International Telecommunications Symposium – ITS2002, Natal, Brazil

Improving simulation performance of Engset's multiservice loss networks

Zdenko Vrdoljak, Josip Gracin and Mladen Kos
University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, HR-10000 Zagreb, Croatia

Abstract— **This paper deals with performance improvement of event-driven simulation of multiservice loss networks by using Engsets's customers model approximation. Model is based on simulation state modulated homogeneous Markov chain. The model obtains equal simulation results as direct model of group of customers. Simulation runs significantly faster, especially for large number of customers considered.**

I. INTRODUCTION

THE primary goal of network engineering is to solve the problem of finding suitable network topology and to obtain optimal sizing of network equipment. Significant progress has already been made in the field of finding suitable network topologies. However, the problem of sizing capacities of transmission systems and sizing of switching nodes is far from being well solved. Current solutions primarily minimize the overall price of switching equipment needed. This is mainly due to extremely complex analytical models describing network traffic and consequently the hardness of extrapolating capacity based on those models.

This paper is concerned with per-connection performance analysis of guaranteed services traffic in multiservice environment. There are two basic problems. One problem of the analysis is calculating end-to-end blocking probabilities for each pair of nodes and each service. The other problem is optimal sizing of resource capacities with objective to satisfy maximum end-to-end blocking probabilities predefined for each service. As we shell discuss later, calculating blocking probabilities involves analysis of complex multidimensional stationary Markov chains. Explicit expressions developed for both Erlang and Engset models are highly complex and require analytical and numerical approximations. Even though blocking probabilities can be calculated from approximations for fixed and alternate routing, extrapolating capacities is extremely difficult. The complexity additionally increases when dynamic routing schemes are considered.

Alternative to the complex analytic approach is simulation-based analysis. Event-driven simulation is the most suitable technique for this problem. It can be used for both performance evaluation and optimum capacity planning, when suitable methods are used. Such an approach has been discussed in [9]. In this paper we are concerned with accurate modeling of the behavior of end users in accordance to Engset's traffic model. To demonstrate quality of the approximation we developed, simulations on two different event-driven simulators were performed. We compare results for Engset's and Erlang's loss network model.

Zdenko Vrdoljak, Josip Gracin and Mladen Kos are with the University of Zagreb, Faculty of Electrical Engineering and Computing, Department of Telecommunications, Zagreb, Croatia, Phone: +385 1 6129 731 Fax: +385 1 6129 832. E-mails: {zdenko, gracin, kos}@tel.fer.hr.

This paper is organized as follows. In Section II we give a short overview of available analytic methods for calculating blocking probabilities and indicate hardness of application of these methods to real networks. In Section III we briefly describe simulation method and capacity planning method. In Section IV we describe Engset's simulation model of customer collections in detail. Section V presents simulation results and comparison to Erlang loss network model. We conclude in section VI.

II. PROBLEMS OF ANALYTIC APPROACH

Homogeneous Markov chains theory is mathematical tool most often used to accurately describe multiservice loss networks. Basically, there are two models: Erlang's model and Engset's model. Due to larger complexity of Engset's model, the most often analyzed model is Erlang's loss network model. It assumes that call establishment requests for a single service arrive according to Poisson process with constant rate λ. On the contrary, Engset's model takes into consideration the fact that call arrival rate is modulated by number of free customers in customers group.

