Cell Capacity of LMDS Systems in Typical Traffic Scenarios

Giovanni Iacovoni and Daniela Pierotti, Ericsson Lab Italy S.P.A.
Davide Parisi, Pharmacy Quality Europe
Marco Isopi, University of Rome "La Sapienza"

This work was carried out as part of the EMBRACE project: http://www.telenor.no/fou/prosjekter/embrace/

ABSTRACT

Broadband access and multimedia technologies are expected to be primary drivers for the telecommunications market of the next few years. The increasing bandwidth needs, together with the plethora of different services and heterogeneous traffic flows, requires accurate methodologies for resource dimensioning, especially in the field of wireless technologies where efficient bandwidth usage is crucial. New methodologies need to be based on up-to-date traffic source modeling and usage scenarios. Therefore, in this article we propose a survey of statistical characterization of single traffic sources, QoS requirements and different traffic demands. When possible, we consider the aggregate traffic resulting from a large number of heavy tailed ON/OFF sources, relying on the theory of long-range dependent flows. This method is particularly useful when time- and resource-consuming simulations are needed in order to dimension the system resources, which is always the case in the multimedia scenarios of future networks. We use such a traffic framework to estimate by simulations the capacity of an LMDS system.

INTRODUCTION

This work has been carried out as part of the European IST Efficient Millimeter Broadband Radio Access for Convergence and Evolution (EMBRACE) project. The main goal of the project consisted of developing a low-cost and effective local multipoint distribution service (LMDS) solution at 40 GHz for the mass market of advanced residential and small/medium business users.

As efficient bandwidth utilization is crucial in LMDS systems, the project needed a proper system capacity evaluation (i.e., an estimate of the maximum number of users that can be served in a cell compatibly with their traffic demands and QoS requirements). Toward this aim first we provide a survey with a complete statistical description of the traffic sources involved in Internet and traditional services. Then we propose a simulative scheme to assess the capacity of the system. As a matter of fact, due to the long-range dependence (LRD) character of Internet traffic, a full analytic treatment is beyond the reach of present techniques, so recourse to simulation is unavoidable for the treatment of realistic situations.

SYSTEM DESCRIPTION AND TRAFFIC CLASSES

The EMBRACE system has a cellular structure (Fig. 1) and operates in the 40 GHz band. Cell diameter is typically in the 2–5 km range. Each cell contains a base station (BS) serving several radio terminals (RTs), and each RT serves a certain number of users that may be organized in a LAN. All cells are connected together by a core asynchronous transfer mode (ATM) network. All traffic, apart from broadcast digital television, carried as MPEG-2 in a dedicated digital video broadcast satellite (DVB-S) transport stream, is transported in IP packets.

The duplex scheme used in the following is frequency-division duplex (FDD).

Downlink channel size is 33 MHz with 39 MHz separation. Uplink channel size is 8 MHz with 10 MHz separation. Vertical and horizontal polarizations can be used to increase capacity.

As far as bandwidth capacities are concerned, we refer to those targeted for one EMBRACE sector: a downlink capacity of 155 Mb/s and an uplink capacity of 25 Mb/s.

The *uplink* access scheme is multifrequency time-division multiple access (MF-TDMA). A set of 10 carrier frequencies, each divided in frames of 26 time slots, is used. Each RT can transmit on only one carrier frequency at a given instant of time, but can hop among different frequencies at different times.

Capacity demands to the BS are made each frame time (i.e., each 12.39 ms). The master cell controller in the BS assigns bandwidth if available, and broadcasts an allocation plan to all the RTs in terms of time slots and frequencies.

The structure of the *downlink* is time-division multiplexing (TDM).

The system guarantees quality of service (QoS) through priority traffic classes, by using

0163-6804/05/$20.00 © 2005 IEEE

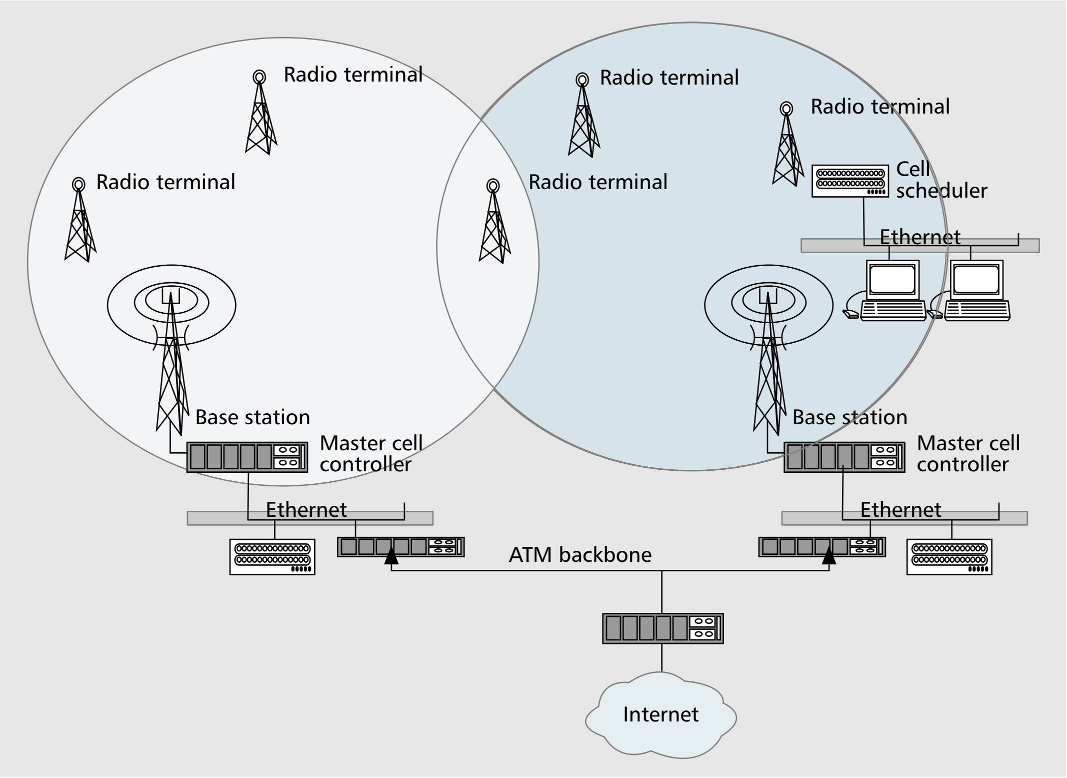

Figure 1. *The EMBRACE system architecture. For simplicity only one radio terminal has been represented in detail, but all the radio terminals are identical.*

The bandwidth is equally shared among all the requesting RTs by a round robin algorithm. If there are not enough time slots for all the RTs the slots will be assigned to randomly chosen terminals. This choice will guarantee that there are no preferred terminals.

multiprotocol label switching (MPLS) routing within the cell interconnection network, and also by diversity routing, that is, the possibility to reroute uplink traffic to the BS that provides the best communication (i.e.,carrier-to-interference ratio, C/I, or bit error rate, BER) so as to counteract heavy fading conditions due to rain.

The traffic classes are described as follows.

Continuous guaranteed (CG) class has the highest priority. Delays are bounded (the 95th percentile must be less than 50 ms). Peak bit rate (PBR) and zero losses are guaranteed during the entire connection lifetime. This class encompasses the two real-time services considered: voice over IP (VoIP) and videoconference.

Available guaranteed (AG) class is of medium priority. Delays are bounded (the 95th percentile must be less than 200 ms). PBR and zero losses are guaranteed. There is substantial difference from the previous class: the PBR, in fact, is guaranteed only for the overall duration of the ON periods. This class includes network games and streaming media (audio and video).

Best effort (BE) class has the lowest priority. Delays are bounded (the 95th percentile must be less than 500 ms), and losses must be less than 10 percent. TCP-based Internet services such as Web browsing, FTP, and file sharing belong to this class. Even if the TCP dynamically adjusts the connection bit rate, the PBR of each BE connection is supposed to be 64 kb/s.

To handle the above traffic classes, both the BS and RT are equipped with three queues, one for each traffic class.

While CG applications make use of UDP as the transport protocol, BE applications rely on TCP packet retransmission to cope with losses. Services in the AG class may be transferred using either UDP or TCP.

The upstream bandwidth is assigned according to the following strategy. The cell scheduler probes the three queues of each RT in strict priority order (i.e., a queue is inspected only once the higher-priority one has been emptied). When the first two (CG and AG) queues have been emptied, the controller considers the BE queue. The bandwidth for BE applications is assigned on a frame basis. The bandwidth is equally shared among all the requesting RTs by a round robin algorithm (one slot at a time is assigned to each RT until all the available slots are used). If there are not enough time slots for all the RTs, the slots will be assigned to randomly chosen terminals. This choice will guarantee that there are no preferred terminals.

Given this policy, the lower-priority services could starve. In our simulations, traffic load percentages are such that this case is never observed. Anyway, the solution to this problem is a call admission control (CAC) mechanism that enforces the respect of predefined bandwidth thresholds between the three classes.

Traffic Sources

In this section we give a statistical description of all the traffic sources considered in the project and refer to a single application session (e.g., a single phone call). In the next section we discuss data concerning average daily usage of a single user and traffic pattern obtained aggregating several users (belonging to the same LAN, for example).

Table 1 summarizes the collected data. The

The bandwidth assigned to VoIP applications is the one provided by the G.723.1 voice compression standard that is 5.3 kb/s during the voice activity period. The additional overhead bandwidth is represented by 40 bytes sent every 4 MAC frames, i.e., every 49.56 ms.

