Journal of Applied Mathematics and Stochastic Analysis, **14**:4 (2001), 421-426.

QUEUEING NETWORKS - CUSTOMERS, SIGNALS, AND PRODUCT FORM SOLUTIONS

by X. Chao, M. Miyazawa, and M. Pinedo

and

INTRODUCTION TO STOCHASTIC NETWORKS

by R.F. Serfozo

A BOOK REVIEW

HANS DADUNA
Institut für Mathematische Stochastik
Universität Hamburg
Bundesstrasse 55, D-20146 Hamburg, Germany
daduna@math.uni-hamburg.de

(Received March, 2000; Revised August, 2001)

Networks of queues (= stochastic networks) have been a field of intensive research over the last three decades. The foundation for this research is classical queueing theory, which dealt mainly with single node queueing systems [4, 5, 8]. There is now a well-developed theory of stochastic networks accompanied by an unbounded set of open problems which originates directly from applications as well as from theoretical considerations. Most of these open problems are easily stated and put into a theoretical framework, but they often require either intricate techniques on an ad hoc basis or deep mathematical methods. Often, both of these approaches have to be combined to tackle successfully the solution of quickly formulated problem.

The development of queueing network theory, which provided application areas with solutions, formulas, and algorithms, commenced around 1950 in the area of Operations Research, with special emphasis on production, inventory, and transportation. The first breakthroughs were the works of Jackson [10] and Gordon and Newell [10]; both papers appeared in *Operations Research.* The second breakthrough in queueing network theory was already connected with Computer Science: the celebrated papers of Baskett, Chandy, Muntz, and Palacios [2], which appeared in the *Journal of the Association for Computing Machinery*, and of Kelly [11]. The two volumes of Kleinrock's book [13, 14] appeared at the same time as [11]. From that time on, queueing network theory and its applications have become intimately connected with performance analysis of complex systems in Computer and Communication Sciences, both at the hardware and software levels. Today's growth of production, manufacturing, and transportation, with information processing and communi-

cation technology results in more and more complex systems, which require even more elaborated models, techniques, and algorithms for a better understanding of their behavior and for prediction of their performance and quality of service.

Perhaps the most influential books associated with the development of stochastic network theory were those by Kelly [12] and Whittle [16]. Later on, Walrand [15] and van Dijk [6] gave summaries of the fundamentals of application-oriented methods and principles. Before going into details of the books under review, it should be mentioned that there is a recent book of El-Taha and Stidham [7] that covers networks of queueing systems. The emphasis there is on the pathwise analysis of the network's development over time, with applications in performance analysis.

Almost all the models and results of network theory reported in the aforementioned books are connected with the principles of separability and product form equilibrium. This roughly means that the spatial steady state distributions of the networks for a given fixed time point can be separated into local steady states concerning the individual nodes. Therefore, these nodes in equilibrium for a fixed time point behave *as if* they were spatially independent (in the open network case) or *as if* they were conditional distributions of independent vector distributions.

It is our experience up to now that, besides direct simulation experiments, only these analytical methods have made stochastic network theory successful in applications. In fact, even in cases where the necessary assumptions for separability and applications of the product form theory are not fulfilled, the analytical product form network theory contributes to performance analysis by providing methods which support classical (heuristic) principles of disaggregation and aggregation to tackle the isolated components. This can be looked upon as an application of the product form models in such cases as a strong *modeling assumption.*

Having observed over twenty years the enormous development of network theory, the time seems to be ripe for trying to summarize the current state of network theory. Indeed, a survey on the state-of-the-art of at least a great part of this area of research is provided in the recent books of Serfozo and Chao, Miyazawa, Pinedo (henceforth abbreviated by **Ser**, **CPM**, respectively), which are the subject of this review. Both books are integral to today's development of analytical methods for queueing networks. But it is obvious that no single book can contain the whole set of models and theory available in the filed. This can easily be seen by the contents of the books, which are mostly nonoverlapping and strongly reflect the authors' individual fields of research. On the other hand, the joint topics of both books elucidate what is now the kernel of analytical (mainly product form) network theory.

Both **Ser** and **CPM** start with *Fundamentals*, which contain classical exponential Jackson networks and their relatives (e.g., closed Gordon-Newell networks, networks with limited access, mixed networks), and discuss their structure and equilibrium behavior in connection with the networks' development in reversed time, utilizing in a first step the structure of time reversal Markov processes (**Ser** Chapter 2, **CPM** Section 2.4). In a second step, there is a detailed investigation of quasi-reversibility in queueing networks at the exponential level: What single queues can be pasted together to build up a network in such a way that the local queues of the network in equilibrium show a behavior which is structurally similar to that in isolation? (**CPM** Chapter 3, 4; **Ser** Chapter 8).