The most simple analysis of loss networks is the one of fixed routing networks. Assume a network with M links (resources) where each link is of capacity C_m expressed as integer multiple of some basic rate ϵ, $\mathbf{C} = \{\mathbf{C_1}, \dots, \mathbf{C_M}\}$. Let there be K different services, each requesting bandwidth $\beta_{k,m}$ on link m, where $\beta_{k,m}$ is an integer multiple of defined basic rate ϵ, $\beta = \{\beta_\mathbf{1}, \dots, \beta_\mathbf{M}\}$. Let r_k be a fixed route used by service k to establish connection and $\mathcal{R}$ set of all routes defined in the network. $\beta_{k,m}$ equals 0 in case link m does not belong to route r_k. Assume that call arrival rate and mean call duration generally depend on the number of established connections of service k on route r_k. Let arrival rate of service k be $\lambda_k(i)$ and mean call duration $1/\mu_k(i)$ where i is the number of service-k connections established on route r_k. Let Markov chain state vector be $\mathbf{n} = \{\mathbf{n_1}, \mathbf{n_2}, \dots, \mathbf{n_k}\}$ and $\mathcal{S}(\mathbf{C})$ set of allowed states corresponding to sharing policy used. The usual sharing policy considered is complete sharing policy (CS) for which

$$\mathcal{S}(\mathbf{C}) = \left\{ \mathbf{n} \in \mathbb{Z}^{+^{\mathbb{K}}} : \beta \cdot \mathbf{n} \leq \mathbf{C} \right\}. \quad (1)$$

It can be easily shown that stationary probability of vector $\mathbf{n}$ is given by

$$\hat{p}(\mathbf{n}) = \prod_{\mathbf{j}=\mathbf{0}}^{\mathbf{K}} \left[\frac{\lambda_\mathbf{j}(\mathbf{0})}{\mu_\mathbf{j}(\mathbf{n_j})} \prod_{\mathbf{i}=\mathbf{1}}^{\mathbf{n_j}-\mathbf{1}} \frac{\lambda_\mathbf{j}(\mathbf{i})}{\mu_\mathbf{j}(\mathbf{i})} \right] \times \mathbf{G}[\mathcal{S}(\mathbf{C})], \quad \mathbf{n} \in \mathcal{S}(\mathbf{C}), \quad (2)$$

where $G[\mathcal{S}(\mathbf{C})]$ is normalization constant given by

$$G[\mathcal{S}(\mathbf{C})] = \sum_{\mathbf{n}\in\mathcal{S}(\mathbf{C})} \prod_{\mathbf{j}=\mathbf{0}}^{\mathbf{K}} \left[\frac{\lambda_{\mathbf{j}}(\mathbf{0})}{\mu_{\mathbf{j}}(\mathbf{n_j})} \prod_{\mathbf{i}=\mathbf{1}}^{\mathbf{n_j}-\mathbf{1}} \frac{\lambda_{\mathbf{j}}(\mathbf{i})}{\mu_{\mathbf{j}}(\mathbf{i})} \right]. \tag{3}$$

The generic product form allows us to deduce Engset and Erlang models for any sharing policy, but fixed routing. Blocking probability for certain service k is obtained by summation of stationary probabilities of blocking states for the service. By certain manipulations it can be shown that blocking probability for service k on route r_k is given by

$$p_{B_k} = 1 - \frac{G[\mathcal{S}(\mathbf{C})]}{G[\mathcal{S}(\mathbf{C})]}, \tag{4}$$

where $\mathcal{S}(\mathbf{C})$ is subset of $\mathcal{S}(\mathbf{C})$ in which service k is not blocked. The main problem with evaluating blocking probability is large complexity of evaluating function $G[\mathcal{S}(\mathbf{C})]$. One of the relaxations to such a large complexity is numerical inversion algorithm developed by Choudhury, Leung and Whitt ([1], [2]).

Another approach to solving the problem of evaluating blocking probabilities is to use approximations. There are several approximations developed. Two most significant ones are classical normal approximation and Erlang fixed point approximation. It can be shown ([4]) that these approximations obtain good results when ratio of link capacity and offered traffic per link remains constant. The main idea of both approximations is the assumption that blocking on each link is independent from others. As shown in [4], there exists a parameter $B_m \in [0, 1\rangle$ such that blocking probability of service class k on route r_k is approximately

$$p_{B_k} \approx 1 - \prod_{m\in r_k} (1 - B_m)^{\beta_{k,m}}. \tag{5}$$

In case of single service network (e.g. telephone network), parameter B_m equals the value of Erlang's loss formula $E_m = E(\rho_m, C_m)$:

$$p_{B_k} \approx 1 - \prod_{m\in r_k} (1 - E_m)^{\beta_{k,m}}, \tag{6}$$

where

$$\rho_m = \sum_{r:m\in r} \frac{\lambda_r}{\mu_r} \prod_{j\in r-\{m\}} (1 - E_j). \tag{7}$$