Class	Application	Ref.	Bit rate (kb/s)	ON periods	OFF periods	Duration pdf	Duration Mean(s)
CG	VoIP	[1]	12			Exponential	180
CG	Video-conference	[2]	387			Exponential	3600
AG	Audio streaming	[3]	20	Exponential M = 0.2 s	Deterministic M = 1.8	Pareto α = 1.6	2400
AG	Network games (residential)	[4]	10	**Packet inter-arrival times** Gumbel M=20 s σ = 6.7 s		Pareto α = 1.6	1800
BE	FTP	[8]	64	Pareto α = 1.15 M = 0.256 s	Exponential M = 15 s	Simulation time	
BE	Web (downlink)	[9]	64	Pareto α = 1.06	Exponential M = 15 s	Simulation time	
BE	Web (uplink)	[10]	64	Lognormal M =0 .041 s σ = 0.011 s	Weibull σ = 29.08 s	Simulation time	
BE	File sharing	[5]	64			Truncated Pareto α = 0.6	278.96

Table 1. *Probability distributions and parameters for all the traffic sources used in the simulations.*

data presented in this table come from either the literature or (more rarely) from guesses based on our experience. Data reported here will be subject to changes due to future charging policies.

The entry *bit rate* in the table refers to the overall application duration for CG class and to ON periods for the other classes.

In the following we describe in detail all the applications and justify the corresponding models.

CG Applications (Voice, Videoconference)

VoIP and videoconference have been modeled as constant bit rate (CBR) sources. A CG service is fully characterized by a constant bit rate and by the probability density functions (pdfs) of the application duration. The CBR reported in Table 1 has been calculated adding the bit rate required by the application itself and the bandwidth needed for the overhead introduced by the RTP/UDP/IP protocols.

The bandwidth assigned to VoIP applications is that provided by the G.723.1 voice compression standard [1]: 5.3 kb/s during the voice activity period. The additional overhead bandwidth is represented by 40 bytes sent every 4 medium access control (MAC) frames (i.e., every 49.56 ms).

The videoconference is assumed to be compliant with H.263-based video [2] with a bit rate of 384 kb/s, which is a good trade-off between quality and bandwidth usage for Common Intermediate Format (CIF) images. Encoders can enforce this bit rate even on the timescales of the MAC frame using appropriate rate control methods. The additional bandwidth needed for overhead consists of 40 bytes sent every 100 ms.

The *duration* of a phone call is known to have an exponential distribution with a mean equal to 3 min. We have assumed that the videoconference is also exponentially distributed. The mean duration has been chosen equal to 1 h.

AG Applications (Audio and Video Streaming, Network Games)

Our reference for the audio streaming application is an empirical study of the Real Audio application [3]. This is one of the most popular streaming media players.

The bit rate R of an audio flow varies between 1 and 20 kb/s, depending on the context of the flow. We choose the highest bit rate to also consider streaming of music files.[1]

Packet sizes show significant regularity and depend on the bit rate: for the bit rate considered here the packet length, L, is around 500 bytes. Within a session the data blocks are emitted according to an ON/OFF scheme; the distribution of the ON period duration is exponential with mean equal to L/R = 0.2 s. The OFF duration may be considered constant and equal to 1.8 s.

Session durations are heavy tailed. The data given in [3] are fitted well by a Pareto distribution with a shape parameter α = 1.6 and a mean equal to 40 min.

The uplink load for audio streaming reduces to a few request packets we have neglected.

[1] *This bit rate is also consistent with the more recent audio coder AMR wideband (standardized in 2001), since characterization tests shown that for music (classical and modern, both instrumental and vocal), its highest bit rate mode (23.85 kb/s) gives equivalent quality to G.722 at 56 kb/s.*

While the statistical properties of audio streaming are well described in the literature, the existing studies on video streaming do not provide suitable statistical models to be used in a simulative setting. Therefore, video streaming is not considered explicitly in our simulation. Incorporating it in our scheme is straightforward once a rich enough collection of data is available.

In [4] network games are modeled as continuous emission of small packets with short interarrival times since the traffic sent consists of the key strokes the players frequently exchange with each other to move, fire, and so on. The Gumbel distribution turns to be a good model for packet interarrival times within a game session. Session durations can be modeled by a Pareto distribution with $\alpha = 1.6$ and mean equal to 30 min.

BE Applications (Web Browsing, FTP and File Sharing)

Although there are a huge number of studies that attempt to analyze and model the load generated by these applications, unfortunately there is no general agreement on a modeling approach, and within a specific approach different papers consider sensibly different parameter values. This is not surprising since it is well known that Internet usage is not characterized by a "typical user" or "typical data patterns."

Besides, the bulk of Internet applications are transferred using TCP. TCP is based on a feedback mechanism that adjusts the source behavior according to the values of round-trip times, which, in turn, reflect the level of congestion in the network. As a result, one cannot simulate a system such as the EMBRACE access network by using data traces collected elsewhere, since that data would reflect the state and topology of that particular network in the particular moment at which they were collected.

Due to the above-mentioned reasons, we decided to model a single BE connection just by considering the most referenced data regarding the distributions of the size of the files exchanged in the Internet, according to different applications. In the next section we show how to scale them to generate an aggregate traffic load.

