Blocking Probability Analysis of Circuit Switched Networks with Long-lived and Short-lived Connections

Meiqian Wang, Shuo Li, Eric W. M. Wong, *Senior Member, IEEE,* Moshe Zukerman, *Fellow, IEEE*

***Abstract*—We consider a circuit-switched network with non-hierarchical alternate routing and trunk reservation involving two types of connections that are modeled as long-lived and short-lived calls. The long-lived calls can be reserved well in advance and the short-lived calls are provided on demand. Therefore, we assume that the long-lived calls have strict priority over the short-lived ones. We develop approximations for the estimation of the blocking probability based on the quasi-stationary approach in two ways. One uses the Erlang fixed-point approximation (EFPA) and the other uses the overflow priority classification approximation (OPCA). We compare the results of the approximations with simulation results and discuss the accuracy of the approximations under different system parameters such as ratio of offered load, number of links per trunk, maximum allowable number of deflections and trunk reservation. We also discuss the robustness of the quasi-stationary approximation to the ratio of the mean holding times of the long-lived and short-lived calls as well as that of EFPA and OPCA to the shape of the holding time distribution. Finally, we demonstrate that OPCA can be applied to a large network such as the Coronet.**

***Index Terms*—blocking probability, circuit switching, alternate routing, trunk reservation, Erlang fixed-point approximation, overflow priority classification approximation, quasi-stationary approximation, OFS**

I. Introduction

Circuit switching has been widely used in telephony and it is envisaged that it will have a renewed and important role in future optical networks [1]–[5]. Circuit switching normally does not require buffering, which is very costly in the optical domain. If the traffic on a circuit-switched network is sufficiently heavy and it is well managed, such a network can guarantee quality of service (QoS) to customers in a way that can lead to efficient trunk utilization and low consumption of energy per bit [6]. In the core Internet, where traffic is heavily multiplexed, it is easier to achieve high utilization and therefore the role of circuit switching at the core is clearly important. Circuit switching can also lead to a green and efficient operation end-to-end for large bursts of data if the amount of data to be transmitted is known in advance.

Circuit-switched networks may encounter traffic overload situations when the offered traffic exceeds the network capacity, and in such situations, there is a need to reject (block, or drop) new calls. The proportion of blocked traffic is defined as blocking probability. Such blocking (or dropping) has adverse implication on QoS perceived by users and therefore blocking probability should not exceed a certain predetermined values. For over a century, operators have considered blocking probability as a key performance measure for circuit switched network design and dimensioning. In today's competitive environment, limiting blocking probability is clearly important.

The authors are with Department of Electronic Engineering, City University of Hong Kong, Hong Kong SAR.

To reduce blocking probability in circuit-switched networks, various approaches for dynamic or alternate routing have been studied [7]–[9], where new calls that cannot be admitted by their primary routes may overflow to other routes that may be more costly in terms of the use of network resources. The use of more and more costly alternate routes may adversely affect blocking probability. Therefore, the number of allowable overflow attempts by a particular call may be limited, so a call is blocked after a finite number of alternate routes attempts.

While traffic flows are transmitted optically on the data plane, the configuration of an optical cross connect (OXC) is done by an electronic control unit. The control units of different OXCs can also communicate with each other and compose the control plane, which is responsible for lightpath set up and tear down, as well as routing, label distribution and state dissemination using protocols such as GMPLS and RSVP-TE [10].

A circuit-switched network with alternate routing, under the assumptions of Poisson call arrivals for any source destination (SD) pair and exponential call holding times, can be modeled as a Markovian overflow loss network. The transition rates between states of a trunk are obtained considering arrival and departure rates of connections on their primary path and those that overflow to the trunk from other fully occupied trunks. The stationary occupancy distribution can, in principle, be obtained by a numerical solution of the steady-state equations of a multidimensional Markov process. Such models usually do not admit product-form solutions [11] and are not amenable to analysis that lead to a scalable solution for realistic size networks. Therefore accurate, robust and scalable approximations are important.

Circuit-switched networks with alternate routing can be classified into two classes: hierarchical and non-hierarchical. In hierarchical networks, the trunks are ranked into several tiers. New calls first attempt to access trunks from the lowest tier, if rejected, they overflow and attempt to access trunks from higher tiers. For hierarchical networks, accurate blocking probability approximations can be obtained using moment matching approaches [12], [13].

A more difficult problem is to accurately approximate blocking probabilities in non-hierarchical networks with *mutual* *overflows* [7], where congestion on a specific trunk causes overflow to the other trunks, and where this overflow loads up the other trunks so that they in turn yield overflow back to the original trunk. Clearly, in such non-hierarchical systems, load dependencies may be much stronger than in hierarchical systems that do not have mutual overflow, and modeling involving independence assumptions is more likely to lead to significant errors. Despite wide applicability and importance, and a century-long research effort, no robust and generic methodology is available for approximation of the blocking probability of general non-hierarchical networks that captures their overflow-induced state dependencies in a scalable way, except for studies for particular applications [14]–[16].

One simple and commonly used approach for approximation of blocking probabilities in non-hierarchical networks is the *Erlang fixed-point approximation* (EFPA) [9], [17]–[20], proposed by Cooper and Katz [21] in 1964. It is based on decoupling a given system into independent subsystems, each of which is modeled by an M/M/K/K queueing system. Convergence and uniqueness of solutions of EFPA is not always guaranteed [9], [22]. Despite its popularity, EFPA is known to introduce two types of errors, Poisson error and independence error [23]. Various attempts to address the errors include [21], [24] that provided means to reduce the Poisson error by moment matching and [25] that tackled the independence error by capturing the correlation between trunks.

In recent years, we have developed a new method called *overflow priority classification approximation* (OPCA) [23], [26]–[28] that can either improve or complement EFPA to achieve more accurate blocking probability approximations for circuit switched networks with alternative routing.

In this paper, we retain the assumption that holding times are independent and follow identical exponential distribution as in most of the published work on analysis of circuit-switched networks. However, we assume that the calls are classified into two types, long-lived (static) and short-lived (dynamic), where the holding times of the long-lived calls are significantly larger than those of the short-lived calls. This assumptions is justified by the fact that the variations in holding times in circuit switched optical networks can be significant. Permanent or semi-permanent connections between major cities or data centers [1], that are used to serve many flows of many users over a long period of time, can be considered as (long-lived) calls where the holding time can be in terms of hours. Examples of such long-lived connections can also include circuit switched connections in the Large Hadron Collider (LHC) network [2], which can be modeled as long-lived calls. On the other hand, on-demand dynamic connections between individual SD pairs of users in the order of seconds or less may be classified as short-lived calls. Classification of calls to long-lived and short-lived have been discussed in various papers (e.g. [29], [30], [31], [32]).

Long-lived connections are likely to have priority over the short-lived ones, as long-lived connections can be booked well in advance, so it is reasonable to assume that their performance will not be affected by the loading of the short-lived ones. Furthermore, long-lived connections between major cities or data centers carry traffic from many users, so it is justifiable for them to have preemptive priority over the short-lived ones. Otherwise, the long-lived connections have to compete with the much more frequent short-lived demands in the same pool of resources, so their blocking probability will increase significantly, which will be detrimental to the many customers they serve. Such priority is also likely to be given to connections that serve the large data bursts generated by the LHC [2] which were the key reason for the design of the LHC network. Then lower priority short-lived connections can share the remaining capacity.

These assumptions justify the use of the so-called quasi-stationary approximation [33]–[35] that are suitable in the cases where changes in system states observed by one type of traffic, due to changes in other traffic type(s), are rare.

To the best of our knowledge, the problem of blocking probability evaluation of circuit switched networks with alternate routing and multiple priorities assigned to short-lived and long-lived calls has not been studied before. In this paper we apply both OPCA and EFPA to evaluate blocking probability of the two traffic types in circuit-switched network with alternate routing. We compare between the results obtained against simulation benchmarks and explain their performance in various different scenarios and parameter ranges. Then we discuss the insight gained into performance tradeoffs as well as design and dimensioning implications.

The remainder of the paper is organized as follows. In Section II, we provide a detailed description of our network model and define notation and basic concepts. Next, in Section III, we provide algorithms based on EFPA and OPCA to evaluate the blocking probability of the model. Then, in Section IV, we provide numerical results over a wide range of parameters for two network topologies (fully meshed and NSF) and discuss performance and design implications. Finally, the paper is concluded in Section V.

II. The Model

We consider a circuit-switched network described by a graph $G(N,E)$ where N is a set of n nodes and E is the set of e arcs. The e arcs correspond to trunks where trunk $i \in E$ carries $C(i)$ links. The N nodes are designated $1,2,3,\ldots,N$, each of them has circuit switching capabilities. We assume that all the nodes have full wavelength conversion capabilities and can switch traffic from any link on one trunk to any other link on an adjacent trunk.

In the context of a core WDM network, a wavelength channel can be viewed as a link. In this case, trunk $i \in E$ is composed of $f(i)$ fibers, each of which supports $w(i)$ wavelengths. Accordingly, trunk $i \in E$ carries $C(i) = f(i)w(i)$ wavelength channels called links. However, if the WDM network is further extended to the metropolitan or local areas, a link can have a sub-wavelength capacity [36]. The assumption above that all nodes can switch traffic from any link on one trunk to any other link on another trunk, in the WDM context implies that we assume that all switches have full wavelength conversion capabilities. In principle, our model can be extended to exclude this assumption, as we can, in our model, split every switch and trunk to multiple "sub-switches" and "sub-trunks" each of which is dedicated to one color wavelength. Then increase the number of allowable alternate routes by a factor of the number of wavelengths. However, this implies a significant increase in the computational complexity of our solutions as the graph that describes the network and the number of alternate routes significantly increase. Katib and Medhi [37] studied by simulations the tradeoff between alternate routing and the number of converters for single priority networks. The focus of this paper is on developing accurate approximations for blocking probability for networks with alternate routing and long-lived and short-lived calls.

Let Γ be a set of directional Source-Destination (SD) pairs. Every directional SD pair $m \in \Gamma$, is defined by its end-nodes. Thus, $m = \{s, d\} \in \Gamma$ represents the directional SD pair s to d. We will distinguish between the term *SD pair* which is an unordered set of the two endpoints: Source and Destination, and the *directional SD pair* that refers to the ordered set: Source-Destination.

The calls are classified according to their priority p ($p = 1, 2$). Long-lived calls ($p = 1$) have preemptive priority over short-lived calls ($p = 2$). For each directional SD pair $m \in \Gamma$, calls of priority p arrive according to a Poisson process with arrival rate $\lambda(m, p)$ [38]. The holding times of calls are assumed exponentially distributed with mean $1/\mu(m, p)$ [39]. We assume that holding times are exponentially distributed for tractability. However, it is well known that in loss systems, blocking probability is highly insensitive to the shape of the holding time distribution and only dependent on the mean value of the holding times. This has been proven for the M/G/K/K system. In Section IV, we have also demonstrated numerically that the blocking probability of our model is also insensitive to the shape of the holding times in Subsection IV-I. Let

$$\rho(m, p) = \frac{\lambda(m, p)}{\mu(m, p)}$$

be the offered traffic (measured in erlangs) for directional SD pair m. We set

$$\rho(p) = \sum_{m \in \Gamma} \rho(m, p).$$

A route between source s and destination d is the sequence of trunks associated with the corresponding arcs in the path between s and d in $G(N, E)$. A path between s and d comprises a sequence of trunks – one on each trunk on the route between s and d.

