
Project Number 288094

# eCOMPASS

eCO-friendly urban Multi-modal route PlANning Services for mobile uSers

STREP
Funded by EC, INFSO-G4(ICT for Transport) under FP7

## eCOMPASS – TR – 032

# Applied Algorithms

P. Gupta and C. Zaroliagis

October 2013

**Lecture Notes in Computer Science**

The LNCS series reports state-of-the-art results in computer science research, development, and education, at a high level and in both printed and electronic form. Enjoying tight cooperation with the R&D community, with numerous individuals, as well as with prestigious organizations and societies, LNCS has grown into the most comprehensive computer science research forum available.

The scope of LNCS, including its subseries LNAI and LNBI, spans the whole range of computer science and information technology including interdisciplinary topics in a variety of application fields. The type of material published traditionally includes

- proceedings (published in time for the respective conference)
- post-proceedings (consisting of thoroughly revised final full papers)
- research monographs (which may be based on outstanding PhD work, research projects, technical reports, etc.)

More recently, several color-cover sublines have been added featuring, beyond a collection of papers, various added-value components; these sublines include

- tutorials (textbook-like monographs or collections of lectures given at advanced courses)
- state-of-the-art surveys (offering complete and mediated coverage of a topic)
- hot topics (introducing emergent topics to the broader community)

In parallel to the printed book, each new volume is published electronically in LNCS Online.

Detailed information on LNCS can be found at
**www.springer.com/lncs**

Proposals for publication should be sent to
LNCS Editorial, Tiergartenstr. 17, 69121 Heidelberg, Germany
E-mail: lncs@springer.com

ISSN 0302-9743

ISBN 978-3-319-04125-4

Lecture Notes in Computer Science
LNCS LNAI LNBI

springer.com

Gupta · Zaroliagis (Eds.)

LNCS 8321

LNCS
8321

Applied Algorithms

ICAA
2014

Prosenjit Gupta
Christos Zaroliagis (Eds.)

# Applied Algorithms

**First International Conference, ICAA 2014
Kolkata, India, January 2014
Proceedings**

ICAA

Springer

# Preface

This volume contains papers accepted for presentation at the International Conference on Applied Algorithms (ICAA 2014) held at the Heritage Institute of Technology, Kolkata, India, during January 13–15, 2014, together with the extended or short abstracts of 7 invited lectures given by Susanne Albers (Technical University of Munich, Germany), Bhargab Bhattacharya (Indian Statistical Institute, Kolkata, India), Gautam Das (University of Texas at Arlington, USA), Dimitrios Gunopulos (University of Athens, Greece), Rina Panigrahy (Microsoft Research, Mountain View, USA), Assaf Schuster (Technion, Israel Institute of Technology, Haifa, Israel), and Christos Zaroliagis (CTI & University of Patras, Greece).

ICAA is a new conference series with a mission to provide a quality forum for researchers working in applied algorithms. Papers presenting original contributions related to the design, analysis, implementation, and experimental evaluation of efficient algorithms and data structures for problems with relevant real-world applications were sought, ideally bridging the gap between academia and industry. Papers were solicited describing original research in a variety of areas including (but not limited to):

- Algorithmic Microfluidics
- Algorithms for VLSI CAD
- Analysis of Algorithms
- Approximation Algorithms
- Big Data Algorithms
- Cloud Computing
- Computational Advertising
- Computational Biology
- Computational Geometry
- Computational Services Science
- Computational Transportation Science
- Cryptography and Security
- Databases
- Data Mining
- Data Structures
- Distributed Algorithms
- Energy Efficient Algorithms
- External Memory Algorithms
- Graph Algorithms
- Graph Drawing
- Hardware Accelerated Algorithms
- Heuristic Search
- Image Processing

- Information Retrieval
- Location Based Services
- Machine Learning
- Parallel Algorithms
- Pattern Recognition
- Railway Optimization
- Randomized Algorithms
- Recommender Systems
- Robotics
- Spatial Informatics
- Social Network Analysis
- Web Intelligence
- Web Mining
- Web Searching

In response to the call-for-papers, 122 submissions from 9 countries were received. The Program Committee (comprising of 31 members from 7 countries) selected 21 papers for presentation. The criteria for selection were perceived originality, quality, and relevance to the subject area of the conference. Considerable effort was devoted to the evaluation of the submissions and to providing authors with helpful feedback. Towards this end, the Program Committee was assisted by 45 external reviewers.

