Error Analysis of Multiservice Single-Link System Studies Using Linear Approximation Model

C. Aswakul
Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University
Bangkok, 10330, Thailand
Email: chaodit.a@chula.ac.th

J. Barria
Department of Electrical and Electronic Engineering
Imperial College London
South Kensington Campus
London, SW7 2AZ, United Kingdom
Email: j.barria@ic.ac.uk

***Abstract*— This paper presents an error analysis of the linear approximation model in a multiservice single-link system with nonlinear equivalent capacity. Two types of error measures have been proposed, namely, the mean error and probabilistic error bounds. Derivation of these error measures reveals that the linear approximation model has two main error sources, i.e., the nonlinearity in equivalent capacity and the fluctuation of system dynamics at the mean operating point. In addition, given the product-form solution of the system with the complete sharing policy, a numerical procedure is derived to facilitate the calculation of proposed error measures. This procedure requires both the time and space complexity of $O(Ck)$, where C is the link capacity and k is the number of call types. Hence, these error measures can be efficiently computed in parallel to all the main system performance parameters (call blocking probabilities and mean revenue rates).**

I. Introduction

One of the major challenges in communication networking is to find the efficient mechanisms that can provide the quality of service (QoS) for a broad range of applications. To deal with the services with *greatly different* characteristics, various forms of *service separation* have been proposed (see, e.g., [1], [2], [3]). The underlying idea of service separation is that only ATM cell or IP packet streams from the same service are allowed to be statistically multiplexed. At a switch or router, a separate buffer is thus allocated to store cells or packets from *each* service. The QoS provision for each buffer (i.e., loss rate, delay and jitter) can then be facilitated by a weighted round robin or weighted fair queueing scheduler and an appropriate assignment of the scheduling weights.

For *dynamic service separation* (also called service separation with dynamic partitions in [1]), the scheduling weight for a given buffer i is made directly proportional to $G_i(n_i)$, which denotes the *equivalent capacity* (known as the *capacity function* in [1]) associated with the buffer i. This equivalent capacity $G_i(n_i)$ is the minimum amount of link capacity needed to achieve the QoS guarantee for buffer i when n_i connections are being served at the buffer. Therefore, to achieve the QoS guarantee for *all* the buffers $i \in \{1, \ldots, k\}$ that share the same link with capacity C, the number of ongoing connections n_i for all i must satisfy the capacity requirement constraint $\sum_{i=1}^{k} G_i(n_i) \le C$. Under service separation, it is worth noting that $G_i(\cdot)$ depends on n_i only, and not n_j $(j \neq i)$. Further, to reflect the economies of scale in statistically multiplexing cell or packet streams, it is known that $G_i(n_i)$ monotonically increases with decreasing slope as n_i increases [4], [5]. For instance, given the on-off source multiplexing model, the stationary Gaussian approximation [5] results in $G_i(n_i) = \alpha_i n_i + \beta_i \sqrt{n_i}$. Therefore, we generally have to cope with a *nonlinear equivalent capacity*.

In the past, given the dynamic service separation with nonlinear equivalent capacity, a few QoS mechanisms have been proposed. For call admission control (CAC), the complete sharing policy has been analyzed [6] in terms of call blocking probabilities of different services. This CAC analysis is then extended to the scenarios of trunk reservation policy [7] to investigate the fairness and prioritization in allocating the link capacity to different services. For network routing, a generalized dynamic alternative routing has been formulated and analyzed in terms of the lower/upper bounds for the mean network revenue rates [8].

The major concept used in solving all the formulated analytical models in [6]–[8] is called the *linear approximation model*, whose principle is to convert the analytical models from the nonlinear domain of equivalent capacity into an approximated linear domain. Consequently, the efficient numerical techniques in the linear domain can then be employed to reduce the involved computational complexity. Efficient numerical techniques are desirable because they need to be repeatedly invoked in network dimensioning procedures. However, the previous studies have only provided empirical investigation on the applicability of linear approximation model. Theoretical error analysis is still needed and becomes the main subject of this paper.

