BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
DEPT. OF TELECOMMUNICATIONS AND MEDIA INFORMATICS

PERFORMANCE EVALUATION OF QOS ARCHITECTURES FOR PACKET-SWITCHED NETWORKS

Mátyás Martinecz

Summary of the Ph.D. Dissertation

Supervised by

Dr. József Bíró and Dr. Zalán Heszberger
High Speed Networks Laboratory
Dept. of Telecommunications and Media Informatics
Budapest University of Technology and Economics

Budapest, Hungary
2008

1 Introduction

We live the convergence of purpose-specific communication networks. The common principle where these various networks seems to meet is the packet-switched transmission, which - besides its many advantages - exposes several problems that should be dealt with. One of the main challenges in connection with packet-switched network is the provisioning of QoS (Quality of Service) assurances; without proper transmission quality ensuring mechanisms the penetration and marketability of novel, value-added services like the Voice over IP (VoIP) or Video on Demand (VoD) is out of the question. This is an important area, that should receive the distinguished attention of not only academic researchers but also telecommunication network operators and content providers.

There are many who tend to disregard the importance of QoS provisioning, especially those who agree with the "philosophy of large bandwidth". They vote for the over-dimensioning of packet-based networks stating that it implicitly assures that the quality requirements will be met. By now they seems to have been right, as the evolution of transmission technologies and the increase in data rates have outpaced the demand for bandwidth-consuming applications. This situation, however, may change in the future with the growing dominance of wireless and mobile telecommunication solutions, where capacity over-provisioning can either be impossible (due to the lack of free frequency bands) or too costly to be a competitive alternative. Naturally, this field has also seen a rapid development regarding the data rates on the air interface, nevertheless it is at least questionable whether available capacities will excess the ever-growing demand for bandwidth on the long run.

Considering the above-mentioned drawbacks of unreasonable overdimensioning a well-tuned QoS provisioning mechanism, e.g. an admission control policy coupled with a fairly accurate but at the same time simple and reliable resource requirement assessment algorithm can be less costly than simply expanding the resources in the network. The adequately precise knowledge of the resource requirements of the already admitted traffic flows may also be very useful as it enables the proper handling of critical or unexpected situations, e.g. when huge capacities become suddenly unavailable due to link failures or when there is an abrupt and unexpected increase in service demand.

In my dissertation I show efficient resource requirement assessment techniques, that are rooted in the Theory of Large Deviations. These new methods are capable of approximating the bandwidth or buffer size need of a certain aggregated traffic flow in case the maximum allowable information loss is prescribed. Also I present combined dimensioning methods that enable the provisioning of multi-level QoS, i.e. packet-level loss and call-level availability, assurances simultaneously. Furthermore I introduce a QoS architecture that can be applied primarily in access aggregation networks to ensure desirable transmission quality for value-added applications in a transparent manner.

2 Research Objectives

My main objective was to develop effective methods for dimensioning the main parameters - namely the buffer space and transmission capacity - of packet servers in case the maximum acceptable level of information loss is prescribed in terms of the workload loss ratio (WLR). Based on these simpler dimensioning techniques I was also aiming to design a combined - so-called multi-level QoS - dimensioning method that can take not only packet-level (i.e. the WLR), but also call-level (i.e. session blocking probability) QoS requirements into consideration. My ultimate goal was to incorporate these new contributions into a reliable, easy-to-implement QoS architecture that can provide satisfactory transmission quality for value-added applications (e.g. VoIP, VoD) primarily in broadband access aggregation network.

The main contributions of my research activity are organized into three thesis groups. In the first thesis group new equivalent capacity estimators are introduced, that are capable of approximating the minimum transmission capacity that is needed to fulfill a certain QoS goal composed in terms of target WLR. These approximation methods can play an essential role in the network capacity dimensioning activity when the network operator wants to provide reliable guarantees on the amount of information loss. The techniques presented in this part differ basically in their expected accuracy and performance, e.g. some of these formulae may give more accurate results than others in case the incoming traffic mix exhibits Gaussian nature, while other formulae may be less accurate, but can be evaluated more easily.

In the second thesis group of the dissertation efficient buffer requirement estimators are presented that are able to approximate the required size of the buffer in case the capacity is given and when a certain WLR level is prescribed. These formulae may serve as complementary techniques of the formulae presented in the first thesis group in the sense that they may be used when the transmission capacity cannot be arbitrarily increased (e.g. when the transmission link is a radio channel with limited frequency resources), but still the WLR requirements have to be met. The performance of the formulae presented here may significantly differ in different network scenarios so I also provide some hints about which formula to use under which circumstances.

The third thesis group of the dissertation contains three novel contributions. The first two are combined multi-level QoS dimensioning techniques, that are capable of estimating the required transmission capacity or buffer space a server node should possess in order to fulfill the prescribed QoS goals composed in terms of WLR and time blocking probability (TBP). These dimensioning methods serve as core elements of the QoS architecture presented in the last thesis of this thesis group. This QoS architecture was designed to operate in broadband access aggregation networks where providing satisfactory transmission quality for value-added applications is crucial. The main building blocks, the applied admission policy and the exploited protocols of this QoS architecture are also thoroughly discussed.

3 Methodology

During my research activity I was focusing on tractable, Large Deviation Theory-based models that are capable of describing single-node packet server systems with buffered communication links. These models were chosen because they are capable tools to characterize the performance of real-life packet routers.

The results in my dissertation were basically obtained through analytical methods. Since the analytical derivation of the formulae contains different simplifying assumptions and approximations, the effects of which are analytically hard to investigate, extensive numerical investigations were carried out to validate the performance of the new contributions.

4 New Results

4.1 Equivalent capacity approximation techniques

Novel, value-added applications may become widespread in the converged multi-service packet-switched networks only if reliable QoS assurances are provided. This statement is unquestionable, however, the ultimate, universal method answering the corresponding "how" question is still sought. The main concern here is - as usual - the cost effectiveness of the system to be implemented: the optimal system operating point on the curve characterizing the relation between resource savings and operational cost (or system complexity) is to be found.

The "heart and soul" of an efficient QoS provisioning mechanism is an effective resource requirement assessment method that is capable of determining resource needs of diverse traffic types with an adequate accuracy. Basically the resource requirement of a network service can be characterized by two parameters: bandwidth (or transmission capacity) and buffer size need. Obviously there is a tradeoff between these two measures, but in general there is always a principle or a constraint that helps to decide which parameter should be considered as given. In many cases the given parameter is the allocated buffer size, constrained by the maximum tolerable delay requirement of the application in question or simply by the technical parameters of the device under configuration. Thus in the majority of cases the minimum service rate of the server that satisfies the loss requirement of the corresponding applications is to be determined.

