



Volume 1, Issue 9, September 2011

Data Warehousing: from Occasional OLAP to Real-time Business Intelligence (Dagstuhl Seminar 11361)	
<i>Markus Schneider, Gottfried Vossen, and Esteban Zimányi</i>	1
Uncertainty modeling and analysis with intervals: Foundations, tools, applications (Dagstuhl Seminar 11371)	
<i>Isaac E. Elishakoff, Vladik Kreinovich, Wolfram Luther, and Evgenija D. Popova</i>	26
Quantum Cryptanalysis (Dagstuhl Seminar 11381)	
<i>Serge Fehr, Michele Mosca, Martin Rötteler, and Rainer Steinwandt</i>	53
Public-Key Cryptography (Dagstuhl Seminar 11391)	
<i>Marc Fischlin, Anna Lysyanskaya, Ueli Maurer, and Alexander May</i>	76

ISSN 2192-5283

Published online and open access by

Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany.

Online available at <http://www.dagstuhl.de/dagrep>

Publication date

January, 2012

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

License

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported license: CC-BY-NC-ND.



In brief, this license authorizes each and everybody to share (to copy, distribute and transmit) the work under the following conditions, without impairing or restricting the authors' moral rights:

- Attribution: The work must be attributed to its authors.
- Noncommercial: The work may not be used for commercial purposes.
- No derivation: It is not allowed to alter or transform this work.

The copyright is retained by the corresponding authors.

Aims and Scope

The periodical *Dagstuhl Reports* documents the program and the results of Dagstuhl Seminars and Dagstuhl Perspectives Workshops.

In principal, for each Dagstuhl Seminar or Dagstuhl Perspectives Workshop a report is published that contains the following:

- an executive summary of the seminar program and the fundamental results,
 - an overview of the talks given during the seminar (summarized as talk abstracts), and
 - summaries from working groups (if applicable).
- This basic framework can be extended by suitable contributions that are related to the program of the seminar, e.g. summaries from panel discussions or open problem sessions.

Editorial Board

- Susanne Albers
- Bernd Becker
- Karsten Berns
- Stephan Diehl
- Hannes Hartenstein
- Frank Leymann
- Stephan Merz
- Bernhard Nebel
- Han La Poutré
- Bernt Schiele
- Nicole Schweikardt
- Raimund Seidel
- Gerhard Weikum
- Reinhard Wilhelm (*Editor-in-Chief*)

Editorial Office

Marc Herbstritt (*Managing Editor*)

Jutka Gasirowski (*Editorial Assistance*)

Thomas Schillo (*Technical Assistance*)

Contact

Schloss Dagstuhl – Leibniz-Zentrum für Informatik
Dagstuhl Reports, Editorial Office
Oktavie-Allee, 66687 Wadern, Germany
reports@dagstuhl.de

Digital Object Identifier: 10.4230/DagRep.1.9.i

www.dagstuhl.de/dagrep

Data Warehousing: from Occasional OLAP to Real-time Business Intelligence

Edited by

Markus Schneider¹, Gottfried Vossen², and Esteban Zimányi³

1 University of Florida, US, mschneid@cise.ufl.edu

2 Universität Münster, DE, vossen@uni-muenster.de

3 Université Libre de Bruxelles, BE, ezimanyi@ulb.ac.be

Abstract

This report documents the outcomes of the Dagstuhl Seminar 11361 “Data Warehousing: from Occasional OLAP to Real-time Business Intelligence”. In the past, data warehousing and analytical processing (OLAP) have produced fundamental and important technologies for the design, management, and use of information systems for decision support. Indeed, many industrial and administrative organizations have successfully used data warehouses to collect essential indicators that help them improve their business processes and their decision making efforts. Recent developments like column stores instead of row stores at the physical level, real-time data warehousing and Business Intelligence applications at the conceptual level, and decision support for new emerging applications have raised new research questions. This seminar has focused on the following five main topics: (i) Real-Time Data Warehouses and Business Intelligence, (ii) Spatio-Temporal Data Warehousing, (iii) Situational Business Intelligence, (iv) Query Processing in Data Warehouses, and (v) Knowledge Extraction and Management in Data Warehouses. These topics were discussed in parallel groups and each group identified open research issues and new challenges.

Seminar 4–9. September, 2011 – www.dagstuhl.de/11361

1998 ACM Subject Classification H.2.7 Database Administration – Data warehouse and repository.

Keywords and phrases Business Intelligence, Data Warehouses, OLAP, Spatio-temporal information, ETL, Service Orientation, Query optimization

Digital Object Identifier 10.4230/DagRep.1.9.1

1 Executive Summary

Markus Schneider

Gottfried Vossen

Esteban Zimányi

License  Creative Commons BY-NC-ND 3.0 Unported license
© Markus Schneider, Gottfried Vossen, Esteban Zimányi

This Dagstuhl Seminar brought together 37 researchers from 14 countries across disciplines that study data warehousing. The seminar can be seen as a successor of the Dagstuhl Perspectives Workshop 04321 “Data Warehousing at the Crossroads” (<http://www.dagstuhl.de/04321>) co-arranged by one of the organizers in 2004. After seven years (in 2011), we felt that it was time to take stock again, determine the status quo, and reflect on the future of data warehousing. Further, the seminar in 2004 was a perspectives workshop with the



Except where otherwise noted, content of this report is licensed under a Creative Commons BY-NC-ND 3.0 Unported license

Data Warehousing: from Occasional OLAP to Real-time BI, *Dagstuhl Reports*, Vol. 1, Issue 9, pp. 1–25

Editors: Markus Schneider, Gottfried Vossen, and Esteban Zimányi



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

	Monday	Tuesday	Wednesday	Thursday	Friday
09:00–10:30	Opening Short Introductions	Parallel Groups	Group Reports	Parallel Groups	Group Conclusions
10:30–11:00	Coffee break				
11:00–12:00	Short Introductions	Parallel Groups	IBM Industrial Presentation	Parallel Groups	Group Conclusions
12:00–14:00	Lunch				
14:00–15:30	Short Introductions	Parallel Groups	Excursion to Trier	Parallel Groups	
15:30–16:00	Coffee break			Coffee break	
16:00–17:30	Short Introductions	Parallel Groups		Parallel Groups	
18:00– . . .	Dinner				

■ **Figure 1** Timetable of the seminar

exclusive goal of predicting the possible future direction of data warehousing and OLAP. This seminar has a different scope and mainly dealt with current and recent research results. This does not exclude to look into the future and to determine what the field has achieved in the meantime.

The participants of the seminar were clustered around five thematic areas as follows

- Real-Time Data Warehouses and Business Intelligence,
- Spatio-Temporal Data Warehousing,
- Situational Business Intelligence,
- Query Processing in Data Warehousing, and
- Knowledge Extraction and Management in Data Warehouses.

This clustering was arranged prior to the seminar so that participants came to the seminar with short introductions describing both themselves and the research topics they work on, as well as identify challenging questions in these thematic areas. The whole Monday was devoted to these introductions.

Tuesday and Thursday were devoted to parallel working groups, each group discussing the state of the art in its thematic area and identifying the open questions. The organizers asked to each group to answer the following four questions

1. Identify most pressing research issues.
2. Is the topic purely industrial or also academic?
3. Where do you expect to be 5 years from now?
4. What would you want a new PhD student to work on?

Further, longer research presentations were given within each group for focusing the groups' work around a common approach. Each group presented a short ongoing report of the work realized in Wednesday morning, and a final presentation of their results on Friday morning.

It is worth noting that the Data Warehouse domain is both an active research domain but also drives a very intense activity in the industrial world. One objective of the seminar was also to establish bridges between these two communities. The seminar attracted participants from major companies active in the Data Warehouse domain (Sybase, HP, IBM, and EMC). Another concrete step in this respect was an industrial presentation given on Wednesday morning by Knut Stolze, from IBM Germany.

All five groups have committed to produce a paper to be published in a special issue of a journal. After discussion among the participants, the organizers started negotiating with

different journals about this possibility, and finally the journal chosen was the International Journal of Data Warehouse and Mining. It is expected that the papers will be submitted to the journal in January 2012 so that the publication of the special issue will be at the end of 2012.

2 Table of Contents

Executive Summary

<i>Markus Schneider, Gottfried Vossen, Esteban Zimányi</i>	1
--	---

Overview of Talks

Service Oriented BI	
<i>Alberto Abello</i>	7
Data Mining for Business Intelligence: from Relational to Graph representation	
<i>Marie-Aude Aufaure</i>	7
Open Research Areas in Data Warehouses	
<i>Cristina Ciferri</i>	8
Spatial Data Warehouses (SDW): Research Topics and Open Issues	
<i>Ricardo Ciferri</i>	8
My Current Research Interests	
<i>Maria Luisa Damiani</i>	9
Towards spatio-temporal datawarehouses	
<i>Maria Luisa Damiani</i>	9
Privacy and performance of cloud data warehouses	
<i>Jerome Darmont</i>	10
BI/OLAP research themes	
<i>Todd Eavis</i>	10
Knowledge Discovery & Management: My Point Of View	
<i>Cécile Favre</i>	10
Research Interests	
<i>Cécile Favre</i>	11
Column Stores and Query Processing for real-time BI	
<i>David Fekete</i>	11
Collaborative Business Intelligence	
<i>Matteo Golfarelli</i>	11
Research on DW, OLAP and Mining	
<i>Leticia I. Gomez</i>	12
Sequential OLAP	
<i>Christian Koncilia</i>	12
Data warehouse mining: what about mining query logs?	
<i>Patrick Marcel</i>	12
Open Business Intelligence	
<i>Jose-Norberto Mazon</i>	13
Mining Association Rules from Data Cubes	
<i>Rokia Missaoui</i>	13
BI Research Overview Gong Show	
<i>Torben B. Pedersen</i>	14

Efficient estimation of Joint Queries from Multiple OLAP Databases	
<i>Elaheh Pourabbas</i>	14
Spatial and Spatio-Temporal Data Warehousing	
<i>Markus Schneider</i>	14
IBM Smart Analytics Optimizer - Technical Introduction	
<i>Knut Stolze</i>	15
Business Intelligence 2.0: A general overview	
<i>Juan Trujillo</i>	15
Data Warehouse design and consistency through trace metamodels on a hybrid approach	
<i>Juan Trujillo</i>	16
Research in Data Warehousing, OLAP & Beyond	
<i>Alejandro Vaisman</i>	16
Near-Real-Time & Evolving ETL	
<i>Panos Vassiliadis</i>	17
Business Intelligence as a Service, in the Cloud	
<i>Gottfried Vossen</i>	17
Some not fully solved problems in data warehouse research	
<i>Robert Wrembel</i>	17
What is Spatio-Temporal Data Warehousing?	
<i>Esteban Zimányi</i>	18

Working Groups

Situational BI	
<i>Stefano Rizzi</i>	18
Situational BI Revisited	
<i>Stefano Rizzi</i>	19
On-demand self-service BI	
<i>Torben B. Pedersen</i>	19
Research Problems in Data Warehousing	
<i>Markus Schneider</i>	19
Query Processing in Data Warehouses	
<i>Anisoara Nica</i>	20
Query Processing in the Elastic Cloud	
<i>Anisoara Nica</i>	20
Knowledge Discovery and Management in Data warehouses	
<i>Rokia Missaoui</i>	20
Knowledge Discovery and Management in Data warehouses: Final report	
<i>Rokia Missaoui</i>	21
Right time BI framework	
<i>Robert Wrembel</i>	21

Open Problems

 Real-Time Data Warehouses and Business Intelligence 21

 Spatio-Temporal Data Warehouses 22

 Situational Business Intelligence 23

 Query Processing in Data Warehouses: Elasticity in Cloud Databases 23


 Knowledge Extraction and Management in Data Warehouses 24

Participants 25

3 Overview of Talks

3.1 Service Oriented BI


Alberto Abello (Universitat Politècnica de Catalunya – Barcelona, ES)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Alberto Abello

The traditional way to manage Information Technologies (IT) in the companies is having a data center, and licensing monolithic applications based on the number of CPUs, allowed connections, etc. This also holds for Business Intelligence environments. Nevertheless, technologies have evolved and today other approaches are possible. Specifically, the service paradigm allows to outsource hardware as well as software in a pay-as-you-go model. In this work, we will introduce the concepts related to this paradigm and analyze how they affect Business Intelligence (BI). We analyze to which extent it is possible to consider Business Intelligence just a service and use techniques like Cloud Computing, Service Oriented Architectures (SOA) and Business Process Modeling (BPM). Finally, we store the other way round. Since special attention must be paid to service companies' characteristics and how to adapt BI techniques to enhance services.

3.2 Data Mining for Business Intelligence: from Relational to Graph representation


Marie-Aude Aufaure (Ecole Centrale – Paris, FR)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Marie-Aude Aufaure

Traditional data mining methods, e.g. clustering, association rules, etc. are based on a tabular representation of input data and mainly ignore the relations between objects. Data, always growing, are mainly unstructured (80% distributed and come from heterogeneous data sources in an unpredictable way. Business Intelligence methods and tools need to take this big data challenge into account. Graphs are everywhere (social networks, web, biology, etc.) and can be considered as a way for managing structured, semi-structured and unstructured data. This structure can be used for traversing linked information, finding shortest paths, doing semantic partition, recommending and discovering potential interesting linked information. Methods and tools can be defined to exploit the graph structure of large repositories such as digital libraries, web, databases or data warehouses with their associated metadata. This presentation focuses on the extraction of graphs from relational databases and their aggregation. Open problems still remain: many graph models exist so the most appropriate one has to be chosen, how can we combine data mining methods with graphs algorithms to find communities that not only takes into account links between individuals, but also their similarities based on their own attributes, how can we summarize and aggregate these graphs and maintain their consistency.

3.3 Open Research Areas in Data Warehouses

Cristina Ciferri (University of Sao Paulo, BR)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Cristina Ciferri

In this talk, I highlight my main research areas of interest, especially those related to data warehousing. Regarding geographic data warehouses, I have proposed: (i) the SB-index and the HSB-index structures; (ii) the Spadawan and the Spatial Star Schema benchmarks; and (iii) a systematic way to investigate spatial data redundancy. Another research area of interest is the storage and recovery of images in data warehousing environments. I have been investigating how to store images and how to perform the ETL process, as well as how to handle aggregation levels and how to provide efficient query processing over image data warehouses. I also have interest in horizontal and vertical fragmentation of data warehouses, view materialization using OLAP signatures and GRASP, and analytical processing over XML documents.

Finally, other research topics include mining of medial data, data integration and provenance, and the proposal of index structures for biological databases, complex data such as image, audio and video, and time series data.

3.4 Spatial Data Warehouses (SDW): Research Topics and Open Issues

Ricardo Ciferri (University of Sao Carlos, BR)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Ricardo Ciferri


In this talk, I presented the main research topics in data warehouse area that the researchers of the Database Group (DBG) of the Federal University of São Carlos (UFSCar), Brazil, have been investigating. The main research topic is spatial data warehouse. For this topic, it was proposed techniques to improve SOLAP query processing performance over star schemas through specialized indices, such as the SB-index and the HSB-index. Also, an improved strategy to process drill-across SOLAP queries was briefly described in this talk. The performance evaluation of SOLAP query processing also was investigated through the proposal of two benchmarks:

The Spadawan Benchmark and The Spatial Star Schema Benchmark.

Nowadays, the Database Group of the Federal University of São Carlos (Brazil) is investigating data warehouses that store vague spatial objects and spatio-temporal data warehouse performance.

3.5 My Current Research Interests


Maria Luisa Damiani (Università di Milano, IT)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Maria Luisa Damiani

This brief contribution is to illustrate my area of research and some research questions related to the use of spatial data warehouses. My current research is carried out in two main areas, both related to spatial and spatio-temporal data management. The first addresses issues related to location-based security and location privacy, i.e., how to use location information to regulate the access to sensitive resources and how to preserve the privacy of personal position data, respectively. The second line is concerned with spatio-temporal data modeling, in particular the conceptual modeling of spatial datawarehouse and the recent paradigm of semantic trajectories, i.e., how to enrich the movement information with semantics. Spatial data warehouses have a potential in all these applications. Spatial data warehousing has been a “promising” area of research since early 2000. The key question is whether this promise has turned into a concrete result. While map visualization is extensively used in conventional data warehouses, it is still not clear whether any real progress has been made in the direction of building real and large scale spatial data warehouses. Moreover, it is not clear whether the paradigm founded on the notion of dimension measure and fact is sufficient to deal with the specific features of novel spatio-temporal representations like semantic trajectories. These are questions that are worth being discussed.

3.6 Towards spatio-temporal datawarehouses


Maria Luisa Damiani (Università di Milano, IT)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Maria Luisa Damiani

The widespread use of positioning technologies makes it relatively easy to track objects across indoor and outdoor settings. Large amounts of data on the traces of entities can be easily collected, possibly reporting not only their time varying positions but also information about the context in which the movement takes place, for example acquired through sensors. In this way the behavior of moving entities can be more easily analyzed. An emerging paradigm for the representation of the behavior of moving entities is that of semantic trajectories. A semantic trajectory is a trajectory which is temporally bounded (i.e., has a start and an end) and which represents the movement carried out by an object for some specific purpose. A semantic trajectory is a sequence of stops and moves where both stops and move are annotated semantically, for example a stop can represent a railway station while the move can be a particular train line. The question we pose is how to obtain aggregated information out from large bulks of semantic trajectories so as to build models of collective behavior. This calls for the definition of novel OLAP algebras defined over sequences of multidimensional data and integrating visual analytics functionalities.

3.7 Privacy and performance of cloud data warehouses

Jerome Darmont (Université Lumière Lyon 2, FR)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Jerome Darmont

We address in this presentation two critical issues in cloud BI. With respect to data privacy and availability, we distribute data over several cloud service providers (CSP) using Shamir's Secret Sharing cryptographic scheme so that:

(1) each CSP cannot exploit the data it stores; (2) a majority of CSP need to band together to break the secret, or fail at the same time to render data unavailable. With respect to query performance, we propose to balance the increase of CPU power (and cost) with classical performance optimization data structures, i.e., materialized views (at a storage cost). Our aim being to find the best cost tradeoff, we designed cost models that help output an optimal configuration of CPU instances and materialized views, thus providing flexible control over both power and budget.

3.8 BI/OLAP research themes


Todd Eavis (Concordia University – Montreal, CA)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Todd Eavis

In this talk I provide an overview of a number of current research issues in the OLAP/BI domain. Topics include: Parallel Computing, Indexing, query ptimization compression, exploitation of GPUs, OLAP languages, native language (Java) querying, Query rewriting, Web-based interfaces, drag and drop BI queries, and conceptual modelling.

3.9 Knowledge Discovery & Management: My Point Of View


Cécile Favre (Université Lumière Lyon 2, FR)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Cécile Favre
URL http://eric.univ-lyon2.fr/~cfavre/DAGSTUHL/-Favre_DAGSTUHL_KnowledgeDiscovery&Management_MyPointOfView.pdf

This talk was prepared in order to begin the discussion in the group. This includes (1) a first list for using Knowledge Discovery and Management in Data Warehouses, (2) the detail about my own works concerning the topics of the group, (3) the links with the other groups, and at the end (4) other issues for the group.

3.10 Research Interests


Cecile Favre (Université Lumière Lyon 2, FR)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Cecile Favre

This is an overview of my past, current and future research work related to datawarehousing. This includes Personalization in Data Warehouses, Data Warehouse Modeling and the combination of Social Networks and Data Warehouses.

3.11 Column Stores and Query Processing for real-time BI

David Fekete (Universität Münster, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© David Fekete

Column Stores, as column-based Database Management Systems (DBMS), have come to be often regarded as alternative to typical row-based DBMS (Row Stores) in Data Warehousing (DWH) and Business Intelligence (BI) scenarios. Column Stores promise to improve read-performance in those read-intensive, analytical (On-Line Analytical Processing, OLAP) scenarios significantly. But especially query processing is an integral part in realizing this speed-up potential and subject to many optimization techniques, like Block Iteration, Late Materialization or Compression. A query executor has to utilize the columnar database structure to enable the promised performance advantages. On the way from “occasional OLAP” to “real-time” BI there are several possible issues with regard to query processing. Those include the contribution of these Column Store technologies to real-time BI or the impact of novel hardware technologies, such as Solid State Drives (SSD) and multi-core CPUs.

3.12 Collaborative Business Intelligence


Matteo Golfarelli (Università di Bologna, IT)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Matteo Golfarelli

Cooperation is seen by companies as one of the major means for increasing flexibility and innovating. Business intelligence (BI) platforms are aimed at serving individual companies, and they cannot operate over networks of companies characterized by an organizational, lexical, and semantic heterogeneity. In this presentation we summarize different approaches to collaborative BI and we propose a framework, called Business Intelligence Network (BIN), for sharing BI functionalities over complex networks of companies that are chasing mutual advantages through the sharing of strategic information. A BIN is based on a network of peers, one for each company participating in the consortium. Peers are equipped with independent BI platforms that expose some querying functionalities aimed at sharing business information for the decision-making process. After proposing an architecture for a BIN, we outline the main research issues involved in its building and operating.

3.13 Research on DW, OLAP and Mining


Leticia I. Gomez (Instituto Tecnológico de Buenos Aires, AR)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Leticia I. Gomez

We discuss different challenges in Business Intelligence. Traditional Data Warehouse and OLAP tools store and manipulate strings, numbers and timestamp data types in a native way. Incorporating new data types in DW enriches the capability of analysis. For instance, spatio-temporal data recorded by GPS and RFID sensors can be used for discovering trajectories hidden patterns. In addition, as new types of data are incorporated into DW and Mining, novel techniques have to be proposed related to physical and logical design of data as well as new forms of human interaction. We showed briefly the idea of some related projects.

3.14 Sequential OLAP

Christian Koncilia (Universität Klagenfurt, AT)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Christian Koncilia

Although Business Intelligence is still a hot topic in computer science, little effort has been put on how to analyze sequences in data warehouses.

In my talk I presented a novel approach which enables the user to analyze sequential data within a standard OLAP environment. This approach consists of two core functions (SOLAP() and RTA() where RTA is short for “Real Time Axis”) which both return a standard OLAP cube. Hence, the user can still use all the well-known OLAP operations like drill-down, roll-up, slice, etc. to analyze the cube. Furthermore, this approach may be applied to all data warehousing solutions.

3.15 Data warehouse mining: what about mining query logs?

Patrick Marcel (Université de Tours – Blois, FR)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Patrick Marcel

This presentation gives a brief overview of the researches lead at the computer science lab of Université François Rabelais Tours on mining OLAP query logs.


Two examples are presented. In the first one, a joint work with Università di Bologna, association rules are extracted from a user’s query log for personalising the user’s queries on the fly.

In the second one, a multi user log is analysed to recommend queries for supporting discovery driven analyses.

Some research perspectives around OLAP log mining are underlined.

3.16 Open Business Intelligence

Jose-Norberto Mazon (Univ. of Alicante, ES)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Jose-Norberto Mazon


Citizens demand an increasingly transparent behavior of public institutions.

Importantly, transparency implies that public data should be available with the aim of providing the greatest benefit to the wider society through an active participation of citizens. Therefore, public data should be freely available to be easily used, reused and redistributed by anyone, i.e. open data. Consequently, open data are generally shared as a raw data which, unfortunately, prevents non-expert citizens from analyzing them to acquire actionable information. Mechanisms that allow citizens to analyze and understand open data in a user-friendly manner are thus highly required.

To this aim, the concept of Open Business Intelligence (OpenBI) is introduced. OpenBI facilitates non-expert users to (i) analyze and visualize open data, thus generating actionable information by means of reporting, OLAP analysis, dashboards or data mining; and to (ii) share the new acquired information as open data to be reused by anyone. As a consequence, OpenBI requires the development of systematic approaches for guiding non-expert users in obtaining and sharing the most reliable knowledge from the available open data. Finally, it is worth noting that OpenBI has to deal with several research challenges: the extraction and integration of heterogeneous open data sources by non-expert users (as citizens), tackling data quality issues in a user-friendly manner, intuitive data visualization, smooth integration of unstructured data, and so on.

3.17 Mining Association Rules from Data Cubes

Rokia Missaoui (Université du Québec en Outaouais, CA)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Rokia Missaoui

Substantial work has been conducted in knowledge discovery from data warehouses. This includes (but is not limited to) outlier detection in multidimensional data, cubegrade generation, constrained gradient analysis, association rule mining, and discovery-driven examination of cubes. To the best of our knowledge, Kamber et al. were the first researchers who addressed the issue of mining association rules (ARs) from multidimensional data. They introduced the concept of metarule-guided mining which consists in using rule templates defined by users in order to guide the mining process.

The objective of the present talk is to show that triadic concept analysis can be used as a formal data mining framework to generate clusters (closed trisets) and triadic association rules from data cubes in an efficient and meaningful way. The merits of triadic association rules over dyadic ones lie in the fact that they represent patterns in a more compact and meaningful way than association rules extracted from one of the possible flattened forms of the data cube.

3.18 BI Research Overview Gong Show


Torben B. Pedersen (Aalborg University, DK)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Torben B. Pedersen

This talk will introduce myself and my research within BI. I have worked on real-time DW and BI, spatio-temporal DWs, ETL frameworks and testing, compressed bitmaps, On-demand integration of cubes and XML, multidimensional schema discovery, contextualized warehouses, semantic web warehousing, and mining of data cubes.

3.19 Efficient estimation of Joint Queries from Multiple OLAP Databases

Elaheh Pourabbas (National Research Council – Rome, IT)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Elaheh Pourabbas

In this study, we investigate the problem of estimation of a target database from summary databases derived from a base data cube. We show that such estimates can be derived by choosing a primary database with the desired target measure but not the desired dimensions, and use a proxy database to estimate the results. The technique we use is linear indirect estimation, commonly used for statistical estimation. We examine two obvious computational methods for computing such a target database, called the “Full cross product” (F) and the “Pre-aggregation” (P) methods. We study the accuracy and computational cost of these methods. While the F method provides a more accurate estimate, it is more expensive computationally than P. Our contribution is in proposing a third new method, called the “Partial Pre-aggregation” method (PP), which is significantly less expensive than F, but is just as accurate. We prove formally that the PP method yields the same results as the F method, and provide analytical and experimental results on the accuracy and computational benefits of the PP method. Then, we consider the problem of how to select a primary and a proxy database given that there are multiple primary databases available with the same measure and multiple proxy databases with the desired target dimensions in order to get the most accurate estimation results.