It is very likely that for a directional SD pair $m \in \Gamma$, there are multiple routes between the source and the destination that do not share a common trunk. Such routes are often called *edge-disjoint paths* or *disjoint paths* [40]–[42]. Edge-disjoint alternate routing is often used to achieve load balancing in optical and other networks [43], [44]. Although using disjoint paths has benefit in reducing blocking probability, non-disjoint paths are often used in practice. Our approximation methods are also applicable to the cases where paths are not disjoint. However, at this stage the strong dependency between trunks in this case which may increase network blocking probability causes our approximation methods to underestimate the blocking probability. Development of algorithms to improve accuracy in the case of non-disjoint paths is still an open problem. In this paper, we only consider disjoint paths in the numerical examples presented in Section IV.

For each $m \in \Gamma$, we designate a route with the least number of hops as the primary path $U(m, 0)$ of the directional SD pair m. If there are multiple routes with the least number of hops, for tractability, the choice is made randomly with equal probabilities. Also, in practice, it may have the advantage of keeping the routing table unchanged. Then considering a new topology where the trunks of the primary path are excluded, the first alternative path for m is chosen to minimize the number of hops in the new topology. Again, a tie is broken randomly. Therefore, all the paths for m, including one primary path and several alternative paths, are edge-disjoint. Let T_m be the maximum number of available alternative paths a directional SD pair m can have based on the network topology.

Furthermore, a maximal number D of overflow attempts to alternate paths are set for calls from all directional SD pairs in Γ. Setting the limit D, implies that a call of the directional SD pair m, can only use

$$T(m) = \min\{T_m, D\}$$

alternative paths. Therefore, before a call is blocked, the procedure continues until all available and allowable $T(m)$ routes are attempted.

It is convenient to maintain the entire set

$$\{U(m, 0), U(m, 1), \ldots, U(m, T(m))\}$$

of alternative routes for the directional SD pair $m \in \Gamma$ in which $U(m, 0)$ is the primary path and $U(m, d)$ is the dth alternate path. This allows for cases where D does not limit the number of usable alternative path.

In our model, the ranking of alternative paths is based on the number of hops and in the case of equality in the number of hops, the rank is chosen randomly. Based on our ranking, if $d(u) > d(v)$ then the number of hops of $U(m, d(u))$ is equal or higher than the number of hops in $U(m, d(v)$. However, in practice, other cost functions (e.g. geographic distance) can be also used for the ranking.

If a request for a call arrives at source node s to the destination node d, and capacity is available on all trunks of the primary path $U(\{s, d\}, 0)$, then this primary path will be used for the transmission of this call.

An arriving call of any type can use any free link on any trunk. When a long-lived call arrives, it can obtain a path on the primary path if no trunk of its primary path has all the links used by long-lived calls. In this case, if it finds all the links in any trunk on its primary path busy, it will preempt a randomly chosen short-lived call. Then the preempted short-lived call will release its resource to the long-lived call and overflow to its next alternate path. If the arriving high priority long-lived call finds that all the links are occupied by higher priority calls on at least one of the trunks of the primary path, it will attempt a route on the first alternate path. In such a case, the long-lived call is said to be *overflowed* from its primary path and to attempt its first alternate path. The same procedure is repeated until the long-lived call exhausts all its D alternate path attempts. Then if it still cannot obtain a path the call will be blocked and cleared of the network.

When a short-lived call arrives, it can obtain a path on the primary path if no trunk of its primary path has all the links used by either long-lived or short-lived calls. Otherwise it will overflow to its first alternate path. Again, the same procedure is repeated. If it is not able to obtain a path in its D alternate path attempts, the call is blocked and cleared of the network.

We realize that call reattempts can affect blocking probability. However, normally our intention is to dimension the network so that the blocking probability is maintained below a certain small value. In this case, the proportion of call reattempts out of the total arriving calls is small and their effect on the blocking probability is negligible.

Considering stability of the network, and recognizing that less resources are used by a call that uses its primary path, priority is given to such calls. To facilitate such priority, a certain number of unoccupied links are reserved for calls attempting their primary path. In particular, if the number of links occupied on trunk j is greater than or equal to a given reservation threshold $RT(j,p)$, the overflowed calls of priority p are not allowed to use that trunk. In the paper, we use trunk reservation to reserve certain number of links to primary path calls to avoid large number of overflowed calls in the network which may cause instability. In all the numerical examples that are presented in the paper (excluding the case of the large Coronet network discussed in Subsection IV-K), both EFPA and OPCA have converged to a unique solution.

III. Blocking Probability Approximations

In this section we describe the approximations we use for blocking probability evaluation for the long-lived and short-lived calls. We use the term 0-call for a call transmitted on its primary path, and the term d-call for a call transmitted on its dth alternate path, for $d = 1,2,\ldots,T(m)$. Accordingly, the term (d,m,p)-call refers to a d-call of priority p, $p = 1,2$ and directional SD pair m, in which for long-lived traffic $p = 1$ and for short-lived traffic $p = 2$. Assume that the arrivals of the (d,m,p)-calls follow a Poisson process with rate $a(d,m,p)$ and the arrivals of the (d,m,p)-calls at trunk $j \in U(m,d)$ also follow a Poisson process with rate $a(d,m,p,j)$. And if j is the first trunk on the path of the (d,m,p)-calls, then $a(d,m,p,j) = a(d,m,p)$. Let $b(d,j,p)$ be the blocking probability for priority p d-calls on trunk $j \in \mathcal{E}$.

The (d,m,p)-calls occur only when $(d-1,m,p)$-calls are blocked. Therefore, we have

$$a(d,m,p) = a(d-1,m,p)\Big(1 - \prod_{j\in U(m,d-1)} (1-b(d-1,j,p))\Big) \tag{1}$$

and $a(0,m,p) = \rho(m,p)$. For a particular trunk along the path $j \in U(m,d)$, we have

$$a(d,m,p,j) = a(d,m,p)\frac{\prod_{i\in U(m,d)}(1-b(d,i,p))}{1-b(d,j,p)} \tag{2}$$

for $d = 0,1,\ldots,T(m)$. For $d > T(m)$ or $j \notin U(m,d)$, $a(d,m,p,j) = 0$.

Let $a(d,j,p)$ be the total offered load of priority p d-calls, on trunk j. They are related with $a(d,m,p,j)$ by

$$a(d,j,p) = \sum_{m\in\Gamma} a(d,m,p,j). \tag{3}$$

Also, let $\tilde{a}(d,j,p)$ be the total offered load of priority p calls that include 0-calls, 1-calls, 2-calls $\ldots$ d-calls, on trunk j. The variables $\tilde{a}(d,j,p)$ and $a(d,j,p)$ are related by

$$\tilde{a}(d,j,p) = \sum_{i=0}^{d} a(i,j,p). \tag{4}$$

Since long-lived traffic has preemptive priority over short-lived traffic, the blocking probability of higher priority long-lived traffic can be evaluated as if it were alone in the network. In other words, it is sufficient to consider a network with a single class of traffic and apply EFPA and OPCA directly to it for the estimation of blocking probability of long-lived traffic. For lower priority short-lived traffic, the available capacity is the leftover of long-lived carried traffic and therefore the blocking probability of short-lived calls in a trunk is dependent on the number of links occupied by long-lived calls in the trunk. To evaluate the blocking probability of the short-lived traffic, we use quasi-stationary approximation in both EFPA and OPCA and calculate the conditional blocking probability of short-lived traffic for each state of long-lived traffic occupancy and then compute the weighted average of these probabilities using the stationary distribution of the long-lived traffic link occupancy.

TABLE I. Summary of Notation

d-call	a call transmitted on its dth alternate path
(d,m,p)-call	d-call of priority p $(p = 1,2)$ and directional SD pair m
$\rho(m,p)$	offered load of SD pair m and priority p
$a(d,m,p)$	offered load of (d,m,p)-calls
$a(d,m,p,j)$	offered load of (d,m,p)-calls on trunk j
$a(d,j,p)$	total offered load of priority p d-calls on trunk j
$\tilde{a}(d,j,p)$	total offered load of priority p up to d-calls on trunk j
$b(d,j,p)$	blocking probability for priority p d-calls on trunk j
$B(m,p)$	blocking probabilities for priority p traffic from SD pair m
$B(p)$	network blocking probability for priority p traffic

A. EFPA

To obtain the estimation of the network blocking probability for the long-lived traffic by EFPA, we begin by setting initial randomly chosen Uniform[0,1) values to trunk blocking probabilities and the probability that a call is admitted on a trunk is 1 minus the trunk blocking probability. With these initial values, we treat the trunks as if they were independent and the probability that a call is admitted to the primary path is the product of probabilities that a call is admitted to all the individual trunks along the path. The primary path blocking probability is 1 minus the probability that a call is admitted to the path. The carried traffic on the primary path for a particular SD pair is the unblocked proportion of the offered traffic, which is also the carried traffic on each trunk along the path attributed to this SD pair. Accordingly, the offered load of the SD pair to a trunk is obtained by the carried traffic attributed to that SD pair divided by 1 minus the trunk blocking probability.

The traffic overflowed to the first alternative path is the traffic offered to the primary path, multiplied by the primary path blocking probability and, in the same way, traffic offered to other alternative paths can also be obtained. Having obtained the traffic offered to every SD pair that uses the trunk, the total traffic offered to the trunk is their sum. Then, we calculate the steady-state probabilities of each trunk in the network and update the trunk blocking probabilities with the ones obtained based on the state probabilities. All these steps compose one iteration of the fixed-point equations. The iterations will continue until the difference of the trunk blocking probabilities of successive two iterations is less than a given value. Having obtained the trunk blocking probabilities, we can obtain network blocking probability for the long-lived traffic.

Let $q(j,1,i)$ be the steady-state probability of having i links busy by long-lived traffic in trunk j. We evaluate the trunk state probability $q(j,1,i)$, for each j and $i \in \{1,\ldots,C(j)\}$ by

$$q(j,1,i) = \left(a(0,j,1) + \mathbf{1}\{RT(j,1) > i-1\} \sum_{n=1}^{D} a(n,j,1)\right) \times q(j,1,i-1)/i, \quad (5)$$

where $\mathbf{1}\{\}$ is the indicator function and $q(j,1,0)$ is set such that $\sum_{i=0}^{C(j)} q(j,1,i) = 1$ is satisfied. The blocking probability, for the long-lived traffic with d overflows, on trunk j is estimated by

$$b(d,j,1) = \begin{cases} q(j,1,C(j)) & d = 0, \\ \sum_{i=RT(j,1)}^{C(j)} q(j,1,i) & d \geq 1. \end{cases} \quad (6)$$

To evaluate the blocking probability for the lower priority short-lived traffic, we use the quasi-stationary approximation (in the sense of [33]–[35], [45]). Such an approximation is often used when the variations in system state observed by one type of traffic are very rare. In our case, as the holding times of long-lived calls are far longer than those of the short-lived calls, so that short-lived calls only rarely observe changes in their service rate during their holding time and can approximately reach steady-state while the number of long-lived calls remains unchanged. Under such conditions, accurate approximation for the blocking probability for the short-lived calls can be obtained by computing the short-lived traffic blocking probability for each state of the long-lived traffic trunk occupancy and then computing the weighted average of these probabilities using the stationary distribution of the long-lived traffic trunk occupancy. A more detailed description of this approximation for our case follows.

