Admissible traffic load of real time class of service for inter-domain peers

Halina Tarasiuk, Robert Janowski, and Wojciech Burakowski
Warsaw University of Technology, ul. Nowowiejska 15/19, 00-665 Warsaw, Poland
{halina; robert; wojtek}@tele.pw.edu.pl

Abstract

The paper[1] deals with the problem of assuring predefined QoS for real time (RT) class of service (CoS), which we define as one of inter-provider or inter-domain classes of service in IP network. We propose an analysis and adequate formulas to obtain admissible traffic load in a case when we map two end-to-end classes of service dedicated for telephony CoS and video conference CoS into one inter-domain RT CoS. The aim of the analysis is to determine the admissible load when the target packet loss ratio and buffer size dedicated for RT CoS are known. The proposed solution accounts for the difference between packet sizes of streams generated by voice (about 100 bytes) and video conference (rather 1500 bytes) applications. We illustrate our studies by simulation results.

1. Introduction

The notion of network service/class of service is not new, and it was successfully used in ATM and some prototypes of IP-based network, as AQUILA[2] [2, 3], GEANT[3]. Briefly speaking, the class of service (CoS) is a service the network offers to the traffic streams. By using this service we can expect that a packet traffic submitted to this service will be transferred according to the guarantees specified for this service. For instance, when we consider best effort service one can expect that the packets submitted to this service may affect unpredictable transfer delay and even may be lost. Therefore, we will speak about, so called, QoS classes of service, the classes that will guarantee for the packets streams specific quality expressed in the form of packet delay characteristics, packet loss characteristics etc.

Nowadays, we observe an evolution of access networks, such as Ethernet/LAN (Local Area Network), xDSL (x Digital Subscriber Line), WLAN (Wireless Local Area Network), UMTS (Universal Mobile Telecommunications System). Such access networks interconnected by the IP core network constitute heterogeneous network environment.

Since we plan to deploy the system over multiple domain network and over different network technologies, the issue of fixing the classes of service is most critical. We should focus on classes of service that should be visible by the user side. The first step is to have a look on the applications we plan to use, more specifically on their requirements corresponding to the QoS.

We state that we need to specify end-to-end classes of service, which will be visible by the applications (end users). The packet traffic from the application is submitted to one of the end-to-end classes of service and the experienced QoS at the packet level depends in straightforward way on the effectiveness and reliability of the service. A list, which covers currently discussed end-to-end classes of service was presented in [5] (see Table 1). In the paper we follow this concept and distinguish between maximum number of 11 end-to-end classes of service, which differ in both predefined QoS objectives and traffic profiles. On the contrary, based on discussions between service providers [5] (see Table 1) we consider only 4 inter-provider classes of service that are: Ctrl - for transferring system information, Real-time (RT) - for transferring the packet streams sensitive to delay and losses characteristics and corresponding to, so called, streaming traffic, None Real Time (NRT) - for transferring the packet streams not sensitive to delay and losses characteristics but requiring a guaranteed transmission rate, and Best Effort (BE).

Taking into account the above, we investigate also the following mapping functions: (1) a function

[1] This work is partially funded by the European Union under contract number 004503 "EuQoS"
[2] AQUILA – Adaptive Resource Control for QoS using an IP-based Layered Architecture, European project, 5FR, No. IST-1999-10077
[3] GEANT – Gigabit European Academic Network

0-7695-2450-8/05 $20.00 © 2005 IEEE

between applications and end-to-end classes of service, (2) a function between end-to-end classes of service and classes of service available in particular network technologies, as WLAN, Ethernet/LAN, IP, etc., (3) a function between end-to-end classes of service and inter-provider/inter-domain classes of service.

In this paper we focus mainly on the third function, which is mapping between end-to-end and inter-domain classes of service.

In particular, we focus on the problem of assuring predefined QoS for real time class of service (RT CoS), which we define as one of inter-provider or inter-domain classes of service. We propose an analysis and adequate formulas to obtain admissible traffic load in a case when we map two end-to-end classes of service dedicated for voice over IP (telephony CoS) and video conference CoS into one RT CoS. The aim of the analysis is to determine the admissible load when the target packet loss and buffer size dedicated for RT CoS are known. The proposed solution accounts for the difference between packet sizes of streams generated by voice (about 100 bytes) and video conference (rather 1500 bytes) applications.

The paper is organized as follows. In section 2 we provide a definition of RT CoS in inter-domain peers. We present the definition in a form of discussion of important factors, which have an impact on the service definition. Then in section 3 we present simulation results and define a problem related to an appropriate method for calculating admissible traffic load for RT CoS. In section 4 we propose an original proposal for calculating admissible traffic load and prove the concept by providing simulation results. Finally, section 5 summarizes the paper.

2. Real time CoS in inter-domain peers

For defining RT CoS in inter-domain peers we need to specify: the QoS objectives for the inter-domain (inter-provider) area, the QoS mechanisms and QoS algorithms, e.g. admission control (AC). We also need to define rules for dimensioning of the service. As we can see in Figure 1, we need to define inter-domain classes of service taking into account resources as link and buffer capacity of output interface located in egress border router (BR) of particular autonomous system (AS).

On inter-domain links we are planning to keep only 4 CoSs, as above mentioned, among them RT CoS will be of our main interest. On the contrary, in the access domains the intention is to keep more classes of service, strictly corresponding to the types of applications. For instance, the telephony and video conference CoSs will be accessible on the access network and they will be aggregated to one, in this case RT, service. Such aggregation is presented on Figure 2.

