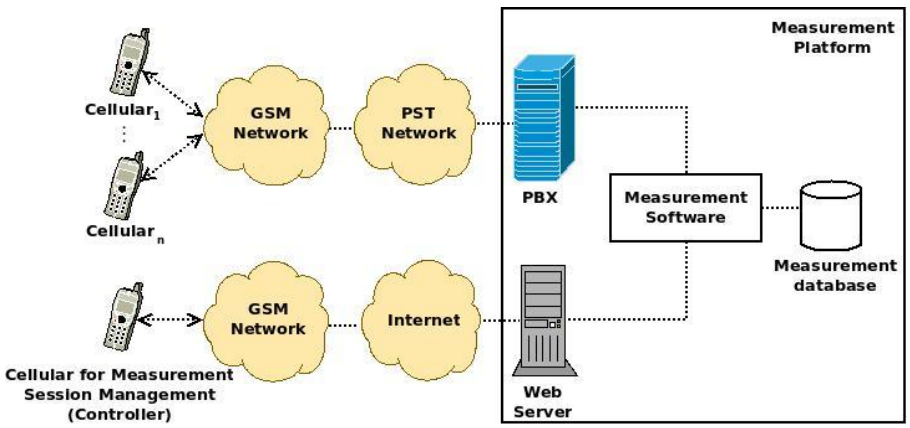


Figure 29: Logical architecture of the measurement platform

4.4 Preliminary Statistical Results

The collected samples were statistically processed. At first a validation process for each measured sample and for each phone model, with the purpose of excluding

those results deriving from possible system problems (e.g. the terminal was off for operational reasons) was performed.

4.4.1 Terminal statistics characterization

For each model SFR, DCR and CSR values, were calculated; the results are reported in Table 3, ordered according to their CSR value. Let us recall that CSR values can be derived once SFR and DCR values are known.

Model	Class	SFR(%)	DCR(%)	CSR(%)
T2	Traditional	1.44	0.65	97.93
S5	Smartphone	0.78	1.60	97.72
T3	Traditional	1.67	0.71	97.65
T1	Traditional	1.62	1.05	97.40
S2	Smartphone	1.63	1.80	96.67
S1	Smartphone	2.72	0.83	96.48
S4	Smartphone	2.77	0.96	96.32
S3	Smartphone	2.83	1.02	96.18
S6	Smartphone	3.21	0.84	95.98
S7	Smartphone	7.49	11.47	82.99

Table 3: SFR, DCR and CSR values

As we are facing dichotomous data (1 in case of successful test, 0 for negative test, for SFR as well as DCR samples), the DCR and SFR quality parameters consist of proportions and are distributed according to a binomial distribution, once the hypothesis that single events are independent from each other holds.

Therefore, from a statistical viewpoint, the values in Table 3 are point estimators of the most relevant statistical parameter of these distributions, namely the mean value.

In contrast to point estimators, we can calculate confidence intervals [45], describing the interval that covers the true parameter value with a certain probability; for instance, when applied to the values in Table 3, a confidence interval represents the interval in which the mean of the underlying distribution lies with a probability, let us say, of 95 percent.

The calculation of the confidence intervals heavily depends on the assumed kind of distribution that in this case is a binomial distribution.

In Figure 30 and Figure 31, the values in Table 3, together with their confidence intervals with confidence level of 95% (5% α-level), are reported.

Let's see that all models share quite variable performance: their SFR and DCR mean values are ranging over wide intervals. It is important to note that each value is calculated over all the valid samples relative to that model.

It is also worth noting that this performance spreading has to be related just to terminal behavior, as all of the other crucial aspects (network, propagation, etc…) are identical for all of the considered devices.

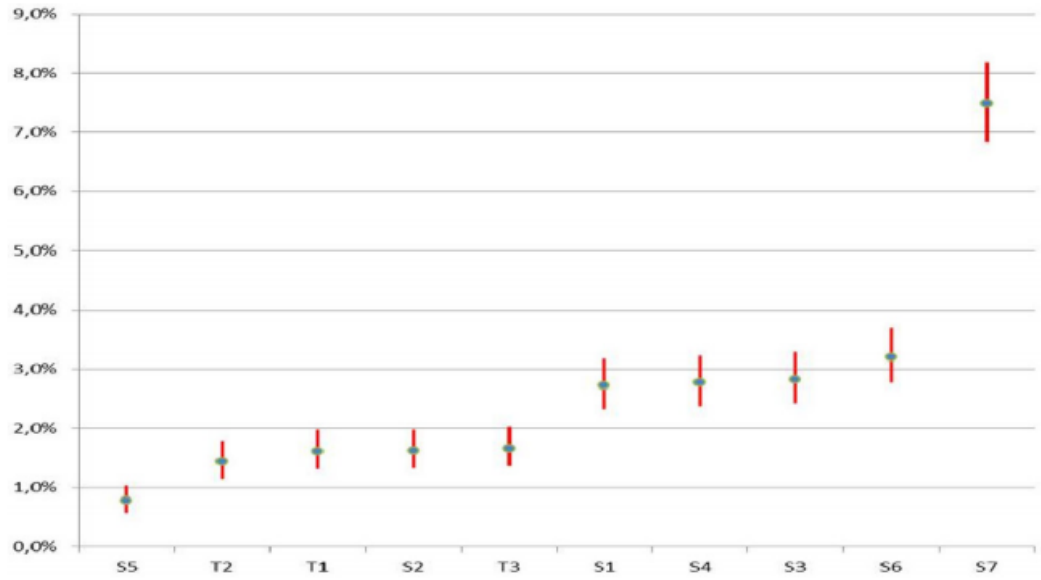


Figure 30: SFR with their 95% confidence interval values

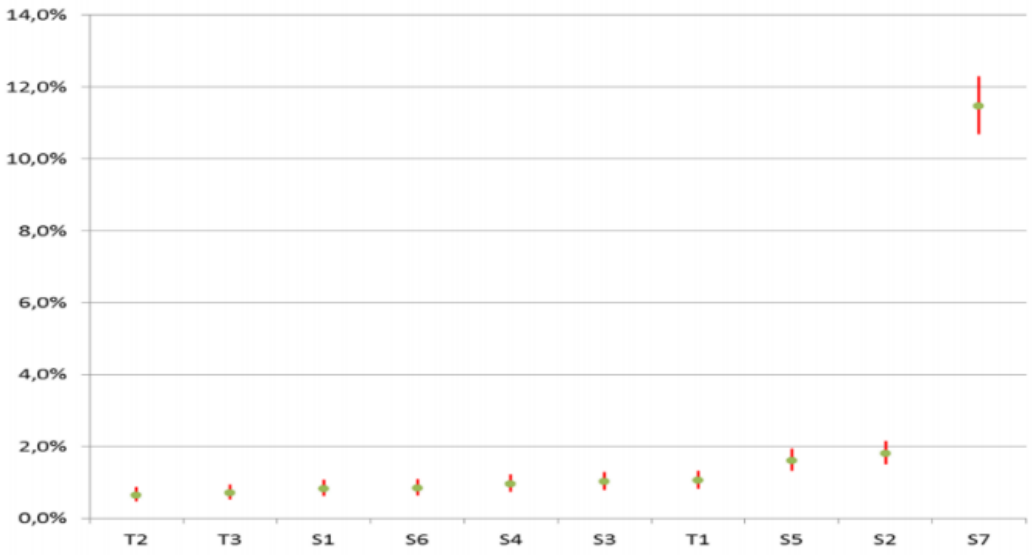


Figure 31: DCR with their 95% confidence interval values

4.4.2 Basic concepts of inferential statistics

As mentioned above, the results presented in Table 3 have been independently calculated for each device. Therefore, these results could provide us with useful hints about the quality ranking among them but they do not still allow us to draw any conclusion about possible statistically sound differences, which would be a better result. In order to reach the latter goal, is necessary to make statistical comparisons among the collected samples and infer from the experimental results which one of the following two mutual excluding hypotheses is true:

- *Null Hypothesis* H_0: cellular performance samples belong to the same statistical distribution; the measured differences are originated by chance;
- *Alternative Hypothesis* H_1: cellular performance samples belong to different statistical distributions; the measured differences are originated by a different value of some statistical parameter (e.g. the mean value of their binomial distribution).

Hypothesis testing can be employed also to draw inferences about one or more populations from given samples (results).

When applying a statistical test procedure, it is possible to make two different kinds of errors, as it may happen in criminal trials. In a criminal procedure, the innocent can falsely be declared guilty or the guilty can wrongly be judged innocent. In the same way, if the null hypothesis H_0 is a true statement about reality but is rejected in favor of the alternative hypothesis H_1, a mistake similar to convicting the innocent has occurred. Such a mistake is known in statistical terms as a false positive or Type I error. For sake of completeness, when the alternative hypothesis H1 is true but is rejected in favor of H_0, the mistake is similar to absolving the guilty. This mistake is known as a false negative or Type II error.

Given the above formulation of H_0 and H_1, a Type I error occurs whenever one decides that the terminals have different performance, when in reality their performance belong to the same distribution. It is worth noting that a consideration of this sort concerns the underlying statistical populations and not the observed sample data. Thus, the disparity among terminals must be great enough to decide with adequate confidence that the underlying populations also differ. A proper

statistical test must account not just for the difference in observed mean levels but also for variability in the data likely to be present in the underlying statistical populations.

The significance level of the test is equivalent to the false positive rate α.

A test conducted at the α = 0.05 level of significance means there is at most a 5% chance or probability that a Type I error will occur in the results. The test is likely to lead to a false rejection of the null hypothesis at most about 5 out of every 100 times the same test is performed.

In order to test which hypothesis is true, we can apply a number of statistical inference methodologies; as the measurement campaign is performed simultaneously (the same time, the same place, the same network) for the 10 device models, the relevant method, namely the "dependent samples" comparison, must be applied.

A second important choice has to be made between parametric or non-parametric methods: the former are more powerful (and should then be preferred), but three conditions must be fulfilled in order to be correctly adopted:

- the samples are independent. The behavior of one device should not influence the behavior of the other ones;
- the samples follow a Gaussian distribution;
- the sample distributions are homoscedastic, which is with the same variance.

The first hypothesis is verified at a good extent, as possible interferences are prevented by keeping a suitable distance among the devices within the cars and by adopting an appropriate measurement procedure (rotating positions, permuted start of the calls, and so on). Also the second condition can be considered verified, as binomial distributions approximate a Gaussian distribution when the number of samples is large enough; moreover, we could be highly tolerant with this hypothesis, will maintaining. The third hypothesis is the more stringent one, in order to safely apply parametric methods, and it is the most critical in this scenario. As a matter of fact, in case of binomial distribution with parameter p, the corresponding variance is equal to *p(1 − p)*; if the comparison regards two phone models having different mean

value *p* (their SFR or DCR values), the associated variances will be correspondingly different and hence the homoscedasticity hypothesis will fail.

Therefore, non-parametric methods, namely the McNemar test [46, 47], in case of comparison between two binomial dependent samples, and the Cochran test [48], in the general case of comparisons among multiple binomial dependent samples, were adopted.

4.4.3 Non-parametric inference methods for dependent samples

Since we have dependent samples, the non-parametric Cochran's Q-method, which is particularly suited for multiple dependent samples, was applied. The two hypotheses to test are:

- null-hypothesis H_0: models obey to the same binomial distribution

$$\pi_i = \pi \quad \forall i, i = 1,\ldots,N$$

  where N=10 is the number of compared phone models;

- alternative hypothesis H_1: at least one of the considered models obeys to a binomial distribution with a mean value different from the others

$$\exists i, i \in \{1,\ldots,N\} : \quad \pi_i \neq \pi$$

In order to apply the Cochran's Q test [48], all device models are compared in the same conditions (in terms of time, place, and network); therefore these 10 dependent samples were aligned in a 10-value row (a block, corresponding to a Test Session); multiple blocks, corresponding to multiple measurement instances gathered along the campaign, provide us with a 10-column matrix (a column for each device model) and as many row as the total number of sessions. It is worth noting that an invalid result for any of the 10 models causes the entire block to be considered invalid.

Once the Cochran's Q test was applied and the null-hypothesis was rejected, we are not aware about which one (or more than one) of the 10 models has a distinct statistical distribution: to address this problem the Unplanned Multiple Comparison Procedures (UMCP) was applied. In this case (absence of an a-priori control group to check against the others), UMCP consists of all the possible N × N paired comparisons; in this scenario of dichotomous samples, each one of these paired comparisons is a McNemar test between the two considered devices [46, 47].

Therefore, the whole family will be composed of 45 McNemar tests *($\frac{N(N-1)}{2}$, N=10)*, each one with an associated p-value. These p-values reflect the error probability for a certain comparison, ignoring the remaining comparisons.

If in each comparison the level of significance is α, then the probability of rejecting a true null hypothesis is *(1 − α)*, and the probability of not making an error in the whole set of comparisons (family) is *(1 − α)* raised to a power equal to the number of comparisons: in our case, $(1 - 0.05)^{45} = 0.01$, a very poor result. To deal with this problem, we resort to adjusted p-values (APVs) methodologies, as they can take into account the accumulated family error and can be directly compared with any chosen significance level α. Among all the possible different choices, the Finner procedure was adopted [49], in order to obtain a family-wise level of significance, α_{tot}, equal to 5%, without requiring a too low significance level α for the single comparison.

4.4.4 List of statistical operations

The measurement campaign has been designed and tailored to obtain the appropriate sample size for the correct inferential statistical answers. We can apply the procedure, described here below, independently and likewise for SFR and DCR key performance parameters. The operations can be summarized as follows:

1. Apply Cochrans Q-method to check for significant statistical differences among terminals. If no statistical differences exist among the terminals, stop the operations; otherwise, go to step 2.
2. Set the "family-wise" significance level equal to 5% ($\alpha_{tot} = 0.05$).
3. Order in ascending order the proportions (whether SFR or DCR) corresponding to the different device models.
4. For all proportions, associate the corresponding terminal, then construct a matrix where the first line (and also the first column) is associated with the terminal which has the minimum value of SFR/DCR, the second row/column with that resulting in the second position and so on.
5. Operate pair-wise comparisons, using McNemar's test (see step 5.a), between any pair of the 10 terminals; populate the square matrix (see step 5.c), described in step 4: set the row and column associated to the two device models of the considered pair to the result of the following step 5.b.

a) McNemar test: for the two generic terminals A and B making up the pair, observe the inconsistencies (that is, when A is 0 and B is 1 or vice versa) 1-0 (X) and 0-1 (Y);
b) find the parameter d^2 as $d^2 = (X - Y)^2/(X + Y)$. For each pair, calculate the value d^2. As d^2 is χ^2 distributed with one degree of freedom, the corresponding p-value for that single comparison (the probability corresponding to that d^2 value) was obtained.
c) Fill the square matrix with the 45 p-values obtained from step 5.b.

6. Order the 45 p-value, keeping the associated pair of terminals with their ranking position.
7. Apply the Finner procedure: if a certain p-value occupies the i-th position in the ranking, it will be compared with the level of significance α_i equal to $\alpha_i = \alpha_{tot}/i$ (Finner value).
8. Determine the j-th critical position, for which the p-value exceeds the Finner value:
9. Consider all the pair of terminals corresponding to p values less than j (i.e. from the first position to the (j−1)-th) significantly different from a statistical viewpoint; vice versa, consider all the pairs of terminals corresponding to p-values from the position j−th to the last one (the 45-th) as statistically indistinguishable (they belong to the same distribution, from an inferential statistics viewpoint).
10. Identify the group of terminals with statistically similar/different quality.

4.5 Relevant Statistical Results

At the end of this analysis, groups of models having different quality characteristics are identified, whereas the individual models within each group are qualitatively identical from a statistical point of view. This conclusion can be drawn with a “family-wise” significance level equal to 5%, as assumed in the second item of the precedent paragraph. The results of this analysis highlight how the quality of service of a voice call depends on the phone model. In this work only the performance difference of

voice service in case of traditional phones and smartphones are reported; for all results see [44].

Class	Average SFR(%)	Average DCR(%)	Average CSR(%)
Traditional	1.6	0.8	97.7
Smartphone	3.1	2.5	94.6

Table 4: Comparison between smartphone and traditional devices

The results shown in Table 4 make evident the differences existing between Traditional phones and Smartphones. However, this problem can tackle from an inferential statistics viewpoint: are these mean differences to be ascribed to chance or traditional phones and smartphones belong to two different distributions?

So, two groups of terminals corresponding to two fictitious models are defined: model T, with the three traditional phone models T1, T2 and T3; model S, with the seven smartphone models S1…S7. As just two models are considered, it is possible to apply McNemar test to model T and model S, considering how many times model T works well (1) whereas model S fails (0) (namely X, according to step 5.a) against how many times model T performs bad (0) whereas model S performs well (1) (namely Y, according to step 5.a). As already mentioned from a practical point of view, we have to compare T1 (and then T2 and then T3) with S1 (and then S2, till S7), summing up the results from each comparison in the grand total X or Y.

When we end up with the application of the McNemar test to the traditional phones and smartphones group, we obtain that our previous (intuitive) conclusion (traditional phones perform better than smartphones, in terms of both SFR and DCR) receive a strong confirm to a very high level of statistical significance. This procedure can be repeated with a subset of the considered smartphone models (e.g. those smartphones having best performance in terms of SFR or DCR); the conclusion (for both SFR and DCR) keeps valid also with subsets of four (or even three) smartphones, with significance level lower than 5% (and even lower than 1%).

These results show how traditional mobile devices have higher global performance than smartphones, when comparing their voice quality performance.

On the other hand, it is worth noting that smartphones provide end-user with richer experience, if compared to the basic voice service, through their full-featured apps and Internet connectivity.

Traditional phones have limited capabilities with respect to smartphones, but nevertheless they cover that limited field of operations with a quality higher than smart devices. It seems that the higher processing power of smartphones is not adequately accompanied by superior characteristics of their Radio Frequency circuitry or basic processing capabilities.

In addition, terminal performance have been calculated according to different driving scenario (extraurban and urban scenario), vehicle speed, time of the day and part of the week (weekdays and weekend).

To conclude the obtained results, submitted to the journal, show that the causes of different quality degradation can only be attributed to the device itself and not to other external factors (e.g. to network, radio or environmental factors).

Chapter 5
Mobile Broadband QoS Evaluation

5.1 Wireless Mobile Environment

Wireless mobile environment poses formidable challenges when attempting to provide reliable, end-to-end data transmission for transport protocols such as TCP.
Despite all the several feature described in the previous chapters, reliable transport protocols such as TCP are tuned to perform well in traditional networks where packet losses occur mostly because of congestion. However, networks with wireless and other lossy links also suffer from significant losses due to bit errors and handoffs [50]. As mentioned in the Chapter 2, TCP responds to all losses by invoking congestion control and avoidance algorithms, bringing to degraded end-to-end performance in wireless and lossy systems.
Packet losses and unusual delays are assumed as the primary cause of congestion in the network. TCP performs well over such networks by adapting to end-to-end delays and congestion losses. The TCP sender uses the received cumulative acknowledgments to determine which packets have reached the receiver, and provides reliability by retransmitting lost packets. In order to properly evaluate the acknowledgements arrival period, the sender estimates a running average of round-trip delay and the mean linear deviation from it. The sender identifies the loss of a packet either by the arrival of several duplicate cumulative acknowledgments or the absence of an acknowledgment for the packet within a timeout interval equal to the sum of the smoothed round-trip delay and four times its mean deviation.
The packet loss events are managed by the TCP dropping its transmission (congestion) window size before retransmitting packets, entering congestion control or avoidance mechanisms (e.g., slow start) and backing off its retransmission timer based on the Karn's Algorithm [51]. These mechanisms lead to a reduction in the load on the intermediate links, thereby controlling the congestion in the network.

Unfortunately, when packets are lost in networks for reasons other than congestion, these measures result in sub-optimal performance, due to a useless reduction in end-to-end throughput. Communication over wireless links is often characterized by sporadic high bit-error rates, and intermittent connectivity due to handoffs. TCP performance in such networks suffers from relevant throughput degradation and very high delays for interactive services [52].

Starting from previous studies and results reported in preceding chapters, in this paragraph a new measurement approach to characterize a mobile broadband access together with QoS parameters is proposed. This novel approach brings to an effective characterization and tuning of measurement test in mobile environment. Such tuning was carried out to implement and set up a real measurement campaign.

Therefore the following list reports a measurements suite useful to characterize QoS mobile broadband access. This suite performs a multilevel measurement calculating the goodput (layer 7 measurement based on HTTP) and the throughput (layer 4 measurement based on TCP):

a) HTTP and TCP download-upload throughput. The throughput is defined as the ratio between the dimension file and the time to transmission accordingly with wireline network scenario with TCP approach. The transmission duration is the period that starts from the instant when the wireless network has received all the information necessary to start the download/upload transmission up to the last bit of the file test is received.
b) Web browsing. The web browsing quality is generally associated with the time it takes to locate and download a web page within a web browser application.
c) HTTP and TCP download-upload failures. Percentage among data transmission failure and the transmission data attempts. Failure can occur when the test file is either not totally transmitted or without error within a fixed time-out.
d) Packet loss measured by Ping (Packet Internet Grouper) technique. Data packets may be discarded due to traffic congestion. This test records the average packet loss that occurs as a percentage of total data sent.

e) Transmission delay in terms of Round Trip Time. It is obtained by means of Echo Request/Reply (Ping) in agreement with protocol ICMP (RFC 792: "Internet Control Message Protocol")
f) Delay variation (Jitter).

5.2 Network Architecture

The QoS measurement architectures carried out in the framework of the AGCOM activities, DELIBERA N. 154/12/CONS [53], in order to monitor the internet accesses, with particular interest for SLA certification have illustrated some characteristics in terms of data treatment and mining. First of all the measurement architecture is based on file download and therefore particular attention has to be given to server sizes that have to be able to simultaneously answer to user requests.

The measurement system boundaries used for Internet access service from mobile station consist, on the one hand, from the user terminal (UE: User Equipment) and the other from its point of connection to the Internet.

Mobile radio networks of different operators are related to the external packet networks, i.e. essentially to the Internet, which exchange information flows through nodes called GGSN (Gateway GPRS Support Node).

So one solution was to install measurement server directly in correspondence of these nodes, but it should be notes that these nodes are in effect an integral part of the network of the operators, and then they belong to.

Moreover, the total number of GGSN nodes to take into consideration the national territory could not be insignificant, with a consequent management and costs increase.

For the stated reasons, it was decided to place the measurement server at the Neutral Access Point (NAP).

Each GGSN node, in fact, routes data traffic to the Internet, using a network provider infrastructure connecting them to the NAP. Therefore in the next figure the structure of the measuring system is outlined. Figure 32 shows that all the traffic exchanged between the NAP and the user terminals transits exclusively on the network provider under test.

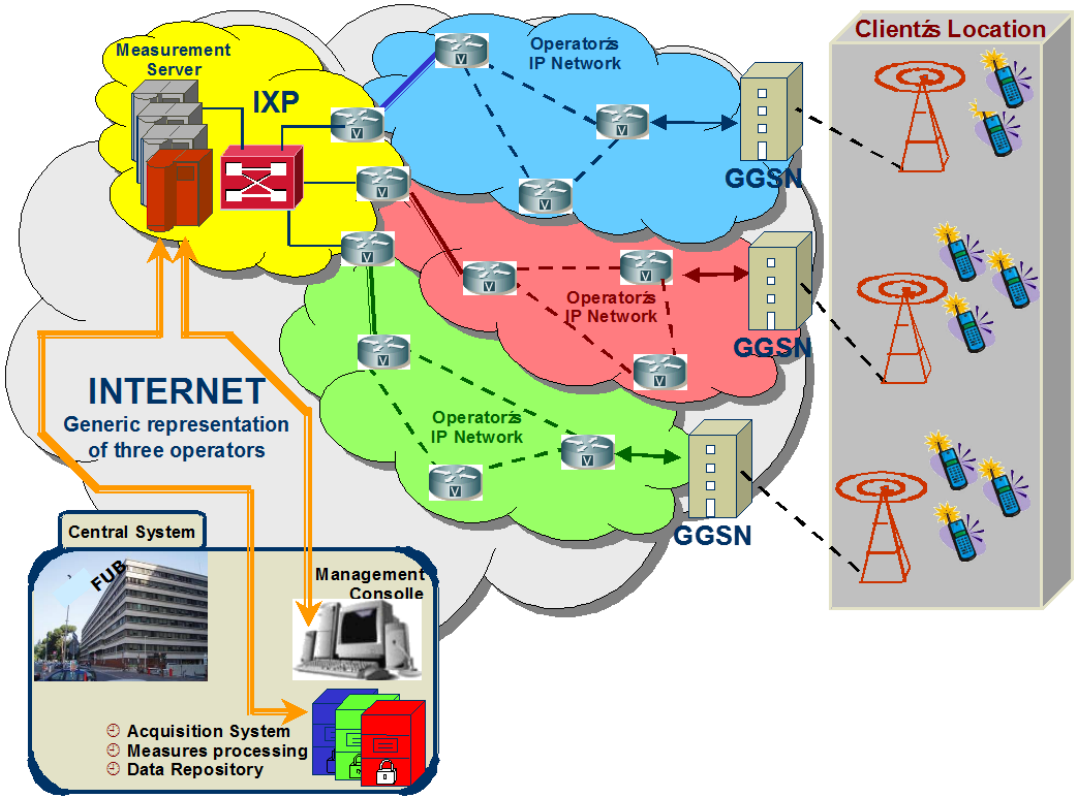


Figure 32: Measurement System Architecture

Within the Internet Exchange Point (IXP) or (NAP), measurement server can be directly connected on the peering LAN, providing a direct link between the measurements devices and mobile networks under test. Therefore tests occurs exclusively on the networks (or network segments) of operators property.

In this way the quality and performance evaluation is referred only to the specific mobile network.

The actual network configuration for the measurement campaigns is based on two measurement server positioned at the NAP of Milan (MIX) and Rome (NAMEX). This location for the primary motivation of network efficiency (in the sense of minimizing, or at least standardize the distance between test points, geographically dispersed throughout the country, and NAP) for the benefit of the measured performance and therefore to protect the interests of operators themselves.

5.3 Wireless Network Performance Measurements

The field campaign of benchmark measurement has the main purpose to verify and acquire performances achievable from actual and available network systems. To reach this goal the best Provider technologies, without any limiting element or operating condition with regard to the following elements, are adopted:

- user equipment;
- radio and access network;
- core network;
- traffic optimization and network management techniques, even using priority mechanisms in the access to shared resources.

Measurement campaign results are acquired by means of through sample tests and are related to Provider network performance. Although network performance is a key element, represents only one of the components which contribute to the perception of quality of service provided to the end customer and its consequent satisfaction.

So the term "Quality of Service" can take many possible meanings, but in this context it will be used in the technical context.

For such test was adopted the same user equipment (in terms of technical features) for all operators' networks, furthermore the best available technology at the beginning of the measurement campaign was adopted.