Even though these results look simple, they require solving the system of non-linear equations. In some cases their application is even more stressing than calculating blocking probabilities from exact equation (4).

When alternate routing is considered, calculating blocking probabilities becomes even more complex since overflow traffic is considered. Difficulty is in non-randomness of overflow traffic. This problem was first studied by Wilkinson [10] and Pratt [7] who analyzed this problem with equivalent random traffic theory. However, when dynamic routing is considered, problem of calculating end-to-end blocking probabilities becomes even more complex.

Even though the blocking probability can be (at least approximately) calculated, the problem of inverting blocking formulas becomes very difficult and imposes use of complicated numerical methods. This all indicates that analytic approach to calculating blocking probabilities in general case is inexhaustible source of applied mathematics research, but still represents a problem to be solved by other approaches, especially when capacity planning is considered. In this paper we consider simulation-based capacity estimation method and analyze Markov chain based models describing large groups of customers in order to improve its performance.

III. Multiservice loss network simulation

Easier way to obtain blocking probabilities is to run event-driven simulation of the multiservice loss networks. The most simple simulation method is direct simulation where each call establishment, termination and rejection event is considered. Blocking probabilities are estimated as ratios of total number of rejections and attempts. In order to well estimate blocking probabilities, large simulation run lengths are needed. However, by increase of processor speeds, time needed to obtain good estimate becomes tolerable.

If fixed routing network is considered, a rational alternative is usage of one of simulation speed-up techniques. One of the most suitable ones is importance sampling technique (see [8]). Even though this technique allows far more precise estimate of blocking probability for fixed routing loss networks, estimating blocking probability for networks with alternate and dynamic routing becomes as tricky as the network mathematical model itself.

Even though simulation is most often used for estimating blocking probability, it can also be used to estimate optimum network resource capacities. One possible capacity estimation method is presented in [9]. This method is based on idea to modulate network resource capacities by measured call blocking probability during simulation as long as targeted GoS (Grade of Service) is unsatisfied. Customers and services are modeled according to Erlang's or realistic Engset's model. In the beginning of the simulation all resource capacities equal 0. First call attempts are rejected and measured blocking probability equals 1. Since the measured GoS is not equal to the targeted one, capacities are increased according to some modulation function as the following call attempts arrive. Thus the measured blocking probabilities decrease and capacities converge to their final values. The procedure of increasing capacities is stopped when GoS for all the services and all customer groups reach targeted value.

In case of relatively small number of customer groups, the most convincing results are obtained when considering Engset's model. This is a well known observation, used in many applications, including equivalence random theory by Wilkinson. Due to this reason we prefer using Engset's model in simulation. However, direct simulation of Engset's model requires large amounts of memory, resulting is rather slow simulation. In order to accelerate the simulation we developed a Markovian approximation model that precisely describes realistic Engset's model. The approximation model is described in the following section.