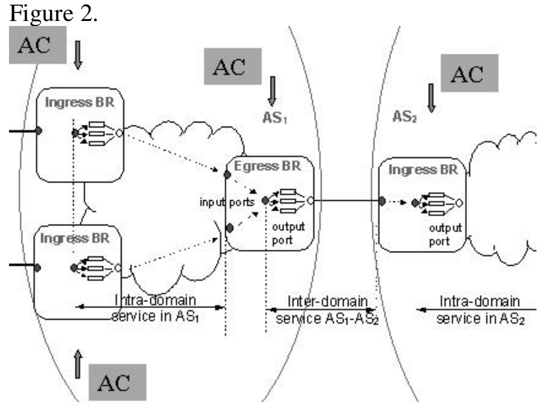

Figure 1. Intra- and inter-domain classes of service

Finally, for the purpose of our discussion we assume two end-to-end CoSs, telephony and video conference (see Figure 2). Now, we focus on main impact of mapping these services to one separate RT CoS in inter-domain peers.

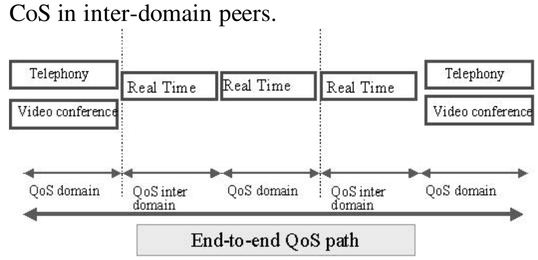

Figure 2. Exemplary mapping between end-to-end and inter-domain classes of service

2.1. Discussion of important factors of RT CoS design

The discussion of important factors of RT CoS design we propose to begin from QoS requirements. As we show in Table 1, QoS requirements for telephony service are specified by very low packet losses, very low packet delay and very low jitter (see also [10, 11]). On the other hand, QoS requirements related to RT interactive service (in our case video conference) are low packet losses, very low packet delay and low jitter.

Proceedings of the Joint International Conference on Autonomic and Autonomous Systems
and International Conference on Networking and Services (ICAS/ICNS 2005)
0-7695-2450-8/05 $20.00 © 2005 IEEE

Table 1. The types of service classes/classes service [5]; xx = 01, 10, or 11

Inter-Provider Service Class	Tolerance to			PHB	End-To-End Service Class	Tolerance To			DSCP	
	Loss	Delay	Jitter			Loss	Delay	Jitter	Name	Value
Ctrl	Low	Low	Yes	CS	Network Control	Low	Low	Yes	CS7	111000
Real Time	VLow	VLow	VLow	EF	Telephony	VLow	VLow	VLow	EF	101110
					Signaling	Low	Low	Yes	CS5	101000
					MM Conferencing	L-M	VLow	Low	AF4x	100xx0*
					RT Interactive	Low	VLow	Low	CS4	100000
					Broadcast Video	VLow	Med	Low	CS3	011000
None Real Time	Low	LIM	Yes	AF	MM Streaming	L-M	Med	Yes	AF3x	011xx0*
					Low Latency Data	Low	L-M	Yes	AF2x	010xx0*
					OAM	Low	Med	Yes	CS2	010000
					High Thruput Data	Low	M-H	Yes	AF1x	001xx0*
Best Effort	NS	NS	NS	DF	Standard	NS	NS	NS	DF	000000

As we can see (Table 1), to meet QoS requirements inside one RT service for both end-to-end classes of service we have to choose the most rigorous requirements as defined for telephony CoS.

Moreover, we assume that our QoS objectives for RT CoS is to offer strict QoS guarantees. As a consequence we have to focus on admission control methods, which support such guarantees. For this purpose in our approach we will consider methods, which support QoS guarantees in a statistical way.

Next factors, which have an impact on RT service design are traffic characteristics. Traffic produced by applications related to telephony and video conference end-to-end CoSs differs in their characteristics. We can point out the following main differences. Flows of telephony and video conference demanding in essentially different bandwidth. Moreover, packets generated by applications of telephony type are rather short (e.g. about 100 Bytes), while packets produced by video are quite larger (as 1500 Bytes).

Consequences of serving two end-to-end classes of service as telephony and video conference by one RT service in inter-domain peers are the following. Packet streams of both end-to-end classes of service will share the same resources as bandwidth and buffer capacities dedicated to RT service in each inter-domain peer. Therefore, we have to take into account an impact of video connections on quality of service perceived by telephony and vice versa.

To guarantee strict QoS objectives for RT service we assume that admission control is performed on-line per flow. New flow is admitted in an invocation process. Admission of new flow requires allocation of resources associated with an inter-domain link. Release of connection requires release of resources.

Setting of any parameter in routers is not required in the invocation process, it is required only in resource provisioning process (in long time scale). Per flow policing is also not required. We assume that policing is applying only at the access networks or edge routers. It is not in a contrary to support per aggregate flow policing related to SLA (Service Level Agreement) between operators.

Now, we need to specify an appropriate admission control rules for the discussed RT CoS.