We thank all those who submitted papers for consideration, as well as the Program Committee members and external reviewers for their invaluable contribution. We thank the management of the Heritage Institute of Technology and the Kalyan Bharathi Trust and the entire Organizing Committee for the excellent arrangements leading up to and during the entire conference. Finally, we gratefully acknowledge the generous financial support received from the TEQIP grant which made this conference a possibility.

January 2014 Prosenjit Gupta
Kolkata Christos Zaroliagis

# Organizing Committee

## Heritage Institute of Technology, Kolkata, India

### Advisors

| | |
|---|---|
| H.K. Chaudhary, Chairman | Kalyan Bharti Trust |
| P.K. Agarwal, CEO | Kalyan Bharti Trust |
| Probir Roy, Executive Director | Kalyan Bharti Trust |
| B.B. Paira, Advisor | Higher Education, Kalyan Bharti Trust |
| D.C. Ray, TEQIP-II Coordinator | Heritage Institute of Technology |
| S.N. Biswas, Deputy Director | Heritage Institute of Technology |
| Sukumar Ghosh, Professor | University of Iowa, USA |

### General Chairs

| | |
|---|---|
| Pranay Chaudhuri, Principal | Heritage Institute of Technology |
| Amitava Bagchi, Professor | Department of Computer Science and Engineering |

### Organizing Chairs

Kalarab Ray
Tapan Chakraborty
Subhashis Majumder

### Subcommittee Chairs

Satarupa Bagchi Biswas
Dinabandhu Bhandari
Arindam Chatterjee
Poulami Das
Aniruddha Dasgupta
Hemanta K. De
Anindita Kundu
Anindya Jyoti Pal
Shilpi Saha
Sujay Saha
Somenath Sengupta
Arvind Srivastava

## Local Hospitality

All faculty members and technical assistants of
The Department of Computer Science and Engineering and
The Department of Information Technology,
Heritage Institute of Technology, Kolkata, India.

# Abstarcts of Invited Talks

# Algorithmic Challenges in Digital Microfluidic Biochips: Protocols, Design, and Test*

Bhargab B. Bhattacharya[1,**], Sudip Roy[2,***], and Sukanta Bhattacharjee[1]

[1] Nanotechnology Research Triangle, Indian Statistical Institute, Kolkata 700108, India
{bhargab,sukanta_r}@isical.ac.in
[2] Dept. of CSE, Indian Institute of Technology Kharagpur, Kharagpur 721302, India
sudipr@cse.iitkgp.ernet.in

**Abstract.** Recent emergence of microfluidic technology has imparted a profound impact on the implementation of miniaturized healthcare chips and systems. In this review article, we will elaborate on several algorithmic challenges that arise while realizing biochemical protocols on a digital microfluidic (DMF) lab-on-a-chip. In particular, we will focus on certain design automation issues of sample preparation, dilution gradient generation, layout planning, and testing of DMF biochips.

* Invited Paper.

** The work of B. B. Bhattacharya was supported by a special grant to Nanotechnology Research Triangle from Indian Statistical Institute, Kolkata, India.

*** This work of S. Roy was supported by Microsoft Corporation and Microsoft Research India under the Microsoft Research India PhD Fellowship Award (2010-2014).

# Energy-Efficient Algorithms

Susanne Albers

Department of Computer Science
Technische Universität München
albers@in.tum.de

**Abstract.** We study algorithmic techniques for energy savings in computer systems. We consider power-down mechanisms that transition an idle system into low power stand-by or sleep states. Moreover, we address dynamic speed scaling, a relatively recent approach to save energy in modern, variable-speed microprocessors.

In the first part of the talk we survey important results in the area of energy-efficient algorithms. In the second part we investigate a setting where a variable-speed processor is equipped with an additional sleep state. This model integrates speed scaling and power-down mechanisms. We consider classical deadline-based scheduling and settle the complexity of the offline problem. As the main contribution we present an algorithmic framework that allows us to develop a number of significantly improved constant-factor approximation algorithms.
The material covered in this talk is contained in [1, 2].

## References

1. S. Albers. Energy-efficient algorithms. *Communications of the ACM*, 53(5):86–96, 2010.
2. S. Albers and A. Antoniadis. Race to idle: New algorithms for speed scaling with a sleep state. *Proc. 23rd Annual ACM-SIAM Symposium on Discrete Algorithms*, 1266–1285, 2012.