Numerous articles and papers deal with the problems of obtaining a good approximation on the capacity requirements of certain applications. Celebrated papers and studies on this topic are delivered by Kelly [7] and Roberts [8]. Courcobetis et al in [14] developed sp-called equivalent capacity formulae for the case when the buffer overflow probability is constrained. Seres et al in [3] developed methods that produce the same results as the formulae of Courcobetis, however have lower computational demand. Likhanov and Mazumdar [5] tried to approximate the WLR under the certain asymptotic regimes in order to form the basis of an improved equivalent capacity formula, while Jamin et al [12] and Courcobetis et al [27] were focusing on the measurement-based parameter acquisition methods that are associated with the equivalent capacity formulae.

In this section several capacity requirement estimation methods that are able to obtain the minimum transmission rate needed by a certain application to fulfill its pre-defined QoS goal composed in terms of the workload loss ratio is going to be introduced. The main difference distinguishing my methods from all others found in the literature is that they account for the workload loss ratio (while other usually regard the saturation probability of the buffer) and that they can be evaluated faster in comparison with the techniques that are based on the straightforward application of the definition of the equivalent capacity. The approximation techniques to be presented in this section basically differ from each other in their accuracy, computational demand and scope of applicability.

Let the stochastic process $X[0,t)$ denote the total amount of workload arriving in the time interval $[0,t)$ from N independent flows at a buffered communication link with buffer size B and transmission capacity C. N is regarded as a scaling factor and in that sense we can identify a per-source buffer $b = \frac{B}{N}$ and per source capacity $c = \frac{C}{N}$. Let us also assume that $X[0,t)$ has stationary increments.

The buffer overflow probability of this finite buffer system ($P(Q > B)$) can be deduced from the proportion of time over which the queue length $Q(C,N)$ is above level B in a queue of infinite buffer. In the many sources asymptotic (MSA) regime the decay rate of the logarithm of the buffer overflow probability is asymptotically linear in the number of sources N in a system where the per-source parameters are kept constant [16] [17]:

$$\lim_{N\to\infty} \frac{1}{N} logP(Q(N,cN) > bN) = \sup_{t>0} \inf_{s>0} \left\{ \frac{\Lambda(s,t)}{N} - s(b+ct) \right\} \stackrel{\text{def}}{=} -I, \tag{1}$$

where $\Lambda(s,t) \stackrel{\text{def}}{=} logE[e^{sX[0,t)}]$ is the cumulant generating function of $X[0,t)$, s and t are free parameters. Equation (1) is often referred to as many sources asymptotic equality of the Theory of Large Deviations.

The practical consequence of the above equation is that for N large one may approximate the buffer overflow probability simply as

$$P(Q(N,C) > B) \approx e^{-NI}, \tag{2}$$

where $-NI$ can be computed as

$$-NI = \sup_{t>0} \inf_{s>0} \left\{ \Lambda(s,t) - s(B+Ct) \right\}. \tag{3}$$

The QoS constraint on the workload loss ratio can be expressed formally as:

$$WLR \stackrel{\text{def}}{=} \frac{E[Q-B]^+}{E[X]} \leq e^{-\gamma}, \tag{4}$$

where X is the random variable characterizing the workload arriving in a time unit.

The minimal transmission capacity requirement - i.e. the minimal service rate at which the prescribed WLR threshold is still not exceeded - is called equivalent capacity and denoted by $C_{equ,WLR}$. Its definition is:

$$C_{equ,WLR} \stackrel{\text{def}}{=} \inf \left\{ C : WLR \leq e^{-\gamma} \right\}. \tag{5}$$

Theses 1. *[J2] [C4] I have developed techniques to estimate the equivalent capacity of a certain traffic aggregate in case the desired QoS level is composed in terms of the WLR. I have proven that some of my solutions preserve or may even outperform the desirable asymptotic properties of the solution of the many sources asymptotic equality (1). I have also pointed out that my formulae can be evaluated faster than other methods found in the literature.*

4.1.1 Theoretical equivalent capacity estimation method

Thesis 1.1. *I have developed the following fix-point equation to approximate $C_{equ,WLR}$*

$$c = sup_{t>0} inf_{s>0} \left\{ \frac{\Lambda(s,t) + F(\gamma, c)}{st} - \frac{B}{t} \right\} \stackrel{\text{def}}{=} f(c) \tag{6}$$

to be solved with respect to c, where $F(\gamma, c) = \gamma - logP\{Q > 0\} + logWLR(0)$. I have shown through analytical reasoning that the solution of (6) retains the desirable asymptotic properties of the solution of the many sources asymptotic equation (1).

The $c = f(c)$ fix point equation is to be solved iteratively, i.e $c_{n+1} = f(c_n)$, where n denotes the iteration step. The iterative process should be continued until the c and $f(c)$ values has a relative difference that is less than a predefined threshold, denoted by ϵ. This "stop condition" of the algorithm can be formally written as:

$$\frac{|c_n - f(c_n)|}{max\,\{c_n, f(c_n)\}} \leq \epsilon. \tag{7}$$

For the evaluation of (6) a double optimization task should be performed in each iteration step. This approach may yield an accurate enough solution faster than the definition-based (see (5)) technique, where finding the C value for the next trial is not straightforward.

4.1.2 Practical and conservative equivalent capacity approximation technique

The equivalent capacity approximation procedure discussed in Thesis 1.1 requires the computation of $P(Q > 0)$ in each iteration step. Estimating $P(Q > 0)$ however is a highly non-trivial task and usually it is advisable to be avoided. Some authors (e.g. Shroff et al in [6]) state that certain buffer saturation probability approximation techniques produce quite accurate result even for $B = 0$, nevertheless those techniques usually require quite a few extra parameters. Thus it seems wise to seek a round-about solution and eliminate $P(Q > 0)$ from the formulae in a way that yields a reliable, reasonably accurate and easily evaluable expression of the equivalent capacity.

The following thesis contains a modified version of the theoretical formula used in the fixed-point equation (6) that applies $P(X > C)$ as a substitute for $P(Q > 0)$. I state that solving the resulting fixed-point equation a reasonable, conservative estimate on the the solution of the fix-point equation (6) can be obtained. Let us denote the solution of the original fix point equation (6) by $\widetilde{C}_{equ,WLR}$, formally written as:

$$\widetilde{C}_{equ,WLR} \stackrel{\text{def}}{=} \lim_{n \to \infty} c_n, \tag{8}$$

where c_n, n positive integer denotes the result of (6) in the nth iteration step.

Thesis 1.2. *I have developed the following practical fix-point equation to estimate $C_{equ,WLR}$*

$$\hat{c} = sup_{t>0} inf_{s>0} \left\{ \frac{\Lambda(s,t) + \hat{F}(\gamma, \hat{c})}{st} - \frac{B}{t} \right\} \tag{9}$$

to be solved with respect to $\hat{c}$, where $\widehat{F}(\gamma, \hat{c}) = \gamma - logP\{X > \hat{c}\} + logWLR(0)$.
I have shown that the solution of (9) is a reasonable upper bound of $\widetilde{C}_{equ,WLR}$. I have also pointed out that this method produces an equivalent capacity estimation with less than 1% relative difference (i.e. $\epsilon = 0.01$) in a couple of iteration cycles in case the aggregated traffic flow consists of several tens or hundreds of ITU-T G.729 coded VoIP flows or MPEG-4 high-quality video flows.

