ASSESSMENT OF TOKEN BUCKET PARAMETERS BY ON-LINE TRAFFIC MEASUREMENTS

Marek Dąbrowski *Wojciech Burakowski*
Institute of Telecommunications, Warsaw University of Technology,
ul. Nowowiejska 15/19, 00-665 Warsaw, Poland, e-mail:
{mdabrow5, wojtek}@tele.pw.edu.pl

Abstract
The paper proposes a method for more precise evaluation of traffic descriptors than a user of QoS oriented packet network is usually able to declare during the pre-connection phase. The recognized approach for traffic characterization is the token bucket mechanism, described by two parameters: token accumulating rate and bucket size. Notice that the values of these parameters constitute a base for evaluation of the effective bandwidth (or an equivalent factor) and, as a consequence, determine the amount of resources the network should dedicate for handling the connection. The typical approach is such that the token bucket parameters are simply declared by the users during the pre-connection phase. However, a user usually does not know a priori the correct token bucket parameters and has a tendency rather for over-dimensioning them. As a consequence, the calculated effective bandwidth is greater (sometimes significantly) than it is really required. In order to mitigate this problem, in this paper we propose a method based on the on-line monitoring of traffic, which allows us to obtain the minimum effective bandwidth.
The method is investigated for the case the users submit the traffic to the network service aimed for handling non-real time variable bit rate traffic without any losses. In this case, the effective bandwidth value could be calculated e.g. by EMW [3] formula. Apart the theoretical considerations, the numerical results illustrating the effectiveness of the approach are also included.

Keywords
Traffic characterization, token bucket, traffic measurements

1. INTRODUCTION

The problem of appropriate characterization of traffic emitted by the users and its policing is the key issue in packet oriented networks supporting QoS (Quality of Service) such as ATM and, recently, IP. The problem becomes more critical in the case of variable bit rate traffic. Despite a lot of efforts for taking into account statistical behaviour of such a traffic, the practical solution for traffic description is, at least for now, to characterize the traffic by the token bucket parameters [6]. The token bucket is the mechanism described by two parameters that are the token accumulating rate (r^T) and bucket size (b^T). Behaviour of the token bucket can be modelled by a queue with the service rate r^T and waiting room b^T. We say that the traffic submitted to such a queue is properly characterized when it is served without losses. However, it is evident that infinite number of pairs (r^T,b^T) satisfy the above condition. Note that the (r^T,b^T) values are "transformed" to the amount of resources the network should allocate for handling the traffic with specified QoS objectives, and that is usually expressed by the effective bandwidth value. Therefore, it is reasonable to search for such (r^T,b^T) that gives the minimum effective bandwidth.

The typical scenario to manage a network service aimed for handling variable bit rate traffic with predefined QoS objectives is based on user traffic declarations and employing the adequate admission control (AC) rules. The AC decides whether new call is admitted/rejected. This decision is based on the current carried traffic, the bandwidth dedicated to the considered service and the traffic declarations of new call. The typical approach, as depicted on Figure 1a assumes that estimation of the carried traffic takes into account the traffic user declarations only. However, the user that must submit the token bucket parameters to the network during its pre-connection phase usually has a tendency to overestimate the values since he wants to avoid potential traffic losses. In addition, the user has rather a little knowledge about the future produced traffic. Notice that the network police this traffic at the entry point. As a consequence of the token bucket over-dimensioning, the allocated network resources for the connection may be greater (even significantly) that are really needed.

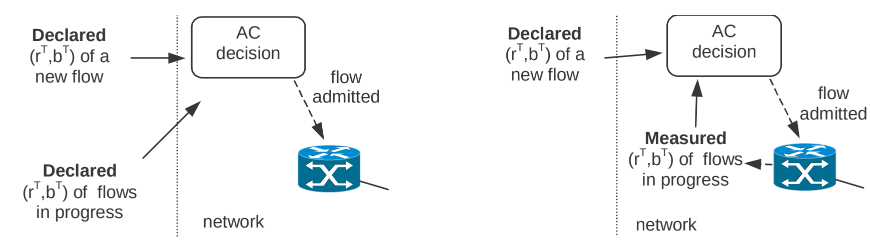


Figure 1. Admission control scheme a) declaration based and b) using token bucket measurements

In this paper we investigate the on-line traffic monitoring method for tuning the optimal values of token bucket parameters, as depicted in Figure 1b. In this way we mitigate the problem of potential network over-dimensioning due to not correct user traffic declarations.