# Algorithms for Transport Optimization Theory and Practice

Christos Zaroliagis[1,2]

[1] Department of Computer Engineering & Informatics, University of Patras, 26504 Patras, Greece
[2] Computer Technology Institute & Press "Diophantus", Patras University Campus, 26504 Patras, Greece
zaro@ceid.upatras.gr

Public or private transport gives rise to several optimization problems, which are typically characterized by high complexity and sheer size, while some of them pose, in addition, real-time response constraints. Efficient algorithms can make a great difference towards an efficient and effective solution of such problems. In this talk, a few important algorithmic approaches are surveyed that are theoretically sound and practically efficient for the transport optimization problems they solve.

In the first part of the talk, robustness issues are investigated for the line planning problem in public transport under a specific uncertainty setting, motivated by recent market regulations in the railway sector. In this setting, a potentially large number of line operators, operating as commercial organizations, offer services to customers, while a central (typically state) authority manages the railway network infrastructure. The line operators act as competing agents for the exploitation of the shared infrastructure and are unwilling to disclose their true incentives (utility functions). The network manager wishes to set up a fair cost sharing scheme for the usage of the shared resources and to ensure the maximum possible level of satisfaction of the competing agents. The challenge is to provide a solution that is robust to the unknown incentives of the line operators, which are neither predictable or quantifiable nor statically describable. Towards this goal, a decentralized incentive-compatible mechanism is presented [1, 2] whose equilibrium point is (provably) the unknown social optimum. An accompanying experimental study of the aforementioned mechanism on both synthetic and real-world data shows fast convergence to the optimum. A wide range of scenarios is also explored, varying from an arbitrary initial state (to be solved) to small disruptions in a previously optimal solution (to be recovered). The experiments with the latter scenario show that the particular mechanism can be used as an online recovery scheme causing the system to re-converge to its optimum extremely fast.

In the second part of the talk, the route planning problem in large-scale road networks is investigated, focusing on two main issues: the efficient representation of such networks in a dynamic environment, and the computation (of not only a single optimal route but) of several source-to-destination alternative routes with specific quality characteristics.

To address the first issue, a new dynamic graph structure [3] is presented that is specifically suited for large-scale transportation networks providing simultaneously three unique features:

- *Compactness*: ability to efficiently access adjacent nodes or edges, a requirement set by all query algorithms in order to meet real-time response constraints.
- *Agility*: ability to change and reconfigure the graph's internal layout in order to improve the locality of the elements, according to a given algorithm.
- *Dynamicity*: ability to efficiently insert or delete nodes and edges.

The practicality and superiority of the new graph structure is demonstrated by an experimental study for shortest route planning in large-scale European and US road networks with a few dozen millions of nodes and edges. The particular structure is the first one that concerns the dynamic maintenance of a large-scale graph with ordered elements using a contiguous memory part, and which allows an arbitrary online reordering of its elements.

To address the second issue, improved methods are presented for computing a set of alternative source-to-destination routes in road networks in the form of an alternative graph [4]. The produced alternative graph is characterized by minimum path overlap, small stretch factor, as well as low size and complexity. Two existing approaches are surveyed and a new one is presented that improves upon those. An accompanying experimental study shows that the new approach can compute the entire alternative graph pretty fast even in continental size networks.

## References

1. A. Bessas, S. Kontogiannis, and C. Zaroliagis. Incentive-Compatible Robust Line Planning. In *Robust and Online Large-Scale Optimization*, Chapter 4, Springer 2009, pp. 85–118.
2. A. Bessas, S. Kontogiannis, and C. Zaroliagis. Robust Line Planning in case of Multiple Pools and Disruptions. In *Theory and Practice of Algorithms in Computer Systems* – TAPAS 2011, Lecture Notes in Computer Science Vol. **6595** (Springer 2011), pp. 33-44.
3. G. Mali, P. Michail, A. Paraskevopoulos, and C. Zaroliagis. A New Dynamic Graph Structure for Large-scale Transportation Networks. In *Algorithms and Complexity* – CIAC 2013, LNCS Volume 7878, pp. 312–323. Springer, 2013.
4. A. Paraskevopoulos and C. Zaroliagis. Improved Alternative Route Planning. In *Proc. 13th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems* - ATMOS 2013, OASICS Vol. 33, pp. 108–122.

# Mining and Analytics of Deep Web Repositories

Gautam Das

University of Texas at Arlington
gdas@uta.edu

**Abstract.** With the proliferation of deep web repositories (e.g., databases or document corpora) hidden behind proprietary web interfaces, e.g., keyword-/form-based search interfaces and hierarchical/graph-based browsing interfaces, efficient ways of exploring contents in such hidden repositories are of increasing importance in a wide variety of applications.