Let us denote the solution of (9) by $\widehat{C}_{equ,WLR}$, formally written as:

$$\widehat{C}_{equ,WLR} \stackrel{\text{def}}{=} \lim_{n \to \infty} \hat{c}_n, \tag{10}$$

where $\hat{c}_n$, n positive integer denotes the result of (9) in the nth iteration step.

For the above-described method $P(X > C)$ and $WLR(0)$ has to be computed in each iteration step. As the calculation of $P(X > C)$ has been extensively studied in the literature it is easy to find an appropriate estimation method among them that works with the given parameters and computationally favorable. While the computation of $WLR(0)$ has been studied by only few authors, there still exist some techniques that require very few parameters and provide quite accurate result with moderate computational complexity. A good summary of such techniques is given by Heszberger et al in [31] .

It was mentioned in the previous section that the fix-point equation (6) may produce an accurate enough solution in fewer iteration steps than the definition based technique. This statement remain also true for the algorithm discussed in Thesis 1.2, however the accuracy of the final result may be worse than that of (6) or the definition based approach due to the applied $P(Q > 0) \approx P(X > C)$ approximation.

4.1.3 Modified equivalent capacity estimation methods

The underlying MSA approximation applied in the previous equivalent capacity estimation formulae (6) and (9) can be refined by applying the Bahadur-Rao theorem [15]. The resulting estimator would produce asymptotically more accurate estimations at the cost of increased parameter demand, i.e. for the Bahadur-Rao pre-factor the second derivative of the logarithmic moment generating function should be known. However, this new parameter can be eliminated from the formulae by applying a second order approximation of $\Lambda(s,t)$ around the optimal s, t values [13].

Based on the results of Courcobetis et al [14] and Seres et al [3] I have designed two such equivalent capacity estimators that has the same parameter demand and computational complexity as my previous estimators, but are capable to outperform the accuracy of those. This improved accuracy is primarily expected in case the aggregate arrival process has Gaussian nature as in this case the second order approximation of $\Lambda(s,t)$ is accurate.

Thesis 1.3. *I have developed the following two fix-point equations to approximate* $C_{equ,WLR}$

$$c^{B\cdot R} = sup_{t>0} inf_{s>0} \left\{ \frac{\Lambda(s,t) + G(F(\gamma, c^{B\cdot R}))}{st} - \frac{B}{t} \right\} \tag{11}$$

and

$$\hat{c}^{B\cdot R} = sup_{t>0} inf_{s>0} \left\{ \frac{\Lambda(s,t) + G(\hat{F}(\gamma, \hat{c}^{B\cdot R}))}{st} - \frac{B}{t} \right\} \tag{12}$$

to be solved with respect to $c^{B\cdot R}$ *and* $\hat{c}^{B\cdot R}$*, respectively, where* $G(x) \stackrel{\text{def}}{=} x - \frac{\frac{1}{2} log 4\pi x}{1+\frac{1}{2x}}$. *I have shown through analytical reasoning that the solutions of (11) have better asymptotic properties than the the solution of the many sources asymptotic equality (1) for aggregated traffic sources exhibiting strong Gaussian nature. I have also pointed out that the solution of (12) is an upper bound of the solution of (11).*

Let us denote the solutions of (11) and (12) by $\widetilde{C}^{B\cdot R}_{equ,WLR}$ and $\widehat{C}^{B\cdot R}_{equ,WLR}$, respectively. Formally written as:

$$\widetilde{C}^{B\cdot R}_{equ,WLR} \stackrel{\text{def}}{=} \lim_{n\to\infty} c^{B\cdot R}_n, \tag{13}$$

and

$$\widehat{C}^{B\cdot R}_{equ,WLR} \stackrel{\text{def}}{=} \lim_{n\to\infty} \hat{c}^{B\cdot R}_n, \tag{14}$$

where $c^{B\cdot R}_n$ and $\hat{c}^{B\cdot R}_n$, n positive integer denotes the result of (11) and (12) in the nth iteration step, respectively.

The (11) and (12) estimators are recommended to use primarily when the aggregate traffic exhibits Gaussian nature, as in that case they can provide more precise estimate of the equivalent capacity. Traffic mixes consisting of many independent traffic sources with identically distributed data rates (and finite variance) generate Gaussian-like traffic.

4.2 Buffer requirement estimators

Capacity dimensioning techniques are not always the sole, appropriate means of ensuring proper transmission quality for premium traffic. In many cases it is not possible or economic (e.g. in radio networks) to handle the increase in traffic demand only by the expansion of physical transmission capacity. More often - especially in case of bursty traffic sources - considerable transmission capacity can be saved by applying a larger buffer instead of raising the transmission rate.

Naturally, the buffer size enlargement has its own limitations: implementing a buffer of appropriate size and speed may have its financial and/or technological limits, but - and that is more important - the enlarged buffer also means higher transmission delay that may eventually result in unacceptable degradation of service quality. This problem becomes specially grave in case of delay-sensitive, value-added applications such as a VoIP or VoD service. These limitations have to be considered upon deciding which parameter (i.e. transmission rate or buffer capacity) of the queuing system is to be tweaked and to what extent that parameter can be changed.

The queue length distributions of various queuing systems have been extensively studied in the literature. In [24] the authors investigate the connection between cell loss probability and output buffer size in ATM switches. The book of Appenzeller et al [25] takes a more abstract, more general approach to this problem; they summarize the main contributions obtained in the field of queue length distribution characterization applying the results of the Large Deviation Theory. In the celebrated paper of Ganesh et al [26] the authors focus on formulating new rules of thumb for sizing router buffers fed by many congestion-controlled (basically TCP) flows.

In the following thesis group techniques that are capable to approximate the minimal buffer space of a queuing system with fixed transmission rate, that is needed to fulfill the prescribed quality level composed in terms of the workload loss ratio will be presented. The first approximation method is based on (1) and the results of Seres et al [3] and Shroff et al [6]. The second formula to be introduced can be considered as a practical upper bound of the measure computed with the first method and it builds upon the result of Borsos and György [4]. The third contribution of mine incorporates the Bahadur-Rao pre-factor [15] in the underlying QoS measure approximation and thus it may yield a more precise result than the previous formulae, especially in case when the aggregated arrival process is Gaussian.

Let the stochastic process $X[0,t)$ denote the total amount of workload arriving in the time interval $[0,t)$ from N independent flows at a buffered communication link with transmission capacity C. Let us also assume that $X[0,t)$ has stationary increments.