IV. Engset's model of customers collections

The obvious incompatibility between Erlang's and Engset's model is the variance of number of active connections established. In case of Erlang's model with no blocking, mean and variance equal $E[X] = \alpha/\beta$, $Var[X] = \alpha/\beta$ where X is Markov chain (birth-death process) in equilibrium (i.e. stationary state). α is call attempts arrival intensity and $1/\beta$ mean call duration. Equivalent Engset's model in equilibrium without blocking, designated by Y, has mean and variance equal $E[Y] = N\frac{\lambda}{\lambda+\mu}$, $Var[Y] = N\frac{\lambda\mu}{(\lambda+\mu)^2}$ where N is number of customers considered in model, $1/\lambda$ mean time between two successive calls initiated by a single customer and $1/\mu$ mean call duration. Comparison of these two simplified models can only be done by considering equal mean of active customers. Poisson arrival process of Erlang's model can be considered as sum of individual Poisson processes coming from each of customers in the group. Let a be offered traffic per customer in Erlang's model so that $\lambda/\mu = Na$. If $\mu = \beta$ then by equalizing mean values we obtain

$$\frac{\alpha}{\mu} = Na = N\frac{\lambda}{\lambda+\mu} \quad \Rightarrow \quad \alpha = N\frac{\lambda\mu}{\lambda+\mu}.$$

Variance ratio equals

$$Var[X]/Var[Y] = 1 + \lambda/\mu. \tag{8}$$

Since λ/μ ratio is about 0.1 during busy traffic hour, variances of both models is almost equal.

More significant is the difference in call arrival process. While Erlang's model generates Poisson call arrival process, call inter-arrival process of Engset's model is modulated by number of free customers in observed collection. Full Engset's model is very difficult to describe by Markov chain. It can be approximated by 3 dimensional Markov chain. The first dimension relates to number of calls initiated from the group, the second dimension to number of customers busy by received calls from other groups and the third dimension relates to number of calls terminated at the same group. However any attempt to find stationary state probabilities fails to obtain simple analytic results what makes the model completely useless. The only way to obtain analytic distribution of number of active customers is to approximate it by normal distribution by using central limit theorem. However, parameters of the normal distribution is difficult to determine analytically. This is why simulation can be used to estimate parameters of the normal distribution.

Simulation technique used for this purpose is most often event-driven simulation technique. In order to realistically simulate customers behaviour according to Engset's model, simulator has to reserve one or more event object for each of the customers. However, number of customers considered may be very large. Even more, the number of physical customers multiplies with number of services present in the network. This all easily results in reservation of large amounts of memory which is unacceptable for practical applications, for instance capacity planning.

This is why we developed a homogeneous Markov chain based approximation model requiring only one event object per customer group. The model assumes that time between two successive calls of a single customer is distributed exponentially regardless of number of calls customer received during the period. We assume that call duration distribution is irrelevant in case the call duration distribution posses rational Laplace transform, due to the "insensitivity" property (e.g. [3]). This is fortunately a very common case.

Each collection can be modeled as an independent Markov chain (birth-death process) $X(t)$. State of the process is unconventionally defined as a pair (n, m) where n is the number of customers that initiated the ongoing call and m is the number of customers receiving a call. The number of customers in the collection is N. Let $\tilde{p}$ be the probability that call initiated from any member of collection terminates in the same collection.

According to regularity property of Markov chains, direct consequence of differentiation of Kolmogorov equation, it follows that state vector $\vec{x}$ probability equals

$$p_{\vec{x}}(t+h) = \sum_{\substack{\vec{x}'\neq\vec{x},\\ \vec{x}'\in\mathbf{S_x}}} p_{\vec{x}',\vec{x}}(h)\cdot p_{\vec{x}'}(t). \tag{9}$$

h is very short period of time and $\mathbf{S_x}$ set of all allowable states

$$\mathbf{S_x} = \left\{ \tilde{\mathbf{x}} = (\mathbf{n},\mathbf{m}) \in \mathbb{Z}^{+\nvDash} : \ltimes + \gtrdot \leq \mathbb{N} \right\}.$$