2.2. Related works

As it was mentioned above a notion of class of service or network services is well known. AQUILA project [2] defined 4 premium network services with strict QoS guarantees and one standard (best effort) service. All premium network services differ in both traffic characteristics and QoS objectives. Two of them PCBR (Premium Constant Bit Rate) and PVBR (Premium Variable Bit Rate) were defined for handling real time traffic. The services with appropriate admission control (AC) algorithms were defined for intra-domain IP network based on DiffServ concept. Therefore, AC algorithms were applied only at edge routers. The AC for the PCBR network service was based on the REM (Rate Enveloped Multiplexing) scheme using a well-known peak rate allocation method. The admissible load for the capacity allocated to PCBR was calculated based on the analysis of an M/D/1/B system depending on the assumed target packet loss ratio and the available buffer size. PCBR was dedicated to constant bit rate traffic (e.g., voice trunks) and was served with the highest priority. Negligible packet delay variation can be assured when at most 10% PCBR load is allowed on the link. The REM scheme was also applied to the PVBR network service, which was designed for effective transfer of streaming variable bit rate traffic, e.g., video applications. Here, the effective bandwidth was calculated by Karl Lindberger's method [8]. The buffer dimensioning rules were adequate for absorbing the so-called packet scale congestion. For this purpose, the analysis of the N*D/D/1 queuing system was applied. For both PCBR and PVBR services in second phase of the project MBAC (Measured Based Admission Control) methods were proposed [3]. Thanks to them traffic descriptors in the case of PVBR service were simplified to peak bit rate and peak bit rate tolerance only.

The reasons to separate PCBR and PVBR services in the AQUILA approach were the following. In a single network service do not mix the flows demanding essentially different bandwidth and do not mix the flows with essentially different traffic profiles (at the packet level). But let us remark that premium

Proceedings of the Joint International Conference on Autonomic and Autonomous Systems
and International Conference on Networking and Services (ICAS/ICNS 2005)
0-7695-2450-8/05 $20.00 © 2005 IEEE

services in AQUILA were defined for intra-domain network.

Since QoS requirements for PCBR and PVBR were defined in the same way as we assume for end-to-end telephony CoS and video conference CoS, respectively, now we will check if rules for admission control of RT CoS could be the same as assumed for PCBR service.

As an example below we provide the rough specification of PCBR service. QoS objectives are strict QoS guarantees; most acceptable for voice codecs, packet losses < 10^{-4}, packet delay < 150msec, jitter < 20msec. Types of connections are point-to-point. Traffic descriptors are based on single token bucket with PBR and PBRT parameters.

The implementation of this service in IP-based network needs the following. Provisioning of resources: made in a static way: C_1 capacity, that is a part of the C link capacity (C_1<C), is dedicated for PCBR, the buffer for this C_1 should be also fixed, e.g. for 10 packets. Admission control function can be based on declarations submitted by the QoS request and we assume, so called, peak rate allocation scheme.

We assume a description of traffic inside each single flow in a form of single token bucket parameters: peak bit rate (PBR) and peak bit rate tolerance (PBRT). These parameters describe traffic of each flow submitted to the network.

New flow is admitted if:

$$PBR_{new} + \sum_{i=1}^{N1} PBR_i \leq \rho C_1 \qquad (1)$$

Where N1 denotes the number of connections in progress and parameter ρ (ρ<1) specifies the admissible load of capacity allocated to the PCBR. The value of ρ can be calculated from the analysis of M/D/1/B system taking into account the target packet loss ratio and the buffer size [2].

$$\rho = \frac{2Buffer}{2Buffer - ln(P_{loss})} \qquad (2)$$

Where Buffer denotes buffer size in packets and P_{loss} target packet loss ratio.

3. Evaluation of admissible traffic load method for RT CoS

The main objective of this section is to evaluate the well known formula for admissible traffic load (see equation (2)) proposed for PCBR service in AQUILA and answer the question if we can apply it for RT CoS defined in section 2. The results obtained by simulation are compared with target values.

For this purpose we run simulations for two cases. In case 1 we assume that resources as link and buffer capacity dedicated for RT CoS are shared between end-to-end CoSs as telephony and video conference. The admissible traffic load for RT CoS is calculated according to the rules as presented in section 2.2 for PCBR service. We assume the buffer size equal to 10 packets and target packet loss ratio (P_{loss}) 10^{-2}. It gives the admissible traffic load ρ=0.81.

Additionally, in case 2 we check an impact of packet size ratio on packet loss ratio.

Simulation topology is depicted in Figure 3. The link capacity C varies from one simulation test to another and equals 512 Kbps, 2Mbps, 10Mbps and 100Mbps. We assume that packets in the system are handling according to FIFO discipline.

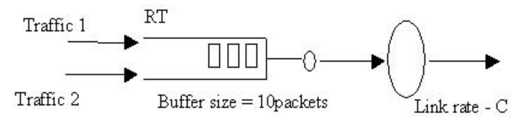

Figure 3. Simulation topology

The system is fed by the worst case traffic of each CoS that is Poisson. The traffic rate of the Poisson streams are determined by the total admissible load (ρ), AC limit for each class (limit – in percent) and link capacity (C) according to ((3) and (4)).

$$\begin{aligned} traffic\ 1\ \ rate &= C\rho\ limit1 \\ traffic\ 2\ \ rate &= C\rho\ limit2 \end{aligned} \qquad (3)$$

and

$$traffic\ 1\ \ rate + traffic\ 2\ \ rate = C\rho \qquad (4)$$

Traffic profiles for case 1 are presented in Table 2.

All simulation results were obtained by using the packet transmission level simulator – ns-2 [ns-2]. We measured the following parameters:

- Mean delay: defined as the arithmetic average of IP packets transfer (queuing+transmission) delays for a fixed set of packets.
- Maximum delay: defined as the maximum delay that a packet experience while passing through the whole path.
- Standard deviation: defined as a square root of variance which is defined as the mean of the squares of the differences between the respective packet delays and their mean.
- Packet loss ratio: For a fixed set of transferred packets, this is the ratio of the number of received packets to the number of sent packets.

Table 2. Traffic profiles for case 1

End-to-end CoS	Inter-domain CoS	Traffic
CoS1 Telephony	RT CoS	Traffic 1: Poissonian type with parameters: mean bit rate, packet size: 60Bytes
CoS2 Video Conference		Traffic 2: Poissonian type with parameters: mean bit rate, packet size: 1500B

The simulations were performed respecting the following rules. Each scenario was simulated for a period of time to send at least 10^6 packets. Simulations of each scenario were repeated 12 times to account for the random nature of the experiment. Obtained results were statistically post-processed to calculate the intervals of confidence with the 0,95 confidence level.

3.1. Impact of link rate

The results of packet losses and packet delays for the case 1 are presented in Table 3. In this case we investigate 4 scenarios. In all scenarios we assume target packet loss ratio equal 10^{-2} (we assume such a value due to simplicity of simulation only), the AC limit for traffic 1 and 2 equal to 10% and 90% respectively.

The results presented in Table 3 indicate that the packet loss ratio is insensitive to the link capacity. It is obvious since the input traffic rates were scaled with the link capacity and the total traffic load was kept constant. However, for all simulation cases the obtained packet loss ratio is higher (about 0.05) than assumed (0.01). This indicates that the packet loss ratio is dependent on the mutual packet sizes of two traffic classes. We observe also that the mean packet transfer delay decreases proportionally to the link capacity e.g. for 2Mbps link it is 4 times smaller than for 512Kbps. The standard deviation of the transfer delay decreases proportionally to the link capacity e.g. for 2Mbps link it is 4 times than for 512Kbps.

The above comments lead to the following conclusions. For the high speed links it is possible to maintain only one class of service in inter-domain peers since the absolute values of the packet delays (mean and the variation) are acceptably small from the QoS requirement point of view.

Table 3. Packet transfer delay and packet loss ratio

End-to-end CoS	Mean Packet transfer delay [s]	Max. Packet transfer delay [s]	Standard deviation	Packet loss ratio
Link rate 512Kbps				
CoS1	0.025	0.184	0.026	$0.053 \pm 6*10^{-4}$
CoS2	0.047	0.204	0.026	$0.053 \pm 6*10^{-4}$
Link rate 2Mbps				
CoS1	0.0063	0.0537	0.007	$0.053 \pm 6*10^{-4}$
CoS2	0.0120	0.0598	0.007	$0.053 \pm 6*10^{-4}$
Link rate 10Mbps				
CoS1	0.0012	0.0099	0.0014	$0.053 \pm 6*10^{-4}$
CoS2	0.0023	0.0108	0.0014	$0.053 \pm 6*10^{-4}$
Link rate 100Mbps				
CoS1	0.00012	0.00104	0.00014	$0.053 \pm 6*10^{-4}$
CoS2	0.00023	0.00107	0.00014	$0.053 \pm 6*10^{-4}$

3.2. Impact of packet size ratio

Since the packet size is characteristic feature of the application type and so the class of service, we checked also an impact of packet size ratio on packet losses.

As in previous scenarios we assume the simulation topology as depicted on Figure 3. The results were partially checked for the link capacity equal to 512Kbps, 2 Mbps, 10Mbps, and 100Mbps and were found to be insensitive to the capacity but the total load of the link. Therefore, simulation results presented in this paper corresponds only to the link capacity equal to 100 Mbps. In this case, simulations were performed for the following target packet loss ratio 10^{-2}, 10^{-3}, and 10^{-4} and admissible traffic loads 0,81, 0,743, and 0,68, respectively. The admissible traffic load was calculated according to the same formula (2) as in the case 1.

In our simulations packet sizes were the following: d_1=100; d2=100, 200, 400, 800, 1200, 1500 Bytes, where d_1 and d_2 are sizes of packets of traffic 1 and traffic 2, respectively. In addition, we assume that $d=d_2/d_1$.

Simulation results enclosed below (Figure 4, Figure 5 and Figure 6) indicate that when two traffic types with different but constant packet sizes share the common buffer (dimensioned in packets), then the packet loss ratio is dependent on the mutual ratio of these packet sizes. Increasing the packet size of one traffic type while maintaining the same total traffic load and the packet size of the other traffic type, increases the packet loss ratio.

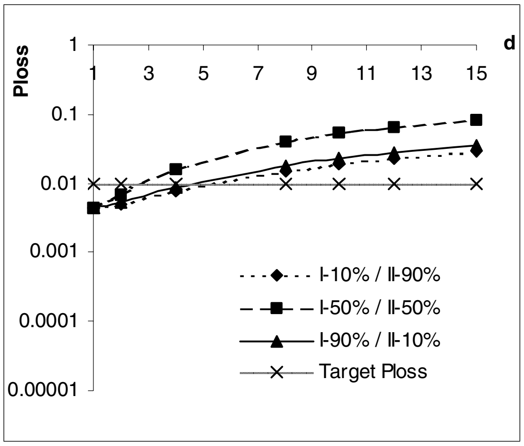

Figure 4. Packet loss ratio vs. packet size ratio of two end-to-end CoSs; target packet loss=10^{-2}; total traffic load= 0.81

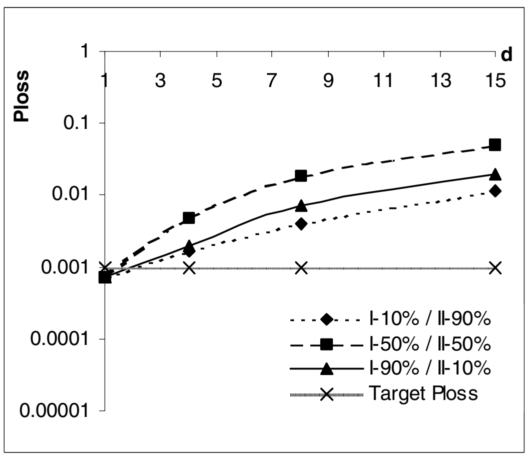

Figure 5. Packet loss ratio vs. packet size ratio of two end-to-end CoSs; target packet loss= 10^{-3}; total traffic load=0.743

In all scenarios the traffic mix that results in highest packet loss ratio is 50% of CoS1 and 50% of CoS2 (I-50% / II-50%). The lowest increase in packet loss ratio is observed for the case when CoS1 contributes 10% and CoS2 contributes 90% (I-10% / II-90%) of total traffic i.e. the majority of packets is long. To explain this behavior we notice that the packet size represents the traffic variability at the smallest time scale. Together with the packet arrival intensity it contributes to the packet losses.

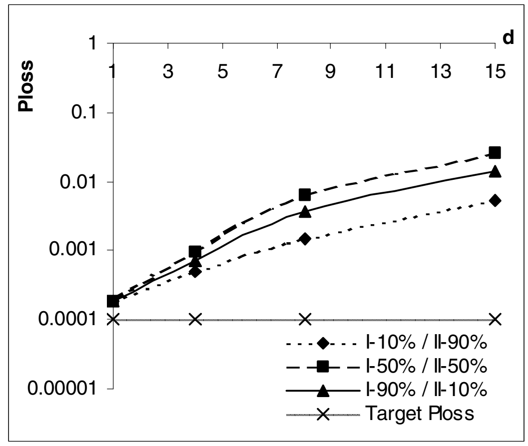

Figure 6. Packet loss ratio vs. packet size ratio of two end-to-end CoSs; target packet loss=10^{-4}; total traffic load= 0.68

Since by varying the ratio of the packet sizes we obtain different packet loss ratios sometimes higher than assumed (see Figure 4, Figure 5, and Figure 6), we need new rule for calculating admissible traffic load that can account of this impact. This is the subject of the next section.

4. Admissible traffic load for RT CoS - new proposal

In this section we provide an adequate formulas to obtain admissible traffic load in a case when we map two end-to-end CoSs telephony and video conference (see MM conferencing in Table 1) into RT CoS. In this case, QoS objectives of RT CoS are very low packet delay, jitter and very low packet losses. The aim of provided analysis is to determine the admissible traffic load when the target packet loss and buffer size dedicated for the real time service are known. The proposed solution accounts for the difference between packet sizes of streams generated by voice (which are about 100 bytes) and video conference (which are rather 1500 bytes for video streams) applications. First approach, which we considered was to obtain admissible load by applying M/D/1/B analysis (see section 3). However, as it was expected, the included simulation results show that when two packet streams of Poissonian type, each of them with different but constant packet sizes share the common buffer (dimensioned in packets), then packet loss probability depends essentially on mutual ratio of

Proceedings of the Joint International Conference on Autonomic and Autonomous Systems
and International Conference on Networking and Services (ICAS/ICNS 2005)
0-7695-2450-8/05 $20.00 © 2005 IEEE

this packet sizes. Increasing the packet size of one stream while maintaining the same traffic load and packet size of the other stream, increases the packet loss ratio. Now, we provide a relatively simple mathematical model that accounts this phenomena. By applying the method RT CoS can assure the target packet loss ratio in variety of traffic parameters (packet size ratio, proportions of the contributed load) by appropriately adjusting the maximum admissible load. It is able to assure the target packet loss ratio even when the packet sizes differ significantly. The effectiveness of the method is illustrated by simulation results.

4.1. System analysis

We now ask how the maximum admissible load of each end-to-end CoS depends on the ratio of packet sizes. The answer to this question is important when we plan to merge different traffic types into one class of service being served in router with FIFO discipline and still assure some QoS guarantees as target packet loss ratio.

We assume the discrete time analysis as in [4] where the time is divided into units (slots) corresponding to the smaller packet transmission time. In the model we assume that the input traffic of both CoSs is Poisson process. The packet sizes are constant equal to 'd_1' and 'd_2', respectively for telephony and video conference CoSs and their ratio (d_2/d_1) is an integer denoted by 'd'. The packets of these CoSs enter the same finite buffer (with buffer size – *Buffer* counted in packets). The studied system is depicted on Figure 7.

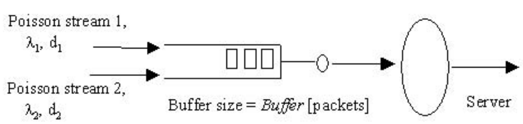

Figure 7. The studied system

We begin with the equation describing the evolution of the system state which is defined as the number of packets in the system without differentiating their types. We look at the system state (number of packets in the system) at the time instants embedded after each packet service completion or after each time slot when the system is empty (see Figure 8).

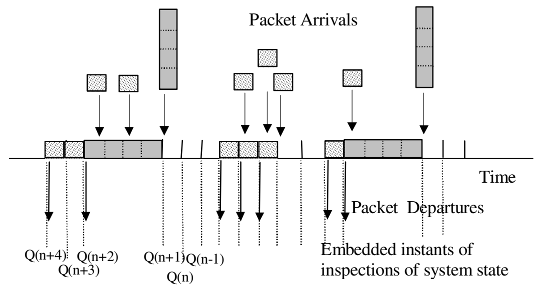

Figure 8. Time evolution of the system state

Then the general system equation in the case of two traffic types is (5):

$$Q(n+1)=\begin{cases} Q(n)-1+A_1+A_2 & \textit{if pkt\_type1\_served} \\ Q(n)-1+\sum_{i=1}^{d} A_1+\sum_{j=1}^{d} A_2 & \textit{if pkt\_type2\_served} \\ A_1+A_2 & \textit{if system\_was\_empty} \end{cases} \quad (5)$$

Where:

- A_1, A_2 random variable describing the number of type 1 (respectively type 2) packet arrivals during one slot,
- $Q(n)$ denotes the system state at the end of n-th embedded time instant.

Using Generating Function approach the equation (5) becomes (6):

$$\begin{aligned} Q(z) &= \frac{Q(z)}{z} A_1(z) A_2(z) \mathrm{Prob}\{\textit{system\_busy\_\&\_pkt\_type1\_served}\} \\ &+ \frac{Q(z)}{z} (A_1(z))^d (A_2(z))^d \,\mathrm{Prob}\{\textit{system\_busy\_\&\_type2\_served}\} \\ &+ A_1(z) A_2(z) \mathrm{Prob}\{\textit{system\_is\_empty}\} \end{aligned} \quad (6)$$

Taking into account that the input is Poisson stream, we obtain (7):

$$\begin{aligned} Q(z) &= \frac{Q(z)}{z} e^{(\lambda_1 d_1+\lambda_2 d_1)(z-1)} \frac{\lambda_1}{\lambda_1+\lambda_2} \\ &+ \frac{Q(z)}{z} e^{(\lambda_1 d_2+\lambda_2 d_2)(z-1)} \frac{\lambda_2}{\lambda_1+\lambda_2} \\ &- \frac{(1-\rho)}{z} \left(e^{(\lambda_1 d_1+\lambda_2 d_1)(z-1)} \frac{\lambda_1}{\lambda_1+\lambda_2} + e^{(\lambda_1 d_2+\lambda_2 d_2)(z-1)} \frac{\lambda_2}{\lambda_1+\lambda_2}\right) \\ &+ e^{(\lambda_1 d_1+\lambda_2 d_1)(z-1)} (1-\rho) \end{aligned} \quad (7)$$

In the above equation the load (ρ) and the arrival intensities (λ_1, λ_2) are related by:

$$\rho_1=\lambda_1 d_1=\lambda_1 \ (\textit{since } d_1=1);\ \rho_2=\lambda_2 d_2=\lambda_2 d; \quad \rho=\rho_2+\rho_2 \quad (8)$$

Furthermore, the percentage contribution of the particular traffic type to the total load is known.

Proceedings of the Joint International Conference on Autonomic and Autonomous Systems
and International Conference on Networking and Services (ICAS/ICNS 2005)
0-7695-2450-8/05 $20.00 © 2005 IEEE

$$\rho_1 = w_1\rho;\ \rho_2 = w_2\rho \quad (9)$$

Rearranging the above equation (7) finally gives us the expression of the Q(z) in the form of (10):

$$Q(z) = \frac{-(1-\rho)\left(e^{(\lambda_1 d_1+\lambda_2 d_1)(z-1)}\frac{\lambda_1}{\lambda_1+\lambda_2} + e^{(\lambda_1 d_2+\lambda_2 d_2)(z-1)}\frac{\lambda_2}{\lambda_1+\lambda_2}\right) + e^{(\lambda_1 d_1+\lambda_2 d_1)(z-1)}(1-\rho)z}{z - \frac{\lambda_1}{\lambda_1+\lambda_2}e^{(\lambda_1 d_1+\lambda_2 d_1)(z-1)} - \frac{\lambda_1}{\lambda_1+\lambda_2}e^{(\lambda_1 d_2+\lambda_2 d_2)(z-1)}} \quad (10)$$

Assuming that the tail probabilities of the queue size distribution function are well approximated by the dominant pole of Q(z), they can be written as (11):

$$\Pr ob\{Q = x\} = C_0(\frac{1}{z_0})^x \quad (11)$$

Further, assuming that the asymptotic constant Co equals 1 [1],[6],[7] the buffer overflow probability can be expressed as (12):

$$Ploss \approx \Pr ob\{Q > Buffer+1\} = \sum_{n=Buffer+2}^{\infty} C_0(\frac{1}{z_0})^n = \frac{1}{z_0-1}(\frac{1}{z_0})^{Buffer+1} \quad (12)$$

