AARHUS UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

Univerzita Karlova v Praze
Matematicko-Fyzikální Fakulta

Evaluation of PhD Dissertation by Jan Bulánek

Department of Computer Science

Gerth Stølting Brodal

Associate professor

Date: 15 August 2014

Direct Tel.: +45 87162387
Mobile Tel.: +45 50595432
E-mail: gerth@cs.au.dk

Web: www.cs.au.dk/~gerth

Sender's CVR no.: 31119103

This is an evaluation of the PhD thesis submitted by Jan Bulánek on June 23, 2014 to the Matematicko-Fyzikální Fakulta, Univerzita Karlova v Praze.

The thesis is entitled "The Online Labeling Problem", and consists of 136 pages written in English. The thesis is based on following three publications, co-authored by Jan Bulánek:

- Jan Bulánek, Michal Koucký, and Michael E. Saks. *On Randomized Online Labeling with Polynomially Many Labels*. In Proceedings of the 40th International Colloquium on Automata, Languages and Programming, ICALP 2013. (Chapter 7)
- Jan Bulánek, Michal Koucký, and Michael E. Saks. *Tight Lower Bounds for the Online Labeling Problem*. In Proceedings of the 44th Symposium of Theory of Computation, STOC 2013. (Chapters 4,8, and 9)
- Martin Babka, Jan Bulánek, Vladimír Čunát, Michal Koucký, and Michael E. Saks. *On Online Labeling with Polynomially many Labels*. In Proceedings of the 20th Annual European Symposium on Algorithms, ESA 2012. (Chapter 6)

All these are leading international theoretical computer science conferences within the area of algorithms – STOC being one of the two worldwide top conferences in the area, and ICALP and ESA being the two top European conferences.

The thesis consists of a brief introduction to the labeling problem (Chapter 1), followed by two parts on upper bounds (Chapters 2-4) and lower bounds (Chapters 5-9). Each part again starts with an introduction (Chapters 2 and 5) with the relevant general context and definitions, and each chapter is then devoted to one case of the problem. Overall a well structured thesis.

The introduction in Chapter 1 is slightly brief, and for a thesis one would normally have expected that the applications would have been discussed in a little bit more detail. On the other hand, the focus of this thesis is the combinatorial intrinsics of the studied problem and not applications, so this is only a minor issue. The problem is important, and has e.g. applications in the context of maintaining a file in secondary memory under random insertions, where we still would like to support efficient scanning of ranges of the file.

The algorithmic problem studied throughout the thesis is the following: Given an online arriving sequence of n distinct numbers from a bounded universe of size r, online assign and reassign labels to these numbers from the label space $\{1,\ldots,m\}$, such that the ordering of the labels is consistent with the ordering of the numbers. Intuitively, smaller values of m will cause more relabelings of the elements to be necessary, whereas smaller values of r will limit the power of an adversary providing the online sequence of numbers. The exact mathematical relationship between these parameters is the topic of this thesis.

Chapter 3 present algorithms achiving an upper bound of $O(n \cdot \log^2 n/(\log m - \log n))$ for the case $4n/3 \leq m \leq n^c$, and $O(n \cdot \log^2 n \cdot \log (m/(m-n))$ for the case $n \leq m \leq 4n/3$. This generalizes previous bounds of $O(n \cdot \log^2 n)$ for $m=O(n)$ by Itai et al. 1981, and $O(n \cdot \log^3 n)$ for $m=n$ by Zhang 1993. The algorithms might not be that novel, but the interesting fact is that their performance is proved optimal by matching highly nontrivial lower bounds, shown in Chapters 8 and 9.

Chapter 4 presents upper bounds for the case of super-polynomial label spaces – more specifically, for $n^{\log n} \leq m \leq 2^n$, an algorithm is presented achieving $O(n \cdot \log n / \log\log m)$. No previous bounds were known for this interval of label sizes, likely because applications primarily need polynomial bounded label sizes. But again the interesting fact is that the achieved bounds are proved optimal by matching lower bounds in Chapter 6.

Chapter 5 gives an overview of the comprehensive list of lower bounds that follows in Chapters 6-9. Each of these sections contains descriptions of adversary constructions and comprehensive analysis of their enforced lower bounds.

Chapter 6 extends an $\Omega(n \cdot \log n)$ lower bound by Dietz et al. 2005 for polynomial sized label spaces to the label range $n^{1+\varepsilon}$ to 2^n, achieving $\Omega(n \cdot \log n/(1+\log\log m - \log\log n))$. The section also provides some missing details in the original paper by Dietz et al. Chapter 7 generalizes the lower bounds to also hold in the randomized case, interestingly without using Yao's min-max principle.

Chapter 8 proves tight lower bounds for the label range $(1+\varepsilon)n \leq m \leq n^{1+o(1)}$.

Chapter 9 contains lower bounds for the case of linear label spaces. Previous lower bounds by Dietz et al. assumed a certain behavior of the algorithms for the lower bound proofs to hold (the relabeling should be "smooth" – a property all existing algorithms share, and an assumption simplifying the proofs significantly). This chapter

AARHUS
UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

eliminates the need for such assumptions, making the lower bounds hold for all deterministic algorithms, and settling that sidestepping the smoothness assumption cannot allow for faster algorithms. Furthermore it is proved that restricting the range of the input numbers does not make the problem asymptotic simpler for linear sized label ranges.

In summary, the thesis of Jan Bulánek consists of a solid collection of coherent original results for an important fundamental algorithmic problem. The thesis is well structured, the results are highly nontrivial and are all mathematical well-founded and shows that the candidate masters the scientific methods in the field.

The thesis clearly fulfills the requirements for awarding the degree of PhD.

Sincerely,

Gerth Stølting Brodal
Associate professor
Aarhus University