SCHLOSS DAGSTUHL
Leibniz-Zentrum für Informatik

Dagstuhl News

January – December 2011

Volume 14
2012

Leibniz Association

Schloss Dagstuhl, the Leibniz Center for Informatics, is operated by a non-profit organization. Its objective is to promote world-class research in computer science and to host research seminars which enable new ideas to be showcased, problems to be discussed and the course to be set for future development in this field.

Welcome

Here are the Dagstuhl News for 2011, the 14th edition of the "Dagstuhl News", a publication for the members of the Foundation "Informatikzentrum Schloss Dagstuhl", the *Dagstuhl Foundation* for short.

The main part of this volume consists of collected resumees from the Dagstuhl Seminar Reports 2011 and Manifestos of Perspectives Workshops. We hope that you will find this information valuable for your own work or informative as to what colleagues in other research areas of Computer Science are doing. The full reports for 2011 are on the Web under URL: `http://drops.dagstuhl.de/dagrep`.

The new guesthouse with 7 more rooms is in use since January. It allows us to run two seminars in parallel. We already make heavy use of this. It also offers more capacity for meetings of projects, of research programmes etc.

Thanks
I would like to thank you for supporting Dagstuhl through your membership in the *Dagstuhl Foundation.* Thanks go to Fritz Müller for editing the resumees collected in this volume.

Reinhard Wilhelm (Scientific Director)

Saarbrücken, July 2012

Contents

Subject Area 1

Data Structures, Algorithms, Complexity

1.1 Theory and Applications of Graph Searching Problems (GRASTA 2011)

Authors: Fedor V. Fomin, Pierre Fraigniaud, Stephan Kreutzer, and Dimitrios M. Thilikos

Graph searching is often referred to, in a more playful language, as a pursuit-evasion game (or, alternatively, cops and robbers game). This is a kind of game where one part is a set of escaping mobile entities, called evaders (or fugitives), that hide in a graph representing a network, and the other part is a number of chasing agents, called searchers (or pursuers), that move systematically in the graph. The game may vary significantly according to the capabilities of the evaders and the pursuers in terms of relative speed, sensor capabilities, visibility, etc. The objective of the game is to capture the evaders in an optimal way, where the notion of optimality itself admits several interpretations.

Graph searching revealed the need to express in a formal mathematical way intuitive concepts such as avoidance, surrounding, sense of direction, hiding, persecution, and threatening. There are many variants of graph searching studied in the literature, which are either application driven, i.e. motivated by problems in practice, or are inspired by foundational issues in Computer Science, Discrete Mathematics, and Artificial Intelligence including

- Information Seeking
- Robot motion planning
- Graph Theory
- Database Theory and Robber and Marshals Games
- Logic
- Distributed Computing
- Models of computation
- Network security

The objective of the seminar was to bring researchers from the widest possible variety of disciplines related to graph searching and we will especially encourage the maximum interplay between theory and applications. The meeting initiated the exchange of research results, ideas, open problems and discussion about future avenues in Graph Searching. As a fruit of this encounter new research results, open problems, and methodologies will appear, especially those of interdisciplinary character.

1.2 Combinatorial and Algorithmic Aspects of Sequence Processing

Seminar No. **11081**
Organizers: Maxime Crochemore, Lila Kari, Mehryar Mohri, and Dirk Nowotka
21.–25. February, 2011 – www.dagstuhl.de/11081

Authors: Maxime Crochemore, Lila Kari, Mehryar Mohri, and Dirk Nowotka

The object of concern of this seminar, *sequences*, implies a large degree of generality. It plays an essential rôle in many fields and constitutes a true cross section area. Hence, the seminar was designed to bring together researchers from different disciplines whose interest are string processing algorithms and related combinatorial problems on words. Scientists working in the following fields were invited to consider the seminar's topic from a wide range of perspectives:

- Combinatorics on Words,
- Computational Biology,
- Stringology,
- Natural Computing,
- Machine Learning.

This Dagstuhl seminar was attended by 40 researchers from 13 countries. Everyone of the five topics above was about equally represented. Given the extremely interdisciplinary approach of this meeting it was an obvious necessity to hold a tutorial on each one of the participating research areas. These tutorials were held over the first and the morning of the second seminar day. They provided a good introduction for the non-specialists and triggered the first scientific discussions and exchanges.

A second (and standard) element of this seminar were regular talks, of course. A total of 15 talks were presented. It has to be noted that one could experience a very productive atmosphere during the whole seminar. All talks were well-attended and accompanied with interesting comments. Plenty of time was reserved for questions and discussions which was actively used by the participants.

The third element of the seminar were open problem sessions which did yield a larger attention to a range of problems, only some of them are included in this report. These open problem sessions provided the ideal ground for the ignition of new research lines and cooperations. Just to mention one example, the paper "On the regularity of iterated hairpin completion of a single word" (arXiv:1104.2385v1) resulted from the collaboration of Steffen Kopecki and Shinnosuke Seki initiated at this Dagstuhl seminar. In the light of such developments, it can be safely claimed that this seminar was a success.

Given the quality of presentations on this seminar and the constructive intensity of discussions between and after the talks, it is self-evident that follow-ups will be attempted.

After this initial meeting of different communities, where common problems were identified, personal contacts established and first cooperations initiated, further events can be sharpened in focus and more on particular cross section topics regarding combinatorial and algorithmic problems in sequence processing.

Finally, we would like to say that the organization of a meeting for researchers of so unusually diverse fields bears a certain risk. However, it can be said that the event turned out better than expected. It was more than worthwhile to have taken that risk. We are grateful to all participants for their contributions to this successful seminar as well as to the staff of Schloss Dagstuhl for their perfect service.

1.3 Packing and Scheduling Algorithms for Information and Communication Services

Authors: Klaus Jansen, Claire Mathieu, Hadas Shachnai, and Neal E. Young

Packing and scheduling are one area where mathematics meets puzzles. While many of these problems stem from real-life applications, they have also been of fundamental theoretical importance. In a *packing* problem given is a set of items and one or more (multi-dimensional) bins. The objective is to maximize the profit from packing a subset of the items, or to minimize the cost of packing *all* items. In a *scheduling* problem, given are a set of jobs and a set of machines. One needs to schedule the jobs to run on the machines (under some constraints) so as to optimize an objective function that depends on the order of the jobs, on their completion times or on the machines by which they are processed.

Storage allocation in computer networks, cutting stock problems in various industries and production planning are only few of the applications of packing and scheduling. With the growing impact of next generation technologies in information and communication services (some examples are Video-on-Demand systems, web applications and wireless networks), practitioners as well as theoreticians seek fast and efficient solutions for new variants of some classic packing and scheduling problems, which are crucial for optimizing the performance of these systems.

Since many of these problems are NP-hard, it is natural to seek efficient approximate solutions. Traditionally, such approximations are obtained by using fundamental tools from combinatorial optimization and mathematical programming. While for some of the problems there exist algorithms which achieve the best possible approximation ratio, one major effort of this community has been to close the gaps in running times between heuristic solutions, which perform well in practice, and algorithms which are provably efficient in terms of approximation ratio, but impractical in use. The large class of approximation schemes for packing and scheduling problems has been the recent target of this effort.

Parameterized complexity uses refined measures for the approximability of a given problem, by referring, e.g., to approximation with instance parameters, by defining performance functions (instead of performance ratios) and by defining the quality of approximation as parameter. Such measures provide further insight to the studied problems and lead to the design of algorithms that work efficiently if the parameters of the input instance are small (even if the size of the input is large). Efficient parameterization for packing and scheduling problems is a major challenge on the way to obtaining practical algorithms.

During the 5 days of the seminar, 24 talks were given by the participants. Five of these talks were two-hour tutorials and 60-minute survey talks on various topics: Kirk Pruhs gave an exciting tutorial on the challenges faced by designers of algorithms for green computing;

Dániel Marx talked about several existing connections between approximation algorithms and fixed-parameter algorithms; Ola Svensson gave an overview of the implications and techniques of two fascinating hardness of approximation results for shops and precedence constraints scheduling; Neal Young talked about using lagrangian-relaxation algorithms to solve packing and covering problems, and Magnús Halldórsson gave an overview of recent analytic work on scheduling wireless links.

The seminar successfully brought together both experts and newcomers from the areas of packing and sequencing, combinatorial optimization, mathematical programming, and parameterized complexity, with many interesting interactions. The talks left plenty of time for discussion in the afternoon. An open problem session was held on Tuesday, and problems raised there were discussed by different groups throughout the seminar and in a research groups session on Friday. A session on current and future trends in scheduling was held on Thursday, and brought up some exciting issues relating to this area.

1.4 Computational Complexity of Discrete Problems

Seminar No. **11121**
Organizers: Martin Grohe, Michal Koucký, Rüdiger Reischuk, and Dieter van Melkebeek
20.–25. March, 2011 – www.dagstuhl.de/11121

Authors: Martin Grohe, Michal Koucký, Rüdiger Reischuk, and Dieter van Melkebeek

Introduction and Goals

Computational models like Turing machines and Boolean circuits work on discrete input data. Even quantum computation and communication studied in the recent past are mainly applied to solve discrete problems. Analysing the computational complexity of such problems with respect to these models is one of the central topics in the theory of computation. Researchers try to classify algorithmic problems according to complexity measures like time and space – both in the uniform and in the nonuniform setting. A variety of specialized computational models have been developed in order to better measure the complexity of certain classes of discrete problems.

Randomness has turned out to be another fundamental measure and added a lot of new intricate questions. Performing probabilistic choices within an algorithm one can design solution strategies for a given computational problem for which there are no obvious deterministic ones. Recently, large effort has been taken to remove randomness from probabilistic algorithms, so called derandomization. Here one tries to develop general techniques that can be applied to a wide range of discrete problems.

Information transfer is investigated according to the amount of communication necessary in different scenarios like 1-way channels or a bounded number of communication rounds. This is a basis for the design of efficient communication protocols. Furthermore, it has been observed that often ordinary computational problems given to a specific computational device can formally be analysed elegantly by concentrating on information flow aspects.

In addition, other computational processes arising in diverse areas of computer science have been studied, each with their own relevant complexity measures. Several of those investigations have evolved into substantial research areas, including:

- approximability (motivated by optimization),
- computational learning theory (motivated by artificial intelligence),
- query complexity (motivated by databases).

The analysis and relative power of basic models of computation remains a major challenge. New lower bound techniques for explicitly defined functions have brought the field a major step forward. For example, close connections have been discovered between circuit lower bounds for certain uniform complexity classes and the existence of pseudorandom generators and the possibility of efficient derandomization.

The seminar "Computational Complexity of Discrete Problems" has evolved out of the series of seminars entitled "Complexity of Boolean Functions," a topic that has been covered at Dagstuhl on a regular basis since the foundation of this research center. Over the years, the focus on nonuniform models has broadened to include uniform ones as well.

A salient feature of the current research in computational complexity is the interpenetration of ideas from different subareas of computational complexity and from other fields in computer science and mathematics. By organizing a generic seminar on computational complexity we have aimed to attract researchers from those various subareas and foster further fruitful interactions.

Organization of the Meeting

47 researchers from all over the world including a good number of young scientists met in Dagstuhl for this seminar. Every day we started with a longer survey talk on recent advances in specific topics that had been selected in advance by the organizers. We thank our colleagues who agreed to prepare and give these presentations:
— Prahladh Harsha took the duty for *threshold functions,*
— Paul Beame for AC^0 *circuits,*
— Troy Lee for *communication complexity,*
— Eli Ben-Sasson for *extractors,* and
— Robin Moser for *constraint satisfaction problems.*

The surveys were followed by shorter talks on new results obtained by the participants. We could schedule 30 of such plenary talks such that enough additional time was left for discussions in smaller groups on a spontaneous basis. In addition, Tuesday evening was devoted to a rump session where everybody could present his favourite open problem.

The first evening the participants could also extend their knowledge and taste in a completely different area, namely arts. We took part in the vernisage of the art exhibit by Irene Zaharoff who presented her colorful paintings in the corridors of the new building. Everybody got so excited that the spontaneous idea to select one of her paintings and support a donation to Dagstuhl was implemented on the spot.

Topics and Achievements

We shortly review the main topics that have been discussed during the meeting. Further details as well as additional material can be found in the abstracts following.

Randomized Computations, Derandomization, and Testing

The complexity of randomized computations and its theoretical foundation was a major issue of the seminar. It showed up in about half of all contributions. Eric Allender discussed the computational power of Kolmogorov-random strings, while Andrej Bogdanov showed how to construct pseudorandom generators for read-once formulas and Pavel Pudlák for read-once permutation branching programs. Improved pseudorandom generators for a

special class of discrete functions called combinatorial checkerboards were given by Thomas Watson.

For randomness extractors improved constructions were presented by Eli Ben-Sasson for the case of two sources and by Xin Li for three sources. For restricted sources that are generated by circuits of constant depth extractors were designed by Emanuele Viola.

Markus Bläser discussed the randomized complexity of identity tests for sparse polynomials, while Beate Bollig did this for integer multiplication in the OBDD model generalizing her deterministic lower bound shown at the previous meeting.

Derandomization techniques were presented by Matthew Anderson for zero tests of multi-linear arithmetic formulas and by Robin Moser for Schöning's satisfiability test of k-CNF formulas. Eldar Fischer considered property testing of monotone formulas.

Communication Complexity

How many bits two parties have to exchange in order to compute a given function if the input is distributed among them? This is a fundamental question for the design of communication protocols. To determine the Hamming distance of two n-bit strings it has been well known that the trivial solution of one party sending his string to the other party is optimal. It was open for quite a while if this still holds if the two parties get the additional information that the distance of their inputs is either small or large (let's say not in the region between $[n/2 - \sqrt{n}, n/2 + \sqrt{n}]$). Oded Regev could resolve this question for general probabilistic protocols with unbounded rounds of communication by proving a linear lower bound.

Allowing the two parties also to exchange quantum bits leads to quantum protocols which – in contrast to quantum computers – are already used in practice. There are examples known that quantum bits can lead to an exponential decrease of the amount of communication. Oded Regev in his second contribution showed that such a separation can even be obtained when comparing 1-way quantum protocols with 2-way classical protocols.

Troy Lee investigated the query complexity of quantum states and a generalization of this problem called state conversion. He defined a new norm for the distance of quantum states and proved that this gives an appropriate measure.

Andrew Drucker showed how fault-tolerant protocols can be used to improve upon more complex objects called probabilistically checkable debate systems. This has implications for the approximability of problems in PSPACE.

Complexity Classes

For classical complexity classes some progress was reported, too. Michael Elberfeld showed how graph problems restricted to instances of bounded tree width can be solved in logarithmic space, while Fabian Wagner was able to solve the isomorphism problem for such instances in LogCFL. Meena Mahajan presented a detailed investigation of arithmetic circuits of logarithmic depth and their relation to logspace.

For Boolean circuits the tradeoff between size and depth has been investigated from the very beginning. Anna Gál showed with the help of the pebble game how the bounds can be substantially improved for layered circuits. Or Meir took a closer lock at the breakthrough result IP=PSPACE and showed an alternate proof that uses fairly general error-correcting codes instead of polynomials.

Another famous result that unbounded circuits of constant depth cannot compute the parity function was taken up by Paul Beame and Johan Håstad trying to get a finer quantitative statement of this property. Paul described a simulation of such circuits by decision trees which has implications how well AC^0 can approximate the parity function, while Johan presented a direct proof for an upper bound on the correlation between parity and functions computable in constant depth. Nicole Schweikardt gave a precise characterization of the locality of order-invariant first-order queries with arbitrary predicates, which are closely related to the complexity class AC^0.

Srikanth Srinivasan showed that computing the determinant over simple noncommutative rings is as hard as computing the permanent in the commutative case, thus establishing a huge complexity gap between commutative and noncommutative domains.

Further Topics

Improving the construction of good error correcting codes by combining classical codes with outer codes was adressed by Amnon Ta-Shma. Matthias Krause presented a new technique to prove security properties of cryptographic hash functions. Philipp Woelfel considered random walks on a line towards a target where the searcher can fix an arbitrary distribution on his probabilistically chosen next movement. A lower bound on the first hitting time was shown matching previously known upper bounds for this problem. Jakob Nordström considered linear invariance properties in the realm of property testing. He investigated the problem to decide on the semantic difference for syntactically different descriptions of linear invariances.

Efficiently learning unknown concepts by queries or taking random samples is another topic of general applicability. An important measure in this respect is the VC-dimension which implies bounds on the minimal achievable additive error for any learning algorithm. Ilan Newman asked the same question for the multiplicative error and showed that the triangular rank of a set system gives a corresponding measure. Kristoffer Hansen considered the class of Boolean functions with a constant degree representation where each variable is used only a bounded number of times. He presented a deterministic polynomial time algorithm for this problem. For Boolean functions the notion of sensivity has turned to be an important measure although it seems to be hard to estimate in many cases. Prahladh Harsha considered polynomial threshold functions and showed the first nontrivial upper bound, which has also implications for the learning complexity of these functions.

Conclusion

Investigating the complexity of discrete problems is one of the fundamental tasks in the theory of computation. On the one hand, new algorithmic techniques and new ways to look

at a problem have led to better algorithms and protocols. On the other hand, typically more demanding is the task to prove lower bounds on the computational complexity of a concrete problem. Progress is still continuing, as seen for example in testing, derandomization and explicit constructions of combinatorial objects like extractors, that improves our knowledge considerably. Despite these significant steps forward that have been achieved in several subareas since our previous meeting three years ago, the general feeling among the participants was that we still have to work hard for many more years to get a good understanding what are the limits of efficient computation.

We like to thank the staff at Dagstuhl who – as usual – provided a marvellous surrounding to make this a successful meeting with ample space for undisturbed interactions between the participants.

