European Journal of Operational Research 000 (2016) 1–9

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

A method of approximate analysis of an open exponential queuing network with losses due to finite shared buffers in multi-queue nodes☆

Miron Vinarskiy*

Institute of Control Sciences of Russian Academy of Science, Computer Science, 3709-A Mariana Way, Santa Barbara, CA 93105, United States

A R T I C L E   I N F O

Article history:
Received 16 March 2016
Accepted 19 September 2016
Available online xxx

Keywords:
Queuing
Queuing network
Multi-queue node
Finite buffer
Losses

A B S T R A C T

We consider a model of an open exponential queuing network where each node comprises several multi-class $M_R/M/1$ queues that share a common waiting space (a buffer) of limited capacity. A customer arriving to a node with fully occupied buffer is lost. An assumption is made that each class input traffic to a node, which is a superposition of the class external Poisson flow and the class flows coming from other nodes, is a Poisson process. Under this assumption a method of an approximate analysis is presented. It is based on solving iteratively a system of non-linear equations for the unknown nodal flow rates. It is shown that the gradient iterations solve the multi-class network equations. For the single-class model we use the direct substitution iterations. In the latter case existence and uniqueness of the solution, obtained by the iterative algorithm, is rigorously proven. It is demonstrated for a few network configurations that the network and node performance characteristics received by analytic approach are close to those obtained by simulation method. Our contribution is a performance evaluation methodology that could be usefully employed in queuing network design.

1. Introduction

Queuing network models with finite buffering in nodes (service centers) have been attractive to researchers for decades. Limiting waiting space in real-life nodal storage leads to a so-called “blocking” when a customer cannot get into the fully occupied buffer. The blocked customer creates a “traffic jam” at the node he is trying to enter, causing a “backward” influence on the traffic at other nodes in the network. It is very difficult, if not impossible, to achieve an exact analytical solution in this case. Various approximate, heuristic and simulation methods for the queuing networks with blocking have been described in the literature (see, e.g., Balsamo, 2011; Gelenbe & Mitrani, 2010; Kim, 2011; Kleinrock, 1974/1975; Osorio & Bierlaire, 2009; Perros, 1994). Lam (1976) considered a packet-switching network model where a blocked customer (packet) is unlimitedly retransmitted from the adjacent node until the nodal buffer becomes open. In Baumann and Sandmann (2016) authors have performed an exact computational analysis for a small two queue tandem network with losses at the first queue and blocking between queues because of a finite buffer at each queue.

Another approach for an approximate analysis is to consider network models with losses, when blocked customers are dropped and leave the network. Such “replacement” method provides a good insight into the functioning of a network with blocking. Customer loss probabilities in the network nodes indicate potential “bottlenecks” and can be used in evaluating an allocation of network service resources.

The relative simplicity of network models with losses makes it possible to investigate them analytically. This is done by decomposing a network into a set of simple nodal models, and integrating the nodal results into network flow balance equations. Following three references demonstrate this approach for queuing networks with single-queue nodes and losses.

Bronshtein and Gertsbakh (1984) analyzed a Jackson-type open exponential queuing network (Kleinrock, 1974/1975) with finite buffers. The assumption was made that the mixture of the original external Poisson stream into a node and the traffic coming from other nodes is a Poisson flow. Under this assumption they presented an approximate analytical analysis based on a system of non-linear equations for nodal traffic rates. They demonstrated

☆ A completed version of the WIP paper “A Model of An Open Exponential Queuing Network with Losses Due To Finite Shared Buffers in Multi-Queue Nodes,” that appeared at SCSC’13 (Summer Computer Simulation Conference), Toronto, Canada, July 2013.

* Fax: +1 805 563 6356.
E-mail address: mironvin@yahoo.com

http://dx.doi.org/10.1016/j.ejor.2016.09.031

by comparison with simulation that for network topologies with a high level of connectivity the analytical results coincide with those from simulation.

Shi (1995) introduced a linear approximation for more general network as an extension of the QNA (Queuing Network Analyzer) approach, developed by Whitt (1983) for general networks of queues with no capacity constraint. His method treats each node separately as GI/G/m/k system partially characterized by the first two moments of the interarrival time and service time distribution. The non-linear network balance equations, that combine the nodal results, are converted to a linear system by using heuristic approach. The relative accuracy of the method was compared to simulation and generally was not exceeded 20%.

Heindl (2003) studied a general queuing network with single-server nodes and Markovian routing without feedback loops and alternative routes between any two given nodes. Markov-modulated Poisson process (MMPP(2)) with two states was used as an input. He exploited existing decomposition formalisms based on renewal processes in order to include semi-Markov processes (SMPs) as the node outputs and convert them to MMPP(2)s. The network was partitioned into individual nodes which were analyzed in isolation as MMPP(2)/G/1/k systems with the complete service time distributions (not only the first two moments). The investigated examples revealed that relative errors of less than 10% may be expected.

The network model under study here is an extension of the Bronstein and Gertsbakh model to the case of multi-queue nodes with multi-class $M_R/M/1$ queues that share a finite common buffer in each node. The single-class network model was briefly introduced in (Vinarskiy, 2013) as a Work-In-Progress research. Buffer sharing policy is Complete Sharing (CS), where no restrictions on buffer occupancy are imposed for any queue. A special class of customers, for example, in a packet switching network, may consist of all customers (packets) between a given source-destination pair of network nodes. In a distributed data processing system the data traffic and control messages can constitute the distinguished classes.