The approach is checked for the case the user submits the traffic to the network service aimed for handling non-real time variable bit rate traffic without any losses. In this case, the effective bandwidth value could be calculated by the EMW [3] formula.

The problem of estimating token bucket parameters by monitoring the traffic was investigated in some papers. For instance, in [4], the measurement-based traffic specification is claimed as giving essential improvement against to a-priori user declarations. The presented algorithm is limited to the case when the token rate is fixed and the searching parameter is the bucket size value only. In [1], the useful notion of burstiness curve of the source was introduced, which represent the relation between the allocated bandwidth and buffer size. This notion was used in [5] for showing the relation between the burstiness curve and the dimensioning of leaky bucket parameters in the case of ATM connections. The proposed algorithm for recursive estimation of the leaky bucket parameters assumes, that the particular connections are scheduled on the output link by the generalized processor sharing (GPS) discipline and takes into account the bound for the acceptable delay. The estimation of the burstiness curves of the video sources was discussed in [2]. However, the proposed algorithm can be rather applied off-line, when the entire traffic is known (e.g. it is stored in the form of video frames trace).

The proposed in our paper mechanism is of on-line type and has the ability to derive the optimum point on the burstiness curve, i.e. the token bucket parameters, which allow us to obtain for the considered traffic the minimum effective bandwidth.

Organization of the paper is as follows. In section 2, we present the considered system in details. The admission control rules for the system are discussed in Section 3. Original algorithm for searching the optimum values of token bucket parameters by on-line traffic monitoring is described in section 4. The numerical results are provided in section 5. Finally, the conclusions are outlined.

2. THE CONSIDERED SYSTEM

In this paper we focus on a QoS network service supported by a packet oriented networks, like ATM or IP QoS, and aimed for handling non-real time variable bit rate traffic with zero losses as QoS objectives. Such a service can be regarded as potentially useful e.g. for handling traffic generated by non-real time video applications like Video on Demand. Let us consider that for this service the network allocates given amount of bandwidth, say C, and associated buffer size, say B. For instance, in the case of IP QoS network based on DiffServ [11] concept the admission control for this service is performed at the edge routers [see e.g. 8]. Note, that in such a network the adequate classification and scheduling mechanisms are implemented in each edge router and, thanks to them, the network resources are partitioned between the supported services.

The traffic control for considered service is based on the preventive approach and is as follows. A user, who wants to establish the connection, must submit to the network appropriate traffic declarations expressed in the form of the traffic descriptors that are the token bucket parameters. Furthermore, the admission control (AC) makes a decision whether this call may be admitted. Below we discuss the limitations of such approach mainly due to traffic characterization by the token bucket.

The token bucket is the mechanism for traffic characterisation/policing and is described by two parameters that are the token accumulating rate (r^T) and bucket size (b^T). It is commonly believed that the worst case of traffic profile which is conformant with the token bucket has the form of the deterministic ON/OFF type. The disadvantage of the traffic characterisation by the token bucket is that its dimensioning does not take into account the statistical nature of traffic. As a consequence, we should fix the same token bucket parameters for traffic with single and with many bursts. This fact has essential impact on effective bandwidth value which is calculated assuming the worst-case traffic pattern (ON/OFF type) instead of the really observed one. In addition, the worst case corresponds to the most critical burst that may occur only once in the considered traffic.

The deterministic ON-OFF model is parameterised by: P (peak rate), T_{on} (time in the ON state) and T_{OFF} (time in the OFF state). The relations between those and the token bucket parameters are the following:

$$T_{ON} = \frac{b^T}{P - r^T}\,,\; T_{OFF} = \frac{b^T}{r^T}\,. \tag{1}$$

Note, that the mean rate of the worst-case ON-OFF source is equal to the token rate, r^T, of the corresponding token bucket.

