PERFORMANCE ANALYSIS OF CALL ADMISSION CONTROL FOR STREAMING TRAFFIC WITH ACTIVITY DETECTION FUNCTION

Kiril Kassev, Yakim Mihov, Boris Tsankov

Abstract: *Admission control is a key issue for quality of service (QoS) provisioning in both wired and wireless communication networks. The call admission control (CAC) algorithm needs to know the source traffic characteristics and the required performance in order to determine whether the connection can be accepted or not and , if accepted, the amount of network resources to allocate. In this paper, we determine the CAC threshold value in case streaming homogeneous ON-OFF traffic flow is considered. An analytical method for packet loss probability evaluation is proposed and numerical examples are presented.*

Keywords: *streaming traffic, ON-OFF traffic model, call admission control, packet loss probability*

ACM Classification Keywords: *C.2.1 Network architecture and Design – Wireless communication, C.2.5 Local and Wide-Area Networks – Access schemes*

Introduction

The necessity of call admission control (CAC) arises together with the wide deployment of connection-oriented packet switching technologies. CAC is the name for a set of tools which has to take a decision whether or not a new connection can be served by the system, in addition to those of the connections that are in progress. If the new connection is admitted it will not deteriorate the bandwidth usage and performance of the connections already established. CAC is a fundamental mechanism for congestion control and QoS provisioning. For this reason, it has been extensively studied in both wired [1] and wireless [2] networks.

The CAC design and performance analysis became an inseparable part of ATM and IP based networks planning [3] and different wireless networks as well. CAC mechanisms can be classified based on various objectives and design options. Among the aims one can list: QoS parameters like call level and packet level congestion probabilities, packet delay, bandwidth guarantee; Optimization of throughput, power allocation, fairness; Controlling handover failure probability, etc. In this paper, the *call blocking* and the *packet dropping* probabilities are considered.

According to the decision time, CAC schemes can be classified as proactive (parameter-based) and reactive (measurements-based). In the former scheme, the arriving call is permitted or rejected based on predictive analytical evaluation of the QoS constraints. In the latter scheme, the CAC decides to permit or reject the call dynamically based on some QoS measurements. Both approaches have advantages and disadvantages. A combination of these two approaches could be used for more effective congestion control to be provided. In this paper, the *proactive CAC* is considered.

The subject of our interest is the traffic flow generated by multiple *variable bit rate* sources and the *bursty traffic* in particular where each source is represented as an ON-OFF source. Our considerations are restricted to *streaming (real-time) traffic* generated by VoIP sources and other multimedia sources as well. There are two consequences of this:

a) It is not possible to compensate the *burst-scale losses* by means of a buffer due to the very stringent restrictions on packet delay and the long average burst duration T_{ON} ;

b) It is reasonable to apply the so called *bufferless fluid flow* or *burst-scale loss approach* ([3], Chapter 12). The buffer (or buffers) used is relatively small and it is dimensioned to cope with the packet-scale losses only.

In this paper the bufferless fluid flow approach is used for CAC parameters determination. The assumption that there is not a buffer at burst level leads to a conservative estimates for packet losses and therefore to the safety side of CAC parameter determination.

The Traffic Model

The bufferless fluid flow model is used quite a while ago [4] – [7] due to its effectiveness and simplicity. Our considerations are restricted to homogeneous traffic sources with the most popular example – the VoIP traffic. Due to implementation of a voice activity detection algorithm into a voice codec of a particular type (i.e. G.729), an *ON-OFF traffic source* is usually characterized by means of the following three parameters: the bit rate during a burst (burst rate) *R*; the mean bit rate *r* and the burst duration T_{ON} . From the obvious relation

$$r = \frac{T_{ON}}{T_{ON} + T_{OFF}} R \tag{1}$$

one can obtain T_{OFF} . In the same time it is evident that an ON-OFF source could be characterized by means of any three out of the four parameters in (1).

Let us suppose *C* represents the total bit rate allocated to streaming traffic transmission, for example UL or DL radio-link capacity of an IEEE WiMAX or 3GPP LTE (Long Term Evolution) base station. We also define the following notations:

- *n* is the maximum number of active traffic sources that can be simultaneous served (or number of transmission resource units);
- *N* is the maximum number of calls (sessions) admitted to the system by the CAC.

The value of *n* is expressed as

$$n = \frac{C}{R} \tag{2}$$

Our aim is to determine *N* as a quantity of CAC. Obviously, there is no sense the value of *N* to be less than *n*. If *N* = *n*, all admitted calls will be served without any packet losses. In order to utilize the ON-OFF traffic behavior due to activity detection function, *N* has to be more than *n*. But if *N* > *C/r*, and the number of admitted calls approaches *N*, the buffer will permanently overflow. Therefore, apparently there is *n* < *N* < *C/r*. The exact value of *N* is the maximum one for which the probability P_{PL} , given packet to be lost is below a prescribed limit.

The system can be in any state (*i, j*), where *i* (*i=0, 1, …, N*) is the number of accepted calls and *j* (*j=0, 1, …, i*) is the number of active calls (number of bursts in progress). The call flow forms a Poisson process with call rate Λ_c and call service time $1/\mu_c$, whereas the burst flow forms a Binomial process with single OFF source burst arrival rate $\lambda_b = 1/T_{OFF}$ and single ON source burst service rate $\mu_b = 1/T_{ON}$. The combination of both processes forms the state-transition diagram shown on Fig. 1.

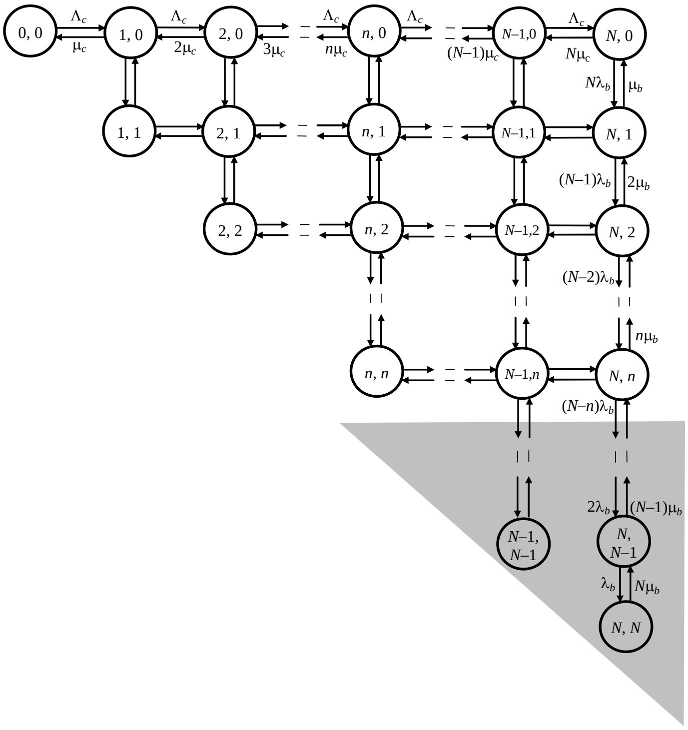

Fig. 1 Two-dimensional state-transition diagram

In the majority of cases when we determine *N* there is a common practice of considering the case *i* = *N* only and evaluate P_{PL} ([3], p. 141). This corresponds to the case where all traffic sources will suffer highest losses and it is related to the most right column on state-transition diagram (Fig. 1). Since we are interested in packet loss probability evaluation, applied for a more realistic case, it is of significant importance to take into account all possible states and perform an analytical evaluation of P_{PL} .

Analytical Evaluation

The state-transition diagram on Fig. 1 presents a two-dimensional *Markov process* where the burst level offered traffic depends on the established call number *i*.

According to the Erlang-B formula, the probability of exactly *i* sources being busy is

$$P(i) = \frac{\frac{A_c^i}{i!}}{\sum_{x=0}^{N} \frac{A_c^x}{x!}} \tag{3}$$

where $A_c = \Lambda_c / \mu_c$.

The conditional probability of *j* sources being active given that *i* traffic sources are busy is

$$P(j/i) = \frac{i!}{j!(i-j)!} \alpha^j (1-\alpha)^{i-j} \tag{4}$$

where $$\alpha = \frac{T_{ON}}{T_{ON} + T_{OFF}} \tag{5}$$

The overall probability $P(i, j)$ is

$$P(i, j) = P(i).P(j/i). \tag{6}$$

The offered rate in state (i, j) is $j.R$. The excess rate in the same state (i, j) is $(j - n).R$. As a consequence, the excess rate mean value is given by

$$\sum_{i=\lceil n \rceil}^{N} \sum_{j=\lceil n \rceil}^{i} R(j-n)P(i, j) \tag{7}$$

We should note that value of n is not necessary to be an integer, and $\lceil n \rceil$ denotes the minimum integer value greater or equal to n.