All BE sources have been modeled as ON/OFF sources, except for file sharing. By recalling that BE EMBRACE sources transmit at CBR during the ON period, data regarding file size are easily translated into the distribution of time duration, resulting in Table 1, where both the distributions and parameter values are reported for all the BE applications.

We now make a few comments about file sharing. File sharing is based on a peer-to-peer (P2P) communication between two remote computers. As an example, the well-known Web site Napster used to allow subscribers to freely share files stored on their own hard disks. Although Napster was closed for copyright related problems, other forms of peer-to-peer sharing are going to become a permanent feature of the Internet, eventually in the form of distributed systems (e.g., Gnutella, Kazaa, eDonkey, the current Napster, no longer free of charge).

The only reference on file sharing we have found is an informal investigation based on Internet traffic data collected at the University of Wisconsin [5]. Therein file sharing represents 27 percent of the overall data traffic. Although the particular scenario considered probably biases these data, we think these applications will occupy quite a large amount of the available bandwidth (i.e., 15–20 percent in a residential scenario).

Data on file sharing shown in Table 1 result from fitting the distribution of file size plotted in [5], where file sizes vary from a few kilobytes to hundreds of megabytes.

Email, chat, and instant messaging are not considered in the present article due to their very low QoS requirements.

> *One cannot simulate a system such as the EMBRACE access network by using data traces collected elsewhere, since that data would reflect the state and topology of that particular network in the particular moment they were collected.*

Traffic Load According to Different Scenarios

In the previous section we describe the statistical features of single applications. In this section we describe the usage patterns typical of different users on different days of the week and at different times of the day. In particular, we consider three scenarios:

- **Residential scenario**: This describes residential user habits during workdays and during evenings and nights (16:00–08:00).
- **Small office/home office (SOHO) scenario**: This concerns users working in SOHOs, on any day of the week, day and evening (08:00–24:00).
- **Business scenario**: This encompasses users during workdays and during days and evenings (08:00–24:00).

Scenarios are characterized by the involved services and the average times users spend on each service. In this article we have considered data collected in Norway [6, references therein].

Even if each scenario refers to a specific part of the day, in order to set conservative traffic estimates we focus on the *busy hour*, when 16 percent of overall daily traffic is thought to be concentrated.[2] We suppose that the statistical distributions are stationary during this period.

CG and AG services in uplink and downlink are inherently symmetric. The only exception is audio streaming, since uplink traffic can be neglected. On the other hand, BE traffic is highly asymmetric, with ratios depending on the particular application.

Single User

In this section we describe the traffic load generated by a single user.

Voice call *interarrivals* are commonly modeled by an exponential distribution. We have extended this modeling assumption to the other CG and AG services.[3] The mean values of the distributions, μ, are given in Table 2 and computed as follows:

$$\mu(s) = \frac{3600\ s}{\dfrac{0.16 \cdot T_a}{T_d}},$$

where:

[2] *A. Odlyzko, "The Low Utilization and High Cost of Data Networks," 1999, http://www.dtc.umn.edu/~odlyzko/doc/networks.html*

[3] *Plots in [3] suggest that audio streaming interarrivals can also be modeled by an exponential distribution.*

Class	Applications	Mean interarrival times (s)			Number of connections per user						Number of users per RT
					Residential		SOHO		Business		
		Residential	SOHO	Business	UP	DOWN	UP	DOWN	UP	DOWN	
CG	Voice	4500	2250	2700	–	–	–	–	–	–	**20**
	Videoconference	135000	27000	22500	–	–	–	–	–	–	10
AG	Audio streaming	9000	9000	9000	–	–	–	–	–	–	20
	Network games	13500	–	–	–	–	–	–	–	–	10
BE	FTP	–	–	–	1	2	2	6	2	8	20
	Web	–	–	–	1	1	3	3	4	4	40
	File sharing	3488	3488	3488	1	3	1	4	1	4	10

Table 2. *Figures on applications during the busy hour (used in the EMBRACE project), when applicable.*

- T_a is the average time a user spends on the application during the considered period of the day (all the values come from [6] except our guess for the audio streaming).
- T_d is the mean duration of one connection (from Table 1).
- $0.16T_a$ is the average time a user spends on the application during the busy hour.

BE applications are more complicated. We assume that FTP and Web browsing are continuously performed during the busy hour, and an exponential interarrival distribution for the file sharing service since it has a model similar to the one used for CG class. By taking the reasonable value of T_a = 30 min, the mean of the interarrival times for file sharing results in the value written in Table 2.

Moreover, we considered that a user may set up several connections at the same time (e.g., loading two Web pages simultaneously), and that the number of simultaneous connections depends on the scenario. The traffic load is different in the two directions (uplink and downlink). This is obviously the case for Web browsing, but an asymmetry can also be reasonably present in the other applications.

Aggregate Traffic per RT

Traffic transmitted and received by each RT results from the aggregation of the loads generated or received by single users.

The aggregation of CG and AG traffic is straightforward once the number of users N performing each service is given.