In this paper, we theoretically investigate the factors that

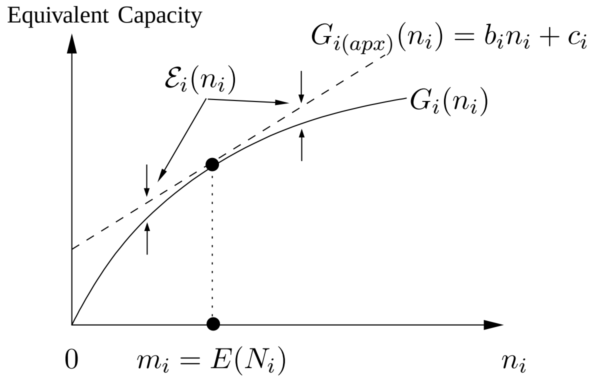

Fig. 1. Approximation of equivalent capacity and resultant error.

influence the linear approximation model's vulnerability to errors. Furthermore, we aim at deriving the error measures that can be efficiently computed. Like [6] and [7], the system considered in this paper is a single link with a finite capacity C accessed by k call types (each from a different service). Call arrivals of each service are characterized by an independent Poisson process and exponentially distributed holding times. The readers are referred to [6] and [7] for a complete expository of the system description and analysis.

The remainder of this paper is organized as follows. In Section II, a mean error ME_i and its approximation $ME_{i(apx)}$ are introduced. The derivation reveals the main factors that contribute to the error of linear approximation model. Section III then considers how to compute for the approximate mean error $ME_{i(apx)}$ recursively. The derivation in Section III is based on the product-form solution of complete sharing policy. Bounds of error probability are then obtained in Section IV. Numerical results are reported and discussed in Section V. Finally, Section VI summarizes all the main findings of this paper.

II. Mean Error of Linear Approximation Model

The essence of linear approximation model is to estimate a nonlinear equivalent capacity $G_i(n_i)$ of call type i $(i = 1, \ldots, k)$ by a linear function $G_{i(apx)}(n_i) = b_i n_i + c_i$, as depicted in Fig. 1. The parameters b_i, c_i are obtained from a tangential line of equivalent capacity at the mean operating point. Note here that $m_i = E(N_i)$ is mean of the number of type-i connections at time t, $N_i(t)$.

Therefore, the error of linear approximation model can be captured by

$$\mathcal{E}_i(n_i) \triangleq G_{i(apx)}(n_i) - G_i(n_i) \quad (1)$$

for $i = 1, \ldots, k$. This error measure $\mathcal{E}_i(n_i)$ is always greater than or equal to 0 since function $G_i(n_i)$ is monotone increasing and concave in practice [4], [5].

For the analytical purpose, suppose that $G_i(n_i)$ is twice-differentiable[1] on the considered range of n_i for all $i = 1, \ldots, k$. In this case, we may represent $G_{i(apx)}(n_i)$ by $G_i(m_i) + G_i'(m_i)(n_i - m_i)$. Based on Taylor's Theorem, if $G_i(n_i)$ is now approximated by the first three terms of the expansion about the mean operating point m_i, then it is possible to write down

$$G_i(n_i) \simeq G_i(m_i) + G_i'(m_i)(n_i - m_i) + \frac{G_i''(m_i)}{2}(n_i - m_i)^2, \quad (2)$$

which leads to

$$\mathcal{E}_i(n_i) \simeq \frac{-G_i''(m_i)}{2}(n_i - m_i)^2. \quad (3)$$

The value of n_i is driven by the stochastic process $\{N_i(t), t \geq 0\}$. Accordingly, the *mean error* of linear approximation model in approximating type-i equivalent capacity can be defined as

$$ME_i \triangleq E[\mathcal{E}_i(N_i)] \quad (4)$$

for $i = 1, \ldots, k$.