From equation (12) we can determine the value of the required decay rate parameter $1/z_0$. This decaying rate ensures that the buffer overflow probability will be below target P_{loss} value. Enforcing this z_0 value to be dominant pole of the Q(z) puts the following condition on the denominator of Q(z) (13).

$$z_0 - \frac{\lambda_1}{\lambda_1+\lambda_2}e^{(\lambda_1 d_1+\lambda_2)(z_0-1)} - \frac{\lambda_2}{\lambda_1+\lambda_2}e^{(\lambda_1 d_2+\lambda_2 d_2)(z_0-1)} = 0 \quad (13)$$

Using the information about the contribution of particular traffic types to total load and their traffic characteristics given in (8) and (9), we can rewrite the above equation in the simpler form that involves just one unknown parameter ρ.

$$z_0 - \frac{w_1\rho}{w_1\rho + \frac{w_2\rho}{d}}e^{(w_1\rho + \frac{w_2\rho}{d})(z_0-1)} - \frac{\frac{w_2\rho}{d}}{w_1\rho + \frac{w_2\rho}{d}}e^{(w_1\rho d + w_2\rho)(z_0-1)} = 0 \quad (14)$$

From this equation we can calculate the total admissible load ρ of both classes of service ($\rho=\rho_1+\rho_2$). After that the admissible load of particular classes of service can be calculated based on the relationship provided in (9).

Below we summarize the method in simple steps that lead to calculation of the admissible load when all the input parameters (*Buffer*, P_{loss}, d_1, d_2, percentage contribution of different types of traffic - w_1, w_2) are given:

1. Given P_{loss} and *Buffer*, determine the parameter z_0 from equation (12).
2. Create the equation (14) taking into account the number of traffic types, their characteristics (intensity, packet sizes) and the assumed input model (Poisson).
3. Solve the equation (14) with respect to ρ which is the total admissible load.
4. Calculate the admissible load of each traffic class based on the information about percentage contribution of different traffic classes - w_1, w_2 (9).

Although the above analysis covers the case of two traffic types related to two end-to-end classes of service, the method is general and might be applied to the case with more traffic types. In these cases the equation (13) must be adjusted accordingly to account for the existence of all traffic types. Each class of service will contribute one term (the generating function of its input traffic) proportionally to its traffic intensity.

The above introduced method is used to calculate the total admissible load of two traffic types so that the required packet loss probability could be assured. The exemplary results of the total admissible load are enclosed in Figure 9. They were obtained for target packet loss ratio equal to 10^{-2}.

4.2. Simulation results

To verify the correctness of the proposed method we once again conducted simulations were the two traffic types were mixed into one buffer served according to FIFO discipline but this time we admitted the traffic from each end-to-end class of service up to the calculated limit. Simulation model and general traffic profiles are the same as described in section 3 for case 2 (impact of packet size ratio). The obtained results of packet loss ratio are depicted in Figure 10, Figure 11, and Figure 12 for target packet loss ratio equal to 10^{-2}, 10^{-3}, and 10^{-4}, respectively. Let us remark that for each value of 'd' and for each of the obtained characteristics we had to calculate admissible traffic load values. To save the space we include only values of admissible traffic loads related to $P_{loss}=10^{-2}$ (see Figure 9).

Taking into account the obtained results we can conclude that the proposed method is able to assure target packet loss ratio even when the packet sizes differ significantly. It proved to be robust in cases with different proportions of the traffic mix (the enclosed results cover 3 cases). It is conservative in the sense it underestimates admissible loads. The primary reason for its conservativeness is that it approximates the packet loss probability with the buffer overflow probability and assumes C_0 constant in equation (11) equal one while in reality it is load dependent and often much less than one.

5. Summary

Merging classes of service which carry traffic with different packet sizes results not only in higher delays but also in higher packet losses. Although the problem with higher delays might become negligible when the classes of service are merged on high speed links, the problem of higher packet losses is insensitive to the capacity of the link and always exists.

To assure that the traffic admitted within e.g. end-to-end telephony CoS can be still guaranteed the same packet losses within inter-domain RT CoS, the maximum admissible load for RT CoS must be calculated based on the model that accounts for differences in packet sizes inside this class.

In this paper we have proposed a relatively simple model of this phenomena and its analysis exploiting Generating Function approach. The obtained results prove the usefulness of the method. In simulations we verify that the traffic which was admitted to the calculated limit was assured the assumed target packet loss ratio. The method serves as an upper bound due to some approximations, which were done to simplify the analysis because we put the main stress on modeling the nature of phenomena and its consequences.

As a further work we plan to investigate the consequences of merging MM streaming, Low Latency Data, OAM and High Throughput Data classes of service into NRT inter-domain CoS. By conducting exhaustive simulation tests we hope to identify any constraints coming from the mergence of these classes of service. Providing the solution for NRT CoS will complete the work on the end-to-end assurance of QoS in inter-domain peers.

References

[1] J. Abate, G. L. Choudhury and W. Whitt. *Asymptotics for steady-state tail probabilities in structured Markov queueing models*. Stochastic Models, 10 (1994) 99-143.

[2] A. Bąk, W. Burakowski, F. Ricciato, S. Salsano, H. Tarasiuk, *A framework for providing differentiated QoS guarantees in IP-based network*, Computer Communications 26 (2003), Elsevier, pp. 327-337.