$p_{\vec{x}',\vec{x}}(h)$ is transition probability through time h equal to

$$p_{\vec{x}',\vec{x}}(h) = q_{\vec{x}',\vec{x}}\cdot h + o(h), \tag{10}$$

where $q_{\vec{n}',\vec{n}}$ is transition rate from state $\vec{n}'$ to state $\vec{n}$, member of transition rate matrix $\mathbf{Q}$ of dimensions $N\times N$ (i.e. infinitesimal generator). $o(h)$ is Landau symbol defined by

$$\lim_{h\to 0}\frac{o(h)}{h} = 0.$$

It is clear that incoming call arrival process changes the number of available (free) customers in observed collection. The same number is also changed by establishing calls inside of observed collection (with probability $\tilde{p}$). Since the current call request process of the observed collection is proportional to number of currently available customers, it is clear that this process is dependent on activity of other collections. Collections thus mutually modulate current call arrival processes intensities. Additionally, intensity is modulated by number of successfully established connections inside and outside of collection.

Assume that current state vector at time point t is $\vec{x} = (n, m)$. m is an independent random variable representing number of customers occupied by receiving call. This variable is modulated by both observed and all other collections. We have no information regarding its behavior. Observe a short period of time h. If h is infinitesimally short, the following are states reachable from state $\vec{x} = (n, m)$ through period h: (n^-, m), (n^+, m), (n^+, m^+), (n^-, m^-), (n, m^-) and (n, m^+). The most interesting are last two states: (n, m^-) and (n, m^+). They are entered only by receiving call from other collections. Since those transition are not controlled by the Markov chain describing collection itself, they are not considered at all. Since the change of variable m cannot be fully controlled by the model, states (n^-, m) and (n^-, m^-) can be considered as one, as well as states (n^+, m) and (n^+, m^+). This is justified because states (n^-, m^-) and (n^+, m^+) will be entered with probability $\tilde{p}$ whenever n changes. By grouping the states we converge to a modulated one-dimensional birth-death process $Y[t, m(t)]$, function of parameter t, simulation time, and outer modulating process

$m(t)$. In order to create a model useful for simulations we have to calculate distribution of time between state shifts, i.e. increase or decrease of number of calls initiated from the observed collection, and probabilities of transitions to first higher and lower states at the moment of state shift.

$Y\left[t, m(t)\right]$ is thus one dimensional Markov chain with intensity transition matrix $\mathbf{Q}[\mathbf{m}(\mathbf{t})]$ modulated by stochastic process $m(t)$ representing number of connection terminating at the collection. Define functions $\delta_{n^-,m}$ and $\delta_{n^+,m}$ as follows:

$$\delta_{n^+,m} = \begin{cases} 1 & n + m(t) < N \\ 0 & n + m(t) = N \end{cases} \qquad \delta_{n^-,m} = \begin{cases} 1 & n > 0 \\ 0 & n = 0 \end{cases} \tag{11}$$

Probability that process Y is at state n at time $t + h$ equals:

$$\begin{aligned} & P\left\{Y[t+h, m(t+h)] = n\right\} = \\ & = P\left\{Y[t, m(t)] = n-1\right\} \cdot p_{n-1,n}[h, m(t)] + \\ & + P\left\{Y[t, m(t)] = n+1\right\} \cdot p_{n+1,n}[h, m(t)] + \\ & + P\left\{Y[t, m(t)] = n\right\} \cdot \\ & \cdot \left\{1 - p_{n,n-1}[h, m(t)] - p_{n,n+1}[h, m(t)]\right\}. \end{aligned} \tag{12}$$

According to regularity property transition probability $p_{n-1,n}[h, m(t)]$ is proportional to current number of available customers and intensity of establishing connection of a single customer:

$$\begin{aligned} & p_{n,n+1}[h, m(t)] = \\ & = \left\{\left[N - m(t) - n\right]\lambda h + o_1(h)\right\}\delta_{n^+,m} = \\ & = q_{n,n+1}^{m(t)} h + o_1(h). \end{aligned} \tag{13}$$

Similarly, $p_{n+1,n}[h, m(t)]$ is proportional to current number of established connections and intensity of closing connection, no matter the number of connections terminating at the observed collection:

$$\begin{aligned} p_{n,n-1}[h, m(t)] &= \left\{n\mu h + o_2(h)\right\}\delta_{n^-,m} = \\ &= q_{n,n-1}^{m(t)} h + o_2(h). \end{aligned} \tag{14}$$

