



UNIVERSITÀ DI PISA

Seventh hands-on: Deterministic Data Stream

Algorithm Design (2021/2022)

Gabriele Pappalardo

Email: g.pappalardo4@studenti.unipi.it

Department of Computer Science

April 2022

1 Introduction

Consider a stream of n items, where items can appear more than once. The problem is to find the most frequently appearing item in the stream (where ties are broken arbitrarily if more than one item satisfies the latter). For any fixed integer $k \leq 1$, suppose that only k items and their counters can be stored, one item per memory word; namely, only $O(k)$ memory words of space are allowed.

Show that the problem cannot be solved deterministically under the following rules: any algorithm can only use these $O(k)$ memory words, and read the next item of the stream (one item at a time). You, the adversary, have access to all the stream, and the content of these $O(k)$ memory words: you cannot change these words and the past items, namely, the items already read, but you can change the future, namely, the next item to be read. Since any algorithm must be correct for any input, you can use any amount of streams and as many distinct items as you want.

Hints:

1. This is "classical" adversarial argument based on the fact that any deterministic algorithm A using $O(k)$ memory words gives the wrong answer for a suitable stream chosen by the adversary.
2. The stream to choose as an adversary is taken from a candidate set sufficiently large: given $O(k)$ memory words, let $f(k)$ denote the maximum number of possible situations that algorithm A can discriminate. Create a set of C candidate streams, where $C > f(k)$: in this way there are two streams S_1 and S_2 that A cannot distinguish, by the pigeon principle.

2 Solution

TODO.