Generalized Loss Models, Loss Networks and Their Applications

Hisashi Kobayashi and Brian L. Mark *
Department of Electrical Engineering
Princeton University
Princeton, New Jersey 08544, USA
Tel: +1 609 258-1984; email: hisashi@princeton.edu

ABSTRACT

There is a renewed interest in the classical Erlang's loss formula and its related topics [4, 12], because the two rapidly growing network segments– i.e., ATM networks and wireless networks–both provide basically connection-oriented services, thus key measures of their QoS (quality of services) are time congestion (or blocking probability) and call congestion (or call loss probability).

In this expository paper, we present our recent results [5, 7, 8] on generalized versions of the Erlang and Engset loss models. We will then show that a generalized loss station (GLS) with multiple Poisson streams is equivalent to an open loss network (OLN). Finally, we report that GLS can be incorporated into product-form queueing networks. Some application examples will be briefly discussed.

I. INTRODUCTION

In the classical Erlang loss model, arriving calls form a Poisson process with rate λ and each call holds one of the m output lines for exponentially distributed time with mean $1/\mu$. The probability that all lines are busy is known as the Erlang loss formula, (or Erlang's B formula):

$$B(m) = \frac{a^m}{m!}\left[\sum_{n=0}^{m}\frac{a^n}{n!}\right]^{-1} \quad (1)$$

where $a = \lambda/\mu$. The Erlang's B formula (1) can be expressed as

$$B(m) = 1 - \frac{G(m-1)}{G(m)}, \quad (2)$$

where $G(m) = \sum_{n=0}^{m}\frac{a^n}{n!}$ is the normalization constant. $B(m)$ is also called *time congestion*, since it represents the proportion of time that all lines are busy. *Call congestion* $L(m)$ is the probability that an arriving call finds all m lines busy. In the Erlang Loss model, the two congestion measures are equivalent, i.e., $L(m) = B(m)$, because of the so-called PASTA (Poisson Arrivals See Time Averages) property.

*Presently with NEC C&C Research Laboratories, 4 Independence Way, Princeton, NJ 08540.

If we replace the Poisson source by a finite source model, we obtain the Engset loss model. The source consists of K independent mini-sources, each of which generates a new call according to a renewal process with rate ν. Then the time congestion is given by the Engset loss formula:

$$E(m,K) = \frac{1}{G(m,K)}\binom{K}{m}b^m = 1 - \frac{G(m-1,K)}{G(m,K)}, \quad (3)$$

where $b = \nu/\mu$, and $G(m,K) = \sum_{n=0}^{m}\binom{K}{n}b^n$. Call congestion $L(m,K)$ and the time congestion satisfy the simple relation $L(m,K) = E(m,K-1)$.

II. GENERALIZED LOSS STATION (GLS)

The classical loss models can be generalized by introducing multi-class calls and by removing the the exponential holding time assumption. We also allow a call to hold simultaneously multiple servers.

Definition 1: Generalized Loss Station (GLS) *A Generalized Loss Station (GLS) is a station with m parallel servers, no storage, and the following additional properties:*

1. **Multi-class calls:** *It serves a set, C, of call classes. Calls of class $c \in C$ arrive according to either (i) a Poisson process with rates λ_c (i.e., a generalized Erlang loss model) or (ii) a finite source model with K_c mini-sources generating renewal arrivals at rate ν_c.*

2. **Simultaneous holding of multiple servers:** *A class c call holds A_c servers simultaneously, when the call is in service. Then, $N_c(t)$, the number of class c calls in service at time t, must satisfy*

$$\sum_{c\in C} A_c N_c(t) \le m. \quad (4)$$

In the finite source model, additional constraints

$$N_c(t) \le K_c, \quad c \in C \quad (5)$$

must be met.

3. **General holding time:** *The holding time distribution for a class c call is generally distributed with mean $1/\mu_c$.*

Let the state process be denoted by $\mathbf{N}(t) = (N_c(t) : c \in \mathcal{C})$, and $P(\mathbf{n})$ be the equilibrium state distribution of $\mathbf{N}(t)$. The GLS has many of the properties associated with product-form queueing networks. Specifically, we extend the notion of quasi-reversibility [3].

Theorem 1: Quasi-Reversibility of GLS
The GLS is quasi-reversible. Furthermore, $\mathbf{N}(t)$ is a reversible Markov process with the following stationary distributions.