3.20 Spatial and Spatio-Temporal Data Warehousing

Markus Schneider (University of Florida – Gainesville, US)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Markus Schneider

Traditional data warehousing and OLAP deal with decision support for alphanumeric data. However, new emerging applications produce new kind of data categories like spatial, spatiotemporal, and biological data. The question is how these data can be made available and leveraged for decision support. The nature of these new data categories is that they lead to complex (variable-length, multi-structured, and hierarchical) data objects and that aggregate operations on them are currently not supported. Spatial data warehousing and Spatial OLAP

(SOLAP) are promising extensions with respect to spatial decision support. A spatial data warehouse is a full-fledged data warehouse with additional capabilities to store, retrieve, query, and analyze spatial data represented as values of spatial data types. In addition, spatial data collections can be spatial connectivity structures like spatial partitions (maps) and spatial graphs (spatial networks) that satisfy certain topological constraints. Spatial OLAP operations are full-fledged OLAP operations that include spatial objects as values of spatial data types in dimensions and measures and that make use of spatial aggregation operations like spatial union, spatial intersection, convex hull, centroid, nearest neighbor, and contour. These features require a basic logical data model for formally defining and integrating needed spatial data warehouse and OLAP concepts. Since spatial data types are abstract data types that hide the internal representation of geometries, the architecture and infrastructure of existing data warehousing and OLAP technology can be used. Extensibility mechanisms at all levels of the architecture of a data warehouse system and an OLAP system are needed to be able to integrate new data categories and aggregation operations. At the conceptual level, a user-centric modeling strategy is needed that abstracts from the logical level and emphasizes user considerations. Unfortunately, a standardized conceptual model for multidimensional data modeling as well as a standardized query interface for data warehouses do not exist. Our main idea is to take the cube metaphor literally as a conceptual user model and to design cube types as abstract data types. An analysis or query language then provides high-level OLAP operations like roll-up, drill-down, slice, etc.

3.21 IBM Smart Analytics Optimizer - Technical Introduction


Knut Stolze (IBM Deutschland – Böblingen, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Knut Stolze

The IBM Smart Analytics Optimizer (ISAO) is a new product designed to enhance and improve IBM's relational database system DB2 for z/OS for data warehouse and BI applications. ISAO builds on several new technologies to dramatically improve query performance for long-running queries. All data is stored in main memory and advanced processor features like SIMD operations are exploited. The presentation gave an overview on the integration aspects for ISAO into DB2 for z/OS. The architecture and data flows for query processing and data loading were explained on a high-level as well.

3.22 Business Intelligence 2.0: A general overview

Juan Trujillo (Univ. of Alicante, ES)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Juan Trujillo


Business Intelligence (BI) solutions allow decision makers to query, understand, and analyze their business data in order to make better decisions.

However, as the technology and society evolve, faster and better informed decisions are required. Nowadays, it is not enough to use only the information from the own organization and making isolated decisions, but rather requiring also to include information present in the web like opinions or information about competitors, while using collective intelligence,

collaborating through social networks, and supporting the BI system with cloud computing. In response to this situation, a vision of a new BI 2.0, based on the evolution of the web and the emerging technologies, arises. However, different researchers differ in their vision of this BI evolution. In this talk, we provide an overview of the aspects proposed to be included in BI 2.0. We describe which success factors and technologies have motivated each aspect. Finally, we review how tool developers are including these new features in the next generation of BI solutions.

3.23 Data Warehouse design and consistency through trace metamodels on a hybrid approach


Juan Trujillo (Univ. of Alicante, ES)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Juan Trujillo

Data warehouses (DW) integrate several heterogeneous data sources in multidimensional structures in support of the decision-making process. The complexity of the DW development process requires to follow a methodological approach in order to be successful. Mainly three approaches have been proposed to tackle this problem in a similar way to software product development, bottom-up, top-down, and an hybrid approach. The hybrid approach, being the one with the most advantages, still suffers from a drawback, requirements and data sources must be reconciliated, accommodating either one or both, to a new DW model. In this process, relationships between requirements, data sources and conceptual elements is lost, since no traceability is included. In turn, this hurts the requirements validation, lowering the user satisfiability (since the validation is not done until the DW is implemented), makes the derivation of the platform specific models incomplete (since currently data types are not included at the conceptual level in DW models), and increases the complexity of the design extraction, transformation and load processes (since the initial relationships between the source tables and the target ones could have been used but was lost). In this paper, we review the proposals for traceability in the Requirements Engineering (RE) and Model Driven Development (MDD) fields and propose a metamodel and a set of Query/View/Extraction (QVT) transformations to include traceability in the DW field, solving the aforementioned problems and enabling impact change analysis in an easy way, increasing the user satisfaction.

3.24 Research in Data Warehousing, OLAP & Beyond

Alejandro Vaisman (Université Libre de Bruxelles, BE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Alejandro Vaisman

In this presentation I summarize my research in the field of Data Warehousing and OLAP, and present current and future research directions I am aimed at following. With respect to the former, I briefly discuss my work on

1. Dimension updates in OLAP;
2. Temporal OLAP;
3. GIS and OLAP integration;
4. Trajectory Mining.

Regarding future research directions, I will comment my work on Spatio-Temporal SOLAP (specifically the temporal extension of the Piet data model), the extension of SOLAP to support Raster data, and my work on RDF and the semantic web, in particular, the need to support semantic web BI.

3.25 Near-Real-Time & Evolving ETL


Panos Vassiliadis (University of Ioannina, GR)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Panos Vassiliadis

This presentation refers to two important topics for the area of ETL: near real time ETL and evolution of ETL scenarios. The talk discusses the area and then moves on to the presenter's high-level point of view on how the problems should be handled. In the end, there is a quick discussion of other alternatives, too.

3.26 Business Intelligence as a Service, in the Cloud


Gottfried Vossen (Universität Münster, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Gottfried Vossen

Business intelligence (BI) has long been provided by tools and software systems that needed local installation. More recently, efforts have successfully been launched to provide BI as a service of the Internet, thereby making BI amenable to cloud computing. We briefly review why this is relevant, what potential drawbacks and problem areas are, and which questions seem most relevant to consider.

3.27 Some not fully solved problems in data warehouse research

Robert Wrembel (Poznan University of Technology, PL)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Robert Wrembel

Despite over 20 years of conducted research in the data warehousing area, there exist several either only partially solved or unsolved research problems. Some of them, which are being researched at the Institute of Computing Science at the Poznan University of Technology, are outline in this presentation. These problems include: (1) handling the evolution of ETL, (2) bitmap compression algorithms on GPUs, and (3) indexing dimension data. The first problem is caused by the evolution of external data sources used for delivering data into a data warehouse. These sources change not only its content but also its structures. Structural changes impact all the layers in a data warehouse architecture. One of the layers includes ETL. In the project that we conduct, we try to develop a framework for semi-automatic or automatic (if possible) adjustment of the ETL layer to structural changes in data sources. The second problem concerns the application of Graphic Processing Units to data processing.

The processing power of GPUs and massive processing parallelism that can be achieved offers a promising framework. In our work, we port various bitmap index compression algorithms to the GPU platform. The third problem concerns the development of efficient indexing techniques for dimensions. In our work we exploit the fact that the most frequent queries in a data warehouse join fact a table with the Time dimension and the fact that dimensions have often a hierarchical structure. So far, we have proposed an index, called Time-HOBI that eliminates the need for joining a fact table with its Time dimension.

3.28 What is Spatio-Temporal Data Warehousing?

Esteban Zimányi (Université Libre de Bruxelles, BE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Esteban Zimányi


In the last years, extending OLAP (On-Line Analytical Processing) systems with spatial and temporal features has attracted the attention of the GIS (Geographic Information Systems) and database communities. However, there is no a commonly agreed definition of what is a spatio-temporal data warehouse and what functionality such a data warehouse should support.

Further, the solutions proposed in the literature vary considerably in the kind of data that can be represented as well as the kind of queries that can be expressed. In this presentation I present a conceptual framework for defining spatio-temporal data warehouses using an extensible data type system. This is based on a taxonomy of different classes of queries of increasing expressive power.

4 Working Groups

4.1 Situational BI


Stefano Rizzi (University of Bologna, IT)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Stefano Rizzi

When decisions have to be made quickly and under uncertainty in complex and dynamic environments, the selection of an action plan requires external business variables coupled with variables on company performance. In situational BI, data from a company's data warehouse are spontaneously correlated with “external” information sources that may come from the corporate intranet, be acquired from some external vendor, or be derived from the internet.

4.2 Situational BI Revisited


Stefano Rizzi (University of Bologna, IT)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Stefano Rizzi

The research group in charge of the situational BI topic describes a typical interaction scenario and the related research challenges.

4.3 On-demand self-service BI


Torben B. Pedersen (Aalborg University, DK)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Torben B. Pedersen

This talk summarizes the work of the Dagstuhl group on on-demand self-service BI. It answers the five questions put forward by the organizers and presents the paper skeleton of the paper to be written by the group.

4.4 Research Problems in Data Warehousing

Markus Schneider (University of Florida – Gainesville, US)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Markus Schneider

Data warehousing and online analytical processing (OLAP) have produced fundamental and important technologies for the design, management, and use of information systems for decision support. This is manifested by their acceptance in many industrial and administrative organizations that have successfully leveraged these technologies to improve their business processes and their decision making efforts. However, these technologies tend to be largely system-centric, that is, system aspects at the logical and physical levels are emphasized and be visible to the user. Users like managers or system analysts are usually supported by sophisticated GUI tools.

But these tools have a logical model like a snowflake schema or a constellation schema as their basis which make it difficult to understand the data. Therefore, a user-centric modeling strategy is needed at the conceptual level that focuses on an appropriate user view on data warehouses and OLAP operations. In general, a consensus has to be achieved for a standardized conceptual model for multidimensional data. The modeling framework to be developed (and consequently the whole data warehouse architecture as well as the OLAP operations) should be extensible so that new kinds of data categories like spatial, spatiotemporal, and biological data can be made available and used for decision support. The nature of these new data categories is that they lead to complex (variable-length, multi-structured, and hierarchical) data objects and aggregate operations on them that are currently not supported. Finally, a standardized query interface is needed both at the conceptual level and at the logical level.

4.5 Query Processing in Data Warehouses

Anisoara Nica (Sybase, An SAP Company, CA)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Anisoara Nica

Joint work of Nica, Anisoara; Eavis, Todd; Fekete, David; Graefe, Goetz; Neumann, Thomas; Petrov, Ilia; Pourabbas, Elaheh; Stolze, Knut

This talk is the first presentation of the Dagstuhl group on “Query Processing in Data Warehouses”. It answers the five questions put forward by the organizers and summarizes the most relevant research problems in query processing in the data warehouse environment. This includes query processing in the elastic cloud as well as parallel query processing.

4.6 Query Processing in the Elastic Cloud

Anisoara Nica (Sybase, An SAP Company, CA)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Anisoara Nica

Joint work of Eavis, Todd; Fekete, David; Graefe, Goetz; Neumann, Thomas; Nica, Anisoara; Petrov, Ilia; Pourabbas, Elaheh; Stolze, Knut

This talk is the second presentation of the Dagstuhl group on “Query Processing in Data Warehouses”. It revises the answers to the five questions put forward by the organizers, and it addresses one of the relevant research issues discussed by the group, namely query processing in the elastic cloud. A central promise of cloud services is elastic, on-demand provisioning. For data-intensive services such as data management, growing and shrinking the set of nodes implies copying data to nodes with temporary membership in a service. At best, a node might retain its copy of the data while it provides another service; at worst, a node that rejoins the database service requires a new copy of data. Many solutions have been proposed to the elasticity for cloud services, but few address the problems raised by “sometimes available copy” nodes. The talk proposes a new approach to elasticity for the environments where sometimes available copies can be used by permanent nodes for query processing.

4.7 Knowledge Discovery and Management in Data warehouses


Rokia Missaoui (Université du Québec en Outaouais, CA)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Rokia Missaoui

The presentation gives the preliminary view of the KDMD group about integrating data, information, and knowledge in data warehouses in a user-centric manner in order to help the end-user get better benefits from data by allowing (i) a unified and meaningful representation of multidimensional data as well as knowledge patterns, and (ii) advanced query mechanisms and guidance to get targeted information and knowledge through information retrieval and data mining techniques.

4.8 Knowledge Discovery and Management in Data warehouses: Final report


Rokia Missaoui (Université du Québec en Outaouais, CA)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Rokia Missaoui

The presentation focuses on some of the seven specific objectives stated in the Wednesday talk by the KDMD group. In particular, a user-centric architecture is presented with some illustrations of advanced OLAP query processing such as cooperative and intensional query management using both data and knowledge.

4.9 Right time BI framework

Robert Wrembel (Poznan University of Technology, PL)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Robert Wrembel

Business Intelligence (BI) has grown to a powerful method for supporting strategic decisions in companies. Data warehouses are the most promising software tool support in that area, providing services to structure, index and query large amounts of business data to find patterns and business opportunities. In traditional data warehousing, large amounts of data are handled by bulk-loading them through an ETL process into the data warehouse's database. This process typically has good throughput at the cost of very high latencies between the data occurring and becoming visible to a business user. Our approach aims at shortening this delay tremendously, without a perceivable impact on throughput. We advocate a two-folded approach: first, we introduce a fast track data path, which uses techniques from event-based systems to analyze incoming data on the fly and raise an alarm in case of a certain pattern appearing. Second, we suggest on-demand view updates, where incoming data is stored in an unstructured way and only structured and indexed upon a query requesting these data. This allows us to avoid handling data that is never queried and keep each loading cycle as lightweight as possible.

5 Open Problems

We present next the position statement of the five groups.

5.1 Real-Time Data Warehouses and Business Intelligence

In a typical BI infrastructure, data, extracted from operational data sources, is transformed and cleansed and subsequently loaded into a data warehouse where it can be queried for reporting purposes. ETL, the process of extraction, transformation, and loading, is a periodic process that may involve an elaborate and rather established software ecosystem. Typically, the actual ETL process is executed on a nightly basis, i.e., a full day's worth of data is processed and loaded during off-hours. Depending on the resources available and the nature of the data and the reporting, ETL may also be performed more frequently, e.g., on an hourly basis.

It is desirable to reduce this delay further and ideally provide reports and business insights at real-time. However, this requires overcoming throughput bottlenecks and improving latency throughout the ETL process. Instead of attempting to incrementally improve this situation, we propose a radically different approach: leveraging a data warehouse's capability to directly import raw, unprocessed records, we defer the transformation and cleaning of data until needed by pending reports. At that time, the database's own processing mechanisms can be deployed to process the data on-demand.

This technique ensures that resources are utilized optimally rather than spent speculatively on processing, potentially irrelevant, data in bulk. Besides excluding irrelevant data from processing all together, a more likely scenario is the case where different types of reports are run at different times of the day, week, or month and require different categories of source data. Again, using an on-demand approach helps optimize resource utilization and improves data freshness substantially.

In addition to running periodic reports, an important component of modern BI architectures are interactive elements such as alerting when detecting outliers or encountering exceptional situations during data processing. While highly efficient, our on-demand approach naturally lacks the continuous processing and thus 'surveying' of data of the traditional technique. In order to capture such events, we augment our on-demand approach with an active component that performs light-weight data screening independent of the ELT processing and may be integrated with a BI dashboard.

Besides outlining an overall architecture, we also developed a roadmap for implementing a complete prototype using conventional database technology in the form of hierarchical materialized views.

5.2 Spatio-Temporal Data Warehouses

Modern organizations nowadays need to use OLAP analytical capabilities together with geographical information. In this direction, SOLAP (standing for Spatial OLAP) aims at exploring spatial data by drilling on maps, in the same way as OLAP operates over tables and charts. However, SOLAP only accounts for discrete spatial data, where spatial objects are represented as geometries. More sophisticated GIS-based decision support systems are increasingly being needed, able to handle more complex types of data, like continuous fields (from now on, fields). Fields describe physical phenomena that change continuously in time and/or space, like temperature, land elevation, land use and population density. They are perceived as having a value at each point in a continuous N-dimensional spatial and/or spatio-temporal domain.

In real-world practice, scientists and practitioners register the values of a field by taking samples at (generally) fixed locations, and inferring the values at other points in space using some interpolation method. Thus, fields can be described by a function that indicates the distribution of the phenomena or feature of interest. The most popular discrete representation for fields is the raster model, where the 2D space is divided into regular squares. The raster model is frequently used for representing soil type, temperature, among other physical phenomena. Other representations have also been proposed, like the Voronoi diagrams or TIN, the latter usually employed to represent an 'Altitude' field.

For adding spatial information to OLAP tools, many models have been proposed, at the conceptual and logical levels. We find this unnecessary from the user's point of view. We believe that a user of a spatio-temporal enabled OLAP system would like to find the data

cube at the conceptual level, independently of the kind of underlying data. Such a model would allow to analyze any type of spatial data, continuous (independently of the underlying representation) or discrete, together with typical alphanumeric discrete OLAP data, using only the classic OLAP operators, like Roll-up, Drill-down, and/or Drill-across. To achieve this, at the logical and physical levels we need different mechanisms to manage these different kinds of data and data representations. That means, the final user only sees the typical OLAP operators at the user level. At the logical level operations that allow to manage different kinds of spatial data are needed, independently of their physical representation. Finally, at the physical level, we need a model that allows managing continuous and discrete data, and a collection of operators over this model. For continuous data, it is at this level where we need to care about raster, Voronoi, and/or TIN representations.

5.3 Situational Business Intelligence

Well-informed and effective decisions often require a tight relationship to be established between the variables on company performance stored in the DW and other data that are not even resident in the company information system. These valuable data may be related for instance to the market, to competitors, or to potential customers; they are sometimes called situational data because they have a narrow focus on a specific business problem and, often, a short lifespan for a small group of users with a unique set of needs. In some cases, situational data can be retrieved (for free or on a fee) in a semi-structured form by accessing registered data providers. In other cases, situational data are chaotically scattered across several, heterogeneous, unstructured sources available on the Web (e.g., opinions expressed by users on social networks, ratings of products on portals, etc.). As such they tend to be highly dynamic in contrast with common enterprise data, which are used to address a large set of business problems and impose a slow and careful management.

The capability of incorporating situational data into the decisional process gives rise to a new class of applications, that in the context of BI 2.0 are often labeled as situational BI or on-demand BI or even collaborative BI. Here we prefer to use the term self-service BI, to emphasize that the search, extraction, and integration of situational data should be accomplished in near-realtime by business users through a continuous interaction with the application, without any mediation or intervention by analysts, designers, or programmers. As also emphasized by Gartner Inc., self-service BI appears to be the big wave in BI for 2011; the key idea is to let end users navigate information in a “surf and save” mode, meaning that data can be stored for reuse or sharing. Among the main applications for self-service BI, we mention brand positioning, pricing, competitor monitoring, policy planning, risk management; the domains involved range from retail, telco, and entertainment to finance and public services such as health and transportation. In all these domains, the strategic information needs are currently not fully satisfied by traditional BI approaches and could be positively affected by including situational data.

5.4 Query Processing in Data Warehouses: Elasticity in Cloud Databases

A central promise of cloud services is elastic, on-demand provisioning. For data-intensive services such as data management, growing and shrinking the set of cloud nodes implies

copying data to nodes with temporary membership in a service. At best, a node might retain its copy of the data while it provides another service; at worst, a node that rejoins the database service requires a new copy of data. Many solutions have been proposed to the elasticity for cloud services, but few address the problems raised by “sometimes available copy” (SAC) nodes with on-demand, incremental updates. We believe that SAC nodes should become up-to-date and useful for query processing incrementally by key range. On-demand, based on the queries being evaluated, additional key ranges for a SAC node become up-to-date, with overall update performance comparable to a traditional high-availability strategy that carries the entire dataset forward, until eventually the entire dataset becomes up-to-date and useful for query processing. The maintenance scheme combines log-based replication and updates with sorting data in transit. Our proposed solution relies on techniques from partitioned B-trees, adaptive merging, deferred maintenance of secondary indexes and of materialized views, and query optimization using materialized views. The paper introduces a family of maintenance strategies for temporarily available copies, the space of possible query execution plans and their cost functions, and query processing techniques for this type of elastic environments.

5.5 Knowledge Extraction and Management in Data Warehouses

The objective of this research is to define techniques towards integrating data, information, and knowledge in data warehouses in a user-centric manner in order to help the end user get better benefits from data by allowing (i) a unified and meaningful representation of multidimensional data as well as knowledge patterns, and (ii) advanced query mechanisms and guidance to provide targeted information and knowledge through information retrieval and data mining techniques.

In this work, we first outline the importance of knowledge (e.g., existing ontology and knowledge discovered from data) in data warehouse management, propose a user-centric architecture for OLAP query processing and then define new solutions towards intensional and cooperative query answering using knowledge and exploiting the peculiarities of data warehouses. Such work aims to help a user (who is not necessarily familiar with a query language or aware about the detailed structure and content of the data warehouse) get a richer and more complete and even compact answer to his/her possibly incomplete or vague OLAP query. For example, an executive may ask for the top sales in a given time period, and the answer could be the concise and semantically rich one: the top sales concern almost all sport products bought by customers from the West Coast.

Participants

- Alberto Abello
Universitat Politècnica de Catalunya – Barcelona, ES
- Marie-Aude Aufaure
Ecole Centrale – Paris, FR
- Cristina Ciferri
University of Sao Paulo, BR
- Ricardo Ciferri
University of Sao Carlos, BR
- Alfredo Cuzzocrea
ICAR-CNR – Rende, IT
- Maria Luisa Damiani
Università di Milano, IT
- Jérôme Darmon
Université Lumière Lyon 2, FR
- Todd Eavis
Concordia Univ. – Montreal, CA
- Lorena Etcheverry
Universidad de la Republica – Montevideo, UY
- Cécile Favre
Université Lumière Lyon 2, FR
- David Fekete
Universität Münster, DE
- Tobias Freudenreich
TU Darmstadt, DE
- Pedro Furtado
University of Coimbra, PT
- Matteo Golfarelli
Università di Bologna, IT
- Leticia I. Gómez
Instituto Tecnológico de Buenos Aires, AR
- Goetz Graefe
HP Labs – Madison, US
- Christian Koncilia
Universität Klagenfurt, AT
- Patrick Marcel
Université de Tours – Blois, FR
- Jose-Norberto Mazón
Univ. of Alicante, ES
- Rokia Missaoui
Université du Québec en Outaouais, CA
- Felix Naumann
Hasso-Plattner-Institut – Potsdam, DE
- Thomas Neumann
TU München, DE
- Anisoara Nica
Sybase, An SAP Company, CA
- Torben B. Pedersen
Aalborg University, DK
- Ilia Petrov
TU Darmstadt, DE
- Elaheh Pourabbas
National Research Council – Rome, IT
- Stefano Rizzi
University of Bologna, IT
- Markus Schneider
University of Florida – Gainesville, US
- Knut Stolze
IBM Deutschland – Böblingen, DE
- Maik Thiele
TU Dresden, DE
- Juan Trujillo
Univ. of Alicante, ES
- Alejandro Vaisman
Université Libre de Bruxelles, BE
- Panos Vassiliadis
University of Ioannina, GR
- Gottfried Vossen
Universität Münster, DE
- Florian M. Waas
EMC – San Mateo, US
- Robert Wrembel
Poznan Univ. of Technology, PL
- Esteban Zimányi
Université Libre de Bruxelles, BE



Uncertainty modeling and analysis with intervals: Foundations, tools, applications

Edited by

Isaac Elishakoff¹, Vladik Kreinovich², Wolfram Luther³, and
Evgenija D. Popova⁴

1 Florida Atlantic University – Boca Raton, US

2 University of Texas – El Paso, US, vladik@utep.edu

3 Universität Duisburg-Essen, DE, luther@inf.uni-due.de

4 Bulgarian Academy of Sciences, BG, epopova@bio.bas.bg

Abstract

This report documents the program and the results of Dagstuhl Seminar 11371 “Uncertainty modeling and analysis with intervals – Foundations, tools, applications”, taking place September 11-16, 2011. The major emphasis of the seminar lies on modeling and analyzing uncertainties and propagating them through application systems by using, for example, interval arithmetic.

Seminar 11.–16. September, 2011 – www.dagstuhl.de/11371

1998 ACM Subject Classification G.1.0 General Computer arithmetic; Error analysis; Interval arithmetic; G.4 Mathematical Software Algorithm design and analysis; Verification; I.6.4 Model Validation and Analysis.

Keywords and phrases Uncertainty modeling – Propagation of uncertainty through and validation for computational models using interval arithmetic – Imprecise probabilities, Sensitivity analysis – Applications to structures, mechatronics, bioinformatics and finance

Digital Object Identifier 10.4230/DagRep.1.9.26

Edited in cooperation with Martin Fuchs (CERFACS – Toulouse, FR)


1 Executive Summary

Isaac Elishakoff

Vladik Kreinovich

Wolfram Luther

Evgenija Popova

License  Creative Commons BY-NC-ND 3.0 Unported license
© Elishakoff, Isaac; Kreinovich, Vladik; Luther, Wolfram; Popova, Evgenija

Verification and validation (V&V) assessment of process modeling and simulation is increasing in importance in various areas of application. They include complex mechatronic and bio-mechanical tasks with especially strict requirements on numerical accuracy and performance. However, engineers lack precise knowledge regarding the process and its input data. This lack of knowledge and the inherent inexactness in measurement make such general V&V cycle tasks as design of a formal model and definition of relevant parameters and their ranges difficult to complete.

To assess how reliable a system is, V&V analysts have to deal with uncertainty. There are two types of uncertainty: aleatory and epistemic.



Except where otherwise noted, content of this report is licensed under a Creative Commons BY-NC-ND 3.0 Unported license

Uncertainty modeling and analysis with intervals: . . . , *Dagstuhl Reports*, Vol. 1, Issue 9, pp. 26–57

Editors: Isaac Elishakoff, Vladik Kreinovich, Wolfram Luther, and Evgenija D. Popova



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Aleatory uncertainty refers to variability similar to that arising in games of chance. It cannot be reduced by further empirical study. Epistemic (reducible) uncertainty refers to the uncertainty resulting from lack of knowledge. An example is the absence of evidence about the probability distribution of a parameter. In this situation, standard methods for modeling measurement uncertainty by using probability distributions cannot be applied. Here, interval methods provide a possible solution strategy.

Another option, mostly discussed in the context of risk analysis, is to use interval-valued probabilities and imprecisely specified probability distributions. The probability of an event can be specified as an interval; probability bounds analysis propagates constraints on a distribution function through mathematical operations. In a more general setting, the theory of imprecise probabilities is a powerful conceptual framework in which uncertainty is represented by closed, convex sets of probability distributions. Bayesian sensitivity analysis or Dempster-Shafer theory are further options.

A standard option in uncertainty management is Monte Carlo simulation. This is a universal data-intensive method that needs random number generators, distributions, dependencies, and a mathematical model (but not a closed analytic solution) to provide accurate results. Compared to interval methods, it yields less conservative bounds, which, however, might fail to contain the exact solution. As an implementation of convolution in probability theory, Monte Carlo methods are complementary to interval approaches.

Additionally, they play an important role in probability bounds analysis, Dempster-Shafer theory, and further approaches combining probabilistic and interval uncertainties.