1.5 Exploiting graph structure to cope with hard problems

Authors: Andreas Brandstädt, Martin Charles Golumbic, Pinar Heggernes, and Ross McConnell

The organizers of the seminar were Andreas Brandstädt, University of Rostock, Germany;
Martin C. Golumbic, University of Haifa, Israel; Pinar Heggernes, University of Bergen,
Norway; and Ross McConnell, University of Colorado, USA. The collector of the presented
seminar material and the editor of the final report was Christian Hundt, University of
Rostock, Germany.

One of the main goals of this Dagstuhl seminar was to gather experts with a common focus
on graph algorithms but with various specializations to attack NP-hard graph problems
using the structure of the input graph. This goal was achieved to a great extent, as the
number of participants of our seminar was above the limit that was given beforehand. The
seminar was granted space for 30 participants, and we had 35 participants on site. Still
there were many experts and young researchers in the field that would like to come but
that could not be invited due to lack of space. The participants that came to our seminar
were from the following countries: 9 from Germany, 7 from USA, 5 from France, 4 from
Israel, 4 from Norway, 1 from Canada, 1 from Korea, 1 from Taiwan, 1 from Turkey, 1
from Greece, and 1 from Great Britain.

By bringing together experts with backgrounds in graph classes, optimization, width
parameters, and parameterized and exact computing, our aim was that several of the
hard problems arising in real applications would eventually find practical solutions. The
seminar program was divided into two parts each day of the seminar: (1) presentation of
new results, and (2) posing and discussions on new problems. In this section, we briefly
summarize the program of the seminar.

On the first day of the seminar, Jesper Nederlof presented new results on solving connectiv-
ity problems parameterized by treewidth in single exponential time, Charis Papadopoulos
presented new results on characterizing the linear clique-width of a class of graphs by
forbidden induced subgraphs, Yngve Villanger presented new kernelization and approx-
imation results on minimum fill-in of sparse graphs, and Martin Milanič presented new
results on hereditary efficiently dominatable graphs. During the problem solving session
of the first day, Pinar Heggernes posed a problem on dense subgraphs on proper interval
graphs, Andreas Brandstädt and Christian Hundt posed problems on k-leaf powers, Dieter
Rautenbach posed a problem on 2-domination on strongly chordal graphs, Pavol Hell
asked which problems that are hard on general digraphs become efficiently solvable on
interval digraphs, and Feodor Dragan introduced and asked questions about the short fill-in
problem.

On the second day of the seminar, Ann Trenk presented a survey on the total linear discrepancy of a poset, Elad Cohen presented new results on vertex intersection graphs of paths on a grid, Yahav Nussbaum presented new results on the recognition of probe proper interval graphs, and Tinaz Ekim presented a survey on polar graphs. During the problem solving session of the second day, Daniel Lokshtanov posed a problem on map graphs, Christian Hundt gave more details on his problem on k-leaf powers, Martin Golumbic posed a problem on the recognition of one-bend EPG and VPG graphs, and Sang-il Oum posed a problem on Bott equivalence between directed acyclic graphs.

On the third day of the seminar, Dieter Rautenbach presented new results on unit interval graphs, R. Sritharan presented new results on finding a sun in graphs, Bernard Ries presented new vertices on coloring vertices of triangle-free graphs, and Wen-Lian Hsu presented new results on PC-trees and planar graphs. In the afternoon of the third day, there was an excursion to Trier and a hike in the close by surroundings of Dagstuhl area.

On the fourth day of the seminar, Pavol Hell presented new results on partitioning chordal graphs, Pim van 't Hof presented new results on contracting graphs to paths and trees, Daniel Lokshtanov presented new results on contracting graphs to bipartite graphs, and Frédéric Maffray presented new results on 3-colorable P_5-free graphs. During the problem solving session of the fourth day, Daniel Lokshtanov posed a problem on the parameterized complexity of protein folding, Yngve Villanger posed a question about forbidden induced subgraphs of circular arc graphs, Christophe Paul asked a question on recognizing circle graphs, and Van Bang Le posed a problem on modified circle graphs.

On the fifth day of the seminar, Feodor Dragan presented new results on graph classes, tree decomposition and approximation algorithms, Elias Dahlhaus presented new results on minimal fill-in ordering of planar graphs in linear time, and Christian Hundt presented new results on the dominating induced matching problem for hole-free graphs. The seminar ended with a brief discussion on all the presented results.

1.6 Graph Drawing with Algorithm Engineering Methods

Seminar No. **11191**
Organizers: Camil Demetrescu, Michael Kaufmann, Stephen Kobourov, and Petra Mutzel
8.–13. May, 2011 – www.dagstuhl.de/11191

Authors: Camil Demetrescu, Michael Kaufmann, Stephen Kobourov, and Petra Mutzel

Automated graph drawing deals with the layout of relational data arising from computer science (database design, data mining, software engineering), and other sciences such as bioinformatics and sociology (social networks). The relational data are typically modeled as graphs, which can be visualized through diagrams drawn in the plane. The main objective is to display the data in a meaningful fashion, (i.e., in a way that shows well the underlying structures) and that often depends on the application domain. Although high quality algorithms exist for many optimization problems that arise in graph drawing, they are often complex and difficult to implement, and theoretically efficient algorithms may have unacceptable runtime behavior even for small-to-medium sized real-world instances. Also large graphs like, e.g., molecular interaction networks, may render exact but complex algorithms infeasible and require approximate or heuristic solutions.

Integrating automated graph drawing techniques into real-world software systems poses several algorithm engineering challenges. To achieve effective implementations, algorithms and data structures designed and analyzed on abstract machine models must be carefully tuned for performance on real hardware platforms. This task is becoming increasingly more difficult due to the impressive growth of data to be visualized in modern applications, as well as their highly dynamic and data-intensive nature. Developers can no longer ignore architectural aspects such as the presence of complex memory hierarchies and multiple cores, which are likely to shape the design of novel algorithmic techniques and the way they will be implemented and engineered in the future.

The aim of this seminar was to bring together researchers from the algorithm engineering and graph drawing communities in order to strengthen and foster collaborations in this area and to identify key research directions for the future.

The seminar was attended by 48 participants from both academia and industry. Much was accomplished, fostered by the productive atmosphere of the Dagstuhl Center. Here we describe some of the more important achievements.

The program consisted of a wide variety of presentations, working group sessions and discussion sessions. The presentations included several survey lectures:

- Beppe Liotta provided a survey on graph visualization paradigms, and discussed general design principles for the realization of effective graph visualization systems.
- Emden Gansner suggested rules in order to get efficient and accurate graph drawing algorithms.
- Ulrik Brandes discussed experimental algorithmics and the relationship between graph drawing algorithms and algorithm engineering.

- Rudolf Fleischer's talk about algorithm engineering and his statement (taken from the definition in the German priority program SPP 1307 *Algorithm Engineering*) that the algorithm engineering cycle should be driven by falsifiable hypotheses, started a lively discussion among the participants.
- Rico Jakob provided a talk on engineering architecture aware algorithms and provided some thoughts about hardware sensitive algorithms. He convinced us that the new computer architectures will strongly influence future algorithmic research.
- Kurt Mehlhorn introduced us into the new and exciting area of slime mould that solves shortest path and network design problems. He would be interested in seeing if slime mould could possible solve graph drawing problems.

Beyond the survey lectures, highlights of the seminar included the two introductory sessions, the open problem sessions, and the working groups.

In two sessions, we have identified over two dozen open problems, which later crystallized into about a dozen well-defined problems, each of which were of interest to several participants. We had working groups on the following topics: Rotating binary trees, feedback arc set convergence, edge bundling models, co-occurence in bipartite graphs, RAC drawings, BRAC drawings, minimum branch spanning tree, cluster tree embedding, point set embeddings, parallel graph drawing, and library of graphs. Participants shared ideas and material using the online seminar Wiki.

The dissemination sessions at the end of the workshop showed that many of the working groups have achieved initial results, which may lead to future publications.

Arguably the most-appreciated features of the Seminar were the lively open discussion sessions, which led to several concrete proposals for the future of the field which, as a result of the workshop, are now being actively pursued.

A big step forward has been done concerning an *online library of graphs*. The graph drawing community would like to have such a library, however, there was no consensus about the requirements on such a graph archive. The working group conducted a survey on requirements for a graph archive during the Dagstuhl seminar. Two groups (Dortmund and Tübingen) presented their ideas and prototypes of such an archive. In order to foster future work and to encourage participation and contributions, it was suggested that the GD proceedings should offer the opportunity to publish papers concerning the library. Moreover, the collection of many benchmark graphs has already begun.

We used the opportunity to bring together experts in algorithm engineering for multi-core algorithms with graph drawing researchers in order to discuss how graph drawing algorithms can be re-engineered to better take advantage of modern computer architecture into account. This working group was inspired by the many different backgrounds of group members. They have discussed how to improve data locality, or exploit multi-core processors, in particular for the widely used Sugiyama drawing method.

Subjectively (from interacting with the attendees) and objectively (from the official feedback data) we believe that the participants enjoyed the great scientific atmosphere offered by Schloss Dagstuhl, and profited from the scientific program and the fruitful discussions. We are grateful for having had the opportunity to organize this seminar. Special thanks are due to Carsten Gutwenger and Karsten Klein for their invaluable assistance in the organization and the running of the seminar.

1.7 Design and Analysis of Randomized and Approximation Algorithms

Seminar No. **11241**
Organizers: Martin E. Dyer, Uriel Feige, Alan M. Frieze, and Marek Karpinski
13.–17. June, 2011 – www.dagstuhl.de/11241

Authors: Martin Dyer, Uriel Feige, Alan M. Frieze, Marek Karpinski

Many, if not most computational tasks that arise in realistic scenarios are computationally difficult, and no efficient algorithms are known that guarantee an exact (or optimal) solution on every input instance. Nevertheless, practical necessity dictates that acceptable solutions be found in a reasonable time. Two basic means for surmounting the intractability barrier are randomized computation, where the answer is optimal with high probability but not with certainty, and approximate computation, where the answer is guaranteed to be within, say, small percentage of optimality. Often, these two notions go hand-in-hand.

The seminar was concerned with the newest developments in the design and analysis of randomized and approximation algorithms. The main focus of the workshop was on the following specific topics: randomized approximation algorithms for optimization problems, approximation algorithms for counting problems, methods for proving approximation hardness, as well as various interactions between them. Here, some new broadly applicable techniques have emerged recently for designing efficient approximation algorithms for various optimization and counting problems as well as for proving approximation hardness bounds.

This workshop has addressed the above topics and some new fundamental insights and paradigms in this area. The 26 regular talks and other presentations delivered at this workshop covered a wide body of research in the above areas. The Program of the meeting and Abstracts of all talks are listed in the subsequent sections of this report.

The meeting was held in a very informal and stimulating atmosphere. Thanks to everyone who made it such an interesting and enjoyable event.

1.8 Coding Theory

Authors: Joachim Rosenthal, M. Amin Shokrollahi, and Judy Walker

This workshop brought together 42 researchers in key areas of coding theory. The seminar had a strong emphasis on interaction and collaboration among researchers. This goal was made clear in a series of five-minute talks by all seminar participants, where they briefly described their research interests and/or described a problem that they were currently working on. The rest of the time was dedicated to more in-depth talks by a selected number of researchers, but the lecture time was kept reasonably short so that long stretches of time were available for seminar participants who wished to discuss further and collaborate in small groups.

The talks fell into several broad categories, the most prominent of which was algebraic coding theory. Algebraic coding theory primarily investigates codes obtained from algebraic constructions. Prime examples of this area of coding theory are codes from algebraic geometry and codes obtained from algebraically constructed expander graphs. This discipline is almost as old as coding theory itself, and has attracted (and continues to attract) some of the brightest minds in the field. Among the most exciting advances in this field in recent years has been the invention of list-decoding algorithms for various classes of algebraic codes; this area began with a landmark paper by Sudan that proposed an algebraic list-decoding scheme for Reed-Solomon codes. List decoding algorithms yield for a received word a short list of codewords that have at most a given distance τ to the received word. The size of the list depends on the distance τ. The methods in this field are mostly algebraic and make use of various properties of multivariate polynomials, or more generally, the properties of "well-behaved" functions in the function field of an irreducible variety. Methods from algebraic geometry are very important in this area. On the computational side the field naturally embeds in the theory of Gröbner bases. There are emerging relationships between this area and codes on graphs, the leading question being whether it is possible to match the superior performance of graph-based codes with list-decoding algorithms, or at least with algorithms that are derived from list-decoding algorithms.

Several talks covered recent advances and current research in some of the most notable questions that relate to algebraic constructions of codes, namely, list decoding; Berlekamp-Massey-like algorithms (for decoding, list-decoding, and Gröbner bases computation); elliptic curve methods and bent and hyperbent functions; rank-metric codes; bounds on codes (semidefinite programming bounds, BCH-like bounds); pivot distributions; properties of specific code constructions such as cyclic orbit codes, classes of self-dual codes, and codes obtained from generalized concatenation; and pseudocodewords, which straddle the boundary between algebraic analysis of codes and message-passing decoding of codes on graphs.

The second area that was covered in the workshop was that of codes on graphs. First proposed by Gallager in the 1950s, the subject of codes on graphs has experienced a huge revival over the past 10–15 years due to the fact that these codes have been shown to have capacity-achieving properties on some channels, and capacity-approaching properties on others. One of the most prominent examples in this area is that of the class of low density parity check (LDPC) codes, which are constructed from sparse bipartite graphs. The sparsity of the graph provides methods for construction of low complexity encoders and decoders. This area is a perfect nurturing ground for cross-fertilization of ideas between computer science, electrical engineering and mathematics. On the engineering side, the amazing simulation results have led some to declare the channel coding problem solved. The main proofs of asymptotic performance of these codes originated in the theoretical computer science community. Despite serious activities in this field many questions remain, related to the performance of graph-based codes and design of good LDPC codes. Several speakers discussed topics related to graph-based codes with applications beyond this field. The topics covered included design of sparse-graph codes for cooperation in communication networks; a new class of iterative decoders for LDPC codes; classes of tail-biting treillises; analysis of treillis pseudocodewords distributions; and bounds on the performance of quantum LDPC codes and their applications to percolations on graphs.

Network coding was another main area covered by the seminar. Network coding theory is concerned with the encoding and transmission of information where there may be many information sources and possibly many receivers. This is a very new area of coding theory and has been around only since 1999. The topic is somewhat far removed from the channel coding problem that Shannon proposed in 1948 and that the other areas of coding theory described in this proposal address. Indeed, there is not yet an agreed-upon formulation of the network coding problem. But there is a notion of linear network coding, and Li, Yeung and Cai showed in 2003 that it is possible to achieve so-called network capacity using linear codes alone. From that basis, much more algebra has been fruitfully introduced into the area. For example, in 2003, Kötter and Médard showed that the linear network codes that achieve a set throughput on a given network are precisely described by the points in a certain algebraic variety associated to the network. The Kötter-Médard approach, of course, relies on knowledge of the network in question. In practice, the network is typically unknown and often continually changing (dynamic). In 2008, Kötter and Kschischang considered random network coding and presented a framework in which the problem of network design is separated from that of code design. The idea is to assume that the network source simultaneously injects k linearly independent vectors from some vector space W into the network. These vectors are combined in various ways and sent through the network, so that the sink receives some linear combinations of them. The mathematical object that is invariant during transmission is the subspace of W spanned by the original k vectors, and so it is natural to consider a code in this context to be a subset of the set $\mathcal{P}(W)$ of all subspaces of W. Since $\mathcal{P}(W)$ is a metric space, the questions of code construction and optimality arise. Some approaches from classical coding theory can therefore be adopted in a fairly straight-forward manner, but deep questions remain, some of which were addressed by seminar speakers. One talk addressed the design and analysis of good end-to-end error-control codes in linear network coding; another topic was the analysis of several strategies for content distribution over network coded systems.

No workshop on such a practical topic as coding theory is complete without the mention of new and emerging applications. Indeed, several of the covered topics had a decidedly practical flavor, as some leading experts in applied coding theory, with an extensive mathematical background, reported about their work and provided insights into the directions the field will be taking in the coming years. Some of the most striking applications of coding-theoretical techniques to practical problems that were discussed included explicit constructions of regenerating codes for distributed storage; using network coding techniques to increase throughput in content distribution; and the design of iterative decoders for LDPC codes with better error floors than the traditional belief propagation decoders.

Subject Area 2

Automata Theory and Circuits

2.1 Design of Reversible and Quantum Circuits

Authors: Kenichi Morita and Robert Wille

The development of computing machines found great success in the last decades. But the ongoing miniaturization of integrated circuits will reach its limits in the near future. Shrinking transistor sizes and power dissipation are the major barriers in the development of smaller and more powerful circuits. To further satisfy the needs for more computational power and further miniturization, alternatives are needed that go beyond the scope of conventional technologies like CMOS. Reversible logic and quantum logic provide a promising alternative that may enhance or even replace conventional circuits in the future. More precisely:

- **Low Power Computation**
  While conventional circuits dissipate energy for each lost bit of information, reversible circuits are information lossless, i.e. theoretically they are not affected by this. Considering the increasing miniaturization, this makes reversible logic interesting for domains like low-power design. Besides this general paradigm, reversible circuits are particularly suited for complementary low-power solutions like adiabatic circuits or on-chip interconnect encoders.
- **Quantum Computation**
  Quantum Computation offers the promise of more efficient computing for problems that are of exponential difficulty for conventional computing paradigms. Considering that many of the established quantum algorithms include a significant Boolean component (e.g. the oracle transformation in the Deutsch-Jozsa algorithm, the database in Grover's search algorithm, and the modulo exponentiation in Shor's algorithm), it is crucial to have efficient methods to synthesize quantum gate realisations of Boolean functions. Since any quantum operation inherently is reversible, reversible circuits can be exploited for this purpose.

However, no real design flow for these new kinds of circuits exists so far. Proposed approaches for synthesis, verification, and test are only applicable for very small circuits and systems. This is crucial since the design for reversible and quantum systems significantly differs from their conventional counterparts. Nearly all concepts and methods developed for conventional hardware design in the last decades have to be redeveloped in order to support the new technologies. Additionally, considering that today researchers are still faced with serious challenges for conventional technologies, it is worth working towards design solutions for reversible and quantum technologies already today.