To obtain the network blocking probability approximation for the short-lived traffic by EFPA, the procedure is similar to that for the long-lived traffic. The difference is the steady-state probabilities and the blocking probabilities of the short-lived traffic are conditional on the number of links that are occupied by the long-lived calls on that trunk.

Let $b(d,j,k,2)$ be the blocking probabilities for the short-lived calls with d overflows for each trunk j when there are $k \in \{0,\ldots,C(j)\}$ links free in trunk j.

First, set: $b(d,j,0,2) = 1$, for each d, and also for $d = 1,2,\ldots,D$, when $k \leq C(j) - RT(j,2)$, set $b(d,j,k,2) = 1$.

Next, to evaluate other $b(d,j,k,p)$ values, we evaluate the trunk state probability $q(j,k,2,i)$ for each trunk j and each state $i \in \{1,\ldots,k\}$ using

$$q(j,k,2,i) = \left(a(0,j,2) + \mathbf{1}\{R(j,k) > i-1\} \sum_{n=1}^{D} a(n,j,2)\right) \times q(j,k,2,i-1)/i, \quad (7)$$

where $R(j,k) = RT(j,2) + k - C(j)$ and $q(j,k,2,0)$ is set such that $\sum_{i=0}^{k} q(j,k,2,i) = 1$ is satisfied.

Then we obtain

$$b(d,j,k,2) = \begin{cases} 1 & k = 0 \text{ or } R(j,k) \leq 0, \\ q(j,k,2,k) & d = 0 \text{ and } R(j,k) > 0, \\ \sum_{i=R(j,k)}^{k} q(j,k,2,i) & d \geq 1 \text{ and } R(j,k) > 0. \end{cases} \quad (8)$$

The blocking probability, for the short-lived traffic with d overflows, on trunk j is estimated by

$$b(d,j,2) = \sum_{i=0}^{C(j)} q(j,1,i) \times b(d,j,C(j)-i,2). \quad (9)$$

Equations (1) – (9) form a set of fixed-point equations which can be solved by successive substitutions.

Having obtained the results of the fixed-point equations, we calculate the blocking probabilities for the long-lived traffic ($p = 1$) and short-lived traffic ($p = 2$) from SD pair m by

$$B(m,p) = 1 - \sum_{d=0}^{D} a(d,m,j,p)(1 - b(d,j,p))/\rho(m,p), \quad (10)$$

where j is the last trunk in the route for the calls of SD pair m that overflow d times. Let $B(p)$ be the network blocking probability for long-lived traffic ($p = 1$) and short-lived traffic ($p = 2$), which is the average of blocking probabilities of all SD pairs, weighted by their offered load.

$$B(p) = \sum_{m \in \Gamma} B(m,p) \times \rho(m,p) / \sum_{m \in \Gamma} \rho(m,p). \quad (11)$$

Algorithm 1 is used to obtain the network blocking probability $B(1)$ for the long-lived traffic.

The relative error is a parameter, set to measure the difference of the substitution results and the iteration will stop when

$$\sum_{j \in \mathcal{E}} |b(d,j,1) - \hat{b}(d,j,1)| < error. \quad (12)$$

In this paper, we set

$$error = 10^{-8}. \quad (13)$$

Algorithm 2 describes the computation of the network blocking probability for the short-lived traffic which is dependent on the state probability $q(j,1,i)$ computed by Algorithm 1.

EFPA is known to introduce two types of errors:

1) **The Poisson error** — EFPA assumes that the traffic offered to any trunk follows a Poisson process whereas

Algorithm 1 Compute $B(1)$ and $q(j,1,i)$ by EFPA

Require: $\rho(m,1)$ for $m \in \Gamma$
  initial: $b(d,j,1) \leftarrow 0$, $\hat{b}(d,j,1) \leftarrow 1$ for $j \in \mathcal{E}$, $d \in \{0,\dots,D\}$
  while $\sum_{d\in\{0,\dots,D\}}\sum_{j\in\mathcal{E}} |b(d,j,1)-\hat{b}(d,j,1)| > error$ **do**
    for $j \in \mathcal{E}$, $d \in \{0,\dots,D\}$, $m \in \Gamma$ **do**
      $\hat{b}(d,j,1) \leftarrow b(d,j,1)$
      compute $a(d,m,1)$ in Eq. (1)
      compute $a(d,m,j,1)$ in Eq. (2)
      compute $a(d,j,1)$ in Eq. (3)
      for $i \in \{1,\dots,C(j)\}$ **do**
        compute $q(j,1,i)$ in Eq. (5)
      end for
      compute $b(d,j,1)$ in Eq. (6)
    end for
  end while
  for $m \in \Gamma$ **do**
    compute $B(m,1)$ in Eq. (13)
  end for
  compute $B(1)$ in Eq. (11)

Algorithm 2 Compute $B(2)$ by EFPA

Require: $\rho(m,1)$, $\rho(m,2)$ for $m \in \Gamma$, $q(j,1,i)$ for $j \in \mathcal{E}$ and $i \in \{1,\dots,C(j)\}$
  initial: $b(d,j,2) \leftarrow 0$, $\hat{b}(d,j,2) \leftarrow 1$ for $j \in \mathcal{E}$, $d \in \{0,\dots,D\}$
  while $\sum_{d\in\{0,\dots,D\}}\sum_{j\in\mathcal{E}} |b(d,j,2)-\hat{b}(d,j,2)| > error$ **do**
    for $j \in \mathcal{E}$, $d \in \{0,\dots,D\}$, $m \in \Gamma$ **do**
      $\hat{b}(d,j,2) \leftarrow b(d,j,2)$
      compute $a(d,m,2)$ in Eq. (1)
      compute $a(d,m,j,2)$ in Eq. (2)
      compute $a(d,j,2)$ in Eq. (3)
      for $k \in \{1,\dots,C(j)\}$, $i \in \{1,\dots,k\}$ **do**
        compute $q(j,k,2,i)$ in Eq. (7)
        compute $b(d,j,k,2)$ in Eq. (8)
      end for
      compute $b(d,j,2)$ in Eq. (9)
    end for
  end while
  for $m \in \Gamma$ **do**
    compute $B(m,2)$ in Eq. (13)
  end for
  compute $B(2)$ in Eq. (11)

in fact the traffic offered by an overflow call is known to have higher variance than a Poisson process [13], when traffic offered to a sequence of trunks on a path may be smoothed out when offered to one trunk due to blocking in another trunk.

2) **The independence error** — EFPA assumes that trunks are mutually independent, whereas they are in fact statistically dependent.

As both the Poisson and the independence errors are related to two effects that are characteristics of circuit switched networks, namely, an effect associated with overflows and the effect associated with the fact that a call requires a multi-hop path to be established. We will henceforth call errors of EFPA caused by these two effects *overflow errors* and *path errors*. While overflow errors cause underestimation of blocking probability (ignoring high variance of overflow traffic and dependence), path error overestimates blocking probability because it ignores the effect of traffic smoothing, and the positive correlation of trunk occupancy along the path that increases the probability to admit calls. These relationships between the errors and their effects are shown in Table II.

TABLE II. EFPA Errors

	Overflow Error	**Path Error**
Poisson Error	underestimate	overestimate
Independence Error	underestimate	overestimate

B. OPCA

OPCA works by using a hierarchical surrogate second system and estimating the blocking probability in the second system by an EFPA-like algorithm. The surrogate system is defined by regarding an overflow loss network as if it were operating under a preemptive priority regime where each call is classified according to the number of times it has overflowed and *junior calls* (calls that experienced less overflows) are given priority. By giving priority to junior calls, the seniors calls that have more "information" about busy paths, namely, paths where at least one trunk is busy (all the links in that trunk is busy), are preempted and overflowed and these senior calls will only attempt alternate paths which they did not visit before. In this way, the surrogate system operates as a hierarchical network where the traffic is strictly prioritized and layered according to how many times it overflows. We remind the reader that the prioritization introduced in the surrogate system is artificially introduced to obtain a more accurate approximation and it is not a feature of the real network.

Despite the fact that the surrogate system may be different from the real system we aim to analyze, the application of EFPA to the surrogate system can, in many cases, provide a better blocking probability approximation for the original problem than the application of EFPA to the original problem due to the following reasons.

1) OPCA avoids the adverse effects on accuracy of mutual overflow.
2) OPCA increases the proportion of the 0-calls in the system and reduces the overflowed traffic. Since 0-calls do not violate the Poisson and independence assumptions, increasing the proportion can reduce the Poisson and independence errors.
3) Having more primary traffic reduces also the independence errors.

The procedure of the blocking probability calculation by OPCA is similar to that of EFPA and the difference between the two is due to the preemptive priority of the surrogate model of OPCA. In OPCA, we first solve the fixed-point equations considering only traffic offered to the primary path that has not overflowed yet, and obtain the network blocking probability for it. Then, we calculate the total traffic comprises the traffic that has not overflowed and the traffic that has overflowed once. We again solve the fixed-point equations and obtain the blocked portion of the total traffic. Subtracting the traffic that was blocked once in the primary path from the total blocked traffic, we obtain the traffic that is blocked twice, which gives the blocking probability for the traffic that has overflowed once. Blocking probabilities of traffic that has overflowed more than once are obtained recursively in a similar way.

In the following we provide detailed information on how to apply OPCA to the present problem of approximating blocking probability of circuit switched networks for the long-lived and short-lived traffic.

We begin by evaluating the trunk state probability $t(d,j,p,i)$ of long-lived traffic ($p=1$) for each trunk j, for d deflections and each state $i \in \{1,\ldots,C(j)\}$ using

$$t(d,j,1,i) = \left(a(0,j,1) + \mathbf{1}\{RT(j,1) > i-1\} \sum_{n=1}^{d} a(n,j,1) \right) \times t(d,j,1,i-1)/i, \quad (14)$$

where $t(d,j,1,0)$ is set to satisfy $\sum_{i=0}^{C(j)} t(d,j,1,i) = 1$.

The average blocking probability $\bar{b}(d,j,1)$ on trunk j, for the long-lived calls with up to and including d overflows, is estimated by

$$\bar{b}(d,j,1) = \frac{\sum_{n=1}^{d} \left(a(n,j,1) \sum_{i=RT(j,1)}^{C(j)} t(d,j,1,i) \right)}{\tilde{a}(d,j,1)} + \frac{a(0,j,1) t(d,j,1,C(j))}{\tilde{a}(d,j,1)} \quad (15)$$

and $\bar{b}(0,j,1)$ is estimated using the Erlang-B formula, i.e. $\bar{b}(0,j,1) = E\left(a(0,j,1),C(j)\right)$. The blocking probability for the long-lived traffic, for d-overflows calls, on trunk j is estimated by

$$b(d,j,1) = \begin{cases} \bar{b}(0,j,1) & d=0, \\ \frac{\bar{b}(d,j,1)\tilde{a}(d,j,1) - \bar{b}(d-1,j,1)\tilde{a}(d-1,j,1)}{a(d,j,1)} & 1 \le d \le D. \end{cases} \quad (16)$$

Note that the blocking probability for the unoverflowed calls is calculated using the Erlang-B formula. Having obtained the trunk blocking probability $b(d,j,1)$, the network blocking probability can be computed by equations (13) and (11). Algorithm 3 is used to compute the network blocking probability for the long-lived traffic.

