TRAFFIC TRUNK PARAMETERS FOR VOICE TRANSPORT OVER MPLS

A. Estepa, R. Estepa and J. Vozmediano
University of Sevilla
C/ Camino de los descubrimientos s/n

Keywords: Voice transport, MPLS Traffic Engineering, VoMPLS, VoIP.

Abstract: Access nodes in NGN are likely to transport voice traffic using MPLS Traffic Trunks. The traffic parameters describing a Traffic Trunk are basic to calculate the network resources to be allocated along the nodes belonging to its corresponding Label-Switched-Path (LSP).
This paper provides an analytical model to estimate the lower limit of the bandwidth that needs to be allocated to a TT loaded with a heterogeneous set of voice connections. Our model considers the effect of the Silence Insertion Descriptor (SID) frames that a number of VoIP codecs currently use. Additionally, two transport schemes are considered: VoIP and VoMPLS. The results, experimentally validated, quantify the benefits of VoMPLS over VoIP.

This work was supported in part by the Spanish *Secretaría de Estado de Universidades y Educación* under the project number TIC2003-04784-C02-02

1 INTRODUCTION

Voice transport over New Generation Networks (NGN) will likely make use of QoS-supporting packet-switching networks. Multi-Protocol-Label Switching (MPL, 2001)(MPLS) is a packet forwarding technique that facilitates the creation of Label-Switched-Paths (LSPs) and allows the use of traffic engineering needed to support the provision of QoS at an optimal cost.

The traffic engineering (TE, 1999) inherent capability of MPLS allows to dynamically route a set of forwarding equivalence classes over a so-called Traffic Trunk (TT) which follows the most adequate path according to its traffic characteristics, available resources in the network and administrative criteria. For the remainder of this paper, a TT will be used to transport a set of voice streams which demand the same QoS and follow the same LSP between two Label-Edge Routers (LERs) as indicated in figure 1.

In order to provide a TT with traffic engineering capabilities, the source LER needs to be aware of a number of its characteristics. Among them, we are interested in the traffic parameters (e.g. mean and peak bit-rates), which can be calculated from the traffic characterization of each voice stream belonging to the TT. These traffic parameters are required to calculate the capacity to be reserved for the TT in each link of the MPLS network and to develop a faithful map of the overall capacity remaining free in the network. Consequently the traffic parameters are a basic input to any constrained-routing algorithm.

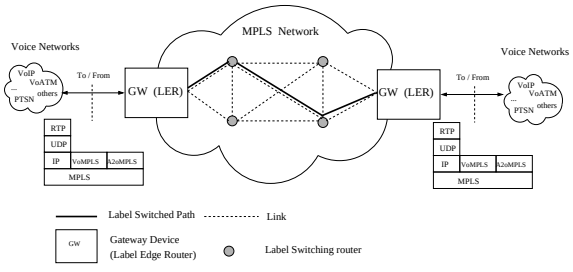

Figure 1: Sample scenario.

The methods used to calculate the optimal capacity to be reserved are usually based in complex analytical models (R. Guérin and Naghshineh, 1991) and are out of the scope of this paper. However, the bandwidth reservation should range from the sum of the conversation's mean bit-rates (stability condition) to

Estepa A., Estepa R. and Vozmediano J. (2006).
TRAFFIC TRUNK PARAMETERS FOR VOICE TRANSPORT OVER MPLS.
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications*, pages 58-64
DOI: 10.5220/0001572800580064

the sum of its peak bit-rates[1]. A common and simple approach to calculate the actual bandwidth reservation is to use the sum of the conversation's peak bit-rates and take this upper limit as the allocation to be requested for the TT. This guarantees that no packet loss occurs at the cost of some over-provisioning of network resources. However, this conservative peak bit-rate approach could potentially cause the rejection of a new TT in the LER's admission procedure (CAC,) in spite of having enough capacity.