The goal of the seminar was to bring together experts in order to present and to develop new ideas and concepts for the design of complex reversible and quantum circuits. In total,

17 presentations together with 1 tool demonstration (of the open source toolkit *RevKit*) and one panel session (on the different opinions on how reversible circuits can help to reduce power consumption during computation) have been conducted within the seminar. This has been accompanied by several working group and proposal preparation meetings. The most important topics which have been discussed were:

- **Design Methods**
  How to (automatically) synthesize reversible and quantum circuits as well as check them for correctness?
  Most of today's synthesis approaches for reversible and quantum circuits still rely on Boolean function descriptions like e.g. permutations, truth tables, or binary decision diagrams. In order to design complex circuits, higher levels of abstractions have to be considered. While for this purpose hardware description languages like VHDL, SystemC, or SystemVerilog have been established in conventional hardware design, high level synthesis of reversible and quantum circuits is just at the beginning. In order to advance this area, ideas about concepts, languages, and synthesis approaches for high level design have been presented and collected at the seminar.

- **Theoretical Consideration**
  How can theoretical studies show us the way for realizing reversible/quantum computers?
  In order to implement efficient reversible/quantum circuits and computers, we still need very basic and theoretical studies on them. This is because the paradigm of reversible/quantum computing has very different natures from that of conventional computing. Therefore, we shall still be able to find many novel and useful ideas for them through theoretical studies. In this seminar, theoretical consideration on various models in several levels have been presented and discussed. These models range from the element level to the software level, which include reversible logic elements and circuits, quantum automata, reversible Turing machines, and reversible programming languages.

- **Physical Realizations and Accuracy of Models**
  How to physically implement the respective circuits?
  How to close the gap between the theoretical models and the physical implementation?
  In order to design reversible and quantum circuits, abstractions of the precise physical realizations are applied. These include the used gate library and the respective cost metrics, but also fault models for testing or abstractions for technology mapping. Due to the progress in the development of physical realizations, these models constantly are subject to modifications which needed to be considered in the design phase. During the seminar, recent achievements in the development of physical realizations have been presented. This built the basis for discussions about updating and refining the applied models and abstractions.

- **Applications**
  How can reversible and quantum circuits be exploited in practically relevant application?
  How to measure and proof the benefits of these emerging technologies (e.g. how to substantiate improvements in the power consumption)?
  So far, design methods have mostly been applied to "academic" examples only. How-

ever, the design of reversible circuits for precise applications is the next logical step. Possible directions (e.g. in the low-power domain) have been discussed at the seminar. This also included discussions of the requirements for such applications and how the benefits can be measured.

Results of the seminar are currently used in the preparation of upcomming scientific papers. As one example, a special issue on the results triggered by the ideas of this seminar is planned for 2013. Furthermore, the discussions encouraged the preparation of proposals for national and international research projects.

Subject Area 3

Verification, Logic, Semantics

3.1 Computing with Infinite Data: Topological and Logical Foundations

Seminar No. **11411**
Organizers: Ulrich Berger, Vasco Brattka, Victor Selivanov, Dieter Spreen, and Hideki Tsuiki
10.–14. October, 2011 – www.dagstuhl.de/11411

Authors: Dieter Spreen

In safety-critical applications it is not sufficient to produce software that is only *tested* for correctness: its correctness ought to be proven formally. This remark also applies to the area of scientific computation. An important example are autopilot systems for aircrafts. The problem is that the current mainstream approach to numerical computing uses programming languages that do not possess sound mathematical semantics. Hence, there is no way to provide formal correctness proofs.

The reason is that on the theoretical side one deals with well-developed analytical theories based on the non-constructive concept of a real number. Implementations, on the other hand, use floating-point realizations of real numbers which do not have a well-studied mathematical structure. Approaches to tackle these problems are currently promoted under the slogan "Computing with Exact Real Numbers".

Well-developed practical and theoretical bases for exact real number computation and, more generally, computable analysis are provided by Scott's Domain Theory and Weihrauch's Type Two Theory of Effectivity (TTE). In both theories real numbers and similar ideal objects are represented by infinite streams of finite objects.

The seminar focused on two problem areas in the realm of computable analysis and computation on infinite streams:

1. Algorithms for stream transforming functions with particular emphasis on (i) logical and category-theoretic methods for the synthesis of provably correct programs, (ii) topological investigations of particular stream representations supporting efficient stream algorithms.
2. Hierarchies and reducibility relations between sets and functions of infinite data as a means of classification. Methods from topology, logic and descriptive set theory were of particular importance in this case.

Infinite streams are infinite words, so there is a close connection to the theory of ω-languages. To study these was a further aim of the seminar.

In the last years much interest in the (logical and topological) structure of the infinite words used to represent continuous data as well as how they code the data space has emerged.

U. Berger and others are developing a constructive theory of digital computation based on co-induction, and are applying it to computable analysis. The aim is to create a mathematical foundation for (lazy) algorithms on analytical data such as real numbers, real

functions, compact sets, etc. Since co-induction admits a particularly elegant formalization, this approach is well suited for computer aided modeling and proving in computable analysis. A concrete goal is to use program extraction from proofs as a practical method for obtaining certified programs in computable analysis.

The theory is based on a category of digit spaces. A typical object in this category is a compact metric space with a set of digits, where each digit is a contracting map. A point is then an infinite sequence of function (digit) compositions and hence represented by the corresponding infinite word. Moreover, for any given finite length, the words of that length over the alphabet of digits define a covering of the given space that allows to exactly locate the points represented by infinite words of digits. The set of uniformly continuous functions between such spaces can be characterized by a combined inductive/co-inductive definition that gives rise, via program extraction, to implementations of such functions as non-wellfounded (lazy) trees.

The theory of digit spaces combines known techniques for implementing and verifying stream processing functions (Edalat, Pattinson, Potts, and others) with ideas from co-induction and co-algebra (Jacobs, Rutten, Bertot, Niqui, and others).

H. Tsuiki is pursuing a similar programme, though, from a different perspective. In his case, e.g., every infinite object is represented by exactly one infinite word over $\{0, 1, \bot\}$, where $\bot$ represents "unknown". It turned out that the encoded space has at most topological dimension n, exactly if there are not more than n occurrences of $\bot$ in the corresponding infinite code words.

Besides the problem of how to represent continuous data in an "optimal" way, it is as well an important task to distinguish between computable and non-computable functions and, in the last case, to estimate the degree of non-computability. Most functions are non-computable since they are not even continuous. A somewhat easier and more principal task is in fact to understand the degree of discontinuity of functions. This is mostly achieved by defining appropriate hierarchies and reducibility relations.

In classical descriptive set theory, along with the well-known hierarchies, Wadge introduced and studied an important reducibility relation on Baire space. As shown by van Engelen et al., von Stein, Weihrauch and Hertling, this reducibility of subsets of topological spaces can be generalized in various ways to a reducibility of functions on a topological space. In this way, the degrees of discontinuity of several important computational problems were classified. It turned out that these classifications refine the so called "topological complexity" introduced in the alternative Blum-Shub-Smale approach to computability on the reals, which is used in complexity considerations in computational geometry.

Recently, reducibilities for functions on topological spaces have been used to identify computational relations between mathematical theorems. This programme, started by V. Brattka et al., can be considered as an alternative to reverse mathematics. Whereas reverse mathematics focuses on set existence axioms required to prove certain theorems, the approach pursued in computable analysis is to identify the computational power required to "compute" certain theorems. The approach, with contributions by M. de Brecht, G. Gherardi, A. Marcone, A. Pauly, M. Ziegler and V. Brattka, has led to deep new insights into computable analysis and has revealed close relations to reverse mathematics, but also some crucial differences.

Motivated initially by decidability problems for monadic second-order logic and Church's synthesis problem for switching circuits, researchers from automata theory (Büchi, Trakhtenbrot, Rabin, Wagner and many others) developed the theory of ω-languages which provides a foundation of specification, verification and synthesis of computing systems. Topological investigations in this theory have led to several hierarchies and reducibilities of languages of infinite words and trees. Instead of being just continuous, the reduction functions are now required to be computable by automata of suitable type. The resulting hierarchies (like, for example, the Wagner hierarchy) have sometimes the advantage that their levels are decidable.

Note that even the study of automata on finite words now involves topological methods in the classical form of profinite topology. Recently, deep relations of profinite topology to Stone and Priestly spaces were discovered by Pippenger and developed by Gehrke, Grigorieff and Pin. Again, suitable versions of the Wadge reducibility seem to play an important role in development of this field.

The seminar attracted 51 participants representing 15 countries and 5 continents, and working in fields such as computable analysis, descriptive set theory, exact real number computation, formal language theory, logic and topology, among them 10 young researchers working on their PhD or having just finished it. The atmosphere was very friendly, but discussions were most lively. During the breaks and until late into night, participants gathered in small groups for continuing discussions, communicating new results and exchanging ideas.

The seminar led to new research contacts and collaborations. The participants are invited to submit a full paper for a special issue of *Mathematical Structures in Computer Science*. At least one submission deals with a problem posted in the discussions following a talk.

The great success of the seminar is not only due to the participants, but also to the staff, both in Saarbrücken and Dagstuhl, who always do a great job in making everything run efficient and smoothly. Our thanks extend to both groups!

Subject Area 4

Geometry, Image Processing, Graphics

4.1 Computational Geometry

Seminar No. **11111**
Organizers: Pankaj Kumar Agarwal, Kurt Mehlhorn, and Monique Teillaud
13.–18. March, 2011 – www.dagstuhl.de/11111

Authors: Pankaj Kumar Agarwal, Kurt Mehlhorn, and Monique Teillaud

Computational Geometry and its Evolution

The field of computational geometry is concerned with the design, analysis, and implementation of algorithms for geometric problems, which arise in a wide range of areas, including computer graphics, CAD, robotics computer vision, image processing, spatial databases, GIS, molecular biology, and sensor networks. Since the mid 1980s, computational geometry has arisen as an independent field, with its own international conferences and journals.

In the early years mostly theoretical foundations of geometric algorithms were laid and fundamental research remains an important issue in the field. Meanwhile, as the field matured, researchers have started paying close attention to applications and implementations of geometric algorithms. Several software libraries for geometric computation (e.g. LEDA, CGAL, CORE) have been developed. Remarkably, this emphasis on applications and implementations has emerged from the originally theoretically oriented computational geometry community itself, so many researchers are concerned now with theoretical foundations as well as implementations.

Seminar Topics

The emphasis of the seminar was on presenting the recent developments in the field as well as identifying new challenges. We have identified a few broad topics, listed below, that cover both theoretical and practical issues in computational geometry and that we believe are some of the most interesting subareas in the field.

- *Theoretical foundations* of computational geometry lie in combinatorial geometry and its algorithmic aspects. They are of an enduring relevance for the field, particularly the design and the analysis of efficient algorithms require deep theoretical insights.
- Various *applications* such as robotics, GIS, or CAD lead to interesting variants of the *classical topics* originally investigated, including convex hulls, Voronoi diagrams and Delaunay triangulations, and geometric data structures. For example, Voronoi diagrams and nearest-neighbor data structures under various metrics have turned out to be useful for many applications and are being investigated intensively.
- Because of applications in molecular biology, computer vision, geometric databases, *shape analysis* has become an important topic. Not only it raises many interesting

geometric questions ranging from modeling and reconstruction of surfaces to shape similarity and classification, but it has also led to the emergence of the so-called field *computational topology*.

- In many applications the data lies in very high dimensional space and typical geometric algorithms suffer from the curse of dimensionality. This has led to extensive work on dimension-reduction and embedding techniques.
- Massive geometric data sets are being generated by networks of sensors at unprecedented spatial and temporal scale. How to store, analyze, query, and visualize them has raised several algorithmic challenges. New computational models have been proposed to meet these challenges, e.g., streaming model, communication-efficient algorithms, and maintaining geometric summaries.
- *Implementation issues* have become an integral part of the research in computational geometry. Besides general software design questions especially *robustness* of geometric algorithms is important. Several methods have been suggested and investigated to make geometric algorithms numerically robust while keeping them efficient, which lead to interaction with the field of computer algebra, numerical analysis, and topology.

Participants

53 researchers from various countries and continents attended the meeting. This high number shows the strong interest of the community for this event. The feedback from participants was very positive.

Dagstuhl seminars on computational geometry have been organized since 1990, lately in a two year rhythm. They have been extremely successful both in disseminating the knowledge and identifying new research thrusts. Many major results in computational geometry were first presented in Dagstuhl seminars, and interactions among the participants at these seminars have led to numerous new results in the field. These seminars have also played an important role in bringing researchers together and fostering collaboration. They have arguably been the most influential meetings in the field of computational geometry.

A session of this Seminar was dedicated to our dear friend Hazel Everett, deceased on July 20th, 2010.

The place itself is a great strength of the Seminar. Dagstuhl allows people to really meet and socialize, providing them with a wonderful atmosphere of a unique closed and pleasant environment, which is highly beneficial to interactions.

Therefore, we warmly thank the scientific, administrative and technical staff at Schloss Dagstuhl!

4.2 Innovations for Shape Analysis: Models and Algorithms

Seminar No. **11142**
Organizers: Michael Breuß, Alfred M. Bruckstein, and Petros Maragos
03.–08. April, 2011 – www.dagstuhl.de/11142

Authors: Michael Breuß, Alfred M. Bruckstein, and Petros Maragos

The notion of *shape* is fundamental in image processing and computer vision, as the shape of objects allows the semantic interpretation of image contents. This is also known from human vision, as humans can recognise characteristic objects solely from their shapes. By *shape analysis* one denotes models and algorithms for detection and processing of shapes in images. It is at the heart of a lot of applications in sciences and engineering.

While the visual quality of an image benefits from a large number of pixels and a high resolution of details, the same assertion does not generally hold for important shape information. As a simple example, when refining an image of $N = n \times n$ pixels towards a higher resolution, the number of pixels in the contour of an object grows just linearly with n, but the image size N grows quadratically. This *linear scaling property* of shape descriptors is an attractive feature of shape analysis methods when considering large images.

Thanks to technological progress made within the last decade one can nowadays acquire high-resolution images with relatively inexpensive, common devices. An important example are digital cameras that allow to acquire images of several megapixels. The size of the image files has grown accordingly. As the trend of developing even more accurate and inexpensive acquisition devices will certainly continue in the next years, the linear scaling property of shape descriptors makes shape analysis methods an even more useful tool in image processing than in the past. However, there are also substantial *new challenges* in shape analysis: Concerning algorithms that allow the processing of the arising large data sets in acceptable time, and with respect to adequate shape analysis models that allow for an efficient algorithmic formulation.

The purpose of this seminar was to meet these challenges by bringing together researchers that are engaged in recent and upcoming developments in shape analysis models and numerical computing. As an example, the field of differential geometry has grown to be important for shape analysis during the last years, while a field like deformable shape modelling just begins to influence shape analysis methods. On the algorithmical side, there are many recent innovations that can be important for shape analysis. As examples, let us mention new broadly applicable, efficient fast marching schemes, or graph-based iterative algorithms. The individual areas in shape analysis and numerical computing share an interest in the described techniques. However, modelling is seen as a hot topic in computer science, while numerical computing is often seen as a mathematical domain. Also the various areas within shape analysis research can benefit from the discussion of models and methods that are modern in their respective fields.

The purpose of bringing together researchers from different disciplines was to explore the benefits of a *cross-disciplinary* point of view.

- Researchers in continuous-scale shape analysis brought their knowledge of differential and variational models and the related methods to the meeting.
- Researchers in discrete shape analysis brought to the meeting their knowledge about the latest techniques in graph-based shape analysis, discrete topology and related optimisation methods.
- Researchers in numerical computing brought to the meeting their knowledge of numerical techniques and of numerical analysis.

As the demands in the individual fields of shape analysis are high, the research groups in which the most interesting techniques are under development are quite specialised. Because of this, there is no regular conference or workshop that serves as a meeting place for an exchange of ideas of these groups.

The seminar was conducted in a conference style, where every contributor gave a talk of about 20 to 25 minutes. There was much time for extensive discussions in between the talks and in the evenings, and as documented by the very positive evaluation there was generally a very open and constructive atmosphere. While it is at the moment this report is written very difficult to identify a new fundamental aspect of shape analysis as a result of the workshop, lots of interesting aspects were discussed. As we believe, these will inspire novel developments in both modelling and algorithms.

4.3 Geometric Modeling

Seminar No. **11211**
Organizers: Thomas Grandine, Stefanie Hahmann, Jörg Peters, and Wenping Wang
22.–27. May, 2011 – www.dagstuhl.de/11211

Authors: Thomas Grandine, Stefanie Hahmann, Jörg Peters, and Wenping Wang

The 8th Dagstuhl seminar was attended by 51 leading researchers coming from 4 continents and 21 countries (!). This high number shows the strong interest of the community for this event. The feedback from participants was very positive. A total of 45 presentations were grouped together into 12 sessions and let a lot of room for stimulating and fruitful discussions in the harmonic Dagstuhl atmosphere.

Geometric Modeling is the branch of Computer Science concerned with the acquisition, representation, modeling and analysis of 3-dimensional geometry. While its combination of technically complex and often interdisciplinary approaches is grounded both in applied mathematics and computer science data structures and theory, applications of the field therefore cover a wide collection of areas from classical product design, virtual prototyping and simulation to computer graphics, scientific visualization, medical imaging, multimedia and entertainment. It is therefore fitting that the seminar was attended by 7 leading scientists and engineers from industry.

The presentations ranged from surface reconstruction, GPU programming, to curve and surface modeling with classical splines, surface meshes and new subdivision methods based on algebraic and differential geometry methods with applications in medical or architectural modeling and in video and gaming industry. As with all previous Dagstuhl Seminars on Geometric Modeling, the conference proceedings will be published as a special issue in an international journal.

