ITC19/ Performance Challenges for Efficient Next Generation Networks 1365-1374
LIANG X.J. and XIN Z.H.(Editors)
V.B. IVERSEN and KUO G.S.(Editors)
Beijing University of Posts and Telecommunications Press

Does repacking improve performance of multiclass loss networks with overflow routing?

Lise George [1], Matthieu Jonckheere [1], and Lasse Leskelä [2]

[1] France Telecom R&D,
38-40 rue du general Leclerc, 92794 Issy-les-Moulineaux Cedex, France
email: {lise.george,matthieu.jonckheere}@francetelecom.com,
[2] Institute of Mathematics, Helsinki University of Technology,
P.O. Box 1100, FI-02015 TKK, Finland
email: lasse.leskela@iki.fi

Abstract. This paper considers multiclass loss networks with monoskill and multiskill servers and overflow routing. Accordingly, an arriving job is assigned to a corresponding vacant monoskill server, if possible. Otherwise the job is routed to a multiskill server, or rejected. We derive an efficiently computable upper bound for the system utilization by studying a modified routing where jobs can be redirected from multiskill to monoskill servers. This repacking policy improves performance also in terms of blocking probability when the service requirements of different job classes are statistically identical. Numerical simulations illustrate that our bounds provide good approximations for the performance of the original system.
Keywords: multiclass loss network, stochastic comparison, overflow routing, repacking

1 INTRODUCTION

In many communication systems, different services can be provided by a combination of monoskill servers, assigned to serving a certain class of jobs, and multiskill servers capable of dealing with all type of jobs. The routing of jobs is usually based on an overflow policy where an arriving job is preferentially routed to a corresponding monoskill server. If all monoskill servers corresponding to the class of the arriving job are busy, the job is routed to a multiskill server, if available, or rejected otherwise. This model fits several telecommunications applications such as call centers, streaming media, grid computing and wireless networks. Dimensioning has become an important economical issue in large call centers with specialized agents dedicated to diverse services [1–3]. As video-on-demand becomes more popular, new optimization methods are needed for streaming media servers that can be specialized in terms of clip size or delay constraints [4, 5]. In the field of grid computing, sharing and specialization of computational tasks between a set of distant workstations and supercomputers may produce considerable capacity gains. In wireless networks, a call can be served by different carriers; network operators may share a chunk of bandwidth for dealing with traffic surges. Dimensioning this type of systems is challenging because the