In the following we describe how to generate the resulting aggregate BE traffic by using the mathematical result which states that aggregating a large number of independent ON/OFF sources with heavy tail distributed ON and/or OFF periods, and with exponent $1 < \alpha \leq 2$, one obtains fractional Gaussian noise (FGN). FGN is a Gaussian model with the LRD property.[4]

An FGN model is expected to be rigorously valid only when the number of superposed sources is large, and it is possible to neglect the TCP effects. The first condition is largely verified in the system under observation. The second hypothesis needs to be further discussed.

As a TCP source adjusts its rate according to network congestion, the decrease of its rate can be considered as if there was a nonempty queue at the sender side, which thus introduces a waiting time. According to teletraffic expert Ilka Norros (VTT, Finland), this queue can be regarded as a *virtual queue*. These virtual queues are distributed among the sources and are not directly observable. This *source buffering* acts as virtual storage, alleviating the *router buffering*. This observation suggests that FGN may be used to model data traffic even when the sources are subjected to TCP rate control. Recent work (e.g., [7]) confirms that FGN can indeed model in a simple and correct way file transmission when there are no dominant connections. This is the case in the EMBRACE system since the mechanism of round robin tends to prevent any users from jamming the available resources.

The FGN is fully described by the mean, variance, and Hurst parameter (representing the degree of correlation over time of FGN), which are given by:

- $M = NCpR$ (1a)
- $\sigma^2 = NCp(1-p)R^2$ (1b)
- $H = (3 - \alpha)/2$ (1c)

R is the bit rate during ON periods,

$$p = \frac{E(T_{ON})}{E(T_{ON}) + E(E(T_{OFF})},$$

α is the exponent of the distribution of the ON periods, and $1 < \alpha \leq 2$.

We used FGN traces to model FTP and Web aggregated downlink traffic. File sharing traffic instead has been generated using the same methodology employed for CG and AG traffic, since the exponent α of the Pareto distribution is less than 1 and therefore the above result on aggregation does not apply.

[4] *LRD for BE traffic has emerged from statistical analysis of measurements made in several working packet networks. These processes have slowly decaying correlations and therefore persist for long periods in a state or region of the phase space, in sharp contrast with the exponential decay of correlations of Markovian processes.*

CAPACITY EVALUATION

In this section we first describe the methodology we used to simulate the EMBRACE cell and then give results regarding capacity evaluation.

SETUP AND SIMULATION METHODOLOGY

By OPNET® we implemented a system cell with the appropriate MAC as described in the EMBRACE system requirements.

In order to generate the uplink and downlink traffic for each RT we have used the data given in Tables 1 and 2. Even if the values of the number of users N is somewhat arbitrary, it is driven by the EMBRACE target [6], which is to aim for widespread use of the technology by having each RT shared by as many users as possible. As far as the number of connections C per user is concerned, there is no available data we could use. Therefore, we made an educated guess based on our observations and estimates of the BE service usage in different scenarios.

We have simulated a single scenario at a time. The number of Web and FTP applications performed by each RT gives the most significant differences between the scenarios. Results are shown in Figs. 2 and 3 and Table 3, and explained in the rest of the article. Every simulation point is the average of five simulation runs. The confidence intervals are too small to appear in the figures. This is due to the large number of events that occur in the simulated time (1 h) of each simulation run. Confidence intervals are discussed below.

In order to evaluate the maximum number of RTs served by a BS in a sector, we performed simulations with an increasing number of RTs until QoS requirements for BE traffic are not fulfilled anymore. In fact, due to priorities, as long as any BE traffic is served, CG and AG traffic fully respect the QoS parameters. At each simulation trial (i.e., for a fixed number of RTs) we tune the uplink and downlink BE buffers length by trial and error in order to fulfill the above-mentioned BE QoS requirements. Considering constant throughput for each RT sets the initial buffer size. Then this value is changed trading loss for delays.

The actual uplink and downlink capacities used in the simulations are 90 percent of the total capacities. Ten percent of the bandwidth, in fact, is left for the "shadow flows" created by those RTs for which BS diversity is provided.

Besides, according to the different traffic scenarios the total downlink capacity is decreased by the amount reserved to DVB channels. For n DVB channels $7 \cdot n$ Mb/s are reserved. In our setting $n = 7$ for the residential scenario, and $n = 3$ for both SOHO and business scenarios.

In our models RTs are independent of each other and have identical traffic distributions.