A special event during the conference was the John Gregory Memorial Award honoring Carl de Boor, Malcolm Sabin and Gershon Elber. This award is presented every three years at Dagstuhl and honors fundamental contributions to the field of geometric modeling.

The organizers thank all the attendees for their contributions and extend special thanks to the team of Schloss Dagstuhl for helping to make this workshop a success. As always, we enjoyed the warm atmosphere of the Schloss, which supports both formal presentations as well as informal exchanges of ideas.

4.4 Scientific Visualization

Seminar No. **11231**
Organizers: Min Chen, Hans Hagen, Charles D. Hansen, and Arie Kaufman
05.–10. June, 2011 – www.dagstuhl.de/11231

Authors: Min Chen, Hans Hagen, Charles D. Hansen, and Arie Kaufman

Scientific Visualization (SV) is the transformation of abstract data, derived from observation or simulation, into readily comprehensible images, and has proven to play an indispensable part of the scientific discovery process in many fields of contemporary science. This seminar focused on the general field where applications influence basic research questions on one hand while basic research drives applications on the other. Reflecting the heterogeneous structure of Scientific Visualization and the currently unsolved problems in the field, this seminar dealt with key research problems and their solutions in the following subfields of scientific visualization:

Biomedical Visualization: Biomedical visualization and imaging refers to the mechanisms and techniques utilized to create and display images of the human body, organs or their components for clinical or research purposes. Computational and algorithmic biomedical imaging is a wide area of research and solution development.The participants presented open problems and some solutions in this research area.

Integrated Multifield Visualization: The output of the majority of computational science and engineering simulations typically consists of a combination of variables, so called multifield data, involving a number of scalar fields, vector fields, or tensor fields. The state of the art in multifield visualization considerably lags behind that of multifield simulation. Novel solutions to multiscale and multifield visualization problems have the potential for a large impact on scientific endeavours and defining open problems and ideas in this subtopic was of keen interest to the seminar.

Uncertainty Visualization: Decision making, especially rapid decision making, typically happens under uncertain conditions. Challenges include the inherent difficulty in defining, characterizing, and controlling comparisons between different data sets and in part to the corresponding error and uncertainty in the experimental, simulation, and/or visualization processes. Refining and defining these challenges and presenting solutions was the focus for participants.

Scalable Visualization: The development of terascale, petascale, and soon to be exascale computing systems and of powerful new scientific instruments collecting vast amounts of data has created an unprecedented rate of growth of scientific data. Many solutions are possible such as trade-offs in speed vs quality, abstractions which provide scalability, novel parallel techniques, and the development of techniques for multivariate visual display and exploration.

However, scaling to the next generation (exascale) platforms may require completely rethinking the visualization workflow and methods. Defining how such architectures influence scientific visualization methods was addressed in this seminar.

4.5 Outdoor and Large-Scale Real-World Scene Analysis. 15th Workshop Theoretic Foundations of Computer Vision

Seminar No. **11261**
Organizers: Frank Dellaert, Jan-Michael Frahm, Marc Pollefeys, and Bodo Rosenhahn
26. June–01. July, 2011 – www.dagstuhl.de/11261

Authors: Frank Dellaert, Jan-Michael Frahm, Marc Pollefeys, Laura Leal-Taixé and Bodo Rosenhahn

The topic of the meeting was *Large-Scale Outdoor Scene Analysis*, which covers all aspects, applications and open problems regarding the performance or design of computer vision algorithms capable of working in outdoor setups and/or large-scale environments. Developing these methods is important for driver assistance, city modeling and reconstruction, virtual tourism, telepresence, and outdoor motion capture. With this meeting we aimed to attain several objectives, outlined below.

A first objective was to take stock of the performance of existing state-of-the-art computer vision algorithms and define metrics and benchmark data-sets on which to evaluate them. It is imperative that we push existing algorithms, which are currently benchmarked or tested with artificial or indoor set-ups, towards *real* applications. Methods of interest are 3D reconstruction, optic flow computation, motion capture, surveillance, object recognition, and tracking. These need to be dragged out of the lab and into the real world. Over the last years the computer vision community has recognized this problem and several groups are increasingly concentrating on the analysis of uncontrolled scenes. Examples include reconstructing large city models from online image collections such as Flickr, or human tracking and behavior recognition in TV footage or video from arbitrary outdoor scenes. An outcome we envision is the definition of appropriate metrics, benchmark sequences, and the definition of a *grand-challenge problem* that exposes algorithms to all the difficulties associated with large-scale outdoor scenes while simultaneously mobilizing the research community.

A second objective, then, was to define what the open problems are and which aspects of outdoor and large-scale scene analysis make the problem currently intractable. In uncontrolled, outdoor settings many problems start to arise, among them harsh viewing conditions, changing lighting conditions, artifacts from wind, rain, clouds or temperature etc. In addition, large-scale modeling, i.e. spanning city-scale areas, contains difficult challenges of data association and self-consistency that simply do not appear in smaller data-sets. Failure of basic building-block algorithms seems likely or even inevitable, requiring system-level approaches in order to be robust to failure. One of difficulties lies in the fact that the observer looses complete control over the scene, which can become arbitrary complex. This also brings with it the challenge to describe the scene in other than purely geometric terms, i.e., perform true scene *understanding* at multiple spatial and temporal

scales. Finally, outdoor scenes are dynamic and changing over time, requiring event learning and understanding as well as integrating behavior recognition. In this, we brought in participants from industry in order to ground the challenges discussed in real-world, useful applications.

The third and final objective was to discuss strategies that address these challenges, by bringing together a diverse set of international researchers with people interested in the applications, e.g. arising from photogrammetry, geoinformatics, driver assistance systems or human motion analysis. Though these people work in different fields and communities, they are unified by their goal of dealing with images and/or video from outdoor scenes and uncontrolled settings. In the workshop we allowed for an exchange of different modeling techniques and experiences researchers have collected. We allowed time for working groups during the workshop that connect people and whose goals are to develop ideas/roadmaps, additionally we allowed young researchers to connect with senior researchers, and in general allow for an exchange between researchers who would usually not meet otherwise.

The seminar schedule was characterised by flexibility, working groups and sufficient time for focused discussions. The participants of this seminar enjoyed the atmosphere and the services at Dagstuhl very much. The quality of this center is unique.

There will be an edited book (within Springer's series on LNCS) following the seminar, and all seminar participants have been invited to contribute with chapters. The deadline for those submissions is in November 2011 (allowing to incorporate results or ideas stimulated by the seminar), and submissions will be reviewed (as normal). Expected publication date is the end of 2012.

4.6 Efficient Algorithms for Global Optimisation Methods in Computer Vision

Seminar No. **11471**
Organizers: Andrés Bruhn, Thomas Pock, and Xue-Cheng Tai
20.–25. November, 2011 – www.dagstuhl.de/11471

Authors: Andrés Bruhn, Thomas Pock, and Xue-Cheng Tai

Most of the leading algorithms in computer vision are based on global optimisation methods. Such methods compute the solution of a given problem as minimiser of suitable cost functional that penalises deviations from previously made assumptions and integrates them over the entire image domain. While their transparent modelling allows for excellent results in terms of quality for many fundamental computer vision tasks such as *motion estimation*, *stereo reconstruction*, *image restoration*, *shape matching*, and *object segmentation*, the minimisation of these cost functional often leads to optimisation problems that are mathematically challenging and computationally expensive.

In the last decade, this fact has triggered a variety of different research directions that try to satisfy the needs of an ever increasing resolution in image and video data as well as the strong real-time demands of industrial applications. These research directions can be roughly divided into four fields that correspond to the different stages of the algorithmic design pipeline:

- *Modelling* (suitable priors, continuous vs. discrete, hybrid approaches)
- *Mathematical Analysis* (convex vs. non-convex, error bounds, well-posedness)
- *Numerical Solvers* (recent techniques, trends in numerical/combinatorial optimisation)
- *Parallelisation* (GPUs, mutli-core, cluster systems, FPGAs)

Since there are no conferences that address all four fields, the goal of this seminar was to identify and address open questions in these four fields in the context of the *entire* design pipeline. To this end, it brought together computer scientists and mathematicians from all stages. Apart from stimulating interdisciplinary research and establishing close collaborations between the different fields by scheduling plenty of time for discussions, the ultimate goal of the seminar was to develop more precise and more efficient algorithms that are conceptually well designed, mathematically well understood and from which all parts are chosen carefully such that they harmonise with each other.

Further aims of the seminar were to establish suitable benchmarks for measuring the performance of each of the stages (model precision, optimisation accuracy, numerical efficiency, parallelisability) and to derive general conceptual guidelines for the design of efficient algorithms that are applicable to a broad class of key problems in computer vision.

The seminar was conducted in a conference style, where every participant gave a talk of about 20 to 25 minutes. There was much time for extensive discussions – directly after

the talks, in dedicated working groups and in the evenings. As documented by the very positive evaluation and the detailed summaries of the working groups, there was a very open and constructive atmosphere. In particular the people appreciated the integration of young researchers in the seminar schedule and the gain of new cross-disciplinary insights from talks and discussions of other participants. While the first observation reflects the fact that the field of efficient algorithms in computer vision is relatively new, the second aspect demonstrates the success of the seminar to bring together people from different communities and establish new ties between the fields. At the moment that this report is written it is very difficult to identify new fundamental issues of efficient algorithms in computer vision. However, lots of interesting aspects were discussed during the talks and in the discussion groups, and many participants established collaborations with people from other fields. Thus we believe this seminar served as an excellent basis to inspire and trigger novel developments in the design of efficient algorithms for global optimisation problems in computer vision.

Finally, it should be mentioned that there was a huge consensus among the participants that there should be a follow-up event to this seminar in the upcoming years. For the current seminar, there will be edited post-proceedings in the Springer LNCS series which gives all participants the opportunity to summarise results from the seminar, discuss open questions and present recent research in the field. The deadline for submission is end of March 2012.

4.7 Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data

Seminar No. **11501**
Organizers: Bernhard Burgeth, Anna Vilanova Bartroli, and Carl-Fredrik Westin
11.–16. December, 2011 – www.dagstuhl.de/11501

Authors: Carl-Fredrik Westin and Bernhard Burgeth

Higher Order Descriptors in Medical Imaging and Engineering

This seminar is the 4th in a series of Dagstuhl Seminars devoted to the visualization and processing of tensor fields and higher order descriptors. Tensor fields play an important role in many different scientific disciplines and application domains such as medical imaging, image processing, fluid dynamics, and structural mechanics. Analysis and visualization of multi-valued data have gained significant importance in scientific visualization and image processing due to rapid advances in medical imaging and and in the engineering industry.

In medical imaging, multi-valued data include diffusion weighted magnetic resonance imaging (dMRI), a medical imaging modality that allows the measurement of water diffusion in tissue *in vivo*. These measurements allow the description of diffusion in living fibrous tissue (e.g., white matter or muscle). The diffusion can be described by a diffusion tensor (i.e., a positive semidefinite 3×3 matrix). It is customary to acquire more complex data than can be described by the tensor model and recently the analysis has been extended to higher-order descriptors (i.e., higher-order tensors or spherical harmonics). There are several open questions how to best analyze and visualize such data.

In addition to tensor data from the medical field, a number of scientific and engineering applications produce tensor fields as results of simulations of physical phenomena. The tensor concept is essential for the physical description of anisotropic behavior, especially in solid mechanics and civil engineering (e.g. stress-strain relationships, inertia tensors, permittivity tensor). The field of engineering faces many open problems in tensor field visualization and processing and novel technology is needed to address these problems.

Seminar Topics and Breakout Sessions

The emphasis of the seminar was on presenting the recent developments in the multidisciplinary field as well as identifying new challenges. We discussed a broad set of topics and challenges that cover both theoretical and practical issues related to analyzing and visualizing fields of tensors and higher order descriptors. During the workshop we discussed

- Higher-order models in dMRI beyond the diffusion tensor
- Higher-order tensors in image processing

- Computational analysis and visualization of airflow dynamics
- Novel differential geometric approaches to brain connectivity from dMRI
- Connectivity concepts in mathematical morphology for tensor fields
- Tensor concepts in structural mechanics and material science
- Visualization of uncertainty
- dMRI in brain studies for clinical applications

This year we scheduled time for breakout sessions that would foster focused discussions in smaller groups. During the first day of the meeting the group defined a list of important topics and open questions. Three of those were chosen and defined the breakout sessions:

- How do we define a suitable Finsler metric from diffusion data?
- How do we define biologically meaningful metrics from diffusion tensor and higher order model diffusion data?
- What are important questions in engineering that can be answered with visualization?

These breakout sessions turned out to be very successful and the groups scheduled extra time in the evenings for continued discussions. The format of the new breakout sessions fits very well in the Dagstuhl environment promoting discussions and interactions. If we get the chance to organize another meeting at Dagstuhl, the breakout sessions will definitely be a part of the schedule.

Outcomes

The participants all agreed that the meeting was successful and stimulating. Seminar participants are already collaborating on a Springer book summarizing the results of the meeting. The Springer book will have about twenty chapters authored by the meeting participants, and we expect the book to be published in early 2013. The participants expressed interest in documenting the discussions in the breakout session in book chapters, in addition to the science described in their regular presentation.

The environment at Schloss Dagstuhl has generated several new scientific collaborations. The work in the engineering breakout session has resulted in a new project of four participants (Stommel, Burgeth, Scheuermann, Hotz) and a submission of a grant proposal to the Landesforschungsförderprogramm (LFFP) des Saarlandes. Meanwhile, the application for the grant has been approved. Three seminar participants who met at the meeting (O'Donnell, Hui Zhang, Schultz), from the USA, Great Britain, and Germany, are collaboratively organizing a workshop on computational diffusion MRI at the conference for Medical Image Computing and Computer-Assisted Intervention 2012.

It was voted that the group will apply for another meeting in this series, and that in addition to the current organizers (Carl-Fredrik Westin, Bernhard Burgeth, Anna Vilanova Bartroli), add Dr. Ingrid Hotz (ZIB – Berlin) as an organizer of the next event.

Subject Area 5

Artificial Intelligence, Computer Linguistic

5.1 Reasoning about Interaction: From Game Theory to Logic and Back

Seminar No. **11101**
Organizers: Jürgen Dix, Wojtek Jamroga, and Dov Samet
06.–11. March, 2011 – www.dagstuhl.de/11101

Authors: Jürgen Dix, Wojtek Jamroga, Dov Samet

The group (48 participants from 13 countries) convened in Dagstuhl in March 2011, for a five day meeting.

The aim of the seminar was to bring together researchers who approach interaction-related phenomena from different perspectives (and with different conceptual tools). We also wanted to identify potentials for coordination, and to discuss general models and methodologies for future research.

Of particular importance was the choice of the participants and the areas they working in, namely: (1) classical game theory, (2) mathematical logic, and (3) economics. While there are some relations between these areas, we felt that more work should be done on the overlapping parts to make tools and methods from one area available in the others (if possible).

In particular, we wanted to find answers to the following questions:

- Are existing models of interaction adequate? Can models used by different disciplines be integrated in a meaningful way?
- How can we use game-theoretical concepts to construct logics that support strategic reasoning? What are the necessary features of such logics?
- How can epistemic-logic reasoning and definitions lead to the definitions of new solution concepts is strategic-form games?
- How can epistemic and strategic logic be adapted to the empirical findings from game theoretic laboratory experiments, manifesting bounded rationality of a variety of types?
- How can issues of computational complexity be addressed vis-a-vis the demand for efficiency/optimality in the design of economic mechanisms under asymmetric information?

The seminar resulted in making the first step towards answering these questions. We did not obtain the ultimate formal answers, especially in the sense of enabling implementation in the form of ready-to-use tools and methodologies. However, researchers with different background shared their views on how games and multi-agent systems can be modeled and reasoned about, which led to several discussions on fundamental questions (like: *what features/concepts are indispensable when analyzing interaction between agents?*). In particular, the issue of whether probabilities (and, more generally: quantities) are necessary to give good account on how agents interact was hotly debated.

The results of the seminar were somewhat constrained by the unbalanced composition of participants. We had invited equally many researchers from computer science (especially computational logic) and economics (game theory). However, while most computer scientists accepted our invitation, the same did only a few economists. This is probably due to the fact that Dagstuhl seminars have an extremely high reputation within computer science, but they are relatively unknown in other disciplines. In consequence, the synergy between different views of interaction occurred only partially. In our opinion, it was especially fruitful on the basic level. That is, economists and computer science logicians learned about the basic models and patterns of analysis used in the other discipline. Even more importantly, they exchanged views on what *research questions* are relevant and viable when analyzing game-like interaction. Most synergy occurred within the subgroup of participants coming from the community of *modal logic in computer science*. Talks on modal logic-related topic triggered intensive discussion and ideas for joint research which are currently being pursued by several participants.

We thank the Dagstuhl staff for a very fruitful and interesting week. We are planning a special issue (in Annals of Math and AI) as a concrete outcome of the seminar. Moreover, it was a general consensus that a follow-up seminar would be highly interesting – this time with more specific topics being the focus. The follow-up is currently in the planning phase.

5.2 Exploration and Curiosity in Robot Learning and Inference

Seminar No. **11131**
Organizers: Peter Dayan, Ales Leonardis, Jan Peters, and Jeremy L. Wyatt
27. March – www.dagstuhl.de/11131

Authors: Jeremy L. Wyatt, Peter Dayan, Ales Leonardis, Jan Peters

Aims and background to the seminar

This seminar was concerned with answering the question: *how should a robot choose its actions and experiences so as to maximise the effectiveness of its learning?*.