Generalized Erlang Loss Station:

$$P(\mathbf{n}) = \frac{1}{G(m)} \prod_{c \in \mathcal{C}} \frac{a_c{}^{n_c}}{n_c!}, \quad \mathbf{n} \in \mathcal{F}(m) \qquad (6)$$

where $a_c = \lambda_c/\mu_c$, $\mathcal{F}(m)$ is the set of states for which (4) holds, and $G(m)$ is the normalization constant defined by

$$G(m) = \sum_{\mathbf{n} \in \mathcal{F}(m)} \prod_{c \in \mathcal{C}} \frac{a_c{}^{n_c}}{n_c!} \qquad (7)$$

Generalized Engset Loss Station:

$$P(\mathbf{n}) = \frac{1}{G(m, \mathbf{K})} \prod_{c \in \mathcal{C}} \binom{K_c}{n_c} b_c{}^{n_c}, \quad \mathbf{n} \in \mathcal{F}(m, \mathbf{K}) \quad (8)$$

where $b_c = \nu_c/\mu_c$, $\mathcal{F}(m, \mathbf{K})$ is the set of states for which (4) and (5) hold, and $G(m, \mathbf{K})$ is given by

$$G(m, \mathbf{K}) = \sum_{\mathbf{n} \in \mathcal{F}(m, \mathbf{K})} \prod_{c \in \mathcal{C}} \binom{K_c}{n_c} b_c{}^{n_c} \qquad (9)$$

proof. See [7, 8].

III. OPEN LOSS NETWORKS (OLN)

Let us now consider a loss network [4] shown in Figure 1. We call this type of loss network an *open loss network* (OLN) [5]. We define the following notation

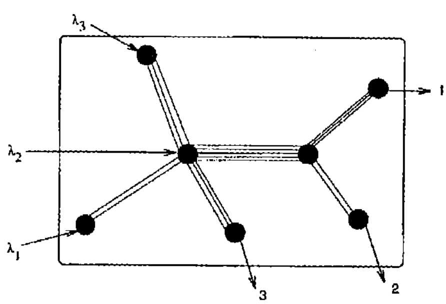

Figure 1: Open Loss Network (OLN)

and properties of the OLN.

1. Let there be L links in the OLN, and $\mathcal{L}$ denotes the set of links. A link ℓ contains m_ℓ lines or servers, $\ell \in \mathcal{L}$
2. A call class $c \in \mathcal{C}$ is defined as a pair (p, τ), where $p \in \mathcal{P}$ is the *path* of a call, and $\tau \in \mathcal{T}$ is the *call type*. Thus, $\mathcal{C} = \mathcal{P} \times \mathcal{T}$. In a loss network, the notion of a class is also referred to as a *route*. [1]
3. The arrival of class c calls to the OLN is a Poisson process with rate λ_c, $c \in \mathcal{C}$.
4. A class $c = (p, \tau)$ call holds $A_{\ell c}$ lines of link ℓ simultaneously for all ℓ in its path p.
5. The holding time of a class c call is a general distribution with mean $1/\mu_c$.

This OLN model provides a general model for a *circuit-switched network* that carries *multi-rate* traffic (i.e., different values of $A_{\ell c}$ for different c) among different classes of call.

We now consider a loss station, which consists of multiple server types $\ell \in \mathcal{L}$. We term this generalized loss station a GLS with *multiple types of servers*. Hence the number of servers is now represented by a vector $\mathbf{m} = \{m_\ell, \ell \in \mathcal{L}\}$. A class c call needs to hold $A_{\ell c}$ type-ℓ servers, $\ell \in \mathcal{L}$.

We now state the following important theorem that relates the OLN to the GLS:

Theorem 2: Equivalence of OLN and GLS, and Reversibility
The OLN is equivalent to the GLS with **multiple server types** [2]*, in which there are m_ℓ type-ℓ servers, $\ell \in \mathcal{L}$. Let $N_c(t)$ be the number of class c calls in progress at time t. Then, the process $\mathbf{N}(t) = \{N_c(t); c \in \mathcal{C}\}$ is reversible, and possesses the stationary distribution*

$$P(\mathbf{n}) = \frac{1}{G(\mathbf{m})} \prod_{c \in \mathcal{C}} \frac{a_c^{n_c}}{n_c!}, \quad \mathbf{n} \in \mathcal{F}(\mathbf{m}) \qquad (10)$$

where $a_c = \lambda_c/\mu_c$, and

$$G(\mathbf{m}) = \sum_{\mathbf{n} \in \mathcal{F}(\mathbf{m})} \prod_{c \in \mathcal{C}} \frac{a_c^{n_c}}{n_c!} \qquad (11)$$

$$\mathcal{F}(\mathbf{m}) = \{\mathbf{n} \geq 0 : \sum_{c \in \mathcal{C}} A_{\ell c} n_c \leq m_\ell\}, \quad (12)$$

proof: The proof is a straightforward extension of that for Theorem 1. Equate the pair (ℓ, c) to the class c in Theorem 1. □.