By combining (3) and (4), the approximate mean error, $ME_{i(apx)}$, can finally be written as $ME_{i(apx)} = -G_i''(m_i) Var(N_i)/2$. Despite of the minus sign here, it should be noted that $ME_{i(apx)} \geq 0$ since $G_i''(m_i) \leq 0$. Therefore, to prevent misunderstanding, we write

$$ME_{i(apx)} = \frac{|G_i''(m_i)|}{2} Var(N_i) \quad (5)$$

for $i = 1, \ldots, k$.

From (5), the error of linear approximation model depends on two factors. First, the second derivative term represents the nonlinearity in equivalent capacity at the mean operating point. Second, how much the system fluctuates around the mean operating point is captured by the variance term. Consequently, the linear approximation model may become erroneous if the considered system operates in the region where the equivalent capacity is highly nonlinear or if there is a large variation in the system dynamics or both.

III. Recursive Computation of Mean Error

In order to calculate $ME_{i(apx)}$ from (5), we need to know both mean $m_i = E(N_i)$ and variance $Var(N_i)$. The mean can be directly obtained in the linear approximation model by

[1] This assumption may be easily justified if an explicit functional form of equivalent capacity can be derived from an analytical model. On the other hand, if the equivalent capacity is obtained from measurements, experiments or computer simulations, then the values of equivalent capacity function $G_i(n_i)$ may be available at only integer values of n_i. In this case, the first and second derivatives of equivalent capacity function $G_i(n_i)$ are here calculated from $G_i'(a) = G_i(\lfloor a \rfloor + 1) - G_i(\lfloor a \rfloor)$ and $G_i''(a) = G_i(\lfloor a \rfloor + 2) - 2G_i(\lfloor a \rfloor + 1) + G_i(\lfloor a \rfloor)$, where $a > 0$ and $\lfloor a \rfloor$ is the largest integer which is not greater than a.

$m_i = \rho_i (1 - B_i)$, where ρ_i is the offered load of type-i call stream and B_i is the probability of blocking type-i calls. In this section, we focus on how to calculate $Var(N_i)$.

The important equation that makes the computation of $Var(N_i)$ possible is the *derivative formula* as derived in [9] from the product-form solution:

$$\frac{\partial B_i}{\partial \rho_i} = \frac{1}{\rho_i^2} \left(E(N_i) - Var(N_i) \right). \tag{6}$$

The derivative formula (6) relates $Var(N_i)$ to $\partial B_i / \partial \rho_i$. Therefore, in the following derivation, we focus on calculating $\partial B_i / \partial \rho_i$. In this respect, recall from the Kaufman-Roberts's recursive algorithm [10], [11] that B_i $(i = 1, \ldots, k)$ can be written as

$$B_i = \sum_{\{c \in \Phi | c + b_i \notin \Phi\}} q(c). \tag{7}$$

In (7), Φ is a set containing all possible numbers of the link capacity units that may be occupied and $q(c)$ is the probability that c link capacity units are occupied in the steady state. Further, b_i is the capacity requirement per type-i call in the linear approximation model. It has been shown in [10], [11] that the distribution $q(c)$ can be obtained from the recursion

$$cq(c) = \sum_{j=1}^{k} \rho_j b_j q(c - b_j) \quad \text{for } c \in \Phi, \tag{8}$$

in conjunction with another two constraints

$$q(c) = 0 \quad \text{for } c \notin \Phi, \tag{9}$$

$$\sum_{c \in \Phi} q(c) = 1. \tag{10}$$

Let us now take $\partial / \partial \rho_i$ throughout (7)–(10) one by one. It then follows from (7) that

$$\frac{\partial B_i}{\partial \rho_i} = \sum_{\{c \in \Phi | c + b_i \notin \Phi\}} q_i'(c), \tag{11}$$

where $q_i'(c)$ denotes $\partial q(c) / \partial \rho_i$ for $i = 1, \ldots, k$. The desired partial derivative $\partial B_i / \partial \rho_i$ can thus be obtained if one can calculate $q_i'(c)$. Now, by taking $\partial / \partial \rho_i$ at both sides of (8), we obtain

$$cq_i'(c) = b_i q(c - b_i) + \sum_{j=1}^{k} \rho_j b_j q_i'(c - b_j) \quad \text{for } c \in \Phi. \tag{12}$$