After calculating all the blocking probabilities for different layers for the long-lived traffic, we move on to calculate the blocking probabilities for the short-lived traffic.

Let $h(j,i)$ for $i \in \{1,\ldots,C(j)\}$ and each trunk j be the probability that there are i number of long-lived calls in trunk j. So that

$$h(j,i) = \left(a(0,j,1) + \mathbf{1}\{RT(j,1) > i-1\} \sum_{n=1}^{D} a(n,j,1) \right) \times h(j,i-1)/i \quad (17)$$

where $h(j,0)$ is set such that $\sum_{i=0}^{C(j)} h(j,i) = 1$ is satisfied.

Let $\bar{b}(d,j,k,2)$ be the blocking probabilities for the short-lived calls with d overflows for each trunk j when there are $k \in \{0,\ldots,C(j)\}$ links free in the trunk j. For $d = 0,1,\ldots,D$, $\bar{b}(d,j,0,2) = 1$. For $d = 1,2,\ldots,D$, when $R(j,k) = k - C(j) + RT(j,2) \le 0$, $\bar{b}(d,j,k,2) = 1$. To evaluate other $\bar{b}(d,j,k,p)$, we evaluate the trunk state probability $t(j,k,d,2,i)$ for each trunk j, each state $i \in \{0,1,\ldots,k\}$ and d using

$$t(d,j,k,2,i) = \left(a(0,j,2) + \mathbf{1}\{R(j,k) > i-1\} \sum_{n=1}^{d} a(n,j,2) \right) \times t(d,j,k,2,i-1)/i, \quad (18)$$

where $t(d,j,k,2,0)$ is set such that $\sum_{i=0}^{k} t(d,j,k,2,i) = 1$ is satisfied, and $t(0,j,k,2,k)$ is estimated using the Erlang-B formula, i.e. $t(0,j,k,2,k) = E\left(a(0,j,2),k\right)$.

Then we obtain

$$\bar{b}(d,j,k,2) = \begin{cases} t(0,j,k,2,k) & d=0, \\ \sum_{i=R(j,k)}^{k} t(d,j,i,2,i) & d \ge 1 \text{ and } R(j,k) > 0, \\ 1 & \text{otherwise.} \end{cases} \quad (19)$$

The averaged blocking probability, for the short-lived traffic with d overflows, on trunk j is estimated by

$$\bar{b}(d,j,2) = \sum_{i=0}^{C(j)} h(j,i) \times \bar{b}(d,j,C(j)-i,2). \quad (20)$$

The blocking probability for the short-lived d-calls, on trunk j is estimated by

$$b(d,j,2) = \begin{cases} \bar{b}(0,j,2) & d=0, \\ \frac{\bar{b}(d,j,2)\tilde{a}(d,j,2) - \bar{b}(d-1,j,2)\tilde{a}(d-1,j,2)}{a(d,j,2)} & 1 \le d \le D. \end{cases} \quad (21)$$

Then the network blocking probability can be computed by equations (13) and (11). Algorithm 4 is used to compute the network blocking probability $B(2)$ for the short-lived traffic.

IV. Numerical Results

We begin this section by comparing the performance of OPCA to the performance of EFPA in approximating the blocking probability for the long-lived and short-lived traffic using simulations. The comparison is performed for a 6-node fully-meshed network and the 13-node National Science Foundation (NSF) network.

Then, we extend our comparison, to consider a range of scenarios and parameter values for each scenario. In all cases considered, we also provide intuitive explanation to the discrepancies between the two approximations and simulation results for the blocking probability as it varies according to the various effects. In particular, we consider traffic effects such as offered load, and the difference between the two types of traffic, in terms of the offered load and mean holding times. The latter is especially important to observe the accuracy and sensitivity of quasi-stationary approximation.

Finally, we consider design factors such as: the number of links per trunk, the maximal allowable number of alternate paths, and the effect of trunk reservation. We also discuss the

Algorithm 3 Compute $B(1)$ and $h(j,i)$ by OPCA

Require: $\rho(m,1)$ for $m \in \Gamma$
  for $d \in \{0,\dots,D\}$ **do**
    initial: $b(d,j,1) \leftarrow 0$, $\hat{b}(d,j,1) \leftarrow 1$ for $j \in \mathcal{E}$
    while $\sum_{j\in\mathcal{E}} |b(d,j,1) - \hat{b}(d,j,1)| > error$ **do**
      for $j \in \mathcal{E}$, $m \in \Gamma$ **do**
        $\hat{b}(d,j,1) \leftarrow b(d,j,1)$
        compute $a(d,m,1)$ in Eq. (1)
        compute $a(d,m,j,1)$ in Eq. (2)
        compute $a(d,j,1)$ in Eq. (3)
        compute $\tilde{a}(d,j,1)$ in Eq. (4)
        for $i \in \{1,\dots,C(j)\}$ **do**
          compute $t(d,j,1,i)$ in Eq. (14)
        end for
        compute $\bar{b}(d,j,1)$ in Eq. (15)
        compute $b(d,j,1)$ in Eq. (16)
      end for
    end while
  end for
  for $m \in \Gamma$ **do**
    compute $B(m,1)$ in Eq. (13)
  end for
  compute $B(1)$ in Eq. (11)
  for $j \in \mathcal{E}$, $i \in \{1,\dots,C(j)\}$ **do**
    compute $h(j,i)$ in Eq. (17)
  end for

Algorithm 4 Compute $B(2)$ by OPCA

Require: $\rho(m,1)$, $\rho(m,2)$ for $m \in \Gamma$, $h(j,i)$ for $j \in \mathcal{E}$ and $i \in \{1,\dots,C(j)\}$
  for $d \in \{0,\dots,D\}$ **do**
    initial: $b(d,j,2) \leftarrow 0$, $\hat{b}(d,j,2) \leftarrow 1$ for $j \in \mathcal{E}$
    while $\sum_{j\in\mathcal{E}} |b(d,j,2) - \hat{b}(d,j,2)| > error$ **do**
      for $j \in \mathcal{E}$, $m \in \Gamma$ **do**
        $\hat{b}(d,j,2) \leftarrow b(d,j,2)$
        compute $a(d,m,2)$ in Eq. (1)
        compute $a(d,m,j,2)$ in Eq. (2)
        compute $a(d,j,2)$ in Eq. (3)
        compute $\tilde{a}(d,j,2)$ in Eq. (4)
        for $k \in \{1,\dots,C(j)\}$, $i \in \{1,\dots,k\}$ **do**
          compute $t(d,j,k,2,i)$ in Eq. (18)
          compute $\bar{b}(d,j,k,2)$ in Eq. (19)
        end for
        compute $\bar{b}(d,j,2)$ in Eq. (20)
        compute $b(d,j,2)$ in Eq. (21)
      end for
    end while
  end for
  for $m \in \Gamma$ **do**
    compute $B(m,2)$ in Eq. (13)
  end for
  compute $B(2)$ in Eq. (11)

robustness of the approximations to the shape of the holding time distribution.

In all scenarios considered, the arrival process of calls for each directional SD pair and each type of traffic follows a Poisson process and the total traffic offered to each directional SD pair is equal to T. The variable T is our measure of traffic load for all cases in both topologies. The shortest path is set to be the primary route for each SD pair, and the alternate routes are pre-assigned ordered by their length. For those routes with the same lengths, the order is chosen randomly and remains unchanged afterwards.

A. Default parameter setting

In many experiments we repeatedly use the same set of parameters with possibly small variations. It is convenient to present them once in this section and through the section only point out the deviations from this default set. This set of parameters will henceforth be referred to as *Default parameter setting* and is described as follows.

For both 6-node fully meshed network and NSF network, both long-lived and short-lived calls arrive according to a Poisson process and the ratio of offered long-lived traffic to that of short-lived traffic is $1:1$. The holding time of both long-lived and short-lived calls are exponentially distributed and the mean holding time of long-lived traffic is 200 times higher than that of the short-lived traffic. The total number of links per trunk is 20. The threshold of long-lived traffic is 16 (80%) and the threshold of the short-lived traffic is 18 (90%). The maximal allowable number of alternate paths are set 4 for both long-lived and short-lived calls in 6-node fully meshed network and 2 in NSF network, respectively.

We have chosen to present the results, for each network, for the long-lived and short-lived calls in two separate figures because they are different by several order of magnitude, and if presented on the same figure as a function of traffic load, for reasonable traffic loads that give acceptable blocking probability to short-lived traffic, in many cases, the blocking probability for the long-lived traffic will be too low for accurate evaluation by simulation.

B. Blocking probability for the long-lived traffic

We evaluate here the blocking probability for the long-lived traffic by EFPA, OPCA and simulations. Since long-lived traffic has preemptive priority over short-lived traffic, the blocking probability for the long-lived traffic can be evaluated as if it were alone in the network. In other words, it is sufficient to consider a network with a single class of traffic. We note that the case of a single class of traffic was considered in [27]. Here we add examples, intuitive explanations and interpretation of the results.

At first, we consider a 6-node fully meshed network model. In such a network, there are in total 15 different SD pairs (or equivalently 30 directional SD pairs). As the offered traffic for each directional SD pair is T, so the total offered traffic in the network is $30T$.

In Fig. 1 (a), we present results for the blocking probabilities obtained by OPCA, EFPA and simulations for long-lived

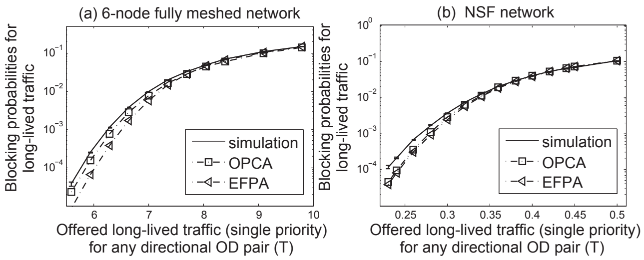


Fig. 1. Blocking probabilities for the long-lived traffic (single priority) versus T in (a) a 6-node fully-meshed network and (b) NSF network.

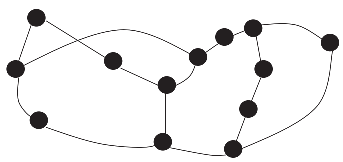


Fig. 2. NSF network topology, each solid line represents a bi-directional trunk between two nodes.

traffic (single priority). We observe in the figure that EFPA and OPCA tend to underestimate blocking probability when the offered load is low. This is due to the fact that in a fully meshed network with low traffic load, and therefore less overflows, long paths will be very rare. Accordingly, overflow error will dominate path errors causing underestimation of blocking probability.

Since the surrogate model of OPCA gives preemptive priority to new calls, and therefore allocates more resources to the primary path traffic, this surrogate model has less overflow traffic than the original model approximated by EFPA, leading to less overflow error, less Poisson error, and therefore less underestimation of blocking probability for the long-lived traffic (single priority). Recall that OPCA is based on approximating the blocking probability of the surrogate model treating each link independently and assuming it is loaded by Poisson arrivals.

Furthermore, we observe that as the traffic load increases, the underestimation for both EFPA and OPCA of the blocking probability is reduced. This is consistent with the fact that in high load overflow probability increases, leading to overflow path length growth, and therefore path error increase. As observed, the path error in the cases of high traffic load may cancel out the overflow error to improve the approximation.