As the number of multiplexed sources in the TT increases, the traffic burstiness is smoothed and the capacity to be reserved should gradually change from the peak bit-rate approach to the mean bit-rate approach, thus making a more effective use of the network resources. However, the current calculus (B, 2002) of the mean-bit rate of the TT is inaccurate, since it is based in the ON-OFF model which does not consider the generation of Silence Insertion Descriptor (SID) frames that a number of voice codecs generate during voice inactivity periods (Estepa et al., 2003). These SID frames mark the end of talkspurts and update the Comfort Noise Generation parameters at the receiver.

Starting from the previous results in (Estepa et al., 2003), we find a more accurate analytical expressions for the traffic parameters of a TT (i.e. mean and peak bit-rates) transporting a set of heterogeneous voice sources when SID-capable codecs are used. We apply them to two possible voice transport schemes: VoIP over MPLS and VoMPLS [2]. This would facilitate the use of the mean bit-rate value as a reference for a effective resource allocation in traffic engineering. In addition, the comparison between these different transport schemes (i.e. VoIP and VoMPLS as observed in figure 1) will let us to assess the bandwidth savings of VoMPLS over VoIP. Our results could be also applied to optimize the off-line analysis of packet loss and delay by using the analytical models to provide a desired QoS level as a function of both the TT mean bit-rate and the number of sources to be multiplexed.

The rest of the paper is structured as follows: section 2 sets the basic models to transport voice over an MPLS cloud and establishes the TT model used throughout the paper. Section 3 calculates the maximum and minimum capacity allocation for a voice TT in a VoIP over MPLS and VoMPLS scenario. Section 4 presents the main results and finally, section 5 concludes the paper.

[1]Within that range, the capacity selected represents a balance between the maximum burst size and the probability of out-of-profile.

[2]The case of A2oMPLS is not addressed in detail because the current implementation agreement does not specify the packetization scheme for the SID frames. However, the findings presented for VoMPLS are still valid for A2oMPLS with some minimum changes

2 MODELS FOR VOICE TRANSPORT IN MPLS

This section addresses two subjects: the characterization of a voice source traffic a in a digital environment, and the means of transporting a set of those conversations belonging to a TT over an MPLS network. Conversely to previous studies, we will not use the ON-OFF model but the more general ON-SID model presented in (Estepa et al., 2005). The main reason for this is the inadequacy of the ON-OFF model to capture the effect of the SID frames in the conversation's mean bit-rate.

2.1 Single Voice Source Model: The ON-SID Model

Low bit-rate codecs are commonly used in the transport of voice over packet-switched networks. Typically, these type of codecs analyze the speech samples generated during a period of time T and generate a information data-unit termed *frame* that can be used at the receiver to faithfully restore the original sequence of speech samples. Low bit-rate codecs are usually equipped with a voice activity detection (VAD) feature which pursues bandwidth savings by avoiding the generation of frames during voice inactivity periods.

Additionally, some audio codecs like G.729, G.723.1 or AMR are also featured with an algorithm which allows, at the beginning of each voice inactivity period, to send SID frames. Reception of a SID frame after a voice frame can be interpreted as an explicit indication of the end of the talk-spurt. In addition, SID frames may be also transmitted at any time during the silence interval to update comfort noise generation parameters. This allows a faithful reproduction of the background noise at the receiver's side, increasing the quality of the conversation at the cost of some additional bandwith (Estepa et al., 2003).

Thus, the voice traffic model to be used in the remainder of this paper will not be limited to the traditional ON-OFF model, but the more general ON-SID model. This model assumes that in the discrete time space $t_i = i \cdot T$ (where T is the codec's frame generation period), the codecs continuously generate frames which can be either of type: ACT (compressed voice), SID (background noise) or NoTX. The latter corresponds to a zero-length frame used to model instants when no frames (ACT nor SID) are being generated. ON and SID periods are exponentially distributed. During voice activity periods ACT frames are generated every T seconds. During voice inactivity

periods, SID frames are generated randomly according to the codec's specific algorithm and to changes in the background-noise signal. Since SID frames generation is a random process, we can use a discrete random variable, X, to indicate the inter-arrival time (in number of periods T) between SID frames as expressed in figure 2. Moreover, we assume that SID frame generation is a renewal process.