This recursive relation for $q_i'(c)$ in (12) is very similar in form as its counterpart in (8). Consequently, the computation of $q_i'(c)$ via (12) can be as efficient as the computation of $q(c)$ via (8). To complete the recursion, two constraints can be stated in parallel to (9) and (10) as follows:

$$q_i'(c) = 0 \quad \text{for } c \notin \Phi, \tag{13}$$

$$\sum_{c \in \Phi} q_i'(c) = 0. \tag{14}$$

It should be noted, however, that (14) does not render for an easy way to invoke the recursion (12). Therefore, it is desirable to obtain another relation for $q_i'(c)$. From the product-form solution, we here derive the value of $q_i'(c)$ at $c = 0$ as

$$q_i'(0) = -\frac{E(N_i) q(0)}{\rho_i}. \tag{15}$$

As a summary, one can apply (12), (13) and (15) to calculate $q_i'(c)$ for all $c \in \Phi$ and $i = 1, \ldots, k$. The developed algorithm for computing $q_i'(c)$ has both the time and space complexity of $O(Ck)$ for the analysis of single-link system with link capacity C and k call types. Obtaining $q_i'(c)$ here thus requires the same level of computational time as obtaining B_i by the algorithm in [6].

After $q_i'(c)$ is obtained for all $c \in \Phi$ and $i = 1, \ldots, k$, one can easily calculate $Var(N_i)$ from (6) and (11). Specifically, for $i = 1, \ldots, k$,

$$Var(N_i) = E(N_i) - \rho_i^2 \sum_{\{c \in \Phi | c + b_i \notin \Phi\}} q_i'(c). \tag{16}$$

Combining (5) and (16) finally gives the desired formula that can be used to calculate the approximate mean error:

$$ME_{i(apx)} = \frac{|G_i''(m_i)|}{2} \left(E(N_i) - \rho_i^2 \sum_{\{c \in \Phi | c + b_i \notin \Phi\}} q_i'(c) \right) \tag{17}$$

for $i = 1, \ldots, k$.

IV. Bounds of Error Probability

So far, the mean value of error measure $\mathcal{E}_i(N_i)$ has been considered via the notion ME_i. To obtain more information on the probability distribution of $\mathcal{E}_i(N_i)$, two types of bounds are obtained in this section.

A. Markov Bound

Since $\mathcal{E}_i(N_i)$ is a random variable that takes only nonnegative values, the Markov Inequality [12] yields that

$$P[\mathcal{E}_i(N_i) < \epsilon] \geq \max\left(0, 1 - \frac{ME_i}{\epsilon}\right) \tag{18}$$

for any value $\epsilon > 0$ and $i = 1, \ldots, k$. The statement (18) ensures, with a degree of certainty, that the error of linear approximation model will be less than a given threshold.

B. Chebyshev Bound

First note that, given $|n_i - m_i| < \delta$ for an arbitrary constant $\delta > 0$, one can obtain a bound on the error measure $\mathcal{E}_i(n_i)$ as

$$\mathcal{E}_i(n_i) < \max\left(\mathcal{E}_i(m_i - \delta), \mathcal{E}_i(m_i + \delta)\right), \tag{19}$$

where $\mathcal{E}_i\,(a < 0) \triangleq \mathcal{E}_i\,(0)$. Since the value of n_i is driven by the random variable N_i, it follows that

$$P[\mathcal{E}_i(N_i) < \max(\mathcal{E}_i(m_i - \delta_i), \mathcal{E}_i(m_i + \delta_i))] \\ \geq P[|N_i - m_i| < \delta] \tag{20}$$

Finally, since N_i has a finite mean and variance in practice, the Chebyshev Inequality [12] yields that $P[|N_i - m_i| < \delta] = 1 - P[|N_i - m_i| \geq \delta] \geq \max[0, 1 - Var(N_i)/\delta^2]$ and (20) results in

$$P[\mathcal{E}_i(N_i) < \max(\mathcal{E}_i(m_i - \delta_i), \mathcal{E}_i(m_i + \delta_i))] \\ \geq \max[0, 1 - Var(N_i)/\delta^2] \tag{21}$$

for any value $\delta > 0$ and $i = 1, \ldots, k$.