IV. MIXED LOSS NETWORKS (MLN)

We define a *closed loss network* (CLN) by replacing the set of Poisson streams by a set of finite sources. In analogy to *open* and *closed* subchains for customer

[1] The GLS model of the previous section is a degenerate case, where the call class $\mathcal{C}$ and the call type $\mathcal{T}$ are equivalent. In an OLN, calls of the same type may take different routes within the network.

[2] Note that the link identifier $\ell \in \mathcal{L}$ in the OLN corresponds to the server type in its equivalent GLS.

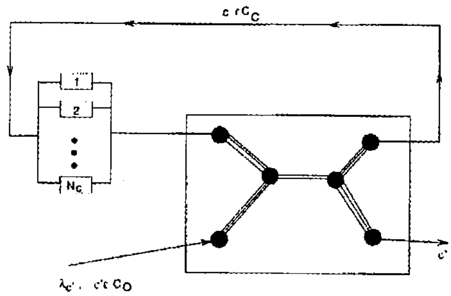

Figure 2: Mixed Loss Network

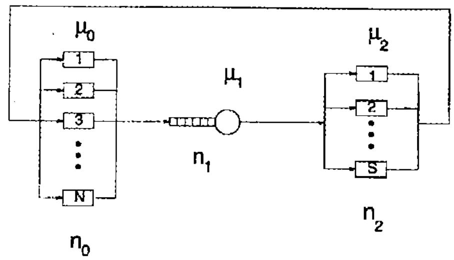

Figure 3: A Simple Queueing-Loss Network

routing in queueing networks, we equate the *chain* of a call to its class and partition the *classes* of a loss network into the set of open classes, $\mathcal{C}_O$, and the set of closed classes, $\mathcal{C}_C$.

An OLN consists entirely of open classes, while a CLN consists entirely of closed classes. A *mixed loss network* (MLN) is a loss network that has both kinds of classes. The MLN may include both generalized Erlang and Engset stations. Denote its state process by $\mathbf{N}(t) = [\mathbf{N}_O(t), \mathbf{N}_C(t)]$, with $\mathbf{N}_O(t) = (N_c(t) : c \in \mathcal{C}_O)$ and $\mathbf{N}_C(t) = (N_c(t) : c \in \mathcal{C}_C)$, where $\mathcal{C}_O$ and $\mathcal{C}_C$ represent the sets of closed subchains and open subchains, respectively. We have the following result for the MLN:

Theorem 3: MLN and Reversibility
The state process of the MLN is a reversible Markov process with equilibrium distribution

$$P(\mathbf{n}) = \frac{1}{G(\mathbf{m}, \mathbf{K})} P_O(\mathbf{n}_O) P_C(\mathbf{n}_C), \quad \mathbf{n} \in \mathcal{F}(\mathbf{m}, \mathbf{K}) \tag{13}$$

where

$$P_O(\mathbf{n}_O) = \prod_{c \in \mathcal{C}_O} \frac{a_c^{n_c}}{n_c!}, \quad P_C(\mathbf{n}_C) = \prod_{c \in \mathcal{C}_C} \binom{K_c}{n_c} b_c^{n_c} \tag{14}$$

with $a_c = \lambda_c/\mu_c \ \ (c \in \mathcal{C}_O), \quad b_c = \nu_c/\mu_c, \ (c \in \mathcal{C}_C)$, *and*

$$\mathcal{F}(\mathbf{m}, \mathbf{K}) = \{\mathbf{n} \geq \mathbf{0} : \ \sum_{c \in \mathcal{C}} A_{\ell c} n_c \leq m_\ell, \ \ell \in \mathcal{L}\} \tag{15}$$

where constraints $n_c \leq K_c$, $c \in \mathcal{C}$ *should be added for class* $c \in \mathcal{C}_C$. *The normalization constant is*

$$G(\mathbf{m}, \mathbf{K}) = \sum_{\mathbf{n} \in \mathcal{F}(\mathbf{m}, \mathbf{K})} P_O(\mathbf{n}_O) P_C(\mathbf{n}_C). \tag{16}$$

We can also show that an OLN component is *quasi-reversible*. We omit the proofs, since they are straightforward generalizations of Theorems 1 and 2.