According to general Markov chains theory, the time the chain resides in state k is distributed exponentially with parameter equal to the k-th diagonal element of transition intensity matrix multiplied by -1. Thus it follows that mean time the process $Y[t, m(t)]$ resides in state n equals:

$$T_{n,m(t)} = \frac{1}{n\mu\delta_{n^-,m} + [N - m(t) - n]\lambda\delta_{n^+,m}}. \tag{15}$$

where $m(t)$ is number of busy customers at moment when process entered state n. At this point we notice a large deviation from the exact model. Namely, since the state residence time is determined at moment of state transition and since event with new state shift time is generated, model behaves like no new connections will be established toward customers inside collection. This is why described simulation model only approximates exact model. However, deviation is not insignificant in stationary state of the model, after simulation advanced for some time. This is because incremental transition intensities are proportional to $N - m(t) - n$ and variation of $m(t)$ is small, i.e. expected change of $m(t)$ during period of time process resides in state n is negligible.

After state shift event was fired, simulation model has to decide whether to initiate a new call or terminate one. Again, according to general Markov chain theory transition probabilities are:

$$\begin{aligned} p_{n^+,m} &= \frac{[N - n - m(t)]\lambda\delta_{n^+,m}}{n\mu\delta_{n^-,m} + [N - n - m(t)]\lambda\delta_{n^+,m}} \\ p_{n^-,m} &= \frac{n\mu\delta_{n^-,m}}{n\mu\delta_{n^-,m} + [N - n - m(t)]\lambda\delta_{n^+,m}} \end{aligned}$$

By defining these parameters, we have described all the mechanisms needed to build simulation model. The most important consequence is that there is only one event needed per collection, while in a conventional model one event is needed per each customer. This significantly improves simulation performance. The overall number of event objects present in the simulator is proportional to number of customer groups thus alleviating priority queuing processing in event lists and event distribution complexity.

The problem arises in case when the birth-death process decides to increase the number of active customers and call blocking occurs. Since customers are expected to retry shortly after blocking, an additional process appears, increasing call establishment attempts intensity. In case when blocking probability is very small it is justified to generate additional events emulating repeated attempts. However, in case when blocking probability is large, a parallel stochastic process should be considered. However, this problem is out of the scope of this paper.

V. Simulation results and comparison

In this section we demonstrate how well our model approximates simulation model in which a separate event thread is used for each customer - direct model. We set up a small scenario consisting of two customer collections connected via link of an infinite capacity. Simulation was run on LATS ATM Simulator [5]. These results have later been verified by the same model implemented in OPNET Modeler [6]. Simulation configuration is described by figure 1. It comprises of 5 collections of customers: observed collection, tree groups generating call arrivals into observed collection and sink, an inactive group receiving calls from the observed collection. Groups are connected by links and nodes of an infinite capacity so that blocking probability due to limited resources is 0. The only blocking that occurs is when a called customer is already busy due to an ongoing call.

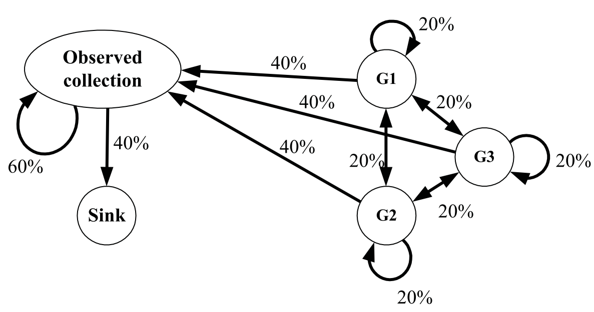

Fig. 1. Simulation configuration.