The dimensioning of the token bucket for the deterministic ON-OFF source is straightforward, but can be rather difficult in the case of realistic traffic sources. Therefore, we can expect that if the users have to specify a-priori the values of token bucket parameters, they will rather tend to over-declare them. Unfortunately, this can lead to too "safe" bandwidth allocations by the AC function than is really needed.

Additionally, a trouble in the dimensioning of the token bucket parameters is caused by the fact, that there is no single unique pair of (r^T,b^T) parameters, which tightly characterize given packet flow. "Tight" characterization means, that setting either r^T or b^T to a lower value could result in occurrence of non-conforming traffic in a monitored flow. In fact, a set of proper token bucket parameter pairs is infinite and can be represented as a function $b^T(r^T)$. As it was stated in [5], the traffic is regarded to be characterized tightly, when the chosen token bucket parameters are the points from the so-called "burstiness curve". This curve is defined as a function $b(e)$, where b and e are the maximum buffer occupancy and the service rate (allocated bandwidth), respectively. The burstiness curve [1] is a continuous, convex, decreasing function and is maximised when e=0, while is minimized (b=0) when $e>P$. Exemplary shape of the burstiness curve is depicted on Figure 2.

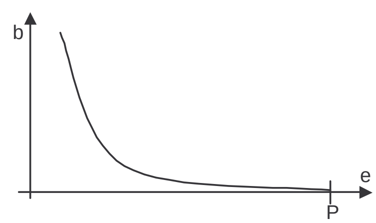


Figure 2. Exemplary shape of the burstiness curve

Taking into account that the proper token bucket parameters should fit the burstiness curve, we deduce that for each r^T, $b^T(r^T) = b(r^T)$. Note that, the resulting worst-case ON-OFF traffic pattern (see eq. (1)) depends on the chosen b^T and r^T values. Thus, the calculated effective bandwidth for the same traffic can differ even essentially. We believe that the algorithm proposed in this paper takes into account the above and allows us for finding the token bucket parameters, which provide the tight description. Additionally, the effective bandwidth is minimized.

3. EFFECTIVE BANDWIDTH CALCULATION

Suppose that bandwidth e_i is dedicated for a flow i (i=1,2,..,K; where K is the number of flows in progress). Sufficient condition to guarantee that buffer overflow never occurs is to allocate buffer space b_i, equal to the maximum buffer occupancy assuming the service rate e_i.

Taking into account the relations between the burstiness curve and the token bucket parameters (r^T,b^T), evident allocations are $e_i=r^T$ and $b_i=b^T$. However, usually a trade-off between the bandwidth (e_i) and buffer (b_i) allocations is possible. In the case of ON-OFF traffic pattern, the maximum buffer occupancy is reached at the end of the ON period. Thus, the required buffer space b_i can be calculated as:

$$b_i(e_i) = P \cdot T_{ON} - e_i \cdot T_{ON} = \frac{b^T \cdot (P - e_i)}{P - r^T} \tag{2}$$

Let us remark, that the bandwidth allocation less than r^T is not permitted since the r^T is the mean rate of the worst-case ON-OFF pattern. Therefore, additional condition is:

$$e_i \geq r^T \tag{3}$$

The function $b_i(e_i)$ is negative-linear with the maximum $b_i=b^T$ when $e_i=r^T$ and the minimum $b_i=0$ when $e_i=P$.