The goal of this seminar is to promote and accelerate the integration of reliable numerical algorithms and statistics of imprecise data into the standard procedures for assessing and propagating uncertainty. The main contributions of this seminar were

- Expressing, evaluating and propagating measurement uncertainties; designing efficient algorithms to compute various parameters, such as means, median and other percentiles, variance, interquantile range, moments and confidence limits; summarizing the computability of such statistics from imprecise data.
- New uncertainty-supporting dependability methods for early design stages. These include the propagation of uncertainty through dependability models, the acquisition of data from similar components for analyses, and the integration of uncertain reliability and safety predictions into an optimization framework.
- Modeling and processing applications from the areas of robust geometrical design, financial simulation and optimization, robotics, mechatronics, reliability and structural safety, bioinformatics and climate science with uncertain input parameters and imprecise data.
- Discussing software for probabilistic risk and safety assessments working with real numbers, intervals, fuzzy numbers, probability distributions, and interval bounds on probability distributions that combines probability theory and interval analysis and makes the newest techniques such as interval Monte Carlo method, probability bounds analysis and fuzzy arithmetic available.
- Promoting a new interval standard for interval arithmetic as explained in the P1788 draft: “This standard specifies basic interval arithmetic operations selecting and following one of the commonly used mathematical interval models and at least one floating-point type defined by the IEEE-754/2008 standard. Exception conditions are defined and standard handling of these conditions are specified. Consistency with the model is tempered with practical considerations based on input from representatives of vendors and owners of existing systems”.

2 Table of Contents

Executive Summary

<i>Elishakoff, Isaac; Kreinovich, Vladik; Luther, Wolfram; Popova, Evgenija</i>	26
---	----

Overview of Talks

Application of Verified Methods to Solving Non-smooth Initial Value Problems in the Context of Fuel Cell Systems <i>Ekaterina Auer</i>	30
Fuzzy Probabilities and Applications in Engineering <i>Michael Beer</i>	31
Asymptotic Stabilization of a Bioprocess Model Involving Uncertainties <i>Neli Dimitrova</i>	32
Robust optimization for aerospace applications <i>Martin Fuchs</i>	33
Verified Solution of Finite Element Models for Truss Structures with Uncertain Node Locations <i>Jürgen Garloff</i>	35
Interval Linear Programming: Foundations, Tools and Challenges <i>Milan Hladík</i>	35
Intervals, Orders, and Rank <i>Cliff Joslyn</i>	36
Interval Computations – Introduction and Significant Applications <i>Ralph Baker Kearfott</i>	37
Integration of Interval Contractors in Hierarchical Space Decomposition Structures <i>Stefan Kiel</i>	38
Degree-Based (Interval and Fuzzy) Techniques in Math and Science Education <i>Olga M. Kosheleva</i>	39
A Comparison of Different Kinds of Multiple Precision and Arbitrary Precision Interval Arithmetics <i>Walter Krämer</i>	40
Towards Optimal Representation and Processing of Uncertainty for Decision Making, on the Example of Economics-Related Heavy-Tailed Distributions <i>Vladik Kreinovich</i>	41
From Processing Interval-Valued Data to Processing Fuzzy Data: A Tutorial <i>Vladik Kreinovich</i>	42
Generating a Minimal Interval Arithmetic Based on GNU MPFR <i>Vincent Lefèvre</i>	43
IPPToolbox – a package for imprecise probabilities in R <i>Philipp Limbourg</i>	43
Constrained Intervals and Interval Spaces <i>Weldon A. Lodwick</i>	43

Verification and Validation Requirements in Biomechanical Modeling and Simulation <i>Wolfram Luther</i>	44
Enclosing solutions of initial-value problems with large uncertainty <i>Arnold Neumaier</i>	45
Characterizing AE Solution Sets to Parametric Linear Systems <i>Evgenija D. Popova</i>	46
What you always wanted to know about decorated intervals <i>John D. Pryce</i>	46
Verified Parameter Estimation for the Thermal Behavior of High-Temperature Fuel Cells <i>Andreas Rauh</i>	47
Verified Add-ons for the DSI toolbox <i>Gabor Rebner</i>	49
Refining Abstract Interpretation-based Approximations with Constraint Solvers <i>Michel Rueher</i>	51
Constraint Programming over Continuous Domains <i>Michel Rueher</i>	51
Managing uncertainty and discontinuous condition numbers in finite-precision geometric computation <i>Peihui Shao</i>	52
Reliable Kinetic Monte Carlo Simulation based on Random Set Sampling <i>Yan Wang</i>	52
The General Interval Power Function <i>Jürgen Wolff von Gudenberg</i>	53
C-XSC – Overview and new developments <i>Michael Zimmer</i>	53
Schedule	55
Participants	57


3 Overview of Talks

The seminar was attended by 33 participants from 8 countries who gave 34 talks. To stimulate debate and cross-fertilization of new ideas we scheduled a mixture of tutorials, contributed talks, a meeting of the IEEE P1788 working group, and software demonstrations. The seminar started with a series of talks aimed at providing a suitable level of introduction to the main areas of discussion and providing a leveling ground for all participants.

The format of the seminar was then a series of contributed presentations on the variety of the seminar topics mentioned above. A lively discussion on the current state of the interval standardization was initiated by the talk on the hot topic of decorated intervals on Tuesday afternoon and continued during the meeting of the IEEE P1788 working group on Thursday afternoon. A session on software tools, held on Wednesday, was followed by software demonstrations on Thursday evening. There was much time for extensive discussions in between the talks, in the evenings, and during the excursion on Wednesday afternoon. The seminar had generally a very open and constructive atmosphere. As a result of the seminar there will be a special issue published in a leading journal that will not only publish papers presented at the seminar, but also provide a roadmap for the future directions of the uncertainty modeling.

3.1 Application of Verified Methods to Solving Non-smooth Initial Value Problems in the Context of Fuel Cell Systems

Ekaterina Auer (Universität Duisburg-Essen, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Ekaterina Auer

Joint work of Auer, Ekaterina; Rauh, Andreas

In many engineering applications, there is a need to choose mathematical models that depend on non-smooth functions. For example, the models for friction or, more broadly, contact dynamics, are not even continuous in general. There are also less obvious situations that call for non-smooth functions, for instance, when naturally arising conditions such as non-positivity of variables have to be taken into account. The task becomes especially difficult if such functions appear on the right side of an initial value problem (IVP). Here, even the definition of the solution depends on the application at hand [2]. Since uncertainty in parameters obstructs many non-smooth tasks additionally, verified methods might prove themselves to be more useful than those from the usual numerics. Besides, they guarantee the correctness of the result within the limitations of a particular model.

The development of verified methods for IVPs with non-smooth right sides has got relatively few attention throughout the last three decades. To our knowledge, there exist no modern publicly available implementation at the moment. In [6], Rihm proposes a suitable definition and a method to enclose the solution to IVPs changing their right sides in dependence on a certain algebraic function. In [1, 4, 5], the authors propose algorithms for systems switching their representation according to graphs containing different ordinary differential equations as vertices and logical conditions as edges. Additionally, a lot of research has been done on generalizing the notion of a derivative for non-smooth functions in the area of verified optimization [3, 7].

In this talk, we give a short overview of the already existing methods for solving IVPs with non-smooth right sides. Next, we develop a generalized derivative definition for a


certain type of continuous functions and its possible modification for bounded non-continuous functions of the same type. Then we use this definition inside the algorithm of the verified solver VALENCIA-IVP. We chose this solver because it needs only first order derivatives of the right side of an IVP. Finally, we demonstrate the applicability of our method in the context of modeling and simulation of high temperature fuel cells.

References

- 1 A. Eggers, M. Fränzle, and C. Herde. Application of Constraint Solving and ODE-Enclosure Methods to the Analysis of Hybrid Systems. In *Numerical Analysis and Applied Mathematics 2009*, volume 1168, pages 1326–1330. American Institute of Physics, 2009.
- 2 A. Filippov. *Differential Equations With Discontinuous Righthand Sides*. Kluwer Academic Publishers, 1988.
- 3 H. Munoz and R. B. Kearfott. Slope Intervals, Generalized Gradients, Semigradients, Slant Derivatives, and Csets. *Reliable Computing*, 10:163–193, 2004.
- 4 N. Nedialkov and M. von Mohrenschildt. Rigorous Simulation of Hybrid Dynamic Systems with Symbolic and Interval Methods. In *Proceedings of the American Control Conference Anchorage*, 2002.
- 5 A. Rauh, C. Siebert, and H. Aschemann. Verified Simulation and Optimization of Dynamic Systems with Friction and Hysteresis. In *Proceedings of ENOC 2011*, Rome, Italy, July 2011.
- 6 R. Rihm. Enclosing solutions with switching points in ordinary differential equations. In *Computer arithmetic and enclosure methods. Proceedings of SCAN 91*, pages 419–425. Amsterdam: North-Holland, 1992.
- 7 M. Schnurr. *Steigungen höherer Ordnung zur verifizierten globalen Optimierung*. PhD thesis, Universität Karlsruhe, 2007.

3.2 Fuzzy Probabilities and Applications in Engineering

Michael Beer (University of Liverpool, GB)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Michael Beer

A key issue in computational engineering disciplines is the realistic numerical modeling of physical and mechanical phenomena and processes. This is the basis to derive predictions regarding behavior, performance, and reliability of engineering structures and systems. In engineering practice, however, the available information is frequently quite limited and of poor quality. A solution to this conflict is given with imprecise probabilities, which involve both probabilistic uncertainty and non-probabilistic imprecision. An entire set of plausible probabilistic models is considered in one analysis. This leads to more realistic results and helps to prevent wrong decisions. In this context fuzzy probabilities and their application in engineering were discussed in the presentation. Usefulness and benefits were demonstrated by means of various practical examples from different engineering fields.

3.3 Asymptotic Stabilization of a Bioprocess Model Involving Uncertainties

Neli Dimitrova (Bulgarian Academy of Sciences, BG)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Neli Dimitrova

The dynamic modeling of anaerobic digestion has recently become an active research area.

This is due to the fact that a mathematical model of the bioreactor can be used as a powerful tool to simulate different operating, control and optimization strategies.

One of the main drawbacks in the modeling and control of the anaerobic digestion lies in the difficulty to monitor on-line the key biological variables of the process and in estimating the expressions of the bacterial growth rates.

Thus developing control systems only based on simple measurements and minimal assumptions on the growth rates that guarantee stability of the process is of primary importance.

We consider the following model of a biological digestion process

$$\begin{aligned}\frac{ds_1}{dt} &= u(s_1^i - s_1) - k_1\mu_1(s_1)x_1 \\ \frac{dx_1}{dt} &= (\mu_1(s_1) - \alpha u)x_1 \\ \frac{ds_2}{dt} &= u(s_2^i - s_2) + k_2\mu_1(s_1)x_1 - k_3\mu_2(s_2)x_2 \\ \frac{dx_2}{dt} &= (\mu_2(s_2) - \alpha u)x_2\end{aligned}\tag{1}$$

with output

$$Q = k_4\mu_2(s_2)x_2,$$

where the phase variables s_1 , s_2 and x_1 , x_2 represent substrate and biomass concentrations respectively, $\mu_1(s_1)$ and $\mu_2(s_2)$ are bacterial growth rate functions, s_1^i and s_2^i are input substrate concentrations, u is the dilution rate (control input), Q is the methane flow rate, α is a homogeneity parameter and k_j , $j = 1, 2, 3, 4$, are coefficients.

This model has been investigated in [1], [2], where some control strategies are proposed and their robustness is illustrated mainly by simulation studies.

The present talk is an overview of authors' results on global asymptotic stabilizability of the dynamic model (1) by a feedback control law.

Practical applications impose the following requirements on the feedback law: dependance on online measurable variables, and robustness under uncertainties in the model coefficients and growth rate functions.

Let the growth rate functions satisfy the following *general assumption*: $\mu_j(s_j)$ is defined for $s_j \in [0, +\infty)$, $\mu_j(0) = 0$, $\mu_j(s_j) > 0$ for $s_j > 0$; $\mu_j(s_j)$ is continuously differentiable and bounded for all $s_j \geq 0$, $j = 1, 2$.

The first result concerns the so-called adaptive asymptotic stabilization of the control system (1).

We extend the system by the differential equation

$$\frac{d\beta}{dt} = -C(\beta - \beta^-)(\beta^+ - \beta)k_4\mu_2(s_2)x_2(s - \bar{s}),\tag{2}$$

where C , β^- and β^+ are appropriate positive constants, and \bar{s} is a previously chosen operating point, representing the biological oxygen demand (which is on-line measurable). Under some

additional (practically meaningful) conditions we have shown in [3] that the feedback control law

$$u \equiv k(s_1, x_1, s_2, x_2, \beta) = \beta k_4 \mu_2(s_2) x_2 = \beta Q$$

stabilizes asymptotically the extended closed-loop system (1)–(2) to an equilibrium point, corresponding to the value \bar{s} , for each starting point $(s_1(0), x_1(0), s_2(0), x_2(0)) \geq 0$. The proposed feedback is robust with respect to model uncertainties only in the case of uncertain growth rates: we assume that instead of the exact functions $\mu_1(s_1)$ and $\mu_2(s_2)$ we know bounds for them, i. e. $\mu_1(s_1) \in [\mu_1(s_1)] = [\mu_1^-(s_1), \mu_1^+(s_1)]$, $\mu_2(s_2) \in [\mu_2(s_2)] = [\mu_2^-(s_2), \mu_2^+(s_2)]$. If any $\mu_j(s_j) \in [\mu_j(s_j)]$, $j = 1, 2$, satisfies the above general assumption, it is shown in [4] that the global stabilizability of the closed-loop system is retained.

Uncertainties with respect to the four coefficients k_j , $j = 1, 2, 3, 4$, are not considered.

Further recent results in [4] extend and improve the above studies. First we avoid the auxiliary differential equation (2), since it cannot be interpreted in terms of process dynamics; second, we assume that not only the growth rates, but also the coefficients k_j are unknown but bounded within compact intervals: $k_j \in [k_j] = [k_j^-, k_j^+]$, $j = 1, 2, 3, 4$.

Denote by $s^{i-} = \frac{k_2^-}{k_1^+} s_1^i + s_2^i$ a lower bound for the input biological oxygen demand. Let be $\beta \in \left(\frac{k_3^+}{s^{i-} \cdot k_4^-}, +\infty \right)$. Take any value of $k_j \in [k_j]$, $j = 1, 2, 3, 4$, and define $\bar{s}_\beta = s^i - \frac{k_3}{\beta k_4}$. It is then proved in [4] that the feedback control law $k(s_1, x_1, s_2, x_2) = \beta k_4 \mu_2(s_2) x_2 = \beta Q$ stabilizes asymptotically the closed-loop system to an equilibrium point, corresponding to \bar{s}_β for each starting point $(s_1(0), x_1(0), s_2(0), x_2(0)) \geq 0$.


An important practical problem is the stabilization of the model (1) to a point, where maximum methane flow rate Q is achieved. This problem is also solved in [3] and [4] by designing an extremum seeking model-based algorithm. Computer simulations illustrate the theoretical studies.

References

- 1 L. Maillert, O. Bernard, and J.-P. Steyer. Robust regulation of anaerobic digestion processes. *Water Science and Technology*, 48(6):87–94, 2003.
- 2 F. Grognaud and O. Bernard. Stability analysis of a wastewater treatment plant with saturated control. *Water Science and Technology*, 53:149–157, 2006.
- 3 N. Dimitrova and M. Krastanov. *Modeling, Design, and Simulation of Systems with Uncertainties*, volume 3 of *Mathematical Engineering*, chapter Nonlinear adaptive control of a bioprocess model with unknown kinetics, pages 275–292. Springer, 2011.
- 4 N. Dimitrova and M. Krastanov. On the asymptotic stabilization of an uncertain bioprocess model. To appear in *Lecture Notes in Computer Science*, 2012.

3.4 Robust optimization for aerospace applications

Martin Fuchs (Cerfacs – Toulouse, FR)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Martin Fuchs
URL <http://www.martin-fuchs.net>

Many modern applications require a profound treatment of uncertainties. Two of the most critical issues to be dealt with are lack of statistical information and the well-known curse of dimensionality. One of the concerned fields is optimization for aerospace applications, where

computational black box models include many uncertain parameters with little data and little information about statistical correlations given. The major research goal is to achieve optimal solutions that can be qualified as robust by accounting for the uncertainties in the high dimensional parameter space.

In lower dimensions, there are several tools to handle lack of information reliably, e.g., p-boxes, Dempster-Shafer structures, or possibility distributions. However, in higher dimensions it may require an intrusive implementation to be efficient to propagate uncertainties through a function. If the uncertainties are propagated through a black box function simulation techniques are often preferred, but they may fail to be reliable in many cases, see [1]. Sensitivity analysis can help to reduce the dimensionality at additional computational cost. The clouds formalism, see [5], combines concepts of intervals, fuzzy sets, and probability theory, in order to deal with both incomplete and higher dimensional information in a reliable and computationally tractable fashion.

Our approach first uses clouds to determine a polyhedral representation of the uncertainties. In other words, we describe the set, in which we search for worst-case scenarios, as a polyhedron. Methods to generate this polyhedron already exist, see [3]. In the second step we solve an optimization problem subject to polyhedral constraints to actually find the worst-case scenario. Our approach to the solution of the problem is inspired by the simulation based Cauchy deviates method for interval uncertainty, see [4]. It turns out to be computationally very attractive and it can be easily parallelized.


The new methods are employed in the context of robust optimization. The worst-case analysis of the previous steps becomes a constraint of an optimization problem formulation, see [2]. The objective is to find an optimum that is safeguarded against uncertain perturbations. To this end we determine the worst-case objective function, i.e., we propagate the uncertainties through the objective function. Thus the extra computational effort to account for robustness amounts to extra objective function evaluations. Hence it is important to use only very few evaluations as the total budget of evaluations in the optimization phase of real-life applications is typically very limited. This gives a new point of view of the Cauchy deviates method. Numerical tests are presented for applications from space system design and aircraft wing shape optimization.

References

- 1 S. Ferson, L. Ginzburg, and R. Akcakaya. Whereof one cannot speak: When input distributions are unknown. Web document, <http://www.ramas.com/whereof.pdf>, 1996.
- 2 M. Fuchs and A. Neumaier. Autonomous robust design optimization with potential clouds. *International Journal of Reliability and Safety*, 3(1/2/3):23–34, 2009.
- 3 M. Fuchs and A. Neumaier. Potential based clouds in robust design optimization. *Journal of Statistical Theory and Practice*, 3(1):225–238, 2009.
- 4 V. Kreinovich and R. Trejo. *Handbook of Randomized Computing*, chapter Error estimations for indirect measurements: randomized vs. deterministic algorithms for 'black-box' programs, pages 673–729. Kluwer, 2001.
- 5 A. Neumaier. Clouds, fuzzy sets and probability intervals. *Reliable Computing*, 10(4):249–272, 2004.

3.5 Verified Solution of Finite Element Models for Truss Structures with Uncertain Node Locations

Jürgen Garloff (HTWG Konstanz, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
 © Jürgen Garloff
 URL <http://www-home.htwg-konstanz.de/~garloff/>

In our talk, we consider finite element models for mechanical truss structures where all of the physical model parameters are uncertain but bounded and are represented by intervals. In this case the application of the finite element method results in a system of parametric linear equations where the parameters vary within given intervals. We aim at finding tight bounds for the solution set of such a parametric system, the so-called parametric solution set. We first consider the case that the parameter dependency is rational and briefly report on a combination of software for the parametric residual iteration written by E. Popova in a *Mathematica* environment and our own software for the tight enclosure of the range of multivariate polynomials over a box.

Then not just the material parameters and applied loads, but also the positions of the nodes are assumed to be inexact and are represented by intervals, a case which does not seem to have previously been considered in the literature. In civil engineering, these uncertainties are often due to imperfections of the fabrication process. The application of the mentioned software for the enclosure of the parametric solution set results in intervals which are too wide for practical purposes. To contract the obtained intervals we employ interval pruning techniques. In the case of a statically indeterminate truss structure, the resulting intervals for the node displacements are still wider than we would like. Therefore, we employ a monotonicity analysis for all the parameters to provide tight guaranteed enclosures.

3.6 Interval Linear Programming: Foundations, Tools and Challenges

Milan Hladík (Charles University – Prague, CZ)

License  Creative Commons BY-NC-ND 3.0 Unported license
 © Milan Hladík

Many practical problems can be formulated in terms of linear programming, and in the others linear programming is sometimes used as a auxiliary technique. Interval liner programming studies such problems that are subject to uncertainties and there are given lower and upper bounds for uncertain quantities. This approach is superior the standard sensitivity analysis since it takes into account more complex perturbations of input data. Contrary to the stochastic programming approach one does not have to care about distributions of uncertain parameters; lower and upper estimates are enough.

We present a survey on interval linear programming according to the very recent paper [1]; just in press. We give a brief overview on the basic problems concerning feasibility, unboundedness and optimality for both weak and strong case, where *weak* means the property for some realization of interval data, and *strong* means validity for each realization. In particular, we list the time complexities (polynomial vs. NP-hard) of these problems. Then we turn our attention to the two fundamental problems studied. The first one is a problem of determining the optimal value range. Depending on the form of the interval linear program, some of the bounds can be computed by an ordinary linear program, but the others


are NP-hard and only a formula using exponential number of ordinary linear programs is known. The second fundamental problem, which is very difficult and still challenging, is to find a tight enclosure of the optimal solution set. This becomes easy in the case of the so called basis stability, i.e., there is a basis optimal for each realization of interval data. Checking basis stability is also a computationally hard problem, but there are quite strong sufficient conditions that may be utilized. Eventually, we state some open problems.

References

- 1 Milan Hladík. *Linear Programming – New Frontiers in Theory and Applications*, 2nd chapter Interval linear programming: A survey. Nova Science Publisher, New York, 2011.

3.7 Intervals, Orders, and Rank

Cliff Joslyn (Pacific Northwest National Lab. – Seattle, US)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Cliff Joslyn

Joint work of Joslyn, Cliff; Hogan, Emilie

Intervals are ordered
Orders have intervals
Orders have ranks
Ranks are intervals

Intervals have order relations defined on them as an important operation. Also, partially ordered sets (posets) and lattices have intervals within them as important sub-structures. In traditional interval analysis, the set on which the intervals are drawn is the real numbers, a special ordered set which is a total order; and the ordering relation used between these real intervals is the “strong” interval order (one interval being entirely below another), in the context of an overall Allen’s algebra.

But in our work in semantic databases and ontology management, it is the more general cases which are demanded. Specifically: 1) the intervals in question are valued in a finite, bounded, generally partially ordered set; and 2) when real intervals ARE used, the conjugate endpoint product order and subset orders are far preferable to the standard strong order (which isn’t even really an ordering relation anyway). The attendant issues have implications for the foundations of interval analysis which we seek to explore with the group.

Depending on time, structure, and the interests of attendees, we can go into more or less depth on the following.

We begin by describing our use of large, finite, bounded posets to represent taxonomic semantic data structures for applications such as ontology clustering and alignment. We then consider the challenges presented by their layout and display.

Our first challenge, the vertical layout of nodes, we have been working on for a while. We observe that rank in posets is best considered as being valued on integer intervals. These integer-valued rank intervals can themselves in turn be ordered (in the endpoint product order), so that an iterative operation is available. Repeated application serves to identify a privileged embedding of the poset to a total preorder preferred to reflect the underlying partial order. We have results about how the height, width, and dimension of the poset changes in repeated application, and prove that we do achieve a final embedding of the original poset to a total preorder. In the process, results about measures of gradedness of posets are also motivated.

Our second challenge we are just beginning to explore, but we will present some preliminary ideas for discussion. We seek to simplify the display of large posets (actually lattices in the first treatment) when only a subset of nodes are specified by a user. A tremendous reduction in complexity can be available when (poset) intervals among the target nodes are identified which are disjoint (pairwise or moreso). The underlying mathematical representation suggested is the graph of the intersection structure of poset intervals, that is, a generalization of an interval graph to poset intervals. Cliques in this graph determine the number of total meets and joins which need to be displayed in the reduced visualization, and thus the amount of compression achievable. But, this approach requires us to have the ordering relation between pairs of poset intervals, that is, to develop an Allen's algebra generalized to poset intervals.

3.8 Interval Computations – Introduction and Significant Applications

Ralph Baker Kearfott (Univ. of Louisiana – Lafayette, US)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Ralph Baker Kearfott

In this tutorial, we first outline the historical motivations and early work in interval arithmetic, then review basic interval arithmetic operations and their properties, including advantages and pitfalls. We conclude with a variety of examples of successful application of interval arithmetic.

Historical motivation and work includes

- 1931 — Classical analysis:** Rosaline Cecily Young developed interval arithmetic to handle analysis of one-sided limits (where $\liminf f \neq \limsup f$) [15].
- 1951 — Roundoff error analysis:** Paul S. Dwyer developed interval arithmetic in the chapter on roundoff error analysis in his numerical analysis text [3].
- 1956 — Calculus of Approximations:** Warmus and Steinhaus developed interval arithmetic to provide a sound theoretical backing to numerical computation [14].
- 1958 — Automatic error analysis:** Teruro Sunaga developed interval arithmetic [11].
- 1959 — Automatic error analysis:** Ray Moore developed interval arithmetic in a report and dissertation to which most modern work on the subject can be traced [5].

Advantages of interval arithmetic include the ability to quickly compute mathematically rigorous bounds on roundoff error and on ranges of functions, where computation of the exact range is NP-hard. Disadvantages are that these bounds may be unusably pessimistic, unless special algorithms are designed.