The packet loss probability is given by the relation

$$P_{PL} = \sum_{i=\lceil n \rceil}^{N} \sum_{j=\lceil n \rceil}^{i} R(j-n)P(i, j) / \sum_{i=1}^{N} \sum_{j=1}^{i} RjP(i, j) \tag{8}$$

The equation (8) can be simplified, and hence

$$P_{PL} = \sum_{i=\lceil n \rceil}^{N} \sum_{j=\lceil n \rceil}^{i} (j-n)P(i, j) / \sum_{i=1}^{N} P(i) \sum_{j=1}^{i} jP(j/i)$$

$$= \sum_{i=\lceil n \rceil}^{N} \sum_{j=\lceil n \rceil}^{i} (j-n)P(i, j) / \sum_{i=1}^{N} P(i) i\alpha \tag{9}$$

In order to perform a comparative analysis, considering the case where $i = N$ ([3], p.141) P_{PL} could be derived from (8)

$$P_{PL} = \sum_{j=\lceil n \rceil}^{N} (j-n)P(j/N) / N\alpha \tag{10}$$

Numerical Results

The current section deals with performance evaluation of analytical model proposed. In order to decrease the bandwidth usage the encoding scheme of each traffic source employs an activity detection function, which is quantitative represented by the activity factor α. Thus, the offered traffic flow A_c is generated by multiple homogeneous ON-OFF sources. Due to the limited amount of system resources available, the maximum number of calls (sessions) admitted to the system depends on the target call (session) blocking probability B, which can be obtained by (3).

Based on the required performance thresholds, such as B and P_{PL}, as well as source traffic characteristics (i.e. A_c), the significant task of CAC is to determine whether the connection can be accepted or not and, if accepted, the amount of network resources to be allocated. Fig. 2 shows the comparison results of the network dimensioning with typical values of the packet losses P_{PL} and activity factor [8], by applying both the model presented in [3] (10) (we will refer to it as "model A") and proposed analytical model (9) (we will refer to it as "model B"). It should be noted that "model A" refers to a case when the system is heavy loaded ($i = N$). Since the network service providers are interested in the system performance evaluation under normal load condition, it is necessary all possible system states to be taken into consideration. Numerical results show that this led to more efficient network resource usage (bandwidth), since the system is not overdimensioned, as it is done by using (10). On the other hand, as expected, silence suppression considerably decreases the transmission resource usage needed to meet the target packet loss probability.

According to [8], the degradation of voice quality (subjective quality measure MOS) is tightly coupled with the packet losses and coding scheme used. In case of using G.729, the admission of 2 % packet losses reduces the MOS from 4.0 to 2.75. Fig. 3 shows network dimensioning for different values of P_{PL}. Study results demonstrate that the same amount of network resource could be allocated to meet call flow demands with higher value of activity factor (α = 0.6), compared to the case when α = 0.45, but there is a trade-off that users will experience poor voice quality (MOS is reduced to a value of 2.75 [8]). This could be applied in case of short-term resource reduction.

