THE TRANSIENT BEHAVIOR OF A SINGLE SERVER QUEUEING PROCESS WITH A POISSON INPUT

LAJOS TAKÁCS
COLUMBIA UNIVERSITY

1. Introduction

There is a huge literature on the stationary behavior of different types of queueing processes but only a few papers deal with their transient behavior. An extensive bibliography of the theory of queues by A. Doig [9] contains about seven hundred papers, most of which deal with stationary queues. The theory of stationary queues is very important because most of the queueing processes are ergodic, that is, starting from any initial state the process tends toward equilibrium irrespective of the initial state. In the state of equilibrium the process shows only statistical fluctuation with no tendency to a certain state. Many queueing processes rapidly approach equilibrium and this explains why we can apply with success the stationary approximation. However, the investigation of the transient behavior of queueing processes is also important, not only from the point of view of the theory but also in the applications. For instance, if we apply the stationary solution instead of the transient solution we are interested in the error of this approximation, and further, even in the case of the stationary process the linear least squares prediction presupposes a knowledge of the transient behavior of the process.

The mechanism of queueing processes is very simple. Customers are arriving at a counter according to a certain probabilistic law (Poisson input, Erlang input, recurrent input, and so forth). The customers will be served by one or more servers following a certain principle (service in order of arrival, random service, priority service, last come first served, batch service, and so forth). The service times are random variables governed by a given probabilistic law. After service the customers depart.

We shall always use the above terminology. Every conceivable process can always be described in this terminology. For instance, in the case of a telephone traffic process the calls, lines, and holding times are replaced by customers, servers, and service times, respectively.

A part of this paper was prepared in the spring of 1959 at the Research Techniques Division of the London School of Economics and Political Science. The paper was completed at Columbia University with the partial support of the Office of Naval Research under Contract Number Nonr-266(33), Project Number NR 042-034.