Current successful significant applications include the following:

- A filter in branch and bound methods** in leading commercial software, such as [9] (and others).
- Constraint solving** and constraint propagation, as in [8, §14] and numerous other works.
- Verified solution of ODEs**, as in [1], [4], etc.
- Computer-aided proofs**, as in [13], [12].
- Chemical engineering**, as in [10].
- PDE problems** such as structural analysis with uncertainties [6], [7] or analysis of photonic crystals [2].
- Numerous others.**

References

- 1 Martin Berz. COSY INFINITY web page, 2000. cosy.pa.msu.edu.
- 2 W. Dörfler, A. Lechleiter, M. Plum, G. Schneider, and C. Wieners. *Photonic Crystals: Mathematical Analysis and Numerical Approximation*. Oberwolfach Seminars. Birkhauser Verlag AG, 2011.
- 3 P. S. Dwyer. Computation with approximate numbers. In P. S. Dwyer, editor, *Linear Computations*, pages 11–35, New York, 1951. Wiley & Sons Inc.
- 4 Youdong Lin, Joshua A. Enszer, and Mark A. Stadtherr. Enclosing all solutions of two-point boundary value problems for ODEs. *Computers in Chemical Engineering*, 2008. (in press).
- 5 R. E. Moore and C. T. Yang. Interval analysis I. Technical Document LMSD-285875, Lockheed Missiles and Space Division, Sunnyvale, CA, USA, 1959.
- 6 R. L. Muhanna and R. L. Mullen. Uncertainty in mechanics problems — interval-based approach. *Journal of Engineering Mechanics*, 127(6):557–566, 2001.
- 7 Rafi L. Muhanna, Hao Zhang, and Robert L. Mullen. Interval finite elements as a basis for generalized models of uncertainty in engineering mechanics. *Reliable Computing*, 13:173–194, 2007.
- 8 Arnold Neumaier. Complete search in continuous global optimization and constraint satisfaction. In A. Iserles, editor, *Acta Numerica 2004*, pages 271–369. Cambridge University Press, 2004.
- 9 Nikolaos V. Sahinidis. BARON: A general purpose global optimization software package. *J. Global Optim.*, 8(2):201–205, 1996.
- 10 Mark A. Stadtherr. Interval analysis: Application to chemical engineering design problems. In Arieh Iserles, editor, *Encyclopedia of Optimization*. Kluwer Academic Publishers, 2001.
- 11 Teruo Sunaga. Theory of interval algebra and its application to numerical analysis. *RAAG Memoirs*, 2:29–46, 1958.
- 12 George G. Szpiro. *Kepler’s Conjecture: How Some of the Greatest Minds in History Helped Solve One of the Oldest Math Problems in the World*. E-Book, June 2003.
- 13 Warwick Tucker. A rigorous ODE solver and Smale’s 14th problem. *Found. Comput. Math.*, 24:53–117, 2002.
- 14 M. (Mieczysław) Warmus and H. Steinhaus. Calculus of approximations. *Bulletin de l’Academie Polonaise des Sciences, Cl. III*, IV(5):253–259, February 1956.
- 15 R. C. Young. The algebra of many-valued quantities. *Math. Ann.*, 104:260–290, 1931.

3.9 Integration of Interval Contractors in Hierarchical Space Decomposition Structures

Stefan Kiel (Universität Duisburg-Essen, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Stefan Kiel

Hierarchical spatial data structures can be used for decomposing geometric objects into simpler primitives [5]. Often used primitives are axis-aligned boxes. Intervals are a natural choice for representing them. Furthermore, interval arithmetic (IA) [1] offers us a way to construct a verified decomposition enclosing the object and to cope with uncertainties in the original model.

However, classical IA often suffers from overestimation which might make object enclosures too wide. Recently, we have presented the framework UniVerMeC (Unified Framework for Verified Geometric Computations) [3] that allowed us to employ more sophisticated

arithmetics like affine arithmetic or Taylor models to reduce the overestimation. Another way to tighten enclosures is to use contractors which identify parts of the decomposition disjoint with the object.

In this talk, we will discuss how UniVerMeC helps us to integrate arbitrary interval contractors into the trees. This is a direct extension of our recent approach [4] which was limited formerly to implicit linear interval estimations [2]. The direct integration into the trees lets an algorithm take advantage of contractors' properties without being aware of their actual use. This decoupling allows us to add or remove different contractors easily. In contrast to domain decompositions used, for example, in global optimization, interval trees have to cover the whole area by nodes properly. That is, we cannot dispose parts not containing the object but have to cover them with WHITE nodes indicating that they are empty. This makes employing interval contractors in trees complicated. We will show how to simplify this task by introducing special inversion nodes into the standard set. This does not change existing tree algorithms because nodes of the new type can be converted exactly into a set of standard nodes.

References

- 1 G. Alefeld and J. Herzberger. *Introduction to Interval Computations*. Academic Press, 1983.
- 2 K. Bühler. Implicit linear interval estimations. In *Proceedings of the 18th spring conference on Computer graphics*, page 132. ACM, 2002.
- 3 E. Dyllong and S. Kiel. A comparison of verified distance computation between implicit objects using different arithmetics for range enclosure. In *Proceedings of SCAN'2010*. Submitted.
- 4 S. Kiel. Verified spatial subdivision of implicit objects using implicit linear interval estimations. In *Curves and Surfaces 2011*, volume 6920 of *LNCS*. Springer, 2011. To appear.
- 5 H Samet. *Foundations of Multidimensional and Metric Data Structures*. Morgan Kaufmann, San Francisco, 2006.

3.10 Degree-Based (Interval and Fuzzy) Techniques in Math and Science Education

Olga M. Kosheleva (University of Texas – El Paso, US)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Olga M. Kosheleva

In education, evaluations of the student's knowledge, skills, and abilities are often subjective. Teachers and experts often make these evaluations by using words from natural language like “good”, “excellent”. Traditionally, in order to be able to process the evaluation results, these evaluations are first transformed into exact numbers. This transformation, however, ignores the uncertainty of the original estimates. To get a more adequate picture of the education process and education results, it is therefore desirable to transform these evaluations into intervals – or, more generally, fuzzy numbers.

We show that this more adequate transformation can help on all the stages of the education process: in planning education, in teaching itself, and in assessing the education results.

Specifically, in planning education and in teaching itself, interval and fuzzy techniques help us:

- better plan the order in which the material is presented and the amount of time allocated for each topic;
- interval and fuzzy techniques help us find the most efficient way of teaching interdisciplinary topics;
- these techniques also help to stimulate students by explaining historical (usually informal) motivations – often paradox-related motivations – behind different concepts and ideas of mathematics and science.


In assessment, interval and fuzzy techniques help:

- to design a better grading scheme for test and assignments, a scheme that stimulates more effective learning,
- to provide a more adequate individual grading of contributions to group projects – by taking into account subjective estimates of different student distributions (and the uncertainty of these estimates), and
- to provide a more adequate description of the student knowledge and of the overall teaching effectiveness.

The talk summarizes, combines, and expands on the ideas and results, some of which published in journals and conference proceedings. These published papers also contain additional technical details and practical examples of using these ideas.

3.11 A Comparison of Different Kinds of Multiple Precision and Arbitrary Precision Interval Arithmetics

Walter Krämer (*Universität Wuppertal, DE*)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Walter Krämer

Joint work of Krämer, Walter; Blomquist, Frithjof; Hofschuster, Werner
URL http://www2.math.uni-wuppertal.de/org/WRST/index_de.html

The current version of the C++ class library C-XSC for verified numerical computing offers quite a lot of different interval data types with different properties. We compared multiple precision data types like staggered precision data types (unevaluated sums of floating-point numbers) as well as arbitrary precision types based on arrays of integers and integer operations. In some respects staggered numbers and operation are restricted by properties of the underlying basic floating-point data type (IEEE double precision) whereas arbitrary precision numbers are only limited by the memory resources available.

We presented some preliminary execution time comparisons and gave some advice when it is appropriate to use a specific multiple/arbitrary precision C-XSC data type. We also discussed the availability of some underlying external packages restricting the use of C-XSC's arbitrary precision data types on some platforms. Several source code examples were presented to demonstrate the ease of use of the data types (due to operator and function name overloading) and the power of the different multiple/arbitrary precision C-XSC packages.

Further features of C-XSC have been discussed in the talk “C-XSC – Overview and new developments” presented by Michael Zimmer during the course of this Dagstuhl seminar. Please refer to the corresponding abstract included in this document.





Keywords: Multiple precision, arbitrary precision, staggered data types, C-XSC, MPFR, MPFI, interval computations.

References

- 1 E. Adams and U. Kulisch. *Scientific Computing With Automatic Result Verification*. Academic Press, Inc., 1993.
- 2 F. Blomquist, W. Hofschuster, and W. Krämer. A modified staggered correction arithmetic with enhanced accuracy and very wide exponent range. In *Lecture Notes in Computer Science LNCS 5492*, A. Cuyt, W. Krämer, W. Luther, and P. Markstein (Eds.), pages 41–67, Springer, 2009.
- 3 L. Fousse, G. Hanrot, V. Lefèvre, P. Pelissier, and P. Zimmermann. MPFR: A multiple-precision binary floating-point library with correct rounding. *ACM Transactions on Mathematical Software*, 33(2), Article 13, 2007.
- 4 M. Grimmer, K. Petras, and N. Revol. Multiple precision interval packages: Comparing different approaches. In *Numerical Software with Result Verification, Lecture Notes in Computer Science*, Vol. 2991, R. Alt, A. Frommer, R.B. Kearfott, W. Luther (Eds.), pages 64–90, Springer, 2004.
- 5 W. Hofschuster and W. Krämer. C-XSC 2.0: A C++ library for extended scientific computing. In *Numerical Software with Result Verification, Lecture Notes in Computer Science*, Volume 2991, Springer, pages 15–35, 2004.
- 6 R. Klatte, U. Kulisch, A. Wiethoff, C. Lawo, and M. Rauch. *C-XSC – A C++ Class Library for Extended Scientific Computing*. Springer, 1993.
- 7 W. Krämer. Multiple precision computations with result verification. In *Scientific Computing With Automatic Result Verification*, Academic Press, pages 325–356, 1993.
- 8 W. Krämer, U. Kulisch, and R. Lohner. *Numerical Toolbox for Verified Computing II – Advanced Numerical Problems (draft)*, chapter 7 Multiple-precision arithmetic using integer operations, pages 210–251, 1998. <http://www2.math.uni-wuppertal.de/wrswt/literatur/tb2.ps.gz>
- 9 R. Lohner. Interval arithmetic in staggered correction format. In *Scientific Computing With Automatic Result Verification*, Academic Press, pages 301–342, 1993.
- 10 N. Revol and F. Rouillier. Motivations for an arbitrary precision interval arithmetic and the MPFI library. *Reliable Computing*, 11:275–290, 2005.
- 11 F. Blomquist, W. Hofschuster, and W. Krämer. C-XSC-Langzahlarithmetiken für reelle und komplexe Intervalle basierend auf den Bibliotheken MPFR und MPFI. Preprint 2011/1, Universitaet Wuppertal, 2011.

3.12 Towards Optimal Representation and Processing of Uncertainty for Decision Making, on the Example of Economics-Related Heavy-Tailed Distributions

Vladik Kreinovich (University of Texas – El Paso, US)

License     Creative Commons BY-NC-ND 3.0 Unported license
© Vladik Kreinovich

Uncertainty is usually gauged by using standard statistical characteristics: mean, variance, correlation, etc. Then, we use the known values of these characteristics (or the known bounds on these values) to select a decision. Sometimes, it becomes clear that the selected characteristics do not always describe a situation well; then other known (or new) characteristics are proposed. A good example is description of volatility in finance: it started with variance, and now many descriptions are competing, all with their own advantages and limitations.


Another good example is the case of heavy-tailed distributions frequently occurring in economics and finance: for these distributions, variance is infinite and thus, we cannot use variance to describe deviations from the mean.

In such situations, a natural idea is to come up with characteristics tailored to specific application areas: e.g., select the characteristic that maximize the expected utility of the resulting risk-informed decision making. As a case study, we found optimal characteristics for measures of deviation and dependence in financial applications – where, for heavy-tailed distributions, traditional variance and correlation cannot be used.

With the new characteristics, comes the need to estimate them when the sample values are only known with interval uncertainty. Algorithms originally developed for estimating traditional characteristics can often be modified to cover new characteristics.

3.13 From Processing Interval-Valued Data to Processing Fuzzy Data: A Tutorial

Vladik Kreinovich (University of Texas – El Paso, US)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Vladik Kreinovich

Some quantities y are difficult (or impossible) to measure or estimate directly. A natural solution is to measure them indirectly, i.e., to measure auxiliary quantities x_1, \dots, x_n which are related to y by a known dependence $y = f(x_1, \dots, x_n)$, and then use the results \tilde{x}_i of these measurements to compute an estimate $\tilde{y} = f(\tilde{x}_1, \dots, \tilde{x}_n)$ for y . Measurements are never absolutely accurate; as a result, the measurement results \tilde{x}_i are, in general, different from the actual (unknown) values x_i of the corresponding quantities. Hence, the estimate \tilde{y} is, in general, different from the desired value y . How can we estimate the difference $\Delta y = \tilde{y} - y$? In some cases, we know the probabilities of different measurement errors $\Delta x_i = \tilde{x}_i - x_i$; however, often, we do not know these probabilities, we only know the upper bounds Δ_i on the measurement errors – upper bounds provided by the manufacturers. In this case, after we know the measurement result \tilde{x}_i , we know that the actual value x_i belongs to the interval $[\tilde{x}_i - \Delta_i, \tilde{x}_i + \Delta_i]$, and we can use interval computations to find the interval of possible values of y .

Often, in addition to the guaranteed bounds Δ_i on the measurement errors, we also have expert estimates on these bounds which hold only with some degree of certainty – and which are described by words from natural language like “small”, “most probably”, etc.. Fuzzy techniques were specifically designed to process such estimates. The main idea is that, for each “fuzzy” (natural-language) property like “small”, and for each real value x , we describe the degree $d(x)$ to which x satisfies this property – e.g., by polling experts. Once we know the degrees $d_i(x_i)$ to which each x_i satisfies the corresponding expert property, we need to combine these degrees into a degree with which all the values x_i satisfy their properties. We show how this need leads to a complex formula called Zadeh’s extension principle, and how from the computational viewpoint, it means that for each real value $\alpha \in [0, 1]$, to find all the values for which $d(y) \geq \alpha$ (“ α -cut” of y), we can apply interval computations to the corresponding α -cuts of x_i . Thus, from the computational viewpoint, processing such fuzzy inputs can be (and usually is) reduced to interval computations.

3.14 Generating a Minimal Interval Arithmetic Based on GNU MPFR

Vincent Lefèvre (*ENS – Lyon, FR*)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Vincent Lefèvre


Searching for the hardest-to-round cases for the correct rounding of some function f on an interval I in a fixed precision can be done efficiently by first approximating the considered function by a polynomial, on which specific algorithms are then applied. One also needs to determine an enclosure of the range $f(I)$, more precisely the exponent range.

Our implementation currently uses Maple and the `intpakX` interval arithmetic package in order to compute both the exponent range and the polynomial approximation. But Maple/`intpakX` has various drawbacks.

The GNU MPFR library has since been available and could be used for our computations in arbitrary precision. But we need an interval arithmetic on top of it. As reliability matters more than performance in this context, we seek to implement a minimal interval arithmetic by generating code on the fly using MPFR. The implementation should be as simple as possible so that it could easily be checked and/or proved formally.

3.15 IPPToolbox – a package for imprecise probabilities in R

Philipp Limbourg (*Universität Duisburg-Essen, DE*)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Philipp Limbourg


Joint work of Limbourg, Philipp; Rebner, Gabor; Auer, Ekaterina; Luther, Wolfram

A lot of researchers working on Dempster-Shafer, imprecise probabilities etc., e. g. in systems reliability engineering, structural reliability and similar fields. While the theory stems from the 60s & 70s, it was originally mainly used for expert systems & automatic reasoning. Imprecise probabilities were brought into uncertainty modelling practice by Sandia laboratories in US (2004).

However, most people use their own, proprietary codes. This talk presents an R and Matlab package – the IPP toolbox, an open-source package for imprecise probability calculations. The package includes R help and two examples (Flood, Reliability). Available on CRAN (package “`ipptoolbox`”), it is the predecessor of the DSI toolbox Application areas: 100+ downloads of Matlab version (R figures not known, at CRAN), applied at Electricité de France R&D, e. g. flood modelling, event trees, design optimization.

3.16 Constrained Intervals and Interval Spaces

Weldon A. Lodwick (*University of Colorado, US*)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Weldon A. Lodwick

Constrained intervals, intervals as a mapping from $[0,1]$ to linear functions with non-negative slopes, and arithmetic on constrained intervals generate a space that turns out to be a cancellative Abelian monoid albeit with a richer set of properties than standard interval arithmetic. This means that not only do we have the classical embedding as developed by H.

Radström and S. Markov but directly the properties of the subset of these polynomials. We study a little of the geometry of the embedding of intervals into a quasi-vector space and some of the properties of the mapping of constrained intervals into a space of polynomials. Thus, there are two parts to this talk. (1) The representation of intervals as linear polynomials with non-negative slopes. (2) The algebraic structure of this new representation. The geometry is mentioned in passing as a way to visualize the embedded space and will not be discussed further. The theoretical reason for considering a new representation of intervals is to have a formalization in (a subset of) polynomial space with the view to evaluate expressions (functions) of intervals. The theoretical reason for considering the algebraic structure of the embedding into a space with inverses is to solve equations. We only look at additive inverses in this presentation.

3.17 Verification and Validation Requirements in Biomechanical Modeling and Simulation

Wolfram Luther (Universität Duisburg-Essen, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license

© Wolfram Luther

Joint work of Luther, Wolfram; Auer, Ekaterina; Chuev, Andrey; Cuypers, Roger; Kiel, Stefan

URL <http://www.scg.inf.uni-due.de/fileadmin/Veroeffentlichungen/papereuromech.pdf>

We give an overview on how accurate and verified methods can be employed for several biomechanical processes. Our focus is on a broad field of patient specific preoperative surgical planning based on the superquadric (SQ) geometrical modeling. We show how to use verified tools to efficiently and reliably implement important parts of the processing pipeline. Furthermore, we describe our current activities to verify SQ-based operations numerically and to validate the models and their parameters by various measurement methods. For this purpose, we replace or combine data types, algorithms and tools with their verified versions where possible. Here, we might consider using C-XSC (Krämer) instead of plain C++, INTLAB (Rump) in addition to Matlab, DSI (Dempster Shafer tool with intervals) instead of IPP (the Imprecise Probabilities toolbox, Limbourg, Rebner) or employing stochastic data types as in Cadna++ (Lamotte et al.) or the static program analyzer Fluctuat (CEA).

Recently, we introduced four classes for the use in V&V assessment, from lowest to highest certification standard. A process implementation that relies on standard floating-point or fixed-point arithmetic with unverified results belongs to Class 4. If the system is to qualify for Class 3, the numerical implementation of the process needs to employ at least standardized IEEE 754-2008 floating-point arithmetic. Furthermore, sensitivity analysis has to be carried out for uncertain parameters and uncertainty propagated throughout the subsystems using various methods. Additionally, a priori/posteriori error bounds should be provided for important sub-processes, condition numbers computed, failure conditions identified. To belong to Class 2, relevant subsystems have to be implemented using tools with result verification or with an accompanying computation of reliable error bounds. The tools should use language extensions, the convergence of numerical algorithms must be proved via existence theorems, analytical solutions, computer-aided proofs or fixed-point theorems. In Class 1, uncertainty is quantified and propagated throughout the process using interval computing. Model parameters are optimized by calibration. The whole system is verified using tools with result verification. Basic numeric algorithms and (special) functions are certified. Alternatively, real number algorithms, analytical solutions or computer-aided

existence proofs are used.


We managed to classify several biomechanical processes. Recently, a dynamical gait simulation based on motion tracking under uncertainty in parameters was described. Processes from a recently completed project PROREOP have been used to perform elements of V&V assessment during the designing step.

Our analysis shows that such assessment should begin with the specification of the process and its sub-processes, the design of the building-blocks and their software modules, the definition of interfaces and data flows, and, finally, the selection and adaptation of appropriate data types and algorithms. PROREOP aimed at developing and evaluating a highly interactive prosthesis planning tool that allowed surgeons to assess 3D imaging data and to use geometrical, mechanical, kinematical and material/surface-specific bone features of the patient as primary sources for their decisions.

In our talk, we focus on an example addressing superquadric (SQ) bone/prosthesis modeling, prosthesis fitting into the medullary space of the routed femoral shaft and the total hip arthroplasty (THA), which was broadly discussed in the PhD-thesis of the fourth author. The following hardware and software blocks are analyzed: Data acquisition by MRI and CT imagery, bone and muscle segmentation (Class 3), SQ modeling with an in/out decision algorithm, distance computation between convex SQs (both Class 1), compound model optimization (Class 3), K-segments algorithms (Class 2), feature extraction with various validation approaches, verified distance computation between compound models (both Class 1), and pose computation (Class 3).

3.18 Enclosing solutions of initial-value problems with large uncertainty

Arnold Neumaier (Universität Wien, AT)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Arnold Neumaier

Joint work of Neumaier, Arnold; Fazal, Qaisra


We consider the enclosure of initial-value problems for ordinary differential equations where the initial condition has a large uncertainty. Uncertainties are described at each time by means of enclosing ellipsoids. Our new approach combines

- work by Chernousko for linear ODEs, who derived for this class differential equations for the parameters of the enclosing ellipsoid,
- work by Kühn based on defect estimates and curvature bounds,
- new results on conditional differential inequalities for validating error bounds, and
- global optimization techniques for verifying the assumptions needed to apply the conditional differential inequalities.

The approach was implemented in Matlab, using automatically generated AMPL files as an interface to optimization algorithms. For simple examples, the performance was essentially the same as the state-of-the-art packages VSPODE, VNODE-LP, and VALENCIA-IVP, while for highly nonlinear problems, it gave much superior error bounds.

3.19 Characterizing AE Solution Sets to Parametric Linear Systems

Evgenija D. Popova (Bulgarian Academy of Sciences, BG)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Evgenija D. Popova

Consider linear systems whose input data are linear functions of uncertain parameters varying within given intervals. Such systems are common in many engineering analysis or design problems, control engineering, robust Monte Carlo simulations, etc., where there are complicated dependencies between the model parameters which are uncertain. Various solution sets to a parametric linear system can be defined depending on the way the parameters are quantified by the existential and/or the universal quantifiers.

We are interested in an explicit description of the so-called AE parametric solution sets (where all universally quantified parameters precede all existentially quantified ones) by a set of inequalities not involving the parameters. The problem is related to quantifier elimination where Tarski's general theory is EXPSPACE hard and a lot of research is devoted to special cases with polynomial-time decidability.

In this talk we present how to obtain an explicit description of AE parametric solution sets by combining a modified Fourier-Motzkin-type elimination of existentially quantified parameters with the elimination of the universally quantified parameters.


Some necessary (and sufficient) conditions for existence of non-empty AE parametric solution sets are discussed, as well as some properties of the AE parametric solution sets.

Explicit description of particular classes of AE parametric solution sets (tolerable, controllable, any 2D) is presented.

Numerical examples illustrate the solution sets and their properties. A comparison to results obtained by quantifier elimination demonstrates the advantage of the presented approach.

3.20 What you always wanted to know about decorated intervals

John D. Pryce (Cardiff University, GB)

License  Creative Commons BY-NC-ND 3.0 Unported license
© John D. Pryce

Work on an Interval Arithmetic Standard has been under way since 2008, in IEEE Working Group P1788. Early on, we chose a “silent” paradigm for function evaluation, namely interval evaluation of a library function partly or wholly outside its domain returns an enclosure of those values that are defined (e.g. $1/[0,1]$ gives $[1,\text{infinity})$; $\text{sqrt}([-2,-1])$ gives the empty set), instead of throwing an exception as in current interval systems.

This raises the question of how to record that such an exceptional event has occurred. This is necessary, since various important interval algorithms need to determine rigorously whether a function, given by an expression, has properties such as being defined, or defined and continuous, everywhere on a box. Examples are branch-and-bound search methods to “pave” space regions defined by inequalities; or validated ODE methods that apply Brouwer's fixed-point theorem. In current interval systems, such information can only be determined to a limited extent, and by clumsy ad hoc means.

Summarising discussions over the past 15 months, this talk aims to show why and how it is feasible to record such data automatically by suitably enhancing the interval versions of

library functions. The data might be stored in global flags similar to the IEEE floating point flags “zerodivide”, etc.; or locally by attaching it as “decorations” to computed intervals. It explains why we favour the latter. It presents the Neumaier-Hayes idea of arranging decoration information in a linearly ordered sequence of values that can be considered to go from “best” to “worst”, and compares several such schemes.

It discusses some problematic and contentious issues such as the status of intersection/union operations. It outlines the proposed Compressed Intervals, which gain speed at the cost of limited decoration support. They use the same space (typically 16 bytes) as undecorated intervals, and suffice for the applications mentioned above.

3.21 Verified Parameter Estimation for the Thermal Behavior of High-Temperature Fuel Cells

Andreas Rauh (Universität Rostock, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license

© Andreas Rauh

Joint work of Rauh, Andreas; Dötschel, Thomas; Auer, Ekaterina; Aschemann, Harald

The thermal behavior of high-temperature solid oxide fuel cell (SOFC) systems is characterized by the interaction of different physical and electro-chemical processes. In particular, these processes take place in the anode and cathode gas manifolds as well as in the interior of the fuel cell stack module.

The chemical reactions of hydrogen at the anode and oxygen at the cathode have to be enabled to generate electricity and process heat simultaneously. The prerequisite for these reactions is both ion conduction through the electrolyte and electron interchange between the electrodes [1]. The fuel cell stack module itself consists of a thermal insulation and an assembly of anode-electrolyte-cathode elements, which are electrically connected in series.

One possible future application of fuel cells is the use in decentralized supply systems for process heat and electricity. For this type of application, it is necessary to operate SOFCs with a time-varying electric load. It is desired to implement operating strategies for variable electric loads with smallest possible battery buffers acting as electric load shaping devices. This leads to the necessity for advanced control strategies minimizing the influence of electric load variations on the resulting changes of the cell temperature and its local distribution.

The decoupling of electric load variations from the thermal behavior of the SOFC is crucial from an application point of view: Mechanical strain introduced by large spatial gradients in the temperature distribution within the fuel cell stack module along with local over-temperatures may lead to an accelerated degradation and — in the worst case — to the destruction of the fuel cell materials. Moreover, the maximization of the efficiency of a high-temperature SOFC is commonly linked to an increase of the overall temperature level in the interior of the stack module [2]. This fact imposes further demands on the accuracy of the mathematical system model used for control synthesis.

The reliable operation of SOFC systems by means of accurate control strategies requires a sufficient knowledge of the ongoing physical processes. Therefore, these processes are expressed in terms of a mathematical system model with explicitly given bounds for the uncertain quantities. This system model has to be usable for the online analysis and prediction of the influence of electric load variations and for its compensation in real time by means of nonlinear feedback control strategies. The same holds for the online estimation

and identification of non-measured internal process variables with the help of state and disturbance observers.

The real-time applicability of the mathematical system models can be achieved by using control-oriented descriptions which take into account the dominating dynamic effects and spatial variations of process variables such as temperatures and partial pressures of the gas fractions. These control-oriented descriptions commonly replace sets of (nonlinear) partial differential equations by finite-dimensional sets of ordinary differential equations and algebraic relations [4].

To parameterize these low-dimensional, control-oriented system models, identification routines are employed. The parameter identification is based on experimental data gathered from a test rig available at the Chair of Mechatronics at the University of Rostock. Classical floating point approaches for parameter identification (which are based on local optimization procedures) cannot be used to incorporate uncertainty in measured data directly in the identification procedure.