Next, we consider an NSF network with 13 nodes and 16 bidirectional trunks. As the number of SD pairs is $(13 \times 12/2) = 78$, the total offered traffic in this case is $156T$. The topology of the NSF network is shown in Fig. 2. The results for the blocking probabilities for the long-lived traffic as a function of T is shown in Fig. 1 (b). We obtain results that have similar behavior to those obtained for the 6-node fully meshed network case in the sense that OPCA outperforms EFPA. But since the number of alternate paths in NSF network is much less than that in 6-node fully meshed network, in NSF network, OPCA outperforms EFPA less than in the case of the 6-node fully meshed network.

C. Blocking probability for the short-lived traffic

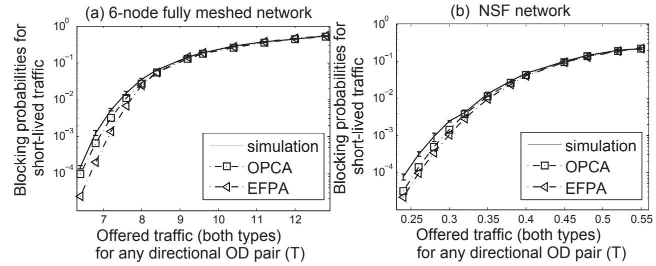


Fig. 3. Blocking probabilities the for short-lived traffic in (a) a 6-node fully-meshed network and (b) NSF network. The ratio of the offered long-lived traffic to offered short-lived traffic is 1:1.

Figures 3 (a) and (b) present the blocking probabilities results we obtained by OPCA, EFPA and simulations for short-lived traffic, as a function of T, in the same 6-node fully meshed and NSF networks described above, but now we consider the two classes of traffic using the networks so T is now the total offered traffic of both long-lived and short-lived calls for each directional SD pair. The parameters are set as in the *Default parameter setting* in IV-A.

As expected, comparing Figs. 3 (a) and (b) with their long-lived traffic (single priority) counterparts Figs. 1 (a) and (b), respectively, we observe significantly lower blocking probability (by several orders of magnitude) for the long-lived traffic than for short-lived traffic. Specifically, observe the blocking probability in Fig 3 (a) for load of $T = 12$ versus the equivalent load of $T = 6$ in Fig. 1 (a). Also, for the NSF network, we compare the blocking probability in Fig. 3 (b) for load of $T = 0.5$ to that of Fig. 1 (b) for load of $T = 0.25$.

Furthermore, we observe that OPCA provides better approximations for both long-lived traffic and short-lived traffic in both networks than EFPA.

If the threshold of long-lived traffic is larger than or equal to that of short-lived traffic, since long-lived calls can preempt short-lived calls and not vice versa, the set of states in which long-lived calls are blocked is a strict subset of the set of states in which short-lived calls are blocked, and since both processes arrive in accordance with Poisson processes, blocking probability for the long-lived traffic must be lower than that of short-lived traffic even if the offered long-lived traffic is much larger than that of short-lived traffic. In general, alternate routing further benefits long-lived traffic, because longer alternate routes of long-lived traffic, that use more network resources per bit than primary path traffic, have a more detrimental effect on short-lived traffic than on long-lived traffic.

To provide some protection to short-lived traffic, we set the threshold of short-lived traffic larger than that of long-lived traffic so that if the offered short-lived traffic is much smaller than that of long-lived traffic, short-lived traffic may have a smaller blocking probability than long-lived traffic. Notice that this threshold difference does not affect the "right" of the long-lived traffic to behave as if it is alone in the system according to their own threshold limitation on overflowed calls. This limitation also provides some protection to short-lived traffic from overflowed long-lived traffic using long alternative path inefficiently.

However, in the present case, the offered short-lived traffic is the same as that of long-lived traffic, thus the blocking probability for the short-lived traffic is larger than that for the long-lived traffic.

The preemptive property of long-lived traffic affects the approximations of short-lived traffic in two distinct ways:

1) The capacity available to short-lived traffic is the leftover of long-lived carried traffic and short-lived traffic can be preempted by long-lived calls. Therefore, short-lived calls may be forced to take longer alternate routes, so the proportion of the overflow traffic in the total short-lived traffic is higher than for long-lived traffic, leading to higher overflow error and path error. For low traffic, since the overflow error is dominant, this higher proportion of overflow traffic in the total short-lived traffic will cause further underestimation of the blocking probability for the short-lived traffic for both EFPA and OPCA. On the other hand, in a high loading scenario, since the path error is dominant in high loading, this higher proportion will cause further overestimation of the blocking probability for the short-lived traffic for both EFPA and OPCA.
2) Long-lived carried traffic, which can be viewed as long-lived background traffic for the short-lived traffic, exhibits dependencies among various trunks. More specifically, the congestion of long-lived traffic on one trunk is likely to cause congestion of long-lived traffic on other trunks. This congestion dependence of long-lived traffic (background traffic) on trunks in turn causes dependence in capacity limitation for short-lived traffic among different links which leads to congestion dependence of short-lived traffic. However, in the quasi-stationary approach, long-lived carried traffic are assumed to be independent. This introduces another kind of dependence error, which does not occur for long-lived traffic, that causes a further underestimation of the blocking probability for the short-lived traffic for both OPCA and EFPA. This effect causes further underestimation for low traffic for both EFPA and OPCA. As the total offered load increases and this effect tends to underestimate the blocking probability obtained by EFPA and OPCA, which will cancel out the overestimation due to the first effect and lead to accurate predictions by both EFPA and OPCA.

D. The effect of the ratio between the offered long-lived traffic and offered short-lived traffic

Figs. 3 (a) and 4 (a)-(d) show the blocking probabilities for the short-lived traffic when the ratios of the offered long-lived traffic to offered short-lived traffic are 1:1, 1:2, 1:5, 2:1 and 5:1, respectively while all the other parameters are kept the same as in the *Default parameter setting* in IV-A. As long-lived traffic is not affected by short-lived traffic, we only

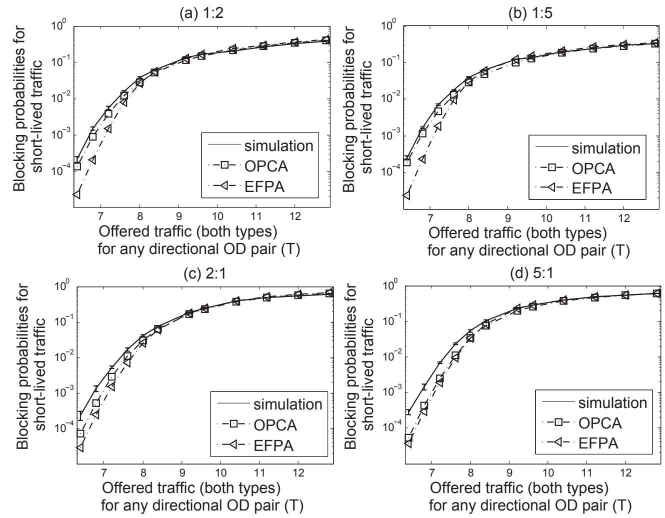


Fig. 4. Blocking probabilities for the short-lived traffic in a 6-node fully-meshed network. The ratio of the offered long-lived traffic to offered short-lived traffic is (a) 1:2, (b) 1:5, (c) 2:1, and (d) 5:1.

consider here the accuracy of OPCA and EFPA for short-lived traffic.

Several observations emerge from Figs. 3 (a) and 4 (a)-(d):

1) OPCA is generally more accurate than EFPA.
2) The accuracy of OPCA in predicting blocking probability for the short-lived traffic decreases when the proportion of long-lived traffic increases.
3) EFPA accuracy is not significantly affected by the ratio of the two traffic types.
4) Both EFPA and OPCA accuracy increases with increased traffic load.

Observations 1 and 4 are consistent with what we have observed in the case when the proportion was 1:1, and the explanations above are applicable. To explain observations 2 and 3, recall that EFPA under light load suffers from the dependency error. This fact is invariant to the proportion between the offered long-lived traffic and offered short-lived traffic. However, OPCA is able to reduce the dependency error of short-lived traffic (notice that the quasi-stationary approximation assumes independence between the long-lived and short-lived traffic for both OPCA and EFPA). Therefore, if the short-lived traffic is reduced, the ability of OPCA to neutralize the dependence effects is also reduced.

Since OPCA has better performance in the cases of low short-lived loading and EFPA has better performance in the cases of high loading, $\max\{OPCA, EFPA\}$ is the best approximation. In [28], we obtained a similar conclusion that $\max\{OPCA, EFPA\}$ is the best approximation for an optical burst switched network.

For NSF network with 13 nodes and 16 trunks, the results for the blocking probabilities for the short-lived traffic are shown in Figs. 3 (b), 5 (a)-(d) when the ratios of the offered long-lived traffic to offered short-lived traffic are 1:1, 1:2, 1:5, 2:1 and 5:1, respectively while all the other parameters are kept the same as in the *Default parameter setting* in IV-A. We obtain results that have similar behavior to those obtained for the 6-node fully meshed network case in the sense that OPCA slightly outperforms EFPA in the cases of low loading.

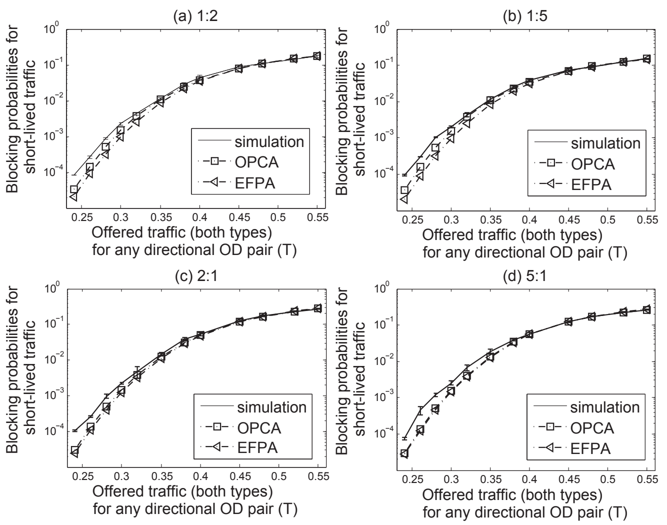


Fig. 5. Blocking probabilities for the short-lived traffic in NSF network. The ratio of the offered long-lived traffic to offered short-lived traffic is (a) 1:2, (b) 1:5, (c) 2:1, and (d) 5:1.

E. The effect of the number of links on each trunk

To examine the effect of the number of links (wavelength channels) on each trunk on blocking probability and on the accuracy of EFPA and OPCA, we increase now the number of links on each trunk to 50 in the 6-node fully meshed network we consider above. In particular, we consider a scenario where the thresholds for long-lived traffic and short-lived traffic are 40 (80%) and 45 (90%), respectively while all the other parameters are kept the same as in the *Default parameter setting* in IV-A.

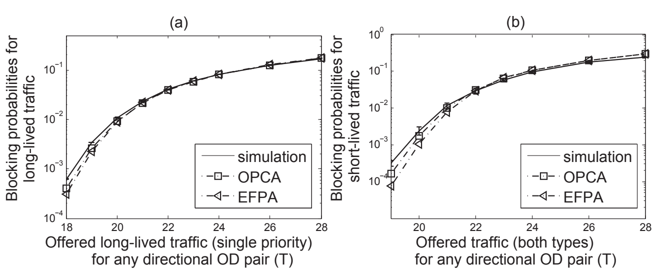


Fig. 6. Blocking probabilities for (a) long-lived traffic (single priority) and (b) short-lived traffic in a 6-node fully-meshed network with 50 links each trunk.

