Simplifying Blocking Probability Estimations for Heterogeneous Voice Sources in VoIP Environments

Reu-Ching Chen
Department of Information Management,
Nan-Kai College
No. 568, Jhongjheng Rd., Caotun Township,
Nantou, Country 542, Taiwan (R.O.C.)
E-mail: che1627@ms18.hinet.net

Abstract

A fast estimation of blocking probability for heterogeneous voice source in VoIP environments is proposed in this paper. In VoIP environments, traffics are generated from various voice sources and the transient system states due to the input traffics are commonly described by MMPP (Markov-Modulated Poisson Process).

MMPP is a generalization of Poisson process and is commonly used in modeling the input process of communication networks. However, as we know, the solution of MMPP model on the estimation of blocking probability is very involved when the total system state number N become large.

In this paper, a generic reduction method is proposed for blocking probability estimation. In our study, a large total number of system states for the original MMPP model is equivalently downsized to an approximating Markov chain model with less total number of system states. This will benefit on easy estimation of blocking probability. Our contribution is focused on system simplification and the blocking probability estimation is asymptotically closed to the original system. The numerical results shows our method is satisfied and this discipline can be widely applied to other high-speed networks for model simplifications.

Keywords: Two-state, voice, reduction, heterogeneous, equivalent model.

1. Introduction

In modern VoIP environments, voice is the vast majority of information exchanging over the telecommunication networks. Voice transmission can be performed either by circuit switching or packet switching.

Circuit switching has the fixed bandwidth and control policy, its blocking probability can be easily calculated by Erlang's formula. However much bandwidth is exhausted in circuit switching technique. In another aspect, less bandwidth is required in packet switching technique. However, the blocking probability estimation in packet switching is much complex than in circuit switching. Packet switching requires the system performance to fit the quality of service (QoS) requirements. Low latency and low blocking rate are the fundamental requirements of QoS. Especially for packets encoded from voice behaving the characteristics of burstiness and delay-sensitive, this results in the complexity of system analysis. Therefore, for QoS guarantee, how to capture the traffic

characteristics of voice is the main issue in VoIP environments. In this paper, we concentrate on blocking probability estimation for voice transmission in VoIP environments.

The MMPP models have been widespread applied in traffic analysis [4]-[8], however no simple methods have been provided for performance estimations. Classical iterative methods, such as the block Gauss-Seidel method, are used to solve the steady state probability. The convergence rate of this method is slow [12].

In our approach, due to the traffic generated by voice source, a generic method is provided to estimate the blocking probability by reducing the MMPP model. A fast estimation for blocking probability is achieved according to the equivalent reduced MMPP model. The dimension of the state space of the Markov chain expressed by the MMPP model is effectively reduced in our scheme. Numerical results presented therein focused on the accuracy of the approximation.

The rest of the paper is organized as follows. After given the model description in the next section, we derive mathematical analysis in Section 3 both for uniform and non-uni-form traffics. In Section 4, numerical and simulation results are presented and Section 5 is the conclusion.

2. Model description

Fig. 1 shows a traffic rate of voice source includes two states in a busy period, i.e., active state and idle state in a conversation cycle. The busy period is defined as the summation of one active state and one idle state. In Fig. 1, the packets are generated only when the voice source is stay in the busy period and no packets are generated when the voice source is stay in the idle period.

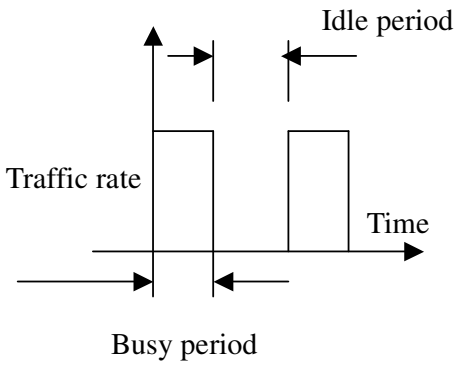


Fig. 1 Voice conversation cycle

In voice conversations, the voice source resides in active state will generate packets. The generating packet is fed to a server with finite queue buffer as depicted in Fig. 2. Where $n\mu$ denotes an n servers with each server owns a service rate equal to μ. Therefore, in our model, the link capacity of the queue is equal to $n\mu$. Since voice transmission is delay-sensitive, to avoid delay occurring caused by the queue buffer, we assume the buffer size is small enough and its effect can be ignored. Therefore, it is reasonable to adopt the M/M/n/n loss queuing model for analysis in our approach. For the M/M/n/n system, if a packet arrives when n servers are occupied, that packet is lost.