Time congestion and *call congestion* for class-c can be expressed in terms of the normalization constant $G(\mathbf{m}, \mathbf{K})$:

1. For an open class c, $c \in \mathcal{C}_O$,

$$B_c(\mathbf{m}, \mathbf{K}) = 1 - \frac{G(\mathbf{m} - \mathbf{A}_c, \mathbf{K})}{G(\mathbf{m}, \mathbf{K})} \tag{17}$$

$$L_c(\mathbf{m}, \mathbf{K}) = B_c(\mathbf{m}, \mathbf{K}), \tag{18}$$

where $\mathbf{A}_c$ is the c-th column of the matrix $\mathbf{A} = [A_{\ell c}]$.

2. For an closed class c call, $c \in \mathcal{C}_C$,

$$E_c(\mathbf{m}, \mathbf{K}) = 1 - \frac{G(\mathbf{m} - \mathbf{A}_c, \mathbf{K})}{G(\mathbf{m}, \mathbf{K})} \tag{19}$$

$$L_c(\mathbf{m}, \mathbf{K}) = E_c(\mathbf{m}, \mathbf{K} - \mathbf{1}_c), \tag{20}$$

where $\mathbf{1}_c$ denotes the unit vector, whose c-th component is unity.

As for numerical methods (exact, approximate and asymptotic) to compute the normalization constants $G(\mathbf{m}, \mathbf{K})$ for different values of $\mathbf{m}$ and $\mathbf{K}$, the reader is referred to our recent book chapter [8] and references therein.

V. QUEUEING-LOSS NETWORKS

First, let us consider a simple three station network connected in a cyclic order as shown in Figure 3. We label the the stations as station 0, 1, and 2, respectively: (a) station 0 represens a finite source model with N mini-sources, with each generating renewal arrivals at rate μ_0; (b) station 1 is a single server with exponential service time of mean $1/\mu_1$; (c) station 2 is a loss station with S servers, and the call holding time has a general distribution with mean μ_1. We assume that $N > S$ (otherwise blocking would not occur at all). This model may approximately represent a situation where a random amount of time (with mean $1/\mu_1$) is spent for processing a call connection request whether or not the request is honored. Even this simplest model cannot be solved by the conventional traffic theory.

By generalizing the above simple model, let us introduce the concept of a *queueing-loss network* (QLN), which is a network (see Figure 4) that contains both queueing subnetwork(s) and loss subnetwork(s). Let a QLN contain a set of queueing sub-

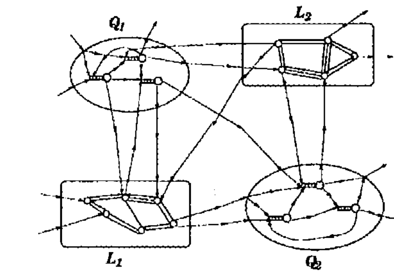

Figure 4: General Queueing-Loss Network (QLN)

networks $\{Q_j;\ j \in \mathcal{J}\}$ and a set of loss subnetworks $\{\mathcal{L}_k;\ k \in \mathcal{K}\}$. These subnetworks may be connected in an arbitrary manner. Arrival traffic from the outside sources are independent Poisson streams. Let $\mathbf{n}_{Q_j}$ and $\mathbf{n}_{\mathcal{L}_k}$ represent the population vectors in these subnetworks. If these vectors have all product form solutions for their equilibrium state distributions, then the joint distribution of the state process $\mathbf{N}(t)$ of the QLN takes the following form in the equilibrium.

$$P(\mathbf{n}) = \frac{1}{G(\mathbf{m},\mathbf{K})} \prod_{j\in\mathcal{J}} P_{Q_j}(\mathbf{n}_{Q_j}) \prod_{k\in\mathcal{K}} P_{\mathcal{L}_k}(\mathbf{n}_{\mathcal{L}_k}) \quad (21)$$

where $P_{Q_j}(\cdot)$ and $P_{\mathcal{L}_k}(\cdot)$ themselves have product forms and are proportional to the marginal distributions of the subnetworks Q_j and $\mathcal{L}_k$, $j \in \mathcal{J}, k \in \mathcal{K}$. The normalization constant $G(\mathbf{m},\mathbf{K})$ and the feasible state set $\mathcal{F}(\mathbf{m},\mathbf{K})$ are defined over the vectors $\mathbf{m} \times \mathbf{K}$.

