Efficient Content Delivery in the Presence of Impatient Jobs

M. Larrañaga\textsuperscript{1,2,6}, O. J. Boxma\textsuperscript{4}, R. Núñez-Queija\textsuperscript{3}, M. S. Squillante\textsuperscript{5}
\textsuperscript{1}CNRS, IRIT, 2 rue C. Carmichel, F-31071 Toulouse, France.
\textsuperscript{2}CNRS, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France
\textsuperscript{3}CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
\textsuperscript{4}Eurandom, Eindhoven University of Technology, P.O. Box 513, Eindhoven, The Netherlands
\textsuperscript{5}Mathematical Sciences Department, IBM Research, Yorktown Heights, NY 10598, USA
\textsuperscript{6}Univ. de Toulouse, INP, LAAS, F-31400 Toulouse, France

Abstract—We consider a content delivery problem in which jobs are processed in batches and may abandon before their service has been initiated. We model the problem as a Markovian single-server queue and analyze two different settings: (1) the system is cleared as soon as the server is activated, \(\mu = \infty\), and (2) the service speed is exponentially distributed with rate \(\mu > \infty\). The objective is to determine the optimal clearing strategy that minimizes the average cost incurred by holding jobs in the queue, having jobs renege, and performing set-ups. This last cost is incurred upon activation of the server in the case \(\mu = \infty\), and per unit of time the server is active otherwise.

Our first contribution is to prove that policies of threshold type are optimal in both frameworks. In order to do so we have used the Smoothed Rate Truncation method which overcomes the problem arising from unbounded transition rates. For our second contribution, we derive the steady-state job-length distribution under threshold policies. The latter yields a characterization of the optimal threshold strategy, which can be easily implemented.

Finally, we present numerical results for our solution across a wide range of parameters. We show that the performance of non-optimal threshold policies can be very poor, which highlights the importance of computing the optimal threshold.

The PhD fellowship of Maialen Larrañaga is funded by a research grant of the Foundation Airbus Group (http://fondation.airbus-group.com/). The research visit of Maialen Larrañaga in CWI and EURANDOM has been funded by EDSYS and SMI. The research of Onno Boxma is funded via the NETWORKS program of the Dutch government.