This seminar was predicated on the assumption that to make significant progress in autonomous robotics, systems level theories of how robots should operate will be required. In recent years methods from machine learning have been employed with great success in robotics, in fields as diverse as visual processing, map building, motor control and manipulation. The machine learning algorithms applied to these problems have included statistical machine learning approaches, such as EM algorithms and density estimation, as well as dimensionality reduction, reinforcement learning, inductive logic programming, and other supervised learning approaches such as locally weighted regression. Most of these robot learning solutions currently require a good deal of supervised learning, or structuring of the learning data for a specific task. As robots become more autonomous these learning algorithms will have to be embedded in algorithms which choose the robot's next learning experience. The problems become particularly challenging in the context of robotics, and even worse for a robot that is faced with many learning opportunities. A robot that can perform both manipulation, language use, visual learning, and mapping may have several quite different learning opportunities facing it at any one time. How should a robot control its curiosity in a principled way in the face of such a variety of choices? How should it choose data on the basis of how much it knows, or on how surprising it finds certain observations? What rational basis should robot designers choose to guide the robot's choice of experiences to learn from? There has been initial progress in several fields, including *machine learning*, *robotics* and also in *computational neuroscience*. In this seminar we brought together these three communities to shed light on the problem of how a robot should select data to learn from, how it should explore its environment, and how it should control curiosity when faced with many learning opportunities.

Machine Learning   This problem of how to explore is one that has been studied both in the context of reinforcement learning (exploration vs. exploitation) and supervised learning (active learning). Within the former the language of the sequential decision making

community is that of MDPs and POMDPs. These are related to Bayesian perspectives on supervised Machine Learning in that we can think of a posterior over hypotheses that results from data that may be seen. In statistics this is related to the idea of conducting pre-posterior analysis. However, the theories from sequential decision making and active learning are currently unintegrated and partial, and it is not clear how they should apply in robotics. The current machine learning preference for Bayesian methods suggests that the ways that model uncertainty can be captured and exploited will be critical. During the seminar we looked for suggestions from this community as to how problems of exploration and curiosity in robotics can be formalised, especially at a systems level.

Robotics Within areas like SLAM (Simultaneous Localisation and Mapping) the problem of how to select data has been addressed, but heuristic measures of exploratory worth are typically employed. Again, the principal formalism is that of Bayesian filtering, within which a POMDP is posed, but typically only the belief filter part is used. Rather than look ahead at all possibilities, heuristics such as information gain are used. There are also other approaches necessary in areas where for example robot learners are learning associations between data from multiple modalities, from time series, and where there maybe limited intervention from humans. Again, learning approaches in contemporary robotics are typically statistical, but there are other approaches. Techniques are also adapted to the domain, such as in the community working on the use of robotics in scientific discovery in the laboratory, where the robot has the ability to determine which experiments to perform, and the methods used employ a great deal of structure and prior knowledge about the domain. There are also challenges for exploration and curiosity from the use of robots for scientific exploration, such as in planetary missions, and in subsea exploration. We looked at how these problems are currently posed, and the challenges they pose for machine learning approaches to data selection.

Animal Cognition and Computational Neuroscience The field of computational neuroscience has many insights to offer roboticists. Animals are forced to consider at each moment how they select data. There are examples of this in studies of motor learning, animal foraging, studies of neurotransmitters, and particular learning circuits, as well as in the study of areas of the brain concerned with action selection. Computational neuroscience has also been strongly influenced by statistical, particularly Bayesian approaches to inference and learning. For example much recent work strongly suggests a Bayesian underpinning to learning in the motor system, and other work has investigated possible neural bases for learning in the face of suprise and uncertainty. Work on reinforcement learning has linked with studies of brain areas such as the Basal ganglia, and there is debate as to whether or not the purpose of certain neurons is to provide a cue for learning in the face of novelty. This is related to the idea of infotaxis as a general mechanism for exploration control in some animal. The connection to the statistical theories from machine learning and optimal control are intriguing. This gives us a strong basis for the hope that a common framework for exploration and curiosity might emerge as a consequence of this seminar.

Summary of objectives   In summary the objectives of this seminar were to:

- Identify the different formulations of exploration and curiosity control, and to categorise robot problems into appropriate classes.
- Share statistical and non-statistical representations suitable for control of curiosity and exploration across communities.
- Identify the links between studies of learning control and motivation in computational neuroscience and formalisations from robotics and machine learning.
- Discuss possible formalisations of the problem of learning one of many possible tasks.
- Identify whether solution classes are heuristic or optimal.

Summary of the seminar program

The seminar was grouped into three themes, roughly according to Marr's levels of description: computational, algorithmic and implementational. Many talks crossed more than one level, but within these themes we were able to organize talks around more specific research areas. These areas were:

1. Ideas from neuroscience about the implementation of exploration and action in the brain.
2. Evidence from the ethology and psychology about the requirements for exploration, and algorithmic frameworks that fit the data on human behaviour.
3. Computational frameworks for intrinsic motivation and the evolution of extrinsic reward functions.
4. Algorithms and properties for specific sub-problems within curiosity and exploration: such as visual object search or the behaviour of greedy algorithms for solving sub-modular problems.
5. Robot implementations of algorithms for control of exploration and curiosity in real tasks.

Summary of the fundamental results

The main findings presented can be grouped into four parts. It is worth stating from the outset that a very large number of the talks, though by no means all, employed a reward based framework. It is not possible in this summary to mention all of the thirty talks given, instead we mention talks that illustrate the common themes of the seminar.

First a tutorial on the Basal Ganglia and it's role in action selection, including for exploration was given by Humphries. In this field there are now a range of computational models that simulate some of the internal workings of the Basal Ganglia. It was clear, however, that there are numerous structures about which little is known, and that many details of the models remain to filled in, or to be tested. Dayan provided evidence that was broadly negative with respect to a Bayesian view of exploration in humans. Dayan showed that human behaviour in a non-stationary bandit task is not better explained by a Bayesian view than by a simple soft-max reinforcement learning model. Sloman argued that the

requirements to support exploration include the need to decompose domains into reuseable patterns. Contrary to a large number of the speakers Sloman argued against a statistical approach to exploration control. In the workshop as a whole statistical methods, with rewards, and often Bayesian inference were dominant, but these talks from biology present evidence that was not always supportive of this dominant approach.

Several different frameworks for intrinsic motivation were given. In several of these (Schmidhuber, Polani, Auer) the idea that exploration is driven by curiosity to enable greater understanding and ability to exploit the environment was central. These approaches can be contrasted with those that are ultimately driven by the need to maximise extrinsic rewards (Tishby, Starzyk). There seemed little question that all of these frameworks are quite general, but no clear unifying account is available. Others (Uchibe, Barto, Elfwing) showed ways to evolve extrinsic reward functions that ultimately contribute to overall agent fitness. Overall the division seems to be between information seeking and value seeking frameworks for self-motivation.

In algorithms the important findings concerned cases where problems that in the general case are intractable can be tackled much more effectively in special cases. Tsotsos showed as part of his talk that some visual search problems are tractable even though in the general case they are not. Krause showed that where problems have a sub-modular property that greedy algorithms can be close to optimal. Dearden showed how for a particular search task that entropic heuristics perform close to the level of more computational expensive information lookahead methods. While the most general algorithmic frameworks to exploration are based on the solution of POMDPs, each of these talks showed that a solution to a simpler problem can often provide very good performance.

In robotic tasks some approaches were necessarily more pragmatic, and this meant that many moved away from a purely reward based framework. Several showed ways of approximating solutions to POMDPs in real robot systems. These included using hierarchical approaches, sampling methods, limited horizon lookahead, or methods that split the problem into parts with, and without state uncertainty (Wyatt, Peters, Martinez-Cantin). While some advocated implementation of the principled frameworks for intrinsic motivation (Pape), problems were often moved away from the common statistical, reward based framework to enable solutions. The benefit of heuristic goal selection methods on top of precise planning approaches to achieving selected goals was demonstrated (Hanheide, Skocaj). A variety of robotic tasks were shown to be tackled with active learning, including motor control and social learning (Peters, Lopes).

The main themes that emerged were that while the dominant paradigm was one that was statistical and reward based, there were alternatives. While there were theoretically rigorous frameworks based on rewards, these were actually not much used by roboticists, who preferred pragmatic approaches. In the middle sit those exploring algorithms that while still approximate, offer some performance bounds relative to that which is optimal, howsoever defined.

5.3 Plan Recognition

Authors: Robert P. Goldman, Christopher W. Geib, Tamim Asfour, and Henry Kautz

Plan recognition, activity recognition, and intent recognition all involve making inferences about other actors from observations of their behavior, i.e., their interaction with the environment and with each other. The observed actors may be software agents, robots, or humans. This synergistic area of research combines and unifies techniques from user modeling, machine vision, intelligent user interfaces, human/computer interaction, autonomous and multi-agent systems, natural language understanding, and machine learning. It plays a crucial role in a wide variety of applications including:

- assistive technology
- software assistants
- computer and network security
- behavior recognition
- coordination in robots and software agents
- e-commerce and collaborative filtering

This Dagstuhl seminar brought together researchers with a wide range of interests and backgrounds related to plan and activity recognition. It featured a substantial set of longer tutorials on aspects of plan and activity recognition, and related topics and useful methods, as a way of establishing a common vocabulary and shared basis of understanding. These were:

- Plan recognition and discourse;
- Plan recognition and psychology;
- Probabilistic methods;
- Plan recognition and learning;
- Grammatical methods and
- Planning and plan recognition.

The common ground constructed by these tutorials provided a basis that individual researchers could build upon when sharing their specific interests and developments.

One challenge to progress in plan recognition is that there has not been a shared agreement about what constitutes plan recognition: what are its inputs and outputs, and what constitutes a good answer. In particular, this has inhibited progress because it is difficult to clearly compare new work in plan recognition with preceding work (quantitative comparisons are almost impossible), there is a paucity of shared data sets, etc. Coming into the seminar, the organizing committee proposed that the field might be improved by the introduction of

a plan recognition competition, modeled on competitions in AI planning (the International Planning Competition), SAT solving, etc. Discussions at the seminar concluded that it would be premature to introduce such a competition at this time. Participants felt that a more productive use of community resources would be to develop a shared repository of plan and activity recognition data sets. A number of participants volunteered to provide their data sets, and there has been movement towards establishing a common public repository.

Plan Recognition: background

The earliest work in plan recognition was rule-based; researchers attempted to come up with inference rules that would capture the nature of plan recognition. However without an underlying formal model these rule sets are difficult to maintain and do not scale well.

In 1986, Kautz and Allen (K&A) published an article, "Generalized Plan Recognition" [7] that framed much of the work in plan recognition to date. K&A defined the problem of plan recognition as the problem of identifying a minimal set of *top-level actions* sufficient to explain the set of observed actions. Plans were represented in a plan graph, with top-level actions as root nodes and expansions of these actions into unordered sets of child actions representing plan decomposition. To a first approximation, the problem of plan recognition was then a problem of graph covering. K&A formalized this view of plan recognition in terms of McCarthy's circumscription. Kautz [6] presented an approximate implementation of this approach that recast the problem as one of computing vertex covers of the plan graph.

A number of early plan recognition systems used techniques such as rule-based systems [9], vertex covering, etc. Such techniques are not able to take into account differences in the *a priori* likelihood of different goals. Observing an agent going to the airport, this algorithm views "air travel," and "terrorist attack" as equally likely explanations, since they explain (cover) the observations equally well.

To the best of our knowledge, Charniak was the first to argue that plan recognition was best understood as a specific case of the general problem of *abduction*, or reasoning to the best explanation [3, e.g.,]. Charniak and Goldman (C&G) [2] argued that, viewing plan recognition as abduction, it could best be done as Bayesian (probabilistic) inference. Bayesian inference supports the preference for minimal explanations, in the case of equally likely hypotheses, but also correctly handles explanations of the same complexity but different likelihoods. For example, if a set of observations could be equally well explained by two hypotheses, theft and bragging being one, and theft alone being the other, simple probability theory (with some minor assumptions), will tell us that the simpler hypothesis is the more likely one. On the other hand, if as above, the two hypotheses were "air travel" and "terrorist attack," and each explained the observations equally well, then the prior probabilities will dominate, and air travel will be seen to be the most likely explanation. There have been many similar approaches to the problem, based on cost minimization, etc.

Another broad area of attack on the problem of plan recognition has been to reformulate it as a parsing problem [10, e.g.,]. Parsing-based approaches to plan recognition promise greater efficiency than other approaches, but at the cost of making strong assumptions about

the ordering of plan steps. The major problem with parsing as a model of plan recognition is that it does not treat partially-ordered plans or interleaved plans well. Approaches that use statisical parsing [11, e.g.,] combine parsing and Bayesian approaches.

Finally, there has been a large amount of very promising work done using variations of Hidden Markov Models (HMMs) [1], techniques that came to prominence in signal processing applications, including speech recognition. These approaches offer many of the efficiency advantages of parsing approaches, but with the additional advantages of incorporating likelihood information and of supporting machine learning to automatically acquire their plan models. Standard HMMs seem to be insufficiently expressive to capture planful behavior, but a number of researchers have extended them to hierarchical formulations, that capture more complicated intentions. Conditional Random Fields [8], dynamic Bayes nets, and other probabilistic models have also been used.

Much of this latter work has been done under the rubric of *activity recognition*. The early work in this area very carefully chose the term *activity* or *behavior recognition* to distinguish it from plan recognition. The distinction to be made between activity recognition and plan recognition is the difference between recognizing a single (possibly complex) activity and recognizing the relationships between a set of such activities that result in a complete plan. Much of the work on activity recognition can be seen as discretizing a sequence of possibly noisy and intermittent low-level sensor readings into coherent actions that could be treated as inputs to a plan recognition system.

Several researchers have been interested in using plan recognition to improve team co-ordination [4, 5]. That is, if agents in a team can recognize what their teammates are doing, then they can better cooperate and coordinate. They may also be able to learn something about their shared environment. For example, a member of a military squad who sees a teammate ducking for cover may infer that there is a threat, so that it also takes precautions.

References

[1] H. Bui and S. Venkatesh and G. A. W. West. Policy recognition in the abstract Hidden Markov Model. In *Journal of Artificial Intelligence Research*, vol. 17, pages 451–499, 2002.

[2] Eugene Charniak and Robert P. Goldman. A Bayesian model of plan recognition. *Artificial Intelligence*, 64(1):53–79, November 1993.

[3] Eugene Charniak and Drew McDermott. *Introduction to Artificial Intelligence*. Addison Wesley, Reading, MA, 1985.

[4] Marcus J. Huber, Edmund H. Durfee and Michael P. Wellman. The Automated mapping of plans for plan recognition. In *Proceedings of the National Conference on Artificial Intelligence*, pages 344–351, 1994.

[5] Gal A. Kaminka, D.V. Pynadath, and M. Tambe. Monitoring teams by overhearing: a multi-agent plan-recognition approach. In *Journal of Artificial Intelligence Research*, vol. 17, pages 83–135, 2002.

[6] Henry Kautz. *A Formal Theory of Plan Recognition*. Technical Report, Department of Computer Science, University of Rochester, May 1987.

[7] Henry Kautz and James F. Allen. Generalized plan recognition. In *Proceedings of the National Conference on Artificial Intelligence*, pages 32–38, 1986.

[8] Lin Liao, Dieter Fox and Henry Kautz. Learning and inferring transportation routines. In *Proceedings of the National Conference on Artificial Intelligence*, 2004.

[9] C.F. Schmidt and N.S. Sridharan and J.L. Goodson. The plan recognition problem: an intersection of psychology and artificial intelligence. In *Artificial Intelligence*, vol. 11, pages 45–83, 1978.

[10] Marc Vilain. Getting serious about parsing plans: a grammatical analysis of plan recognition. In *Proceedings of the National Conference on Artificial Intelligence (AAAI-90)*, pages 190–197, 1990.

[11] David V. Pynadath and Michael P. Wellman. Accounting for context in plan recognition with application to traffic monitoring. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pages 472–481, 1995.

5.4 Organic Computing – Design of Self-Organizing Systems

Seminar No. **11181**
Organizers: Kirstie Bellman, Andreas Herkersdorf, and Michael G. Hinchey
1.–6. May, 2011 – www.dagstuhl.de/11181

Authors: Kirstie Bellman, Andreas Herkersdorf, Mike Hinchey, Christian Müller-Schloer, Hartmut Schmeck, Rolf Würtz

Organic Computing (OC) has become a challenging vision for the design of future information processing systems: As the systems become increasingly more complex, powerful, cheaper and smaller, our environment will be filled with collections of autonomous systems. Autonomous systems are equipped with intelligent sensors and actuators to be aware of their environment, to communicate, and to organize themselves in order to perform the required actions and services. However, these abilities result in even greater system complexity, which we will not be able to explicitly design and manage in every detail, nor are we able to anticipate every possible configuration. Nevertheless, these services have to be as robust, safe, flexible, and trustworthy as possible. In particular, a strong orientation of these systems towards human needs – as opposed to a pure implementation of the technologically possible – is absolutely central.

So far, the OC community, mainly driven by the priority research program of the German Research Foundation (DFG), successfully proposed and – at least partially – established a common nomenclature and terminology for terms like emergence, self-organization, selfadaptation, robustness and flexibility within an interdisciplinary research community.

Quantitative metrics for emergence and self-organization were introduced and applied. Observer controller structures have been established as a common architectural pattern for OC systems within a wide spectrum of applications ranging from traffic control, to Systems on Chip, to collaborative robot systems, to wireless sensor networks. Roles and applicability of different types of supervised and reinforcement-based technical learning techniques were investigated and adapted to OC needs.

Despite the progress in understanding the implications and exploiting the potentials of the OC paradigm, a number of key challenges and research questions still remain. In particular, the planned 2011 OC seminar shall shed light on the various notions of design within the OC context. Design in the classical sense follows a hierarchical top-down constraint propagation starting from a purely functional specification. All eventual environmental influences and disturbances have to be anticipated by the designer at "design time". Due to this anticipatory nature the resulting system is rigid and not able to sufficiently react to run time events.

Complex systems in nature often develop bottom-up due to the self-organizing capabilities of their components. Each component and the system as a whole react to the demands of the environment. In doing so, they are guided by the principles to survive as an individual

(selfishness) and the necessity to co-operate (altruism). In technical life-like OC systems we must provide some control by a higher-level entity (finally the user) guiding the bottom-up decisions of the components into a globally desirable direction.