Let us define the buffer space requirement - i.e. the minimal buffer size at which the prescribed WLR threshold ($e^{-\gamma}$) is still not exceeded - as:

$$B_{req,WLR} \stackrel{\text{def}}{=} \inf\left\{B : WLR \le e^{-\gamma}\right\}. \tag{15}$$

Theses 2. *I have developed techniques to estimate the buffer requirement of a certain traffic aggregate in case the desired QoS level is composed in terms of the WLR. I have analytically proven that some of my solutions preserve or under special circumstances outperform the asymptotic properties of the solution of the many sources asymptotic equality (1). I have also pointed out that my formulae have lower computational complexity and thus can be evaluated faster than other methods found in the literature.*

4.2.1 Theoretical buffer requirement estimation method

Thesis 2.1. *I have constructed the following formula to approximate* $B_{req,WLR}$

$$B_{req,WLR} \approx sup_{t>0} inf_{s>0} \left\{\frac{\Lambda(s,t) + F(\gamma, C)}{s} - Ct\right\} \stackrel{\text{def}}{=} \widetilde{B}_{req,WLR}, \tag{16}$$

where

$$F(\gamma, C) = \gamma - logP\{Q > 0\} + logWLR(0).$$

I have shown that $\widetilde{B}_{req,WLR}$ *retains the asymptotic properties of the solution of the many sources asymptotic equality (1). Furthermore I have pointed out that (16) has lower computational complexity than other methods found in the literature and thus can be evaluated faster.*

For the evaluation of (16) a double optimization has to be carried out with respect to the free parameters of s and t. This method is equivalent (i.e. it yields the same result) with the the so-called implicit method, where the buffer requirement is sought based on its definition presented in (15). By following the implicit method, however, a triple optimization should be carried out as the expected WLR has to be approximated through solving a two-dimensional optimization task in each iteration step. As a conclusion the explicit buffer requirement estimator presented in (16) is computationally more feasible than the natural, implicit estimation technique.

4.2.2 Practical and conservative buffer requirement approximation technique

Finding a good estimate on $P(Q > 0)$ for the evaluation of (16) may be a problem. Therefore I have proposed another method that eliminates the need for obtaining $P(Q > 0)$ while still providing a reasonable and computationally tractable approximation on the buffer requirement of the queuing system. The main idea of this method is that $P(Q > 0)$ is substituted with its lower bound, $P(X > C)$ which results in a formula that gives a conservative upper bound on the buffer requirement estimated with (16).

Thesis 2.2. *I have developed the following practically tractable approximation for $B_{req,WLR}$:*

$$\widetilde{B}_{req,WLR} \leq sup_{t>0} inf_{s>0} \left\{ \frac{\Lambda(s,t) + \widehat{F}(\gamma, C)}{s} - Ct \right\} \stackrel{\text{def}}{=} \widehat{B}_{req,WLR}, \tag{17}$$

where

$$\widehat{F}(\gamma, C) = \gamma - logP\{X > C\} + logWLR(0).$$

I have also shown that $\widehat{B}_{req,WLR}$ is an upper bound of $\widetilde{B}_{req,WLR}$.

For the evaluation of the buffer requirement estimator presented in (17) the bufferless saturation probability (i.e. $P(X > C)$) needs to be obtained. This measure can be approximated much easier than $PQ > 0$, which makes this method more tractable in practice.

4.2.3 Modified buffer requirement estimation methods

The methods in Thesis 2.1 and 2.2 use an underlying $P(Q > B)$ estimation directly rooted in the MSA equality (1). This WLR estimation can however be further improved by applying the Bahadur-Rao theorem.

In the following thesis I will present two buffer requirement estimators that apply improved underlying $P(Q > B)$ approximation techniques that incorporate the Bahadur-Rao pre-factor. Using the B-R pre-factor may yield more precise estimation on $P(Q > B)$ and as a consequence a more precise estimation on the buffer requirement.

An important property of the formulae to be presented is that they still manifest in the same, simple, explicit forms as the other ones already introduced in this section, and therefore their computational complexity remains in the same order of magnitude as of those.

Thesis 2.3. *I have constructed the following estimators for $B_{req,WLR}$:*

$$B_{req,WLR} \approx sup_{t>0} inf_{s>0} \left\{ \frac{\Lambda(s,t) + G(F(\gamma, C))}{s} - Ct \right\} \stackrel{\text{def}}{=} \widetilde{B}^{B-R}_{req,WLR} \tag{18}$$

and

$$\widetilde{B}^{B \cdot R}_{req,WLR} \leq sup_{t>0} inf_{s>0} \left\{ \frac{\Lambda(s,t) + G(\widehat{F}(\gamma, C))}{s} - Ct \right\} \stackrel{\text{def}}{=} \widehat{B}^{B \cdot R}_{req,WLR}, \tag{19}$$

where $G(x) \stackrel{\text{def}}{=} x - \frac{\frac{1}{2}log4\pi x}{1+\frac{1}{2x}}$.
I have shown that $\widetilde{B}^{B-R}_{req,WLR}$ *has better asymptotic properties than the solution of the many sources asymptotic equality (1) in case the aggregated traffic exhibits strong Gaussian nature. I have also pointed out that* $\widehat{B}^{B\cdot R}_{req,WLR}$ *is a tractable, reasonable upper bound of* $\widetilde{B}^{B-R}_{req,WLR}$.

It can be observed that Formulae (18) and (19) - while in theory being more precise than the buffer requirement estimation formulae (16) and (17) - manifest in a very similar form as those. This also implies that the computational complexity of these formulae is very close to each other. This advantageous property has a price, however. To be able to incorporate the original B-R pre-factor into the buffer saturation probability estimation a second-order approximation had to be carried out. This second-order approximation is only precise in case the arrival process is Gaussian and so the formulae presented in this section should primarily be used in case the aggregated arrival process is well-characterized by a Gaussian process, for example a fractional Brownian motion.

4.3 Techniques for providing multi-level QoS guarantees

The results presented in the previous sections can be applied to solve capacity planning and buffer dimensioning problems in case the QoS requirement of the network flows is composed in terms of the workload loss ratio. Real applications, however, compose their QoS needs not only in terms of loss, but often also in terms of delay and/or delay variation. This statement is especially true in case of value-added, multimedia services such as a Video-on-demand service, which is becoming popular nowadays. This implies that more sophisticated methods - that are capable of taking the diverse QoS needs of traffic flows into consideration - should be applied for real network dimensioning activities.

Another important aspect worth to be considered upon dimensioning is the call-level behavior of the expected traffic, i.e. the understanding of the birth-death process of network flows. Taking this kind of traffic dynamics into account makes the whole dimensioning process more precise and economical and also guarantees that the service availability will reach the desired level.

Dimensioning methods found in the literature usually investigate traffic demand on only one, distinct granularity level. For instance papers and books dealing with the dimensioning issues of telephony (or in general circuit switched) networks usually focus on the call-level granularity level and aim to formulate bounds on the expected availability measure, often on the time blocking probability. Famous basic results in this field are attached to the names of Erlang [9], Engset [10] and Kleinrock [11].