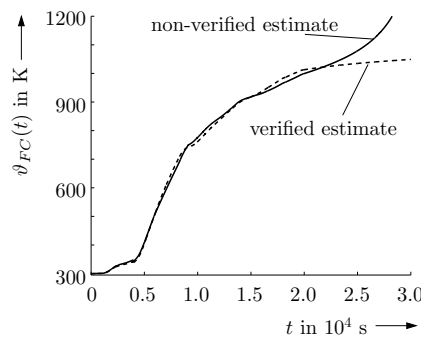
For this reason, a first version of a verified interval-based parameter identification routine is presented in this contribution. This routine is applied to determine those parameter ranges which are consistent with measured data under explicit consideration of their uncertainty. The interval-based identification routine is a generalization of a verified optimization procedure developed by the authors for the design of reliable optimal controllers for uncertain continuous-time and discrete-time processes [3]. Moreover, the identification results obtained by the above-mentioned interval-based routine are compared with a classical floating point implementation for the thermal subsystem of the SOFC currently involving a total number of 26 a-priori unknown or uncertain parameters.

A simplified thermal system model is given by the scalar nonlinear differential equation

$$\dot{\vartheta}_{FC} = \frac{1}{c_{FC} m_{FC}} \left[\frac{1}{R_A} (\vartheta_A - \vartheta_{FC}) - (c_{N_2} \zeta_{N_2,C} + c_{O_2} \zeta_{O_2}) \cdot \dot{m}_{CG} \cdot (\vartheta_{FC} - \vartheta_{CG,in}) - (c_{H_2} \dot{m}_{H_2} + c_{H_2O} \dot{m}_{H_2O} + c_{N_2} \dot{m}_{N_2,A}) \cdot (\vartheta_{FC} - \vartheta_{AG,in}) + \frac{\Delta H_m(\vartheta_{FC}) \dot{m}_{H_2}}{M_{H_2}} \right],$$

which describes the temperature ϑ_{FC} in the fuel cell stack. This temperature is assumed to be homogeneously distributed. It depends on the mass flow rate \dot{m}_{CG} of cathode gas (CG) as well as the stoichiometrically balanced mass flow \dot{m}_{H_2} of hydrogen, \dot{m}_{H_2O} of vaporized water, and $\dot{m}_{N_2,A}$ of nitrogen at the anode. The inlet temperature of the cathode gas is given by $\vartheta_{CG,in}$, the corresponding anode inlet temperature is denoted by $\vartheta_{AG,in}$. Second-order polynomial approximations $c_x(\vartheta_{FC}) = \sum_{i=0}^2 \alpha_{x,i} \vartheta_{FC}^i$ are used to describe the temperature dependencies of all heat capacities, $x \in \{N_2, O_2, H_2, H_2O\}$, as well as the reaction enthalpy $\Delta H_m(\vartheta_{FC})$. According to Fig. 1, the estimation of the parameters of the thermal system model by a verified optimization routine leads to a system parameterization that is asymptotically stable beyond the range of the temperature values used for the parameter identification. In contrast, the simulated system output shows an unstable time response if the parameters are identified on the basis of a purely non-verified local optimization procedure.

Finally, an outlook has been given on how to employ the uncertain dynamic system model in real time for robust online control. This control framework will be designed by using interval arithmetic techniques guaranteeing both accuracy and stability in a rigorous way. For that purpose, the control synthesis is based, for example, on sliding mode techniques to guarantee stability of the system dynamics in spite of parameter uncertainties [5].







■ **Figure 1** Comparison of simulated temperatures for verified and non-verified parameter identification

References

- 1 R. Bove and S. Ubertini (Eds.). *Modeling Solid Oxide Fuel Cells*. Springer-Verlag, Berlin, 2008.
- 2 A. Gubner. Non-isothermal and dynamic SOFC voltage-current behavior. In *Solid Oxide Fuel Cells IX (SOFC-IX): Vol. 1, Cells, Stacks and Systems*, S. C. Singhal and J. Mizusaki, Eds. The Electrochemical Society, pages 814–826, 2005.
- 3 A. Rauh, J. Minisini, E. P. Hofer, and H. Aschemann. Robust and optimal control of uncertain dynamical systems with state-dependent switchings using interval arithmetic. *Reliable Computing*, 15(4):333–344, 2011.
- 4 A. Rauh, T. Dötschel, H. Aschemann. Experimental parameter identification for a control-oriented model of the thermal behavior of high-temperature fuel cells. In *Proc. of IEEE Intl. Conference on Methods and Models in Automation and Robotics MMAR 2011*, Miedzyzdroje, Poland, 2011.
- 5 V. Utkin. *Sliding Modes in Control and Optimization*. Springer-Verlag, Berlin, Heidelberg, 1992.

3.22 Verified Add-ons for the DSI toolbox

Gabor Rebner (Universität Duisburg-Essen, DE)

License     Creative Commons BY-NC-ND 3.0 Unported license
© Gabor Rebner

The Dempster-Shafer with Intervals (DSI) toolbox [1], which is based on the imprecise probability toolbox (IPP) [5], is an extension for MATLAB that provides algorithms for verified computing with basic probability assignments. Throughout this talk, we use the term verification in its narrow sense of referring to a mathematical proof for correctness of a result obtained by a computer calculation. One problem of computer calculations is the appearance of rounding errors, which are unavoidable because of the finite nature of floating-point arithmetic. Even more important is to allow for uncertainty in a simulation model. Verified methods offer a solution in these cases, motivating us to use them in combination with probabilistic methods. In this talk, we present a new verified implementation of Markov set chains (MSC) [4], a C-XSC [3] interface to MATLAB and an improvement in the evaluation of monotonic and non-monotonic system functions in DSI. Furthermore we demonstrate the ability of DSI to handle uncertainty under Dempster-Shafer theory in a software presentation.

This talk is structured as follows. First, we review the DSI toolbox and its ability to handle uncertain data in a verified way. Next, we discuss our verified implementation of MSC. MSC extend classic Markov chains by defining uncertain transition matrices and initial vectors. For example, a nuclear power plant cannot be modeled by Markov chains because of the deterministic behavior of the applied transition matrices. By utilizing MSC, we are able to describe the system's operation time and environmental influences in terms of a nondeterministic transition matrix. Besides, the use of verified algorithms allows us to cope with the limitations of floating point arithmetic. Such algorithms provide an enclosure which is guaranteed to contain the exact result.

We illustrate the functionality of our new implementation using a close-to-life example. In the next part of our contribution, we present an implementation of a C-XSC to MATLAB interface. The goal of this add-on is to make algorithms written in C-XSC accessible in DSI. To avoid conversion errors, we use the MATLAB build-in interface to access MATLAB memory directly. We discuss this implementation by considering the example of the error function. Finally, we demonstrate the ability of DSI to sample monotonic and non-monotonic functions in a verified way. In order to minimize the overestimation and the computation time, we implemented a monotonicity test using automatic differentiation provided by INTLAB [6]. To compute a tight enclosure of the solution space of a function which is monotonic and contains exclusively basic arithmetic operations $\{+, -, *, /\}$ and their compositions, we utilize floating point arithmetic with directed rounding. Otherwise, we have to use interval arithmetic to get a verified enclosure of the solution. We close our talk by giving illustrative examples of sampling $f(x) = x \cdot x - x$ with x uniform distributed in two to three.

In the course of the software presentation, we demonstrate the ability of DSI to handle basic probability assignments (BPA) [2] in a verified way.


This demonstration is split into two parts. First we show the option to define cumulative distribution functions with uncertain bounds. As an example we use the triangle distribution with the uncertain mean, the lower, and the upper bound. The second part of our demonstration deals with (non-)monotonic function propagation. We take the function $f(x) = \sin(x^2) + x^3$ with x triangularly distributed and having uncertainties in the bounds and the mean. Furthermore, we illustrate the ability of DSI to detect erroneous user inputs. Finally, we demonstrate the C-XSC to MATLAB interface.

References

- 1 E. Auer, W. Luther, G. Rebner, and P. Limbourg. A Verified MATLAB Toolbox for the Dempster-Shafer Theory. In M. Arnaud, editor, *Proceedings of the Workshop on the Theory of Belief Functions*, Brest, April 2010. <http://bfas.iutlan.univ-rennes1.fr/belief2010/html/papers/p83.pdf> <http://udue.de/DSI>.
- 2 S. Ferson, V. Kreinovich, L. Ginzburg, D.S. Myers, and K. Sentz. Constructing probability boxes and Dempster-Shafer structures. Sand Report SAND2002-4015, Sandia National Laboratories, 2003.
- 3 R. Hammer, M. Hocks, U. Kulisch, and D. Ratz. *C++ Toolbox for Verified Computing*. Springer-Verlag Berlin Heidelberg New York, 1995.
- 4 D. Hartfiel. *Markov Set-Chains*. Lecture notes in mathematics; 1695. Springer, 1998. ISBN 3-540-64775-9.
- 5 P. Limbourg. Imprecise Probability Propagation Toolbox (IPP Toolbox). <http://www.uni-due.de/il/ipptoolbox.php>.
- 6 S. Rump. INTLAB – INTerval LABoratory. In Tibor Csendes, editor, *Developments in Reliable Computing*, pages 77–104. Kluwer Academic Publishers, Dordrecht, 1999. <http://www.ti3.tu-harburg.de/rump/>.

3.23 Refining Abstract Interpretation-based Approximations with Constraint Solvers

Michel Rueher (Université Nice Sophia Antipolis, FR)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Michel Rueher

Joint work of Ponsini, Olivier; Michel, Claude; Rueher, Michel

Main reference O. Ponsini, C. Michel, M. Rueher, “Refining Abstract Interpretation-based Approximations with Constraint Solvers,”

URL <http://hal.archives-ouvertes.fr/hal-00623274/fr/>


Programs with floating-point computations are tricky to develop because floating-point arithmetic differs from real arithmetic and has many counterintuitive properties. A classical approach to verify such programs consists in estimating the precision of floating-point computations with respect to the same sequence of operations in an idealized semantics of real numbers.

Tools like Fluctuat –based on abstract interpretation– have been designed to address this problem. However, such tools compute an over-approximation of the domains of the variables, both in the semantics of the floating-point numbers and in the semantics of the real numbers. This over-approximation can be very coarse on some programs. We show that constraint solvers over floating-point numbers and real numbers can significantly refine the approximations computed by Fluctuat.

Keywords: Program verification; Floating-point computation; C programs; Abstract interpretation-based approximation; Interval-based constraint solvers over real and floating-point numbers

3.24 Constraint Programming over Continuous Domains

Michel Rueher (Université Nice Sophia Antipolis, FR)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Michel Rueher

In this tutorial, we first recall the basics of the constraint programming framework, a *branch & prune* schema which is best viewed as an iteration of two steps:

1. *Pruning the search space*
2. *Making a choice to generate two (or more) sub-problems*

The pruning step is based on partial consistencies. It *reduces an interval* when it can prove that the upper bound or the lower bound does not satisfy some constraint. We outline the intuitions behind the most common partial consistencies and we analyze the relationship between these consistencies and interval arithmetic techniques. We detail some critical implementation issues of Hull-consistency, Box-consistency and Quad-consistency.

The branching step *splits the interval* associated to some variable in two intervals (often with the same width). We give a short overview of the most effective search heuristics used in this process.


To conclude, we illustrate the powerful *refutation capabilities* of local consistencies on two applications: *boosting OBR* [2] in global optimisation; *refining approximations* (see <http://hal.archives-ouvertes.fr/hal-00623274/fr/>) in program verification.

References

- 1 F. Benhamou and L. Granvilliers. *Handbook of Constraint Programming*, chapter Continuous and Interval Constraints. Elsevier 2004.
- 2 A. Goldsztejn, Y. Lebbah, C. Michel, and M. Rueher. Capabilities of constraint programming in safe global optimization. *Reliable computing* 15(1):47–59, 2011.
- 3 R. B. Kearfott. Interval analysis: Unconstrained and constrained optimization. *Encyclopedia of Optimization*, pages 1727–1730, 2009.
- 4 P. Van-Hentenryck, D. McAllester, and D. Kapur. Solving polynomial systems using branch and prune approach. *SIAM Journal on Numerical Analysis*, 34(2):797–827, 1997.

3.25 Managing uncertainty and discontinuous condition numbers in finite-precision geometric computation

Peihui Shao (*Université de Montréal, CA*)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Peihui Shao


Joint work of Shao, Peihui; Stewart, Neil

Roundoff and representation errors in geometric computation may be modelled as epistemic or aleatory processes. In either case, if it is required that the computed result have correct topological form, then problems such as computing regularized Boolean operations are fundamentally ill-conditioned. In this paper it is shown that if we wish to drive the process through a discontinuity in the condition number, and if we wish to prove rigorous theorems, then use of an exception-handling mechanism analogous to those used in programming languages cannot be avoided.

The above observations show that traditional approaches to proving robustness, when computation is done using ordinary IEEE floating-point arithmetic, are inappropriate. We discuss the nature, in the context described above, of the theorems that should be proved, and we give a very simple result that illustrates our approach.

3.26 Reliable Kinetic Monte Carlo Simulation based on Random Set Sampling

Yan Wang (*Georgia Institute of Technology, US*)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Yan Wang

The kinetic Monte Carlo (KMC) method has been widely used in simulating rare events such as chemical reactions or phase transitions. The lack of complete knowledge of transitions and the associated rates is one major challenge for accurate KMC predictions. In this work, a reliable KMC (R-KMC) mechanism is developed to improve the robustness of KMC results, where propensities are interval estimates instead of precise numbers and sampling is based on random sets instead of random numbers. A multi-event algorithm is developed for event selection, and the system time is advanced based on best- and worst-case scenarios. The weak convergence of the multi-event algorithm towards traditional KMC is demonstrated with an interval master equation.

References

- 1 Y. Wang. Reliable kinetic Monte Carlo simulation based on random set sampling. In *Proc. 2011 ASME International Design Engineering Technical Conferences & The Computer and Information in Engineering Conference*, Aug. 29-31, 2011, Washington, DC, Paper No.DETC2011-48575.

3.27 The General Interval Power Function

Jürgen Wolff von Gudenberg (Universität Würzburg, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license

© Jürgen Wolff von Gudenberg

Joint work of Heimlich, Oliver; Nehmeier, Marco; Wolff von Gudenberg, Jürgen

The purpose of this article is to find a general power function for use in interval arithmetic, particularly with regard to the upcoming interval arithmetic standard which is being developed by the IEEE working group P1788.

Apparently, in the history of mathematics exponentiation has sometimes been used more or less as an abbreviatory notation and several definitions from different backgrounds, i.e., algebra and analysis, have been combined in people's minds.

For the set of real numbers these definitions agree on many points, but not on all. Points at issue are treated differently depending on the context, which poses problems searching for a common standard.

One contentious issue with the definition of general exponentiation is the assignment of real result to powers with negative base and rational exponent. We discuss three variants each of which has its merits and drawbacks.

The interval extensions are presented for each of the three variants, and efficient algorithms are implemented in INTLAB.

Finally, we recommend two of the three variants for interval libraries.

3.28 C-XSC – Overview and new developments

Michael Zimmer (Universität Wuppertal, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license

© Michael Zimmer

C-XSC is a powerful C++ library for verified numerical computing. It provides many useful data types, among them real and complex intervals, (interval) vectors and matrices and a long accumulator for the computation of dot product expressions in high accuracy. Also included is a toolbox with implementations of many useful verified algorithms. In the past few years, there have been some major extensions to the C-XSC library, especially with regard to the use of C-XSC in High Performance Computing. In our presentation, we gave a short overview of its features and then focused especially of the new developments in recent versions.

The first addition is the ability to compute dot products and dot product expressions in K -fold double precision. The desired precision can be changed at runtime, so that the user can choose between higher accuracy or higher speed for each computation. The second major addition is the optional support of the BLAS library for all vector-vector, matrix-vector and

matrix-matrix products in double precision. The use of these algorithms can drastically increase the performance of such operations.

Another major addition are new data types for sparse matrices and vectors. These allow to work with such vectors and matrices in a very efficient way, both in terms of memory consumption and computing speeds. The new data types provide an easy to use through the use of operator overloading. The sparse data types are based on widely used data structures which makes it easy to write interfaces to other sparse matrix software.

Also among the new features are new data types for multiple and arbitrary precision arithmetic, which were covered by the talk of Walter Krämer during the course of this Dagstuhl seminar.

The last presented new feature was the drastically improved thread safety of C-XSC, which makes it easy for the user to parallelize C-XSC programs for multicore machines, for example by using OpenMP. Finally, the talk also give a small outline of future developments of the C-XSC library.

During this Dagstuhl seminar, Gabor Rebner presented a talk about add-ons for the DSI toolbox using a self build interface between Matlab/Intlab and C-XSC using Matlabs Mex compiler to make use of the error function implemented in C-XSC. In a future work, a full interface between Matlab/Intlab and C-XSC might be possible.

In an additional software presentation at this Dagstuhl seminar, the installation of C-XSC on an example system (64 bit Macbook Pro running Mac OS X Lion) was demonstrated. Many of the options for compilation and optimization for the compilation of the core library and of C-XSC programs were covered in detail during this presentation. It should be stressed that a default installation of C-XSC only requires to start an installation script, accept the license by typing **yes** and then hitting enter 9 times to accept all the default settings.

Special thanks to Frithjof Blomquist as well as to some of our recent Master/Diploma students (Falko Sieg, Sascha Habicht, Frank Roitzsch, Christian Doescher, Daniel Kreuer, Michael Hirdes and Daniel Dakowski) for contributing to the development of C-XSC.

References

- 1 Weblink to C-XSC:
http://www2.math.uni-wuppertal.de/wrswt/xsc/cxsc_new.html
- 2 G. Alefeld and J. Herzberger. *Introduction to Interval Computations*. Academic Press, 1983.
- 3 L. S. Blackford, J. Demmel, J. Dongarra, I. Duff, S. Hammarling, G. Henry, M. Heroux, L. Kaufman, A. Lumsdaine, A. Petitet, R. Pozo, K. Remington, and R. C. Whaley. An updated set of basic linear algebra subprograms (BLAS). *ACM Trans. Math. Soft.*, 28(2):135–151, 2002.
- 4 A. Cuyt, W. Krämer, W. Luther, and P. Markstein. Numerical validation in current hardware architectures. In *Lecture Notes in Computer Science LNCS 5492*, Springer, 2009.
- 5 R. Hammer, M. Hocks, U. Kulisch, and D. Ratz. *Numerical Toolbox for Verified Computing I: Basic Numerical Problems*. Springer, 1993.
- 6 W. Hofschuster and W. Krämer. C-XSC 2.0: A C++ library for extended scientific computing. In *Numerical Software with Result Verification, Lecture Notes in Computer Science*, Volume 2991/2004, Springer, pages 15–35, 2004.
- 7 W. Krämer, M. Zimmer, and W. Hofschuster. Using C-XSC for high performance verified computing Accepted for publication (PARA2010).
- 8 U. Kulisch. *Computer Arithmetic and Validity: Theory, Implementation, and Applications*. De Gruyter Studies in Mathematics, 2008.
- 9 R. E. Moore, R. B. Kearfott, and M. J. Cloud. *Introduction to Interval Analysis*. SIAM, Philadelphia, 2009.

- 10 S. M. Rump. INTLAB – Interval Laboratory. *Developments in Reliable Computing*, pages 77–104, 1999.
- 11 S. M. Rump. Verification methods: Rigorous results using floating-point arithmetic *Acta Numerica*, 2010.
- 12 M. Zimmer, W. Krämer, G. Bohlender, and W. Hofschuster. Extension of the C-XSC library with scalar products with selectable accuracy. *Serdica Journal of Computing*, 4(3):349–370, 2010.
- 13 M. Zimmer, W. Krämer, and W. Hofschuster. Sparse matrices and vectors in C-XSC. *Reliable Computing*, 14:128–160, 2010.

4

 Schedule

Monday, September 12, 2011

9:00–9:30: Welcome
 9:30–10:00/10:15–10:45: *Ralph Baker Kearfott*, Interval Computations – Introduction and Significant Applications
 10:45–11:15: *Olga M. Kosheleva*, Degree-Based (Interval and Fuzzy) Techniques in Math and Science Education
 11:15–12:15: *Michel Rueher*, Constraint Programming over Continuous Domains
 14:00–14:30: *Wolfram Luther*, Verification & Validation Requirements in Biomechanical Modeling and Simulation
 14:30–15:00: *Andreas Rauh*, Verified Parameter Estimation for the Thermal Behavior of High-Temperature Fuel Cells
 15:30–15:30: *Ekaterina Auer*, Application of Verified Methods to Solving Non-smooth Initial Value Problems in the Context of Fuel Cell Systems
 16:00–17:00: *Vladik Kreinovich*, Tutorial: Foundations in Fuzzy Modeling
 17:00–17:30: *Michael Beer*, Fuzzy Probabilities and Applications in Engineering
 17:30–18:00: *Weldon A. Lodwick*, Constrained Intervals and Interval Spaces

Tuesday, September 13, 2011

9:00–9:30: *Jürgen Garloff*, Verified Solution of Finite Element Models for Truss Structures with Uncertain Node Locations
 9:30–10:00: *Milan Hladik*, Interval Linear Programming: Foundations, Tools and Challenges
 10:15–10:45: *Ralph Baker Kearfott*, A Review of Techniques for Handling Model Uncertainty in Interval-Based Global Optimization Procedures
 10:45–11:15: *Evgenija D. Popova*, Characterizing AE Solution Sets to Parametric Linear Systems
 11:15–11:45: *Olivier Mullier*, Under-approximation of the Range of Vector-Valued Functions having Different Dimensions for Domain and Codomain
 11:45–12:15: *Nathalie Revol*, Solving and Verifying Efficiently the Solution of a Linear System
 14:00–14:30: *Martin Fuchs*, Robust Optimization for Aerospace Applications
 14:30–15:00: *Neli Dimitrova*, Asymptotic Stabilization of a Bioprocess Model Involving Uncertainties
 15:00–15:30: *Jean-Luc Lamotte*, Analysis of Electronic Circuit with Point of View of Uncertainty
 16:00–16:30: *John D. Pryce*, Decorations for Dummies (1)

16:30–17:00: *Heinrich Rommelfanger*, Describing Vague Data by Fuzzy Intervals - Intelligent Ways for Solving Real-World Decision Problems and for Saving Information Costs

17:00–17:30: *Yan Wang*, Reliable Kinetic Monte Carlo Simulation based on Random Set Sampling

Wednesday, September 14, 2011

9:00–9:30: *Walter Krämer*, A Comparison of Different Kinds of Multiple Precision and Arbitrary Precision Interval Arithmetics

9:30–10:00: *Vincent Lefèvre*, Generating a Minimal Interval Arithmetic Based on GNU MPFR

10:15–11:15: *Philipp Limbourg and Gabor Rebner*, IPP Toolbox and Verified Add-ons for the DSI toolbox

11:15–11:45: *Jürgen Wolff von Gudenberg*, The General Interval Power Function

11:45–12:15: *Michael Zimmer*, C-XSC - Overview and New Developments

13:45–21:00: Excursion to Trier (guided tour, sight seeing, winery, dinner and wine tasting)

Thursday, September 15, 2011

9:00–9:30: *Cliff Joslyn and Emilie Hogan*, Intervals, Orders, and Rank

9:30–10:00: *Peihui Shao and Neil Stewart*, Managing Epistemic Uncertainty when Condition Numbers are Discontinuous

10:15–10:45: *Stefan Kiel*, Integration of Interval Contractors in Hierarchical Space Decomposition Structures

10:45–11:15: *Götz Alefeld*, Error Bounds for Nonlinear Complementary Problems with Band Structure

11:15–11:45: *Michel Rueher*, Using CSP Refutation Capabilities to Refine AI-based Approximations

14:00–14:30: *Vladik Kreinovich*, Towards Optimal Representation and Processing of Uncertainty for Decision Making, on the Example of Economics-Related Heavy-Tailed Distributions

14:30–15:00: *John D. Pryce*, Decorations for Dummies (2)

15:00–18:00: P 1788 Interval Standard Group (*N. Revol et al.*)

19:30–20:30: Software Presentation

Friday, September 16, 2011

9:00–9:30: *S. Rump*, Error Estimation of Floating-Point Summation and Dot Product

9:30–10:00: *Arnold Neumaier and Qaisra Fazal*, Enclosing Dynamical Systems with Large Initial Uncertainties

10:30–11:15: Strategy discussion

11:15–12:00: Closing session

Acknowledgements

We would like to thank the staff of Schloss Dagstuhl for their help in organizing this seminar and for the excellent facilities. Thanks go to Martin Fuchs for his help in collecting abstracts of the talks and other related materials for these proceedings.

Participants

- Götz Alefeld
KIT – Karlsruhe Institute of
Technology, DE
- Ekaterina Auer
Universität Duisburg-Essen, DE
- Michael Beer
University of Liverpool, GB
- Neli Dimitrova
Bulgarian Acad. of Sciences, BG
- Martin Fuchs
Cerfacs – Toulouse, FR
- Jürgen Garloff
HTWG Konstanz, DE
- Milan Hladík
Charles University – Prague, CZ
- Cliff Joslyn
Pacific Northwest National Lab. –
Seattle, US
- Ralph Baker Kearfott
Univ. of Louisiana –
Lafayette, US
- Stefan Kiel
Universität Duisburg-Essen, DE
- Olga M. Kosheleva
University of Texas – El Paso, US
- Walter Krämer
Universität Wuppertal, DE
- Vladik Kreinovich
University of Texas – El Paso, US
- Jean-Luc Lamotte
UPMC – Paris, FR
- Vincent Lefevre
ENS – Lyon, FR
- Philipp Limbourg
Universität Duisburg-Essen, DE
- Weldon A. Lodwick
University of Colorado, US
- Wolfram Luther
Universität Duisburg-Essen, DE
- Olivier Mullier
Supélec – Gif-sur-Yvette, FR
- Arnold Neumaier
Universität Wien, AT
- Evgenija D. Popova
Bulgarian Acad. of Sciences, BG
- John D. Pryce
Cardiff University, GB
- Andreas Rauh
Universität Rostock, DE
- Gabor Rebner
Universität Duisburg-Essen, DE
- Nathalie Revol
ENS – Lyon, FR
- Heinrich Rommelfanger
Univ. Frankfurt am Main, DE
- Michel Rueher
Université de Nice, FR
- Siegfried M. Rump
TU Hamburg-Harburg, DE
- Peihui Shao
Université de Montréal, CA
- Neil Stewart
Université de Montréal, CA
- Yan Wang
Georgia Inst. of Technology, US
- Jürgen Wolff von Gudenberg
Universität Würzburg, DE
- Michael Zimmer
Universität Wuppertal, DE



Quantum Cryptanalysis

Edited by

Serge Fehr¹, Michele Mosca², Martin Rötteler³, and
Rainer Steinwandt⁴

1 CWI – Amsterdam, NL, Serge.Fehr@cwi.nl

2 IQC, University of Waterloo, and Perimeter Institute, CA, mmosca@iqc.ca

3 NEC Laboratories America, Inc. – Princeton, US, mroetteler@nec-labs.com

4 Florida Atlantic University – Boca Raton, US, rsteinwa@fau.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 11381 “Quantum Cryptanalysis”. The first section gives an overview of the meeting, including organizational aspects. Subsequently abstracts of presentations at the meeting are provided (in alphabetical order).