In this way, the former top-down design process dissolves into a balanced run-time negotiation between top-down constraints and bottom-up opportunities. The ultimate consequence of this would mean a total replacement of the design process (at design-time) to controlled self-organization (at runtime).

The 2011 OC seminar was held to answer questions resulting from this shift from design-time to run-time. Is OC a realistic or even desirable vision? How can we replace rigid human designtime control by self-adaptive run-time control without stifling the creativity of the emergent bottom up processes? How can we balance top-down control and bottom-up emergence? Beyond these theoretical questions it is a goal of the seminar to define a number of concrete OC demonstrators – or even a common demonstrator – to be pursued in the sequel.

5.5 Computer Science & Problem Solving: New Foundations

Authors: Iris van Rooij, Yll Haxhimusa, Zygmunt Pizlo, and Georg Gottlob

This Dagstuhl seminar brought together a group of computer scientists and psychologists to discuss their perspectives on problem solving. The seminar was inspired by two previous Problem Solving workshops in 2005 and 2008 at Purdue University, USA. These workshops, organized primarily by psychologists, laid bare some fundamental theoretical questions in problem solving research. The organizers believed that research on these questions could benefit from more involvement of computer scientists in the area of problem solving. This motivated the organization on this seminar, which aimed to bring together computer scientists and psychologists to help build new formal foundations for problem solving research.

Of the 36 participants at the seminar about half were computer scientists and the other half were psychologists, though many identified as interdisciplinary researchers (e.g., cognitive scientists). To facilitate cross-disciplinary perspectives, computer science and psychology talks were alternated in the program of the seminar. There were 7 longer featured talks and 15 shorter talks.

On Day 1 of the seminar, Iris van Rooij opened the seminar by explaining its history and motivation. She discussed how computational complexity theory gives a formal framework for quantifying the difficulty of solving different types of search problems (e.g., the Traveling Salesman Problem and Minimum Spanning Tree), but that no analogous formal framework exists yet for quantifying the difficulty of solving so-called 'insight' problems (e.g, the Nine-dot problem), or more generally, quantifying the difficulty of representing a problem in the right way. The question of how one could develop such a formal framework was an overarching theme of the seminar. All participants were invited to think about this question. The topic resurfaced in several talks and workgroup discussions.

The featured talks on Day 1 were by Todd Wareham and Bill Batchelder. Wareham presented novel ideas on how a formal theory of 'insight difficulty' may take shape. His analysis was based on existing ideas in the psychology literature, such as the Representational Change Theory of Knoblich and et al., and his formalisms were inspired by Gentner's Structure-Mapping Theory. Batchelder presented a list of 19 classic examples of insight problems. These problems served as illustrations of problems that are not 'search problems' in the sense of Newell and Simon, yet for which problem solving researchers should nevertheless like to be able to model and explain the processes involved. Four shorter talks on Day 1 were given by Georg Gottlob, Sarah Carruthers, Sashank Varma and Jakub Szymanik. Gottlob introduced conceptual tools from computational complexity theory,

graph theory and probabilistic computation that could inspire new ways of thinking about problem solving. Carruthers presented novel experimental data on how humans solve the graph problem Vertex Cover. Varma presented a methodology for modeling the resource requirements of different brain areas invoked during problem solving. Szymanik presented a generalization of the Muddy Children Problem and explained how its solution can be modeled using logic.

On Day 2 of the seminar there were 3 featured talks. In the first featured talk, Niels Taatgen presented the ACT-R modeling architecture and illustrated how it could model the development of more general problem solving skills as a re-combination of more basic skills. Rina Dechter and Ken Forbus each gave a different AI perspective on problem solving in their featured talks. Dechter presented several sophisticated algorithmic techniques for solving NP-hard problems, such as Constraint Satisfaction and Bayesian Inference. Forbus proposed to consider 'analogy' as a new foundation for problem solving research and illustrated his perspective using the Companion framework. There were 3 short talks on this day. The two short talks by Johan Kwisthout and Marco Ragni (like Wareham's talk on Day 1) touched clearly on the theme of the seminar. Kwisthout proposed a formal framework for capturing the notion of 'relevance' when it comes to finding a suitable problem representation, and Ragni proposed a framework for quantifying the *a priori* difficulty of problem items on an IQ test based the notion of 'representational transformation'. In the third short talk, Jelle van Dijk gave a designers' perspective on problem solving. Van Dijk made the case that much real-world problem solving is probably best studied from an embodied embedded cognitive perspective. The official program for this day was closed with a working group discussion on meanings of common terms used throughout the talks.

Day 3 opened with a featured talk by Dedre Gentner. The talk by Gentner complemented the talk by Forbus on Day 2 as she laid out the experimental evidence for the idea that analogical thinking (comparison and matching) lies at the foundation of human learning and reasoning. The featured talk was followed by two short talks, one by Liane Gabora and one by Daniel Reichman. Gabora presented a perspective on problem solving that is quite unlike the traditional view of problem solving as search through a well-defined space for a well-defined solution. Her perspective is that (creative) problem solving can perhaps best be seen as the recognition and actualization of a solution that before only existed in a state of potentiality. Reichman presented a theoretical computer science perspective on the well-known phenomenon of speed-accuracy tradeoffs in psychology. He proposed that algorithmic techniques from computer science can help predict what shape curves describing speed-accuracy tradeoffs will have in a variety of experimental conditions. In the afternoon of Day 3 there was no official program, and instead participants enjoyed the surroundings of Schloss Dagstuhl and/or went for a hike on one of the hills near the Schloss.

Day 4 started with the featured talk by Yun Chu, who gave an overview of the psychological research on insight problem solving (the talk had originally been scheduled for Day 1, but due to unforeseen circumstances Chu could not arrive at the seminar earlier). The talk by Chu helped build further common ground between the computer scientists and psychologists as it explained in more detail common paradigms and concepts used in psychological research on problem solving. The rest of the day consisted of two workshop sessions aimed at stimulating the formation of new interdisciplinary perspectives and collaborations

and several short talks. Ute Schmid presented a framework for what Chomsky called a 'competence level model' of learning to problem solve, based on analytical inductive functional programming. Ulrike Stege gave a survey of typical computer science problems that are or could be used to investigate human problem solving strategies and pointed out some research challenges. Among them is the problem that researchers may think their participants are solving the problem that they posed, but the participants may in fact be solving an altogether different problem which the participants *think* the researchers have posed. Brendan Juba presented a new formal framework for heuristic rules based on PAC semantics. Nysret Musliu presented the concept of a (hyper)tree decomposition, a concept that can be utilized in algorithmic techniques for solving NP-hard problems efficiently. Jered Vroon presented a non-standard formalism in which problem solving is regarded as producing a solution rather than as a search through a search space. Last, Zyg Pizlo presented new algorithmic ideas for modeling human performance on the Traveling Salesman Problem based on the notion of multiresolution-multiscale pyramids. The day closed with a session in which participants brainstormed about novel interdisciplinary collaborations and open problems in the field. Some of these ideas were presented the same day, others were presented in the morning session of Day 5.

The morning of the last day of the seminar, Day 5, was reserved for short presentations of new collaborative ideas that the participants came up with, as well as the presentation of new ideas and open problems. The seminar closed with a wrap-up session in which participants reflected on the process and outcomes of the seminar. To conclude, the seminar was successful in several ways: (1) It has resulted in a renewed awareness of how computer science and psychology can complement each other in the study of problem solving; (2) it has created a new impetus for more involvement of computer scientists in contemporary problem solving research; (3) it has created more common ground between computer science and psychologist in the study of problem solving; (4) it has produced several novel ideas on how to conceptualize 'problem solving' and, in particular, 'problem solving by insight;' (5) it has produced several novel ideas on how to formalize these new conceptualizations; (6) it has produced concrete suggestions for new experimental paradigms for studying problem solving in the lab; (7) it has inspired new cross-disciplinary collaborative research projects; and last but not least (8) it has provided the groundwork on which follow-up Dagstuhl seminars can build in the future. With this seminar, the organizers hope to have contributed to an increased and sustained collaborative research effort between computer science and psychology in the domain of problem solving.

Subject Area 6

Software Technology

6.1 Feature-Oriented Software Development (FOSD)

Seminar No. **11021**
Organizers: Sven Apel, William Cook, Krzysztof Czarnecki, and Oscar Nierstrasz
09.–14. January, 2011 – www.dagstuhl.de/11021

Authors: Sven Apel, William R. Cook, Krzysztof Czarnecki, Oscar M. Nierstrasz

Seminar Motivation

Feature orientation is an emerging paradigm of software development. It supports partially or completely automated generation of software systems in a domain based on features—units of functionality covering the domain. The key idea of feature orientation is to emphasize the similarities of a family of software systems for a given application domain (e.g., database systems, banking software, and text processing systems) by treating features as first-class entities throughout the entire development lifecycle and across all the software artifacts, with the goal of reusing software artifacts among the family members. For example, features of a database system could be transaction management, query optimization, and multi-user operation, and those of a text processing system could be printing, spell checking, and document format conversions.

A key software engineering challenge is that a feature does not necessarily map cleanly to an isolated module of code. Rather, it may affect ("cut across") many components of a modular software system. For example, the feature transaction management would affect many parts of a database system, e.g., query processing, logical and physical optimization, and buffer and storage management.

The concept of feature orientation is still in its infancy. However, a growing community of researchers have been working on it for years, and there are related, well-known concepts of software engineering with well-populated research communities, e.g., software product lines, aspect-oriented software development, service-oriented architecture, and model-driven engineering. The main goal of the Dagstuhl seminar on FOSD was to gather researchers and practitioners who are active in these different communities to discuss the roots, state of the art, and future directions of FOSD research and practice and to strengthen the identity of the feature orientation community. We think that this seminar met this goal. An overview of the seminar organization and a summary of results are given below.

Seminar Organization

As a warm-up for the seminar we conducted a survey on FOSD. The idea was to ask the emerging community what they think FOSD was about. We asked the following seven questions:

    1. What do you think are the distinguishing concepts and ideas of FOSD?

2. What do you think are the major challenges in FOSD?
3. Which success stories of FOSD do you know?
4. What is missing in FOSD to adopt it in industry?
5. Is FOSD sufficiently visible in the software engineering community?
6. What do you expect to get out of the week?
7. What format and what kind of activities are you interested in (tutorials, demos, talks, breakout groups, brainstorming, social events, etc.)?

Based on the responses of 27 participants (available at the seminar's website), we prepared an introductory presentation on FOSD that aimed at "synchronizing" the participants, which is especially important in a field that is still in its infancy. After the self-introductions of all of the 49 participants and the introductory presentation, we allocated slots for the "hot" topics in the field of FOSD. On Monday, we had a discussion session of feature modularity. Tuesday was dedicated entirely to feature interactions. On Thursday, we had a mix of discussions sessions on industry adoption, the relation of FOSD to other development paradigms, as well as automatic product generation based on FOSD. On Tuesday and Thursday, we had demo sessions in the evening; on Wednesday, we had breakout sessions and a social event. Finally, on Friday, we had two wrap-up sessions, one concluding the individual discussions of the breakout groups and one summarizing the seminar and discussing results and further action items.

Seminar Results

From the organizers' perspective, the seminar was successful, although the large number of participants pushed the Dagstuhl concept to its limits. The topic attracted a lot of interest (the seminar was fully booked), and during the seminar there were many very lively and sometimes quite controversial discussions. Many participants contributed actively by organizing mini-tutorials, discussion sessions, and breakout groups. The results of the discussion sessions and the breakout groups are available at the seminar's website.

The participants used the seminar as an opportunity to learn about each others work and to establish collaborations, which will bear fruit in the years to come. As a first tangible outcome, we would like to point out the list of resources that the seminar's participants developed in a team effort:

- Key FOSD papers
- Annotated bibliographies in the portal researchr.org
- A suite of benchmark problems
- Teaching material on FOSD

The details of this list are described on the seminar's website. Further discussion points were how to promote FOSD in the future, how to further strengthen the community, and how to collaborate in an efficient manner.

In summary, we conclude that the seminar was constructive and largely met its goals. Dagstuhl provided a productive and interactive atmosphere. It was certainly a key event in the maturation of the FOSD community.

6.2 Bidirectional Transformations "bx"

Authors: Zhenjiang Hu, Andy Schürr, Perdita Stevens, James Terwilliger

This Dagstuhl Seminar was the second meeting bringing together 39 researchers from 13 countries across disciplines that study bidirectional transformations. The first one was the GRACE International Meeting on Bidirectional Transformations held in December 2008 near Tokyo, Japan [K. Czarnecki, J. N. Foster, Z. Hu, R. Lämmel, Andy Schürr, and J. F. Terwilliger. Bidirectional Transformations: A Cross-Discipline Perspective.*ICMT 2009*, 260–283]. The GRACE meeting consisted of short introductions from each of the participants on their background and work, followed by some longer presentations and demonstrations on some representative technologies from each field, concluding in some open discussion time. A primary takeaway from the GRACE meeting was an opportunity for each discipline to get some initial exposure to each other.

The Dagstuhl seminar intended to go a step further and begin to identify commonalities between the disciplines and start to set a cross-disciplinary research agenda. The first part of the seminar consisted of tutorials from each of the four represented disciplines on the various bidirectional transformation solutions that field has to offer. The second part consisted of cross-disciplinary working groups dedicated to investigating specific examples of commonality between solutions or identifying requirements, terminology, or scenarios that may reach across fields. There were also a number of sessions reserved for participants to give position statements regarding their individual fields of research.

Participation at both the Dagstuhl and GRACE seminars came from four disciplines: (1) Programming Languages (PL), (2) Graph Transformations (GT), (3) Software Engineering (SE), and (4) Databases (DB). Each of the first three disciplines made up about 2/7 of the participants, while databases took the remaining 1/7 out of about 39 participants. Representation from the database field was, nevertheless, an improvement over the turnout from the GRACE meeting.

Tutorials

The first part of the workshop was allocated for representatives from each of the four disciplines giving deep tutorials of about two hours length on the various solutions to bidirectional transformation problems offered by that discipline. In all cases these tutorials were presented by groups of scientists, who were responsible to offer the attendants of the other disciplines a basic introduction into their field as well as to highlight different aspects of ongoing research activities. The PL-related tutorial consisted of five parts presented by Nate Foster, Robert Glück, Zhenjiang Hu, Benjamin Pierce, and Janis Voigtländer. It contained a general introduction to the whole research area of bx. Furthermore, it addressed

a large variety of different aspects ranging from specific types of bx programming languages on one hand to model/graph transformation approaches in the software-engineering world and lense-based view definition approaches in the database system world on the other hand. As a consequence this set of (sub-)tutorials also was responsible for highlighting already existing links between the existing bx subcommunities. The DB-tutorial offered by Jean-Luc Hainaut and James Terwilliger nicely complemented the preceding tutorial. A survey of various opportunities for using bx transformations in database engineering processes as well as for specifying views by means of bx technologies was presented. The GT-related tutorial afterwards given by Andy Schürr and Frank Hermann focused on one special category of bidirectional graph transformations. Its two parts dealt with practical as well as theoretical aspects of bx graph transformations. Finally, the SE-related tutorial of Krzysztof Czarnecki and Stephan Hildebrandt concluded the first 2.5 days long introduction to the field of bx with a discussion of model synchronization and incremental change propagation techniques from the point of view of the software engineering community. For a more detailed description of the four (sets of) tutorials the reader is referred to their abstracts collected in this Proceedings.

Working Groups

After the ground had been prepared by the above mentioned list of tutorials the participants had the opportunity to present specific research results and challenges in the form of 27 short presentations. Furthermore, six working groups were organized related to cross-disciplinary topics that were identified during the first days of the Seminar. The summaries of three of these working groups only finally made into the Proceedings, but all of them were very helpful to deepen our understanding of common challenges and triggered new research cooperations across the four disciplines. The list of addressed topics included aspects like the design of bx benchmarks, the identification of typical bx application scenarios and their requirements as well as options for the joint development of a new generation of bx languages that would incorporate and combine elements from solutions developed in different disciplines. The individual position statements covered a very broad spectrum of topics; they were used to intensify already started discussions of the first tutorial days and gave further input for the identification of new workshop topics. Again the reader is referred to the abstracts collected here for more details.

Summary and Future Work

We went into the seminar knowing that longer-term ideas like a common research agenda or a benchmark would take longer than a single week. The participants decided on several follow-up actions to keep work progressing:

- A follow-up meeting in the same style as the GRACE and Dagstuhl seminars to continue collaborating on a cross-disciplinary research agenda
- Workshops at conferences associated with each of the disciplines to work toward specific, targeted goals (a first one has already been scheduled associated with GTTSE

2011, and will focus on developing a benchmark[1]; a second follow-up event has just been accepted as a satellite workshop for ETAPS.)

- Tutorials and other assorted smaller, education-minded events at conferences to continue bringing awareness of bidirectional solutions from other disciplines, as well as awareness of the general BX effort
- Smaller-scale research cooperations that combine techniques from different fields like merging concepts from bidirectional programming languages and triple graph grammars as envisaged in one of the seminar's working groups.

In particular, one goal of the upcoming seminars and workshops is to increase participation from the database community. The bidirectional transformation problem has origins deep in the database community, but now has grown so that solutions are being driven from many different directions in different fields across computer science. The plan is to hold some of the tutorials or workshops at database venues to help solicit more ideas and opportunities for collaboration; details will be made available once they are scheduled.

---

[1]http://www.di.univaq.it/CSXW2011/

6.3 Self-Repairing Programs

Seminar No. **11062**
Organizers: Mauro Pezzè, Martin C. Rinard, Westley Weimer, and Andreas Zeller
06.–11. February, 2011 – www.dagstuhl.de/11062

Authors: Mauro Pezzè, Westley Weimer, Andreas Zeller

Dagstuhl seminar 11062 "Self-Repairing Programs" included 23 participants and organizers from research and industrial communities. Self-Repairing Programs are a new and emerging area, and many participants reported that they initially felt their first research home to be in another area, such as testing, program synthesis, debugging, self-healing systems, or security. Over the course of the seminar, the participants found common ground in discussions of concerns, challenges, and the state of the art.

