



on Theoretical Aspects of Computer Science 2009 (Freiburg), pp. 51–62
-conf.org

LOWER BOUNDS FOR MULTI-PASS PROCESSING OF MULTIPLE DATA STREAMS

Institut für Informatik, Goethe-Universität Frankfurt am Main,
Robert-Mayer-Str. 11–15, D-60325 Frankfurt am Main, Germany
E-mail address: schweika@informatik.uni-frankfurt.de
URL: <http://www.informatik.uni-frankfurt.de/~schweika>

ABSTRACT. This paper gives a brief overview of computation models for data stream processing, and it introduces a new model for multi-pass processing of multiple streams, the so-called *mp2s-automata*. Two algorithms for solving the set disjointness problem with these automata are presented. The main technical contribution of this paper is the proof of a lower bound on the size of memory and the number of heads that are required for solving the set disjointness problem with *mp2s-automata*.

1. Introduction

In the basic data stream model, the input consists of a stream of data items which can be read only sequentially, one after the other. For processing these data items, a memory buffer of limited size is available. When designing data stream algorithms, one aims at algorithms whose memory size is far smaller than the size of the input.

Typical application areas for which data stream processing is relevant are, e.g., IP network traffic analysis, mining text message streams, or processing meteorological data generated by sensor networks. Data stream algorithms are also used to support query optimization in relational database systems. In fact, virtually all query optimization methods in relational database systems information about the number of distinct values of an attribute or the self-join size of a relation and these pieces of information have to be maintained while the database is updated. Data stream algorithms for accomplishing this task have been introduced in the seminal paper [2].

Most parts of the data stream literature deal with the task of performing one pass over a stream. For a detailed overview on algorithmic techniques for this scenario we refer to [23]. Lower bounds on the size of memory needed for solving a problem by a one-pass algorithm are usually obtained by applying methods from communication complexity (see, e.g., [2, 20]). In some problems it is known that the memory needed for solving the problem by a one-pass algorithm is at least linear in the size n of the input. For some of these problems one-pass algorithms can still compute good approximate answers while

Editorial Board:

- Susanne Albers
- Michael Mitzenmacher
- Madhavan Mukund
- Wolfgang Thomas
- V Vinay
- Pascal Weil (chair)
- Reinhard Wilhelm

STACS 2009
Symposium on Theoretical Aspects of Computer Science 2009 (Freiburg), pp. 51–62
<http://drops.dagstuhl.de/opus/volltexte/2009/1857>

Schloss Dagstuhl - Leibniz Center for Informatics has decided to establish a new series of non-commercial online conference proceedings called Leibniz International Proceedings in Informatics:

- New series of conference proceedings
- High-quality conferences
- All fields of informatics
- Copyright stays with the author
- Open access
online and free of charge for the reader
- Timely publication
- Low publication fee
kept to a minimum, due to a sharing of the workload between Dagstuhl and the conference organizers

The idea was initiated by the conferences STACS (stacs-conf.org) and FSTTCS (www.fsttcs.org).



Publications Portal:
drops.dagstuhl.de/lipics

Detailed Information:
www.dagstuhl.de/en/publications/lipics
www.dagstuhl.de/lipics_announce.pdf

Contact:
editorial board (chair): Prof. Pascal Weil
chair@lipics.dagstuhl.de
publication office: Dr. Jutta Huhse-Merz
office@lipics.dagstuhl.de



Leibniz
International
Proceedings in
Informatics –
LIPIcs

Leibniz International Proceedings in Informatics

Published by LZI
Schloss Dagstuhl



2009

26th International Symposium on Theoretical Aspects of Computer Science (STACS 2009)

Susanne Albers, Jean-Yves Marion (Eds.)

2008

IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2008)

Ramesh Hariharan and Madhavan Mukund and V Vina (Eds.)

25th International Symposium on Theoretical Aspects of Computer Science (STACS 2008)

Susanne Albers, Pascal Weil (Eds.)