Reversibility and quasi-reversibility are intimately connected with some *local balance*, *partial balance* or *detailed balance* structure. These definitions vary over the different literature sources, and are often adapted to the special models and processes under consideration. Similar results are achieved here (**Ser** Section 1.5, 2.8, 2.9, 3.9,

and throughout Chapter 8, **CPM** Section 2.3, Chapter 10). All these notions are the underlying modeling features leading to explicit descriptions of the steady state behavior of the described Markovian state processes. As indicated above, the first attempt to explicitly solve the equilibrium (global balance) equations for stochastic networks by Jackson, Gordon and Newell resulted in independent state vectors or conditioned distributions of independent vectors; the product form is visible in the formulas.

Now there are some extensions of the notion of product forms which apply to different extensions of the network types (e.g., if there are customers of different types circulating in the network, the queue length process is no longer Markovian). Often, the type of a sequence process in such single systems shows an *internal* product form, and applying suitable (new) routing concepts leads to separable steady states, which are again similar to those found in Gordon-Newell or Jackson networks. Furthermore, if service times are not exponential, neither the queue length nor type processes are Markovian. Using the additional assumption, that the service discipline is symmetric in a specified sense, yields the two-step procedure: First, there is an internal product form for the single node behavior (i.e., an additional factor occurs for each customer, with nonexponential service request in a system describing either residual service requests or ages. Secondly, there is an *external* product form over the nodes as observed in more elementary systems as well (**CPM** Chapter 6, **Ser** Chapter 3). The latter property is called *decomposability* in **CPM** Section 2.5. A further distinction made there takes note of a structural property of many network models dealt with in both books, which do not explicitly show spatial separability over the nodes. The steady state can be expressed as a product of two factors, one of which relates to the service process all over the network. The other factor expresses the asymptotic routing behavior.

In **CPM** Section 2.5, the term *functional product form* is used in contrast to *distributional product form* for the classical separability. Using this term, the so-called Whittle network's steady state (the nomenclature seems to be due to Serfozo) is of functional product form, but not of distributional product form (**Ser** Section 1.5). Many further examples of this situation are evaluated in **CPM** Chapter 9. The main advantage of these seemingly less explicit steady state results is that the factor in the functional product form, which relates to the overall service processes, is usually obtained directly in the modeling process via specification, (e.g., the service intensities). The factor concerned with the routing is usually obtained by solving generalized traffic equations. In contrast to the classical networks' case, these are no longer linear, due to:

1. the occurrence of negative customers, which may annihilate other customers present;
2. customers' coalescence; or
3. the spontaneous birth of new customers due to arrivals of external signals or internal jumps in the system's state (**CPM** Chapter 9, **Ser** Chapter 7).

Discussing many different modeling features obtainable by these principles in a detailed way is in my opinion at the heart of **CPM**. Conversely, **CPM** provides us with a survey on queueing networks with additional features and structures as indicated by the subtitle and the sequence of chapter headings; 6. Multiple customer classes and arbitrary service times, 7. Networks with batch services and negative signals, 8. Batch arrivals, batch services and concurrent movements. The problems, which emerge when batch arrivals and batch services are included as modeling devices, become apparent, even for the simplest examples, Ex. 2.18 of **Ser** and Sections 2.6,

2.7 in **CPM**.

The central feature in **Ser** is the application of point process theory and methods to evaluate the structure of network processes (**Ser** Chapter 4, 6, partly 8). Because, traditionally, network processes in applications are chosen to be Markovian, the author often reduces the necessary theory to the case of point processes generated by a specific embedded jump sequence in such Markovian processes. This tends to shorten the proofs, but makes the definitions not directly comparable to the standard texts of point process literature. Readers would likely benefit from having parallel access to [1] with the general definitions - but there is only limited space dedicated to network theory and their specific problems.

Exploiting the point process results, especially Palm theory, brings detailed investigation into a customer's individual behavior. This is a main topic in **Ser**, Sections 4 an d6. In **CPM**, the main topic is to derive steady state probabilities for various networks, and the individual customers' behavior is only sketched when proving some PASTA (Poisson Arrivals See Time Averages) or ASTA results by reduction to compute elementary conditional probabilities and respectively flows. In **Ser**, an in-depth study of possible applications of Palm probabilities is provided to determine sojourn and travel times even behind expectations. The latter can be obtained already from the steady state behavior via Little-type formulas (**Ser** Chapters 5, 6); much of the effort is devoted to obtain special mean transfer times between different node subsets of the networks.

Beside these different views regarding the main issue, which are presented to some extent in both books, there are some chapters in each book not covered by the other. In **Ser**, Chapter 9, a discussion of space-time Poisson models, which from a queueing theorist's viewpoint, are generalizations of networks of infinite server queues and their processes. In fact, there are many other applications; special attention is paid to systems of particles evolving in space and time. A certain extension of such models is presented in Chapter 10, but the main topic there is an investigation of spatial queueing systems. Here the service is not provided to the customers only at specific nodes; customers obtain service in regions of interest. Generalizations of Jackson and Whittle networks are given, and the asymptotic and equilibrium behavior of these systems is investigated. Additional throughputs and mean sojourn times in a spatial sector are those in computers.

CPM in Chapter 12 is devoted to discrete time networks of queues. Models considered here can be looked upon to a great part as discrete time counterparts of the continuous time systems (with additional features investigated before). Because Walrand's S-queues fit into the class of models under consideration, this chapter covers at least, in principle, the class of models investigated in [17]. Reading [17] should be done with the emphasis on the many applications presented there; finding a companion text is recommended. Another book concerned with discrete time networks is [3], which is more oriented towards applications. The network applications can mostly be looked upon as examples for aggregation/disaggregation principles.

Some topics that are not covered by either book under review are:

1. Diffusion approximation,
2. fluid approximations,
3. in-depth development of numerical algorithms for evaluating closed network models with the normalization constants (partition functions),
4. stochastic ordering techniques for queueing networks,
5. the asymptotics of networks when the number of nodes and the number of customers grow to infinity, which usually is strongly connected with

bottleneck analysis,

6. systems with an infinite number of queues and customers (some connections to this problem can be found in **Ser** Chapters 9, 10),
7. nonexponential networks without assuming servers to be symmetric, optimization and control of networks, especially stochastic dynamic optimization, which has found much interest in the literature.

Obviously, all the these topics cannot be covered by a single book. This is due to the rich fields of research and applications connected with networks of queues. Both books mainly present the sections of network theory where their authors have worked in over the last years, some of them over time. Therefore, both books collect an interesting summary of what one might view as the heart of network theory. Even if a reader might have different favorite topics with respect to queueing network theory and its application, he or she will benefit from both texts - at least many of the chapters of both books serve as an addendum compiled for convenience. Looking back on the common parts of both books described in the review, there seems to be an interesting conclusion: The core of network theory is still what was first collected in the chapters on general networks in the book of Kelly [12]. Clearly, the actual research and recently developed theory has evolved far and away from those early systems and results. But all that was elaborated there, and the fundamental product form results, were still evoked many times in the books under review.

References

[1] Baccelli, F. and Brémaud, P., *Elements of Queueing Theory*, Springer, New York 1994.

[2] Baskett, F., Chandy, M., Muntz, R. and Palacios, F.G., Open, closed and mixed networks of queues with different classes of customers, *J. of the Assoc. for Compu. Mach.* **22** (1975), 248-260.

[3] Bruneel, H. and Kim, B.G., *Discrete-Time Models for Communication Systems including ATM*, Kluwer Academic Publications, Boston 1993.

[4] Cohen, J.W., *The Single Server Queue*, second edition, North-Holland Publishing Company, Amsterdam-London, 1982.

[5] Cooper, R.B., Queueing theory, In: *Stochastic Models*, vol. 2 of *Handbooks in Operations Research and Management Science* (ed. by D.P. Heyman and M.J. Sobel), North-Holland, Amsterdam (1990), Chapter 10, 469-518.

[6] van Dijk, N.M., *Queueing Networks and Product Forms-A Systems Approach*, John Wiley and Sons, Chichester 1993.

[7] El-Taha, M. and Stidham, Jr., S., *Sample-Path Analysis of Queueing Systems*, Kluwer Academic Publisher, Boston 1999.

[8] Gnedenko, B.W. and König, D., *Handbuch der Bedienungstheorie*, volume 2, Akademie-Verlag, Berlin 1984.

[9] Gordon, W.J. and Newell, G.F., Closed queueing networks with exponential servers, *Opers. Res.* **15** (1967), 252-267.

[10] Jackson, J.R., Networks of waiting lines, *Opers. Res.* **5** (1957), 518-521.

[11] Kelly, F., Networks of queues, *Adv. of Appl. Prob.* **8** (1976), 416-432.

[12] Kelly, F.P., *Reversibility and Stochastic Networks*, John Wiley and Sons, Chichester - New York - Brisbane - Toronto 1979.

[13] Kleinrock, L., *Queueing Theory*, volume I, John Wiley and Sons, New York 1975.

[14] Kleinrock, L, *Queueing Theory*, volume II, John Wiley and Sons, New York 1976.
[15] Walrand, J., *An Introduction to Queueing Networks*, Prentice-Hall International Editions, Englewood Cliffs 1988.
[16] Whittle, P., *Systems in Stochastic Equilibrium*, Wiley, Chichester 1986.
[17] Woodward, M.E., *Communication and Computer Networks: Modelling with Discrete-Time Queues*, IEEE Computer Society Press, Los Alamitos, CA 1994.

Queueing Networks - Customers, Signals, and Product Form Solutions
by **X. Chao, M. Miyazawa, and M. Pinedo**
Publisher John Wiley and Sons
Publication Year 1999
ISBN 0471-98309-8
Price: $135.00

Introduction to Stochastic Networks
by **R.F. Serfozo**
Publisher Springer-Verlag
Publication Year 1999
ISBN 0-387-98773-8
Price: $69.75