V. Numerical Results

A. Study of Mean Error

This section investigates the degree of accuracy as obtained from the approximate mean error $ME_{i(apx)}$ in (17), which is resulted from the expansion of $G_i\,(n_i)$ in (2) by only the first three terms in the Taylor's series. Like [6], we set $k = 2$ and use the equivalent capacity of the form $G_i\,(n_i) = \alpha_i n_i + \beta_i \sqrt{n_i}$. This quadratic form of equivalent capacity, derived from the stationary Gaussian traffic model [5], allows us to vary the nonlinearity in equivalent capacity conveniently by changing two parameters α_i and β_i. That is, function $G_i(n_i)$ should be less (more) nonlinear with higher value of α_i (β_i). Two experiments of complete sharing policy are here reported.

In the first experiment, $(\rho_1, \rho_2) = (10, 35)$ Erlang, $C = 150$ Mbits/s, $\alpha_1 = 0, \ldots, 3$ Mbits/s, $\alpha_2 = 1.5 + \alpha_1/2$, $\beta_i = 21.2132 - 7.0711\alpha_i$ for $i = 1, 2$. In the second experiment, $\alpha_2 = 3 - \alpha_1$ and all other settings remain the same as in the first experiment. The settings of (α_i, β_i) are such that $G_i\,(n_i)$ always pass through the origin $(0, 0$ Mbits/s$)$ and the point $(50, 150$ Mbits/s$)$. Examples of $G_i\,(n_i)$ can be found in Fig. 1 of [6]. The resultant exact and approximate values of ME_1, ME_2 are plotted as a function of α_1 in Figs. 2 and 3 for the first and second experiments, respectively.

Figs. 2 and 3 suggest that both ME_1 and ME_2 can be well captured by the approximation in (17) over the whole range of equivalent capacity considered here. This finding is useful because it implies that one needs not worried about other higher (than second) moment terms in the Taylor's series to derive for a meaningful error indicator for the linear approximation model. The importance of this implication is increased especially when one realizes that, in practice, calculating the variance is difficult enough and other higher moments are even harder (if not impossible at all) to obtain.

B. Remarks on Bounds of Error Probability

The complete sharing policy is again assumed. The link capacity is set to 150 Mbits/s. Loads offered to this link are fixed at $(\rho_1, \rho_2) = (10, 35)$ Erlang. Again, β_i is set to $21.2132 - 7.0711\alpha_i$ for $i = 1, 2$. The value of (α_1, α_2) is set to $(0, 1.5)$ Mbits/s and the resultant relationship between the exact value and bounds of $P\,[\mathcal{E}_i\,(N_i) \geq \epsilon]$ is depicted by Figure 4.

Figure 4 suggests that the exact value of $P\,[\mathcal{E}_i\,(N_i) \geq \epsilon]$ is well captured by its bounds, where the Chebyshev bound is found to be tighter than the Markov bound.

VI. Conclusions

In this paper, an error analysis has been carried out for the linear approximation model as applied to the multiservice single-link system studies with nonlinear equivalent capacity. The notion of mean error ME_i has been introduced and its approximation $ME_{i(apx)}$ has been obtained in (5). This approximate mean error $ME_{i(apx)}$ can capture the two main factors that influence the error of linear approximation model, namely, (i) the nonlinearity in equivalent capacity at the mean operating point as quantified by the second derivative term in (5) and (ii) the fluctuation of system dynamics around the mean operating point as captured by the variance term in (5).