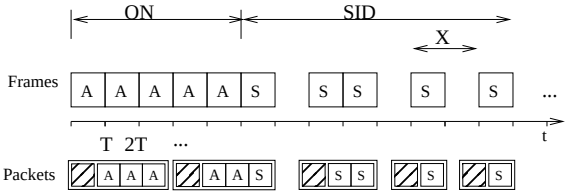

Figure 2: ON-SID frame generation model for VoIP and N_{fpp}=3.

Additionally, we also assume that during voice activity periods, to compensate the excess of overhead of layer protocols (H), ACT frames are usually sent to the network in groups of N_{fpp} consecutive frames per packet. Note that this also causes a packetization delay that limits the maximum acceptable value of N_{fpp}.

2.1.1 Mean Bit-rate of a Single Voice Source

According to the ON-SID model, a packetized voice stream transmitted with VAD capable codecs which transmit SID frames exhibits a mean bit rate of:

$$r = \rho \cdot p + (1 - \rho) \cdot r_{SID} \quad (1)$$

where ρ is the conversation mean activity rate, p is the peak rate and r_{SID} is the mean rate during voice inactivity periods caused by the transmission of SID frames. For those codecs which do not generate SID frames, obviously $r_{SID} = 0$.

The factors of equation 1 depend on the transport scheme used (i.e. VoIP and VoMPLS,) and will be addressed in next subsections.

2.2 Alternatives for Voice Transport in MPLS

This paper addresses two possible ways of voice transport over an MPLS TT, depending on whether the tributary conversations come from VoIP or are directly taken from the payload of the VoIP packets; that is, the transport of codec frames directly over MPLS or VoMPLS.

2.2.1 VoMPLS

The implementation agreement defined by the MPLS-FrameRelay Alliance (VoM, 2001) describes how to transport voice directly over MPLS. The method is illustrated in figure 3 and can be summarized in the following ideas:

- A number of voice calls may be transported over an LSP. The multiplexing structure consists of a mandatory Outer Label, zero or more Inner Labels, and one or more VoMPLS Primary Subframes consisting of a 4-octet Header (HDR) and variable length Primary Payload each, as shown in figure 3.
- Each Primary Subframe may be associated with a different voice connection. A Primary Payload is either a sequence of encoded ACT voice frame(s) or a single SID frame.
- Within the header of a Primary Subframe, the length field is indicated in multiples of 4 octets. Thus, up to three padding octets may be inserted in each subframe depending on the codec's frame size and the number of codec's frames carried in the Primary Subframe.
- A Primary Payload contains the traffic that is fundamental to the operation of a connection identified by a Channel Identifier (CID). It includes ACT and SID frames. Primary Payloads are variable-length subframes.
- Control Subframes may be sent to support the Primary Payload (e.g., dialled digits for a primary payload of encoded voice) and other control functions (like RTP-timestamps). Control and Primary Subframes are not mixed together in the same multiplexing frame. Thus, Control Subframes will not be considered in the present study since they belong to the signalling plane.

The header's Channel ID (CID) allows up to 248 VoMPLS calls to be multiplexed within a single LSP so the Inner Labels will not be considered in the rest of the paper.

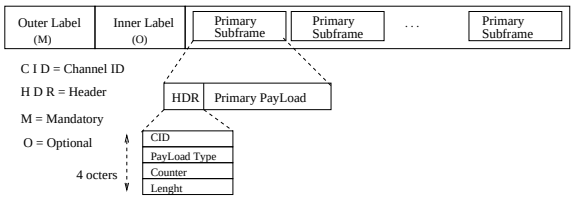

Figure 3: VoMPLS traffic trunk format.

The aforementioned implementation agreement also establishes the maximum N_{fpp} allowed value for each codec (e.g. in VoMPLS there is a maximum value of N_{fpp}=6 for the G.729B codec, while for the G.723.1 codec N_{fpp} is forced to be 1).

2.2.2 VoIP over MPLS

The protocols involved in the transport of IP packets are the Real Time Protocol (RTP) and the UDP protocol, resulting in a RTP/UDP/IP header of 40 octets per each IP packet.