In this way the following goals were achieved:

- device invariance ensures that the user equipment affects in the same way the performance of the various networks;
- best technology ensures that measured network performance are not limited by user equipment characteristics (e.g. speed download or upload);

As mentioned, this aspect is only one component of the Quality of Service from the technical point of view, which, in turn, is only one of the factors that contribute to the real customer experience and its overall satisfaction.

5.4 Measurement Campaign Implementation

In the following section is described how accurately these tests are carried out.

Samples were acquired according to the technique of “drive-test”.

Samples acquisition is performed by equipping a vehicle with proper measurement appliances: a computer, connected to a battery of phones or USB dongles, themselves equipped with appropriate software that allows communicating with the instrument, bringing a whole range of information e.g. the status of the network, radio access physical parameters, etc. The measuring instrument was provided by SwissQual. The instrument has slots dedicated to each USB flash drives to ensure fairness during the measurement campaign.

Accordingly with architecture represented in Figure 32 user equipment implement probes connecting with measurement servers located at the edge of the provider mobile access network under evaluation. The adopted measurement system links together performance indicators, radio access physical parameters and scenario information to which they refer (the time of day, geographic location, the cell to which it is connected, and so on).

The above mentioned scenario information are useful in the processing phase, during measurement analysis, allowing a meaningful aggregation of row tests according to relevant elements that affect quality of service.

It is possible to outline some of the benefits coming from field measurements:

- performance evaluations are made from an external point with respect to the network, representing the user's perception;
- can be used to compare the results obtained for different networks since measurements can be made in the same place and at the same time;
- points where there is no coverage are adequately taken into account;

On the other hand it possible to summarize drawbacks of a field measurement campaign as follows:

- Probes setup and configuration (user device and how to use it) is not indicative of how the user actually uses its own terminal;

- in order to obtain adequate accuracy and statistical significance, it is necessary a large number of measurement samples;
- measurement paths must be representative of the user behavior from the geographical and temporal point of view. Both the density of users that services offered differs in different areas, moreover locations of measurement cannot always be considered representative of the entire network. This means that each part of a path may have to be weighted in a different way.

It shall be also taken into account:

- different user equipment;
- different conditions in which the terminals are used, such as in a car, walking, at home, in the office, on the train as well as different modes of use.

The measurement campaign should also be repeated at predefined time intervals, in this way to take account of any change of the network conditions.

In this case the test campaigns are repeated twice a year and the operating mode chosen is nomadic (which emulates only one of the possible circumstances of actual use of the user equipment).

The "nomadic" operating condition consists of static and outdoor tests, realized through an instrumentation placed on a moving vehicle, which, during the execution of the test, and for the duration of the test itself, is stopped in the measurement point.

5.5 Measuring Points Selection

The campaign consisted in outdoor drive-test measurements that were made across the entire national territory.

The measurements objective is to explore how performance vary both within and between areas with different demographic characteristics.

To this aim the map of municipal area of the identified cities object of measurement campaign is divided into elementary areas, formed by square of side equal to 500 meters, called pixels (picture element).

As part of the total pixels of a city, a basic subset in which to perform tests (defined as a function of population density) was identified; measurements were carried out within such pixel. The number of pixel is predetermined for each city and it is randomly chosen from among those belonging to the subset.

The criterion for uniform distribution of measurement pixels is obtained from the set of total pixels (500m x 500m), corresponding to the selected municipalities; this set consists of 16562 pixels. From this set, 30% of pixels are selected by ordering a function of population density, thus obtaining a subset of 4969 pixels. This subset represents the set from which they are extracted randomly measurement points. For each measurement campaign and for each city, the number of pixels is determined based on the average of two values. The first is obtained by calculating 20% of pixel with a density greater than 1875 ab/sq km, with a minimum of 10 pixels for each city. In this way 6% of total pixel of the municipalities of the regional capitals is selected.

The second comes from the distribution of 1000 pixel in 20 cities with respect to the population density of the corresponding region, imposing a minimum of 10 measurement points for the city. Table 5 shows the number of total pixel for the 20 cities, along with those for the subset and selections above described.

REGIONE	Popolazione nelle 20 regioni (%)	COMUNE	Superficie territoriale totale (%)	Popolazione nelle 20 città (%)	Numero pixel comuni	Numero pixel SUBSET (30%; Dens media > 1875)	Numero pixel 1 (20% - min. 10)	Numero pixel 2 (popolazione regionale - min. 10)	Numero pixel SELEZIONE
LAZIO	9.4%	ROMA	31.7%	28.3%	5,154	1,504	301	94	198
LOMBARDIA	16.4%	MILANO	4.4%	13.6%	728	515	103	164	134
CAMPANIA	9.6%	NAPOLI	2.8%	9.8%	497	340	68	96	82
PIEMONTE	7.4%	TORINO	3.2%	9.3%	518	360	72	74	73
SICILIA	8.3%	PALERMO	3.8%	6.7%	646	382	76	83	80
LIGURIA	2.7%	GENOVA	5.9%	6.2%	996	299	60	27	44
EMILIA-ROMAGNA	7.3%	BOLOGNA	3.4%	3.9%	567	198	40	73	57
TOSCANA	6.2%	FIRENZE	2.5%	3.8%	410	198	40	62	51
PUGLIA	6.7%	BARI	2.8%	3.3%	478	220	44	67	56
VENETO	8.1%	VERONA	5.0%	2.7%	799	184	37	81	59
FRIULI-VENEZIA GIULIA	2.0%	TRIESTE	2.0%	2.1%	368	132	26	20	23
CALABRIA	3.3%	REGGIO DI CALABRIA	5.7%	1.9%	985	108	22	33	28
UMBRIA	1.5%	PERUGIA	10.9%	1.7%	1,786	110	22	15	19
SARDEGNA	2.8%	CAGLIARI	2.1%	1.6%	354	91	18	28	23
ABRUZZO	2.2%	PESCARA	0.8%	1.3%	138	79	16	22	19
TRENTINO-ALTO ADIGE	1.7%	TRENTO	3.8%	1.2%	628	97	19	17	18
MARCHE	2.6%	ANCONA	3.0%	1.1%	504	59	12	26	19
BASILICATA	1.0%	POTENZA	4.2%	0.7%	698	42	10	10	10
MOLISE	0.5%	CAMPOBASSO	1.3%	0.5%	223	27	10	10	10
VALLE D'AOSTA	0.2%	AOSTA	0.5%	0.4%	85	24	10	10	10
	100.0%	TOTALE	100.0%	100.0%	16,562	4,969	1,006	1,012	1,013

Table 5: Selection criterion of measuring points

5.6 Key Performance Indicators (KPI)

To characterize the mobile access networks as accurately as possible performance measurements are represented by some technical parameters (such KPIs, Key Performance Indicator). Starting from the measured values of these parameters will be possible to typify quality aspects of Internet Services.

In this analysis, the interest is essentially focused on the objective quality of service, achieved precisely through directly measurable metrics, such as speed, delay, jitter or packet loss, rather than the subjective (i.e. how the user perceives the service level of a certain network application or a service).

Services provided by mobile system are common divided into three families: voice, SMS and data services. But there is an important difference between these three categories.

As for voice and SMS services the entire category consists mainly of a single essential service, in case of data services we are faced with a multitude of services such as browsing, streaming, e-mail, etc.) and applications (such as YouTube, Facebook, Google, etc..) available to the user.

From the objective observation point of view, this has a substantial impact in the methodology construction for QoS evaluation, in fact it is difficult to identify common parameters that are both objective (i.e. independent of subjective aspects associated to user preferences) and direct (i.e. able to express the immediate effect on the perceived quality of that particular service).

In addition, the number and type of services based on Internet access has grown enormously in recent years and the future seems destined to grow.

Rather than follow this rapid evolution and differentiation of services, defining specific parameters for a single service, it was decided to identify a set of parameters common and objective, on the basis of which it may be inferred a qualitative evaluation (good, acceptable, poor) of a single service.

According to ETSI [54, 55] some KPI have been chosen for evaluation of network performance and quality of mobile broadband service, in terms of:

a) Data transmission throughput,
b) Data transmission unsuccesful rate,
c) Packet delay,
d) Packet loss,
e) Jitter.

These KPIs can be measured with facility and with a good degree of objectivity at the same time, summarizing key features of the mobile network being analyzed. On the basis of their accurate statistical description is also possible analyze the individual services behavior in terms of quality, also from a user point of view.

For example, considering a file transfer (upload or download), the key parameter is the throughput. Otherwise taking into account one of the most common service the web browsing, the interest is focused in how many time is necessary to completely download the web page. While in case of real-time communications, such as VoIP services, is necessary to have a guaranteed rate, but a significant effect is played by the transmission delay variation (jitter).

Mentioned indicators are related to the Quality of Service during the correct execution and delivery of the service itself, providing an indication of proper

maintaining of service quality during the complete period of time requested by the user (Service Integrity).

However quality of service is completed with two other important aspects:

- Service Accessibility: after the network access phase, user requests access to a service. This access can be denied for a variety of causes, such as the lack of free radio channels or available links between the Base Station (BS) and the mobile network switching center (MSC).
- Service Retainability: indicates if the requested service has been maintained until user wants to use it, or if the close was forced by the network independently of the user will. Service Retainability is valued by the failure rate and packets loss rate indicators.

5.7 Measurement Suite Definition

During every test suite, a test cycle (of a fixed duration equal to 20minutes) composed of atomic tests is running, providing measurement of following KPI:

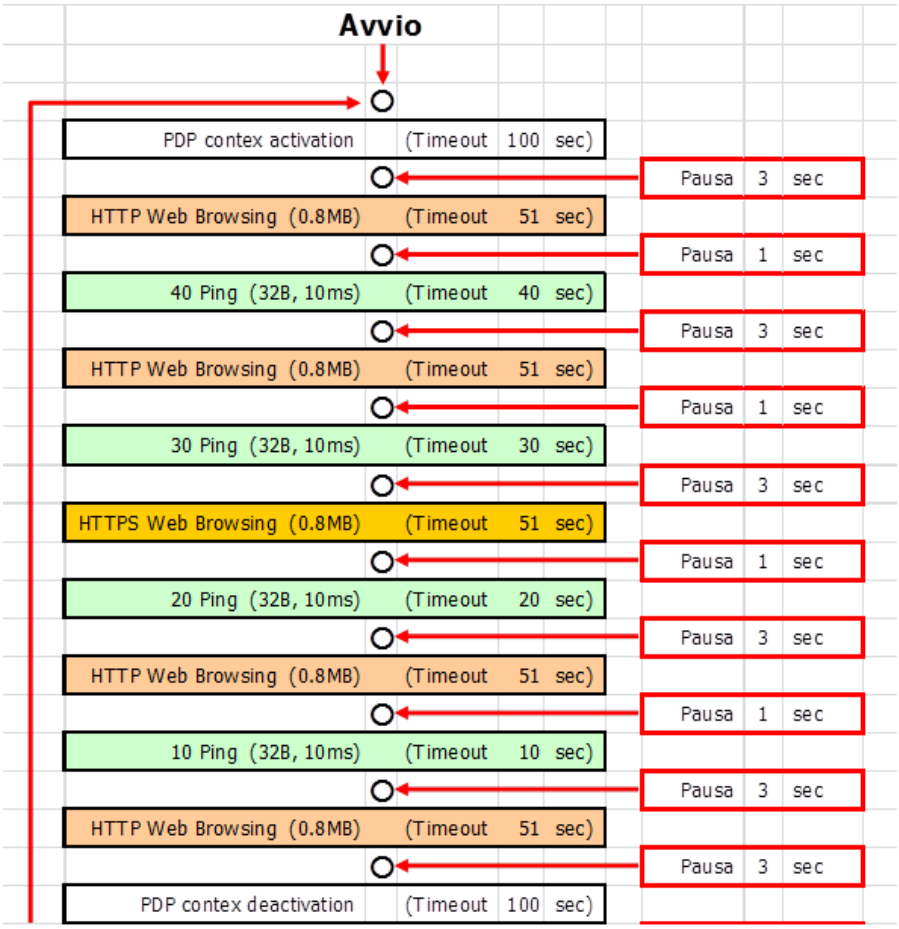
- FTP upload data transmission and failure rate;
- HTTP download data transmission and failure rate;
- HTTP download web page browsing and failure rate;
- Data transmission delay (Round Trip Time);
- Packet loss rate;
- Jitter.

Regarding the data rate measurements three different indicators, as previously mentioned, are used to determine:

- File download: an MP3 test file is downloaded, the file size is 3Mbyte and adopted protocol is HTTP (HTTP Download test);
- File upload: an MP3 test file is uploaded, the file size is 1Mbyte and adopted protocol is FTP (FTP upload test);
- WEB browsing: time to download a web page defined by ETSI, the web reference page size is 800kbyte and it is composed of texts, images, compressible and non-compressible data.

To measure the packet transmission delay, ping technique has been adopted. Measurements are carried out by ICMP packets (Internet Control Message Protocol) having a size of 32 byte.

In the next figure test cycle description implemented in the measurement campaign is reported. As shown in Figure 33 each measurement test reported a timeout value, and between measurement tests a pause is regulated.