For the considered system, the effective bandwidth is calculated using the EMW method [3][7] which requires that the link and buffer allocations should be in proportion to the total link capacity C and the total buffer size B, i.e.

$$\frac{e_i}{b_i} = \frac{C}{B} \tag{4}$$

It can be proved, that if (4) holds then the following equivalence takes place: $\sum_i e_i = C \Leftrightarrow \sum_i b_i = B$. This simplifies the admission control rules since instead of using two criteria for accepting new flow, which are: $\sum_i e_i \le C$ *and* $\sum_i b_i \le B$, now the sufficient condition is only:

$$\sum_i e_i \le C \tag{6}$$

Substituting (2) to (4) and taking into account inequality (3), we obtain the final formula for effective bandwidth e_i:

$$e_i = \max\left(\frac{P}{1 + \frac{B}{C \cdot b^T}\left(P - r^T\right)}, r^T \right) \tag{7}$$

However, keep in mind that infinite set of (r^T, b^T) can tightly characterize the same traffic flow. Therefore, we aim to find such traffic description (r^T_{OPT}, b^T_{OPT}) that minimize the effective bandwidth calculated by (7). For showing where the minimum occurs we consider three cases of token bucket parameter settings, corresponding to 3 exemplary points on the burstiness curve, (r^T_i, b^T_i), i=1,…,3, as shown in Figure 3. Corresponding link and buffer allocations (e_i, b_i), i=1,…,3, are also depicted on the same figure.

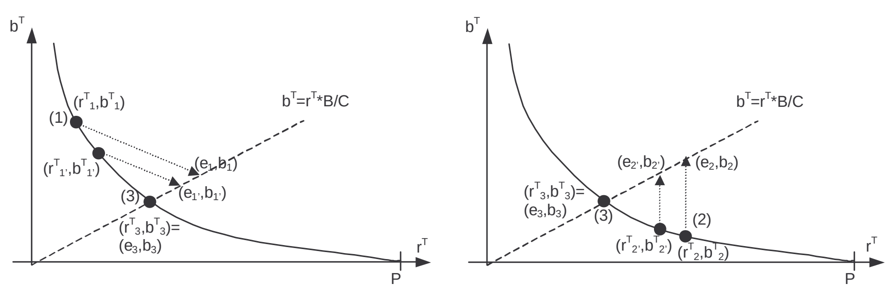


Figure 3. Illustration of the existence of optimal token bucket parameters

(1) Token bucket parameters are (r^T_1,b^T_1), such that $\frac{r_1^T}{b_1^T} < \frac{C}{B}$. Corresponding bandwidth and buffer allocations are denoted by (e_1,b_1). By formulas (2) and (4), the point (e_1,b_1) must be located on the intersection of the line crossing the points (r^T_1,b^T_1) and $(P,0)$, with the line $b^T/r^T=B/C$. It is evident, that one can find some points $(r^T_{1'},b^T_{1'})$, for which the corresponding allocation of bandwidth $e_{1'}<e_1$. Thus, we conclude that (r^T_1,b^T_1) is not the optimum.

(2) Token bucket parameters are (r^T_2,b^T_2), such that $\frac{r_2^T}{b_2^T} > \frac{C}{B}$. Now, applying the formulas (2) and (4) would lead to the allocation of bandwidth $e_2<r^T_2$. However, such allocation does not satisfy the condition (3) and we have to fix $e_2=r^T_2$ and over-allocate the buffer space, $b_2>b^T_2$. Once again, there are some points $(r^T_{2'},b^T_{2'})$, for which $e_{2'}<e_2$.

(3) Token bucket parameters are (r^T_3,b^T_3), such that $\frac{r_3^T}{b_3^T} = \frac{C}{B}$. Now, the condition (4) is satisfied and we can choose $e_3=r^T_3$ and $b_3=b^T_3$. There is no such point $(r^T_{3'},b^T_{3'})$, for which the corresponding allocation $e_{3'}$ is smaller than e_3. We conclude, that in this case the optimal traffic characterization (r^T_{OPT},b^T_{OPT}) satisfies:

$$\frac{r_{OPT}^T}{b_{OPT}^T} = \frac{C}{B} \tag{8}$$

Geometric interpretation of the condition (8) is illustrated in Figure 3 where it is shown that the optimal token bucket parameters are located on the intersection of the burstiness curve with the line $r^T/b^T=C/B$.

4. ON-LINE ALGORITHM FOR FINDING OPTIMAL TOKEN BUCKET PARAMETERS

In this section we present the algorithm for estimation of the optimal token bucket parameters by on-line traffic monitoring. This algorithm should be implemented as additional functionality e.g. of the ingress edge router in IP QoS network following the DiffServ concept.