Packets are generated every $T \cdot N_{fpp}$ seconds during voice activity periods. During voice inactivity periods, SID frames are packed according to the RFC 3551 packetization scheme, where only those SID frames consecutively generated may be carried in the same packet, up to the maximum of N_{fpp}.

Each VoIP stream is multiplexed in a Subframe of the MPLS TT multi-frame structure. A new TT multi-frame is sent whenever any conversation of the TT needs to send a new IP packet. The multi-frame has an Outer Label as indicated in next section.

2.3 Aggregation of Heterogeneous Voice Sources in a Traffic Trunk

For the aggregation of the voice sources in a single TT we consider the multiplexing model illustrated in Figure 4, where a set of K classes of m_i homogeneous ON-SID sources feed a multiplexer (which can be a LER) where a TT of real-time voice sources is created. Each one of the m_i voice streams belonging to the same class i will have a common value of N_{fpp} and codec, and consequently, the same ON-SID parameters; namely, the peak bit-rate (p_i) and mean bit-rate (r_i).

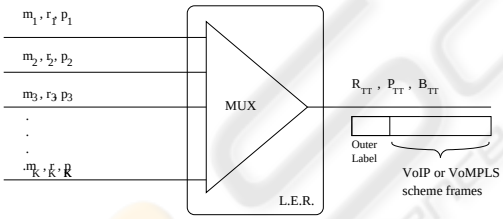

Figure 4: Voice multiplexing model.

According to figure 4, the traffic characterization should include the Outer Label of the TT. Therefore, traffic parameters defining the traffic profile are:

- Traffic Trunk's Mean Bit-rate: this parameter is the minimum service rate that guarantees stability in the system and thus, is the minimum capacity that should be allocated for the TT. It includes the sum of all the conversations mean bit-rate for the class i plus the mean bit-rate caused by the Outer Label of the TT (R_{OL}).

$$R_{TT} = \sum_{i=1}^{K} m_i \cdot r_i + R_{OL} \qquad (2)$$

- Traffic Trunk's Peak bit-rate: is the sum of all the peak rates plus the peak bit-rate of the Outer Label of the trunk supposed that all sources are ON (P_{OL}). This is the maximum capacity that should be allocated to the TT to avoid packet loss.

$$P_{TT} = \sum_{i=1}^{K} m_i \cdot p_i + P_{OL} \qquad (3)$$

- Traffic Trunk burst-size: this parameter can be considered to be free. The reason for this is that the allocated capacity (C) for the TT must be a value greater than R_{TT} (to be stable) and smaller than P_{TT} (to take advantage of statistical multiplexing). For C=P_{TT}, the buffer size (B) needs to be only big enough to store one voice packet from each conversation, while for C=R_{TT}, B should be large enough in order to queue all the instant traffic in order to bound the potential packet loss. An excellent paper reviewing this tradeoff is (Procissi G, 2002).

Since the relation between B and the QoS depends on the multiplexing analytical model used in the study (i.e. either fluid model or MMPP), our goal is to find the values of P_{TT} and R_{TT} for VoMPLS and VoIP trunks. The next section is devoted to this task.

To account the Outer Label influence in the TT mean bit rate, we make the following assumption: a new MPLS frame is generated every T_{min}=$min\{i = 1, ..k; Ti\}$ whenever there is any source generating a new frame (i.e. voice frames or SID frames). For the peak bit-rate calculation, we assume that a new trunking frame is generated every T_{min}. Thus, the values of P_{OL} and R_{OL} result as follow:

$$P_{OL} = \frac{H_{OL}}{T_{min}} \qquad (4)$$

$$R_{OL} = \frac{H_{OL}}{T_{min}} \cdot G_{TX} \qquad (5)$$

where G_{TX} is the probability of having at least one voice source generating a new frame at T_{min}.

3 NEW VALUE OF THE TRAFFIC PARAMETERS FOR MPLS TRANSPORT

This section is devoted to finding out the analytical expression for the traffic parameters as indicated in equations 3 and 2 of previous subsection. In our approach, we first find analytical expressions for p_i and r_i for both transport schemes under study: VoIP over MPLS, and VoMPLS.

Table 1: Codec characteristics.

Codec	Mode	L_{ACT}	L_{SID}	T (ms)	$E[X]$	P_1
G.729	-	10	2	10	7.33	0
G.723.1	6.3	24	4	30	13.05	0.27
	5.3	20	4	30		
AMR	4.75	12	5	20	7.47	0
	12.2	31	5	29	7.47	0

3.1 Mean and Peak Bit-rate for VoIP

The peak rate of a VoIP conversation depends upon both the codec characteristics and the number of frames per packet (N_{fpp}). Thus, it is clearly given by:

$$p_i = \frac{H + N_{fpp}L_{ACT}}{N_{fpp}T} \tag{6}$$

where H is the header size of the protocol layers involved in the transport service (i.e. 40 octets), L_{ACT} is the voice frame size and T is the frame generation period of a given codec. Table 1 shows the characteristics of some VoIP codecs.

Regarding the r_{SID} member of equation 1, an analytical expression for the VoIP transport may be found in (Estepa et al., 2005). The deduction was based in the separation of the contribution of the header and the SID frames to the mean bit-rate so $r_{SID} = R_H + R_{fr}$. The contribution of the SID frames can be obtained by application of the Elementary Renewal Theorem (ERT) which states that the SID frames arrival long-term rate is the inverse of the expected inter-arrival time ($E[X] \cdot T$).

$$R_{fr} = \frac{L_{SID}}{T \cdot E[X]} \tag{7}$$

where L_{SID} is the size of a SID frame.

In VoIP, the contribution of the packet header generated during inactive periods follows the packet generation pattern imposed by the RFC 3551, where one packet header is sent every non-consecutive SID frame ($X = x > 1$). For consecutive SID frames, one packet header is sent every N_{fpp} frames, so both cases must be considered. Since the mean time between SID frames is given by ($E[X] \cdot T$), the header contribution (R_H) can be expressed as:

$$R_H = P_1 \cdot \frac{H}{N_{fpp} \cdot T \cdot E[X]} + (1 - P_1)\frac{H}{T \cdot E[X]} \tag{8}$$

where P_1 stands for the probability of having two time-consecutive SID frames (i.e. $P(X = 1)$).

Thus, for the VoIP case we have an overall mean bit-rate of:

$$\begin{aligned} r_i &= \rho \cdot p_i + \frac{1-\rho}{T \cdot E[X]} \\ &\cdot \left(L_{SID} + H \cdot \left(1 + \frac{P_1(1 - N_{fpp})}{N_{fpp}} \right) \right) \end{aligned} \tag{9}$$

3.2 Mean and Peak Bit-rate for VoMPLS

When compared to the VoIP case introduced above, VoMPLS transportation shows three main changes:

1. The header size (H) only accounts for one HDR header, with a size of 4 octets instead of the VoIP header of 40 octets.
2. The padding phenomenon may add extra octets to the packets generated during voice activity periods or SID periods.
3. The packetization scheme forces that one primary subframe may carry only one SID frame. This implies changes in R_H when compared to the VoIP case.

According to the first and second items, an extra load of $(N_{fpp}L_{ACT}) \mod 4$ octets needs to be added in the ON periods. Thus, p_i is:

$$p_i = \frac{H + N_{fpp}L_{ACT} + (N_{fpp}L_{ACT}) \mod 4}{N_{fpp} \cdot T} \tag{10}$$

On SID periods, an extra load of $L_{SID} \mod 4$ octets needs to be added to equation R_{fr}. Applying again the renewal theorem and taking into account that only one SID frame can travel in the subframe, we can redefine:

$$R_{fr} = \frac{L_{SID} + (L_{SID} \mod 4)}{T \cdot E[X]} \tag{11}$$

and,

$$R_H = \frac{H}{T \cdot E[X]} \tag{12}$$

So the mean bit-rate for VoMPLS will be:

$$r_i = \rho \cdot p_i + (1 - \rho) \cdot \frac{H + L_{SID} + (L_{SID} \mod 4)}{T \cdot E[X]} \tag{13}$$

where all the information units are measured in octets.

3.3 Traffic Trunk Parameters

According to equations 2 and 3, the lower and upper limits of the bandwidth reservation for the TT can be readily calculated, since p_i and r_i have been deduced in sections 3.1 and 3.2 for VoMPLS and for VoIP over MPLS respectively.

However, the traffic parameters of the TT should also include the effect of the MPLS Outer Label in the mean and peak bit-rates. To do this, we have to compute the probability of having at least one voice source generating a new frame at T_{min} (G_{TX}). This can be calculated from the probability that no source from any class generates a packet.

In the VoIP over MPLS case, and according to equation 8, it is given by:

$$G_{TX} = 1 - \prod_{i=1}^{K} \left(1 - \frac{\rho \cdot T_{min}}{N_{fpp} \cdot T_i} - \frac{T_{min}(1-\rho)}{E[X_i] \cdot T_i} \left(1 + \frac{P_1 \cdot (1 - N_{fpp})}{N_{fpp}}\right)\right)^{m_i} \quad (14)$$

In the VoMPLS case, we should consider that packetization scheme forces that one primary subframe may only carry one SID frame. Thus, G_{TX} results into:

$$G_{TX} = 1 - \prod_{i=1}^{K} \left(1 - \frac{\rho \cdot T_{min}}{N_{fpp} \cdot T_i} - \frac{(1-\rho) \cdot T_{min}}{E[X_i] \cdot T_i}\right)^{m_i} \quad (15)$$

Experimental values show that, when more than 5 sources are multiplexed, G_{TX} is greater than 0.95, and for more than 10 sources, the probability increases up to 0.99. Thus, we may assume G_{TX}=1 for a large number of multiplexed sources (i.e. more than 10).

4 VALIDATION AND NUMERICAL RESULTS

This section presents the results of a comparative study between VoIP over MPLS and VoMPLS. This allows to quantify the benefits of VoMPLS, and validates the equations presented in the previous section.

4.1 Experiment Setup

Following the methodology found in (Estepa et al., 2003), both edges of 14 conversations which took place between males and females speakers, were recorded from an ISDN line in a low-noise office environment (i.e SNR $>$ 20dB.) The 600 minutes of PCM audio files obtained were encoded using the G.729B codec. This codec, highly available in any VoIP environment, holds the capability of generating SID frames and is widely referenced in the literature, so it will let us to compare our results with previous studies. The output of the codec was processed to obtain the sequence of types of frames generated (ACT, SID or NoTXN for none.) This information was stored in a file -*ftype* files- for each conversation, and all of them were processed to experimentally find the proper parameters to be used in the models (i.e. activity rate ρ, $E[X]$, P_1.)

The *ftype* files were also split into 120 pieces of five-minutes-long conversations. This database of five-minute pieces of speech conformed a pool from which N were randomly chosen to feed a simulator. Each simulation was repeated 40 times to provide accurate measures of the TT mean bit-rate at the output of the LER.

The mean bit-rate obtained in our simulations is then compared to those provided by the analytical expressions of R_{TT}, for both VoIP over MPLS and VoMPLS, respectively.

4.2 Numerical Results

A total of N=20 homogeneous voice sources were multiplexed following the procedure explained above. Figure 5 plots R_{TT} as obtained from simulations and from our analytical ON-SID model for both VoIP and MPLS TTs. Additionally, it shows R_{TT} when r_{SID}=0, as is traditionally assumed by the one-source ON-OFF model.

In the VoIP over MPLS case the differences between our analytical ON-SID model prediction and the simulation results, measured at the LER's output, range between 1% (N_{fpp}=1) and 4% (N_{fpp}=6). For VoMPLS those differences vary from 0%(N_{fpp} =1) to 5% (N_{fpp} =6). This validates the analytical results in both cases.

When using the traditional ON-OFF model, the VoIP over MPLS case shows differences ranging from 11% to 31%, at the same measuring point. In the VoMPLS case, with the same frame-generation model, the differences range between 11% and 19%. This means that the SID-frames effect, known to be non-negligible in VoIP, should also be taken into account in the VoMPLS case.