Another group of researchers focus on the inner dynamics of traffic flows and investigate traffic behavior on the packet granularity level. Their aim usually is to obtain bounds or approximations on the expected loss occurring in the system under study. Celebrated papers dealing with this topic can be coupled with the work of Kelly, Courcobetis and Roberts just to name a few pioneers in this research area [7][14] [8].

In the following thesis group a combined dimensioning method aiming to ensure simultaneous call- and packet level QoS guarantees is going to be shown. This technique is capable of providing guaranteed availability composed in terms of time blocking probability, and at the same time attaining the desired loss level composed in terms of workload loss ratio.

Based on the combined dimensioning technique a QoS architecture designed for providing multi-level QoS assurances in broadband access networks is also going to be introduced. The basic building blocks, exploited protocols and the recommended admission control scheme is also discussed.

Let the call-level dynamics be described by the Engset model [10] (see Figure 1). Let N denote the server capacity (e.g. number of linecards), P denote the size of the population, λ denote the service (or call-) initiation intensity and h denote the holding time of a call. The system accepts calls when the number of ongoing calls is

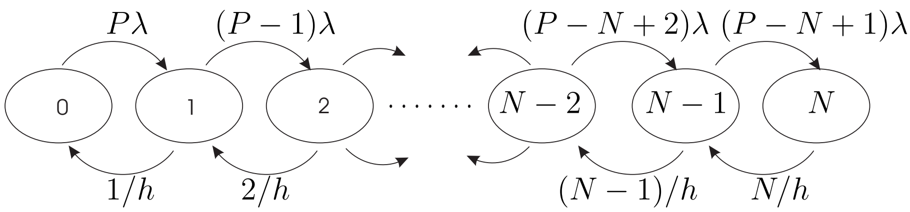

Figure 1: The applied call-level (Engset) model

less than N, and rejects every call that arrives during the period when all linecards are busy.

Let us define the combined equivalent capacity (C^*) as the minimum transmission capacity (for a given buffer size) that ensures that both the expected WLR level remains under a certain threshold $e^{-\gamma}$ and the TBP level does not exceed β.

$$C^* \stackrel{\text{def}}{=} inf\left\{C : WLR \leq e^{-\gamma}, TBP \leq \beta\right\} \tag{20}$$

Similarly, define the combined buffer requirement (B^*) as the minimum buffer space (for a given transmission capacity) that ensures that both the expected WLR level remains under a certain threshold $e^{-\gamma}$ and the TBP level does not exceed β.

$$B^* \stackrel{\text{def}}{=} inf\left\{B : WLR \leq e^{-\gamma}, TBP \leq \beta\right\} \tag{21}$$

Theses 3. ***[J1] [C2] [C3]***

I have developed combined resource dimensioning techniques, that build upon the dimensioning formulae introduced in the previous thesis groups and are able to account for both call-level availability and packet-level loss prescriptions. I have also designed a robust QoS architecture that simplifies resource management related tasks through the exploitation of the potentials of the combined dimensioning methods. The architecture applies a distributed admission control mechanism with flexible policy to provide transparent, reliable multi-level QoS guarantees for premium services in the access aggregation segment of packet-switched networks.

4.3.1 Capacity dimensioning method ensuring multi-level QoS

Thesis 3.1. *I have developed the following fix-point equation to approximate the combined equivalent capacity*

$$c^* = sup_{t>0} inf_{s>0} \left\{ \frac{\Lambda_{N^*}(s,t) + K}{st} - \frac{B}{t} \right\} \tag{22}$$

to be solved with respect to c^**, where* B *is the buffer size,* s *and* t *are free parameters,* $K \in \{F(\gamma, c^*), \hat{F}(\gamma, c^*), G(F(\gamma, c^*)), G(\hat{F}(\gamma, c^*))\}$ *and* $\Lambda_{N^*}(s,t)$ *is the cumulant generating function of* $X_{N^*}(0,t]$*, the random variable denoting the workload arriving from*

N^* sources in time interval $(0, t]$. The number of sources (N^*) should be computed with

$$N^* = inf \left\{ N : \frac{\begin{pmatrix} P \\ N \end{pmatrix} (\lambda h)^N}{\sum_{j=0}^{N} \begin{pmatrix} P \\ j \end{pmatrix} (\lambda h)^j} \leq \beta \right\}. \tag{23}$$

Simply spoken the method described in Thesis 3.1 tells us that if the population size is P and calls arrive with λ intensity and h lifetime to the system then we should provide a transmission channel that accepts maximum N^* simultaneous flows (or calls) to ensure that the TBP remains under the threshold. According to the methods discussed in Theses 1 N^* flows require at least (22) capacity in order to fulfill the WLR prescription.

The above-described dimensioning methods work properly in case the models characterizing the call and packet level dynamics of the offered traffic load are precise enough and the packet server applies a simple admission control policy admitting the service of only N^* simultaneous flows.

4.3.2 A buffer sizing method providing multi-level QoS

A similar theorem can be composed for the combined buffer requirement, B^*.

Thesis 3.2. *I have constructed the following formula to estimate B^**

$$B^* \approx sup_{t>0} inf_{s>0} \left\{ \frac{\Lambda_{N^*}(s,t) + L}{s} - Ct \right\}, \tag{24}$$

where C is the transmission capacity, s and t are free parameters, $L \in \{F(\gamma, C), \hat{F}(\gamma, C), G(F(\gamma, C)), G(\hat{F}(\gamma, C))\}$ and $\Lambda_{N^*}(s,t)$ is the cumulant generating function of $X_{N^*}(0,t]$, the random variable denoting the workload arriving from N^* sources in time interval $(0,t]$. The number of sources (N^*) should be computed with (23).

The combined buffer dimensioning technique operates reliably if the call and packet level dynamics of the offered traffic are precisely described and only N^* flows are admitted to enter the buffered link simultaneously.

4.3.3 QoS architecture for broadband access networks

Thesis 3.3. *I have constructed a robust QoS architecture that simplifies resource management related tasks through the exploitation of the potentials of the combined dimensioning methods introduced in Thesis 3.1 and Thesis 3.2. The architecture applies*