In Fig. 6 (a) we provide the results obtained for the blocking probabilities for the long-lived traffic. We can observe that the accuracy of EFPA is improved comparing to the case of 20 links per trunk shown in Fig. 1 (a). The improvement in accuracy is achieved because of the following reasons.

1) When the number of links on each trunk increases, the variance of the overflow traffic decreases, leading to a lower Poisson error.
2) The increase of the number of links on each trunk also reduces the proportion of overflowed traffic and therefore reduces the overflow error, which also increase the accuracy of EFPA.

We also observed that OPCA, in general, is superior to EFPA, so it is sandwiched between EFPA and the simulation results.

Fig. 6 (b) shows the blocking probability for the short-lived traffic. We observe again that the accuracy of EFPA is improved comparing to the case of 20 links per trunk shown in Fig. 3 (a) and OPCA is generally sandwiched between EFPA and the simulation results. The reasons for the improvement in EFPA results are the same as those discussed in the case of the long-lived traffic blocking probability evaluation.

Notice also that for the blocking probability evaluation for both long-lived and short-lived traffic, OPCA still outperforms EFPA in the case of 50 links per trunk. This together with the improved accuracy as the number of links on each trunk increases from 20 to 50, provide some evidence that OPCA can be accurate as the network capacity scales upwards and performs even better than for networks with lower capacity.

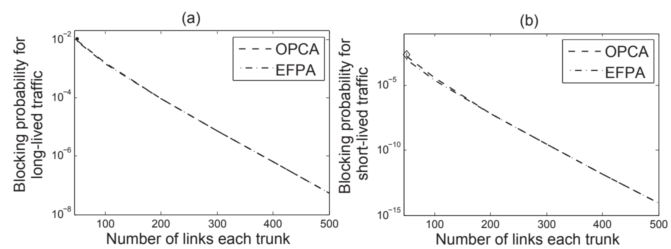


Fig. 7. Blocking probabilities for (a) long-lived traffic (single priority) and (b) short-lived traffic in a 6-node fully-meshed network versus number of links on each trunk.

We further increase the number of links on each trunk in 6-node fully meshed network and in Fig. 7 (a), we provide the results obtained by EFPA and OPCA for the blocking probabilities for the long-lived traffic versus number of links on each trunk. The threshold of long-lived traffic is kept at 80% and the maximum allowable alternate paths is kept at 4. Let C be the number of links on each trunk. The offered long-lived traffic (Erlangs) is $0.4C$ for each directional SD pair, so the total traffic per SD pair in both directions is $0.8C$. As increasing the traffic and the number of links on each trunk at the same rate, will decrease the proportion of overflowed traffic, thus in large capacity networks, only negligible traffic is overflowed, so that almost all end-to-end paths in our fully meshed networks are single link (the network approximately turns into a fixed-routing network), this will mean that approximately all links will be 80% utilized. In fixed-routing large capacity networks, EFPA is accurate based on Kelly [9]. Since there is only negligible overflow in large capacity networks, as mentioned in the Introduction, we have the condition under which model is based on fixed routing, i.e. no overflow is allowed, OPCA is reduced to EFPA. This is consistent with the results presented in Fig. 7 (a).

Fig. 7 (b) shows the blocking probability for the short-lived traffic versus the number of links per trunk. The thresholds of long-lived traffic and short-lived traffic are set as before at

80% and 90%, respectively while all the other parameters are kept the same as in the *Default parameter setting* in IV-A. The offered traffic load for are both long-lived and short-lived calls for each directional SD pair are $0.2C$. Although, in general, we observe similar results for short-lived traffic blocking probability to those obtained for long-lived traffic, we notice a slight difference between the EFPA and OPCA results for the cases of 50 and 100 links per trunk, which indicates that some overflows at these level occur which lead to conditions where EFPA underestimate the blocking probability slightly more than OPCA.

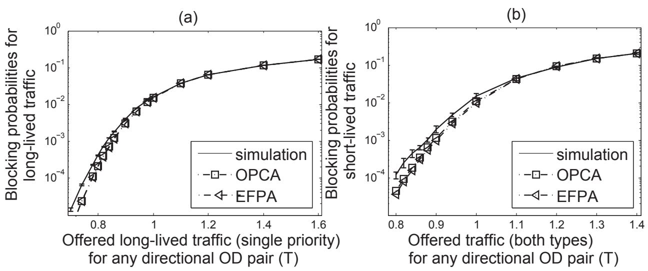


Fig. 8. Blocking probabilities for (a) long-lived traffic (single priority) and (b) short-lived traffic in NSF network with 50 links each trunk.

We also increase the number of links per trunk for the NSF network to 50. In particular, we consider a scenario where the thresholds for long-lived traffic and short-lived traffic are 40 (80%) and 45 (90%), respectively while all the other parameters are kept the same as in the *Default parameter setting* in IV-A. In Fig. 8 (a) we provide the results obtained for the blocking probabilities for the long-lived traffic. And the blocking probabilities for the short-lived traffic is shown in Fig. 8 (b). The trends and behavior of the results presented for the case of NSF network are consistent with the results provided for the 6-node fully meshed network case.

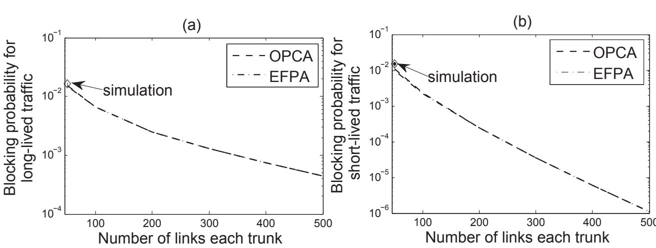


Fig. 9. Blocking probabilities for (a) long-lived traffic (single priority) and (b) short-lived traffic in NSF network versus number of links per trunk.

We also increase the number of links on each trunk in the NSF network, and in Fig. 9 (a), we again provide the results obtained by EFPA and OPCA for the blocking probabilities for the long-lived traffic versus number of links per trunk. These results are based on having the offered long-lived traffic (Erlangs) to be $0.02C$ for each directional SD pair, the threshold and the maximum allowable alternate paths are kept the same as in the *Default parameter setting* in IV-A.

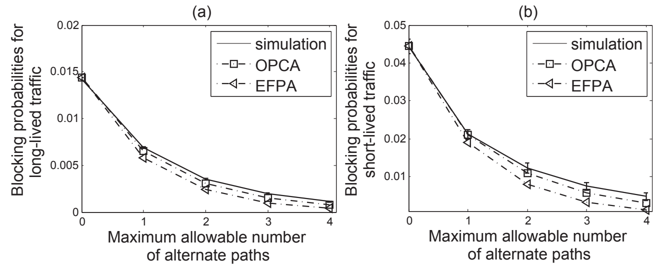


Fig. 10. blocking probability for (a) long-lived traffic (single priority) and (b) short-lived traffic for a 6-node fully-meshed network

Then we consider the NSF network with long-lived and short-lived traffic where the thresholds of long-lived traffic and short-lived traffic are set as before at 80% and 90%, respectively while all the other parameters are kept the same as in the *Default parameter setting* in IV-A. Both the offered long-lived traffic and offered short-lived traffic for each directional SD pair are $0.01C$. In Fig. 9 (b) we present the blocking probabilities for the short-lived traffic versus number of links per trunk obtained by EFPA and OPCA.

Comparing Figs. 9 (a), 7 (a), and Figs. 9 (b), 7 (b), the results for NSF network are generally consistent with those for 6-node fully meshed network. The results based on OPCA and EFPA are almost identical. The small discrepancy observed for short-lived traffic in the case of fully meshed network does not exist in the present case because under NSF the allowable number of alternate routes is smaller so this scenario is closer to fixed-routing network than the fully-meshed alternate-routing network, in which case we already know that if large capacity is available on trunks both OPCA and EFPA are very accurate.

F. *The effect of maximal allowable number of alternate paths*

Here we examine how the blocking probabilities is affected by the maximal allowable number of alternate paths. The maximal allowable number of alternate paths D limits how many times traffic can overflow. Traffic that already overflowed D times is not allowed to overflow again and will be blocked and cleared of the network. For single class networks with light traffic, increasing D appropriately means more opportunities to overflow and benefits the system by reducing the blocking probability. However, when the offered load in the network is high, increasing D may not reduce the blocking probability because of the inefficiency associated with having the average number of links used per call unnecessarily long.

Fig. 10 (a) demonstrates the effect of maximal allowable number of alternate paths on the blocking probability for the long-lived traffic obtained by simulation, EFPA and OPCA. The offered long-lived traffic is 6.3 Erlangs and the threshold is 16 (80%). Fig. 10 (b) demonstrates the effect of maximal allowable number of alternate paths on the blocking probability for the short-lived traffic obtained by simulation, EFPA and OPCA. We focus on the 6-node fully meshed network as the low average node degree of NSF network restricts the number of alternate paths and therefore the number of overflows. The

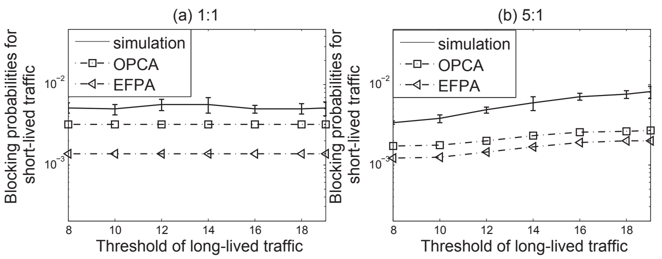


Fig. 11. Blocking probability for short-lived traffic in a 6-node fully-meshed network with long-lived and short-lived traffic. The ratio of the offered long-lived traffic to offered short-lived traffic is (a) 1:1 and (b) 5:1.

offered traffic load for both long-lived and short-lived traffic are 3.6 Erlangs. We change the maximal allowable number of alternate paths, while keeping all the other parameters the same as in the *Default parameter setting* in IV-A.

We observe that there is a clear benefit, in the present example, of 6-node fully meshed network, to increase the maximal number of overflow to at least 2. After that, the rate of decrease in the blocking probability for both long-lived and short-lived traffic slow down as D increases, due to the inefficiency of the long alternate paths.

G. *The effect of trunk reservation*

In general, thresholds are applied to reserve channels for the primary path traffic and prevent the network from being crowed by the overflow traffic. Also, the threshold of long-lived traffic can protect the short-lived traffic from being preempted by long-lived overflow traffic which requires longer paths to establish a call, uses more resources, and may cause congestion. Threshold of long-lived traffic should be chosen carefully because small threshold for long-lived traffic will cause many overflow calls unable to enter the network, leading to large blocking probability for the long-lived traffic while large threshold invites too much overflow traffic, congesting the network and preempting short-lived calls.

We again consider a 6-node fully meshed networks and the offered long-lived traffic and offered short-lived calls are both 3.6 Erlangs. We change the threshold of long-lived traffic, while keeping all the other parameters the same as in the *Default parameter setting* in IV-A.

For this case, Fig. 11 (a) illustrates the effect of threshold for long-lived traffic on the blocking probability for the short-lived traffic.

We can observe that when the both offered long-lived traffic and offered short-lived traffic are equal, varying threshold of long-lived traffic does not affect the blocking probability for the short-lived traffic. This is because the blocking probability for the long-lived traffic is very small and therefore the change of the carried load of long-lived traffic (background traffic of short-lived traffic) caused by the change of threshold for long-lived traffic is also very small, so its effect on blocking probability for the short-lived traffic is negligible.