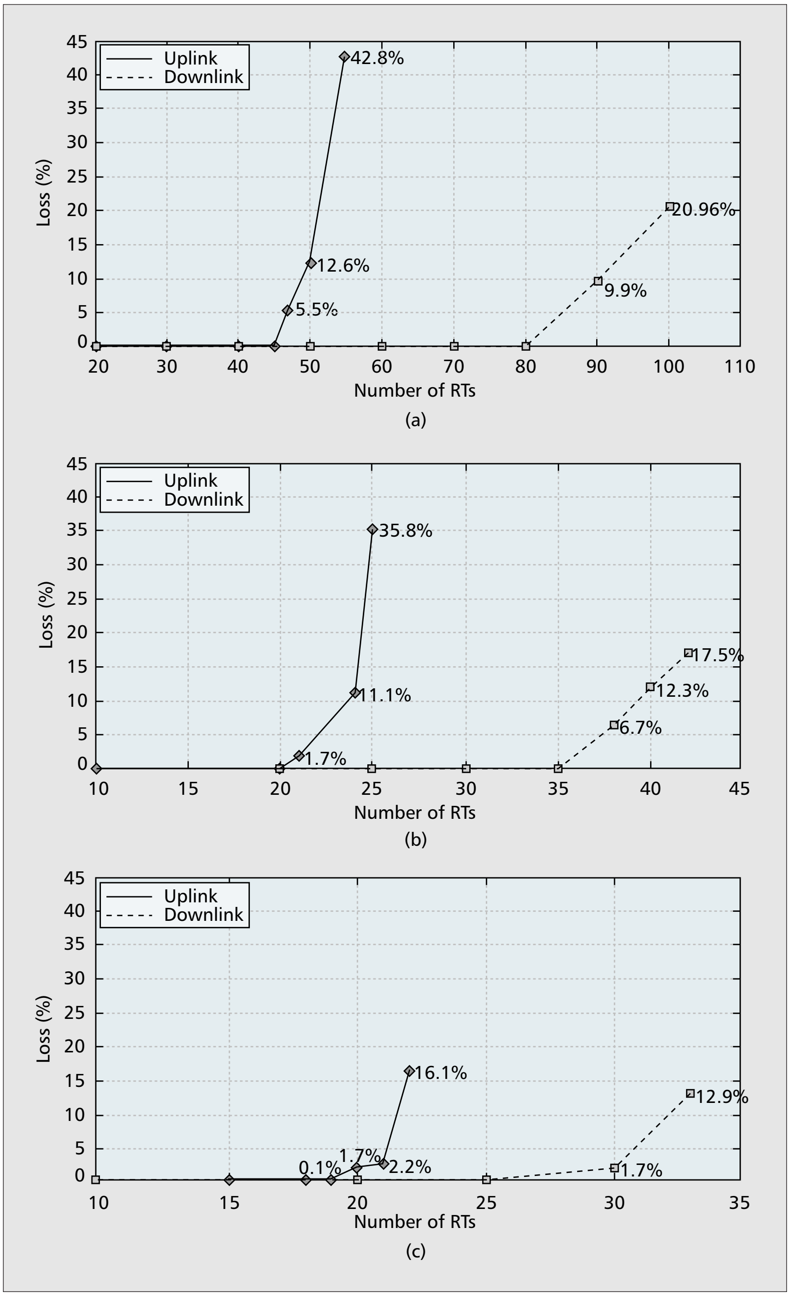

Figure 2. *BE losses (percent) for uplink and downlink traffic, for a) residential; b) SOHO; c) business scenarios.*

RESULTS AND CONCLUSIONS

Figure 2 shows the percentage of BE packet losses versus the number of active RTs for uplink and downlink, for all the three scenarios. Losses are caused by buffer overflow. Since the IP packet is segmented into several MAC cells, it is enough that only one segment is lost to waste the entire packet.

By comparing Fig. 2 with the corresponding plots concerning delays (not shown here due to lack of space), we have found that the maximum number of RTs due to loss constraints is smaller than that due to delay constraints for all the scenarios. Consequently, our criterion for cell capacity evaluation is based on BE packet losses.

In Fig. 2 one can notice that uplink and downlink loss percentages behave differently: while downlink losses vary slowly and linearly with the number of RTs, uplink losses show abrupt changes after a "critical" number of RTs. This kind of behavior is due to LRD. Very large buffers used in the downlink together with the stronger statistical multiplexing are likely the cause for the sizable reduction of this effect in the downlink direction.

® *OPNET Modeler Version 7 is a registered trademark of OPNET Technologies Inc.: http://www.opnet.com.*

	Residential		SOHO		Business	
	Uplink (kb/s)	Downlink (kb/s)	Uplink (kb/s)	Downlink (kb/s)	Uplink (kb/s)	Downlink (kb/s)
CG	112	112	531	531	630	630
AG	13	24	0	11	0	11
BE	81	697	116	1889	129	2319
Total	206	833	647	2431	759	2960

Table 3. *Average traffic generated by an RT during the busy hour. Bit rates are net; that is, they do not take into consideration the bandwidth used for overhead (MAC and physical layers).*