4.1. Searching for the bucket size assuming that the token rate is fixed

First of all, we recall the algorithm for finding the minimum required bucket size, if the token rate is fixed. The method described in [4] is based on the operation of the modified leaky bucket mechanism, with the leak rate r^T and unlimited bucket size. It was shown, that the required bucket size in the case the token bucket mechanism with rate r^T is equal to the maximum occupancy of the leaky bucket mechanism with the same rate. For each arriving packet the algorithm calculates the leaky bucket occupancy, say d, and the new value for the token bucket size b^T, as:

$$d = max(d + (L - r^T * \tau)\ ,\ L) \qquad (9)$$
$$b^T = max(b^T\ ,\ d)$$

, where L is the packet length and τ is the time elapsed since the previous packet arrival. Notice, that the algorithm produces the exact solution for the minimum required bucket size for traffic submitted up to the considered time moment.

4.2. Approximated shape of the function $r^T(b^T)$.

The shape of the function $r^T(b^T)$ can be approximated by the linear segments, assuming that a number of points of this characteristic is known. For illustration of this approach let us consider, for instance, the approximation of the function from Figure 4. In this case, the function is approximated by three linear segments, crossing the points Q_i, i=1,…,4. Notice, that the point O', which is located on the intersection of one of these linear segments with the line $b^T/r^T=B/C$, is close to the point O, which is located on the intersection of the burstiness curve with the above mentioned line.

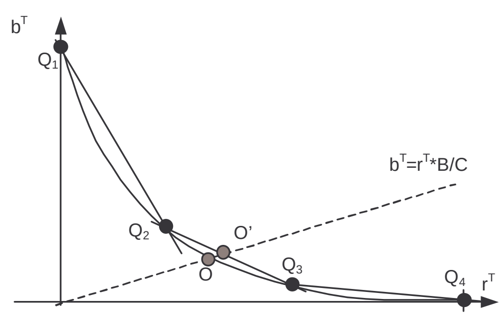


Figure 4. Example of approximation of the burstiness curve by the linear segments. O – optimal point, O' – approximated optimal point

The base of the algorithm is the modified leaky bucket mechanism, as described in section 4.1. Notice, that this algorithm is designed for finding exactly one point on the burstiness curve, when one of the coordinates (namely the r^T) is fixed. Thus, we need to choose a number of points (say n) and set their r-coordinates. Let us denote the values of token rates for the selected points as x_i, i=1,…,n. Then, we need to monitor the flow with n modified leaky buckets, to obtain the corresponding values of bucket sizes, which will be denoted as y_i,

i=1,..,n. Thus, on each packet arrival, the formulas from (9) are executed n times, independently for each of n modified leaky buckets.

Let as assume that the r-coordinates (x_i) are evenly spaced over the interval *(0, P)*:

$$x_i = \frac{(i-1)\cdot P}{n-1}, \quad i = 1,...,n \tag{10}$$

If the proposed algorithm is applied on-line to the monitored traffic, in each time moment it gives us the information about the coordinates of n points from the burstiness curve, $Q_i=(x_i, y_i)$, i=1,…,n,. By connecting the points Q_i with the linear segments, we obtain the demanded approximation of the burstiness curve. Notice, that we know a priori the coordinates of the n-th point that are *(P,0)*.

It is obvious that by selecting larger number of points we can improve the accuracy of the approximation. Unfortunately, this requires more leaky bucket instances. Therefore, the number of points should be carefully chosen considering the trade-off between the accuracy and complexity.