The above result can be proved by noting the fact that each OLN can be replaced by its equivalent GLS, and that a GLS can be viewed as a generalized IS (infinite server) station, since no queue will be formed at the station. We have also shown that each GLS is a quasi-reversible queue, hence the overall state process $\mathbf{N}(t)$ of the QLN is a reversible process. When a loss subnetwork is an MLN and contains a finite number of mini-sources within the subnetwork itself, we can further decompose the subnetwork into an OLN part and IS stations (K_c mini-sources are represented by an IS station of class c.) Each of these component stations are quasi-reversible, as we observed before.

By combining these observations we find that the QLN is a generalized queueing network. The only difference from the conventional queueing networks is that it now contains GLSs as its components. We have already noted that a GLS can be treated a generalized IS station, and is a quasi-reversible station. Thus, we can conclude that the QLN has a product form solution.

By referring to the simple model of Figure 3, the steady state distribution is given by

$$P(n_0, N-n_0-n_2, n_2) = \frac{1}{G(S,N)} \frac{1}{n_0!} \left(\frac{\mu_1}{\mu_0}\right)^{n_0+n_2} \frac{1}{n_2!} \left(\frac{\mu_0}{\mu_2}\right)^{n_2} \quad (22)$$

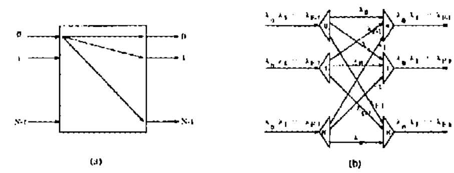

Figure 5: (a) $N \times N$ WGR; (b) Its demultiplexer-multiplexer representation

where

$$\begin{aligned} G(S,N) &= \sum_{0\le n_0+n_2\le N, n_2\le S} \frac{1}{n_0!} \left(\frac{\mu_1}{\mu_0}\right)^{n_0+n_2} \frac{1}{n_2!} \left(\frac{\mu_0}{\mu_2}\right)^{n_2} \\ &= \sum_{k=0}^{N} \frac{1}{k!} \sum_{n_2=0}^{\min\{k,S\}} \binom{k}{n_2} \left(\frac{\mu_1}{\mu_0}\right)^{k} \left(\frac{\mu_0}{\mu_2}\right)^{n_2} \end{aligned} \quad (23)$$

Time congestion and call congestion at station 2 are then obtained by

$$E(S,N) = 1 - \frac{G(S-1,N)}{G(S,N)} \quad (24)$$

$$L(S,N) = E(S,N-1) \quad (25)$$

$$= 1 - \frac{G(S-1,N-1)}{G(S,N-1)} \quad (26)$$

VI. APPLICATIONS OF GENERALIZED LOSS MODELS

The generalized loss station (GLS) model and the open loss network (OLN) model we discussed in the previous sections significantly enlarges the class of network services for which we can construct analytical models. Any connection-oriented service with multiple service rates is a good candidate for the GLS or loss network model. We briefly discuss below a few obvious examples.

A. All-optical networks (AON)

An all-optical network (AON)[1, 6], which connects a large number of users via wavelength grating router (WGR), can be modeled as a loss network with fixed route. In Figure 5 we show an input-output relation of a WGR and its demultiplexer-multiplexer representation. The $N \times N$ WGR (or static wavelength router) connects input port σ_m to output port σ_n ($m, n = 0, 1, \ldots, N-1$) using wavelength λ_f

$$n = m + f \ \ (\text{modulo } N). \quad (27)$$

From the above relation we readily find that if $N = F$, there exists a ***unique wavelength*** that routes a call originating from a source node to some destination node. If $F < N$, there may not exist a wavelength that can carry the call, i.e., a connection between some pair of nodes may not be possible. If $F > N$, there may be more than one wavelength that can carry the call.

Consider a simple network in which the N input and output ports of a WGR are connected to N LANs (local area networks). Such interconnection of many LANs by a WGR may, therefore, constitute a MAN (metropolitan area network). The AON prototype discussed by Alexander et al. [1] is such an example. Then if we can represent the call generations from each source (i.e., LAN) as a Poisson process, the product-form formula based on the generalized Erlang model holds.