Seminar 18.–23. September, 2011 – www.dagstuhl.de/11381

1998 ACM Subject Classification E.3 Code Breaking, F.2 Analysis of Algorithms and Problem Complexity, G.2 Discrete Mathematics, G.3 Probability and Statistics

Keywords and phrases Security of cryptographic schemes, quantum algorithms, computational hardness assumptions

Digital Object Identifier 10.4230/DagRep.1.9.58

Edited in cooperation with Florian Speelman

1 Executive Summary

Serge Fehr

Michele Mosca

Martin Rötteler

Rainer Steinwandt

License  Creative Commons BY-NC-ND 3.0 Unported license
© Serge Fehr, Michele Mosca, Martin Rötteler, and Rainer Steinwandt

Motivation and Goals

Cryptography aims at providing tools for securing information and preventing critical information-processing operations from adversarially provoked malfunction. These are very crucial objectives in today’s society where the importance of information is steadily increasing. As such, great effort is put into studying and implementing cryptographic schemes that offer privacy-protecting solutions for various tasks, and, wittingly or unwittingly, many people rely on cryptography in daily life. However, most of the cryptographic schemes that are currently in use rely on computational hardness assumptions that fail to hold in the presence of a quantum computer (like the hardness of factoring large integers or of computing discrete logarithms in certain cyclic groups). Thus, if brought to fruition, a large scale quantum computer will have a poignant impact on the security of cryptographic schemes. The following extreme opinions are commonly encountered:



Except where otherwise noted, content of this report is licensed under a Creative Commons BY-NC-ND 3.0 Unported license

Quantum Cryptanalysis, *Dagstuhl Reports*, Vol. 1, Issue 9, pp. 58–75

Editors: Serge Fehr, Michele Mosca, Martin Rötteler, and Rainer Steinwandt



DAGSTUHL
REPORTS

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

- *We'll deal with it later.* Subscribers to this point of view argue that because quantum computers currently are still in their infancy they do not pose a threat to existing schemes. Further, as it is unclear whether they will ever scale beyond the size of a handful qubits, it is not necessary to change security parameters of currently deployed practical systems, not to mention the need to switch to other systems with different (new) hardness assumptions.
- *Fight quantum with quantum.* Proponents of this view point out that the laws of quantum mechanics offer the possibility of information-theoretic security from first principles. Quantum key establishment is a prominent example of this line of research and (for shorter distances) has reached remarkable maturity. However, one should note that classical cryptographic components are often involved here, too— e.g. for ensuring authenticity, or even for encryption in order to reduce the amount of key material.

Both opinions have their merit, but offer no satisfying options for today's design of mid- and long-term secure cryptographic solutions, where a typical user does not have access to quantum links with other users, etc, as needed in quantum protocols. Completely ignoring the threat of a quantum computer looming on the horizon, means taking a systemic risk for real life applications. At the same time, there is no need for a panic type of reaction “just” because of an asymptotic threat for existing cryptographic infrastructures. Unfortunately, various cryptographic proposals made for “post quantum” cryptography build on hardness assumptions which have never been seriously cross-checked with experts in the design of quantum algorithms. This situation is rather unsatisfying and the goal of this Dagstuhl seminar was to pave the road towards a sound exploration of hardness assumptions and cryptographic protocol design where an adversary may use quantum algorithms. Loosely speaking, the idea was to “find and characterize quantum-resilience”: bringing together cryptographic experts with an interest in quantum computing and experts on quantum computing with an interest in cryptography, we want to study complexity and hardness assumptions of classical cryptographic schemes from a quantum perspective. We aim at the design of practical cryptographic schemes with tangible evidence for their “post quantum” security that goes beyond the mere non-existence of quantum attacks according to the current state of the art.

The seminar aimed at understanding the exact potential of quantum attacks on today's cryptographic schemes. This question is closely related to the question of plausible quantum computational assumptions. Motivating examples of such assumptions can be found in a cryptographic scheme of Regev from 2009 and a candidate one-way function suggested by Moore, Russell and Vazirani in 2007: the former is a classical public key scheme based on the hardness of the unique shortest vector problem for lattices. It can be argued to be resilient against quantum attacks by relating security guarantees to a hidden subgroup problem in dihedral groups for which, despite much effort by experts on quantum algorithms, no polynomial quantum algorithm has been found. Moore et al.'s proposal rests on an argument from lower bounds on the size of a quantum memory that would be required for the standard quantum approach to graph isomorphism by reducing again to a hidden subgroup problem.

Seminar Organization

A total of 41 scientists from across the world, including both young and senior researchers, visited Dagstuhl for this seminar. To ensure fruitful discussions between experts in quantum computing and in cryptography, the invited participants were chosen such that there is enough common ground/research experience to communicate with colleagues in the other

“camp”. We scheduled the talks with sufficient buffer to have time left for interaction during the talks and for discussions in smaller groups between the talks. Details of the schedule kept changing during the seminar, reflecting the dynamic nature of this meeting. For Wednesday afternoon no talks were scheduled and some participants took advantage of this free afternoon for a hiking trip, some for an excursion to Trier, and others for more discussions.

Topics and Achievements

As anticipated, one of the central topics of the seminar was the hardness of cryptographically relevant computational problems in the presence of quantum attacks: a number of talks addressed classical computational problems and the availability or non-availability of efficient quantum algorithms for these. Moreover, specific cryptographic proposals were discussed which were designed to offer resistance against adversaries with access to quantum computers. Security guarantees of such schemes may rely on some suitable computational hardness assumption, but also on other technological restrictions imposed on the attacker, or solely on the correctness of quantum mechanics. Talks on additional topics, specifically on efficient implementations, foundations of quantum computing and quantum information theory completed the program of the seminar. More details on the individual talks can be found in the abstracts following this introduction.

Looking at the extensive, fruitful, and passionate discussions in the seminar, it is fair to say that this meeting successfully fostered the exchange of two research communities. The presented talks and ensuing discussions added to our understanding of particular cryptographic constructions in the presence of quantum computers. Directions for future work on “quantum-resistant” cryptographic schemes have been indicated, and we hope that follow-up meetings will offer the opportunity to deepen the collaboration between quantum computing and cryptography and therewith help to advance the state-of-the-art in “post quantum” cryptography.

2 Table of Contents

Executive Summary

Serge Fehr, Michele Mosca, Martin Rötteler, and Rainer Steinwandt 58

Overview of Talks


Post-quantum cryptanalysis <i>Daniel J. Bernstein</i>	63
Quantum computing on encrypted data <i>Anne Broadbent</i>	63
Constructing elliptic curve isogenies in quantum subexponential time <i>Andrew Childs</i>	64
A quasipolynomial-time algorithm for the quantum separability problem <i>Matthias Christandl</i>	64
Free randomness amplification <i>Roger Colbeck</i>	65
Security against quantum side information in randomness amplification <i>Serge Fehr</i>	65
Quantum money with classical verification <i>Dmitry Gavinsky</i>	65
Computing the unit group, class group and compact representations in algebraic function fields <i>Sean Hallgren</i>	66
Random quantum circuits are approximate poly-designs <i>Aram W. Harrow</i>	66
Quantum money <i>Avinatan Hassidim</i>	66
Schemes for establishing keys with quantum eavesdroppers <i>Peter Hoyer</i>	67
Quantum fingerprints that keep secrets <i>Tsuyoshi Ito</i>	67
Complexity implications of quantum field theory <i>Stephen P. Jordan</i>	68
Simplified instantaneous non-local quantum computation with applications to position-based cryptography <i>Robert Koenig</i>	68
State-of-the-art branchless techniques for elliptic curve scalar multiplication <i>Tanja Lange</i>	69
Techniques for quantum circuit optimization <i>Dmitri Maslov</i>	69
Decoding random linear codes in $\tilde{O}(2^{0.054n})$ <i>Alexander May</i>	69

The McEliece cryptosystem resists quantum Fourier sampling attacks	
<i>Cris Moore</i>	70
Proof of plaintext knowledge for code-based cryptosystems	
<i>Kirill Morozov</i>	70
What can you hide in qutrit chains?	
<i>Daniel Nagaï</i>	71
Quantum algorithms for the hidden shift problem of Boolean functions	
<i>Maris Ozols</i>	71
Self-testing for sequential CHSH games	
<i>Ben Reichardt</i>	71
Quantum adversary lower bounds by polynomials	
<i>Jeremie Roland</i>	72
Improvements on circuit lattices	
<i>Igor A. Semaev</i>	72
On the hidden shifted power problem	
<i>Igor Shparlinski</i>	73
The garden-hose game and application to position-based cryptography	
<i>Florian Speelman</i>	73
Certifable quantum dice	
<i>Thomas Vidick</i>	73
Participants	75

3 Overview of Talks

3.1 Post-quantum cryptanalysis


Daniel J. Bernstein (University of Illinois at Chicago, US)

License  Creative Commons BY-NC-ND 3.0 Unported license
 © Daniel J. Bernstein
URL <http://cr.yp.to/talks/2011.09.22/slides.pdf>

This talk surveyed the pre-quantum and post-quantum cryptographic landscape, and highlighted some examples of cryptanalytic challenges, including challenges where the best available attack algorithms should be accelerated by quantum computers.

3.2 Quantum computing on encrypted data

Anne Broadbent (University of Waterloo, CA)

License  Creative Commons BY-NC-ND 3.0 Unported license
 © Anne Broadbent

We show that any two-party quantum computation, specified by a unitary which acts simultaneously on the registers of both parties, can be securely implemented against any specious (quantum semi-honest) adversary, with the only additional assumption that the parties have access to an ideal quantum SWAP gate. This establishes that unitaries alone are universal for private two-party evaluation of unitaries, thus answering an open question of Dupuis, Nielsen and Salvail.


We first give a simple protocol for computing the $\pi/8$ gate in a client-server scenario, where the client holds the encryption key for an encrypted qubit held by the server. The client need only prepare a single random auxiliary qubit (chosen among four possibilities), and exchange classical communication. This construction improves on previous work, which requires either multiple auxiliary qubits or two-way quantum communication. We show security against any adversarial server. We then show how to promote this protocol to be secure against both parties, without introducing any extra assumptions. Combined with [1], this shows our main result.

References

- 1 F. Dupuis and J.B. Nielsen and L. Salvail. Secure Two-Party Quantum Evaluation of Unitaries against Specious Adversaries. In *Proceedings of Crypto 2010*, pp. 685–706, LNCS, Vol. 6332, 2010.

3.3 Constructing elliptic curve isogenies in quantum subexponential time

Andrew Childs (University of Waterloo, CA)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Andrew Childs

Joint work of Childs, Andrew; Jao, David; Soukharev, Vladimir


Main reference A. M. Childs, D. Jao, V. Soukharev, “Constructing elliptic curve isogenies in quantum subexponential time,” arXiv:1012.4019

URL <http://arxiv.org/abs/arXiv:1012.4019>

Given two elliptic curves over a finite field having the same cardinality and endomorphism ring, it is known that the curves admit an isogeny between them, but finding such an isogeny is believed to be computationally difficult. The fastest known classical algorithm takes exponential time, and prior to our work no faster quantum algorithm was known. Recently, public-key cryptosystems based on the presumed hardness of this problem have been proposed as candidates for post-quantum cryptography. In this work, we give a new subexponential-time quantum algorithm for constructing isogenies between two such elliptic curves, assuming the Generalized Riemann Hypothesis (but with no other assumptions). Our algorithm is based on a reduction to a hidden shift problem, and represents the first nontrivial application of Kuperberg’s quantum algorithm for the hidden shift problem. This result suggests that isogeny-based cryptosystems may be uncompetitive with more mainstream quantum-resistant cryptosystems such as lattice-based cryptosystems.

3.4 A quasipolynomial-time algorithm for the quantum separability problem

Matthias Christandl (ETH Zürich, CH)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Matthias Christandl

Joint work of Brandao, Fernando; Christandl, Matthias; Yard, Jon

Main reference F. Brandao, M. Christandl, J. Yard, “A quasipolynomial-time algorithm for the quantum separability problem,” Proc. 43rd annual ACM Symposium on Theory of Computing (STOC’11), pp. 343–352.

URL <http://dx.doi.org/10.1145/1993636.1993683>

We present a quasipolynomial-time algorithm for solving the weak membership problem for the convex set of separable, i.e. non-entangled, bipartite density matrices. The algorithm decides whether a density matrix is separable or whether it is ϵ -away from the set of the separable states in time $\exp(O(\epsilon^{-2} \log |A| \log |B|))$, where $|A|$ and $|B|$ are the local dimensions, and the distance is measured with either the Euclidean norm, or with the so-called LOCC norm. The latter is an operationally motivated norm giving the optimal probability of distinguishing two bipartite quantum states, each shared by two parties, using any protocol formed by quantum local operations and classical communication (LOCC) between the parties.

3.5 Free randomness amplification

Roger Colbeck (Perimeter Institute – Waterloo, CA)

License © ⓘ ⊖ Creative Commons BY-NC-ND 3.0 Unported license
 © Roger Colbeck
Joint work of Colbeck, Roger; Renner, Renato
Main reference R. Colbeck, R. Renner, “Free randomness can be amplified,” arXiv:1105.3195v2 [quant-ph]
URL <http://arxiv.org/abs/1105.3195v2>

In many cryptographic tasks, it is assumed that parties have a source of local random bits. I will consider a scenario in which this assumption is weakened, imagining that instead a malicious adversary can influence the source to some extent. In such a scenario, it is known that no classical protocol can use such a source to generate even a single uniform bit. However, I will show that with a quantum protocol this can be done. Furthermore, this protocol remains secure even if quantum theory is one day superseded, provided that the new theory is non-signalling.

3.6 Security against quantum side information in randomness amplification

Serge Fehr (CWI – Amsterdam, NL)

License © ⓘ ⊖ Creative Commons BY-NC-ND 3.0 Unported license
 © Serge Fehr
Joint work of Fehr, Serge; Gelles, Ran; Schaffner, Christian
Main reference S. Fehr, R. Gelles, C. Schaffner, “Security and Composability of Randomness Expansion from Bell Inequalities,” arXiv:1111.6052v2 [quant-ph]
URL <http://arxiv.org/abs/1111.6052v2>

We consider the problem of randomness amplification from Bell inequalities, as initially proposed by Colbeck and worked out by Pironio et al. We show that in this setting, security against quantum side information comes for free.

Specifically, we show that a lower bound on the (worst case) min-entropy that is obtained under the assumption that the adversary holds no (quantum nor classical) side information, also holds in case the adversary holds quantum side information. This in particular implies that the bounds obtained by Pironio et al. extend to the setting with quantum side information.

3.7 Quantum money with classical verification


Dmitry Gavinsky (NEC Laboratories America, Inc. – Princeton, US)

License © ⓘ ⊖ Creative Commons BY-NC-ND 3.0 Unported license
 © Dmitry Gavinsky
Main reference D. Gavinsky, “Quantum Money with Classical Verification,” arXiv:1109.0372v1 [quant-ph]
URL <http://arxiv.org/abs/1109.0372>

We construct a quantum money scheme that allows verification through classical communication with bank. This is the first demonstration that a secure quantum money scheme exists that does not require quantum communication for coin verification.

3.8 Computing the unit group, class group and compact representations in algebraic function fields

Sean Hallgren (Penn State University, US)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Sean Hallgren

Joint work of Eisentraeger, Kirsten; Hallgren, Sean

Number fields and global function fields have many similar properties. Both have many applications to cryptography and coding theory, and the main computational problems for number fields, such as computing the ring of integers and computing the class group and the unit group, have analogues over function fields. The complexity of the number field problems has been studied extensively and these problems have been the source of some exponential speedups by quantum computation. In this paper we study the analogous problems in function fields. We show that there are efficient quantum algorithms for computing the unit group, the class group and for solving the principal ideal problem in function fields of arbitrary degree. We show that compact representations exist, which allows us to show that the principal ideal problem is in NP. Unlike the number field case, we are also able to show that these compact representatives can be computed efficiently.

3.9 Random quantum circuits are approximate poly-designs

Aram W. Harrow (University of Washington, US)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Aram W. Harrow


Joint work of Brandao, Fernando G.S.L.; Harrow, Aram W.; Horodecki, Michal

An ϵ -approximate t -design is a distribution over unitaries such that t copies of a unitary from the distribution cannot be distinguished from t copies of a Haar-uniform unitary with bias greater than ϵ . In other words, designs look like they are uniformly random if we look at low-degree polynomials of their entries.

We prove that random circuits on n qubits of length $O(n^2 t^5 \log(1/\epsilon))$ are approximate t -designs. Previously random circuits were only known to be 3-designs, and efficient constructions of t -designs were only known for $t \leq n/\log(n)$. Our proof uses tools from many-body theory and from representation theory.

3.10 Quantum money

Avinatan Hassidim (MIT – Cambridge, US)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Avinatan Hassidim

Joint work of Aaronson, Scott; Farhi, Eddie; Gosset, David; Kelner, Jon; Lutomirski, Andy; Shor, Peter
URL <http://www2.lns.mit.edu/~avinatan/publications.html>

One of the problems in classical security is that information can be copied: passwords can be stolen, songs can be pirated, and when you email an attachment, you still have the original. One implication is that E-commerce requires communicating with a server (e.g. the credit card company or PayPal) whenever one makes a transaction. One could hope that the no-cloning theorem would help circumvent this and enable a physical quantum state to

function like money. Such money could be used in transactions both in person and on a future “Quantum Internet,” not requiring contact with a central authority.

In the talk I will survey some recent progress on quantum money. I will present an impossibility result for a certain family of schemes, and a scheme which is based on ideas from knot theory.

3.11 Schemes for establishing keys with quantum eavesdroppers

Peter Hoyer (University of Calgary, CA)

License © ⓘ ⊕ Creative Commons BY-NC-ND 3.0 Unported license

© Peter Hoyer

Joint work of Brassard, Gilles; Høyer, Peter; Kalach, Kassem; Kaplan, Marc; Laplante, Sophie; Salvail, Louis

Main reference G. Brassard, P. Høyer, K. Kalach, M. Kaplan, S. Laplante, L. Salvail, “Merkle Puzzles in a Quantum World,” Proc. 31st Annual International Conference on Cryptology (CRYPTO 2011), pp. 391–410, LNCS Vol. 6841, 2011.

URL http://dx.doi.org/10.1007/978-3-642-22792-9_22

In 1974, Ralph Merkle proposed the first unclassified scheme for secure communications over insecure channels. When legitimate communicating parties are willing to spend an amount of computational effort proportional to some parameter N , an eavesdropper cannot break into their communication without spending a time proportional to N^2 , which is quadratically more than the legitimate effort. We showed in an earlier paper that Merkle’s schemes are completely insecure against a quantum adversary, but that their security can be partially restored if the legitimate parties are also allowed to use quantum computation: the eavesdropper needed to spend a time proportional to $N^{3/2}$ to break our earlier quantum scheme. Furthermore, all previous *classical* schemes could be broken completely by the onslaught of a quantum eavesdropper and we conjectured that this is unavoidable.

We give two novel key establishment schemes in the spirit of Merkle’s. The first one can be broken by a quantum adversary that makes an effort proportional to $N^{5/3}$ to implement a quantum random walk in a Johnson graph reminiscent of Andris Ambainis’ quantum algorithm for the element distinctness problem. This attack is optimal up to logarithmic factors. Our second scheme is purely classical, yet it cannot be broken by a quantum eavesdropper who is only willing to expend effort proportional to that of the legitimate parties.

3.12 Quantum fingerprints that keep secrets

Tsuyoshi Ito (University of Waterloo, CA)

License © ⓘ ⊕ Creative Commons BY-NC-ND 3.0 Unported license

© Tsuyoshi Ito

Joint work of Gavinsky, Dmitry; Ito, Tsuyoshi

Main reference D. Gavinsky, T. Ito, “Quantum Fingerprints that Keep Secrets,” arXiv:1010.5342v1 [quant-ph]

URL <http://arXiv.org/abs/1010.5342>


We consider the task of quantum fingerprinting (Buhrman, Cleve, Watrous, and de Wolf 2001) with an additional cryptographic requirement; namely, we require that if a fingerprint state is received by a malicious party, Eve, she cannot use the state to extract much classical information about the message. We show that there exists a fingerprinting scheme which encodes an n -bit classical message into an $O(\log n)$ -qubit fingerprint state such that

1. (Correctness) From two fingerprint states, one can decide whether they are made from the same message or not with error ε constant < 1 .
2. (Hiding property) The accessible information in the fingerprint state about the message is at most a constant independent of n or ε .
3. (Efficient construction) Construction is probabilistic, using $\text{poly}(n)$ bits of randomness, but these random bits are not required to be kept secret.

We also show a variation of this construction for equality testing in one-way communication model, where both the error probability and the accessible information tend to 0 as n tends to infinity.

3.13 Complexity implications of quantum field theory

Stephen P. Jordan (NIST – Gaithersburg, US)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Stephen P. Jordan

Joint work of Jordan, Stephen P.; Lee, Keith S. M.; Preskill, John

The field of post-quantum cryptography is based on the assumption that the set of efficiently decidable problems is BQP. However, BQP is derived from the quantum circuit model, which is based on physics as we understood it in the 1930s. Today, our best physical model is the Standard Model of particle physics, which is a relativistic quantum field theory. Relativistic quantum field theories are defined using a formalism different from that used to define the quantum circuit model, and they exhibit new phenomena, such as particle creation, as well as new problems, such as renormalizability and continuum limits. Thus, it is necessary to carefully examine whether quantum field theories give rise to computational power beyond the standard quantum circuit model. We provide some evidence in the negative, showing that a certain quantum field theory (massive phi-fourth theory) can be efficiently simulated by standard quantum computers.

3.14 Simplified instantaneous non-local quantum computation with applications to position-based cryptography

Robert Koenig (IBM TJ Watson Research Center, US)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Robert Koenig

Joint work of Beigi, Salman; Koenig, Robert

Main reference S. Beigi, R. Koenig, “Simplified instantaneous non-local quantum computation with applications to position-based cryptography,” New Journal of Physics 13 (2011), 093036, arXiv:1101.1065v3 [quant-ph]

URL <http://arxiv.org/abs/1101.1065v3>

Instantaneous measurements of non-local observables between space-like separated regions can be performed without violating causality. This feat relies on the use of entanglement. Here we propose novel protocols for this task and the related problem of multipartite quantum computation with local operations and a single round of classical communication. Compared to previously known techniques, our protocols reduce the entanglement consumption by an exponential amount. We also prove a linear lower bound on the amount of entanglement required for the implementation of a certain non-local measurement. These results relate

to position-based cryptography: an amount of entanglement scaling exponentially in the number of communicated qubits is sufficient to render any such scheme insecure.

Furthermore, we show that certain schemes are secure under the assumption that the adversary has less entanglement than a given linear bound and is restricted to classical communication.

3.15 State-of-the-art branchless techniques for elliptic curve scalar multiplication

Tanja Lange (Technische Universiteit Eindhoven, NL)

License  Creative Commons BY-NC-ND 3.0 Unported license


© Tanja Lange

URL <http://hyperelliptic.org/tanja/vortraege/11/blackboard/images.html>

Researchers on elliptic-curve cryptography have investigated uniform branchless computations to avoid side-channel attacks and to make better use of SIMD instructions on modern CPUs. This talk surveys some of these techniques including our high-speed high-security elliptic curve signatures scheme Ed25519. The motivation to give this talk for this audience is that the same ideas should help to speed up breaking the ECDLP on a quantum computer

3.16 Techniques for quantum circuit optimization

Dmitri Maslov (NSF – Arlington, US)

License  Creative Commons BY-NC-ND 3.0 Unported license

© Dmitri Maslov

Joint work of Golubitsky, Oleg; Maslov, Dmitri

Main reference O. Golubitsky, D. Maslov, “A Study of Optimal 4-bit Reversible Toffoli Circuits and Their Synthesis,” IEEE Transactions on Computers, in print. arXiv:1103.2686

URL <http://arxiv.org/abs/1103.2686>

I will briefly discuss some known circuit optimization techniques and then concentrate on the in-depth analysis of the peep-hole optimization technique. In particular, I will discuss the efficient ways of synthesizing and accessing minimized/optimal implementations of small functions, illustrated with the optimal synthesis of any given and all 4-bit reversible functions. This presentation is based on the results reported in arXiv:1103.2686.

3.17 Decoding random linear codes in $\tilde{O}(2^{0.054n})$

Alexander May (Ruhr-Universität Bochum, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license

© Alexander May

Joint work of May, Alexander; Meurer, Alexander; Thomae, Enrico

Main reference Alexander May, Alexander Meurer, Enrico Thomae, “Decoding Random Linear Codes in $\tilde{O}(2^{0.054n})$,” In Advances in Cryptology (Asiacrypt 2011), Lecture Notes in Computer Science, Springer-Verlag, 2011.

URL http://dx.doi.org/10.1007/978-3-642-25385-0_6

Decoding random linear codes is a fundamental problem in complexity theory and lies at the heart of almost all code-based cryptography. The best attacks on the most prominent


code-based cryptosystems such as McEliece directly use decoding algorithms for linear codes. The asymptotically best decoding algorithm for random linear codes of length n was for a long time Stern's variant of information-set decoding running in time $\tilde{O}(2^{0.05563n})$.

Recently, Bernstein, Lange and Peters proposed a new technique called *Ball-collision decoding* which offers a speed-up over Stern's algorithm by improving the running time to $\tilde{O}(2^{0.05558n})$.

In this work, we present a new algorithm for decoding linear codes that is inspired by a representation technique due to Howgrave-Graham and Joux in the context of subset sum algorithms. Our decoding algorithm offers a rigorous complexity analysis for random linear codes and brings the time complexity down to $\tilde{O}(2^{0.05363n})$.

3.18 The McEliece cryptosystem resists quantum Fourier sampling attacks

Cris Moore (University of New Mexico – Albuquerque, US)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Cris Moore


Joint work of Dinh, Hang; Russell, Alex

Since Shor's algorithm breaks RSA cryptography, it makes sense to look for post-quantum cryptosystems: cryptosystems that can be carried out with classical computers today, but which will remain secure even if and when quantum computers are built.

In this talk I will give an introduction to the McEliece and Niederreiter public-key cryptosystems, which are based on error-correcting codes, and argue that they are possible candidates for post-quantum cryptography. Specifically, I will show that they are immune to quantum algorithms based on the natural reduction to the Hidden Subgroup Problem, where we construct a coset state and perform strong Fourier sampling on it. This does not rule out other quantum (or classical) attacks on these systems, but it suggests that additional algorithmic ideas would be needed to break them.