Why Self-Repairing Programs?

Recent years have seen considerable advances in automated debugging. Today, we have techniques that automatically determine problem causes — in the source code, in program input, in the change history, or in internal data structures. While these approaches make it considerably easier to find the causes of defects, their precision is still insufficient to suggest a single concrete course of action — a human in the loop is still required to design and apply the patch. At the same time, there is an ongoing need for self-healing systems — systems that can recover from failures and even reconfigure themselves such that the failure no longer occurs. Most research efforts in this direction, though, assume planned recovery — that is, well-defined recovery strategies for anticipated failures.

An alternative is to explore self-repairing systems from a generic perspective — that is, to develop techniques that repair systems that are as generic, unassuming, and non-intrusive as program analysis and debugging. The idea is to determine actual fixes to state, to configuration, or to code. These fixes can be seen as guidance for the developer on how to fix the problem and evolve the software. However, fixes can also be deployed automatically, and effectively lead to programs that fix themselves. Such techniques may be particularly useful for orphaned systems that are no longer maintained or for critical software for which downtime is extremely expensive or even unacceptable. In these situations, there is no time to wait for a human developer to find and fix the bug. A synthesized patch can form a first line of defense against failures and attacks, a "first aid" approach to buy time while more expensive or manual methods are deployed. At the same time, automatically generated fixes provide a much richer diagnostic quality then simple fault localization, and thus may dramatically reduce the time it takes to debug a problem.

Goals of the Seminar

This main goal of this seminar was to provide knowledge exchange, mutual inspiration, and opportunities for collaborations for a rapidly developing field. The seminar aimed to bring

together researchers in dynamic program analysis, automated debugging, specification mining, software survival techniques, and autonomic computing to increase awareness of these issues and techniques across relevant disciplines (program analysis, debugging and self-adaptive systems), and to discuss:

- how to monitor systems to detect abnormal state and behavior
- how to generate fixes and how to choose the best fixes
- how to deploy them in real-life systems and how to deal with the issues that arise when automatically correcting errors in software systems

Format and Presentations

The seminar started with summary presentations to bring all participants up to the speed on the state of the art and establish a common terminology. Subsequent activities alternated between technical presentations and plenary discussion sections. The seminar participants also split into two groups based on the self-identified focus areas of "The Architecture of Self-Repairing Systems" and "Validating Automated Repairs via Testing and Specification". Some evenings featured demonstrations or special-interest talks.

Common Concerns and Insights

As a whole, the group identified four challenge areas and opportunities for self-repairing programs: Architecture, Redundancy, Efficiency and Trust. In terms of Architecture, there was an acknowledgment that overall progress could be made by tackling particular problems or subdomains (e.g., fixing only atomicity violations or fixing only web applications, etc.) and potentially combining solutions later. There was a broad realization that redundancy is important on many levels: as a source of comparison for finding bugs or specifications; as a source of repair components; and as a main component of self-healing or self-adaptive systems at the architectural level. In terms of efficiency, the speed of the repair process — including the time required to validate a candidate repair — was of some concern, although many current techniques take minutes rather than hours to produce repairs. Trust was perhaps the most universally accepted issue: a notion that it is the responsibility of the repair process to provide an assurance argument, backed up by evidence, that would give a user or developer confidence that a repair can be applied safely. As general guidelines, we felt that an automatically-generated repair should not (or should at most minimally) regress the program by impairing functionality, and that applying such a repair should not be worse than doing nothing.

The group also identified two cross-cutting concerns related to correctness and evidence. The first was a notion that the evidence used to produce a repair (e.g., a few test cases or a partial specifications used for fault localization or repair construction) might be different from the evidence used to validate a final candidate repair (e.g., a larger test suite or a more complete specification). In addition, emphasis was placed on a clear characterization of common versus anomalous (or incorrect) behavior, possibly via a learned specification.

Challenge Areas Identified

The participants also identified a number of challenge areas or difficult tasks. By far the most popular was a notion of benchmarking. While the group acknowledged that the field is still quite new, and that formal benchmarks may not be appropriate, there was a desire for representative instances of programs with defects, tests that demonstrate those defects, normal regression tests, and indications of how humans fixed those defects.

The second challenge identified was the need for low-overhead, continuous monitoring to learn formal specifications for correct behavior, detect anomalies, and validate a system after repair deployment. The third challenge was to provide "just-in-time" repairs that were as quick as the auto-correction in Word or Eclipse. A fourth challenge related to documenting repairs or otherwise equipping them with evidence and arguments that would give confidence that they fix the system without causing additional harm.

A number of additional concerns were identified but were supported by a smaller segment of the participants. These included focusing on the economic value of repairs (e.g., targeting high severity defects or measuring the effort saved), the desire to repair programs even if a regression test suite is note available, the desire to have tools that succeed or fail with certainty (i.e., rather than producing incorrect repairs), some notion of automated repair techniques fixing 10% (or 50%, or 70%) of all reported bugs with some level of confidence, and the desire to improve automated fault localization techniques and allow them to report causes, not just locations.

6.4 Models@run.time

Authors: Uwe Aßmann, Nelly Bencomo, Betty H. C. Cheng, and Robert B. France

To date, research on model-driven engineering (MDE) has mainly focused on the use of models during software development. This work has produced relatively mature techniques and tools that are currently being used in industry and academia to manage software complexity during development. Research on models@run.time seeks to extend the applicability of models and abstractions to the runtime environment, with the goal of providing effective technologies for managing the complexity of evolving software behaviour while it is executing.

As is the case for many software development models, a runtime model is often created to support reasoning. However, in contrast to development models, runtime models are used to reason about the operating environment and runtime behaviour for some purpose, for example, to determine an appropriate form of adaptation, and thus these models must capture abstractions of runtime phenomena. Different runtime dimensions need to be balanced when adapting software at runtime, including efficient use of resources (time, memory, energy), context-dependencies (time, location, platform), and personalization concerns (quality-of-service specifications, profiles). The hypothesis can be stated as follows: Models@run.time that provide meta-information on these dimensions during execution enables the development of technologies that automate (1) runtime decision-making and (2) safe adaptation of runtime behaviour. Thus, we anticipate that this technology will play an integral role in the management of autonomic systems and self-adaptive systems.

The problems targeted by the models@run.time community are multi-faceted and thus tackling them requires expertise from a variety of research areas. The Dagstuhl Seminar models@run.time brought together a diverse set of researchers and practitioners with a broad range of expertise, including MDE, software architectures, reflection, self-adaptive systems, validation and verification, middleware, robotics and requirements engineering.

The following gives the objectives of the seminar and describes the extent to which they were met :

1. Objective: To identify and document the potential benefits of Models@run.time including benefits associated with their use in adaptive and autonomic systems. Extent Met: One of the working groups (Group 3: Uses and Purposes of M@RT) was charged with coming up with use cases that demonstrate the benefits associated with the use of models@run.time. The report produced by this group discusses the types of significant software systems that can benefit from use of models@run.time.
2. Objective: To reach a common understanding of the terminology and associated concepts that underpin the use of different models once a system is deployed. Extent

Met: Each working group defined the terminology and the concepts used in the descriptions of their primary outcomes.

3. Objective: To identify a set of key research challenges that must be tackled to address the real-world problems posed by self-adaptive systems within the next five (5) years (the research roadmap). Extent Met: Each working group identified, discussed, and described the key challenges in its focus area. The identified challenges will be included in the roadmap we plan to publish.

4. Objective: To also identify associated technology transfer strategies to ensure that the research in this area has impact on industrial practice and associated methodologies and tool-sets. Extent Met: Group 2 (Runtime updating/adaptation mechanisms) focused on discussing and analysing the effectiveness of technologies that have been developed to support models@run.time.

5. Objective: To publish a collection of articles containing the roadmap, as well as papers from the participants. Extent Met: Plans have been put in motion for publishing peer-reviewed papers from participants in a LNCS State-of-the-Art Survey Volume on Models@run.time.

The seminar consisted of participant presentation and working group sessions. Monday and Tuesday morning were "speed dating" days, in which everyone (with the exception of some of the organizers, to better manage time) presented a 10-min introductory talk. The presenters covered a wide range of topics including adaptive cyber-physical systems, self-evolution, requirements-driven runtime adaptation, safe and trustworthy autonomous robots, adaptive and self-managing software, runtime variability and architectural reconfiguration at runtime. We also had two longer presentations by persons from industry (ERICSSON). The presenters were Joe Armstrong and Paer Emanuelsson. From Tuesday to Thursday participants worked in working groups that focused on particular aspects of models@run.time research. At the start of each day the organizers summarized the activities that were to take place and the deliverables that were to be produced at the end of the day. At the end of each day, each group presented their deliverables.

On Wednesday evening we had a productive panel where the following questions were discussed: What are the compelling business models for models@run.time? What are the killer applications? What are the obstacles to deployment of models@run.time systems? What are key enabling technologies for models@run.time (e.g., standards or component models).

Subject Area 7

Distributed Computation, Networks, Architecture

7.1 Multi-Core Memory Models and Concurrency Theory

Seminar No. **11011**
Organizers: Hans J. Boehm, Ursula Goltz, Holger Hermanns, and Peter Sewell
03.–01. January, 2011 – www.dagstuhl.de/11011

Authors: Hans J. Boehm, Ursula Goltz, Holger Hermanns, Peter Sewell, Christian Eisentraut, and Malte Lochau

The current and future trend to multi-core and many-core computing systems suggests that within the next decade, concurrent multi-threaded programming will continue to replace sequential programming as the standard programming paradigm. However, concurrency and modern computer architecture do not go together easily:

- Current programming language memory models are still incomplete. Mainstream languages such as Java increasingly promise sequential consistency for data-race-free programs. However, how data races can be handled in a way that supports reasonable performance, security, and debugability, is currently completely unknown.
- Hardware specifications are so informal that it is very hard to know whether we have a correct implementation of the language specs (if we knew how to specify those fully). It is not clear that existing ISAs, which have a long history, are a good match for the language semantics we are developing. There is an argument that this misalignment continues to encourage languages to support complex low-level constructs.
- The concurrent algorithms that are now being developed, and which are key to exploiting multiprocessors (via high-performance operating systems, hypervisor kernels, and concurrency libraries), are very subtle, so informal reasoning cannot give high confidence in their correctness.
- While there is a rich literature on concurrency theory, and extensive prior work on software verification for concurrency software (including process calculi, model-checking, temporal logics, rely-guarantee reasoning, and separation logic), almost all of it assumes a global notion of time, unrealistic though that is for these systems and algorithms.

Concurrency theory has a long tradition in investigating the foundational principles of concurrent computation, devoted to parallelism and causality of operations. It has created a wealth of models and notions readily applicable in this semantic challenge. Recent developments in the research communities of programming languages and concurrency theory, respectively, indeed show a growing trend towards cross-fertilization.

This seminar has fostered cross-fertilization of different expertises that will help to develop novel practical and, at the same time, theoretically sound paradigms for multi-core programming. It brought together concurrency theorists, memory model experts, computer systems architects, compiler experts, and formal verification researchers. The aim of the seminar was to adress in particular:

1. Survey of problem domain: state of the practice in multi-core-programming and state of the art in memory consistency models.
2. Application of concurrency theory approaches to reveal underlying concepts of parallelism, reordering, causality, and consistency.
3. Cross-fertilization across formal approaches to memory consistency models and semantics of multithreaded programming.
4. Attack points for formal analysis and computer aided programmer support.

Many of the questions that stood at the outset of this seminar have not been conclusively answered, thus yielding many potentials for further investigation. However, what makes this seminar uniquely successful is that it initiated a vivid exchange between a multitude of different scientific and industrial communities. During the seminar, it became clear that the current and future challenges of multi-core programming and memory models design in software and hardware can only be solved if the communities continue to exchange ideas and will learn from each other.

7.2 Learning from the Past: Implications for the Future Internet and its Management?

Seminar No. **11042**
Organizers: Gabi Dreo Rodosek, Aiko Pras, Henning Schulzrinne, and Burkhard Stiller
25.–28. January, 2011 – www.dagstuhl.de/11042

Authors: Gabi Dreo Rodosek, Aiko Pras, Henning Schulzrinne, Burkhard Stiller

The attendees of the seminar discussed a hot topic of various research initiatives, namely the Future Internet, and its management. Can we make assumptions for the Future Internet if we question ourselves how the present Internet needs to be re-engineered by analyzing its current major limitations? Although discussions like the clean-slate vs. evolutionary approach are taking place, and are certainly vision-based, an alternative approach to investigate why certain decisions failed or succeeded seems to be promising as well.

Looking back, we recognize that the Internet architecture and protocols are, in theory, defined by about 5000 RFCs ("standards"), but in practice depend on a much smaller number of core assumptions, protocols and technologies. The Internet is largely defined by a modest set of protocols with numerous extensions and modifications: IPv4 and IPv6 at the network layer, UDP and TCP at the transport layer, and a handful of standardized and a large number of proprietary application protocols, for applications from email and file transfer to gaming. Along with these data-delivering protocols, control plane protocols such as BGP, configuration protocols such as DHCP and management protocols such as SNMP are needed to keep the Internet running.

Many of these protocols are now at least a decade, in some cases three decades, old. Some have aged gracefully; others clearly have reached the end of their useful lifetime. Also, numerous other protocols and architectures have been proposed in the literature and many were even standardized over the past decades, but most have had very limited or no practical impact. Unfortunately, the design of new protocols is still largely based on folklore and ideas passed on informally during discussions at conferences, IETF standardization meetings and PhD seminars. It is relatively easy to stuff all ideas accumulated into a new protocol, but it seems much harder to destil the lessons learned into future designs, so that we do not loose some of the core properties that have made the Internet successful. Not only in the design of protocols, the same is true also for the development of management approaches. Although it seems that we have recognized that the management of the Future Internet needs to be addressed in the design phase, it is largely ignored so far, maybe because we do not know how to approach it.

Therefore, learning from the past would give us valuable insights for the design of the Future Internet, and its management. The goal of the seminar was to discuss these questions and identify the key findings.

Overview of Talks

The seminar investigated core network architectural issues and core protocols as well as management approaches, to define precisely what made design decisions succeed against their competitors at the time of design, and which choices have turned out to be prescient and where we are reaching the end of the extensibility. The discussion was motivated by the following talks:

- Georg Carle, TU München
  Network Management for Measuring, Processing, and Reacting
- Gabi Dreo Rodosek, Universität der Bundeswehr München
  Learning from the Past: KISS principle for the Future Internet, and Self-Management
- Olivier Festor, INRIA – Nancy Grand Est
  A Biased Focus on 20 Years Research in Network Management
- Sape J. Mullender, Alcatel-Lucent Bell Labs – Antwerpen
  New Protocols for Accessing Services, Files, Content and more
- Aiko Pras, University of Twente
  Learning from the past – why OSI failed
- Danny Raz, Technion – Haifa
  Networking and Computation Resources: Can They Mix?
- Ramin Sadre, University of Twente
  Is The Future Unpredictable?
- Peter Schoo, Fraunhofer Institut SIT, München
  Management in FI needs increasingly more and better means for anomaly detection
- Henning Schulzrinne, Columbia University
  What did we learn? Random thoughts on protocol design
- Burkhard Stiller, Universität Zürich
  Communications + Management/Operations + Security* != Communications, Management/Operations, Security*
- Hermann de Meer, Universität Passau
  Problems in cross layer optimizations contrasted to E2E communications

Lessons Learnt resp. the Key Findings Identified

In the following the key findings of the seminar during the discussions are summarized:

1. We need a better understanding of design trade-offs.
2. New applications can change traffic characteristics in a few months. In the past decade several applications dramatically changed the way how the Internet is used. Nobody has actually foreseen the success of P2P networks, and especially Youtube and Facebook. Thus, the question is whether it is possible to design a Future Internet without having any ideas what the "next big things" could be. If thus the traffic changes are unpredictable, then we need to establish a fast and stable infrastructure without any assumptions on the traffic.

3. The waterfall model does not work in practice in communications, for sure, software is not a "one-time instance", changes will occur for some time. Thus, versions are needed, and for protocols we may arrive at the same iterative refinement approach.
4. The use of formal specifications in case of OSI was rather harmful than it helped. OSI failed because nobody needed it, and options have been considered to be harmful.
5. Implementations from the beginning are necessary. However, with first to the market means to loose quality.
6. Full decoupling of planes (management, user, control) is good in an "old-style telco world", however, it will not work in the Future Internet.
7. There is no future for a centralized management (in most cases). It is necessary to move the research effort towards self-management approaches.
8. Future protocols should have built-in management possibilities (Management-by-Design).
9. The focus of management must concentrate on algorithms to automate it, the supporting data and data feeding will follow afterwards.
10. Effective management is the management that disappears or is invisible, respectively.
11. The need for self-configuration in access networks, programmable nodes (measurement is an important case on layer 3).
12. Fundamentally new protocols and approaches are needed that are additionally suitable for addressing replicated objects, which are mutable, as well as trust.
13. Assumptions for DiffServ/IntServ changed, there is no need anymore for end-to-end Quality-of-Service (QoS), except for Label Switched Paths.
14. There is no need for Network Address Translation anymore due to IPv6.
15. IP addresses are unsuitable for the mobility aspect. Mobile IP has failed.

Conclusions

All discussions have shown that a lot of various concepts that have been developed are nowadays, or will be in the near future, more or less obsolete due to the enormous amount of bandwidth being available in network backbones or migration to IPv6. These insights of those discussions have shown very interesting combinations of effects and consequences as well as mechanisms and their implementation alternatives. Based on these observations a scientific paper is under preparation.

