Report on *The Online Labeling Problem* by Jan Bulánk

John Iacono
School of Engineering
New York University
Brooklyn, New York, USA

August 14, 2014

The thesis concerns the following problem: given a set of ordered data, and a one-dimensional place to store the data (such as an array) figure out the best way to store the data if it must be kept in order subject to insertions and deletions. If, for example, n elements of sorted data occupy the first n elements of an array, and a new minimum element is added, all n elements of the array must be moved in order to make room for the new element and preserve the order. However, a crucial thing allowed is that some array locations can be left vacant in order to provide room for future insertions without having to move all of the data. It is this tradeoff between the amount of vacant space in the array and the amount of data movement required for an insertion that is the primary subject of this thesis.

This problem is well-known and goes by two names: the *order maintenance problem* and the *list labeling problem*. The former name is usually used in the scenario described above where once actually stores data in an array with blank elements, and the latter is usually used when the primary interest is the numbered locations (the *labels*) that the data elements are assigned to.

The most relevant parameters are the amount of data (n), the size of the label space/array, (m). The classic decades-old results in this area are that if m is linear in n, then insertion takes $O(\log^2 n)$ amortized relabelings, and that if m is polynomial in n, then insertion takes $O(\log n)$ amortized relabelings.

In its obvious application, which is to maintain a dictionary, it is not a very efficient method, as other structures such a binary search trees perform better with logarithmic time and linear space using traditional algorithm analysis. They do so by not storing the data consecutively in memory, and in fact through an abstraction known as the the pointer model, largely ignore where in memory the data is.

However, recent attempts to model modern memory hierarchies on a computer, such as the *cache-oblivious model* view all of memory as a one dimensional array and the main design paradigm is to maintain data locality as much as possible. That is, memory locations that are accessed consecutively in an algorithm should be located close together in memory to the largest extent possible. When data needs to be added or removed, it should happen at (or between) specific memory locations to maintain locality and the online labeling problem is exactly the tool that is needed to do this. As such, it has become a standard subroutine in the development of cache-oblivious algorithms.

Having established the importance of the topic of the thesis and its applications to surrounding areas, I would like to turn my attention to the specific results.

The thesis contains the first real lower bounds for the problem. Some previous bounds existed, but they only applied to a restricted class of algorithms, called *smooth*. Whether or not, for example, it was possible to have a $o(\log^2 n)$ linear-space non-smooth algorithm was a major open question—I am aware of at least one serious attempt to show a $O(\log n)$ amortized algorithm for the linear-size case, something this thesis shows is impossible. This thesis provides tight lower bounds for almost all universe sizes, with the unresolved sizes not being the ones of primary interest. The lower bounds presented are non-trivial and a genuine scientific accomplishment. The main technique is to use an adversary argument, using the intuitive idea that by inserting into areas where there is little free space to move data, one can force any algorithm to perform badly. However, the details are subtle and required a number of insights to get to work.

This thesis is a clear scientific accomplishment. The problem addressed is of fundamental importance, and the thesis resolves almost all of the open questions in the area. The questions remaining are of secondary importance to those solved in the thesis.

The thesis demostrates the author's ability for creative scientific work.

John Iacono