Note that, due to the padding phenomenon in VoMPLS, at N_{fpp}=2 R_{TT} is smaller than at N_{fpp}=3. It also demonstrates that N_{fpp}=2 is an interesting working point for the G.729 codec, achieving less delay with lower bandwidth consumption than N_{fpp}=3.

Figure 6 reveals that using VoMPLS instead of VoIP over MPLS yields bandwidth savings ranging

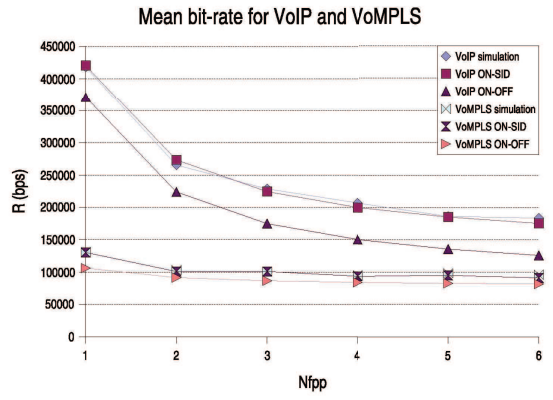

Figure 5: Experimental and analytical values of R_{TT}.

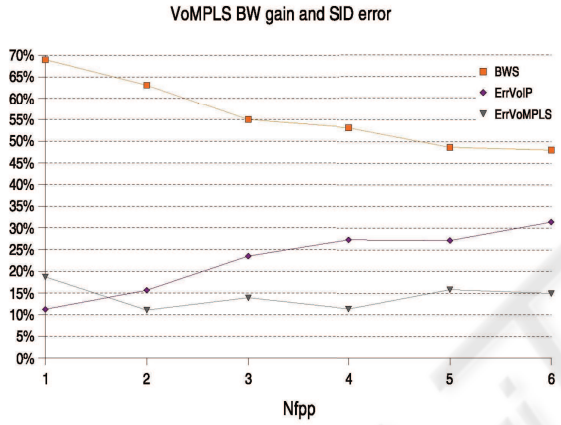

Figure 6: Bandwidth saving and ON-OFF error.

from 69% (N_{fpp}=1) to 48% (N_{fpp}=6).

5 CONCLUSIONS

Traffic engineering needs accurate traffic parameters in order to calculate the optimal capacity allocation for a Traffic Trunk. We have provided analytical expressions for the mean and peak bit-rate of a Traffic Trunk loaded with a mix of heterogeneous voice sources for both the VoIP and VoMPLS transport models. Conversely to the ON-OFF based models, the model used for voice sources captures the effect of the SID frames generated by a number of modern voice codecs. We show that the SID-frames effect has to be considered in the VoMPLS case, too. This conveys an improvement in the accuracy of results, which show a quantitative gain in the bandwidth necessary to transport voice trunks when compared to VoIP.

The calculation of required bandwidth to be allocated for a voice TT with QoS commitments is in progress at the time of writing this paper. This subject as well as the A2oMPLS transport case are left for further study.

REFERENCES

(1999). Requirements for traffic engineering over mpls. *RFC 2702*.

(2001). Multiprotocol label switching architecture. *RFC 3031*.

(2001). Voice over mpls- bearer transport implementation agreement. *MPLS Forum I.A.1.0*.

B, G. (2002). Voice over internet protocol (voip). In *Proceedings of the IEEE*. IEEE Press.

Estepa, A., Estepa, R., and Vozmediano, J. (2003). Packetization and Silence Influence on VoIP Traffic Profiles. *Lecture Notes in Computer Science*, 2899(1):331–339.

Estepa, A., Estepa, R., and Vozmediano, J. (2005). Accurate prediction of voip traffic mean bit rate. *IEE Electronic Letters*, 8(10):644–647.

Procissi G, Garg A., G. M. S. M. (2002). Token bucket characterization of long-range dependent traffic. *Computer Communications. Ed. Elsevier*, 25:1009–1017.

R. Guérin, H. A. and Naghshineh, M. (1991). Equivalent capacity and its application to bandwith allocation in high-speed networks. *JSAC. IEEE*, 9(7).