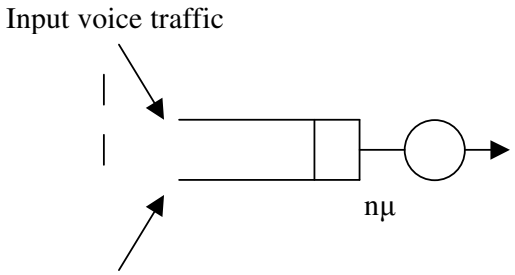


Fig. 2 M/M/n/n queuing model

3. Mathematical analysis

Fig. 3 shows the two-state MMPP Markov model corresponds to Fig. 2, where the total number of state for the original model equal to 2(n +1), where n is a positive integer indicating the total number of packets resident in the system. We assume each packet is served by one server with service rate μ, and nμ servers can serve at most n packets at one time. In our scheme, the 2(n+1) states are reduced to k+1 states, where k is an positive integer and its magnitude is much less than n. The advantage of the reducing method is based on the fact that the calculation of blocking probability is simplified since the dimension of the system state space is reduced. For convenience, we call each of the k+1 states as cluster state.

From Fig. 3, the system states are described by the two-dimensional Markov chain with metric state (x, y), where the first parameter x is a random variable indicating the system state with state space set {0,1}, in which, 0 indicates the idle state and 1 indicates the active state. The second parameter y is a random variable indicating the total number of customers in the system with set {0, 1, 2, … ,n }. For instance, state (1,9) indicates the system is in active state and the total number of customers equal to 9.

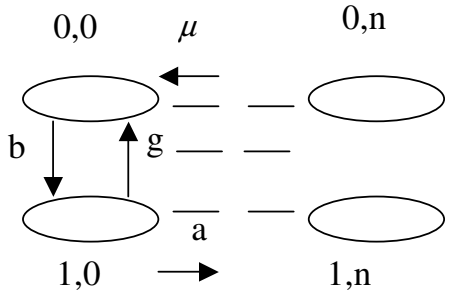


Fig. 3 Two-state Markov model

Recall that packets can only be generated during the time interval when the voice source is in active state. In our study, the original 2(n+1) states of MMPP Markov model shown in Fig. 3 (including both of active and idle state) is reduced to n+1 states as Fig. 4 depicted. It is noted the arrival rate a in Fig. 3 is replaced by r as Fig. 4 depicted with the following equation. i.e.,

$$r = \frac{b}{b+g} a \text{ ----------(1)}$$

Equation (1) can be validated from the two state Markov chain model as Fig. 5 shown under the assumption that the system is stationary and ergodic. Let P_{ON} be the steady state probability for the "ON state", similarly, let P_{OFF} be the steady state

probability for the "OFF state". Then applying the balance equation to Fig. 5 and the normalization constraint, we have

$$P_{ON}g = P_{OFF}b \quad \text{----------(2)}$$

and

$$P_{ON} + P_{OFF} = 1 \quad \text{----------(3)}$$

Solving for P_{ON} by conjunction of Eqs. (2) and (3), we obtain

$$P_{ON} = \frac{b}{g+b}.$$

Since the total arrival rate equal to a when the state is staying in ON state and equal 0 in OFF state. Therefore the equivalent arrival rate r in the reduced MMPP model (as shown in Fig. 4) is equal to the partial fraction of the arrival rate in the original MMPP model (as depicted in Fig. 3) in ON state.

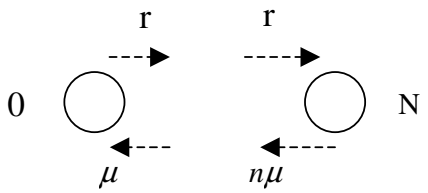


Fig. 4 Reduced MMPP model with n+1 state

An advanced simple model is obtained From Fig. 4 by partitioning the n+1 states into k+1 states by associations (we call k+1 clusters in the following sections) with k be much less than n.

It is noted that the total number of the states for the MMPP model and the reduced model are n+1 and k+1 respectively since the start number of the state is 0. In reducing a MMPP queueing model containing large number of n+1 states into a small number of k+1 states. For simplicity we consider the uniform distribution in the following.

In this case, the distributions of the arrival rate for the ON/OFF transition rates are assumed to be constant as shown in Fig. 4 and Fig. 5 for the original MMPP model and reduced equivalent model respectively. Then, from symmetry point of view, we can select the states numbered 0, m'th, 2m'th, 3m'th,…,etc. as the pilot states. For each pilot state, we make decision to decide either the m-1 states that located at its left hand side or right hand side should align to the pilot state to constitute a cluster or not.