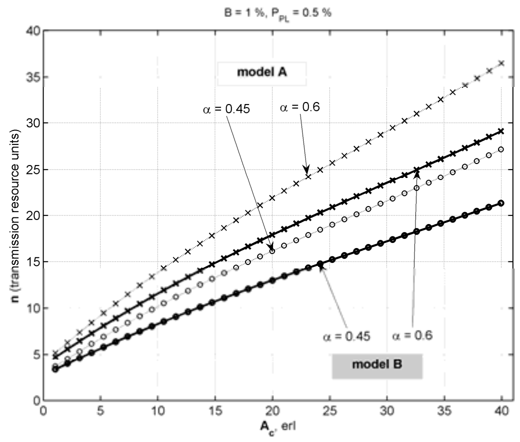

Fig. 2 Network dimensioning – CAC models comparison

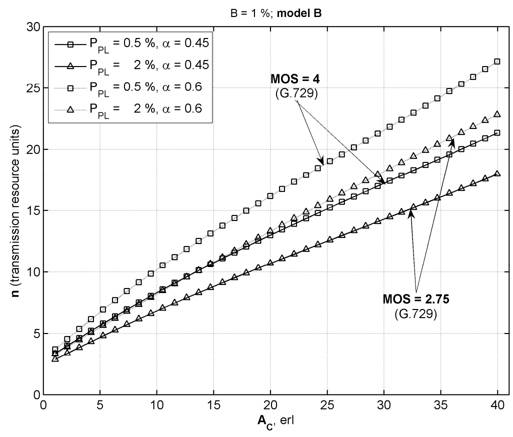

Fig. 3 Network dimensioning – model B

Conclusion

In this paper, an analytical method for quantitative analysis of a call admission control mechanism is proposed. The method takes into account the more realistic case study of bursty traffic arrival. A comparative analysis with a model with similar capabilities, suggested in the literature, has been performed. The results obtained demonstrate that analytical method proposed is efficient, especially when it is applied for wireless access networks dimension, since it does not overdimension the network in terms of necessary bandwidth (transmission resource units). This is of significant importance, since wireless communications resources are scarce and expensive. The method developed can be used in cross layer design of wireless networks, where considering the application requirement the more efficient resource allocation is achieved [9]. It is based on the Erlang model on the call level although an assumption of the Engset model [10] is also applicable with corresponding numerical complexity. The model could be extended by considering a heterogeneous traffic case.

Acknowledgement

This paper is a part of research work in the context of the research project "Optimal telecommunication resource allocation considering cross-layer interaction", funded by the Bulgarian Ministry of Education and Science with grant DVU01/0109, as well as a research project funded by the Research and Development Sector of the Technical University of Sofia with grant 080910dni-7.

Bibliography

[1] H. G. Perros and K. M. Elsayed, "Call admission control schemes: a review", *IEEE Communications Magazine*, vol. 34, no. 11, Nov. 1996, pp. 82-91.

[2] M. H. Ahmed, "Call admission control in wireless networks: A comprehensive survey", *IEEE Communications Surveys*, vol. 7, no. 1, 2005, pp. 50 – 69.

[3] J. M. Pitts and J. A. Schormans, *Introduction to IP and ATM Design and Performance* (Second edition), Chichester, England: John Wiley & Sons, 2000.

[4] J. Y. Hui, "Resource allocation for broadband networks", *IEEE J. on Selected Areas in Communications*, vol. 6, no. 9, Dec. 1988, pp. 1598-1608.

[5] R. J. Gibbens, F. P. Kelly and P. B. Key, "A decision-theoretic approach to call admission control in ATM networks", *IEEE J. on Selected Areas in Communications*, vol. 13, no. 6, Aug. 1995, pp. 1101-1114.

[6] M. Reisslein, K. W. Ross and Rajagopal, "Guaranteeng statistical QoS to regulated traffic: the single node case", *Proc. 18th, IEEE INFOCOM*, 1999, pp. 1061-1072.

[7] G. Mao and D. Habibi, "Loss performance analysis for heterogeneous ON-OFF sources with application to Connection Admission Control", *IEEE/ACM Trans. on Networking*, vol. 10, no. 1, Feb. 2002, pp. 125-138.

[8] A. P. Makropoulou, F. A. Tobagi, M. J. Karam, "Assessing the quality of voice communications over Internet backbone", *IEEE/ACM Transactions on Networking*, Oct 2003, vol. 3, no. 5, pp. 747-760.

[9] E. Pencheva, I. Atanasov and D. Marinska, "Cross Layer Design of Application-level Resource Management Interfaces", *Proc. of IEEE International Workshop on Cross Layer Design IWCLD'2009*, June 2009, Spain, pp. 1-5.

[10] V. G. Vassilakis, G. A. Kallos, I. D. Moscholios and M. D. Logothetis, "The wireless Engset multi-rate loss model for the call-level analysis of W-CDMA networks", *IEEE 18th Int. Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2007)*, 2007, pp. 1-5.

Authors' Information

Kiril Kassev – Assistant Professor, Department of Communication Networks, Technical University of Sofia, 8 Kliment Ohridski Blvd., Sofia-1000, Bulgaria; e-mail: kmk@tu-sofia.bg

Yakim Mihov – MSc Student, Department of Communication Networks, Technical University of Sofia, 8 Kliment Ohridski Blvd., Sofia-1000, Bulgaria; e-mail: yakim_mihov@abv.bg

Boris Tsankov – Professor, Department of Communication Networks, Technical University of Sofia, 8 Kliment Ohridski Blvd., Sofia-1000, Bulgaria; e-mail: bpt@tu-sofia.bg