Fig. 11 (b) shows the effect of changing threshold of long-lived traffic on the blocking probability for the short-lived traffic when the ratio of the offered long-lived traffic and offered short-lived traffic is 5:1. The offered loads are set to be 6 and 1.2, for long-lived and short-lived traffic, respectively, and all the other parameters are the same as in the *Default parameter setting* in IV-A. We can observe that when long-lived traffic is much higher than short-lived traffic, changing threshold of long-lived traffic will increase the blocking probability for the short-lived traffic, due to the fact that the blocking probability for the long-lived traffic is in a range that increasing threshold of long-lived traffic will considerably reduce the blocking probability and increase the carried load of long-lived traffic (which is the background traffic of short-lived traffic).

H. *Robustness of the quasi-stationary approximation*

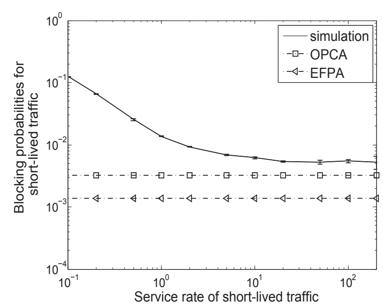


Fig. 12. The average blocking probabilities for the short-lived traffic with different values of μ_2 for a 6-node fully-meshed network with 20 links each trunk and $\mu_1 = 1$.

As stated in Section III, when the holding times of long-lived calls are far longer than those of the short-lived calls, we can use the quasi-stationary approximation to obtain accurate results. We use a 6-node fully-meshed network with 20 links each trunk to illustrate by how much the holding time of long-lived traffic should be longer than that of short-lived traffic, for the quasi-stationary approximation to be accurate. The result is shown in Fig. 12. In the scenario we consider, both the offered long-lived traffic and offered short-lived traffic are 3.6 Erlangs, and the mean holding time of long-lived calls $(1/\mu_1)$ is 1.

As expected, when the holding times of short-lived calls $(1/\mu_2)$ are larger than or close to those of long-lived calls, the short-lived calls cannot reach steady-state while the number of long-lived calls remains unchanged, so the quasi-stationary approximation is inaccurate. However, when the holding times of short-lived calls are significantly shorter than those of long-lived calls (e.g. by more than two orders of magnitude, namely $\mu_2 > 100$), the blocking probability for the short-lived traffic becomes invariant to further increase in the ratio μ_2/μ_1, indicating that short-lived traffic may approximately reach steady-state while its background state (due to long-lived traffic activities) remains approximately unchanged, so the quasi-stationary approximation can lead to accurate results. The errors shown in Fig. 12 in this condition are mainly due to overflow error and path error discussed above. From the figure, we can observe that our approximation methods work well when the average holding times of short-lived calls is less than 5% of the average holding times of long-lived calls.

I. The effect of the shape of the holding time distribution

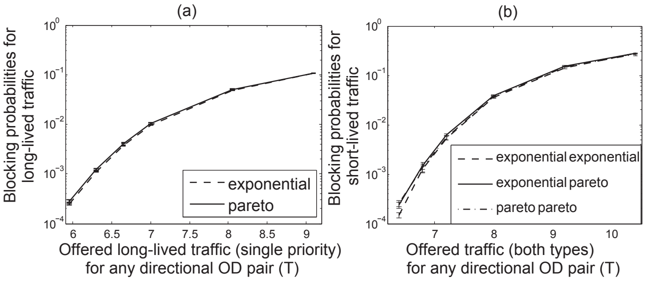


Fig. 13. The average blocking probabilities for (a) long-lived traffic (single priority), and (b) short-lived traffic, considering different service time distributions for a 6-node fully-meshed network with 20 links in each trunk.

The results presented above are based on the assumption that the holding times of both the long-lived and short-lived traffic are exponentially distributed. It is therefore important to examine the robustness of the approximations to the shape of the holding time distribution. To this end, we compare the results obtained under the exponential assumptions versus results obtained under heavy-tailed holding time distribution, where we maintain the same mean for the two alternatives.

In particular, we consider our heavy-tailed holding times, to follow a Pareto distribution with a complementary distribution function (CDF) presented in Subsection H-1 in [28], with δ (seconds) being the scale parameter and minimum holding time and γ being the shape parameter.

In our simulation we set $\delta = 0.5$ for long-lived traffic and $\delta = 0.0025$ for short-lived traffic and $\gamma = 2$ for both types of traffic. All the other parameters are kept the same as in the *Default parameter setting* in IV-A.

Fig. 13 (a) shows the simulation blocking probability for the long-lived traffic (single priority) for 6-node fully meshed network with holding time exponential distributed and Pareto distributed. The two curves are very close to each other and their confidence interval are overlapped, which shows that blocking probability for the long-lived traffic (single priority) is insensitive to the holding time distribution.

Fig. 13 (b) shows the simulation blocking probability for the short-lived traffic for 6-node fully meshed network of three cases:

1) exponential exponential – holding time of both long-lived and short-lived calls are exponentially distributed
2) exponential Pareto – holding time of long-lived traffic is exponentially distributed while that of short-lived traffic is Pareto distributed
3) Pareto Pareto – holding time of both long-lived and short-lived calls are Pareto distributed.

The closeness of the simulation results of the three cases illustrates that the blocking probability for the short-lived traffic for 6-node fully meshed network is insensitive to the holding time distribution shape.

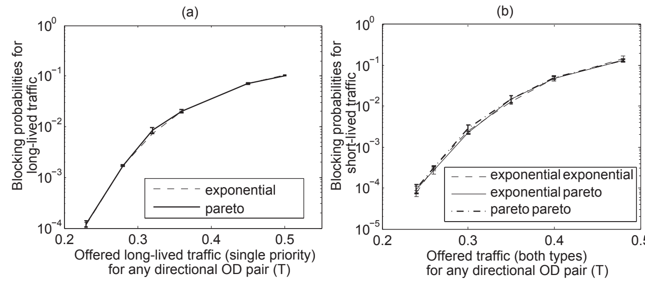


Fig. 14. The average blocking probabilities for (a) long-lived traffic (single priority), and (b) short-lived traffic, considering different service time distributions for the NSF network with 20 links in each trunk.

We have produced equivalent simulation results for the case of the NSF network topology. The results are presented in Fig. 14 (a) and (b). We observe similar behavior as in the case of fully meshed networks that provide further evidence that the assumption of exponential holding time distribution is reasonable and that the blocking probability is not very sensitive to the shape on the holding time distribution.

These results are consistent with equivalent results obtained in [28] for OBS networks.

J. Computational complexity of the algorithms

TABLE III. Computational complexity of the algorithms

Trunk capacity	Running time of EFPA (seconds)	Running time of OPCA (seconds)	Memory of EFPA (bytes)	Memory of OPCA (bytes)
C = 20	0.173369	0.413720	66840	89040
C = 100	0.431417	3.034318	95640	137040
C = 1000	278.515913	1065.012123	419640	677040

It is difficult to provide general analytical results for computational complexity of OPCA and EFPA because both require fixed-point iterations to converge. Nevertheless, we provide numerical examples that illustrate the time and memory complexity of the algorithms.

In Table III, we provide information on running times and memory usage of both EFPA and OPCA required for the network blocking probability computation in a six-node fully meshed network for three cases representing different trunk capacity values.

We have observed from these numerical examples, that for a small network, OPCA requires more computation time and more memory than EFPA, but the overall computing resources are manageable. As we demonstrate in the next subsection OPCA is also applicable to the large Coronet network while the equivalent EFPA results are unattainable.

K. The Coronet

We demonstrate here that OPCA is applicable to large scale networks such as the Coronet, shown in Fig. 15, while simulation results are computationally prohibitive for such

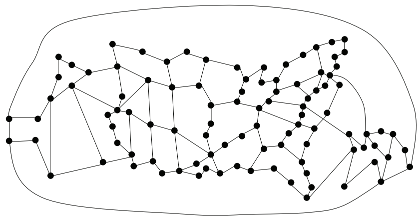


Fig. 15. The Coronet topology.

large scale networks. With all of the 9900 SD pairs in the network and one alternative path for each SD pair, the network blocking probabilities for long-lived and short-lived traffic can be obtained by OPCA within reasonable running time, as shown in Fig. 16. The number of links per trunk is 50 and the thresholds for long-lived traffic and short-lived traffic are 40 (80%) and 45 (90%). The maximum allowable number of alternate paths is 1. All the other parameters are kept the same as in the Default parameter setting in subection IV-A. The running times used to calculate the network blocking probabilities in the Coronet is about 121.734439 seconds by OPCA, obtained using MATLAB 7.6.0 executed on a desktop PC with IntelR CoreTM 2 Quad @ 3 GHz CPU, 4 GHz RAM and 32-bit operating system.

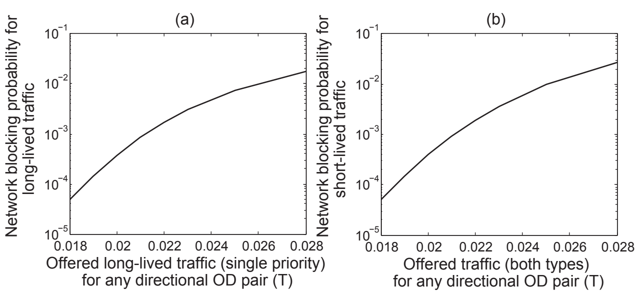


Fig. 16. The blocking probabilities obtained by OPCA for (a) long-lived traffic (single priority), and (b) short-lived traffic for the Coronet with 50 links per trunk.

V. Conclusions

In this paper, we have considered a circuit-switched network with long-lived and short-lived calls where the long-lived calls can preempt the short-lived ones. We use EFPA and OPCA combined with the quasi-stationary approximation to estimate the blocking probabilities. The results demonstrate that in most cases, OPCA can estimate the blocking probabilities reasonably well, and generally, better than EFPA. As long-lived calls provides background traffic for short-lived ones, the ratio of their offered load also affects the accuracy of the approximations. Reduction of offered long-lived traffic together with increase of offered short-lived traffic will improve the accuracy of OPCA, while that of EFPA is not much improved. However, when the number of links on each trunk increases, the performance of EFPA is improved. Allowing more alternate path traffic, either by increasing the maximum allowable alternate paths or the long-lived traffic threshold, is beneficial under light traffic. However, when the network is fully occupied, it is important to restrict alternate path traffic. We have observed that the quasi-stationary approximation requires that the mean holding time of long-lived calls is at least 20 times longer than that of short-lived calls. Nevertheless, this is not a very restrictive requirement if long-lived calls represent static calls and short-lived calls represent dynamic ones. We have also demonstrated that approximating blocking probability based on the exponential holding time assumption is not very sensitive to the shape of the holding time distribution, and is fairly accurate also for heavy-tailed holding time distributions. We have illustrated by numerical examples that, for small network, OPCA requires more computation time and more memory than EFPA, but the overall computing resources are manageable. For a large scale network such as the Coronet, we have demonstrated that OPCA is also applicable while the equivalent EFPA results are unattainable.

References

[1] A. Mahimkar *et al.*, “Bandwidth on demand for inter-data center communication,” in *Proc. The 10th ACM Workshop on Hot Topics in Networks (HotNets-X)*, Cambridge, MA, Nov. 2011.