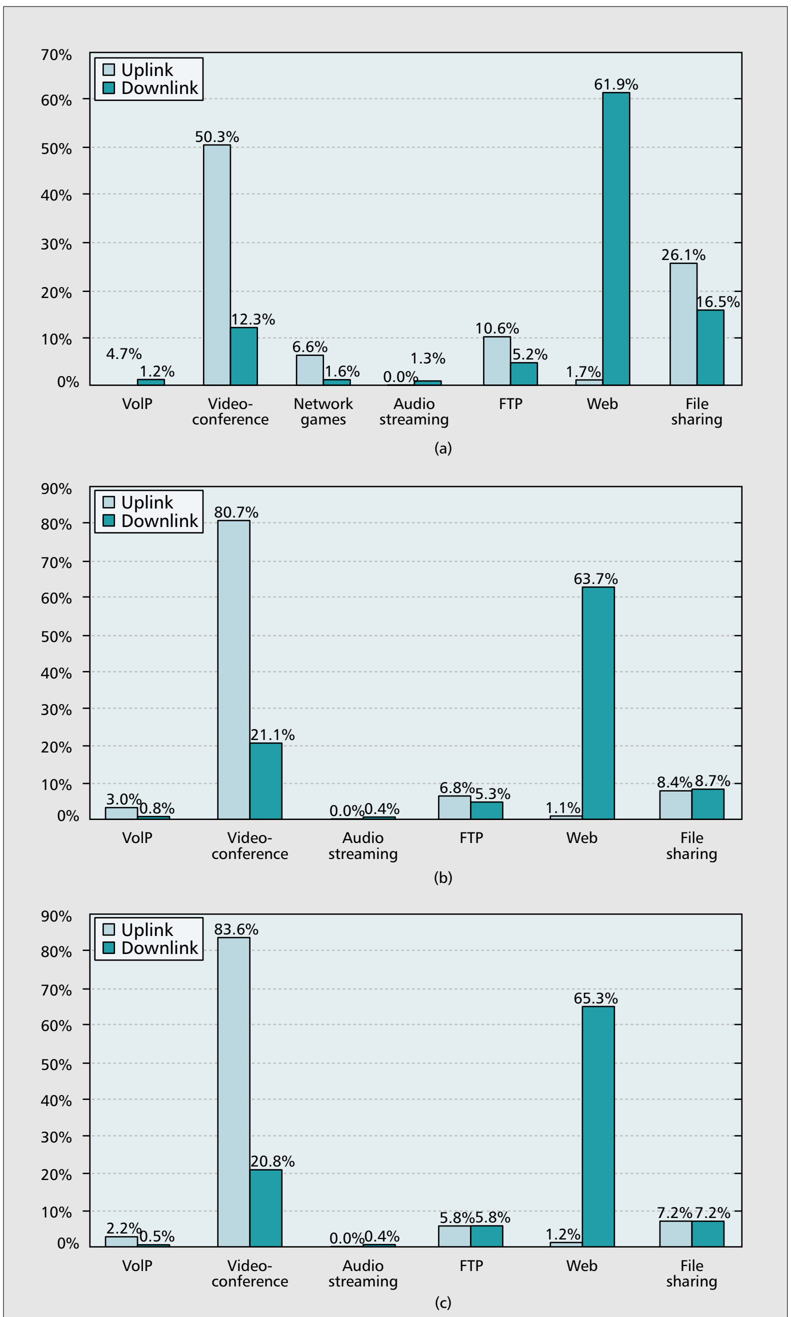

Figure 3. *Mean percentages of total RT bandwidth (for both traffic directions) used by each service for a) residential; b) SOHO; c) business scenarios.*

In order to take into account the above-mentioned difference we use two different criteria to settle the maximum number of RTs that can be deployed in a cell.

In the downlink we impose that the number of terminals for which losses are 10 percent or more is the 95th percentile of the distribution of active terminals. This criterion guarantees that the probability of having more than 10 percent loss is only 0.05. For instance, this criterion leads to accepting a maximum of 90 active RTs in the residential scenario.

In Fig. 2 one can notice that uplink losses show abrupt changes with the number of active terminals. It is therefore advisable to choose a number of RTs smaller enough than that for which the onset of this behavior is observable and losses increase dramatically. For this reason we impose that the 95th percentile of the distribution of active terminals coincides with the rightmost point in the graph for which we observe zero losses. In this way we guarantee that with probability 0.95 we are far from the critical point and have zero losses. For instance, this criterion leads to accepting a maximum of 46 active RTs in the residential scenario.

In order to evaluate the actual number of RTs that can be deployed in a cell, we need to consider that not all the RTs are active simultaneously. An RT is considered active if it is sending or receiving traffic. The probability for an RT to be active at a given time is $p = 0.25$, and each terminal is activated independent of the others, so the number of active terminals has a binomial distribution.

In conclusion, in a residential scenario, given the downlink bandwidth, an operator could deploy up to 311 RTs in a cell, but the uplink capacity cannot support more than 151 RTs.

We observed a similar result for the other two scenarios: in the SOHO setup we obtained for the downlink and uplink capacity 126 and 59 RTs, respectively, while in a business scenario we obtained 101 and 53 RTs.

With the above operating points, for all three scenario, the transmission buffer sizes are set to 25 kbytes in the RT and 5 Mbytes in the BS.

Therefore, with the given uplink and downlink sector bandwidth and scenarios, it would be an efficient solution to deploy two uplink sectors for each downlink one.