The model can be used for performance evaluation of computer networks and distributed data processing systems. Output queuing structures in shared-memory switches/routers are good examples of such nodes (Aello, Kesselman, & Mansour, 2008). In this application, a packet memory pool is shared among output ports. Another example could be a distributed data processing system with nodes implemented as shared-memory architecture multiprocessor service centers. Processors can either perform identical tasks or have different functionalities. A customer (data request) can travel between nodes in order to get access to a distributed database. All node processor queues are kept in a finite area of node shared memory. Upon completing its service by a node processor, a customer can leave the system from the node or continue service at either the next node or at the same node by different processor.

To solve the model we use a decomposition of the network into separate nodal models combining the nodal results in a system of non-linear equations for the unknown nodal flow rates. It is shown that the system can be solved iteratively. The Newton–Raphson iterations (Ortega & Rheinboldt, 1970) are used to solve the multi-class network flow balance equations. For the single-class network model we use the direct substitution iterations, and a proof that iterations converge to a unique solution is provided. In both cases the solution for the nodal flow rates is used to receive several all-network and node performance measures.

It should be noted that traditional queuing network models use a notion of a node as a service facility with one or multiple servers associated with a queue. The novelty of our network model with losses is in considering multi-queue nodes with limited common waiting spaces. Each queue is served by one dedicated server. To the best of our knowledge, the only network model known for multi-queue nodes with finite shared buffers is the Lam's packet-switch network model (Lam, 1976). It focuses on the communication protocol details, such as retransmission of blocked packets and time-outs. No packet losses are allowed.

The remainder of this paper is organized as follows. In Section 2, we provide a formal description of the multi-class network model, the node product-form state distribution, and concentrate on the network flow balance equations. The Newton–Raphson iterations are formally introduced to solve the equations. To reduce the computational complexity of solving the multi-class flow balance equations we have transformed them to the non-linear equations for the node non-blocking probabilities. The latter are solved by the Newton iterations, which rapidly converge in this case to a solution even for many flow classes. The required network performance measures are defined as well.

Section 3 introduces the direct substitution iterations procedure to solve the single-class network equations. The proof of convergence of the iterations to a unique solution is presented.

Section 4 is devoted to a comparison of the numerical results obtained analytically with simulation results. The simulation code is written in C++ and simulates a processing of $\sim 10^6$ customers through the network. For multi-queue nodes a few network configurations are investigated including single-class networks and multi-class ones.

2. Multi-class network model

We consider an open queuing network with W nodes and R distinguished customer classes. Each node is connected to at least one node, i.e., there are no isolated nodes. Let us denote by $R_i(R_i \le R)$ the number of customer classes traveling through the node-i ($i=1, 2, \ldots, W$) and use upper index-r to specify the class-r ($r=1, 2, \ldots, R_i$) variables. The node-i comprises $Q_i > 1$ of the multi-class $M_{R_i}/M/1$ queues sharing finite common buffer of size N_i units "Fig. 1". The buffer contains all Q_i queues, including customers in service. Exponential queuing system q ($q=1, 2, \ldots, Q_i$) in node-i serves customers of all classes with rate μ_q, and its queuing discipline is FCFS (first come first served). A customer of any class arriving at a node when its buffer is fully occupied is lost and leaves the network. If there are free spots in the buffer of node-i, then an arriving class-r customer joins the qth queue system with probability $\alpha_{iq}^{(r)} (\sum_{q=1}^{Q_i} \alpha_{iq}^{(r)} = 1)$. Customers initially arrive to the network from R external sources, that generate Poisson flows with rates $\lambda_0^{(r)}, r = 1, 2, \ldots, R$. The class-$r$ flow is distributed between nodes according to probabilities $p_{0i}^{(r)}$, $i=1, 2, \ldots, W$, $\sum_{i=1}^{W} p_{0i}^{(r)} = 1$. A class-$r$ customer, who completed his service in node-i, is transferred to node-j with routing probability $p_{ij}^{(r)} (i, j = 1, 2, \ldots, W, \sum_{j=1}^{W} p_{ij}^{(r)} \le 1)$, or completes his service in the network and leaves with probability $p_{iE}^{(r)} = 1 - \sum_{j=1}^{W} p_{ij}^{(r)}$.

Fig. 1 shows the class-r traffic in a node in the network. Node-i receives the "original" class-r Poisson flow from the class-r external source with the rate $\lambda_0^{(r)} p_{0i}^{(r)}$. Other nodes in the network and possibly the node-i itself produce a "secondary" flow (dashed line) to the node-i. Superposition of the original and secondary flows forms the class-r input flow to the node-i with the rate $\lambda_i^{(r)}$. A part of this flow with the rate $\lambda_{iL}^{(r)}$ is lost. The rest goes through the node-i and then splits into a secondary flow with probability $1 - p_{iE}^{(r)} = \sum_{j=1}^{W} p_{ij}^{(r)}$ and traffic with flow rate $\lambda_{iE}^{(r)}$ exiting the network after the node-i.

The main step in the network analysis is to determine $\lambda_i^{(r)}$ ($i=1, 2, \ldots, W, r=1, 2, \ldots, R_i$). To do so we will use the following natural assumption: the superposition of the original class-r external Poisson flow to the node-i and all secondary class-r flows from nodes in the network to the node-i is a Poisson process.

Please cite this article as: M. Vinarskiy, A method of approximate analysis of an open exponential queuing network with losses due to finite shared buffers in multi-queue nodes, European Journal of Operational Research (2016), http://dx.doi.org/10.1016/j.ejor.2016.09.031