In addition, a computationally efficient procedure (by (12), (13), (15) and (17)) has been obtained to calculate $ME_{i(apx)}$ with the time and space complexity of $O\,(Ck)$, where C is the link capacity and k is the number of call types. The derived formula for $ME_{i(apx)}$ can thus be computed in parallel to the underlying solution of all the main system performance

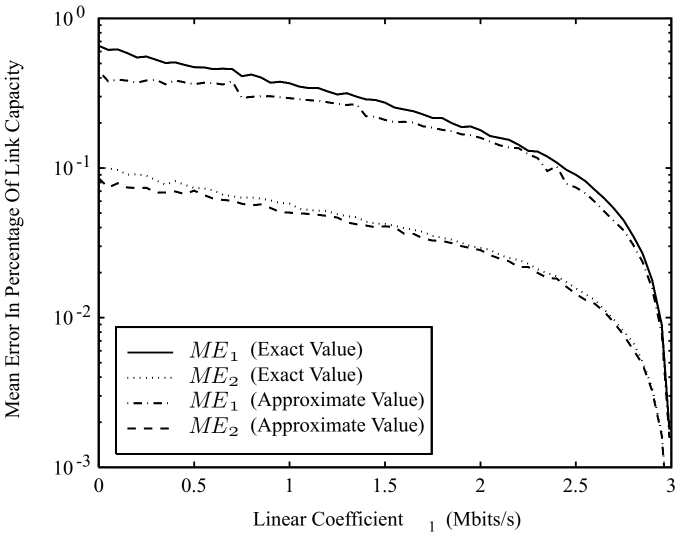

Fig. 2. Exact and approximate mean errors in the first experiment.

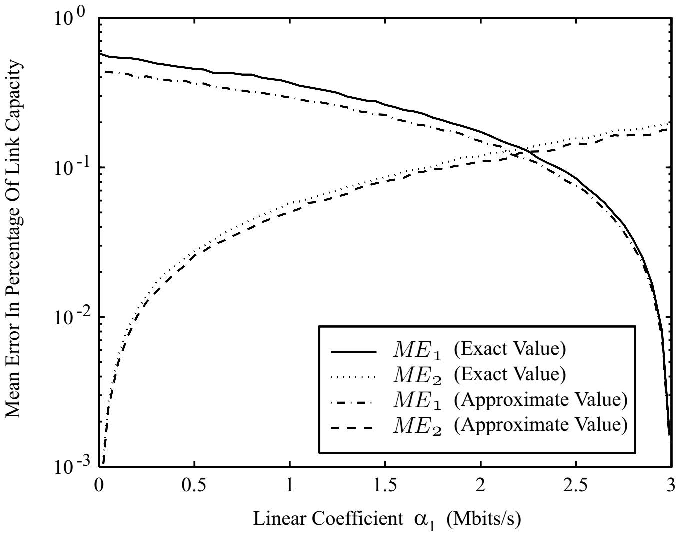

Fig. 3. Exact and approximate mean errors in the second experiment.

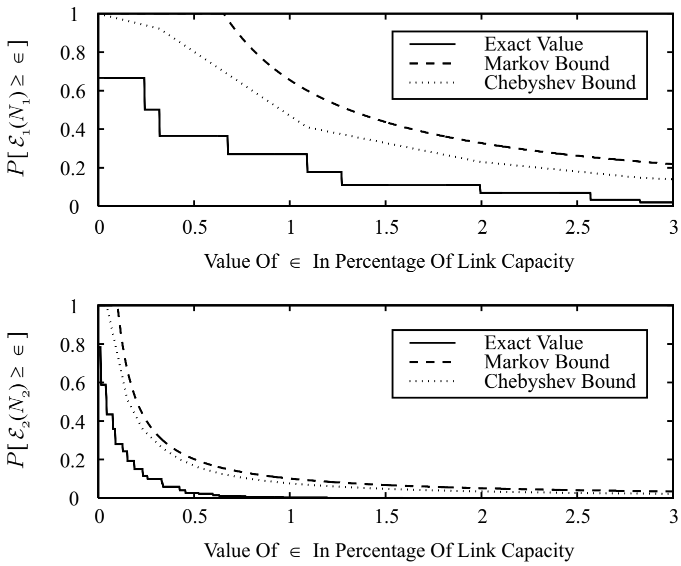

Fig. 4. Bounds of error probability.

parameters (call blocking probabilities and mean revenue rates).