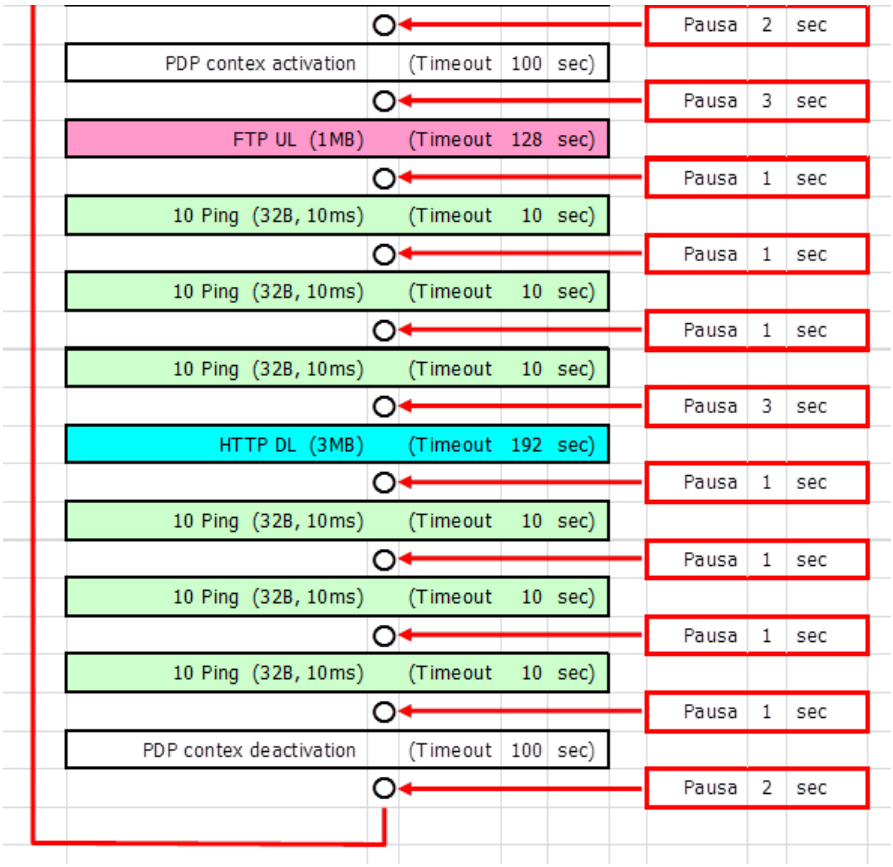


Figure 33: Test cycle flow diagram

5.8 Network Results

All the results obtained from the measurement campaign were published on the AGCOM's website (www.misurainternetmobile.it).

In this chapter some results carried out on a subset of measurement point are reported. In particular some aggregate results related to the network performance measurement delivered by the four mobile networks Italian operator are shown; it is important to underline that this data were acquired during a trial phase of the official campaign. Results reported in this paragraph outline the state of mobile broadband performance, as above mentioned the official and complete results set obtained during measurement campaign is reported on AGCOM website. In particular Figure 34 shows the probability density function of ftp upload tests.

The average upload speed is about 1.6 Mb/s.

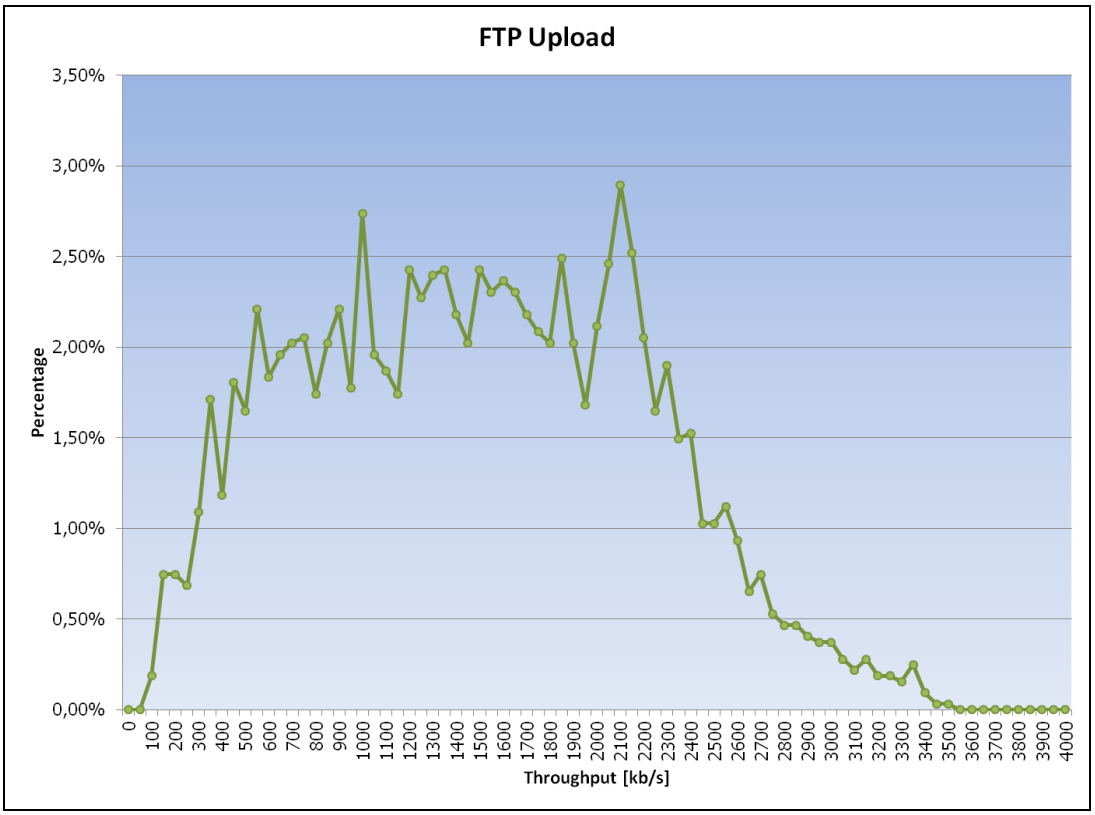


Figure 34: FTP Probability Density Function

In the next figure the probability density function of http download tests performance is illustrated. The average download speed for the four networks is about 6.2 Mb/s.

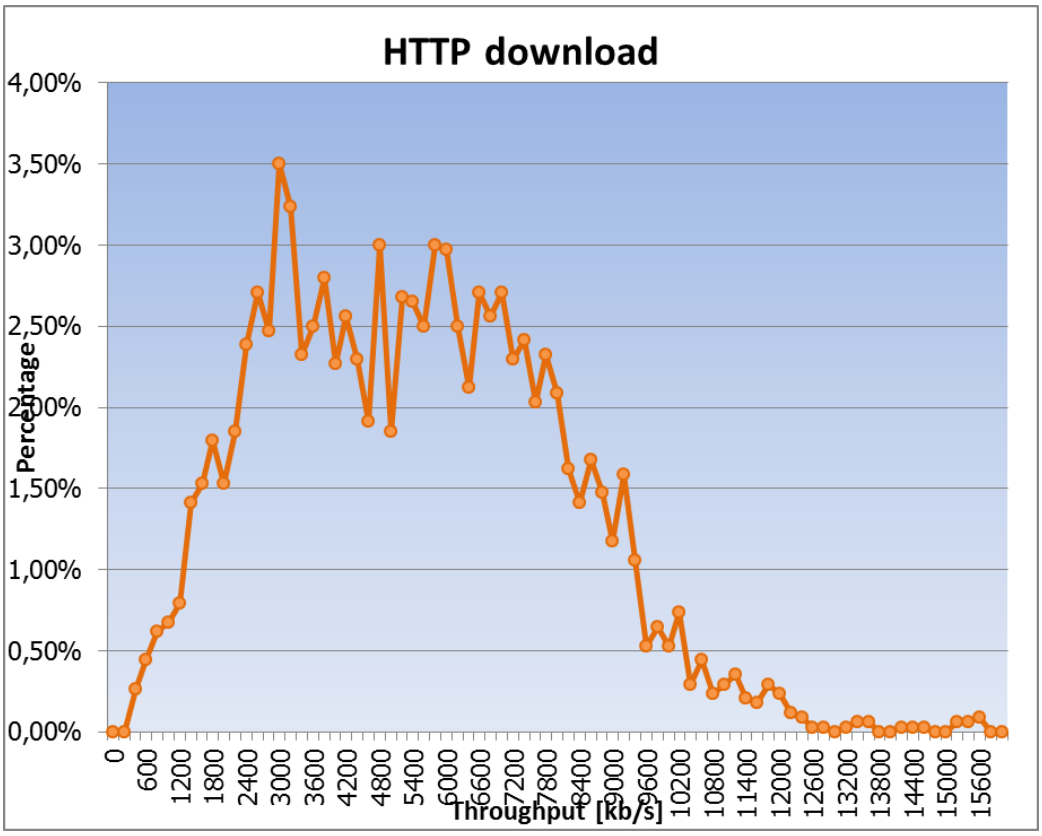


Figure 35: HTTP Probability Density Function

Figure 36 reports the trend to download a standard web page of 800 kbyte.

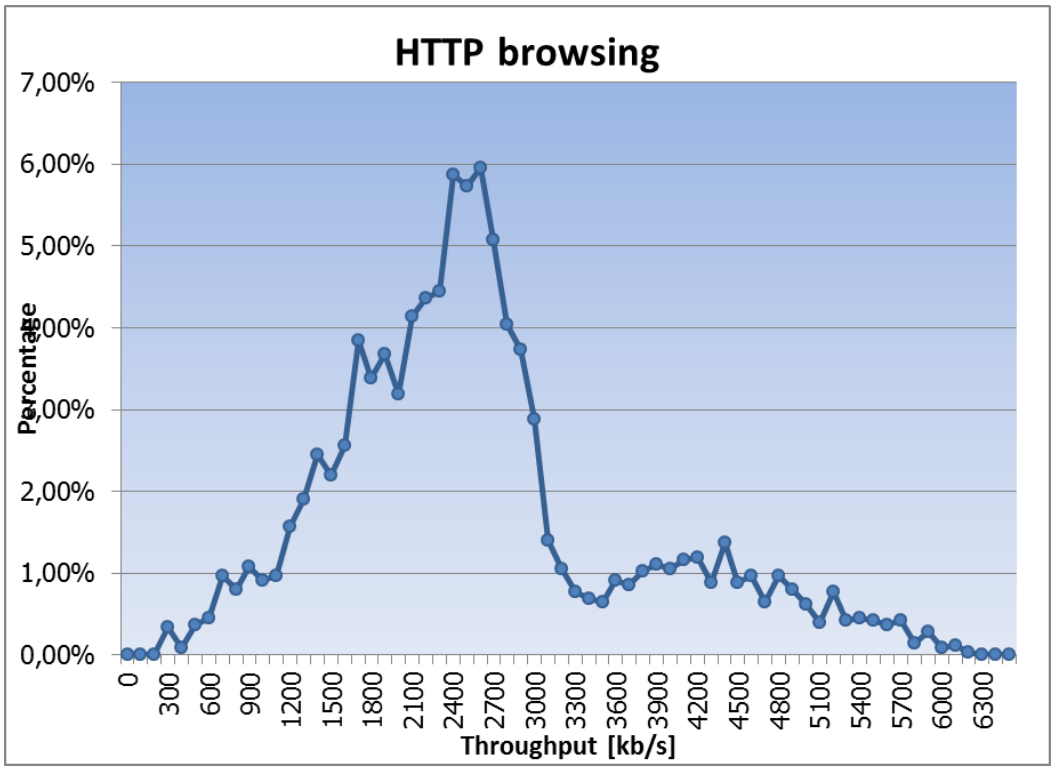


Figure 36: Browsing Probability Density Function

The difference between results shown in Figure 35 and Figure 36 depends on the file size to be downloaded; in the first case the file size is 3Mbyte while in the second case the web reference page size is 800kbyte.

As illustrated in Chapter 2 the TCP congestion control mechanism requires a transient before reaching the steady state and the highest throughput. Furthermore the transient period depends directly from the network latency. As a consequence larger files lead to better exploit the TCP steady state, achieving higher throughputs in specific data transfer.

Based on investigations and results reported in Chapter 2 a multisession test suite, as proposed in Chapter 3, has been implemented.

This is one of the major contributions of this work, overcoming TCP limitation especially for high RTT, strongly relevant in radio mobile environment.

The multisession test consists of 5 parallel download sessions performed by HTTP protocol. Test runs for 1 minute and at the end of such period the number of bytes transferred over all connection is computed. Cumulative curves reported in Figure 37 show how the multisession technique permits to much better exploit the line capacity (orange line) with respect to download with a single connection (blue line).

In this way it is possible to fully exploit the line capacity, achieving an effective bandwidth estimation. Figure 37 shows how the multisession approach can be adopted to evaluate the line capacity exceeding the limit of the TCP protocol described in the Chapter 3.