3.19 Proof of plaintext knowledge for code-based cryptosystems

Kirill Morozov (Kyushu University, JP)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Kirill Morozov

We present a zero-knowledge (ZK) proof of plaintext knowledge for the code-based McEliece and Niederreiter public-key encryption schemes. It applies to both their original and randomized (IND-CPA) versions. Our proof uses Stern's ZK identification scheme.

3.20 What can you hide in qutrit chains?

Daniel Nagaj (Slovak Academy of Sciences – Bratislava, SK)

License © © © Creative Commons BY-NC-ND 3.0 Unported license
© Daniel Nagaj

Joint work of Caha, Libor; Bravyi, Sergey; Nagaj, Daniel

Instead of looking at QMA-completeness of problems involving chains of high dimensional qudits, we ask whether we could encode interesting problems in a chain of low-dimensional particles. We simplify the question, focusing on translationally invariant systems of qutrits in 1D with nearest neighbor interactions (projector terms), asking whether they are unfrustrated, i.e. whether there exists a state satisfying all of the local conditions (annihilated by all of the local terms). We present an interesting system with a unique ground state which is rather entangled, but still has a polynomial gap in the energy spectrum. It differs significantly from the usual history-state construction used in QMA-completeness results, and is related to a language of properly bracketed expressions.

3.21 Quantum algorithms for the hidden shift problem of Boolean functions

Maris Ozols (University of Waterloo, CA)

License © © © Creative Commons BY-NC-ND 3.0 Unported license
© Maris Ozols

Joint work of Ozols, Maris; Roetteler, Martin; Roland, Jeremie

Main reference M. Ozols, M. Roetteler, J. Roland, “Quantum rejection sampling,” arXiv:1103.2774v3 [quant-ph]

URL <http://arxiv.org/abs/1103.2774v3>

We discuss several quantum algorithms for attacking the following problem: given oracle access to $f(x+s)$ where $f(x)$ is a known Boolean function, determine the hidden shift s . One of our approaches is a new quantum state generation technique—quantum rejection sampling. We use semidefinite programming to express the query complexity of our algorithm in terms of “water-filling” properties of the Fourier spectrum of f .

3.22 Self-testing for sequential CHSH games

Ben Reichardt (University of Waterloo, CA)


License © © © Creative Commons BY-NC-ND 3.0 Unported license
© Ben Reichardt

Joint work of Reichardt, Ben; Unger, Falk; Vazirani, Umesh

By collecting statistics on sequential CHSH games, we can gain confidence that the provers are playing according to an ideal strategy.

3.23 Quantum adversary lower bounds by polynomials

Jeremie Roland (Université Libre de Bruxelles, BE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Jeremie Roland


Joint work of Magnin, Loïck; Roland, Jeremie

The polynomial method and the adversary method are the two main techniques to prove lower bounds on quantum query complexity, and they have so far been considered as unrelated. Here, we show an explicit reduction from the polynomial method to the multiplicative adversary method. The proof goes by introducing a new type of adversary method, which generalizes the polynomial method. We then show that this adversary bound can be obtained from the multiplicative adversary bound by taking the limit $c \rightarrow \infty$, where $c > 1$ is the maximum factor by which the adversary progress function can increase after each query.

Interestingly, it is also known that the additive adversary method can be obtained from the multiplicative method by taking the limit $c \rightarrow 1$, and this new result therefore provides a clear picture of the relation between the different lower bound methods. It also gives new hope to prove lower bounds on variations of problems such as *collision* and *ED*, for which the only known lower bounds are proved by the less flexible polynomial method.

3.24 Improvements on circuit lattices

Igor A. Semaev (University of Bergen, NO)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Igor A. Semaev

Circuit Lattice (CL) is a hardware tool for solving sparse Boolean equations. That is where the number of variables in each particular equation is bounded, though the overall number of variables is large. One introduces a guess on some variables and the device signals out whether it is wrong. CL may be constructed as a combination of wires and switches(transistors) on a semiconductor crystal and used for key search by brut force in cryptanalysis. In contrast to a conventional computer, the transistors are not necessarily synchronized.

By gluing some of the cipher initial equations one may produce, at the expense of enlarging the number of local solutions to particular equations, a system with a reduced key search space. Now it is unlikely to implement CL for that on one semiconductor crystal. So whether the parallelism of quantum computing may replace electric potential expansion in CL is an interesting question.

In this talk I plan to explain some recent improvements to Circuit Lattices as reducing the number of required transistors and its new architecture more suitable for implementing by modern computer industry.

3.25 On the hidden shifted power problem

Igor Shparlinski (Macquarie University – Sydney, AU)

License © © ⊖ Creative Commons BY-NC-ND 3.0 Unported license
© Igor Shparlinski

Joint work of Bourgain, Jean; Konyagin, Sergei; Shparlinski, Igor

We consider the problem of recovering a hidden element s of a finite field \mathbb{F}_q of q elements from queries to an oracle that for a given $x \in \mathbb{F}_q$ returns $(x + s)^e$ for a given divisor $e \mid q - 1$. This question is motivated by some applications to pairing based cryptography.

Using Lagrange interpolation one can recover s in time $ep^{o(1)}$ on a classical computer. In the case of $e = (q - 1)/2$ an efficient quantum algorithm has been given by W. van Dam, S. Hallgren and L. Ip.

We describe some techniques from additive combinatorics and analytic number theory that lead to more efficient classical algorithms than the naive interpolation algorithm, for example, they use substantially fewer queries to the oracle. We formulate some questions and discuss whether quantum algorithms can give further improvement.

3.26 The garden-hose game and application to position-based cryptography

Florian Speelman (CWI – Amsterdam, NL)

License © © ⊖ Creative Commons BY-NC-ND 3.0 Unported license
© Florian Speelman

Joint work of Buhrman, Harry; Fehr, Serge; Schaffner, Christian; Speelman, Florian

Main reference H. Buhrman, S. Fehr, C. Schaffner, F. Speelman, “The Garden-Hose Game: A New Model of Computation, and Application to Position-Based Quantum Cryptography,” arXiv:1109.2563v2 [quant-ph]

URL <http://arxiv.org/abs/1109.2563v2>

We study position-based cryptography in the quantum setting. We examine a class of protocols that only require the communication of a single qubit and $2n$ bits of classical information. To this end, we define a new model of communication complexity, the garden-hose model, which enables us to prove upper bounds on the number of EPR pairs needed to attack such schemes.

This model furthermore opens up a way to link the security of quantum position-based cryptography to traditional complexity theory.

3.27 Certifiable quantum dice

Thomas Vidick (University of California – Berkeley, US)

License © © ⊖ Creative Commons BY-NC-ND 3.0 Unported license
© Thomas Vidick

Joint work of Vazirani, Umesh; Vidick, Thomas

We introduce a protocol through which a pair of quantum mechanical devices may be used to generate n bits of true randomness from a seed of $O(\log n)$ uniform bits. The bits generated are certifiably random based only on a simple statistical test that can be performed by the user, and on the assumption that the devices obey the no-signaling principle. No other assumptions are placed on the devices’ inner workings. A modified protocol uses a seed of

$O(\log^3 n)$ uniformly random bits to generate n bits of true randomness even conditioned on the state of a quantum adversary who may have had prior access to the devices, and may be entangled with them.

Participants

- Daniel J. Bernstein
University of Chicago, US
- Anne Broadbent
University of Waterloo, CA
- Harry Buhrman
CWI – Amsterdam, NL
- Andrew Childs
University of Waterloo, CA
- Matthias Christandl
ETH Zürich, CH
- Roger Colbeck
Perimeter Inst. – Waterloo, CA
- Serge Fehr
CWI – Amsterdam, NL
- Dmitry Gavinsky
NEC Laboratories America, Inc.
– Princeton, US
- Sean Hallgren
Penn State University, US
- Aram W. Harrow
University of Washington, US
- Avinathan Hassidim
MIT – Cambridge, US
- Peter Hoyer
University of Calgary, CA
- Tsuyoshi Ito
University of Waterloo, CA
- Stacey Jeffery
University of Waterloo, CA
- Stephen P. Jordan
NIST – Gaithersburg, US
- Robert Koenig
IBM TJ Watson Res. Center, US
- Tanja Lange
TU Eindhoven, NL
- Frédéric Magniez
University Paris-Diderot, FR
- Loïck Magnin
University Paris-Diderot, FR
- Dmitri Maslov
NSF – Arlington, US
- Alexander May
Ruhr-Universität Bochum, DE
- Cris Moore
University of New Mexico –
Albuquerque, US
- Kirill Morozov
Kyushu University, JP
- Michele Mosca
University of Waterloo, CA
- Daniel Nagaj
Slovak Academy of Sciences –
Bratislava, SK
- Maris Ozols
University of Waterloo, CA
- Anupam Prakash
University of California –
Berkeley, US
- Ben Reichardt
University of Waterloo, CA
- Martin Rötteler
NEC Laboratories America, Inc.
– Princeton, US
- Jérémie Roland
Université Libre de Bruxelles, BE
- Alexander Russell
University of Connecticut –
Storrs, US
- Leonard J. Schulman
CalTech – Pasadena, US
- Igor A. Semaev
University of Bergen, NO
- Igor Shparlinski
Macquarie Univ.- Sydney, AU
- Rolando Somma
Los Alamos National Lab., US
- Florian Speelman
CWI – Amsterdam, NL
- Rainer Steinwandt
Florida Atlantic University –
Boca Raton, US
- Barbara Terhal
RWTH Aachen, DE
- Wim van Dam
University of California – Santa
Barbara, US
- Thomas Vidick
University of California –
Berkeley, US
- Arne Winterhof
RICAM – Linz, AT



Public-Key Cryptography

Edited by

Marc Fischlin¹, Anna Lysyanskaya², Ueli Maurer³, and
Alexander May⁴

¹ TU Darmstadt, DE, marc.fischlin@gmail.com

² Brown University – Providence, US

³ ETH Zürich, CH, maurer@inf.ethz.ch

⁴ Ruhr-Universität Bochum, DE, alex.may@ruhr-uni-bochum.de

Abstract

From September 25th till September 30th, 2011, the Dagstuhl Seminar 11391 about “Public-Key Cryptography” took place at Schloss Dagstuhl. The meeting hosted 33 international researchers and incited active discussions about recent developments in this area.

Seminar 25.–30. September, 2011 – www.dagstuhl.de/11391

1998 ACM Subject Classification D.4.6 Security and Protection

Keywords and phrases Fully-Homomorphic Encryption, Leakage-Resilience, Constructive Cryptography

Digital Object Identifier 10.4230/DagRep.1.9.76

Edited in cooperation with Alexander Meurer

1 Executive Summary

Marc Fischlin

Anna Lysyanskaya

Ueli Maurer

Alexander May

Cryptography is the science of protecting data in presence of malicious parties. Without cryptography e-commerce, e-banking and e-government would not be possible. Indeed, the most prominent application of cryptography today is the SSL/TLS protocol to secure e-mail and web communication. But soon citizens will also use cryptography on large scales on identity cards, passports and health cards.

Cryptography is a relatively new area in computer science, with the first modern and scientific approaches dating back to the mid 70’s, and the first large-scale scientific conferences in this area in the early 80’s. Since then, cryptography has evolved as its own sub area in computer science, with intersections with many areas like number-theory or complexity theory.

Cryptography has a good tradition within the Dagstuhl Seminar series, with the first meeting about cryptography held in 1993, and subsequent seminars on this topic about every 5 years. In 2007 and 2012 a seminar for the sub area of “Symmetric Cryptography” is added, inciting us to coin the seminar here “Public-Key Cryptography” for sake of distinction.

The seminar brought together 33 of the leading scientists in the area of public-key cryptography. The participants came from all over the world, including countries like the US, Great Britain, Israel, France, or Italy. Among the affiliations Germany lead the number with 10 participants, followed by the US with 7, and Switzerland with 6.



Except where otherwise noted, content of this report is licensed under a Creative Commons BY-NC-ND 3.0 Unported license

Public-Key Cryptography, *Dagstuhl Reports*, Vol. 1, Issue 9, pp. 76–94

Editors: Marc Fischlin, Anna Lysyanskaya, Ueli Maurer, and Alexander May



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The program contained 28 talks, each of 25-30 minutes, and a panel discussion about the field's future, with a free afternoon on Wednesday for social activities and half a day on Friday for traveling. Before the seminar we asked the participants to present very recent and ongoing work which, ideally, should not have been published or accepted to publication yet. Most of the participants followed our suggestion and to a large extent the presentations covered topics which have not even been submitted at the time.

The topics of the talk represented the diversity of public-key cryptography. As expected and envisioned, there was quite a number of talks about encryption schemes (such as homomorphic encryption) and their use for the cloud scenario. To further this area has been stated as one of the goals of the seminar. Presentations about this topic included improvements for such encryption schemes, e.g., the even more general functional encryption was covered comprehensively, as well as their applicability. Another well-represented area of the seminar touched the intended question of looking into more leakage-resilient alternatives like learning with errors (LWE) or lattice-based constructions. Discussions during and after the talks were lively.

The goal of the seminar was to incite new research in the area of public-key cryptography, with the explicit goal to enhance the areas of computing on encrypted data, leakage-resilience, and hash functions. We —and seemingly also the participants— enjoyed the possibility to further discuss fresh topics like constructive cryptography. Overall, the personal feedback of the participants to us was very positive, with the wish to repeat such a seminar.

The organizers would like to thank Alexander Meurer for collecting all abstracts of this seminar report. Finally, the organizers, also on behalf of the participants, would like to thank the staff and the management of Schloss Dagstuhl for providing the surrounding for a very pleasant and fruitful seminar.

2 Table of Contents

Executive Summary

<i>Marc Fischlin, Anna Lysyanskaya, Ueli Maurer, and Alexander May</i>	76
--	----

Overview of Talks


Generic Algorithms for Hard Subset Sums and its Application to Linear Codes <i>Anja Becker</i>	80
Functional Re-encryption and Collusion-Resistant Obfuscation <i>Melissa Chase</i>	80
Active Security in General Secure Multi-Party Computation via Black-Box Groups <i>Yvo Desmedt</i>	81
Leftover Hash Lemma, Revisited <i>Yevgeniy Dodis</i>	81
Functional Encryption: Extensions and Implications <i>Pooya Farshim</i>	82
Resource-based Corruptions and the Combinatorics of Anonymity <i>Juan A. Garay</i>	82
All-But-Many Lossy Trapdoor Functions <i>Dennis Hofheinz</i>	83
Constructing a Pseudorandom Generator Requires an Almost Linear Number of Calls <i>Thomas Holenstein</i>	84
How to Garble Arithmetic Circuits <i>Yuval Ishai</i>	84
Cover and Decomposition on Elliptic Curves <i>Antoine Joux</i>	85
LWE is Lossy <i>Eike Kiltz</i>	85
Constructive Cryptography – A New Paradigm for Security Definitions and Proofs <i>Ueli Maurer</i>	85
A New Information Set Decoding Algorithm <i>Alexander Meurer</i>	86
Code Obfuscation with a Stateless Hardware Token <i>Joern Mueller-Quade</i>	86
Oblivious Transfer with Anonymous Access Control and Pricing <i>Gregory Neven</i>	87
On the Joint Security of Encryption and Signature, Revisited <i>Kenny Paterson</i>	87
Hidden Vector Encryption Fully Secure Against Unrestricted Queries – No Query Left Unanswered <i>Giuseppe Persiano</i>	88

Commitments and Efficient Zero- Knowledge from Hard Learning Problems <i>Krzysztof Pietrzak</i>	88
Careful with Composition: Limitations of Indifferentiability and Universal Composability <i>Thomas Ristenpart</i>	89
Pseudo Random Functions and Lattices <i>Alon Rosen</i>	89
Identification and Signatures Based on NP-Hard Problems of Indefinite Quadratic Forms <i>Claus Peter Schnorr</i>	90
Security of Blind Signatures Revisited <i>Dominique Schroeder</i>	90
Integrity Notions for Encryption Schemes <i>Bjoern Tackmann</i>	91
The equivalence of the random oracle model and the ideal cipher model, revisited. <i>Stefano Tessaro</i>	91
Secure Computation with Corruptible Setups <i>Vassilis Zikas</i>	92
Working Groups	92
Open Problems	93
Panel Discussions	93
Scientific Output	93
Participants	94

3 Overview of Talks

3.1 Generic Algorithms for Hard Subset Sums and its Application to Linear Codes

Anja Becker (*University of Versailles, FR*)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Anja Becker

Joint work of Becker, Anja; Coron, Jean-Sébastien; Joux, Antoine;

Main reference A. Becker, J.S. Coron, A. Joux, “Improved Generic Algorithms for Hard Knapsacks,” in *Advances in Cryptology – EUROCRYPT 2011*, pp. 364–385, LNCS Vol. 6632.

URL http://dx.doi.org/10.1007/978-3-642-20465-4_21


At Eurocrypt 2010, Howgrave-Graham and Joux described an algorithm for solving hard knapsacks of density close to 1 in time $\mathcal{O}(2^{0.337n})$ and memory $\mathcal{O}(2^{0.256n})$, thereby improving a 30-year old algorithm by Shamir and Schroepel. In this talk we will present the following:

Using the simple observation that a binary vector can be represented by two overlapping vectors with coefficients in $\{-1, 0, 1\}$, we can obtain a better algorithm of running time $\mathcal{O}(2^{0.291n})$.

Furthermore, this technique can be directly applied to improve information set decoding attacking linear codes such as used in McEliece public key encryption.

3.2 Functional Re-encryption and Collusion-Resistant Obfuscation

Melissa Chase (*Microsoft Research – Redmond, US*)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Melissa Chase

Joint work of Chandran, Nishanth; Chase, Melissa; Vaikuntanathan, Vinod;

Main reference N. Chandran, M. Chase, V. Vaikuntanathan, “Collusion Resistant Obfuscation and Functional Re-encryption,” *IACR eprint archive*, 2011/337, 2011.

URL <http://eprint.iacr.org/2011/337>

We introduce a natural cryptographic functionality called functional re-encryption.

Informally, functional re-encryption allows an untrusted server to transform a ciphertext encrypted for Alice into a ciphertext encrypted for one of n recipients, depending on the message. In particular, a functional re-encryption function for a function F will transform an encryption of message m intended for Alice into an encryption of m intended for recipient $F(m)$.


In many settings, one might require that the program implementing the functional re-encryption functionality should reveal as little as possible about the input secret key SK and the function F . Furthermore, ideally we would obtain an even stronger guarantee: that this information remains hidden even when some of the n recipients may be corrupted.

To formalize these issues, we introduce the notion of collusion-resistant obfuscation and define this notion with respect to average-case secure obfuscation (Hohenberger et al. – TCC 2007). We show that this notion of functional re- encryption can be achieved for any function F with polynomial-size domain, by providing a direct construction from bilinear pairings.

This is joint work with Nishanth Chandran and Vinod Vaikuntanathan.

3.3 Active Security in General Secure Multi-Party Computation via Black-Box Groups

Yvo Desmedt (*University College London, GB*)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Yvo Desmedt


At CRYPTO 2007, Desmedt et al introduced a construction for a multi-party multiplication protocol for black-box non-Abelian groups. The security achieved was against a passive adversary controlling t parties using the unconditionally secure model. This left as an open problem how to achieve security against active attacks. The construction was based on a reduction to a new planar graph coloring problem.

In this presentation, we extended above to the case of general adversary structure. Our focus was on, for such an adversary, to achieve security against an active adversary. Our solution is the first n -party protocol that achieves multiparty computation using as building block a black-box non-Abelian group solution, which is secure against an active attacker, tolerating any adversary structure satisfying the property that no union of three subsets from the adversary structure covers the whole player's set.

Our protocol uses Maurer's Verifiable Secret Sharing (VSS) but preserves roughly the essential simplicity of the graph-based approach of Desmedt et al. This implies that each shareholder can avoid having to rerun the full VSS protocol after each local computation. The reduction of the need to use VSS may have consequences to secure multiparty computation beyond the fact we use non-Abelian groups.

3.4 Leftover Hash Lemma, Revisited

Yevgeniy Dodis (*New York University, US*)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Yevgeniy Dodis
Joint work of Barak, Boaz; Dodis, Yevgeniy; Krawczyk, Hugo; Pereira, Olivier; Pietrzak, Krzysztof; Standaert, Francois-Xavier; Yu, Yu
Main reference B. Barak, Y. Dodis, H. Krawczyk, O. Pereira, K. Pietrzak, F.-X. Standaert, Y. Yu, "Leftover Hash Lemma, Revisited", Published at CRYPTO'2011.
URL <http://eprint.iacr.org/2011/088>

Randomness extractors are procedures used to extract nearly perfect randomness from any "imperfect" source of randomness, such as physical sources, biometrics, etc. Such extractor have found numerous applications in many areas of computer science.

A very simple and elegant randomness extractor is given by the famous Leftover Hash Lemma (LHL), which states that a random universal hash function is a good extractor.

Unfortunately, despite their numerous applications, LHL-based extractors suffer from the following two drawbacks. First, the maximum number of extracted bits is at least $2 \cdot \log(1/e)$ less than the amount of entropy in the source, where e is the desired statistical distance from uniform, which could be large for low-entropy sources.

Second, the description length of a random universal hash function (called the "seed") is linear in the length of the imperfect source, which could be very large.

Quite surprisingly, we show that both limitations of the LHL — large entropy loss and large seed — can often be overcome (or, at least, mitigated) in various quite general scenarios. First, we show that "entropy loss" could be halved from $2 \log(1/e)$ to $\log(1/e)$ for the setting of deriving secret keys for most cryptographic applications. Second, we study the soundness


of the natural *expand-then-extract* approach, where one uses a pseudorandom generator (PRG) to compress the description length of the extractor seed. We show that, although the expand-then-extract-approach is *not* sound in general, any counter-example implies an efficient construction of public-key encryption from a PRG.

This suggests that the sample-then-extract approach is likely secure when used with 'practical' PRGs, despite lacking a reductionist proof of security!

The paper can be found at <http://eprint.iacr.org/2011/088>, and is also mentioned in the New Yorker magazine (October 2011).

3.5 Functional Encryption: Extensions and Implications

Pooya Farshim (TU Darmstadt, DE)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Pooya Farshim

We propose extensions of functional encryption (FE) to more complex application scenarios and, in doing so, shed additional light on the properties of simulation-based and indistinguishability-based security notions for FE schemes.

We first study chosen-ciphertext security in the functional setting. We find (somewhat surprisingly) that a CCA-secure FE scheme can be generically built from a CPA-secure one. This transformation, in contrast to the usual CHK transformation (which we also show to hold in the functional setting) does not rely on a delegation mechanism. We then propose an extension of functional encryption to probabilistic functionalities. We call this primitive probabilistic-functional encryption (PFE), and discuss the correct security notions for it. We present constructions of PFE schemes, both generic and concrete, from standard FE schemes. Finally we consider two homomorphic extensions for functional encryption: one notion allows arbitrary computations over a restricted portion of encrypted data, while the other allows restricted computation over the entire encrypted data (i.e., allows functional re-encryption). We identify a class of functionalities for which a meaningful definition of security for the first primitive can be formulated, and show that a large and practically relevant subclass can be securely realized via the KEM/DEM paradigm. We also describe two generic constructions of the second primitive. As a corollary, we show that a homomorphic encryption scheme supporting a large class of circuits can be generically built from a semantically secure FE scheme.

3.6 Resource-based Corruptions and the Combinatorics of Anonymity

Juan A. Garay (AT&T Research – Florham Park, US)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Juan A. Garay

Main reference Submitted for publication.

In the setting of cryptographic protocols, the corruption of a party has been viewed as a simple, uniform and atomic operation, where the adversary decides to get control over a party and this party immediately gets corrupted. In this paper, motivated by the fact that different players may require different resources to get corrupted, we put forth the notion of *resource-based corruptions*, where the adversary must invest some resources in order to do so.

If the adversary has full information about the system configuration then resource-based corruptions would provide no fundamental difference from the standard corruption model. However, in a *resource anonymous* setting, in the sense that such configuration is hidden from the adversary, much is to be gained in terms of efficiency and security.

We showcase the power of anonymity in the setting of secure multiparty computation (MPC) with resource-based corruptions and prove that anonymity can effectively be used to circumvent known impossibility results. Specifically, if OPT is the corruption budget that violates the completeness of MPC (the case when half or more of the players are corrupted), we show that by using anonymity, the completeness of MPC can be made to hold against an adversary with as much as a $B \cdot OPT$ budget, for any constant $B > 1$. This result requires a suitable choice of parameters (in terms of number of players and their hardness to corrupt), which we provide and further prove other tight variants of the result when the said choice is not available. Regarding efficiency gains, we show that anonymity can be used to force the corruption threshold to drop from $1/2$ to $1/3$, in turn allowing the use of much more efficient (information-theoretic) MPC protocols.

We achieve the above through a series of technical contributions:

- The formulation of the notion of *inversion effort preserving* (IEP) functions which is a type of direct-sum property, and the property of *hardness indistinguishability*. While hardness indistinguishability enables the dissociation of parties' identities and the resources needed to corrupt them, IEP enables the discretization of adversarial work into corruption tokens;
- the modeling of the corruption process in the setting of MPC through *corruption oracles* as well as the introduction of a notion of reduction to relate such oracles;
- the abstraction of the corruption game as a combinatorial problem and its analysis,

all of which may be of independent interest.

3.7 All-But-Many Lossy Trapdoor Functions

Dennis Hofheinz (KIT – Karlsruhe Institute of Technology, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Dennis Hofheinz

Main reference D. Hofheinz, “All-But-Many Lossy Trapdoor Functions,”
IACR Cryptology ePrint Archive, 2011/230, 2011.

URL <http://eprint.iacr.org/2011/230>

We put forward a generalization of lossy trapdoor functions (LTFs). Namely, all-but-many lossy trapdoor functions (ABM-LTFs) are LTFs that are parametrized with tags. Each tag can either be injective or lossy, which leads to an invertible or a lossy function. The interesting property of ABM-LTFs is that it is possible to generate an arbitrary number of lossy tags by means of a special trapdoor, while it is not feasible to produce lossy tags without this trapdoor.


Our definition and construction can be seen as generalizations of all-but-one LTFs (due to Peikert and Waters) and all-but-N LTFs (due to Hemenway et al.).

However, to achieve ABM-LTFs (and thus a number of lossy tags which is not bounded by any polynomial), we have to employ some new tricks. Concretely, we give two constructions that employ “disguised” variants of the Waters, resp. Boneh-Boyen signature schemes to make the generation of lossy tags hard without trapdoor. In a nutshell, lossy tags simply correspond to valid signatures. At the same time, tags are disguised (i.e., suitably blinded) to keep lossy tags indistinguishable from injective tags.

ABM-LTFs are useful in settings in which there are a polynomial number of adversarial challenges (e.g., challenge ciphertexts). Specifically, building on work by Hemenway et al., we show that ABM-LTFs can be used to achieve selective opening security against chosen-ciphertext attacks. One of our ABM-LTF constructions thus yields the first SO-CCA secure encryption scheme with compact ciphertexts ($O(1)$ group elements) whose efficiency does not depend on the number of challenges. Our second ABM-LTF construction yields an IND-CCA (and in fact SO-CCA) secure encryption scheme whose security reduction is independent of the number of challenges and decryption queries.

3.8 Constructing a Pseudorandom Generator Requires an Almost Linear Number of Calls

Thomas Holenstein (ETH Zürich, CH)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Thomas Holenstein

Joint work of Holenstein, Thomas; Sinha, Makrand

We show that a black-box construction of a pseudorandom generator from a one-way function needs to make $\Omega(\frac{n}{\log(n)})$ calls to the underlying one-way function. The bound even holds if the one-way function is guaranteed to be regular. In this case it matches the best known construction due to Goldreich, Krawczyk, and Luby (SIAM J. Comp. 22, 1993), which uses $O(\frac{n}{\log(n)})$ calls.

3.9 How to Garble Arithmetic Circuits

Yuval Ishai (Technion – Haifa, IL)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Yuval Ishai

Joint work of Applebaum, Benny; Ishai, Yuval; Kushilevitz, Eyal;


Main reference B. Applebaum, Y. Ishai, E. Kushilevitz, “How to Garble Arithmetic Circuits,” Proceedings of FOCS 2011.

Yao’s garbled circuit construction transforms a boolean circuit $C : \{0, 1\}^n \rightarrow \{0, 1\}^m$ into a “garbled circuit” \hat{C} along with n pairs of k -bit keys, one for each input bit, such that \hat{C} together with the n keys corresponding to an input x reveal $C(x)$ and no additional information about x . The garbled circuit construction is a central tool for constant-round secure computation and has several other applications.

Motivated by these applications, we suggest an efficient arithmetic variant of Yao’s original construction. Our construction transforms an arithmetic circuit $C : \mathbb{Z}^n \rightarrow \mathbb{Z}^m$ over integers from a bounded (but possibly exponential) range into a garbled circuit \hat{C} along with n affine functions $L_i : \mathbb{Z} \rightarrow \mathbb{Z}^k$ such that \hat{C} together with the n integer vectors $L_i(x_i)$ reveal $C(x)$ and no additional information about x . The security of our construction relies on the intractability of the learning with errors (LWE) problem.

3.10 Cover and Decomposition on Elliptic Curves


Antoine Joux (University of Versailles, FR)

License  Creative Commons BY-NC-ND 3.0 Unported license
 © Antoine Joux
Joint work of Joux, Antoine; Vitse, Vanessa
Main reference A. Joux, V. Vitse, “Cover and Decomposition Index Calculus on Elliptic Curves made practical. Application to a seemingly secure curve over F_{p^6} ,” IACR Cryptology ePrint Archive, 2011/020, 2011.
URL <http://eprint.iacr.org/2011/020>

We present a new “cover and decomposition” attack on the elliptic curve discrete logarithm problem, that combines Weil descent and decomposition-based index calculus into a single discrete logarithm algorithm. This attack applies, at least theoretically, to all composite degree extension fields, and is particularly well-suited for curves defined over \mathbb{F}_{p^6} . We give a real-size example of discrete logarithm computations on a curve over a 151-bit degree 6 extension field, which would not have been practically attackable using previously known algorithms.

3.11 LWE is Lossy


Eike Kiltz (Ruhr-Universität Bochum, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
 © Eike Kiltz
Joint work of Bellare, Mihir; Kiltz, Eike; Peikert, Chris; Waters, Brent
Main reference M. Bellare, E. Kiltz, C. Peikert, B. Waters, “Identity-Based (Lossy) Trapdoor Functions and Applications,” IACR Cryptology ePrint Archive, 2011/479, 2011.
URL <http://eprint.iacr.org/2011/479>

We show that, under an appropriate choice of parameters, the Learning With Errors (LWE) function is a lossy trapdoor function.

3.12 Constructive Cryptography – A New Paradigm for Security Definitions and Proofs

Ueli Maurer (ETH Zürich, CH)

License  Creative Commons BY-NC-ND 3.0 Unported license
 © Ueli Maurer
Joint work of Maurer, Ueli; Renner, Renato
Main reference U. Maurer, R. Renner, “Abstract Cryptography,” in Proceedings of Innovations in Computer Science – ICS 2010.


Constructive cryptography, an application of abstract cryptography proposed by Maurer and Renner, is a new paradigm for defining the security of cryptographic schemes such as symmetric encryption, message authentication codes, public-key encryption, key-agreement protocols, and digital signature schemes, and for proving the security of protocols making use of such schemes. Such a cryptographic scheme can be seen (and defined) as constructing a certain resource (e.g. a channel or key) with certain security properties from another (weaker) such resource. For example, a secure encryption scheme constructs a secure channel from an authenticated channel and a secret key.

The term “construct”, which is defined by the use of a simulator, is composable in the sense that a protocol obtained by the composition of several secure constructive steps is itself secure. This is in contrast to both the traditional, game-based security definitions for cryptographic schemes and the attack-based security definitions used in formal-methods based security research, which are generally not composable.

Constructive cryptography allows to take a new look at cryptography and the design of cryptographic protocols. One can give explicit meaning to various types of game-based security notions of confidentiality, integrity, and malleability, one can design key agreement, secure communication, certification, and other protocols in a modular and composable manner, and one can separate the understanding of what cryptography achieves from the technical security definitions and proofs, which is useful for didactic purposes and protocol design.

3.13 A New Information Set Decoding Algorithm

Alexander Meurer (Ruhr-Universität Bochum, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Alexander Meurer


Joint work of May, Alexander; Meurer, Alexander; Thomae, Enrico
Main reference A. May, A. Meurer, E. Thomae, “Decoding Random Linear Codes in $O(2^{0.054n})$,” *Advances in Cryptology – ASIACRYPT 2011*, pp. 107–124, LNCS Vol. 7073.
URL http://dx.doi.org/10.1007/978-3-642-25385-0_6

Decoding random linear codes is a fundamental problem in complexity theory and lies at the heart of almost all code-based cryptography. The best attacks on the most prominent code-based cryptosystems such as McEliece directly use decoding algorithms for linear codes. The asymptotically best decoding algorithm for random linear codes of length n was for a long time Sterns variant of information-set decoding running in time $\mathcal{O}(2^{0.05563n})$.

Recently, Bernstein, Lange and Peters proposed a new technique called Ball-collision decoding which offers a speed-up over Sterns algorithm by improving the running time to $\mathcal{O}(2^{0.05558n})$. In this talk, we present a new algorithm for decoding linear codes that is inspired by a representation technique due to Howgrave-Graham and Joux in the context of subset sum algorithms. Our decoding algorithm offers a rigorous complexity analysis for random linear codes and brings the time complexity down to $\mathcal{O}(2^{0.05363n})$.

3.14 Code Obfuscation with a Stateless Hardware Token

Joern Mueller-Quade (KIT – Karlsruhe Institute of Technology, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Joern Mueller-Quade

Code obfuscation is one of the most powerful concepts in cryptography. It could yield functional encryption, digital rights management, and maybe even secure cloud computing. However, general code obfuscation has been proven impossible and the research then focused on obfuscating very specific functions, studying weaker security definitions for obfuscation, and using tamper-proof hardware tokens to achieve general code obfuscation. Following this last line this work presents the first scheme which bases general code obfuscation of multiple programs on one single stateless hardware token.

Our construction is proven secure in the UC-framework and proceeds in three steps:

1. We construct an obfuscation scheme based on fully homomorphic encryption (FHE) and a hybrid functionality conditional decrypt, which decrypts the result of a homomorphic computation given a proof that the computation was performed as intended. One difficulty of the first step are possible decryption errors in the FHE. These decryption errors can occur whenever the randomness for the encryption is chosen maliciously by the receiver of the obfuscated code. Such decryption errors then could make a real obfuscated computation distinguishable from a black box use of the non-obfuscated program.


2. Given two common reference strings (CRS) we construct a UC-protocol realizing the functionality conditional decrypt with a stateless hardware token.

As the token is stateless it is resettable by a dishonest receiver and the proofs given to the token must be resettablely sound. One additional difficulty occurs when the issuer of the token can be corrupted. A malicious token can be stateful and it cannot be prevented that it aborts after a hardwired number of invocations. To prevent adaptive behavior of a malicious token the data of the receiver has to be hidden from the token and the proofs given to the token must even hide the size of the program and the length of the computation.

3. Last we construct a protocol constructing a CRS with a stateless hardware token. Care has to be taken here to not let the token learn anything about the resulting CRS which could not be simulated, because the very same token will later be used in a protocol based on the security of this CRS.

3.15 Oblivious Transfer with Anonymous Access Control and Pricing


Gregory Neven (IBM Research – Zürich, CH)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Gregory Neven

This talk unifies several schemes in a line of work that combines oblivious transfer (OT) protocols with anonymous credentials. It shows how to enhance a basic OT protocol (presented at Eurocrypt 2007) to support priced records (Financial Cryptography 2010) and anonymous access control with known (ACM CCS 2009) and hidden policies (PKC 2011).

3.16 On the Joint Security of Encryption and Signature, Revisited

Kenny Paterson (Royal Holloway University – London, GB)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Kenny Paterson

Joint work of Paterson, Kenneth G.; Schuldt, Jacob C.N.; Stam, Martijn; Thomson, Susan
Main reference K.G. Paterson, J.C.N. Schuldt, M. Stam, S. Thomson, “On the Joint Security of Encryption and Signature, Revisited,” *Advances in Cryptology – ASIACRYPT 2011*, pp. 161–178, LNCS Vol. 7073.
URL http://dx.doi.org/10.1007/978-3-642-25385-0_9

We revisit the topic of joint security for combined public key schemes, wherein a single keypair is used for both encryption and signature primitives in a secure manner. While breaking the principle of key separation, such schemes have attractive properties and are sometimes used in practice. We give a general construction for a combined public key scheme having joint security that uses IBE as a component and that works in the standard model. We provide a more efficient direct construction, also in the standard model. We then consider

the problem of how to build signcryption schemes from jointly secure combined public key schemes. We provide a construction that uses any scheme to produce a triple of schemes signature, encryption and signcryption that are jointly secure in an appropriate and strong security model.

3.17 Hidden Vector Encryption Fully Secure Against Unrestricted Queries – No Query Left Unanswered

Giuseppe Persiano (University of Salerno, IT)

License  Creative Commons BY-NC-ND 3.0 Unported license

© Giuseppe Persiano

Joint work of De Caro, Angelo; Iovino, Vincenzo; Persiano, Giuseppe;

Main reference A. De Caro, V. Iovino, G. Persiano, “Hidden Vector Encryption Fully Secure Against Unrestricted Queries,” IACR Cryptology ePrint Archive, 2011/546, 2011.

URL <http://eprint.iacr.org/2011/546>

Predicate encryption is an important cryptographic primitive that enables fine-grained control on the decryption keys. Roughly speaking, in a predicate encryption scheme the owner of the master secret key can derive secret key sk_P , for any predicate P from a specified class of predicates. In encrypting a message, the sender can specify an attribute vector and the resulting ciphertext \tilde{X} can be decrypted only by using keys sk_P such that $P(\tilde{x}) = 1$.


Our main contribution is the *first* construction of a predicate encryption scheme that can be proved *fully* secure against *unrestricted* queries by probabilistic polynomial-time adversaries under non-interactive constant sized (that is, independent of the length ℓ of the attribute vectors) hardness assumptions on bilinear groups of composite order.

Specifically, we consider *hidden vector encryption* (HVE in short), a notable case of predicate encryption introduced by Boneh and Waters. In a HVE scheme, the ciphertext attributes are vectors $\vec{x} = \langle x_1, \dots, x_\ell \rangle$ of length ℓ over alphabet Σ , keys are associated with vectors $\vec{y} = \langle y_1, \dots, y_\ell \rangle$ of length ℓ over alphabet $\Sigma \cup \{\star\}$ and we consider the $\text{Match}(\vec{x}, \vec{y})$ predicate which is true if and only if, for all i , $y_i \neq \star$ implies $x_i = y_i$. Previous constructions restricted the proof of security to adversaries that could ask only *non-matching* queries; that is, for challenge attribute vectors \vec{x}_0 and \vec{x}_1 , the adversary could ask only for keys of vectors \vec{y} for which $\text{Match}(\vec{x}_0, \vec{y}) = \text{Match}(\vec{x}_1, \vec{y}) = \text{false}$.

Our proof employs the dual system methodology of Waters, that gave one of the first fully secure construction in this area, blended with a careful design of intermediate security games that keep into account the relationship between challenge ciphertexts and key queries.

3.18 Commitments and Efficient Zero- Knowledge from Hard Learning Problems

Krzysztof Pietrzak (IST Austria – Klosterneuburg, AT)

License  Creative Commons BY-NC-ND 3.0 Unported license

© Krzysztof Pietrzak

Joint work of Jain, Abhishek; Pietrzak, Krzysztof; Tentes, Aris


I'll first show a simple (non-interactive) perfectly binding string commitments scheme whose security (i.e. hiding property) relies on the learning parity with noise (LPN) problem.

Next, I'll give an efficient zero-knowledge proof of knowledge (a Σ -protocol) for any linear function of the secret used to generate LPN instances.

Combining these results, we get a very simple string commitment scheme which allows to efficiently (but interactively) open any linear function (e.g. a subset) of the committed string, while revealing no other information. We borrow ideas from Stern [CRYPTO'93], and for the special case where one opens an "empty" commitment, our protocol can be seen as a "dual" version of Stern's public-key identification protocol.

3.19 Careful with Composition: Limitations of Indifferentiability and Universal Composability

Thomas Ristenpart (*University of Wisconsin – Madison, US*)

License  Creative Commons BY-NC-ND 3.0 Unported license

© Thomas Ristenpart

URL <http://eprint.iacr.org/2011/339>

We exhibit a hash-based storage auditing scheme which is provably secure in the random-oracle model (ROM), but easily broken when one instead uses typical indifferntiable hash constructions. This contradicts the widely accepted belief that the indifferntiability composition theorem applies to any cryptosystem. We characterize the uncovered limitation of the indifferntiability framework by showing that the formalizations used thus far implicitly exclude security notions captured by experiments that have multiple, disjoint adversarial stages. Examples include deterministic public-key encryption (PKE), password-based cryptography, hash function nonmalleability, key-dependent message security, and more. We formalize a stronger notion, reset indifferntiability, that enables an indifferntiability-style composition theorem covering such multi-stage security notions, but then show that practical hash constructions cannot be reset indifferntiable. We discuss how these limitations also affect the universal composability framework. We finish by showing the chosen-distribution attack security (which requires a multi-stage game) of some important public-key encryption schemes built using a hash construction paradigm introduced by Dodis, Ristenpart, and Shrimpton.

3.20 Pseudo Random Functions and Lattices

Alon Rosen (*The Interdisciplinary Center – Herzliya, IL*)

License  Creative Commons BY-NC-ND 3.0 Unported license

© Alon Rosen

Joint work of Banerjee, Abhishek; Peikert, Chris; Rosen, Alon

Main reference A. Banerjee, C. Peikert, A. Rosen, "Pseudorandom Functions and Lattices," IACR Cryptology ePrint Archive, 2011/401, 2011.


URL <http://eprint.iacr.org/2011/401>

We give direct constructions of pseudorandom function (PRF) families based on conjectured hard lattice problems and learning problems. Our constructions are asymptotically efficient and highly parallelizable in a practical sense, i.e., they can be computed by simple, relatively *small* low-depth arithmetic or boolean circuits (e.g., in NC^1 or even TC^0). In addition, they are the first low-depth PRFs that have no known attack by efficient quantum algorithms.

Central to our results is a new “derandomization” technique for the learning with errors (LWE) problem which, in effect, generates the error terms deterministically.

3.21 Identification and Signatures Based on NP-Hard Problems of Indefinite Quadratic Forms

Claus Peter Schnorr (Goethe-Universität Frankfurt am Main, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Claus Peter Schnorr

Joint work of Hartung, Rupert J.; Schnorr, Claus Peter;

Main reference R.J. Hartung, C.P. Schnorr, “Identification and signatures based on NP-hard problems of indefinite quadratic forms,” in J. Math. Crypt., 2(4), pp. 327–341.

Quadratic form cryptography covers both lattice cryptography and factoring cryptography. While the lattice problems **SVP** and **CVP** of finding shortest and closest lattice vectors are only **NP**-hard for large, non constant dimension n representation and equivalence problems for quadratic forms are already **NP**-hard for dimension $n = 3$. While integers N can be factored in subexponential time solving **NP**-hard problems, with a reduction from **3SAT** that merely linearly increases the bit length, require exponential time as **3SAT** requires exponential time for all known algorithms.

The following problems **CBR** of finding small representations of integers and **CBE** of finding small equivalence transforms are **NP**-hard for indefinite, quadratic forms of arbitrary dimension $n \geq 3$.

CBR: Given an n -ary quadratic form f and an integer $m \in \mathbb{Z}$

Find a representation $x \in \mathbb{Z}^n$ with a given bound, i.e., $f(x) = m$.


CBE: Given two equivalent n -ary quadratic forms f_0, f_1

Find an equivalence transform $T \in GL_n(\mathbb{Z})$ with a given bound.

We present a practical identification scheme and a corresponding signature scheme on quaternary quadratic forms ($n = 4$) for which the best known attacks require to solve **NP**-hard problems and require exponential time. We use **CBE** of anisotropic forms.

3.22 Security of Blind Signatures Revisited

Dominique Schroeder (University of Maryland – College Park, US)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Dominique Schroeder

Joint work of Schroeder, Dominique; Unruh, Dominique

Main reference U. Schroeder, D. Unruh, “Security of Blind Signatures Revisited,” IACR Cryptology ePrint Archive, 2011/316, 2011.

URL <http://eprint.iacr.org/2011/316>

We revisit the definition of unforgeability of blind signatures as proposed by Pointcheval and Stern (Journal of Cryptology 2000). Surprisingly, we show that this established definition falls short in two ways of what one would intuitively expect from a secure blind signature scheme: It is not excluded that an adversary submits the same message m twice for signing, and then produces a signature for $m' \neq m$. The reason is that the forger only succeeds if *all* messages are distinct. Moreover, it is not excluded that an adversary performs k signing queries and produces signatures on $k + 1$ messages as long as *each* of these signatures does not pass verification with probability 1.

Finally, we proposed a new definition, honest-user unforgeability, that covers these attacks. We give a simple and efficient transformation that transforms any unforgeable blind signature scheme (with deterministic verification) into an honest-user unforgeable one.

3.23 Integrity Notions for Encryption Schemes

Bjoern Tackmann (ETH Zürich, CH)

License © © © Creative Commons BY-NC-ND 3.0 Unported license
© Bjoern Tackmann

Joint work of Maurer, Ueli; Rueedlinger, Andreas; Tackmann, Bjoern;

The fundamental goal of encryption schemes is to protect the confidentiality of the encrypted plaintext messages. Some applications, however, require the encryption scheme to additionally guarantee some type of integrity: A prime example is the Authenticate-then-Encrypt transformation used, e.g., in TLS.

The security definitions for such confidentiality or integrity guarantees that appear in the literature are mostly game-based and do not provide any composability guarantees. Moreover, their exact semantics often remain unclear.

In this work, we use the approach of constructive cryptography for a systematic treatment of confidentiality and integrity, questioning the justification for the existing game-based security properties. We translate previous game-based notions into integrity guarantees of channels, and find that some of the considered notions are too weak, such as INT-PTXT, some are appropriate, and others are too strong, such as INT-CTXT. Some notions have semantics that appear inappropriate for symmetric encryption, such as IND-CCA.

3.24 The equivalence of the random oracle model and the ideal cipher model, revisited.

Stefano Tessaro (University of California – San Diego, US)

License © © © Creative Commons BY-NC-ND 3.0 Unported license
© Stefano Tessaro

Joint work of Holenstein, Thomas; Künzler, Robin; Tessaro, Stefano;

Main reference T. Holenstein, R. Künzler, S. Tessaro, “The equivalence of the random oracle model and the ideal cipher model, revisited,” in Proc. 43rd ACM Symp. on Theory of Computing (STOC’11), pp. 89–98, 2011.

URL <http://dx.doi.org/10.1145/1993636.1993650>

We consider the cryptographic problem of constructing an invertible random permutation from a public random function (i.e., which can be accessed by the adversary). This goal is formalized by the notion of indistinguishability of Maurer et al. (TCC 2004). This is the natural extension to the public setting of the well-studied problem of building random permutations from random functions, which was first solved by Luby and Rackoff (Siam J. Comput., ’88) using the so-called Feistel construction.

The most important implication of such a construction is the equivalence of the random oracle model (Bellare and Rogaway, CCS ’93) and the ideal cipher model, which is typically used in the analysis of several constructions in symmetric cryptography.


Coron et al. (CRYPTO 2008) gave a rather involved proof that the six-round Feistel construction with independent random round functions is indistinguishable from an invertible random permutation. Also, it is known that fewer than six rounds do not suffice for

indifferentiability. The first contribution (and starting point) of our paper is a concrete distinguishing attack which shows that the indifferentiability proof of Coron et al. is not correct. In addition, we provide supporting evidence that an indifferentiability proof for the six-round Feistel construction may be very hard to find.

To overcome this gap, our main contribution is a proof that the Feistel construction with eighteen rounds is indifferentiable from an invertible random permutation. The approach of our proof relies on assigning to each of the rounds in the construction a unique and specific role needed in the proof. This avoids many of the problems that appear in the six-round case.

3.25 Secure Computation with Corruptible Setups

Vassilis Zikas (University of Maryland – College Park, US)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Vassilis Zikas

Joint work of Katz, Jonathan; Kiayias, Aggelos; Zikas, Vassilis; Zhou, Hong-Sheng

Universally composable (UC) protocols satisfy strong and desirable security properties. Unfortunately, soon after the introduction of the UC framework it was shown that in the “plain” model most cryptographic tasks cannot be realized without an honest majority. Researchers since then have therefore proposed various forms of “trusted setup”, and have shown many setups that are *complete* and can thus be leveraged to securely carry out any desired task.

With only a few notable exceptions, past work has viewed these setup assumptions as being implemented by some ideal, incorruptible entity. In reality, however, setups would likely be carried out by some mechanism that could be subverted, or by some party that could be compromised. Most prior work provides no guarantees in such cases.

We propose here a clean, general, and generic approach for modeling potential corruption of setups within the UC framework, where such corruption might be fail-stop, passive, or arbitrary and is in addition to possible corruption of the parties. We also show several results regarding feasibility in this model for these corruption types (and their combinations) for different specifications of the corruptible sets. For example, we show that given m complete setups, any t of which might be actively corrupted, general secure computation is possible iff $t < m/2$ even when arbitrarily many parties are actively corrupted.

4 Working Groups

Because one of the most renowned conferences in our area, Eurocrypt, had the submission deadline on Sept 30th, some researchers took the opportunity to prepare their submission with attending colleagues and submit from Dagstuhl. Our feedback was that, being able to work with one’s co-authors face to face, was in fact one of the nice side effects of this seminar. We currently know of more than half a dozen submissions to Eurocrypt which have been worked out to the final state during the seminar.

5 Open Problems

The interest in hash functions, one of the three main areas of interest of the seminar, was decent but smaller than expected, with only a few talks covering this topic. Here, the participants mainly discussed model-related issues with “idealized” hash functions (so-called random oracles), which are not known to yield the desired impact on the standardization process. Instead, the general idea of “constructive cryptography”, presented by one of the organizers, Ueli Maurer, as a new view on cryptography incited many discussions. We found the occasion at Dagstuhl a perfect setting to put such a topic to scrutiny.

6 Panel Discussions

The panel discussion on the “Future of Public-Key Cryptography” revealed that the area of (public-key) cryptography is by no means a closed area, despite the fact that basic primitives for digital signatures and encryption are known for decades now. Instead, new challenges arise permanently through the changing environment in which crypto is used. Some participants expressed their wish to work more towards bridging applied security and cryptography but the general consensus was that the current seminar is —and should be— the platform for public-key cryptography as a whole.

7 Scientific Output

As for now, we know of at least the following four publications that, to some extent, originated from joint collaborations at the Dagstuhl seminar.

- Yevgeniy Dodis, Eike Kiltz, Krzysztof Pietrzak, and Daniel Wichs, “Message Authentication, Revisited”, accepted at Eurocrypt 2012.
- Anja Becker, Antoine Joux, Alexander May, and Alexander Meurer, “Decoding Random Binary Linear Codes in $2^{n/20}$: How $1 + 1 = 0$ Improves Information Set Decoding”, accepted at Eurocrypt 2012.
- Dennis Hofheinz, “All-But-Many Lossy Trapdoor Functions”, accepted at Eurocrypt 2012.

Participants

- Anja Becker
University of Versailles, FR
- Johannes Blömer
Universität Paderborn, DE
- Jan Camenisch
IBM Research – Zürich, CH
- Melissa Chase
Microsoft Res. – Redmond, US
- Yvo Desmedt
University College London, GB
- Yevgeniy Dodis
New York University, US
- Pooya Farshim
TU Darmstadt, DE
- Marc Fischlin
TU Darmstadt, DE
- Juan A. Garay
AT&T Res. – Florham Park, US
- Dennis Hofheinz
KIT – Karlsruhe Institute of Technology, DE
- Thomas Holenstein
ETH Zürich, CH
- Yuval Ishai
Technion – Haifa, IL
- Antoine Joux
University of Versailles, FR
- Eike Kiltz
Ruhr-Universität Bochum, DE
- Anja Lehmann
IBM Research – Zürich, CH
- Stefan Lucks
Bauhaus-Universität Weimar, DE
- Ueli Maurer
ETH Zürich, CH
- Alexander May
Ruhr-Universität Bochum, DE
- Alexander Meurer
Ruhr-Universität Bochum, DE
- Jörn Müller-Quade
KIT – Karlsruhe Institute of Technology, DE
- Gregory Neven
IBM Research – Zürich, CH
- Tatsuaki Okamoto
NTT Labs. – Tokyo, JP
- Kenneth G. Paterson
RHUL – London, GB
- Giuseppe Persiano
University of Salerno, IT
- Krzysztof Pietrzak
IST Austria –
Klosterneuburg, AT
- Thomas Ristenpart
University of Wisconsin –
Madison, US
- Alon Rosen
The Interdisciplinary Center –
Herzliya, IL
- Claus Peter Schnorr
Goethe-Universität Frankfurt am
Main, DE
- Dominique Schroeder
University of Maryland – College
Park, US
- Björn Tackmann
ETH Zürich, CH
- Stefano Tessaro
University of California – San
Diego, US
- Bogdan Warinschi
University of Bristol, GB
- Vassilis Zikas
University of Maryland – College
Park, US