Finally, the error of linear approximation model has been bounded probabilistically by the Markov and Chebyshev Inequalities. The Markov bound and Chebyshev bound can be calculated from the mean error ME_i and the variance $Var\,(N_i)$, respectively. The computationally efficient recursion in Section III should also thus be readily applicable to obtain both of these bounds in practice. Given the numerical results herein reported (as well as all others in [13]), it is expected that the error measures in this paper can be used as the standard criteria of judgement on when the CAC with dynamic service separation (which is a nonlinear CAC) can be acceptably approximated by a linear model.

ACKNOWLEDGMENT

The authors would like to thank all reviewers for helpful suggestions. C. Aswakul is grateful to the Anandamahidol Foundation Scholarship and Chulalongkorn University's Ratchadaphiseksomphot Endownment Fund, Thailand.

REFERENCES

[1] K. W. Ross, *Multiservice Loss models for Broadband Telecommunications Networks*, London: Springer-Verlag, 1995.
[2] K. Sriram, "Methodologies for bandwidth allocation, transmission scheduling, and congestion avoidance in broadband ATM networks," *Computer Networks and ISDN Systems*, Vol. 26, 1993, pp. 43–59.
[3] R. Bolla, F. Davoli, and M. Marchese, "Bandwidth allocation and admission control in ATM networks with service separation," *IEEE Communications Magazine*, May 1997, pp. 130–137.
[4] E. Gelenbe, X. Mang, and R. Onvural, "Bandwidth allocation and call admission control in high-speed networks," *IEEE Communications Magazine*, May 1997, pp. 122–129.
[5] R. Guérin, H. Ahmadi, and M. Naghshineh, "Equivalent capacity and its application to bandwidth allocation in high-speed networks," *IEEE Journal on Selected Areas in Communications*, Vol. SAC-9, September 1991, pp. 968–981.
[6] C. Aswakul and J. Barria, "Performance analysis of single-link system with nonlinear equivalent capacity," *IEEE Communication Letters*, Vol. 4, No. 2, February 2000, pp. 71–73.
[7] C. Aswakul and J. Barria, "Analysis of dynamic service separation with trunk reservation policy," *IEE Proceedings - Communications*, Vol. 149, No. 1, February 2002, pp. 23–28.
[8] C. Aswakul and J. Barria,"Performance analysis of ATM routing with nonlinear equivalent capacity: symmetric case," *IEEE International Conference on Computer Communications and Networks (ICCCN) 2001*, 15-17 October 2001, Arizona, USA.
[9] S. Jordan and P. P. Varaiya, "Throughput in multiple service, multiple resource communication networks," *IEEE Transactions on Communications*, Vol. 39, No. 8, August 1991, pp. 1216–1222.
[10] J. S. Kaufman, "Blocking in a shared resource environment," *IEEE Transactions on Communications*, Vol. COM-29, October 1981, pp. 1474–1481.
[11] J. W. Roberts, "A service system with heterogeneous user requirement" In G. Pujolle, ed., *Performance of Data Communications Systems and Their Applications*, Amsterdam: North-Holland Publishing Company, 1981, pp. 423–431.
[12] S. Ross, *A First Course in Probability*, 4th ed., Singapore: Prentice Hall, 1994.
[13] C. Aswakul, *Performance Analysis of Routing in ATM Networks with Nonlinear Equivalent Capacity Functions*, PhD Thesis, Imperial College of Science, Technology and Medicine, University of London, U.K., 2000.