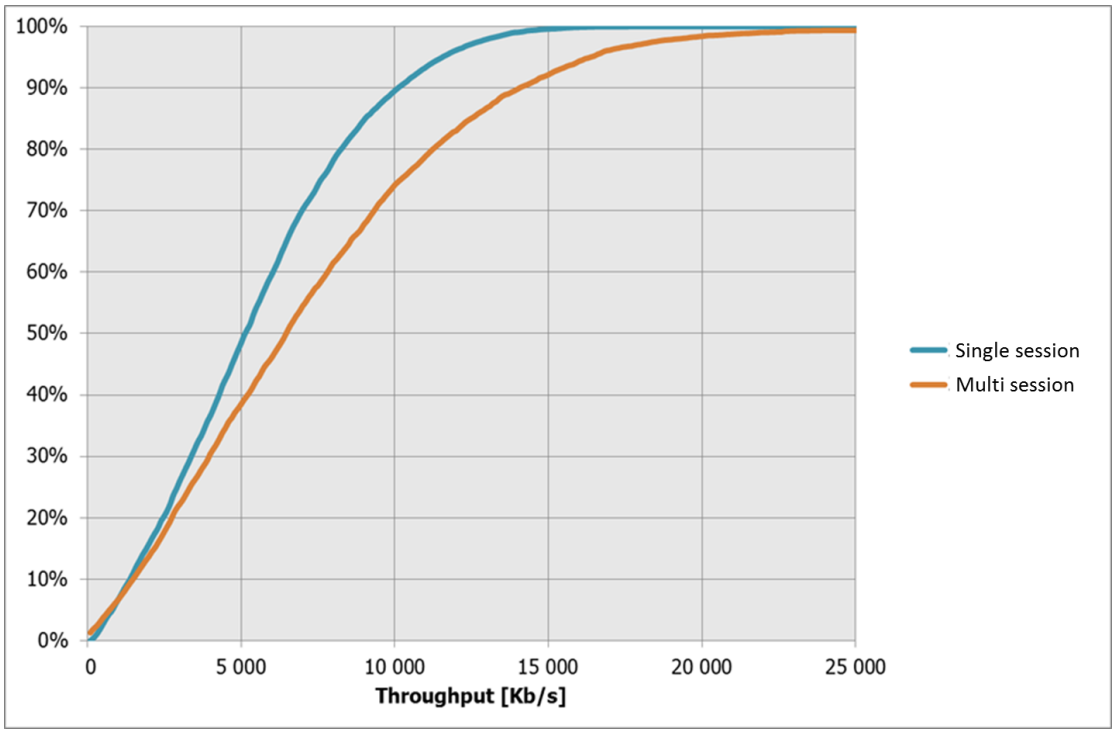


Figure 37: HTTP Cumulative Distribution Function

Chapter 6
Channel Characterization: the Interference Between DVB-T and LTE Systems

Focusing on SLA at physical layer, investigation on interference between LTE and DVB-T systems has been investigated and reported in this chapter, in particular regarding the impact on QoS.

As mentioned in the previous chapter the world of mobile Telecommunication, in the last 20 years, has been characterized by a surprising technological evolution that has never slowed down. Fourth generation systems are now started and this leads to innovative technological developments as well as large economical investments.

The fourth generation of mobile systems, whose base is the standard 3GPP-Long Term Evolution (LTE), guarantees higher data speed as well as high quality new multimedia services and will ensure convergence towards all-IP networks. Due to the spectrum refarming process, the most likely designated frequency bands where the new services can operate are the ones at present used by UMTS and GSM, as well as new bands at 2.6 GHz. However, the best frequency band option to open up these services would be, definitely, the digital dividend portion of the UHF spectrum (from 790 MHz to 862 MHz), resulting from the digital switchover (DSO) process from analog to digital television (DTV).

For both mobile operators and customers the benefits of carrying out this choice will be numerous including larger coverage areas and best penetration into buildings. All of this, of course, operating at the same radiated power or lower.

During the last World Radio Conference (WRC-2007) was indicated to allocate the 790-862 MHz frequency band to mobile services in several world regions. Such position was also supported by main international radio communication organizations (CEPT, FCC, etc.). However, in order to make these systems fully interoperable, the problem of coexistence between DVB-T channels, operating in the lower part of the UHF band (470-790 MHz) and signals of future mobile

communication systems, operating in the above mentioned band, must be carefully analyzed and evaluated and results of these studies should be taken into account in the implementation of the new broadband mobile networks.

In the transition process from analogue to digital television system (switch-off) through which frequency resources were made available that can be allocated to different services from those of broadcast services and to meet the European Commission's recommendations contained within the Decision 2010/267/EU of 6/5/2010, an experimental investigation aimed at reuse of the UHF-TV band for mobile services, has been conducted. The signal propagation characteristics in this frequency band provide an high quality of service and a low cost implementation of the access networks. This has encouraged a large and very heated discussion about the possible consequences that may result due to the introduction of these technologies, especially taking into account that they will be implemented on adjacent bands to those currently used by TV operators.

In order to provide a relevant contribution to the debate in this particular technological environment, it is crucial to identify the technical limitations arising from the use of a specific technology on these frequency bands. These technological limitations have been characterized from the analysis and evaluation of the interference effects that might arise from the coexistence in the UHF range of technologies of different nature. On the basis of experimental findings will be possible to give indications about the type of technological solutions to be used in order to minimize any interference effects.

The goal of this study is to define the most appropriate technological parameters and the conditions that these parameters must comply in order to guarantee the coexistence of different technologies when used for signals transmitting on the reserved band for the digital dividend adjacent the UHF band, as regard the LTE technology.

The present chapter intends to investigate the interference issues especially considering the effects of signal emitted by new generation mobile stations on existing DVB-T systems, whose quality of service should still be guaranteed. In particular this analysis focuses on the above mentioned coexistence problem taking

into account the interference impact on DVB-T signals and evaluating, for some reference scenarios, the Protection Distance (*dP*) between LTE and DVB-T stations [56].

In particular, after a brief description of the UHF band regulatory aspects, the concept of Protection Distance and an analytical approach to estimate it, is proposed.

6.1 UHF Band Regulatory Aspects

By the year 2012, the transition process from analogue to digital television systems is completed. Until this time the two signals have coexisted in the same frequency bands: VHF band III (174-230 MHz) and UHF bands IV and V (470-862 MHz). The switch off process has made possible to release the last portion of the UHF band V that ranging from 790 to 862 MHz (or from channel 61 to channel 69). Starting from the CEPT (European Conference of Postal and Telecommunications Administrations) decision to implement networks for mobile telecommunication systems in both TDD or FDD duplexing technique in the 790-862 MHz band, several studies have been conducted [57].

In case of TDD implementation (or mixed TDD/FDD) has been set a guard band of 7 MHz, between the UHF channel 60 (782-790 MHz) and the 13 blocks of 5 MHz channels (see Figure 38), where radio mobile signals will be transmitted.

This channelization meets the requirements of country authorities having problems of spectral harmonization with neighboring countries or that have already allocated part of the UHF band V to other services.

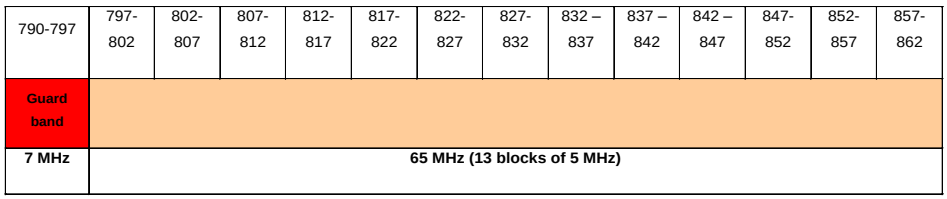


Figure 38: CEPT Channelization proposal in case of TDD duplexing technique implementation

In case of using the FDD technique, two 30 MHz bands have been established: one for downlink and one for uplink communications, separated by a duplex gap of 11

MHz (see Figure 39). Each of these bands is split into 6 blocks of 5 MHz. The downlink frequency band starts at 791 MHz, while the corresponding uplink frequency band starts at 832 MHz, with a resultant guard band of 41 MHz.

The reason of this choice is to protect the DVB-T broadcast signal transmitted on 60 UHF channel from interference produced by the uplink transmission of mobile ECN (Electronic Communication Networks) terminals, which represents a critical aspect for DVB-T receiving systems. From this point of view, the adoption of the above mentioned frequency allocation, determines a 41 MHz virtual guard band. In addition, the remaining 11 MHz duplex gap could be exploited by other unspecified services.

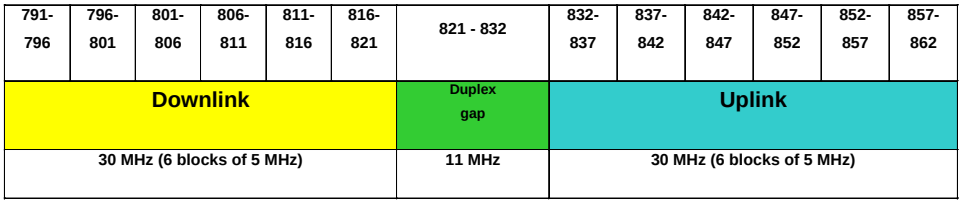


Figure 39: CEPT Channelization proposal in case of FDD duplexing technique implementation

6.2 Interference Scenarios and Block Edge Mask

In order to ensure the co-existence between digital television broadcasting and 3G/4G cellular technologies, operating in adjacent frequency bands, the interference problem must be addressed.

It is obvious that the main problem is represented by the co-channel and adjacent channel interference between the DVB-T signal and the 3G-4G systems in adjacent geographical areas, and from adjacent channel interference between the two systems in the same geographical area.

For convenience, the 3G-4G systems are collectively referred to as ECN (Electronic Communication Networks).

The DVB-T system is the most penalized because:

- the field strength of the interfering ECN BS is larger than the typical heights of the buildings on which the receiving DVB-T antennas are positioned;

- the antenna gain of the fixed DVB-T receivers is higher than that of mobile ECN devices;
- at the edge of the covered area by the DVB-T transmitter, the desired signal is very low.

CEPT has proposed solutions to optimize the use of the 790-862 MHz band in addition to the guard bands, including the adoption of the block edge mask (BEM) both the ECN BS and the ECN TS (UE) [57].

The approach of the BEM is to define emission limits for the components in band (in-block) and out of band (out-of-block) of the interfering signal.

A block edge mask is constituted by the in-block component that is a threshold that cannot be exceeded by the in-band power of the interfering signal, and by the out-of-block component composed of a baseline threshold, which is a level of power that must not be passed by the interfering signal out of band.

The threshold levels (for both components) are given by the highest values of the signals emissions of the ECN systems taken into consideration by the CEPT.

The BEM recommended by CEPT are also dependent on the type of duplexing technique (TDD or FDD), and the values are relative to the spectral emission mask of the LTE signal 10 MHz.

Limiting the power interfering with emission masks is necessary for the coexistence of several signals that occupy contiguous frequency bands.

However, the use of the only BEM is not enough to guarantee complete protection of the DVB-T signal.

Therefore, is necessary introducing mitigation spectral techniques such as:

- co-siting between DVB-T BS and ECN BS as the coverage area of the ECN BS coincides with the area of maximum emission of DVB-T signal; this represents the best technique;
- cross polarization: horizontal for DVB-T and vertical to the ECN (keeping in mind that future ECN Mobile Terminals could use MIMO techniques with both polarizations);
- adjustment of power, height, pattern and direction of the interfering BS on the effective radiant field existing DVB-T (especially for UHF channels 59 and

60), even if you have to take into account a large number of ECN BS operating in these channels and therefore of the economic impact resulting from the modification of the Base Stations;

- installation of rejection filters on DVB-T receivers located near the ECN BS, useful to minimize the interference on a single channel and not on the entire band, since it is not possible to limit the out of band emissions. In addition their cost is not negligible.

6.3 Protection Distance Concept and Evaluation approach

In fixed and mobile cellular networks the interference between uplink and downlink transmissions, among different operators serving the same geographical area and operating in adjacent channels, can generate extremely high noise due to inadequate filtering, unwanted modulation products, improper tuning, or poor frequency control, in either the reference channel or the interfering channel, or both. The purpose of the available spectrum optimization is important the study of the protection ratio, and therefore the study of the interference level between the two signals, as a function of some important parameters such as the distance between the DVB-T receiver and the ECN Base Station (terminal station) and the value of the frequency signals offset interfered-interfering.

In particular in this paragraph the Protection Distance (*dP*) concept between LTE and DVB-T stations will be introduced; setting an appropriate distance between the victim and the interfering system can help to mitigate these undesired effects.

The *dP* parameter is defined as the minimum distance between an LTE (BS or UE) antenna and a DVB-T receiving antenna in order to make the interference effects at the DVB-T front-end acceptable in terms of quality of service.

Recommendation ITU-R SM.337-5 [58] provides a general method to evaluate the spatial distance between an interfering and a victim antenna for radio systems, as well as their frequency separation in order to bind the interfering signal under an acceptable power level. According to this method, to evaluate *dP*, in addition to the selection of the adopted radio propagation model (Gaussian, Rayleigh, etc.), the following physical parameters must be taken into account:

- power and spectral distribution of the desired signal at the receiver or the signal that would otherwise be received without interference;
- power and spectral distribution of the interfering signal at the receiver;
- path loss contribution to signal attenuation;
- characteristics of the receiver (noise figure, demodulation technique).

The basic scenario taken into consideration analyzes the case of a desired DVB-T signal received at the antenna with power P_d, which is interfered by a signal, with power P_r, emitted by a mobile system (ECN LTE for example) on a carrier at distance Δf from the DVB-T carrier frequency and adding, on the desired signal (exclusively in the reference channel of the receiver), an interfering signal with power level equal to Pi (see Figure 40).

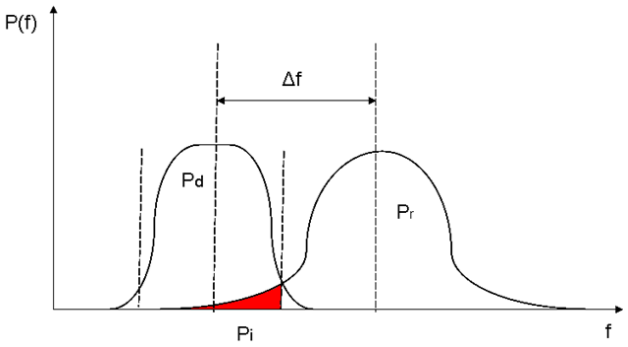


Figure 40: Wanted/interfering signal modeling

The power Pi depends on two kinds of factors: the spectral one and the spatial one.

The spectral factor depends on the power density spectrum of the interfering signal and the frequency response of the receiver.

It is expressed by the off-channel rejection factor (OCR(Δf)), which represents an index of the receive selectivity:

$$OCR(\Delta f) = -10\log\frac{\int_{-\infty}^{+\infty} P(f)\left|H(f+\Delta f)\right|^2 df}{\int_{-\infty}^{+\infty} P(f)df} \qquad (7)$$

In this expression:

- P(f) is the power density spectrum of the interfering signal;
- H(f) is the frequency response of the DVB-T receiver in RF band;

- Δf is the difference between the carrier frequencies of the interfering and the desired signals.

The spatial factor depends on the antennas distance between the victim receiver system and the interfering one, the propagation model chosen and the statistical distribution of the signal, and is expressed by the path loss L_p. If the diffraction propagation model (see the worst case reported in the Recommendation ITU-R SM337-5, [58]) is adopted, L_p is given by:

$$L_p = L_{FS} - L_{DIF/FS}$$

$$L_{FS} = 32,44 + 20\log_{10} f_{MHz} + 20\log_{10} d_{iv} - G_T - G_R \quad (8)$$

where:

- L_{FS} is the attenuation due to free space (dB);
- $L_{DIF/FS}$ is the attenuation due to diffraction (dB), and is usually a negative term;
- d_{iv} is the distance between the interfering and the victim antennas, and is expressed in km.

The goal in this work is evaluate how the above mentioned factors affect the interfering power, and consequently calculate dP.

In the following a brief description of the approach adopted to calculate the Protection Distance of the DVB-T signal in presence of an interfering radio mobile signal (BS and UE) is given.

Taking into account the minimum C/N ratio (CNR_{min}) required for a DVB-T signal good quality reception and the two mentioned factors, for each value of power level P_d it is possible to calculate the correspondent maximum interfering power P_i that does not cause critical interference on the DVB-T signal (P_i*). In general, the interfering power is calculated by the formula:

$$P_i = P_t + G_r - L_p - OCR(\Delta f) = P_r + G_r - OCR(\Delta f) \quad (9)$$

where P_t is the EIRP (dBW) of the interfering transmitter, G_r is the receiver antenna gain (dBi) and $P_r = P_t - L_p$ is the power level of interfering signal (dBW) at the receiver.

According to the expression (9) the interfering power depends on the power P_r, which is function of the distance between the antenna transmitting the interfering

signal and the antenna of the victim system (DVB-T), but also by the carrier frequency offset Δf.

As evidentiated in Figure 40, P_i is the portion of the mobile signal spectral power causing interference on the DVB-T signal. It represents the co-channel or adjacent channel interference depending whether the frequency offset Δf is null or equal to one (or more) channel spacing. Once the values of P_d and CNR_{min} are fixed, the maximum allowable interfering power P_i*, in the DVB-T signal bandwidth, can be calculated as reported in expression (10):

$$\frac{P_d}{(N+P_i)} \geq CNR_{min} \equiv \frac{S_x}{N}$$

$$P_i \leq N\left(\frac{P_d}{S_x} - 1\right) \equiv \frac{(P_d - S_x)}{CNR_{min}} \equiv P_i \quad (10)$$

According to expression (10), in order to not have a clear degradation of TV signal quality, the expression $P_i \leq P_i$* should be fulfilled.

Using the value P_i* in expression (9), it is possible to extrapolate the L_p parameter and, consequently, the distance d_{iv} from expression (8) that, in this case, is equivalent to the Protection Distance (dP).

To empirically evaluate a solution of the above mentioned expression, simulation software has been implemented using MATLAB tool.

6.4 Physical characteristics of the DVB-T and ECN systems

6.4.1 DVB-T signal

As earlier mentioned, DVB-T is the victim signal in this simulation analysis.

A DVB-T receiver is characterized by several parameters (see Table 6), including receiver sensitivity (S_x) and the minimum C/N ratio (CNR_{min}) required for a DVB-T signal good quality reception. Table 7 provides a list of values for CNR_{min} extracted from the ETSI EN 300 744 V 1.6.1 recommendation [59].

Operational Frequency	470-862 MHz
Reference Frequency	790 MHz
Bandwidth	8 MHz
Receiver Noise Figure	7 dB
Reference location probability	95%
Environment	urban, suburban, rural

Table 6: DVB-T system parameters for a generic user terminal (fixed/portable/mobile)

Concerning the modulation scheme, taking into account the high computational time, in this study only two different configurations have been considered. The first one, more robust to interferences, operating with 8k carriers, modulated 16QAM 2/3, assuming a Gaussian channel model and an SNR_{min}=11.4dB (DVB-T-1 configuration); the second one, a little more sensible, operating with 8k carriers, modulated 64QAM 2/3, assuming a Gaussian channel model and an SNR_{min}=16.7dB (DVB-T-2 configuration).

The DVB-T receiver is assumed to operate with an ideal receive filter H(f) and have a sensitivity S_x=-77.2 dBm [60].

		Required C/N (dB) for BER = 2×10^{-4} after Viterbi QEF after Reed-Solomon (see note 2)			Bitrate (Mbit/s) (see note 3)			
Constel-lation	**Code rate**	**Gaussian Channel (AWGN)**	**Ricean channel (F_1)**	**Rayleigh channel (P_1)**	**Δ/T_U = 1/4**	**Δ/T_U = 1/8**	**Δ/T_U = 1/16**	**Δ/T_U = 1/32**
QPSK	1/2	*3,5*	*4,1*	*5,9*	*4,98*	*5,53*	*5,85*	*6,03*
QPSK	2/3	*5,3*	*6,1*	*9,6*	*6,64*	*7,37*	*7,81*	*8,04*
QPSK	3/4	*6,3*	*7,2*	*12,4*	*7,46*	*8,29*	*8,78*	*9,05*
QPSK	5/6	*7,3*	*8,5*	*15,6*	*8,29*	*9,22*	*9,76*	*10,05*
QPSK	7/8	*7,9*	*9,2*	*17,5*	*8,71*	*9,68*	*10,25*	*10,56*
16-QAM	1/2	*9,3*	*9,8*	*11,8*	*9,95*	*11,06*	*11,71*	*12,06*
16-QAM	2/3	*11,4*	*12,1*	*15,3*	*13,27*	*14,75*	*15,61*	*16,09*
16-QAM	3/4	*12,6*	*13,4*	*18,1*	*14,93*	*16,59*	*17,56*	*18,10*
16-QAM	5/6	*13,8*	*14,8*	*21,3*	*16,59*	*18,43*	*19,52*	*20,11*
16-QAM	7/8	*14,4*	*15,7*	*23,6*	*17,42*	*19,35*	*20,49*	*21,11*
64-QAM	1/2	*13,8*	*14,3*	*16,4*	*14,93*	*16,59*	*17,56*	*18,10*
64-QAM	2/3	*16,7*	*17,3*	*20,3*	*19,91*	*22,12*	*23,42*	*24,13*
64-QAM	3/4	*18,2*	*18,9*	*23,0*	*22,39*	*24,88*	*26,35*	*27,14*
64-QAM	5/6	*19,4*	*20,4*	*26,2*	*24,88*	*27,65*	*29,27*	*30,16*
64-QAM	7/8	*20,2*	*21,3*	*28,6*	*26,13*	*29,03*	*30,74*	*31,67*
NOTE 1: Figures in italics are approximate values. NOTE 2: Quasi Error Free (QEF) means less than one uncorrected error event per hour, corresponding to BER = 10^{-11} at the input of the MPEG-2 demultiplexer. NOTE 3: Net bit rates are given after the Reed-Solomon decoder.								

Table 7: CNR_{MIN} FOR A DVB-T SYSTEM

6.4.2 LTE signal

LTE systems, in this simulation study, are the interferer sources and are assumed to transmit on a 5 MHz bandwidth channel. Transmitting antennas of base station and user terminal systems are supposed to be at an height of 30m and 1.5m respectively. Being unable to get the spectra effective numerical representation of the interfering signals, the power density spectrum of LTE signals have been approximated using the spectrum Block Emission Masks (BEM) reported on the ETSI recommendations, ETSI TS 136 104 V8.7.0 for LTE base stations (see Table 8) and ETSI TS 136 521-1 V8.3.1 for LTE UE (see Table 9) [61, 62].

Frequency offset of measurement filter -3dB point, Δf	Frequency offset of measurement filter centre frequency, f_offset	Minimum requirement	Measurement bandwidth (Note 1)
0 MHz ≤ Δf < 5 MHz	0.05 MHz ≤ f_offset < 5.05 MHz	$-7dBm-\frac{7}{5}\cdot\left(\frac{f\_offset}{MHz}-0.05\right)dB$	100 kHz
5 MHz ≤ Δf < 10 MHz	5.05 MHz ≤ f_offset < 10.05 MHz	-14 dBm	100 kHz
10 MHz ≤ Δf ≤ Δf_{max}	10.05 MHz ≤ f_offset < f_offset_{max}	-13 dBm	100 kHz

Table 8: LTE BS (5 MHz): 3GPP Spectrum Emission Mask

Spectrum emission limit (dBm)/ Channel bandwidth							
Δf_{OOB} (MHz)	1.4 MHz	3.0 MHz	5 MHz	10 MHz	15 MHz	20 MHz	Measurement bandwidth
± 0-1	-10	-13	-15	-18	-20	-21	30 kHz
± 1-2.5	-10	-10	-10	-10	-10	-10	1 MHz
± 2.5-2.8	-25	-10	-10	-10	-10	-10	1 MHz
± 2.8-5		-10	-10	-10	-10	-10	1 MHz
± 5-6		-25	-13	-13	-13	-13	1 MHz
± 6-10			-25	-13	-13	-13	1 MHz
± 10-15				-25	-13	-13	1 MHz
± 15-20					-25	-13	1 MHz
± 20-25						-25	1 MHz

Table 9: LTE UE (5 MHz): 3GPP Spectrum Emission Mask

6.5 Results for Protection Distance

In the simulation tests two propagation scenarios have been considered:

- the "large urban environment", where the LTE BS interfering antenna height is 30 meters, the LTE UE interfering antenna height is 1.5 meters and DVB-T interfered antenna height is 30 meters;

- the "small urban environment" where the LTE BS interfering antenna height is 30 meters, the LTE UE interfering antenna height is 1.5 meters and DVB-T interfered antenna height is 20 meters.

To calculate the attenuation of LTE signal, reference is made to Recommendation ITU-R SM.337-5, 2007 [58]. For both scenarios the power of interfering LTE signals (BS and EU) at the DVB-T receiver has been calculated as a function of the distance between interfering and interfered antennas, adopting different modulation format (64QAM and 16QAM) and considering several power levels of DVB-T signal (P_d spanning from -70dBm to -50dBm) received at the fixed rooftop receiving antenna front-end.

The value P_d=-70dBm corresponds roughly to the power of the DVB-T signal to the edge of the DVB-T transmitter macrocell (in an urban environment is about 28 km).

Note that, as reported in paragraph 6.1, the FDD duplexing solution for 790-870 MHz UHF frequency band guarantees an effective virtual guard band (41 MHz) between UHF DVB-T channel 60 and LTE UE first block FDD uplink. For this reason, in the following it will be analyzed only the TDD duplexing technique case study for LTE network implementation.

6.5.1 LTE UE Interfering

It should be underlined that, when a mobile network provider need of using a TDD duplexing technique and have to transmit in a channel adjacent to the UHF DVB-T channel 60, the UE transmitter signal should be compliant to the LTE signal emission mask specified in the CEPT Report 30 (see Figure 41).

In this case, simulating the worst case (that is when the DVB-T receiving antenna is at the edge of the cell coverage of the transmitter and the signal is 64QAM modulated), it results that the protection distance between the LTE-UE and the antenna of the DVB-T terminal fixed roof is in the order of just a few meters.

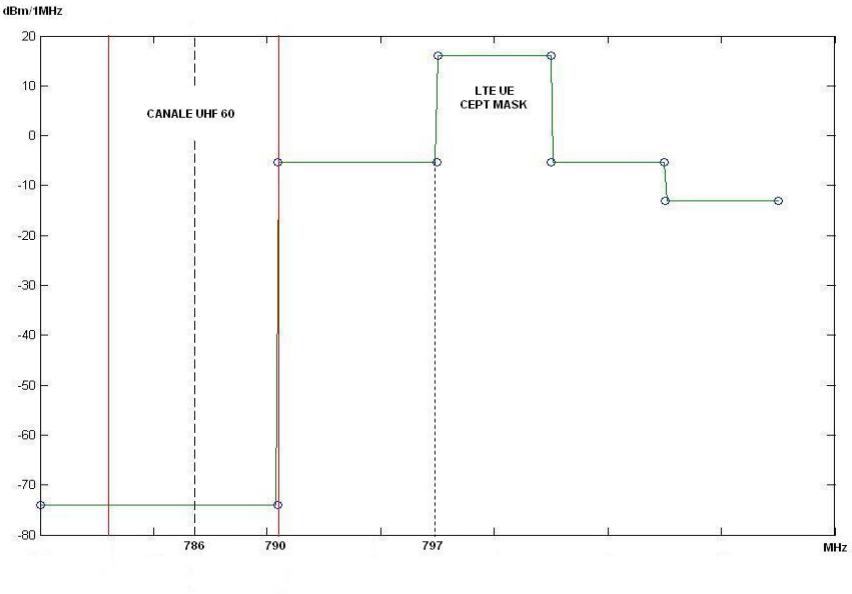


Figure 41: CEPT mask for a LTE UE (TDD) 5 MHz signal adjacent to UHF 60 channel

This result is possible as a consequence of the blocking mask severe requirements recommended by CEPT for the LTE-UE signal. In fact, as it is shown in Figure 36, a quite consistent guard band (7 MHz) between the upper edge of the UHF channel 60 and the lower edge of the first available LTE channel (797-802 MHz) has been fixed. In addition, the out of band emission limit for LTE signal for frequencies below 790 MHz is equal to -74 dBm, that is very close to the sensitivity level of a DVB-T receiver (about -77 dBm).