If the static router is converted into a dynamic router by use of wavelength converters (frequency changers), then the simple product-form no longer holds, and some approximation technique is called for, as in the three-stage switching network. See [9] for a further discussion and numerical examples.

B. ATM network model and admission control

The product-form solution applies to any network with fixed route. Hence, the connection-oriented service in an ATM network can be modeled as a loss network. Different service rates for different users (e.g. virtual paths) can be explicitly handled by proper choice of the integers A_c, or $A_{\ell c}$. The call admission control (CAC) may make use of the projected QoS (quality of service) that can be estimated by the time and call congestion formulae we have. For VBR (variable bit rate) traffic, the service rates parameters A_c will not be constant; however, the so-called *effective bandwidth* of the call can be used as an approximate value for A_c.

C. Cellular radio networks

The channel assignment problem in a cellular radio network can be easily modeled as a GLS system. If the number of mobile users is not so large, the finite source model (i.e., Engset model and its generalized version) should be used. The model formulation equally applies whether the cellular network adopts FDM (e.g., AMPS), TDMA (e.g., GSM) or CDMA (e.g., IS-95). The model applies to both the forward link (down link) and reverse link (up link). The GLS model formulation is most appropriate when the services provide different rates, i.e, data services as well as voice services. The bandwidth or channel assignment problem in a satellite communication services can be formulated in a similar manner.

ACKNOWLEDGEMENTS

This work has been supported, in part, by the National Science Foundation's Program on Communications and Networking, New Jersey Commission on Science and Technology, and the Ogasawara Foundation for the Promotion of Science and Engineering.

REFERENCES

[1] Alexander, S. B. et al. (1993). "A Precompetitive Consortium on Wide-Band All-Optical Networks," *Journal of Lightwave Technology*, Vol. 28, No. 5/6,pp. 714-735, May/June 1993.

[2] Cohen, J. W. (1957). "The Generalized Engset Formulae," *Philips Telecommunication Review*, 18 (4): 158-170, November 1957.

[3] Kelly, F. P. (1979). *Reversibility and Stochastic Networks* John Wiley & Sons.

[4] Kelly, F.P. (1991). "Loss Networks." *The Annals of Applied Probability*, 1:319-378.

[5] Kobayashi, H., and B. L. Mark (1994), "On Queuing Networks and Loss Networks," *Proc. of the 28th Annual Conference on Information Sciences and Systems*, pp. 794-799, March 1994., Princeton University, N. J.

[6] Kobayashi, H., and I. P. Kaminow (1996). "Duality Relationships among 'Space', 'Time' and 'Wavelength' in All-Optical Networks," *Journal of Lightwave Technology*, pp. 344-351, March 1996.

[7] Kobayashi, H., and B. L. Mark (1995), "A Unified Theory for Queuing and Loss network Models," *Proc. 2nd IEEE Malaysia International Conference on Communications*, Nov. 20-23, 1995, Langkawi, Malaysia.

[8] Kobayashi, H., and B. L. Mark (1997), "Product-Form Loss Networks," in J. Dshalalow (Ed.), *Frontier in Queueing: Models and Applications in Science and Engineering*, CRC Press, 1997, pp. 147-195.

[9] Kobayashi, H., B. L. Mark and Y. Osaki, "Call Blocking Probability of All-Optical Networks," *Proc. the 1995 IEEE Broadband Switching Conference*, April 1995, Poznan, Poland, pp. 186-200.

[10] Kogan, Y. (1994). "Asymptotic Solution of Generalized Multiclass Engset Model," in J. Labetoulle and J. Roberts (Eds.) *Proc. ITC 14*, Vol. 1b, pp. 1239-1250, Elsevier, Amsterdam, 1994.

[11] Roberts, J. W. (ed.) (1992) *Performance Evaluation and Design of Multiservice Networks: COST 224 Project Report.* Commission of the European Communities, 1992.

[12] Ross, K. W. (1995). *Multiservice Loss Models for Broadband Telecommunication Networks*, Springer.

[13] Syski, R. (1986). *Introduction to Congestion Theory in Telephone Systems* (2nd edition). Amsterdam·New York·Cambridge: North-Holland

[14] Whitt, W. (1985). "Blocking when Service is Required from Several Facilities Simultaneously," *AT&T Technical Journal*, 64, pp. 1807-1856, 1985.