The mean traffic loads for each RT are shown in Table 3. One can notice that the uplink traffic generated by an RT consists mainly of CG applications (i.e., real-time services), while downlink traffic is mostly jammed by BE services that are inherently asymmetric. AG class is almost empty, although it encompasses services that are expected to grow, such as video streaming (which we have not simulated for the reasons explained above).

When the simulation encompasses the maximum number of active RTs, the bandwidth is shared among the different applications as shown in Fig. 3, where the percentages of the total uplink and downlink bandwidth are given for all the services.

In this setting, as far as confidence intervals are concerned, let us consider in Fig. 3, residential scenario, the most demanding application, the Web downlink, modeled by the FGN process as explained above. From standard theory, for one simulation run lasting T s, the standard deviation of the sample mean $\overline{X}$, $stdv[\overline{X}]$, is equal to

$$\sigma\left(\frac{T}{t_s}\right)^{H-1},$$

where t_s is the simulation timescale. For 90 active RTs, by using Eq. 1 and the parameter values shown in Table 1, and recalling that t_s = 12.39 ms, we get $stdv[\overline{X}]/M = 0.022$. Considering five simulation runs further reduces variability. Therefore, in Fig. 1, for 95 percent confidence interval, the mean bandwidth for Web downlink, residential scenario, can deviate by approximately 1 percent. Since all other applications are less demanding in terms of bandwidth and/or exhibit less variability, all the other confidence intervals are smaller. Similar conclusions hold for the other scenarios.

In conclusion, this work shows how freely available data on Internet traffic and recent advances in traffic modeling can be effectively used in dimensioning network elements before production and deployment. The pace with which new usage patterns of traditional and emerging services change will require updates of the actual values of the parameters.

This work shows how freely available data on Internet traffic and recent advances in traffic modeling can be effectively used in dimensioning network elements before production and deployment.

References

[1] ITU-T Rec. G.723.1, "Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kb/s," Mar. 1996.

[2] ITU-T Rec. H.263, "Video Coding for Low Bit Rate Communication," Feb. 1998.

[3] A. Mena and J. Heidemann, "An Empirical Study of Real Audio Traffic," *Proc. IEEE INFOCOM*, Tel Aviv, Israel, Mar. 2000.

[4] M. Borella, "Source Models of Network Games Traffic," *Proc. Networld+Interop '99 Engineer's Conf.*, Las Vegas, NV, May 1999.

[5] "Internet Traffic Flow Size Analysis," http://net.doit.wisc.edu/data/flow/size

[6] IST-1999-11571 EMBRACE D2, "System Requirements," http://www.telenor.no/fou/prosjekter/embrace/Open_Publications/Deliverable_D2.pdf

[7] X. Wang *et al.*, "Network Traffic Modeling Using Connection-Level Information," *Proc. DIMACS Wksp. Internet and WWW Measurement, Mapping and Modeling*, Rutgers, NJ, Feb. 2002.

[8] A. K. Jena, P. Pruthi, and A. Popescu, "Resource Engineering for Internet Applications," *Proc. 7th IFIP ATM Wksp.*, Antwerp, Belgium, June 1999.

[9] M. Molina, P. Castelli, and G. Foddis, "Web Traffic Modeling Exploiting TCP Connections' Temporal Clustering through HTML-REDUCE," *IEEE Network*, vol. 14, no. 3, May/June 2000.

[10] S. Khaunte, and J. Limb, "Statistical Characterization of a World Wide Web Browsing Session," Tech. rep. GIT-CC-97-17, College of Comp., GA Inst. of Technology, 1997.

Biographies

GIOVANNI IACOVONI (giovanni.iacovoni@ericsson.com) received his Laurea degree in electronic engineering from the University of Rome "La Sapienza" in 1989 and a Ph.D. in electrical engineering from the University of Pavia in 1998. He has been working in Ericsson Lab Italy since 1990. Currently he is a technical coordinator in the Research and Innovation Unit. His current research interests are in digital image/speech compression and transport, video transcoding for real-time applications, TCP analysis, traffic modeling, and cell capacity evaluation of both wired and wireless access systems.

MARCO ISOPI received his Laurea degree in mathematics from the University of Rome and his PhD. degree in mathematics from New York University. He was attache de recherche at Ecole Polytechnique, Paris. He currently teaches mathematics at the University of Rome "La Sapienza." His work is in the theory of stochastic processes and applied probability.

DAVIDE PARISI received his degree with honors in electronic engineering from the University of Rome "Tor Vergata" in 1999. He worked with the research consortium CoRiTel and with Ericsson Lab Italy. Now he is with Pharmacy Quality Europe (PQE). ßHis main research interests are in wireless access systems, Internet application requirements, traffic sources modeling, medium access control, and TCP.

DANIELA PIEROTTI received her Master's degree and Ph.D. (physics) from the University of Rome "La Sapienza." Since 2000 she has been working as a system engineer in the Research and Innovation Department at Ericsson Lab Italy. Her main interests are in wireless telecommunications, statistical analysis and modeling of IP traffic, and network synchronization.