6.5.2 LTE BS Interfering

A different result is obtained when the DVB-T receiving antenna is interfered by an LTE Base Station. In the considered simulation tests, the LTE BS, can transmit a signal in a frequency band adjacent to the UHF channel 60. In this case the LTE BS signal must be compliant with the constraints imposed by the CEPT recommendation, reported in Table 10, where out-of-band emissions are subject to significant restrictions.

In Figure 42 the emission mask of an LTE BS 5 MHz signal (EIRP: 43dBm/5MHz) is shown in case the protection of UHF channel 60 is required (case A in Table 10).

Case	Frequency range of out-of-block emissions	Condition on base station in-block E.I.R.P., P (dBm/10MHz)	Maximum mean out-of-block EIRP	Measurement bandwidth
A	For DTT frequencies where broadcasting is protected	P ≥ 59 dBm	0 dBm	8 MHz
		36 dBm ≤ P < 59 dBm	P-59 dB	8 MHz
		P < 36 dBm	- 23 dBm	8 MHz
B	For DTT frequencies where broadcasting is subject to an intermediate level of protection	P ≥ 59 dBm	10 dBm	8 MHz
		36 dBm ≤ P < 59 dBm	P-49 dB	8 MHz
		P < 36 dBm	-13 dBm	8 MHz
C	For DTT frequencies where broadcasting is not protected	No condition	22 dBm	8 MHz

Table 10: BASELINE REQUIREMENTS – BS BEM OUT-OF-BLOCK EIRP LIMITS OVER FREQUENCIES OCCUPIED BY BROADCASTING

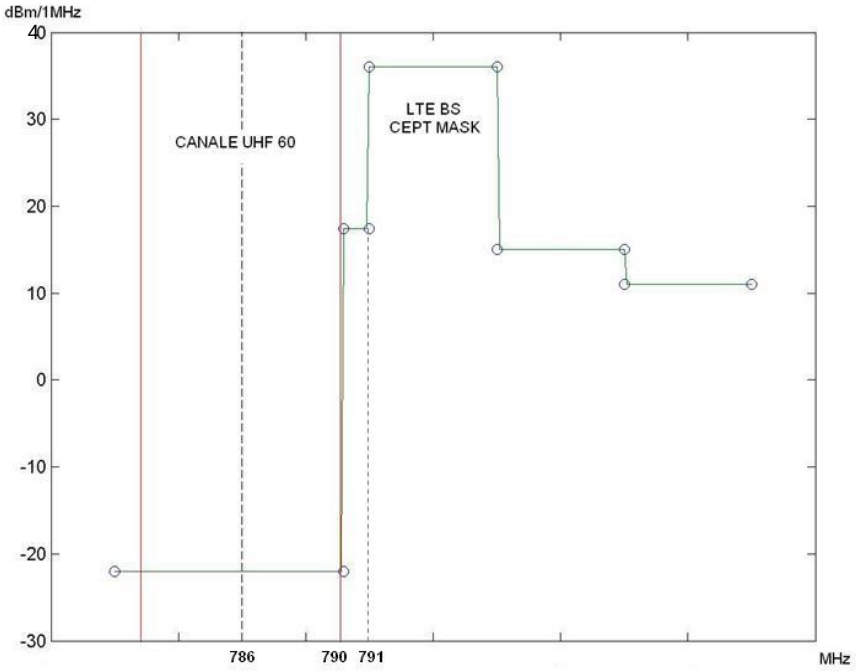
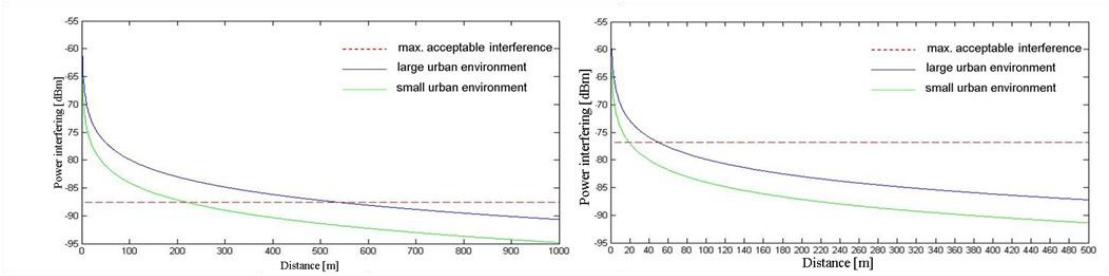


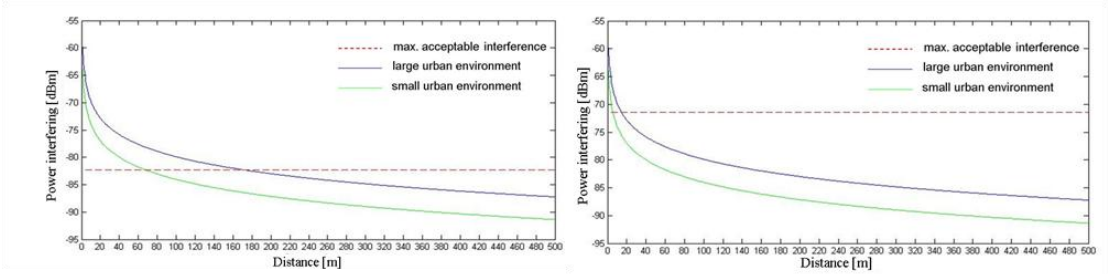
Figure 42: CEPT mask for a LTE BS (FDD) 5 MHz signal adjacent to UHF 60 channel

In Figure 43 and Figure 44 results obtained respectively in large and small urban environment reported in [63, 64] are shown. In particular figures show the power of an interfering LTE BS antenna as a function of its distance (in meters) from fixed roof DVB-T antenna and considering different received powers (in dBm) and DVB-T signal modulation formats, calculated for both the large and small urban environment. In each graph the dotted line indicates the maximum interference level of LTE signal power acceptable at DVB-T front-end antenna.



a) Pd=-70dBm b) Pd=-60dBm

Figure 43: DVB-T 64QAM



a) Pd=-70dBm b) Pd=-60dBm

Figure 44: DVB-T 16QAM

As a consequence, dP results from the intersection of the dotted line with the curve correspondent to the type of environment under observation. For example, dP value for a large and a small urban environment, in presence of a 16QAM DVB-T signal with -60dBm power level received at the fixed rooftop receiving antenna front-end, is respectively equal to 15 meters and 6 meters.

Finally, in Figure 45 have been reported values of the protection distance in function of the DVB-T signal power received at fixed roof antenna for both the signal modulation formats (64QAM and 16QAM) and the type of propagation environment (large and small urban environment).

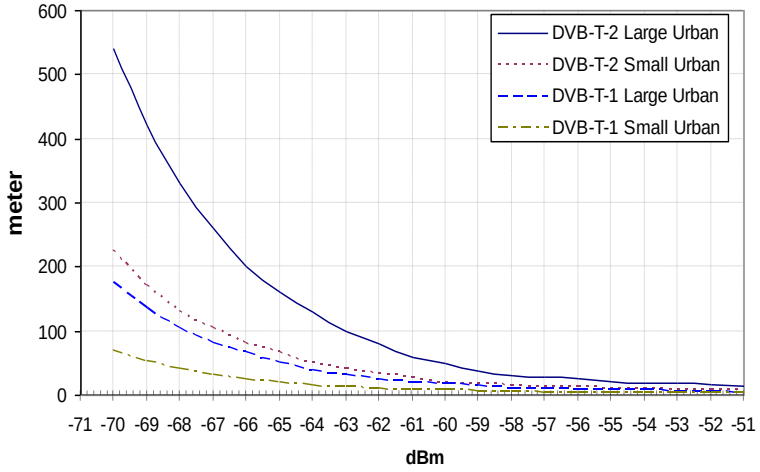


Figure 45: Protection Distance – Simulation Results

Conclusions

The huge capacity offered by the optical fiber and 4G mobile accesses brings out new aspects regarding the measurement and the exploitation of the bandwidth. In fact due to the intrinsic behavior of the Internet protocols, the effective bandwidth available to the user can be much lower than the line capacity. Furthermore the evaluation of the Quality of Service and in particular the measurement of bandwidth requires further insights, since we have to distinguish to which OSI Layer we refer to. The question is if we wish to measure either the line capacity or the available bandwidth at disposal of the user, since this issue has important consequences on the Service Level Agreement verification, especially between customers and their Internet Service Providers (ISPs).

During these studies a detailed analysis on the bandwidth measurement techniques both for fixed and mobile accesses has been realized. Initial investigations focused on the study of the throughput dependence on the network delay (RTT) for different TCP protocol implementations. In particular in this PhD dissertation, an experimental investigation on the measurement of the available ultra-bandwidth delivered to Gigabit Passive Optical Networks (GPON) users and the limitations of Internet protocols is reported. The obtained results underline that the huge capacity offered by the GPON points out the enormous differences that can be showed among the available and actually exploitable bandwidth. The main issues, linked to bandwidth exploitation and related measurement, are raised from the particular protocol implementation and from the specific protocol layer considered. In fact while the physical layer capacity can reach value of 100 Mb/s and more, the bandwidth provided to the user (i.e. either throughput at transport layer or goodput at application layer) can be much lower and strongly dependent on network delay and/or measurement implementation.

After having identified and analyzed the limitations of TCP especially for networks with high bandwidth-delay product, a TCP multisession measurement has been implemented.

This dissertation shows how it is possible to better exploit the physical layer capacity by adopting multiple TCP connections to avoid the bottleneck of a single connection. The obtained results show how this approach permits to fully exploit the line capacity in a TCP environment by means of parallel flows.

In addition, starting from TCP limitations, is underlined the requirement of a Quality of Service verification performed at several OSI layer, starting from layer 1 (line capacity), to layer 7 (represented also as quality of user experience). The goal of this approach is to fully characterize the service provided to end user.

Starting from the experience of the previous measurements, in the context of Service Level Agreements (SLAs) verification, this PhD dissertation proposes a method to verify multilayer SLA measuring simultaneously the throughput and the line capacity. The proposed Multi-Level QoS measurement has been described in Chapter 2 and both implemented and experimentally evaluated as part of this work. A detailed analysis of results is reported in Chapter 3. The proposed method performs a set of TCP and UDP measurements, giving the opportunity to characterize performance at different OSI layer and correlating them with the actual user experience. This approach enable to simultaneously measure the goodput and the line capacity in order to evaluate the effective bandwidth at disposal of the user and the one offered by the access technology provided by the ISP.

In particular the obtained results show how the proposed algorithm based on TCP/UDP (with few iterations) can permit to fully exploit the GPON capacity measuring both the goodput and the line capacity.

In order to complete the multilevel measurement suite is reported an experimental investigation on QoS measurements in terms of throughput versus access capacity, including a correlation in terms of Quality of Experience (QoE) evaluated for some Italian web TV adopting both a laboratory ultra broadband (GPON) set-up and a field 3G/HSPA mobile environment. The results carried out in this dissertation could be useful in the definition of novel Service Level Agreements among ISPs, users and content operators. These measurements are also correlated with subjective user tests in terms of MOS (Mean Opinion Score) to verify the suitability of some video web services for two opposite broadband scenarios: ultra broadband with GPON

access and mobile broadband with 3G/HSPA dongle. This investigation provides some guideline in terms of minimum average access capacity to watch television over IP.

In order to characterize mobile broadband accesses and starting from previous lab results, a measurement approach and related QoS parameters, has been investigated and proposed. This is one of the major outputs of this work, overcoming TCP limitation especially for high RTT, strongly relevant in radio mobile environment, by means of multisession test suite.

The obtained results have contributed to the implementation of the measurement suite adopted in the measurement campaigns carried out under the project AGCOM DELIBERA N. 154/12/CONS: DISPOSIZIONI IN MATERIA DI QUALITA' E CARTE DEI SERVIZI DI COMUNICAZIONI MOBILI E PERSONALI.

It is well known that wireless mobile QoS behavior is much more complex than the wireline one since much more information are needed to characterize the transmission performance due to random behavior of the radio channel.

Besides the mobile network problems, due to random behavior of the radio channel, other degradation are due to limited and shared resources between users and services; furthermore some issues related to the user behavior and in particular to the user terminals are present.

To this aim, in order to consider mainly the aspect related to the user terminal, an investigation related to the voice service, which is the most simple and consolidated service, has been carried out. Results highlight how focusing on mobile phones the QoS of mobile voice services perceived by end-users depends on the specific mobile phone model. Terminals performance were measured testing several models of mobile phone in the same network conditions, time of day and location by "drive-test", in terms of the two most important KPIs (Key Performance Indicators); the Setup Failure Rate (SFR), which is the probability that the establishment of the call is unsuccessful, and the Drop Call Rate (DCR), which is the probability that the call is suffering unnecessary shutdowns.

To conclude, the main research activity is focused on the quality evaluation of Internet access in fixed and mobile networks. This work proposes and develops a

reliable measure in both access networks, contributing to the definition of a monitoring network to measure the QoS of Internet access.

References

[1] “Qualità dei servizi di comunicazioni mobili e personali: Relazione sulla definizione delle metriche di qualità dei servizi mobili e personali e delle metodologie di misura”, Approfondimenti MisuraInternetMobile.it, FUB-AGCOM.

