On Online Labeling with Polynomially Many Labels

Martin Babka[1], Jan Bulánek[1,2], Vladimír Čunát[1,⋆],
Michal Koucký[2,3,⋆⋆], and Michael Saks[4,⋆⋆⋆]

[1] Faculty of Mathematics and Physics, Charles University, Prague
[2] Institute of Mathematics, Academy of Sciences, Prague
[3] Department of Computer Science, Aarhus University
[4] Department of Mathematics, Rutgers University
{babkys,vcunat}@gmail, {bulda,koucky}@math.cas.cz, saks@math.rutgers.edu

Abstract. In the online labeling problem with parameters n and m we are presented with a sequence of n *keys* from a totally ordered universe U and must assign each arriving key a label from the label set $\{1, 2, \dots, m\}$ so that the order of labels (strictly) respects the ordering on U. As new keys arrive it may be necessary to change the labels of some items; such changes may be done at any time at unit cost for each change. The goal is to minimize the total cost. An alternative formulation of this problem is the *file maintenance problem*, in which the items, instead of being labeled, are maintained in sorted order in an array of length m, and we pay unit cost for moving an item.

For the case $m = cn$ for constant $c > 1$, there are known algorithms that use at most $O(n \log(n)^2)$ relabelings in total [9], and it was shown recently that this is asymptotically optimal [1]. For the case of $m = \theta(n^C)$ for $C > 1$, algorithms are known that use $O(n \log n)$ relabelings. A matching lower bound was claimed in [7]. That proof involved two distinct steps: a lower bound for a problem they call *prefix bucketing* and a reduction from prefix bucketing to online labeling. The reduction seems to be incorrect, leaving a (seemingly significant) gap in the proof. In this paper we close the gap by presenting a correct reduction to prefix bucketing. Furthermore we give a simplified and improved analysis of the prefix bucketing lower bound. This improvement allows us to extend the lower bounds for online labeling to the case where the number m of labels is superpolynomial in n. In particular, for superpolynomial m we get an asymptotically optimal lower bound $\Omega((n \log n)/(\log \log m - \log \log n))$.

Keywords: online labeling, file maintenance problem, lower bounds.

⋆ The first three authors gratefully acknowledge a support by the Charles University Grant Agency (grant No. 265 111 and 344 711) and SVV project no. 265 314.

⋆⋆ Currently a visiting associate professor at Aarhus University, partially supported by the Sino-Danish Center CTIC (funded under the grant 61061130540). Supported in part by GA ČR P202/10/0854, grant IAA100190902 of GA AV ČR, Center of Excellence CE-ITI (P202/12/G061 of GA ČR) and RVO: 67985840.

⋆⋆⋆ The work of this author was done while on sabbatical at Princeton University and was also supported in part by NSF under grant CCF-0832787.

L. Epstein and P. Ferragina (Eds.): ESA 2012, LNCS 7501, pp. 121–132, 2012.
© Springer-Verlag Berlin Heidelberg 2012

1 Introduction

In the online labeling problem with parameters n, m, r, we are presented with a sequence of n *keys* from a totally ordered universe U of size r and must assign each arriving key a label from the label set $\{1, 2, \ldots, m\}$ so that the order of labels (strictly) respects the ordering on U. As new keys arrive it may be necessary to change the labels of some items; such changes may be done at any time at unit cost for each change. The goal is to minimize the total cost. An alternative formulation of this problem is the *file maintenance problem*, in which the items, instead of being labeled, are maintained in sorted order in an array of length m, and we pay unit cost for moving an item.

The problem, which was introduced by Itai, Konheim and Rodeh [9], is natural and intuitively appealing, and has had applications to the design of data structures (see for example the discussion in [7], and the more recent work on cache-oblivious data structures [3,5,4]). A connection between this problem and distributed resource allocation was recently shown by Emek and Korman [8].

The parameter m, the *label space* must be at least the number of items n or else no valid labeling is possible. There are two natural range of parameters that have received the most attention. In the case of *linearly many labels* we have $m = cn$ for some $c > 1$, and in the case of *polynomially many labels* we have $m = \theta(n^C)$ for some constant $C > 1$. The problem is trivial if the universe U is a set of size at most m, since then we can simply fix an order preserving bijection from U to $\{1, \ldots, m\}$ in advance. In this paper we will usually restrict attention to the case that U is a totally ordered set of size at least exponential in n (as is typical in the literature).

Itai et al. [9] gave an algorithm for the case of linearly many labels having worst case total cost $O(n \log(n)^2)$. Improvements and simplifications were given by Willard [10] and Bender et al. [2]. In the special case that $m = n$, algorithms with cost $O(\log(n)^3)$ per item were given [11,6]. It is also well known that the algorithm of Itai et al. can be adapted to give total cost $O(n \log(n))$ in the case of polynomially many labels. All of these algorithms make no restriction on the size of universe U of keys.

For lower bounds, a subset of the present authors recently proved [1] a tight $\Omega(n \log(n)^2)$ lower bound for the case of linearly many labels and tight bound $\Omega(n \log(n)^3)$ for the case $m = n$. These bounds hold even when the size of the universe U is only a constant multiple of m. The bound remains non-trivial (superlinear in n) for $m = O(n \log(n)^{2-\varepsilon})$ but becomes trivial for $m \in \Omega(n \log(n)^2)$.

For the case of polynomially many labels, Dietz at al. [7] (also in [11]) claim a matching lower bound for the $O(n \log(n))$ upper bound. Their result consists of two parts; a lower bound for a problem they call *prefix bucketing* and a reduction from prefix bucketing to online labeling. However, the argument giving the reduction seems to be incorrect, and we recently raised our concerns with one of the authors (Seiferas), who agrees that there is a gap in the proof.

This paper makes the following contributions: