Introduction

In recent years there has been an increased interest in the analysis of queueing networks with finite capacity queues. This is probably due to the realization that these queueing networks are useful in modelling computer systems, communication networks, and flexible manufacturing systems. However, with the exception of the proceedings of a workshop on queueing networks with finite capacity queues, there has not been any other publication, and in particular a journal publication, which gives the state of the art in this research area. We are hoping that this special issue will fill in this gap.

Queueing networks with finite capacity queues are subject to *blocking*. That is, the flow of jobs through one node may be stopped for a moment if a destination node has reached its full capacity. (Please note that this type of blocking is not related to the notion of blocking in teletraffic, where an arriving job is blocked, i.e. lost, if the node is full.) The set of rules that dictate when a node becomes blocked or unblocked is commonly referred to as the *blocking* mechanism. There are basically only a few blocking mechanisms that have been extensively studied in the literature. Unfortunately, each author (and that goes for the editors of this special issue) has chosen to give different names to each of these blocking mechanisms. As a result, there may be as many as four different names for the same blocking mechanism! In order to alleviate this problem we asked the authors who contributed to this special issue to use the names given below. Each name was chosen so as to reflect a special feature of the blocking mechanisms. It is hoped that these names will be widely used in the future.

Blocking-after-service: A job upon service completion at node i attempts to join destination node j. If node j at that moment is full, the job is forced to wait in node i, in front of the server until it enters destination node j. The server remains blocked for this period of time and it cannot serve any other jobs waiting in the node. This blocking mechanism is known by the names: type 1 blocking, transfer blocking, production blocking, and non-immediate blocking.

Blocking-before-service: A job in node i declares its destination (say node j) prior to starting service. If node j is full, the server of node i becomes blocked, i.e. it cannot serve any jobs. When a departure occurs from destination node j, the server of node i becomes unblocked and the job begins receiving service. This blocking mechanism is known in the literature by the names: type 2 blocking, communication blocking, immediate blocking, and service blocking.

I.F. Akyildiz was born in Istanbul, Turkey in 1954. He received his Vordiplom, Diplom Informatiker and Doctor of Engineering degrees in Computer Science from the University of Erlangen Nürnberg, West Germany, in 1978, 1981 and 1984, respectively. Currently, he is an Associate Professor in the School of Information and Computer Science at Georgia Institute of Technology. He is a co-author of a textbook entitled *Analysis of Computer Systems* published by Teubner Verlag in the fall of 1982. He is an Editor for *Computer Networks and ISDN Journal*. His research interests are in the area of performance evaluation, computer networks, distributed systems, and computer security. Dr. Akyildiz is a senior member of IEEE and a member of ACM (Sigops and Sigmetrics) and GI (Gesellschaft für Informatik).

H.G. Perros received the B.Sc. degree in Mathematics in 1970 from Athens University, Greece, the M.Sc. degree in Operational Research with Computing from Leeds University, England, in 1971, and the Ph.D. degree in Operations Research from Trinity College Dublin, Ireland, in 1975. From 1976 to 1982 he was an Assistant Professor in the Department of Quantitative Methods, University of Illinois at Chicago. In 1979 he spent a subbatical term at INRIA, Rocquencourt, France. In 1982 he joined the Department of Computer Science, North Carolina State University, as an Associate Professor, and since 1988 he is a Professor. During the academic year 1988/89, he was on a sabbatical leave of absence first at BNR, Research Triangle Park, North Carolina, and subsequently at the Laboratoire MASI, University of Paris 6, France. He was the conference chairman of PERFORMANCE '86 and ACM SIGMETRICS 1986, and he was also the conference co-chairman of the First International Workshop on Queueing Networks With Blocking. His research interests are in the area of computer and communication performance modelling. He is a member of the IFIP Working Group 7.3.

North-Holland
Performance Evaluation 10 (1989) 149–151

Repetitive-service: A job upon service completion at node i attempts to join node j. If node j is full, the job receives another service at node i. This is repeated until the job completes a service at node i at a moment when node j is not full. This type of blocking is known in the literature by the names: type 3 blocking, and rejection blocking. In the above definition, it was assumed that the destination is fixed. In general, one can distinguish two cases: *fixed destination* and *random destination*. In the first case, once the job's destination has been selected it cannot be altered. That is, each time the job completes a service, it attempts to enter the same destination. In the second case, each time the job completes a service, a new destination node is chosen independently of the node chosen the previous time.

This special issue comprises five papers, two short communications, and a bibliography of the relevant literature. These contributions are briefly summarized below.

Van Dijk introduces a method for computing an upper bound on the throughput in closed queueing networks with exponential servers. The method consists of considering an aggregate open network of which throughput acts as an upper bound for that of the closed one. He gives necessary conditions for the existence of a simpler upper bound and for a bound on the error. The method is illustrated by a simple example. Several approximations have been proposed for such systems before but they are rarely theoretically justified. This paper is the first attempt to check the approximation accuracy a priori. Van Dijk's results can be used in the future in the investigation of throughput and error bounds for general queueing networks, even though the verification of the proposed necessary conditions may prove to be extremely difficult.

Kouvatsos and Xenios propose an approximation algorithm for the analysis of closed (and open) queueing networks with multiple servers, generalized exponential service times, and repetitive-service blocking. The approximation method is based on the maximum entropy methodology. The authors show through numerical experimentation that the algorithm has a good accuracy.

Frein and Dallery propose a method for analyzing approximately cyclic queueing networks with blocking-before-service. Their approximation method involves a node by node decomposition of the queueing network under study. Numerical validation shows that the algorithm gives good estimates for the througput and mean queue-length.

Brandwajn and Sahai introduce a speed-up technique for the approximate analysis of queueing networks with blocking-before-service. This technique is an improvement of an earlier approximation introduced by Brandwajn and Jow, which has been shown to give good results. The main idea is to decompose the queueing network under investigation into successive two-node subsystems which are then solved using a 'back and forth sweep' technique. The speed-up technique reduces the computation cost by a factor of up to 2. The authors also present two methods for the solution of the two-node subsystems. Method 1 works if the first node has a buffer which is larger than the buffer of the second node and Method 2 works well in the reverse case. The improvement on the convergence of the iterative scheme is illustrated through numerical examples.

Hillier and So consider the problem of where to place a fixed number of additional servers in order to maximize the throughput of a tandem queueing network with blocking-after-service. Their general conclusion is that the interior nodes should be given preference over the end nodes for receiving an additional server.

Ammar and Gershwin analyze fork/join networks with exponentially distributed service times and finite buffers using the concept of duality. In particular, they obtain equivalencies between fork/join queueing networks with finite buffers and queueing networks with blocking but without the fork/join operation.

Onvural obtains some product-form solutions for multi-class queueing networks with blocking using the notion that if the state space of a reversible Markov process is truncated, the resulting Markov process is also reversible.

Finally, in order to further enhance this special issue it was decided to include a bibliography by Perros of the relevant papers in the area of queueing networks with finite capacity queues. This list of references is an updated version of an earlier bibliography published by the same author.

We would like to thank the Editor-in-Chief Martin Reiser and the Managing Editor Werner Bux for allowing us to edit this special issue. We would also like to thank all the authors who

submitted to this special issue for their interest. Although many fine papers could not be included (we received over 20 papers) we anticipate seeing them in other publications. Finally, we would like to express our appreciation to all referees for their diligent and timely efforts and their substantive comments on the quality and appropriateness of the submitted papers.

I.F. Akyildiz
H.G. Perros
Atlanta and Raleigh
September 1989