

## ABSTRACT OF THE DISSERTATION

### Modeling Forking Nodes With State-Dependent Routing Application to Asymmetric Networks

by

Essia Hamouda El Hafsi

Doctor of Philosophy, Graduate Program in Computer Science  
University of California, Riverside, December 2005  
Dr. Mart Molle, Chairperson

In this work, we focus on the performance analysis of local scheduling policies at a “forking node” in a network, which must decide how to distribute traffic across multiple alternate paths leading to the final destination. Our problem formulation is general enough to allow each output port on the forking node to be a multi-server “link group” (for example, multiple wavelengths transmitting independent data streams over a single optical fiber), and to account for differences in the mean network transit delay along each path in the routing decisions. After evaluating several routing policies on the basis of average packet delay, ease of implementation in network switches, and compatibility with tractable analytical network performance models, we find that Join-the-Shortest-Queue with bias ( $JSQ + b$ ) scheduling has significantly better performance than any of the alternatives. Moreover, if the input process to a forking node with  $JSQ + b$  scheduling is Poisson, then standard time series analysis techniques show that its two outputs are very close to being independent Poisson processes.

Thus, if we can find a sufficiently accurate and efficient “offline” analytical performance model for a  $JSQ + b$  forking node, then we can extend the applicability of product-form queueing networks to include such forking nodes. For this reason, we develop several ways to model the performance of a  $JSQ + b$  node, and show how to modify the standard solution algorithm for open product-form networks to incorporate our results.

Unfortunately, an exact analysis of  $JSQ + b$  scheduling using standard techniques for continuous time Markov chains is not practical because of the complex structure of the associated state transition matrix. Therefore, we developed a variety of simplified models from which we can generate upper and lower bounds to the statistics of the original system. First, we present a one-dimensional generalized birth-death model to handle the case where the internal structure of the forking node is a single shared queue serving both outputs. Our one-dimensional model lumps together all states with equal numbers of customers present; the bounds are obtained by a novel method for assigning the aggregate service completion rate and routing probability to each lumped state. We then present a two-dimensional matrix geometric model to handle the case where the internal structure of the forking node includes separate queues for each output port. Our two-dimensional model reduces the size of the state space by dropping states with low occupancy probabilities; the bounds are obtained by modifying the state transitions that would have taken the system into a truncated state. Our two-dimensional model also includes a novel state-space truncation algorithm, which allows us to decide which states can be dropped while maintaining a specified level of accuracy. Both models can be solved to obtain closed-form expressions for the stationary probabilities. We compute bounds for several performance measures and compare the models. We also study the sensitivity of the overall delay to the routing bias, and provide a closed-form expression for optimizing the routing bias as a function of the delay difference between the two downstream paths.

Finally, we use the theory of Markov Decision Processes to determine the optimal routing policy for a forking node. Our analysis shows that a threshold routing is optimal, and the shape of the optimal switch curve is equivalent to  $JSQ + b$  scheduling for some value of the bias  $b$ .