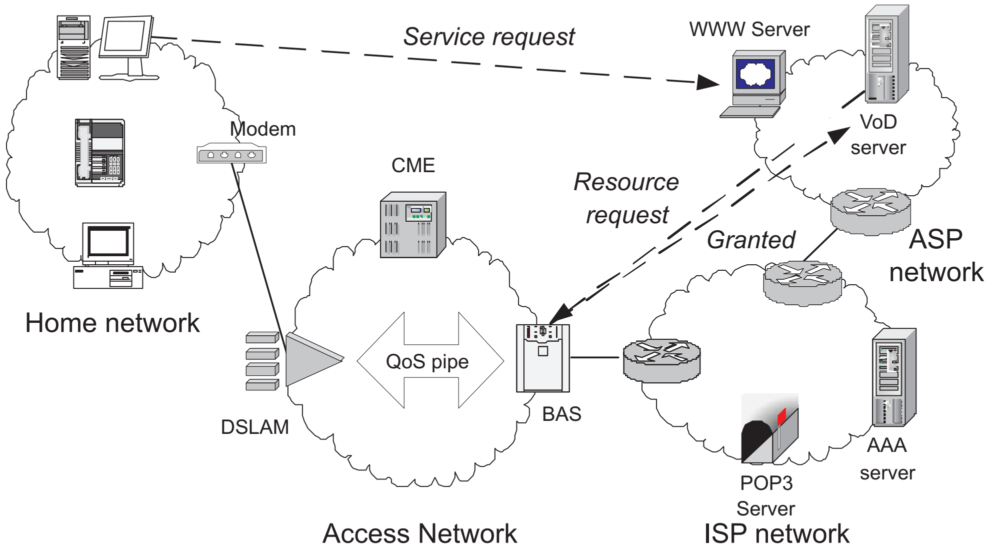

Figure 2: The model of the recommended QoS architecture

a distributed admission control mechanism with flexible policy to provide transparent, reliable multi-level QoS guarantees for premium services in the access aggregation segment of packet-switched networks.

The main concept of the architecture is that an interface is provided between the NAP (Network Access Provider) and the ASP (Application Service Provider) through which QoS- and resource reservation-related information may be exchanged in a manner that is fully transparent to subscribers. The core element (see Figure 2) is a central entity - called CME (Central Management Entity) which performs two tasks: it defines a logical overlay network upon the physical topology and manages resource (re-)allocation related tasks. The predefined overlay network consists of logical point-to-point trunks between the edge nodes of the AN. These logical channels are called QoS pipes, as there are one such channel for each QoS class between each access node - edge node pair. To gather the needed information for the operation of the QoS architecture certain functions of currently available protocols such as the RSVP or the DHCP are also exploited.

The majority of traffic management related tasks - including admission control, traffic metering, shaping, policing and pricing - are delegated to the Broadband Access Servers (BASs). Broadband Access Servers are located on the border between the AN operator's and the ISPs' networks, and so all traffic addressed to a certain subscriber should travel through these network node. They are less in numbers than the DSL Access Multiplexers (DSLAMs), are usually located at the premises of the AN operator and have increased intelligence compared to DSLAMs. These properties make them the proper candidate for accomplishing extra functions, such as the ones mentioned above.

The nodes located in the inner region of the AN are ATM or Ethernet switches with limited intelligence and function sets. Their task is to forward traffic incoming at a certain port to another outgoing port. Also they should be configurable to handle traffic according to their priority implicitly indicated by their QoS class. The scheduling mechanism of these switches should support bandwidth and buffer space allocation on PVP or VLAN ID basis. This latter functionality ensures the realization of the independent QoS pipes with dedicated resources discussed earlier, as these pipes are to be identified by their PVP or VLAN ID.

The fourth element of the architecture consist of the already mentioned DSLAMs. These ATM or Ethernet switches are located at the boundary of the AN close to the subscribers' home networks and have only one task, i.e. to aggregate (distribute) and forward traffic arriving from (traveling to) the subscribers. These nodes should also be capable of handling different QoS classes according to their priority. They also serve as end-points of the QoS pipes.

The operation of the architecture is the following. The CME entity creates the initial logical overlay network by establishing QoS pipes between the edge nodes of the AN. The creation of QoS pipes is in fact the configuration of DSLAM, switch and BAS nodes so that each of these would be aware of how to handle traffic belonging to different PVPs or VLANs and how much capacity (fraction of service cycle) and buffer space is dedicated to a certain PVP or VLAN. The initial amount of resources allocated to each QoS pipe is determined with the aid of the combined dimensioning method discussed in Thesis 3.1 and Thesis 3.2.

Once the logical overlay network is created traffic management related activities are carried out primarily at the BAS nodes. These nodes continuously monitor the load of the QoS pipes (that are starting at them) and make admission control decisions in a manner that is transparent to both the subscriber and the Application Service Provider (ASP).The traffic monitoring and metering is used for measurement-based equivalent capacity and/or buffer requirement estimations, i.e. the aggregated peak and mean data rate of the traffic is measured and the cumulant generating function of the aggregated traffic is approximated with the method described in [27]. The measured parameters accompanied by the prescribed QoS goals are then fed to one of the resource requirement estimator algorithms discussed in Theses 1 and Theses 2.

The transparency of the admission control process is ensured through additional functionality of the BAS, namely DHCP traffic monitoring and RSVP message capturing. DHCP monitoring is needed to couple assigned IP addresses and QoS pipes groups, in other words to determine through which DSLAM the subscriber with a certain IP address can be reached.

The RSVP capturing is used for impersonating the subscriber according to the following process. The BAS captures RSVP messages addressed to the subscriber initiating the service and checks whether the resource requirements described in the RSVP message can be fulfilled in the QoS pipe that would transport the service. Then, in accordance with the result of this check, the BAS answers in lieu of the subscriber whether the request can be granted. For the admission control decision the BAS need to couple the destination IP address of the RSVP message with the corresponding QoS pipe and this is why the DHCP monitoring functionality is needed in the BAS nodes. Also the BAS needs to be aware of the current load of that pipe and this is done by the monitoring and resource requirement assessing function.

The recommended admission control process is the following:

- *Step 1* The subscriber in its home network would like to connect to the Internet. Therefore, it requests IP address from its ISP (Internet Service Provider) using the widespread DHCP protocol.

- *Step 2* The DHCP communication is monitored by the BAS that is located on the boundary of the AN operator's and the ISP's networks. The BAS stores the IP address that has been assigned to the subscriber in its local database and also the identifier of the DSLAM which the subscriber is connected to. (These pieces of information are needed later to determine the identifier of the QoS pipes through which the subscriber can employ a certain type of service.)

- *Step 3* The subscriber wants to make use of a premium service provided via the Internet. She sends a service requests to the web server of the ASP by clicking on the appropriate hyperlink for example.

- *Step 4* The web server forwards the requests to the content server that is aware of the statistical characteristics and QoS requirements of the required service.

- *Step 5* The content server requests network resources by sending an RSVP message to the subscriber.

- *Step 6* The RSVP message is captured by the BASs. It looks up the DSLAM identifier of the subscriber based on the IP address and determines the appropriate QoS class which the new session would belong to based on the required QoS parameters in the RSVP message. Finally, the BAS checks resource availability in the appropriate QoS pipe.

- *Step 7* The BAS makes its admission decision. According to actual resource availability three types of actions may be taken at this point by the BAS. These are

- If there is enough capacity available in the appropriate pipe, the access should be granted. The BAS informs the content server about the positive decision in an RSVP message. In parallel, its database that tracks available capacities in the QoS pipes is refreshed (e.g. available capacity decreased by the peak rate of the newcomer flow).
- If there is not enough capacity available in the appropriate pipe, but the pipe can be expanded to a size that can cope with the new request, the BAS requests the expansion of the pipe from the CME. The CME performs the expansion and informs the BAS about it. The BAS then sends an RSVP message to the content server telling that its resource request is granted. In parallel, the database of the BAS that tracks available capacities in the QoS pipes is refreshed.
- If there is not enough capacity in the appropriate QoS pipe and it can not be expanded to cope with the newcomer flow, the resource request is rejected and the content server is informed about the decision in an RSVP message. The database of the BAS that tracks available capacities in the QoS pipes - of course - remains unchanged.