4.3. Finding optimal traffic characterization

In this section we focus on a method for deriving the coordinates of the optimal point approximation, say $O'=(r^T_{opt}{'},b^T_{opt}{'})$, as above defined. Note, that the algorithm presented in section 4.2 allows us for finding the n points, Q_i, i=1,…,n, with the coordinates (x_i,y_i). Thus, the burstiness curve is approximated by *(n-1)* linear segments. Each of these lines should cross the points $Q_i(x_i,y_i)$ and $Q_{i+1}(x_{i+1},y_{i+1})$ and is described by:

$$y = K_i \cdot x + L_i \tag{11}$$

, where:

$$\begin{cases} K_i = \dfrac{y_i - y_{i+1}}{x_i - x_{i+1}} \\ L_i = y_{i+1} - x_{i+1} \cdot K_i \end{cases} \tag{12}$$

The searched point $O'=(r^T_{opt}{'},b^T_{opt}{'})$ is located on the intersection of the line $b^T/r^T=B/C$ with one of the *(n-1)* lines defined by (11). The coordinates of the intersection points, say (x^r_i,y^r_i), i=1,...n-1, are calculated as follows:

$$\begin{cases} x^r_i = \dfrac{L_i}{-K_i + B/C} \\ y^r_i = \dfrac{B}{C} \cdot x^r_i \end{cases} \tag{13}$$

Due to the concave shape of the burstiness curve, we are looking for the point *O'* among the *(n-1)* points defined by (13), which has the greatest values of both coordinates, so:

$$r_{opt}^{T}{}' = \max\left(x_i^r\right) \qquad b_{opt}^{T}{}' = \max\left(y_i^r\right) \tag{14}$$

5. SIMULATION EXPERIMENTS

The effectiveness of the investigated in this paper algorithm was illustrated by the packet-level simulation experiments. The simulated traffic was generated using the trace files. Two types of traffic were taken into account. One type was represented by the MPEG4 video traces from [9]. These traces, which originally contain the record of the video frames emitted periodically, for the purpose of the experiments were converted into a packet stream form. For the second type of traffic the LAN traffic trace from [10] was employed. Table 1 presents the mean and peak bit rate values of the used traces.

Table 1. Trace parameters and link and buffer capacities assumed in the simulations

Trace file	Mean rate [Mbps]	Peak rate [Mbps]	Duration	*C* [Mbps]	*B* [MB]
Star Wars med	0.08	0.94	1 h	10	0.1
Jurrasic med	0.27	1.7	1 h	10	0.1
ARDNews	0.72	3.4	15 min	10	0.1
Star Wars high	0.28	1.9	1 h	10	0.1
Soccer	1.1	3.6	1 h	10	0.1
Formula med	0.29	1.4	0.5 h	10	0.1
Jurrasic high	0.77	3.3	1 h	10	0.1
Silence high	0.58	4.4	1 h	10	0.1
Bellcore Oct89Ext4	0.026	10	1 million pkts	100	1
Bellcore Oct89Ext	0.026	10	1 million pkts	100	1

In the first experiment, we have obtained the shapes of the burstiness curves of selected traffic patterns, i.e. from "*Star Wars medium quality*" and "*Bellcore Oct89Ext*". For this purpose, we were supported by the algorithm from section 4.1. The simulation was repeated assuming different values of the token rate r^T. The results are presented on Figure 5. The two upper graphs depict the experimentally obtained shapes of the burstiness curves of a) the video and b) of the LAN trace. Notice that as it was expected the obtained burstiness curve is a convex, decreasing function of r^T. The two lower graphs show the effective bandwidth values (calculated by (7)) corresponding to each of the points from the obtained burstiness curve. The assumed values of the link and buffer capacities (*C* and *B*), are collected in Table 1. Note, that dimensioning the token bucket parameters in non-optimal way, although still properly from the point of view of traffic characterization, can significantly (in the worst case, about 2 times for video and even 10 times for LAN) overestimate the required effective bandwidth.