There are two key challenges: one on the proper understanding of interfaces, and the other on the efficient exploration, e.g., crawling, sampling and analytical processing, of very large repositories. In this talk, we focus on the fundamental algorithmic developments in the field, including web interface understanding, sampling, and data analytics over deep web repositories with various types of interfaces and containing structured or unstructured data. In the case of sampling, the objective is to draw sample tuples according to a pre-determined distribution over the repository. An example is simple random sampling which features uniform distribution over all documents/tuples in the deep web repository. The collected sample can be later used for analytical processing, mining, etc. In the case of analytics, the objective boils down to the ability to efficiently estimate aggregates (COUNT, SUM, MIN, MAX, etc.) over the deep web repository e.g., the total number of products in Amazons database, or the total number of documents indexed by Google. We shall discuss various recently developed efficient algorithms for sampling and aggregate estimation over such web repositories.

# Efficient Distance Measures for Social Networks

Rina Panigrahy

Microsoft Corp.
Mountain View, CA
rina@microsoft.com

**Abstract.** A fundamental operation in large graphs such as social networks is computing the distance between two nodes. While the shortest path graph distance is one standard measure of distance, there may be other candidates such as s-t min cuts, effective resistance (or commute time), and Katz distance [2] that may better capture the similarity between the users represented by the nodes. This gives rise to two lines of questioning: first, what is a good distance measure between nodes that captures user similarity, and second, can it be computed efficiently in real time? Since the standard shortest path algorithms are expensive to run on large graphs in real time, we study an approach [1] that estimates it in real time by moving the time-consuming shortest-path computation offline, and looking up only the precomputed values at query time and performing simple and fast computations on these precomputed values. More specically, during the offline phase we compute and store a small sketch for each node in the graph, and at query-time we look up the sketches of the source and destination nodes and perform a simple computation using these two sketches to estimate the distance. The algorithm is a modification of distance oracles by Thorup and Zwick [4] that samples a small set of seed nodes and only stores the nearest few seeds from each node as its sketch. Our experiments show that this algorithm is able to estimate distances in large graphs within a small additive error. We also study alternate distance measures that are easy to compute in real time and looks not only the shortest path distance but also takes into account the number of such paths. We also compare these different distance measures and evaluate which correlates most to user similarity [3].

## References

1. A. Das Sarma, S. Gollapudi, M. Najork, and R. Panigrahy. A sketch-based distance oracle for web-scale graphs. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 401–410. ACM, 2010.
2. L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.
3. R. Panigrahy, M. Najork, and Y. Xie. How user behavior is related to social affinity. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 713–722. ACM, 2012.
4. M. Thorup and U. Zwick. Approximate distance oracles. *Journal of the ACM (JACM)*, 52(1):1–24, 2005.

# Monitoring Distributed, Heterogeneous Data Streams: The Emergence of Safe Zones

Daniel Keren[1], Guy Sagy[2], Amir Abboud[2], David Ben-David[2], Assaf Schuster[2], Izchak Sharfman[2], and Antonios Deligiannakis[3]

[1] Department of Computer Science, Haifa University
[2] Faculty of Computer Science, Israeli Institute of Technology
[3] Department of Electronic and Computer Engineering, Technical University of Crete

**Abstract.** In many emerging applications, the data to be monitored is of very high volume, dynamic, and distributed, making it infeasible to collect the distinct data streams to a central node and process them there. Often, the monitoring problem consists of determining whether the value of a global function, which depends on the union of all streams, crossed a certain threshold. A great deal of effort is directed at reducing communication overhead by transforming the monitoring of the global function to the testing of *local* constraints, checked independently at the nodes. Recently, *geometric monitoring* (GM) proved to be very useful for constructing such local constraints for general (non-linear, non-monotonic) functions. Alas, in all current variants of geometric monitoring, the constraints at all nodes share an identical structure and are, thus, unsuitable for handling heterogeneous streams, which obey different distributions at the distinct nodes. To remedy this, we propose a general approach for geometric monitoring of heterogeneous streams (HGM), which defines constraints tailored to fit the distinct data distributions at the nodes. While optimally selecting the constraints is an NP-hard problem, we provide a practical solution, which seeks to reduce running time by hierarchically clustering nodes with similar data distributions and then solving more, but simpler, optimization problems. Experiments are provided to support the validity of the proposed approach.

# Table of Contents

## Abstracts of Invited Talks



## Invited Talks



## Contributed Talks