5 Applicability of the Results

The equivalent capacity estimation techniques proposed in the first group can be effectively used to reserve adequate capacity for certain traffic aggregates and also to build reliable admission control methods upon them. They also form as a basic element of the combined capacity planning method introduced in the third thesis group.

The formulae presented in the second thesis group can be applied to approximate the buffer size demand of an aggregated network traffic. They should primarily be used when the transmission capacity of the server is limited or too expensive to be expanded (e.g. it requires extra radio resources). They serve as complementary methods of the techniques presented in the first thesis group.

The combined dimensioning methods presented in the third thesis group can be used in case a network operator wants to ensure that value-added services perceive not only satisfactory transmission quality, but also appropriate availability in terms of time blocking probability. These combined resource need assessment techniques form the basis of the QoS architecture presented in Thesis 3.3. The Qos architecture aims to provide simple, efficient and transparent quality guarantees for premium network flows traversing broadband access network by applying simplified network management mechanisms, admission controlling and exploiting the functionalities of widely used protocols.

The majority of the presented results of the dissertation have been published in Hungarian and international forums. Two of these publications won best paper award.

References

[1] S.R.S. Varadhan, "Asymptotic probability and differential equations", *Comm. Pure Appl. Math.* vol. 19, pages 261-286., 1966.

[2] C. Bouchat, S. van den Bosch, T. Pollet, "QoS in DSL Access", *IEEE Communications Magazine* vol 41. no. 9, Page(s): 108–114, November 2003.

[3] G. Seres, A. Szlávik, J. Zátonyi and J. Bíró, "Quantifying resource usage - a large deviation-based approach", *IEICE Transactions on Communications (Japan)*, E85-B(1):938–952, 2002.

[4] A. György, T. Borsos, "Estimates on the packet loss ratio via queue tail probabilities", *Global Telecommunications Conference 2001., GLOBECOM '01.*, IEEE, Volume 4, Issue 2001, Page(s): 2407 - 2411 vol.4, 2001.

[5] N. Likhanov and R. R. Mazumdar, "Cell loss asymptotics in buffers fed with a large number of independent stationary sources", *INFOCOM '98. Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies*, Volume 1, Page(s):339 - 346, March 1998.

[6] H.S. Kim, N.B. Shroff, "Loss probability calculations and asymptotic analysis for finite buffer multiplexers", *IEEE/ACM Transactions on Networking*,Volume 9, Issue 6, Page(s):755 - 768, December 2001.

[7] F. P. Kelly. "Notes on effective bandwidths" *Stochastic Networks: Theory and Applications*, vol 4, pp. 141–168, Oxford University Press, 1996.

[8] T. Bonald, P. Olivier, J. Roberts, "Dimensioning high speed IP access networks", *Proceedings of ITC-18*, pages 241–251, Aug. 2003.

[9] A. K. Erlang, "Solution of some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges", *Elektrotkeknikeren*, vol 13, 1917.

[10] E. Pinsky, A. Conway, Wei Liu, "Blocking formulae for the Engset model ", *IEEE Transactions on Communications*, Volume: 42, Issue: 6, page(s): 2213-2214, Jun 1994

[11] L. Kleinrock, "Queueing Systems. Vol I: Theory", pp. 93–94, John Wiley & Sons, New York - London - Sydney - Toronto, 1975.

[12] S.Jamin , P. B. Danzig, S. J. Shenker, L. Zhang , "A measurement-based admission control algorithm for integrated service packet networks", *IEEE/ACM Transactions on Networking (TON)*, Volume 5 , Issue 1, Pages: 56 - 70, 1997

[13] M. Montgomery and G. de Veciana, “On the relevance of time scales in performance oriented traffic characterizations”, *Conference on Computer Communications (IEEE Infocom)* ,Volume 2, Page(s): 513 - 520, March 1996.

[14] C. Courcobetis, V. A. Siris and G. D. Stamoulis, “Application of the many sources asymptotic and effective bandwidths to traffic engineering”, *Telecommunication systems*, 12:167-191, 1999.

[15] R. R. Bahadur and R. Rao, “On deviations of the sample mean”, *The Annals of Mathematical Statistics*, Vol. 31, No. 4, page(s): 1015-1027. December 1960.

[16] D. D. Botvich and N.G. Duffield, “Large deviations, the shape of the loss curve, and economics of scale in large multiplexers”, *Queuing Systems*, Vol. 20, page(s): 293-320, 1995.

[17] C. Courcobetis and R. Weber, “Buffer overflow asymptotics for a buffer handling many traffic sources”, *Journal of Applied Probability*, Vol. 33, page(s): 886-903, 1996.

[18] J. Bíró, Z. Heszberger, M. Martinecz, Sz. M. Kis, “Novel Equivalent Capacity Approximation Through Asymptotic Loss Analysis”,*HET-NETs '04, First International Working Conference on Performance Modelling and Evaluation of Heterogenous Networks*, page(s): 123-129,July 2004.

[19] A. K. Parekh, R. G. Gallager, “A Generalized Processor Sharing approach to Flow control in Integrated Services Networks: The Single Node Case”, *IEEE/ACM Transactions on Networking (TON)*, Volume 1, Issue 3, Pages: 344 - 357, 1993

[20] J.-Y. Le Boudec, “Application of Network Calculus to Guaranteed Service Networks”, *IEEE Transactions on Information Theory*, 44(3):1087-1096, May 1998.

[21] J.-Y. Le Boudec and P. Thiran, “Network Calculus: A Theory of Deterministic Queuing Systems for the Internet”, *LNCS: Lecture Notes in Computer Science* , Springer, 2001.

[22] R. L. Cruz, “A Calculus for Network Delay. Part I: Network Elements in Isolation and Part II: Network Analysis” , *IEEE Transactions on Information Theory*, 37(1):114-141, Jan. 1991.

[23] M. Fidler, “An End-to-End Probabilistic Network Calculus with Moment Generating Functions ”, *IEEE International Workshop on Quality of Service (IWQoS)*, Jun. 2006.

[24] H. Bruneel, B. G. Kim, "Discrete-time Models for Communication Systems Including ATM", *Kluwer Academic Publishers*, ISBN:0792392922, 1993.