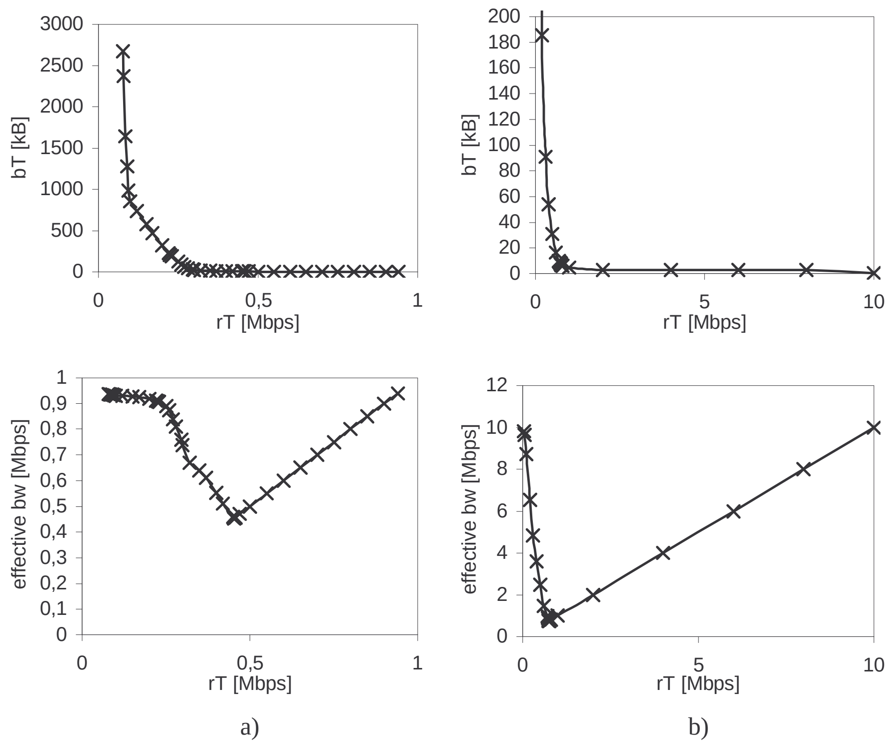
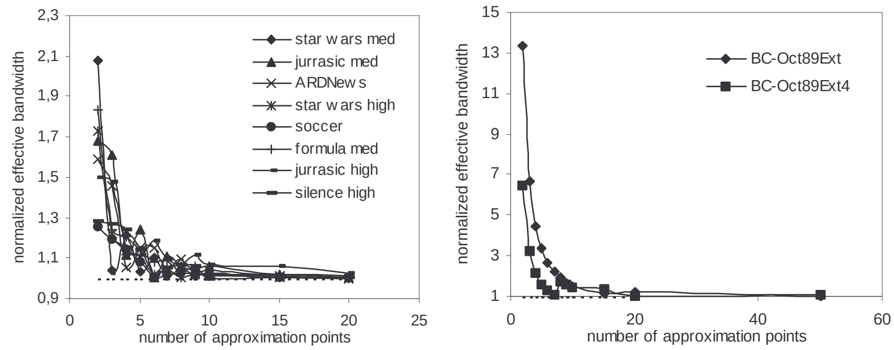


Figure 5. The burstiness curves (upper graph) and the corresponding effective bandwidth (lower graph) for a) video traffic trace and b) LAN traffic trace

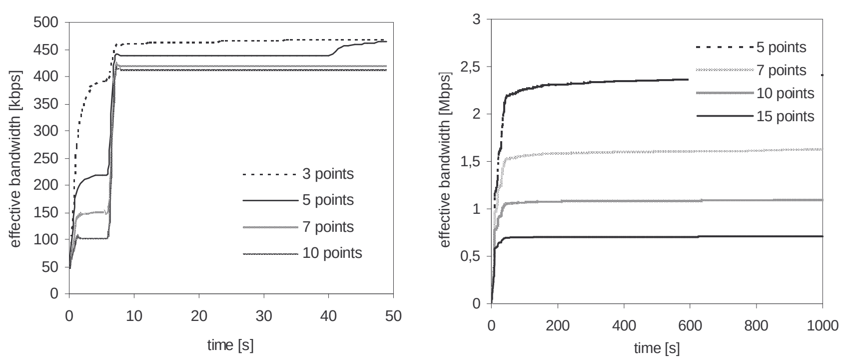
For showing the effectiveness of the algorithm proposed in the paper (section 4.2 and 4.3), its operation was implemented into the simulation program. The ability of the algorithm for getting close to the optimum depends on the assumed number of approximation points *n*. Therefore, the simulation experiments were performed for all traces from Table 1, and for different number of approximation points. Figure 6 shows the normalised effective bandwidth characteristic vs. number of approximation points. The effective bandwidth is normalized in relation to the lowest value (the closest to the optimum) obtained for a given traffic during all the simulation experiments. It was observed that about 5 approximation points is usually sufficient for finding the token bucket parameters close to the optimum.



