Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Discrete Algorithms 4 (2006) 255–276
JOURNAL OF DISCRETE ALGORITHMS
www.elsevier.com/locate/jda

Online competitive algorithms for maximizing weighted throughput of unit jobs

Francis Y.L. Chin [a,1], Marek Chrobak [b,2], Stanley P.Y. Fung [a,1], Wojciech Jawor [b,2], Jiří Sgall [c,*,3], Tomáš Tichý [c,3]

[a] *Department of Computer Science, The University of Hong Kong, Hong Kong*
[b] *Department of Computer Science, University of California, Riverside, CA 92521, USA*
[c] *Mathematical Institute, AS CR, Žitná 25, CZ-11567 Praha 1, Czech Republic*

Available online 13 April 2005

Abstract

We study an online unit-job scheduling problem arising in buffer management. Each job is specified by its release time, deadline, and a nonnegative weight. Due to overloading conditions, some jobs have to be dropped. The goal is to maximize the total weight of scheduled jobs. We present several competitive online algorithms for various versions of unit-job scheduling, as well as some lower bounds on the competitive ratios.

We first give a randomized algorithm RMIX with competitive ratio of $e/(e-1) \approx 1.582$. This is the first algorithm for this problem with competitive ratio smaller than 2.

Then we consider s-bounded instances, where the span of each job (deadline minus release time) is at most s. We give a 1.25-competitive randomized algorithm for 2-bounded instances, matching the known lower bound. We also give a deterministic algorithm EDF_α, whose competitive ratio on s-bounded instances is $2 - 2/s + o(1/s)$. For 3-bounded instances its ratio is $\phi \approx 1.618$, matching the known lower bound.

* Corresponding author.
E-mail addresses: chin@cs.hku.hk (F.Y.L. Chin), marek@cs.ucr.edu (M. Chrobak), pyfung@cs.hku.hk (S.P.Y. Fung), wojtek@cs.ucr.edu (W. Jawor), sgall@math.cas.cz (J. Sgall), tichy@math.cas.cz (T. Tichý).

[1] Supported by an RGC Grant HKU7142/03E.
[2] Supported by NSF grants CCR-9988360 and CCR-0208856.
[3] Partially supported by Institute for Theoretical Computer Science, Prague (project 1M0021620808 of MŠMT ČR), grant IAA1019401 of GA AV ČR and grant 201/05/0124 of GA ČR.

1570-8667/$ – see front matter
doi:10.1016/j.jda.2005.03.005

In s-uniform instances, the span of each job is exactly s. We show that no randomized algorithm can be better than 1.25-competitive on s-uniform instances, if the span s is unbounded. For $s = 2$, our proof gives a lower bound of $4 - 2\sqrt{2} \approx 1.172$. Also, in the 2-uniform case, we prove a lower bound of $\sqrt{2} \approx 1.414$ for deterministic memoryless algorithms, matching a known upper bound.

Finally, we investigate the multiprocessor case and give a $1/(1 - (\frac{m}{m+1})^m)$-competitive algorithm for m processors. We also show improved lower bounds for the general and s-uniform cases.

Keywords: Scheduling; Online algorithms; Buffer management

1. Introduction

The online *bounded delay buffer problem* has been recently introduced [2,14,15] to model the trade-offs arising in managing buffers for storing packets in QoS networks. In this formulation, packets arrive and are buffered at network switches. At each integer time step, only one packet can be sent along the link. Each packet is characterized by its QoS value, which can be thought of as a benefit gained by forwarding the packet. Network switches can use this QoS value to prioritize the packets. In order to control the end-to-end delay, each packet has also a deadline that specifies the latest time when the packet can be sent. In overload conditions, some packets will not be sent by their deadline. Such packets do not contribute to the benefit value, and can as well be dropped. The objective is to maximize the total value of the forwarded packets, that is the *weighted throughput*.

It is easy to see that this buffer management problem is equivalent to the following *unit-job scheduling* problem. We are given a set of n unit-length jobs, with each job j specified by a triple (r_j, d_j, w_j) where r_j and d_j are integral release times and deadlines, and w_j is a non-negative real weight. We have a single machine, i.e., one job can be processed at each integer time. We use the term *weighted throughput* or *gain* for the total weight of the jobs completed by their deadline. The goal is to compute a schedule for the given set of jobs that maximizes the *weighted throughput*. In Graham's notation, this problem can be described as $1|p_j = 1, r_j| \sum_j w_j U_j$.

In this paper we focus on the online version of unit-job scheduling, where each job arrives at its release time. At each time step, an online algorithm needs to schedule one of the pending jobs, without the knowledge of the jobs that will be released later in the future. An online algorithm $\mathcal{A}$ is called *R-competitive* if its gain on any instance is at least $1/R$ times the optimum (offline) gain on this instance. The smallest such value R is called the *competitive ratio* of $\mathcal{A}$. The competitive ratio is commonly used as a performance measure for online algorithms, and we adopt this measure in this paper.

For unit jobs, some restrictions on instances have been proposed in the literature [2,9,14,15]. In *s-bounded instances*, the span of the jobs (defined as the difference between the deadline and the release time) is at most s, and in *s-uniform instances* the span of each job is exactly s. In the context of QoS buffer management, these cases would correspond to models where, in order to reduce the end-to-end delay, only a small amount of delay is allowed at each node [14].