Groups G1, G2 and G3 have 10000 customers each. They direct 40% of overall call attempts to the observed collection, close 20% inside themselves and exchange remaining 40% among themselves. Observed collection closes 60% of all calls inside itself and sends remaining 40% to the sink group. Sink group does not generate any traffic.

First we compare means and standard variations of number of connections established from the observed collection to the sink obtained by direct simulation and approximation model. As can be seen from figure 2, mean number of connections established in direct simulation matches mean number of connections established in simulation with approximation model for different number of customers in the observed collection. Figure 3 shows negligible difference between means and standard deviations obtained for different numbers of customers in the observed collection. One should take into account statistical error introduced by limited number of sample points used to calculate means and standard deviations.

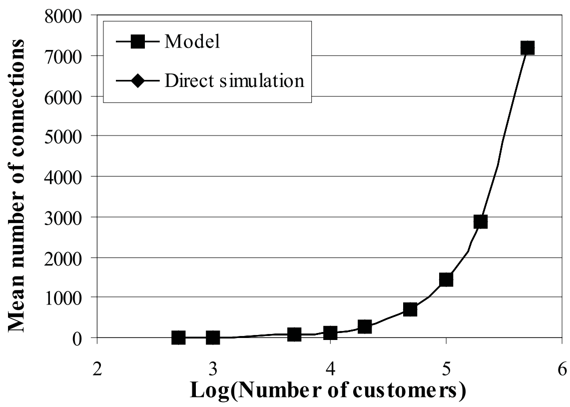

Fig. 2. Mean number of active connections for direct simulation and simulation model vs. customers group size.

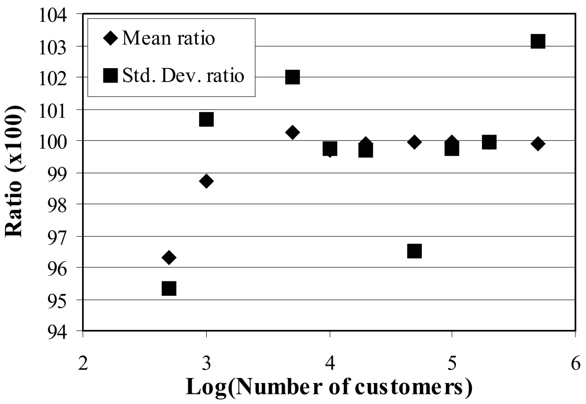

Fig. 3. Ratio of mean and variance of number of active connections for direct simulation and simulation model vs. customers group size.

Previous figures show how well statistical parameters match for both direct simulation and approximation simulation model. Figure 4 shows how well relative frequencies (i.e. probability density functions - PDFs) match for two simulation models.

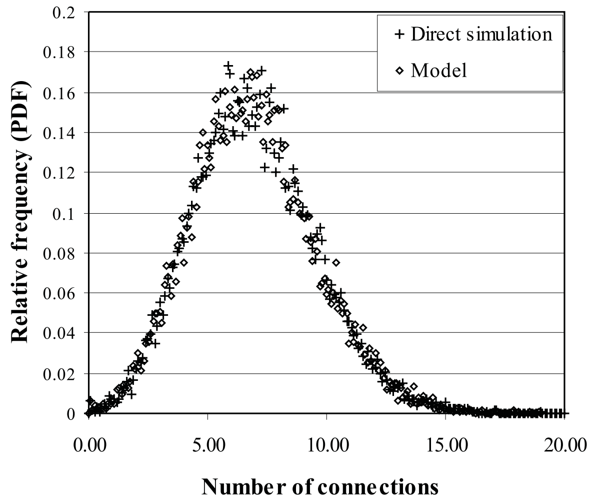

Fig. 4. Relative frequency (PDF) of number of established connections for customers group size
$N = 1000$.