[2] A. Barczyk, “World-wide networking for LHC data processing,” in *Proc. NFOEC’12*, Mar. 2012, pp. NTu1E.1 321–329.

[3] K. Sato and H. Hagesawa, “Optical networking technologies that will create future bandwidth-abundant networks,” *J. Opt. Commun. Networking*, vol. 1, no. 2, pp. A81–A93, Jul. 2009.

[4] P. O’Reilly, “The case for circuit switching in future wide bandwidth networks,” in *Proc. IEEE ICC’ 88*, vol. 2, Jun. 1988, pp. 899–904.

[5] G. Weichenberg, V. Chan, and M. Medard, “Design and analysis of optical flow-switched networks,” *IEEE/OSA J. Opt. Commun. Networking*, vol. 1, no. 3, pp. B81–B97, Aug. 2009.

[6] G. Shen and R. Tucker, “Energy-minimized design for IP over WDM networks,” *IEEE/OSA J. Opt. Commun. Networking*, vol. 1, no. 1, pp. 176–186, Jun. 2009.

[7] A. Girard, *Routing and Dimensioning in Circuit-Switched Networks*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc, 1990.

[8] A. Inoue, H. Yamamoto, and Y. Harada, “An advanced large scale simulation system for telecommunications networks with dynamic routing,” in *Network Planning in the 1990s*, 1989.

[9] F. P. Kelly, “Blocking probabilities in large circuit-switched networks,” *Adv. in Appl. Probab.*, vol. 18, no. 2, pp. 473–505, Jun. 1986.

[10] A. Farrel, A. Satyanarayana, A. Iwata, N. Fujita, and G. Ash, “Crankback Signaling Extensions for MPLS and GMPLS RSVP-TE,” Internet Engineering Task Force, Jul. 2007. [Online]. Available: http://www.ietf.org/rfc/rfc4920.txt

[11] F. P. Kelly, *Reversibility and Stochastic Networks*. Chichester: Wiley, 1979.

[12] L. Kosten, “On the blocking probability of graded multiples,” *Nachrichtentech Elek*, vol. 14, pp. 5–12, Jan. 1937.

[13] R. I. Wilkinson, “Theories of toll traffic engineering in the USA,” *Bell Syst. Tech. J.*, vol. 35, no. 2, pp. 421–514, Mar. 1956.

[14] J. F. Pérez and B. V. Houdt, “Markovian approximations for a grid computing network with a ring structure,” *Stochastic Models*, vol. 26, no. 3, pp. 357–383, 2010.

[15] J. Anselmi, Y. Lu, M. Sharma, and M. S. Squillante, “Improved approximations for the Erlang loss model,” *Queueing Syst. Theory Appl.*, vol. 63, pp. 217–239, Dec. 2009.

[16] K. Jung, Y. Lu, D. Shah, M. Sharma, and M. S. Squillante, “Revisiting stochastic loss networks: structures and algorithms,” *SIGMETRICS Perform. Eval. Rev.*, vol. 36, pp. 407–418, Jun. 2008.

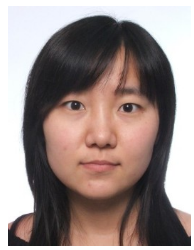
[17] F. P. Kelly, “Loss networks,” *Annals of Applied Probability*, vol. 1, no. 3, pp. 319–378, Aug. 1991.

[18] D. Mitra, “Asymptotic analysis and computational methods for a class of simple circuit-switched networks with blocking,” *Adv. in Appl. Probab.*, vol. 19, no. 1, pp. 219–239, Mar. 1987.

[19] E. W. M. Wong, A. K. M. Chan, and T. S. Yum, “Analysis of rerouting in circuit-switched networks,” *IEEE/ACM Trans. Networking*, vol. 8, no. 3, pp. 419–427, Jun. 2000.

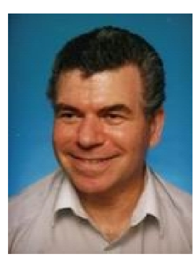
[20] E. W. M. Wong, T. S. Yum, and K. M. Chan, "Analysis of the *m* and m^2 routings in circuit-switched networks," *Eur. Trans. Telecommun.*, vol. 6, no. 5, pp. 613–619, Dec. 1995.
[21] R. B. Cooper and S. S. Katz, "Analysis of alternate routing networks with account taken of the nonrandomness of overflow traffic," Bell Telephone Lab, Tech. Rep., 1964, memo.
[22] W. Whitt, "Blocking when service is required from several facilities simultaneously," *AT&T Tech. J.*, vol. 64, no. 8, pp. 1807–1856, Oct. 1985.
[23] E. W. M. Wong, A. Zalesky, Z. Rosberg, and M. Zukerman, "A new method for approximating blocking probability in overflow loss networks," *Comput. Netw.*, vol. 51, no. 11, pp. 2958–2975, Aug. 2007.
[24] L. Delbrouck, "The uses of Kosten's systems in the provisioning of alternate trunk groups carrying heterogeneous traffic," *IEEE Trans. Commun.*, vol. 31, pp. 741–749, Jun. 1983.
[25] A. J. Coyle, W. Henderson, and P. G. Taylor, "Reduced load approximations for loss networks," *Telecommun. Syst.*, vol. 2, pp. 21–50, 1993.
[26] E. W. M. Wong, M. Y. M. Chiu, Z. Rosberg, M. Zukerman, S. Chan, and A. Zalesky, "A novel method for modeling and analysis of distributed video on demand systems," in *Proc. IEEE ICC 2005*, vol. 1, Seoul, Korea, May 2005, pp. 88–92.
[27] G. Raskutti, A. Zalesky, E. W. M. Wong, and M. Zukerman, "Enhanced blocking probability evaluation method for circuit-switched trunk reservation networks," *IEEE Commun. Lett.*, vol. 11, no. 6, pp. 543–545, Jun. 2007.
[28] E. W. M. Wong, J. Baliga, M. Zukerman, A. Zalesky, and G. Raskutti, "A new method for blocking probability evaluation in OBS/OPS networks with deflection routing," *IEEE/OSA J. Lightwave Technol.*, vol. 27, no. 23, pp. 5335–5347, Dec. 2009.
[29] C. Qiao, Y. Mei, M. Yoo, and X. Zhang, "Polymorphic control for cost-effective design of optical networks," *Eur. Trans. Telecommun.*, vol. 11, pp. 17–26, 2000.
[30] A. L. Chiu *et al.*, "Network design and architectures for highly dynamic next-generation IP-Over-Optical long distance networks," *J. Lightwave Technol.*, vol. 27, pp. 1878–1890, 2009.
[31] A. Elwalid, D. Mitra, I. Saniee, and I. Widjaja, "Routing and protection in GMPLS networks: From shortest paths to optimized designs," *J. Lightwave Technol.*, vol. 21, pp. 2828–2838, 2003.
[32] M. Youssef, E. Doumith, and M. Gagnaire, "Power-aware multi-rate WDM network design under static/dynamic traffic," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, Dec. 2011, pp. 1–6.
[33] A. Leon-Garcia, R. H. Kwong, and G. F. Williams, "Performance evaluation methods for an integrated voice/data link," *IEEE Trans. Commun.*, vol. 30, no. 8, pp. 1848–1858, Aug. 1982.
[34] O. Boxma, A. Gabor, R. Nunez-Queija, and H.-P. Tan, "Performance analysis of admission control for integrated services with minimum rate guarantees," in *Proc. 2nd NGI '06*, Apr. 2006, pp. 41–47.
[35] W. Song and W. Zhuang, "QoS provisioning via admission control in cellular/wireless LAN interworking," in *Proc. BroadNets 2005*, vol. 1, Oct. 2005, pp. 543–550.
[36] M. Baldi, G. Marchetto, and Y. Ofek, "A scalable solution for engineering streaming traffic in the future Internet," *Comput. Netw.*, vol. 51, Oct. 2007.
[37] I. Katib and D. Medhi, "Adaptive alternate routing in WDM networks and its performance tradeoffs in the presence of wavelength converters," *Opt. Switching Networking*, vol. 6, no. 3, pp. 181–193, 2009.
[38] V. B. Iversen, *TELETRAFFIC ENGINEERING and NETWORK PLANNING*. [online]ftp://ftp.dei.polimi.it/users/Flaminio.Borgonovo/Teoria/teletraffic_Iversen.pdf/, 2010.
[39] W. Feller, *An Introductionto Probability Theory and Its Applications*. New York: Wiley, 1968.
[40] A. Itai, Y. Perl, and Y. Shiloach, "The complexity of finding maximum disjoint paths with length constraints," *Networks*, vol. 12, no. 3, pp. 277–286, 1982.
[41] M. Middendorf and F. Pfeiffer, "On the complexity of the disjoint paths problem," *Combinatorica*, vol. 13, pp. 97–107, 1993.
[42] C. Chekuri and S. Khanna, "Edge disjoint paths revisited," in *Proceeding SODA '03*, PA, USA, 2003, pp. 628–637.
[43] J. Wolf, A. Viterbi, and G. Dixon, "Finding the best set of k paths through a trellis with application to multitarget tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, no. 2, pp. 287–296, Mar. 1989.
[44] C. Xin, C. Qiao, and S. Dixit, "Traffic grooming in mesh WDM optical networks-performance analysis," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 9, pp. 1658–1669, Nov. 2004.
[45] F. Delcoigne, A. Proutière, and G. Régnié, "Modeling integration of streaming and data traffic," *Perform. Eval.*, vol. 55, no. 3-4, pp. 185–209, 2004.

Meiqian Wang received her Bachelor degree and Master degree in Department of Electronic Engineering at Harbin institute of Technology in 2006 and 2009, respectively. She is now working towards the PhD degree at City University of Hong Kong. Her research interest lies in performance evaluation in circuit switching and burst switching networks.



Shuo Li received the B.Sc. degree in electronic and communication engineering from City University of Hong Kong, Hong Kong, in 2009. She is currently working toward the Ph.D. degree in the department of Electronic Engineering at City University of Hong Kong. Her research interests are performance evaluation of telecommunication networks.

Eric W. M. Wong (S'87–M'90–SM'00) received the B.Sc. and M.Phil. degrees in electronic engineering from the Chinese University of Hong Kong, Hong Kong, in 1988 and 1990, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Massachusetts, Amherst, in 1994. In 1994, he joined the City University of Hong Kong, where he is now an Associate Professor with the Department of Electronic Engineering. His research interests include the analysis and design of telecommunications networks, optical switching and video-on-demand systems.



Moshe Zukerman (M'87–SM'91–F'07) received the B.Sc. degree in industrial engineering and management and the M.Sc. degree in operations research from the Technion – Israel Institute of Technology, Haifa, Israel, and the Ph.D. degree in engineering from University of California, Los Angeles, in 1985. He was an independent Consultant with the IRI Corporation and a Postdoctoral Fellow with the University of California, Los Angeles, in 1985–1986. In 1986–1997, he was with the Telstra Research Laboratories (TRL), first as a Research Engineer and, in 1988–1997, as a Project Leader. He also taught and supervised graduate students at Monash University in 1990–2001. During 1997-2008, he was with The University of Melbourne, Victoria, Australia. In 2008 he joined City University of Hong Kong as a Chair Professor of Information Engineering, and a group leader. He has served on various journal editorial boards and conference technical program committees.