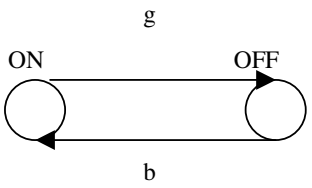


Fig. 5 ON/OFF state model

Without loss of generality, the n+1 states are divided into k+1 clusters where each cluster contains m states with the exception that the left end cluster (i.e., the first cluster) and the right end cluster (i.e., the last cluster). Hence, the total number of states in the left end and right end cluster depend on which one of the left-aligned or right-aligned structures are selected. Therefore we have the following relation between parameters n, m and k.

$$\left\lfloor \frac{n}{m} \right\rfloor = k \quad \text{----------(4)}$$

At here, n, m and k are all positive integers, where $\lfloor x \rfloor$ is the ceiling of the x. Hence, in our approach, the magnitude of n is very large, i.e., $m\left\lfloor \frac{n}{m} \right\rfloor \approx n$.

It is noted, different selected value of m corresponds to different number of system states to be solved. Large value of m corresponds to small number of state equations necessary to solve, however less accuracy in system performance estimations. Small value of m selected corresponds to more involved system states to be solved but will benefit on more accurate in system performance estimations. Generally, accuracy requirements and calculation complexity are tradeoff in system performance estimations. Therefore adequate selection for the value of k is crucial to network designers.

The blocking probability for the original MMPP model as shown in Fig. 4 is obtained from the well-known birth-death Markov chain [1], i.e.,

$$P_n = \frac{r^n}{\sum_{s=0}^{n} \frac{n!}{s!} \mu^{n-s} r^s} \quad \text{----------(5)}$$

similarly, the blocking probability for the reduced left-aligned Markov chain is

$$P_L = \frac{r^k}{\sum_{t=0}^{k} \frac{k!}{t!} r^t (m\mu)^{k-t}} \quad \text{----------(6)}$$

Similarly, the blocking probability of the right-aligned model is

$$P_R = \frac{r^k}{r^k + \sum_{q=0}^{k-1} \left\{ \prod_{h=1}^{k-q} [(k-h)m+1]\mu \right\} r^q} \quad \text{----------(7)}$$

Applying Eqs. (6)-(7), we have the following useful Lemma for blocking probability estimation.

Lemma: For an equivalent n+1 states Markov Modulated Poisson process model as shown in Fig. 4, the blocking probabilities are over-estimated both for the reduced left-aligned and right-aligned structures. For the reduced system containing k+1 states with $k > m$ and $m > 2$. We claim the left-aligned model is prior than the right-aligned model in the estimation of blocking probability, equivalently, the following relation is hold.

$$P_n < P_L < P_R \quad \text{----------(8)}$$

The reason for $m > 2$ is from the fact that the efficiency of association is zero for $m = 1$ and $k > m$ will effectively highlight the reducing efficiency.

<Proof>

The detail proof is depicted in appendix.

In the estimation of blocking probability, the relation (8) indicates the left-aligned method is prior than the right-aligned method. Eq. (6) is powerful in the calculations of blocking probability estimation when the total number of states of the MMPP model is very large since large value of n makes solution complicated

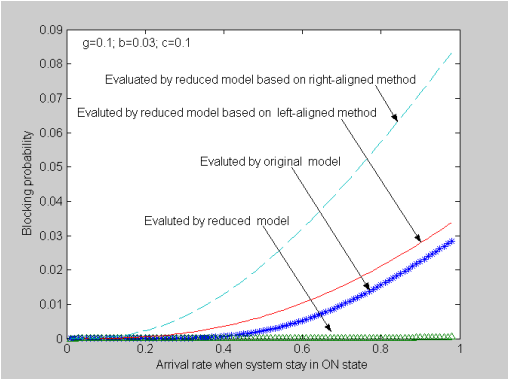
and unreachable.

Hence, in our study, Poisson arrival and exponential service rate are adopted for model analysis. By using similar method, other traffic distributions of input arrivals and different service distribution can also be taken into account in the reducing process for performance estimations (e.g., Parento's distribution in Ethernet). Owning to space limitation, other distributions for input stream and service rate are not presented in this paper.

