Optimal routing among ·/M/1 queues with partial information
Extended Abstract

E. ALTMAN[*], T. JIMÉNEZ[†], R. NÚÑEZ-QUEIJA[‡], U. YECHIALI[§]

In many applications in which routing decisions have to be made, the information relevant to the decision maker is only partially available. This is true, in particular, in high speed communication networks in which information about the state of a down stream node may be subject to a considerable delay, thus making that information irrelevant at decision epochs. This situation may occur even in case the down stream link is quite close but — links being unidirectional — the effect of routing decisions on the down stream node appears much quicker than the return of information about that node's congestion state.

We analyze dynamic routing choices between two paths. We model path delays using two ·/M/1 queues, with state dependent arrivals. The latter dependence is due to the fact that the arrival to any queue depends on the routing decision, which, in turn, depends on the congestion state of the path. Specifically, we consider the problem in which the controller taking the routing decisions can observe only one of the queues. In spite of lacking precise information about the second queue, the controller can probabilistically evaluate the influence of its routing strategies on the distribution of the congestion state of the second queue.

We concentrate on routing strategies of a random threshold type. These are characterized by two parameters: (n, r). If the number of jobs in the first path X_1 is less than n at an instance of arrival, the arriving job is sent to path 1. If $X_1 = n$, then it is routed to path 1 with probability r and to path 2 with the complementary probability. If $X_1 > n$, then it is

[*]INRIA, 2004 route des Lucioles, B.P. 93, 06902 Sophia Antipolis, France, also with CESIMO, Universidad de Los Andes, Mérida, Venezuela (altman@sophia.inria.fr)

[†]CESIMO, Universidad de Los Andes, Mérida, Venezuela (tania@ing.ula.ve)

[‡]CWI, P.O.B. 94079, 1090 GB Amsterdam, The Netherlands (sindo@cwi.nl); also with Eindhoven University of Technology

[§]Department of Statistics and Operations Research, School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel (uriy@post.tau.ac.il)

routed to path 2.

We study and analyze two types of dynamic routing control methods. In the first, we wish to determine the threshold parameters that minimize the average expected delay in the system. We call this the **Socially optimal routing** problem. Secondly, we consider **Individually optimal routing**. In this type of dynamic control we wish to obtain a routing strategy that minimizes the expected delay of each individual job. (The minimization of the expected delay for a given job does not take into account the impact of the individual routing decision on the delays that will be experienced by future jobs.) This type of routing approach has been advocated in the context of ad-hoc networks [1]. Although the routing decisions may be taken by a single controller, this control problem can be formulated as an equivalent game-theoretic model in which each job chooses its own route. We shall use the basic game-theoretic optimality concept of Nash equilibria [2]. Here, a Nash equilibrium is a set of routing strategies for all jobs such that no single job can decrease its weighted expected delay by deviating from its routing strategy.

The individual optimization problem was also considered in [2] in a different context: customers have to make a choice between two gas stations on a highway. Following a game theoretic analysis, a numerical solution is presented in [2] to obtain the performance for given candidate solutions. In our paper we provide a verifiable condition to check whether an equilibrium indeed exists within the class of threshold policies. When the service rates at the two stations are equal to each other, it is verified numerically (for $n = 1, 2, \ldots, 18$) that a threshold equilibrium exists, which is in agreement with the results in [2]. However, when the service rates at the two stations differ, such an equilibrium need no longer exist, as is shown in an example.

The structure of the talk is as follows. We first formulate — for a fixed (n, r) policy — the underlying Quasi Birth and Death process and derive the balance equations for the two-dimensional steady state probability distribution. The marginal distribution of X_2, the number of jobs in queue 2, is then derived for the case when $r = 1$. We then develop expressions for the joint Probability Generating Functions (PGF) of the queue lengths for any $r \in (0, 1]$. These results are then used to the study of the global (social) optimization and the individual optimization problems.

References

[1] P. Gupta and P. R. Kumar. A system and traffic dependent adaptive routing algorithm for Ad Hoc networks. In *Proceedings of the 36th IEEE Conference on Decision and Control*, pages 2375–2380, San Diego, USA, Dec. 1997.

[2] R. Hassin. On the advantage of being the first server. *Management Science*, 42:618–623, 1996.