a) b)

Figure 6. The normalized obtained effective bandwidth of different a) movie traces, and b) LAN traces, vs. number of points in the approximation of the burstiness curve

Finally, Figure 7 presents dynamic behaviour of the discussed algorithm. The basic requirement for application of the algorithm in the on-line regime is such that the time needed for assessment of effective bandwidth should be short. The convergence speed depends on the dynamics of the particular traffic, especially from the point of view of worst-case traffic occurrence. Let us recall that the token bucket is dimensioned for the worst case traffic. The results from Figure 7 were obtained during the simulation of selected video and LAN traces. One can observe, that the measured effective bandwidth is stabilized in really short interval after the connection starts. For example, in the case of the "Star Wars" video trace, the searched optimal token bucket parameters have been found within the first minute of the movie.



a) b)

Figure 7. Danamic changes of the evaluated effective bandwidth during the connection. Traffic was generated according to a) "Star Wars" trace and b) BC-Oct98Ext trace

6. SUMMARY

In this paper, we presented the novel algorithm for finding the token bucket parameters by on-line traffic monitoring. From the infinite set of token bucket parameter pairs (r^T,b^T), which properly and tightly characterize the monitored traffic, the algorithm allows to find such pair (r^T_{OPT},b^T_{OPT}), which results in the minimum value of the effective bandwidth. The method was investigated for the case when the users submit the traffic to the network service aimed for handling non-real time variable bit rate traffic without any losses. In this case, the effective bandwidth value is calculated by the EMW [3] formula. The effectiveness of the proposed method was verified by the simulations using different types of traffic traces.

REFERENCES

[1] S.Low, P.Varaiya, "A simple theory of traffic and resource allocation in ATM", In proceedings of IEEE GLOBECOM '91, volume 3, December 1991

[2] C.Tryfonas, A.Varma, S.Varma, "Efficient algorithms for computation of the burstiness curve of video sources", 16th ITC, Edinburgh, 1999

[3] A.Elwalid, D.Mitra, R.H.Wentworth, "A new approach for allocating buffers and bandwidth to heterogenous, regulated traffic in an ATM node", IEEE Journal on Selected Areas in Communications, Vol. 13, No. 6, August 1995

[4] P.P.Tang, T.Y.C.Tai, "Network traffic characterization using token bucket model", In Proceedings of IEEE INFOCOM'99

[5] S.Valaee, "A recursive estimator of worst-case burstiness", IEEE/ACM Transactions on Networking, Vol. 9, No. 2, April 2001

[6] J.Roberts, U.Mocci, J.Virtamo (ed.), "Final Report COST 242, Broadband network teletraffic: Performance evaluation and design of broadband multiservice networks", LNCS 1155, Springer, 1995

[7] P.Tran-Gia, N.Vicari (ed.), "Final Report COST 257, Impact of new services on the architecture and performance of broadband networks", compuTEAM, Wuerzburg, 2000

[8] C.Brandauer, W.Burakowski, M.Dabrowski, B.Koch, H.Tarasiuk, "AC algorithms in AQUILA QoS IP network", in proceedings of 2nd Polish-

German Teletraffic Symposium, Gdańsk, 2003. To appear in European Transactions in Telecommunications

[9] F.H.P.Fitzek, M.Reisslein, "MPEG-4 and H.263 video traces for network performance evaluation", Technical report TKN-00-06, Technical University Berlin, October 2000

[10] W.E.Leland, D.V.Wilson, "High time-resolution measurement and analysis of LAN traffic: Implications for LAN interconnection", Proceedings IEEE INFOCOM'92, pp.1360-1366, April 1991

[11] S.Blake et al., "An architecture for Differentiated Services", RFC2475, IETF, December 1998