[25] G. Appenzeller, I. Keslassy, N. McKeown, "Sizing Router Buffers", *ACM SIGCOMM*, Volume 34, Issue 4, Pages: 281 - 292, October 2004

[26] A. Ganesh, N. O'Connell, D. Wischik, "Big Queues (Lecture Notes in Mathematics)", *Springer*, ISBN:3540209123, May 1, 2008

[27] C. A. Courcoubetis, A. Dimakis, G. D. Stamoulis, "Traffic equivalence and substitution in a multiplexer withapplications to dynamic available capacity estimation", *IEEE/ACM Transactions on Networking*, Volume 10, Issue 2, page(s) 217-231, ISSN: 1063-6692, Apr 2002

[28] G. Mao and D. Habibi. "Loss performance analysis for heterogeneous on-off sources with application to connection admission control" *IEEE/ACM Transactions on Networking (TON)*, 10(1):125–138, 2002.

[29] I. Norros, "A storage model with self-similar input", *Queuing Systems*, Volume 16, Numbers 3-4, pages 387-396, September, 1994

[30] R. J. Gibbens, Y. C. Teh, "Critical time and space scales for statistical multiplexing in multi service networks", *Proceedings of the International Teletraffic Congress ITC'16*, pages 87-96, Edingburgh, Scotland, 1999.

[31] Z. Heszberger, J. Zátonyi, and J. Bíró, "Performance bounds for rate envelope multiplexing", *Performance Evaluation*, Volume 48, Issue 1, May 2002, Pages 87-101. Special Issue on ATM & IP Networks: Performance Modelling and Analysis

[32] F.H.P. Fitzek and M. Reisslein, "MPEG-4 and H.263 Video Traces for Network Performance Evaluation", *IEEE Network*, 15(6):40-54., 2001

[33] András Kern, Mátyás Martinecz and Zalán Heszberger, "Architecture and Configuration of Broadband Access Networks Supporting Multimedia Applications", *The Tenth IEEE Symposium on Computers and Communications*, June. 2005

[34] "Dynamic Host Configuration Protocol (DHCP)", *RFC 2131*, www.ietf.org

[35] "Resource ReSerVation Protocol (RSVP)", *RFC 2205*, www.ietf.org

[36] "An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks", *RFC 2205*, www.ietf.org

Publications

Journal papers

[J1] M. Martinecz, A. Kern, Z. Heszberger and J. Bíró, "Architecture and Configuration of Broadband Access Networks Supporting Multimedia Applications", *IJCA Journal Special Issue.* pages 34–46, April, 2007.

[J2] M. Martinecz, J. Bíró and Z. Heszberger, "Novel Techniques for Assessing Resource Requirements in Packet-based Networks", *Híradástechnika - Telecommunications - Hungarian periodical, Selected Papers.* pages 24–29, June, 2005.

[J3] J. Bíró, A. Gulyás and M. Martinecz, "Parsimonious Estimates of Bandwidth Requirement in Quality of Service Packet Networks", *Performance Evaluation*, Volume 59, Issues 2-3, p. 159-178, February 2005.

[J4] M. Martinecz, J. Bíró and Z. Heszberger, "Újszerű erőforrásigény-becslő módszerek csomagkapcsolt hálózatokban", *Híradástechnika - Telecommunications - Hungarian periodical.* pages 13–18, Sept, 2004.

Conference papers

[C1] M. Martinecz, J. Bíró, Z. Heszberger, "A Logical Channel Dimensioning Method Enabling Simple QoS Provision in Aggregation Networks", *World Telecommunications Congress (WTC) 2006* , Budapest, Hungary, May 2006.

[C2] M. Martinecz, A. Kern, J. Bíró, "Multi-level QoS Guarantees in Broadband Access Networks", In *HET-NETs '05, First International Working Conference on Performance Modelling and Evaluation of Heterogenous Networks*, Ilkley, West Yorkshire, U.K., July 2005.

[C3] A. Kern, M. Martinecz and Z. Heszberger, "Architecture and Configuration of Broadband Access Networks", In Proc. of *ISCC'2005, The Tenth IEEE Symposium on Computers and Communications*, volume 1, pages 172–181, La Manga del Mar Menor, Cartagena, Spain, June, 2005.

[C4] M. Martinecz, J. Bíró, Z. Heszberger, Sz. M. Kis, "Novel Equivalent Capacity Approximation Through Asymptotic Loss Analysis", In *HET-NETs '04, First International Working Conference on Performance Modelling and Evaluation of Heterogenous Networks*, Ilkley, West Yorkshire, U.K., July 2004.

[C5] J. Bíró, Z. Heszberger, M. Martinecz, N. Felicián and Octavian Pop, "Towards a Framework of QoS Measure Estimates for Packet-Based Networks", In Proc. of *IEEE International Conference on Communications - ICC 2004*, volume 4, pages 2231–2235, Paris, France, June, 2004.

[C6] J. Bíró, Z. Heszberger and M. Martinecz, "A Family of Performance Bounds for QoS Measures in Packet-Based Networks", *Lecture Notes in Computer Science - LNCS 3042 (Proc of IFIP Networking 2004)*, pages 1108–1119, Athens, Greece, May, 2004.

[C7] J. Bíró, Z. Heszberger, and M. Martinecz, "Equivalent Capacity Estimators for Bufferless Fluid Flow Multiplexing", In Proc. of *IEEE Global Telecommunications Conference - Globecom 2003*, volume 7, pages 3706–3710, San Francisco, CA, USA, Dec. 2003.

[C8] J. Bíró, Z. Heszberger, T. Dreilinger, A. Gulyás, and M. Martinecz, "Parsimonious Estimates of Bandwidth Requirement in Quality of Service Packet Networks", In Proc. of *First International Working Conference on Performance Modeling and Evaluation of Heterogenous Networks - HET-NETs '03*, pages 69/1-69/9, Ilkley, UK, 21-23 July 2003.

[C9] J. Bíró, Z. Heszberger, F. Németh, M. Martinecz, "Bandwidth Requirement Estimators for QoS Guaranteed Packet Networks", In Proc. of *International Network Optimization Conference - INOC 2003*, pages 41–49, Evry-Paris, France, May, 2003.

[C10] J. Bíró, Z. Heszberger, G. Kún and M. Martinecz, "Advanced QoS Provision for Real-Time Internet Traffic", In *IEEE International Packet Video Workshop - Packet Video 2003*, pages 26–28, Nantes, France, April, 2003.

[C11] J. Bíró, Z. Heszberger, M. Martinecz, "Measurement-based QoS Provision for VoIP Traffic", In *Proceedings of Transcom 2003 Conference*, pages 195-200, Zilina, Slovakia, June 2003.

[C12] J. Zátonyi, Z. Heszberger, J. Bíró, M. Martinecz and B. Héjj, "Providing Guaranteed QoS for Real-Time Internet Traffic", In *Proceedings of Polish-Czech-Hungarian Workshop on Circuit Theory and Applications*, pages 65-70, Warsawa, Poland, September 2002.

[C13] B. Héjj, M. Martinecz, "Minőségbiztosított forgalmak elemzése és szimulációs vizsgálata IP hálózatokban", In *Proceedings of National Student Conference on Information Sciences (OTDK)*, Győr, April 2003.