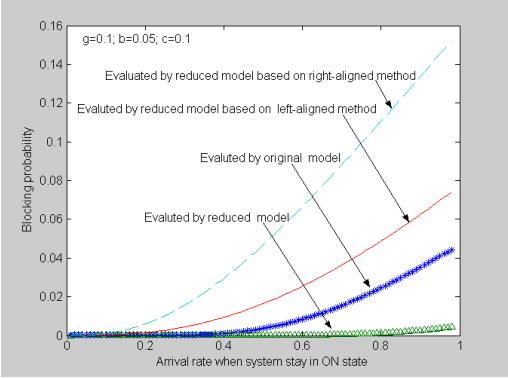
4. Numerical and simulation results

In this case, the arrival rate is constant for each state. Fig. 6(a) shows the blocking probability as functions of input arrival rate for service rate c equal to 0.1, the "ON state" transition probability rate b equal to 0.05 and the "OFF state" transition probability rate g equal to 0.1. Similarly, Fig. 6(b) shows for the same values of c and g except b = 0.03. It is noted, the blocking probability is under-estimated for the reduced model, and over-estimated for the right-aligned and left-aligned models. The blocking probability of the left-aligned model is more approached to the blocking probability of the original model than the right-aligned model as the Lemma depicted previously. Hence, it is clear that the blocking probability is an increasing function of traffic load as desired.

It is noted, the system with larger value of the "ON state" probability rate has higher blocking probability than less value of the "ON state" probability rate, i.e., the blocking probability of Fig. 6(a) is greater than Fig. 6(b). This is reasonable from Eq. (5) for larger value of arrival rate corresponds to larger value of blocking probability.



(a)



(b)

Fig. 6 Blocking probability comparisons for uniform condition with n=9, arrival rate=0.1

Our simple model has the benefit to fast estimate the blocking probability without need concerning all the system states. This will facilitate on the calculation cost in system performance estimation. Although more numbers of clusters we take, more accuracy we get, i.e., as the total number of cluster increases, the more accurate results we obtained. However, more clusters will

induce more state equations to be solved, this will be non-practical in economic consideration. The tradeoff between total number of selected clusters and calculation cost is reasonable. Although we focus on the fast estimation of the blocking probability, nevertheless the other parameters such as delay, throughput can also be calculated easily in the same method.

In our scheme, the pilot states have been selected first and then we make decision for which of the adjacent states (located at left and right hand side of the pilot state) are selected to associate with its pilot state. Our solution is based on the M/M/n/n queuing model in which the service rate and arrival rate are assumed to be uniform. E.g., arrival rate equal to a for each state and the service rate equal to u for state 1 and equal to iu for state i respectively.

5. Conclusions

An efficient asymptotic estimation for blocking probability in VoIP environments has been derived. In real time environments, the asymptotic estimates of blocking probability is compared with the results of the original model. The voice model describing bursty traffic belonging to MMPP are solved by reducing the numbers of system states into small number of clusters. In this paper, we illustrate the blocking probability of a complex queuing model can be easily estimated by simple model that is downsized from the original models. Only the uniform distribution is considered. The non-uniform condition can be treated in the same manner. Our results shows the left-aligned method is better than the right-aligned method for uniform condition.

Consequently, the developed method can be widely applicable in simplifying any traffic distributions (i.e., the Ethernet traffic, which is Parento's distributions).The results presented here are satisfied. Our methods are simple and can easily applied in performance estimations for high speed network.

Appendix

Proof of Lemma

To validate the relation $P_n < P_L < P_R$ in Eq. (4), We first prove $P_L < P_R$. By expending Eq. (6) and Eq. (7), we have

$$P_L = \frac{r^k}{km\mu \quad (k-1)m\mu \quad \ldots \quad m\mu + \ldots + r^{k-1}km\mu + r^k}$$

----------(9)

$$P_R = \frac{r^k}{[(k-1)m+1]\mu \quad [(k-2)m+1]\mu \quad \ldots\mu + \ldots + r^{k-1}[(k-1)m+1]\mu + r^k}$$

----------(10)

It is noted, Eqs. (5) and (6) are function of r. Then the coefficient of r^i in the denominator of Eq. (5) can be expressed as

$$\frac{k!}{i!}(mu)^{k-i} = km\mu \,.(k-1)m\mu \,.(k-2)m\mu)m\mu\ldots(i+1)m\mu$$

----------(11)

similarly, the coefficient of r^i in the denominator of Eq. (6) is

$$[(k-1)m+1]\mu \,.[(k-2)m+1]\mu \,.[(k-3)m+1].....[jm+1]\mu$$

----------(12)

Therefore, the corresponding p'th term in Eq. (7) and (8) is $(k-p+1)m\mu$ and

$[(k-p)m+1]\mu$ respectively.