As stressed before, the most important advantage of the approximation model in contrast to the direct simulation is simulation speed. Graph in figure 5 shows that simulation execution time when using direct simulation is at least 2 times greater than in the case of approximation model. As the number of simulated customers increases, simulation execution time ratio increases. After certain number of customers is reached, direct simulation becomes unacceptably slow and ratio of execution times becomes greater than 10. In our case this difference is reached for the total number of 4 million customers in the network.

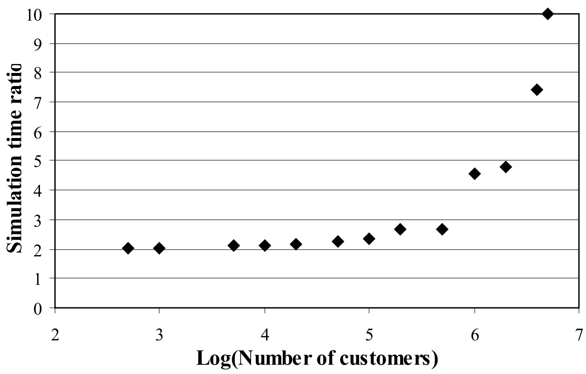

Fig. 5. Comparison of simulation execution times.

VI. Conclusion

Analytic approach to planning capacities in multiservice loss networks is too complex for a suitable practical applications. This is why we emphasize the importance of using the simulation in solving this complex problem. A capacity planning algorithm is reviewed in order to demonstrate how simulation can be used instead of analytic approach.

The main aim is estimating capacities in the network as accurately as possible. When a conventional event-driven simulation, a separate event thread for each customer in the simulation. This results in low performance of the simulation.

In order to improve simulation performance, we have developed a simple simulation model that very accurately describes direct model in which each customer is simulated separately. In order to demonstrate this improvement we ran dozens of simulations with both direct and approximate model. As shown, approximate model behaves quite accurately, with negligible deviations in comparison to direct model. This model allows us to design capacities of large multiservice network with millions customers according to Engset's loss model by using simulation-driven method described.

REFERENCES

[1] G. L. Choudhury, K. K. Leung, and W. Whitt, "An inversion algorithm for loss networks with state-dependent rates," in *Proceedings of IEEE INFOCOM'95*, 1995, pp. 513–521.

[2] G. L. Choudhury, K. K. Leung, and W. Whitt, "An algorithm to compute blocking probabilities in Multi-Rate Multi-Class Multi-Resource Loss Models," in *Advances in Applied Probability*, vol. 27, pp. 1104–1143, 1995.

[3] J. S. Kaufman, "Blocking in a Shared Resource Environment," in *IEEE Transactions on Communications*, vol. COM-29, pp. 1474–1481, October 1981.

[4] F. P. Kelly, "Loss Networks," in *Annals of Applied Probability*, vol. 1, pp. 319–378, 1991.

[5] M. Kos, Z. Vrdoljak, "Simulation of ATM Network Behavior with the Event Driven ATM Simulator," in *Proceedings of MEleCon 2000*, Vol. 1, pp. 202–205, March 2000.

[6] OPNET Technologies, Inc., OPNET Modeler, www.opnet.com.

[7] C. W. Pratt, "The Concept of Marginal Overflow in Alternate Routing," Fifth International Teletraffic Congress, pp. 51-58, New York, June 1967.

[8] P. E. Lassila, J. T. Virtamo "Nearly Optimal Importance Sampling for Monte Carlo Simulation of Loss Systems," *ACM Transactions on Modeling and Computer Simulation*, vol. 10, No. 4, pp. 326–347, October 2000.

[9] Z. Vrdoljak, G. Kovacevic, M. Kos "Connection-level simulation-based capacity planning method," *Proc. of SPIE, Internet Performance and Control of Network Systems II*, R.D. van der Mei, F. Huebner-Szabo de Bucs Eds., vol. 4523, pp. 314–325, August 2001.

[10] R. I. Wilkinson, "Theories for Toll Traffic Engineering in the U.S.A.," The Bell System Technical Journal, Vol. 35, pp. 421-514, March 1956.