[3] C. Brandauer, W. Burakowski, M. Dabrowski, H. Tarasiuk, and B. Koch, *AC algorithms in AQUILA QoS IP Network*, Euro. Trans. On Telecommunications, Accepted paper 31 August 2004.

[4] H. Bruneel and B.G.Kim, *Discrete-time models for communication systems including ATM*, Boston: Kluwer Academic Publishers, 1993.

[5] Kwok Ho Chan, Jozef Babiarz, *Inter-Provider Service Classes*, Inter-provider QoS Meeting#2, 28 January 2005, US, http://cfp.mit.edu/qos/slides.html.

[6] C.S. Chang and J.Thomas, *Effective bandwidth in high speed digital networks*, IEEE Journal on Selected Areas in Communications, vol. 13, August 1995, pp. 1091-1114.

[7] F. Kelly, *Effective bandwidth in multiclass queues*, Queueing Systems, vol. 9, 1991, pp. 5-16.

[8] K. Lindberger, *Dimensioning and design methods for integrated ATM networks*, Proc. of 14-th International Teletraffic Congress, Antibes, 1994.

[9] www.isi.edu/nsnam/ns

[10] Telecommunication Standardization Sector of ITU, ITU-T Recommendation G.1010, *End-user multimedia QoS categories*, 2001.

[11] Telecommunication Standardization Sector of ITU, ITU-T Recommendation Y.1541, *Network Performance objectives for IP-based services*, 2002.

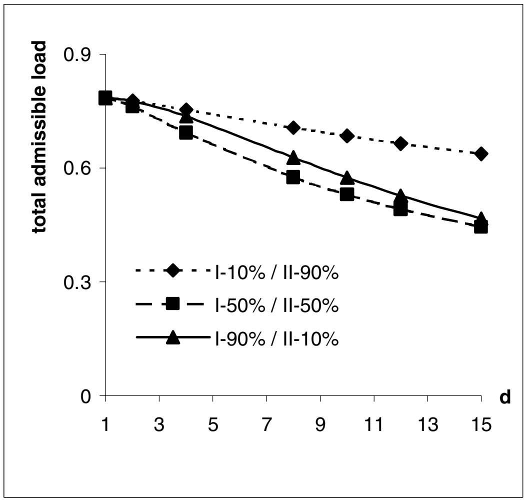

Figure 9. Total admissible load vs. packet size ratio of two end-to-end CoSs; target Ploss=10^{-2}

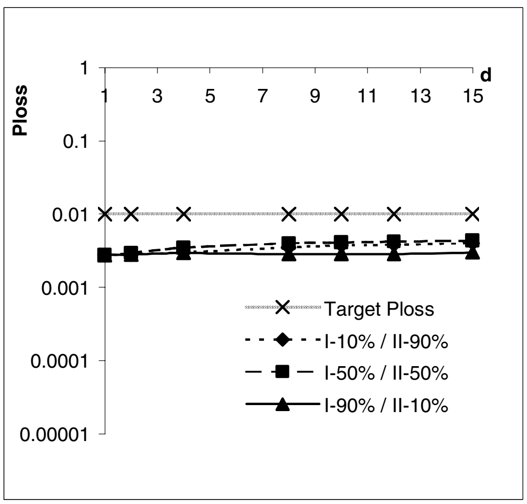

Figure 10. Packet loss ratio vs. packet size ratio of two end-to-end CoSs; target Ploss=10^{-2}

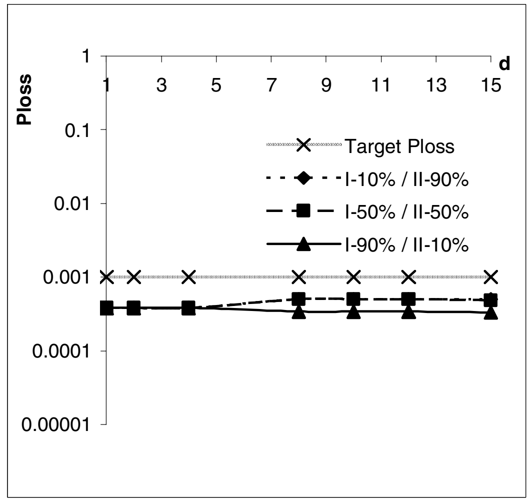

Figure 11. Packet loss ratio vs. packet size ratio of two end-to-end CoSs; target Ploss=10^{-3}

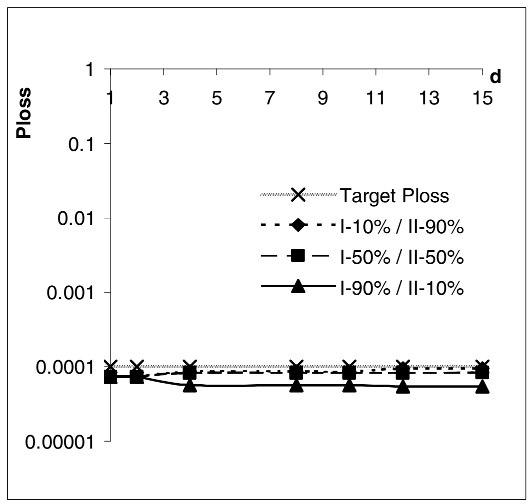

Figure 12. Packet loss ratio vs. packet size ratio of two end-to-end CoSs; target Ploss=10^{-4}