計算機システムの制御と評価 17 － 1
（１９８２．１０．４）

QUEUEING NETWORK MODELS AND THEIR APPLICATIONS - A SURVEY

Hisashi Kobayashi, Science Institute, IBM Japan Ltd.

Summary:

As the complexity of today's data processing systems grows, system design, performance tuning and capacity planning can no longer be done effectively by brute-force simulation as we have done in the past. The need for a more systematic and economical way of evaluating and predicting system performance has been recognized for several years. In fact, we have witnessed considerable progress in performance evaluation methodology. For instance, queueing network theory, which originated in operations research back in the early 1960's, has been substantially extended by the computer performance community, and is now widely used by system designers and performance analysts.

We view a computer system as a multiple-resource system, where the resources are the CPU, memory, auxiliary storage, and I/O channels and devices. Jobs or programs demand services from these resources or servers. A primary function of the operating system is to manage the use of resources among many porgrams. Most performance problems are therefore related to queueing delay caused by contention for resources. A queueing network (or network of queues) representation provides us with a basic mathematical framework in dealing with resource allocation/contention problems in such a system.

In this tutorial presentation we will review a general class of queueing network models in which a simple "product form" solution exists for the joint queue size distribution. We will then discuss their applications to computer systems and communication networks. Various computational algorithms related to these models will also be discussed.

Many of the assumptions that we must make in such queueing models may seem unrealistic or oversimplified, yet the prediction results obtained are often found to agree surprisingly well with either the actual observations or the estimates obtained through more realistic simulations. These agreements are not coincidental; some measures of performance are rather insensitive or robust with respect to the distributional forms of service time and to the scheduling rules adopted. It is important for the system analysts to acquire some feeling as to when bold assumptions can be introduced without seriously affecting the prediction results.

There is an important constraint in queueing theory that intrinsically limits its applications to multiple resource systems-that is, a job (or customer) cannot occupy more than one resource (or server) simultaneously. Certainly this rule is violated in modeling a computer system: A job in execution holds at least main-memory space and the CPU simultaneously. The hierarchical modeling approach may sometimes overcomes difficulties of this kind.

References:

Baskett, F., K. M. Chandy, R. R. Muntz and F. G. Palacios (1975), "Open, Closed and Mixed networks of Queues with Different Classes of Customers," J. ACM. 22(2), 248-260.

Buzen, J. P. (1973), "Computational Algorithms for Closed Queueing Networks with Exponential Servers", CACM 16(9), 527-531.

Chandy, K. M. and C. H. Sauer (1980), "Computational Algorithms for Product Form Queueing Networks,"CACM 23(10), 573-583.

Kleinrock, L. (1976), Queueing Systems, Vol.Ⅱ: Computer-Applications Theory, Wiley, New York.

Kobayashi, H. (1976), "A Computational Algorithm for Queue Distribution via the Polya Theory of Enumeration", IBM Research Report RC-6154, August 1976.

Kobayashi, H. (1978a), Modeling and Analysis: An Introduction to System Performance Evaluation Methodology, Addison-Wesley, Reading, Mass.

Kobayashi, H. (1978b), "System Design and Performance Analysis Using Analytic Models" in K. M. Chandy and R. T. Yeh (Eds.). Current Trends in Programming Methodology, Vol.Ⅲ: Software Modeling, 72-114, Prentice-Hall, Englewood Cliffs, N. J.

Kobayashi, H. (1981), "Computational Algrithms for Markovian Queueing Networks", IBM Research Report RC-8820, April 1981. Also to appear in Probability Theory and Computer Science to be published by Academic Press (1982).

Kobayashi, H. and A. G. Konheim (1977), "Queueing Models for Communication System Analysis" (Invited Paper), IEEE Trans. Comm. COM-25, No.1 (January) 2-29.

Kobayashi, H. and M. Reiser (1975), "On Generalization of Job Routing Behavior in a Queueing Network Model," IBM Research Report RC-5252, February 1975.

Reiser, M. (1981), "Mean-Value Analysis and Convolution Method for Queue-Dependent Servers in Closed Queueing Networks," Performance Evaluation 1 (1). North-Holland Publishing Co, 7-18.

Reiser, M. (1982), "Performance Evaluation of Data Communication Systems" (Invited paper), Proceedings of IEEE, 70(2), 171-196.

Reiser, M. and H. Kobayashi (1975), "Queueing Networks with Multiple Closed Chains: Theory and Computational Algorithms," IBM Journal of Research and Development 19, (May), 282-294.

Reiser, M. and S. S. Lavenberg (1980), "Mean Value Analysis of Closed Multichain Queueing Networks", JACM 22 (April), 313-322.