[2] Afanasyev, A., et al.: Host-to-Host Congestion Control for TCP. In: IEEE Communications Surveys & Tutorials, Vol. 12, NO.3, Third Quarter 2010.

[3] Janey C. Hoe, “Improving the Start-up Behavior of a Congestion Control Scheme for TCP”, laboratory for Computer Science Massachusetts Institute of Technology.

[4] Lakshman, T. V., Madhow, U., and Suter, B., “TCP/IP performance with random loss and bidirectional congestion”, IEEE/ACM TRANSACTIONS ON NETWORKING, 2000, 8, (5), pp. 541-555.

[5] D. Katabi, M. Handley, C. Rohrs, “Congestion Control for High Bandwidth-Delay Product Networks”, SIGCOMM’02, August 19-23, 2002, Pennsylvania, USA.

[6] T. V. Lakshman, Member, “The Performance of TCP/IP for Networks with High Bandwidth-Delay Products and Random Loss” IEEE, and Upamanyu Madhow IEEE/ACM TRANSACTIONS ON NETWORKING, 1997. 5, (3), pp. 336-350.

[7] Mellia, M., Lo Cigno R., Neri, F., “Measuring IP and TCP behavior on edge nodes with Tstat,” Computer Networks, 2005, 47, (1), pp. 1-21.

[8] Bolletta, P., Rea, L., “Monitoring of the User Quality of Service: Network Architecture for Measurements and role of Operating System with consequences for optical accesses” ONDM 2011 – Bologna, Italy February 2011.

[9] Bolletta, P. et al: “On the Impact of Operative Systems Choice in End-user Bandwidth Evaluation: Testing and Analysis in a Metro-access Network”, IARIA

ACCESS 2010 Sept. 2010 Valencia, Spain.

[10] Jansen S., and McGregory, A., “Measured Comparative Performance of TCP Stacks”, PAM Conference, Boston (USA) March 31-April 1, 2005.

[11] Postel, J., Reynolds, J., “File Transfer Protocol”, STD 9, RFC 959, October 1985.

[12] Postel, J., “Transmission Control Protocol”, RFC 793, September 1981.

[13] King, J., “Parallel FTP Performance in a High-Bandwidth High-Latency WAN”, SC2000, November 2000.

[14] Rea, L., Mammi, E., “Italian QoS Monitoring network: impact on SLA control”, Networks 2012, Rome October 15-18, 2012.

[15] T. Henderson, Boeing, S. Floyd, A. Gurtov, “The NewReno Modification to TCP's Fast Recovery Algorithm”, RFC 6582, Aprile 2012.

[16] M. Allman, V. Paxson, and W. Stevens, “RFC2581-TCP congestion control,” RFC, 1999

[17] R. Stankiewicz, P. Cholda, A. Jajszczyk, “QoX: What is Really?” IEEE Communication Magazine, vol. 49, iss. 4, pp. 148-158, April 2011.

[18] A. Valenti, A. Rufini, S. Pompei, F. Matera, , S. Di Bartolo, C. Da Ponte, D. Del Buono, G.M. Tosi Beleffi, “QoE and QoS Comparison in an Anycast Digital Television Platform Operating on Passive Optical Network” Networks 2012 - Roma – October 15-18, 2012.

[19] F. Matera, et al., “Comparison between objective and subjective measurements of quality of service over an Optical Wide Area network”, European Transactions on Telecommunications vol. 19, 2008, pp. 233-245.

[20] R. K. P. Mok, E. W. W. Chan, R. K. Chang, “Measuring the Quality of Experience of HTTP Video Streaming” 12th IFIP/IEEE IM 2011 Conference, Dublin May 23-27 2011.

[21] K. Seshadrinathan, R. Seshadrinathan, A. Bovik, L. K. Cormack “Study of Subjective and Objective Quality Assessment of Video” IEEE Transactions on Image Processing, vol. 17, 2009, pp. 1-16.

[22] M. Fiendler, T. Hassfeld, P. Tran-Gia, “A Generic Quantitative Relationship

between Quality of Experience and Quality of Service" IEEE Networks, march-April 2010, pp. 36-41.

[23] P. Casas, M. Seufert, R. Schatz "YOUQMON: A System for On-line Monitoring of YouTube QoE in Operational 3G Networks" ACM SIGMETRICS Performance Evaluation Review, vol.41, no.2, pp. 44-46, 2013.

[24] P. Casas, R. Schatz, T. Hoßfeld "Monitoring YouTube QoE: Is Your Mobile Network Delivering the Right Experience to your Customers?" in Proceedings of IEEE Wireless Communications and Networking Conference, Shanghai, China, 2013.

[25] L. Rea, et al, "On the Quality of Service of fixed-line broadband access network: the Italian experience", FITCE 2011, Italy.

[26] ETSI EG 202 057-4 V1.2.1 (2008-07), "Speech Processing, Transmission and Quality Aspects (STQ); User related QoS parameter definitions and measurements; Part 4: Internet Access".

[27] F. Matera, et al, "Comparison between objective and subjective measurements of quality of service over an Optical Wide Area network",, European Transactions on Telecommunications vol. 19, 2008, pp.23.

[28] Iperf (ver. 2.0.4) URL: http://sourceforge.net/projects/iperf.

[29] W. Stevens. RFC 2001: TCP slow start, congestion avoidance, fast retransmit, and fast recovery algorithms, January 1997.

[30] http://www.tcptrace.org/manual.html

[31] D. Borman R. Braden, "-TCP Extensions for High Performance", RFC 1323.

[32] S. Feyzabadi and J. Schönwälder, "Identifying TCP Congestion Control Mechanisms using Active Probing", Computer Science Department, Jacobs University Bremen, Germany.

[33] G. Forget, R. Geib, R. Schrage, "Framework for TCP Throughput Testing", RFC 6349, August 2011.

[34] M. Allman, "TCP Congestion Control with Appropriate Byte Counting (ABC)", RFC 3465, February 2003.

[35] Yang Richard, Simon Lam, “General AIMD Congestion Control”, Department of Computer Sciences”, University of Texas at Austin.

[36] MPLANE Deliverable 3.1, “Basic Network Data Analysis”, http://www.ict-mplane.eu/.

[37] MPLANE Deliverable 5.1 - First Data Collection Track Record, “Bandwidth measurements in gigabit passive optical networks”, http://www.ict-mplane.eu/.

[38] ITU-T Recommendation G. 984.1, “Gigabit-capable Passive Optical Network (GPON): general characteristics”, March 2003.

[39] A. Valenti, et al, “Experimental implementation of efficient multi cast processes: towards Carrier Ethernet networks and all-optical multi cast”, IEEE ICTON 2011, Stoccolma.

[40] S. Pompei et al, “Experimental implementation of an IPTV architecture based on Content delivery Network managed by VPLS technique”, IEEE RNDM 2010, Moskow.

[41] A. Valenti et al, “Experimental Investigation of Quality of Service in an IP All-Optical Network Adopting Wavelength Conversion” IEEE/OSA J. Optical Communications and Networking, Vol. 1, Issue 2, pp. A170-179, July 2009.

[42] F. Matera. A. Valenti, A. Rufini, “Complete Digital Television Platform Based on Optical Fiber Access Architecture”, Fotonica 2013.

[43] AA.VV. Speech Processing, Transmission and Quality Aspects (STQ); User related QoS parameter definitions and measurements; Part2: Voice telephony, Group 3 fax, modem data services and SMS, ETSI EG 202 057-2 V1.3.1, 2009.

[44] A. Rufini, E. Tarantino, M. Bianchi and G. Riva, “On the characterization of QoS perceived by end-users of mobile voice services”, Networks 2012.

[45] AA.VV. Speech Processing, “Transmission and Quality Aspects (STQ); QoS aspects for popular services in GSM and 3G networks; Part 6: Post processing and statistical methods”, ETSI TS 102 250-6 V1.2.1, 2004.

[46] S. Garcia et al, “Advanced nonparametric tests for multiple comparisons in

the design of experiments in computational intelligence and data mining: Experimental analysis of power", Information Sciences 180 (2010), Elsevier, pp. 2044-2064.

[47] J. Derrac et al., "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms", Swarm and Evolutionary Computation 1 (2011), Elsevier, pp. 3-18.

[48] W. G. Cochran, "The comparisons of percentages in matched samples", Biometrika, vol. 37, No.3/4, Dec. 1950 - pp. 256-266.

[49] H. Finner. "On a monotonicity problem in step-down multiple test procedures", Journal of the American Statistical Association 88 (1993), pp. 920-923.

[50] H. Balakrishnan, V. N. Padmanabhan, S. Seshan and R. H. Katz, "A Comparison of Mechanisms for Improving TCP Performance over Wireless Links", IEEE / ACM Transactions on Networking, December 1997.

[51] P. Karn and C. Partridge, "Improving Round-Trip Time Estimates in Reliable Transport Protocols", ACM Transactions on Computer Systems, 9(4):364–373, November 1991.

[52] R. Caceres and L. Iftod, "Improving the Performance of Reliable Transport Protocols in Mobile Computing Environments", IEEE Journal on Selected Areas in Communications, 13(5), June 1995.

[53] DELIBERA N. 154/12/CONS, "DISPOSIZIONI IN MATERIA DI QUALITA' E CARTE DEI SERVIZI DI COMUNICAZIONI MOBILI E PERSONALI".

[54] ETSI TS 102 250-5 V2.2.1 (2011-04), "Speech and multimedia Transmission Quality (STQ); QoS aspects for popular services in mobile networks; Part 5: Definition of typical measurement profiles".

[55] ETSI ES 202 765-4 V1.1.1 (2010-10), "Speech and multimedia Transmission Quality (STQ); QoS and network performance metrics and measurement methods; Part 4: Indicators for supervision of Multiplay services".

[56] ECC report 138 (2009), "Measurements on the performance of DVB-T

receivers in the presence of interference from the mobile service (especially from UMTS)".

[57] Progira Radio Communications AB (2009), "Interference from future mobile network services in frequency band 790-862 MHz to digital tv in frequencies below 790 MHz".

[58] Recommendation ITU-R SM.337-5, (2007), "Frequency and distance separations".

[59] Recommendation ETSI EN 300 744, (2009), "Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television (DVB-T)".

[60] CEPT Report 30, (2010), "The identification of common and minimal (least restrictive) technical conditions for 790-862 MHz for digital dividend in the European Union".

[61] Recommendation ETSI TS 136 104 V8.7.0, (2009), "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception".

[62] Recommendation ETSI TS 136 521-1 V8.3.1, (2009), "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) conformance specification; Radio transmission and reception; Part 1: Conformance testing".

[63] M. Celidonio, L. Pulcini, A. Rufini, "LTE and DVB-T Coexistence: A Simulation Study in the UHF Frequency Band", Journal of Communication and Computer (JCC), 2011.

[64] A. Aloisi, M. Celidonio, L. Pulcini, A. Rufini, "Experimental Study on Protection Distances between LTE and DVB-T stations operating in adjacent UHF Frequency Bands", Wireless Telecommunications Symposium, NY, 2011.

List of Acronyms

Term	Definition
NGN	Next Generation Networks
NGAN	Next Generation Access Networks
QoS	Quality of Service
QoE	Quality of Experience
SLA/ASLA	Service Level Agreement/Application Service Level Agreement
FTTx/FTTCab/ FTTB/FTTH	Fiber to the X/Fiber to the Cabinet/Fiber to the Building/ Fiber to the Home
DSL	Digital Subscriber Line
ADSL	Asymmetric Digital Subscriber Line
VDSL/VHDSL	Very-high-bit-rate Digital Subscriber Line
IEEE	Institute of Electrical and Electronic Engineers
ETSI	European Telecommunications Standards Institute
ITU	International Telecommunication Union
CEPT	European Conference of Postal and Telecommunications Administrations
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
BDP	Bandwidth Delay Product
FTP	File Transfer Protocol
HTTP	Hypertext Transfer Protocol
OSI	Open Systems Interconnection
ISP	Internet Service Provider
SSthresh	Slow Start threshold
SS	Slow Start
CA	Congestion Avoidance
CWND/CWIND	Congestion Window
RWND/RWIN	Receive Window
ACK	Acknowledgment
RTT	Round Trip Time

WAN	Wireless Area Network
FIFO	Fist In First Out
SACK	Selective Acknowledgment
GPON	Gigabit Passive Optical Networks
ADSL	Asymmetric Digital Subscriber Line
MOS	Mean Opinion Score
DSLAM	Digital Subscriber Line Access Multiplexer
SYN	Synchronization
OLT	Optical Line Terminal
ONT	Optical Network Termination
RSTP	Rapid Spanning Tree Protocol
DVB-T	Digital Video Broadcasting – Terrestrial
VLC	Video Lan Client
VLM	Video Lan Management
VPLS	Virtual Private Lan Service
SFR	Setup Failure Rate
DCR	Drop Call Rate
CSR	Call Successful Rate
KPI	Key Performance Indicator
UMCP	Unplanned Multiple Comparison Procedure
Ping	Packet Internet Grouper
ICMP	Internet Control Message Protocol
UE	User Equipment
BS	Base Station
GGSN	Gateway GPRS Support Node
MSC	Mobile Switching Centre
NAP	Neutral Access Point
IXP	Internet Exchange Point
LTE	Long Term Evolution
UMTS	Universal Mobile Telecommunications System
GSM	Global System for Mobile Communications
UHF	Ultra High Frequency
VHF	Very High Frequency
TDD	Time Division Duplexing

FDD	Frequency Division Duplexing
ECN	Electronic Communication Networks
BEM	Block Edge Mask
MIMO	Multiple Input and Multiple Output
OCR	OFF-Channel Rejection Factor
EIRP	Effective Isotropic Radiated Power