Therefore

$(k-p+1)m\mu = km\mu - pm\mu + m\mu >$

$km\mu - pm\mu + \mu = [(k-p)m+1]\mu$

for m > 2 (as described in Lemma).

This results induce the value of the denominator of Eq. (5) is greater than the value of the denominator of Eq. (6), i.e., $P_L < P_R$.

Next we prove $P_n < P_L$ in the following.

Since we assume the system load is under the non-saturation condition, then arrival rate r is less than 1. The value of the numerator of Eq. (6) is less than the numerator of Eq. (7) from the fact that r < 1 and n > k. We only need to compare the value of the denominator of Eq. (6) and (7). The coefficient of r^i in the denominator of Eq. (6) is

$$\frac{n!}{i!}\mu^{n-i} \text{ ----------(13)}$$

for large value of n, we have

$$\frac{n!}{i!}\mu^{n-i} = \frac{(km)!}{i!}\mu^{km-i} = \frac{km(km-1)(km-2)...(km-k)(km-k-1)...1}{i!}\mu^{km-i}$$

----------(14)

Therefore, the requirement for $P_n < P_L$ is equivalent to guarantee Eq. (14) is greater than Eq. (11), i.e.,

$$\frac{km(km-1)(km-2)...(km-k)(km-k-1)...1}{i!}\mu^{km-i} > \frac{k!}{i!}\mu^{k-i}m^{k-i}$$

this implies

$$\frac{km(km-1)(km-2)...(km-k)...(k+1)}{1}\mu^{k(m-1)} > (m)^{k-i}$$

----------(15)

where m > 2 and k > m as defined in the Lemma.

Applying the fact that n is a large positive integer and n = km with k > m, then Eq. (15) is true when μ is greater than 0.09 for k = 10 and i = 3, where A corresponds to the left term of the Eq. (15) and B corresponds to the right term of Eq. (15). It is noted that large value of k corresponds to lower bound of value of μ in satisfying the requirement of Eq. (15).

Therefore

$P_n < P_L$ The proof of the Lemma is completed.

References

[1] Kleinrock, L. , Queueing System, Vol. 1: Theory, Wiley, New York 1975.

[2] Kathleen S. Meier-Hellstern, "The Analysis of a Queue Arising in Overflow Models," IEEE Trans. On Communications, Vol. 37, No. 4, April 1989.

[3] T. Yoshihara, S. Kasahara, and Y. Takahashi, "Practical Time-Scale Fitting of Self-Similar Traffic with Markov-Modulated Poisson Process," Telecommunication Systems 17, 185-211, 2001.

[4] Shoji Kasahara, "Internet Traffic Modeling: Markovian Approach to Self-Similar Traffic and Prediction of Loss Probability for Finite Queues," IEICE Trans. Commun. Vol. E84-B, No. 8, pp. 2134+, 2001.

[5] M. Leung, J. Lui, and D. Yau. “Adaptive proportional delay differentiated services: Characterization and performance evaluation,” IEEE/ACM Trans. on Networking, 9(6):801--817, 2001.

[6] Aimin Sang and San-qi Li, "A Predictability Analysis of Network Traffic," INFOCOM, 2000, pp. 342-351.

[7] C. Courcoubetis, A. Dimakis, and G. D. Stamoulis. “Traffic equivalence and substitution in a multiplexer,” IEEE INFOCOM'99, pp. 1239-1247, 1999.

[8] A. Borella, F. Chiaraluce and F. Meschini, “Statistical Multiplexing of Random Processes in Packet Switching Networks,” IEE Proc.-Commun, Vol. 143, No. 5, Oct. 1996.

[9] Victor Firoiu, Jean-Yves Le Boudec, Don Towsley and Zhi-Li Zhang, “Theories and models for internet quality of service,” Proceedings of the IEEE may 2002.

[10] M. Nomura, T. Fujii, N. Ohta, “Basic characteristics of variable rate video coding in ATM access environment,” IEEE J. Sec. Area. Comm. June 1989.

[11] Rachel Levi and Adam Shwartz, “Throughput-Delay Tradeoff with Impatient Arrivals,” Proceedings of the 23rd Allerton Conference on Communications, Control and Computing, Allertor, IL 1994.

[12] P. Davis, Circulant Matrices, John Wiley and Sons, New York 1979.

[13] G. Calinescu, A. Chakrabarti, H. Karloff, and Y. Rabani, ” Improved approximation algorithms for resource allocation,” In Proceedings of the 9th Integer Programming and Combinatorial Optimization Conference, 2002.