7.3 Information Management in the Cloud

Authors: Anastassia Ailamaki, Michael J. Carey, Donald Kossmann, Steve Loughran, Volker Markl

Cloud computing is emerging as a new paradigm for highly scalable, fault-tolerant, and adaptable computing on large clusters of off-the-shelf computers. Cloud architectures strive to massively parallelize complex processing tasks through a computational model motivated by functional programming. They provide highly available storage and compute capacity through distribution and redundancy. Most importantly, Cloud architectures adapt to changing requirements by dynamically provisioning new (virtualized) compute or storage nodes. Economies of scale enable cloud providers to provide compute and storage powers to a multitude of users. On the infrastructure side, such a model has been pioneered by Amazon with EC2, whereas software as a service on cloud infrastructures with multi-tenancy has been pioneered by Salesforce.com.

The Dagstuhl seminar on Information Management in the Cloud brought together a diverse set of researchers and practitioners with a broad range of expertise. The purpose of this seminar was to consider and to discuss causes, opportunities, and solutions for technologies, and architectures that enable cloud information management. The scope ranged from web-scale log file analysis using cluster computing techniques to dynamic provisioning of resources in data centers, covering topics from the areas of analytical and transactional processing, parallelization of large scale data and compute intensive operations as well as implementation techniques for fault tolerance.

The seminar consisted of keynotes, participant presentations, demos and working groups. The first two seminar days consisted of a keynote by Helmut Krcmar on "Business Aspects of Cloud Computing" as well as 33 short presentations on various aspects of cloud computing. On the evening of the second day, the participants formed working groups on economic aspects, programming models, benchmarking. The third day of the seminar consisted of two keynotes, by Dirk Riehle on "Open Source and Cloud Computing" and by Donald Kossmann on "Benchmarking". After these keynotes, working groups discussed their respective topics. In the evening, an industrial panel with Miron Livny, Steve Loughran, Sergey Melnik, Russell Sears, and Dean Jacobs discussed research challenges in Cloud Computing from an industrial point of view. On the fourth day, a keynote by Miron Livny discussed Cloud Computing from a distributed systems and high-performance computing point of way. After the keynote, a demo session presented the following systems:

- HyPer: A Cloud-scale Main Memory Database System (Team from TUM)
- Asterix and Hyrax (Team from UCI)

- Stratosphere (Team from TU Berlin, HU Berlin and HPI)
- Myriad Parallel Data Generator (Team from TU Berlin)

After these demos, working groups continued during the day and presented their results in the evening. The last day of the seminar, participants continued in working groups and discussed further collaborations with respect to papers and project proposals. During this day, several abstracts for papers have been prepared, and discussions about several joint research project proposals have started.

The organizers hope that the seminar has helped to organize the research space in cloud computing and identified new research challenges. We look forward towards research collaborations and papers that were bootstrapped during this intensive week.

Subject Area 8

Embedded Systems

8.1 Science and Engineering of Cyber-Physical Systems

Seminar No. **11441**
Organizers: Holger Giese, Bernhard Rumpe, Bernhard Schätz, and Janos Sztipanovits
01.–04. November, 2011 – www.dagstuhl.de/11441

Authors: Holger Giese, Bernhard Rumpe, Bernhard Schätz, Janos Sztipanovits

Today, a new category of engineering systems is emerging that combines physical processes with computational control in a holistic way: Cyber-physical systems (CPS). Cyber Physical Systems are engineered systems of synergistically interacting physical and computational components. The key property of these systems is that functionality and salient system properties are emerging from an intensive interaction of physical and computational components. As the computational components are aware of their physical context, they are intrinsically distributed, (time)-synchronizing, have to cope with uncertainty of sensoric-input and need to produce real-time reactions. Consider an unmanned aerial vehicle (UAV) with active wings. In such an UAV, a cyber-physical system may consist of an embedded controller, monitoring the airflow over the wing surface and electromechanical actuators modulating the airflow to ensure laminar flow such that the vehicle is capable of extreme maneuvers. Unlike more traditional embedded systems, full-fledged CPSs are often designed as networks of interacting elements including autonomous automotive systems, medical monitoring, process control systems, distributed robotics, and automatic pilot avionics. The question naturally arises: are cyber-physical systems fundamentally different such that they need a different fundamental science, a different development approach, or is the current approach sufficient and no new research is necessary? We argue that new science, new techniques, and a new view are necessary. Traditional separation along engineering disciplines in the design of such systems leads to various quality, maintainability and evolutionary problems: thus, integrated theories and engineering techniques are urgently needed. The technology is pervasive, transcends industrial sectors and serves as the engine of innovation for new generation of products. CPS is also a disruptive technology that transforms established industries, may create new ones and possibly rearranges the status quo of development in entire industrial sectors. Current industrial experience tells us that we have reached the limits of our knowledge regarding integration of computers and physical systems. These shortcomings range from technical limitations in the scientific foundations of cyber-physical systems through the engineering processes to the way we educate engineers and scientists that support cyber-physical system design. However, besides the National Science Foundation initiative in the US, the topic is currently addressed by initiatives such as intelligent and autonomic automobiles, ambient intelligence, self-organizing embedded systems, plant-control and reparation, self-optimizing mechatronic systems, 'smart' power grids, in-home medical assistance devices, etc. This seminar focused on the scientific foundations and the engineering aspects of cyber-physical systems by bringing together researchers from both academia and industry to discuss the new scientific foundations and engineering principles for the vastly emerging field of CPS.

Subject Area 9

Modelling, Simulation, Scheduling

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

Seminar No. **11371**
Organizers: Isaac Elishakoff, Vladik Kreinovich, Wolfram Luther, and Evgenija D. Popova
11.–16. September, 2011 – www.dagstuhl.de/11371

Authors: 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 biomechanical 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.

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 incertitude 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".

Subject Area 10

Cryptography, Security

10.1 Online Privacy: Towards Informational Self-Determination on the Internet

Authors: Simone Fischer-Hübner, Chris Hoofnagle, Kai Rannenberg, and Michael Waidner

While the collection and monetization of user data has become a main source for funding "free" services like search engines, on-line social networks, news sites and blogs, neither privacy-enhancing technologies nor its regulations have kept up with user needs and privacy preferences.

The aim of this Dagstuhl Perspectives Workshop is to raise awareness for the actual state of the art of on-line privacy, especially in the international research community and in ongoing efforts to improve the respective legal frameworks, and to deliver soon after the workshop a Dagstuhl Manifesto providing recommendations to industry, regulators, and research agencies for improving on-line privacy. In particular we have examined how the basic principle of informational self-determination, as promoted by European legal doctrines, could be applied to infrastructures like the Internet, Web 2.0 and mobile telecommunication networks.

It was deemed necessary and timely to bring together a *broad spectrum of key contributors* in order to promote both legally and commercially viable foundations for a balanced on-line privacy:

- *Academia* (specifically data security, privacy, cyber-law, and privacy-influential technologies & services),
- *Public sector* (data protection officers, organizers of relevant research programs, relevant civil rights organizations), and
- *Industry* (providers of communication solutions, browsers and apps; data aggregation and web analytics companies; providers of major Internet and mobile Internet services)

This workshop and its planned Dagstuhl Manifesto have four goals, aside from galvanizing an emerging research community:

1. *Provide a big picture of on-line privacy, which can be understood widely*
   Because of swift progress in the mobile Internet, on-line social networks, and on-line advertisements, it is a challenge for non-experts (and perhaps even experts themselves) to understand the current state of on-line privacy including the technologies and systems to collect personal information on-line.
2. *Compile the industry and engineering options to improve on-line privacy*
   On-line privacy depends on the technologies and systems used to access Internet/Web 2.0 services as well as on the services provided to users. Therefore industry has a strong influence.

3. *Update the respective legislative and regulative authorities on their options for enforcing practical, commercially viable informational self-determination of users in global infrastructures (e.g. EU's Privacy Directive to be revised in 2011)*
   Access to personal information is critical to self-determination; it is also seen as a right that serves a policing function among information-intensive firms. However, legal and business structures have often foreclosed rights of access, or made them impracticable for consumers to exercise.

4. *Foster industry's and academia's research for creating effective on-line privacy technologies, components, and systems that promote informational self-determination*
   Corresponding to additional risks for on-line privacy, new approaches are required in research to again establish adequate levels of on-line privacy.

This workshop has been structured into four parts, for each part, a topic responsible has been assigned:

- Part 1: Current S-o-A of on-line privacy w.r.t. informational self-determination
  Responsible: Alma Whitten, Google Research, Great Britain
- Part 2: Industry & engineering options to improve on-line privacy
  Responsible: Michael Waidner, ex-IBM CTO Security, then TU Darmstadt, Germany
- Part 3: Recommendations for improving regulations of online privacy
  Responsible: Caspar Bowden, Microsoft WW Technology Office, Great Britain
- Part 4: Recommendations for research to improve the S-o-A of online privacy
  Responsible: Kai Rannenberg, Goethe University Frankfurt, Germany

A Dagstuhl Manifesto will conclude this workshop according to `http://www.dagstuhl.de/en/program/dagstuhl-perspectives/`.[1]

---

[1](Editorial) The manifesto of 11061 is available at `http://dx.doi.org/10.4230/DagMan.1.1.1`.

10.2 Verifiable Elections and the Public

Authors: R. Michael Alvarez, Josh Benaloh, Alon Rosen, and Peter Y. A. Ryan

This seminar brought together leading researchers from computer and social science, policymakers, and representatives from industry to discuss the issue of "Verifiable Elections and the Public." The purpose was to present new research, develop new interdisciplinary approaches for studying election technologies, and to determine ways to bridge the gap between research and practice. This seminar built upon the foundation provided by an earlier Dagstuhl seminar in 2007: Frontiers of Electronic Voting, Seminar number 07311, www.dagstuhl.de/07311.

The initial sessions of the seminar were devoted to a conceptual discussion of verifiable voting, and to a summary of the apparent obstacles associated with implementing innovations in election technology. There was a general sense from most seminar participants that while great progress has been made in development of verifiable voting systems, there has not been as much progress towards testing, implementing, and deploying these new voting systems. Additionally, the research community would like to be more involved in policymaking and the practice of election administration. In particular, a panel discussion regarding obstacles to innovation was quite productive, outlining several reasons for this feeling that insufficient progress has been made, including politics, a lack of interest on the part of voters, legal and regulatory confusion, a lack of sensitivity to the training and incentives of election officials, and a sense that some efforts to innovate have been overly ambitious and complex.

After a productive discussion of obstacles to innovation, seminar participants heard talks about a variety of recent e-voting and Internet voting trials and projects. These included talks on developments in Australia, Brazil, India, Estonia, Switzerland, and Norway, as well as discussion of voting technology implementations in two U.S. counties: Cuyahoga, Ohio and Sarasota, Florida. These presentations provided a great deal of real-world information on technological and practical issues regarding the implementation of new voting systems throughout the world.

Substantial time was devoted to the presentation of new voting systems. Some of these presentations regarded innovative new conceptual and hardware schemes, including new protocols for elections and ideas like using smartphones as voting platforms. Other presentations focused on advancement and elaboration of existing voting systems, for example further development of voting systems like Helios, Wombat, Prêt à Voter, and Scantegrity. All of these presentations documented the significant progress that has been made in the scientific community, in terms of development and elaboration of important cryptographic and procedural protocols for voting, as well as new ideas for potential uses of technology in elections.

One of the most exciting new developments since the earlier 2007 Dagstuhl seminar has been the implementation and testing of some of the new voting systems that are under development. These include implementations of Helios and Wombat, and also a systematic usability and understandability project regarding Prêt à Voter. These efforts are providing important data that is aiding in the continued development of these and other related new voting systems.

Voting online continues to expand throughout the world, as was widely discussed during talks on projects in Estonia, Norway, and Switzerland. And many of the talks about new voting systems regarded new protocols that can be deployed online, like the extension of Scantegrity to remote online use ("Remotegrity"). Presentations about these projects came from social scientists, technologists, and policymakers.

At the same time, there continue to be important questions raised from researchers about voting online, focusing largely on concerns about the security of online voting – specifically including the challenges of making online voting coercion-resistant in a practical, convincing, and usable way. These concerns fueled much discussion during the seminar, and it is clear that more research about the voting systems being currently deployed, and those proposed for use in the near future, is needed.

Concern is growing in the research community about how to maximize the impact of the considerable body of research that has accumulated in recent years. Seminar participants raised concerns about ways to improve the science of studying election technology, as well as methods to improve connections and collaborations between the scientific and policymaking communities. These issues will continue to intensify in the near future, and we hope that the discussions at this Dagstuhl seminar will fuel progress in the development of new scientific opportunities for research and dissemination, as well as closer collaboration between scientists and policymakers.

10.3 Security and Rewriting

Authors: Hubert Comon-Lundh, Ralf Küsters, and Catherine Meadows

Security is a fundamental problem in computer science. Because of the possible catastrophic problems that can arise from poor security, the ability to mathematically prove and formally verify the security of computer systems is vital. Research has been ongoing in this area since the 1970's and has been the subject of many Dagstuhl seminars, including (in the last three years) "Theoretical Foundations of Practical Information Security" (November 2008)[2], and "Formal Protocol Verification Applied" (October 2007)[3].

Research on formal proofs of security has demonstrated that rewriting techniques, including completion, narrowing, unification, play a central role in this area, for example:

- Formally modeling the properties of cryptographic primitives: fundamental properties of the cryptographic primitives are presented as algebraic theories and used as a basis for security analysis.
- Automatically proving security protocols: both the protocol and the attacker's possible actions can be modeled as a rewrite system and unification algorithms play a central role in the security analysis of such systems.
- Formally specifying and verifying security policies: the (possibly infinite) set of allowed transitions may be a represented as a finite rewriting system. The views on a document or a class of documents may be specified by tree automata.
- Modeling and analysis of other security-critical applications: rewrite techniques are used to model and analyze the security of web services, APIs and systems for access control.

The goal of this seminar was (i) to bring together researchers who have a background in rewriting techniques and researchers who have a background in security applications (or both) (ii) to answer, among others, the following questions:

- Are there specific problems in rewriting that stem from security applications and would deserve some further research? For instance, do the algebraic theories of cryptographic primitives enjoy some specific properties? Are there restrictions that are relevant to the applications and that would yield more efficient unification/rewriting algorithms? Which new challenges does the addition of an arbitrary attacker context bring? What are the specific problems on tree automata that are brought by security applications?
- What are the limits/successes/failures of rewriting techniques in security applications?
- What are the emerging research areas at the intersection of security and rewriting?

---

[2] http://www.dagstuhl.de/08491
[3] http://www.dagstuhl.de/07421

10.4 Quantum Cryptanalysis

Seminar No. **11381**
Organizers: Serge Fehr, Michele Mosca, Martin Rötteler, and Rainer Steinwandt
18.–23. September, 2011 – www.dagstuhl.de/11381

Authors: 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:

- *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.

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.

10.5 Public-Key Cryptography

Seminar No. **11391**
Organizers: Marc Fischlin, Anna Lysyanskaya, Ueli Maurer, and Alexander May
25.–30. September, 2011 – www.dagstuhl.de/11391

Authors: Marc Fischlin, Anna Lysyanskaya, Ueli Maurer, and 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.

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 extend 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.

10.6 Forensic Computing

Authors: Felix C. Freiling, Dirk Heckmann, Radim Polcák and Joachim Posegga

After a brief introduction by the organizers, the seminar started off with a sequence of 3 slide/5 minute talks by all participants stating their research interests, their background and their expectations towards the seminar. In the afternoon, two introductory talks by Dieter Gollmann ("Access control — principles and principals") and Stig Mjolsnes ("ICT and forensic science") paved the way for a common understanding of the open questions in the area and the relation of forensic computing to computer security.

Wednesday morning commenced with a first introductory law talk by Focke Höhne ("Introduction to German IT Forensics Law"). It was followed by two insightful technical talks from presenters who had considerable practical experience in the area: Glenn Dardick and Kwok Lam.

The afternoon was spent on a pleasant hike to a nearby village where the Dagstuhl office had organized delicious traditional coffee and cake. On the way back to Schloss Dagstuhl a group of adventurers separated from the main party to explore the woods around Wadern. They only managed to return to Dagstuhl in time because of modern navigation technology (paper maps provided by the Dagstuhl office). Reasons for the failure of more traditional technology (iPhones, etc.) were discussed in the evening in the wine cellar.

Thursday saw a mix of legal and technical talks: Herbert Neumann raised many questions during his presentation of practical (law) case studies while Viola Schmid presented a proposal for a "Casebook on Cyber Forensics". Harald Baier discussed the deficits of forensic hash functions and Felix Freiling shared some of his experiences from teaching digital forensics. After lunch Michael Spreitzenbarth presented an overview over mobile phone forensics while Radim Polčàk gave some background on the issues of data retention relevant in different countries. Joshua James pointed out the necessity to overcome the traditional separation of sciences and encouraged more interaction between computer science and law.

Finally, Johannes Stüttgen introduced the method of "Selective Imaging" to improve the digital evidence collection process.

Friday morning hosted a series of three talks from computer science, law and practice. Stefan Kiltz spoke about techniques to seize transient evidence in networks, Sven Schmitt gave an overview of digital forensics at the German federal police (BKA), and Nicolas von zur Mühlen sparked many discussions during his presentation on transborder searches.

Overall, the seminar was well-received by the participants. They particularly liked the interdisciplinary approach, which is documented by the results of the final Dagstuhl survey: Almost all participants stated that the seminar led to "insights from neighboring fields or

communities" and that they made "new professional contacts like an invitation to give a talk or to join an existing project or network".

The organizers also identified room for improvement: Only about one-third of the participants came from law. This points to a fundamental problem for future seminars since — similar to participants from industry — it is rather untypical for academics in law or for international practicioners to spend an entire week at a seminar or workshop.

In possible future seminars, the set of relevant topics should been broadened to include legal aspects of IT forensics in enterprises. This would substantially enlarge the set of interested international academics and further nourish community building which is currently vital to the field.

10.7 Secure Computing in the Cloud

Seminar No. **11491**
Organizers: Benny Pinkas, Ahmad-Reza Sadeghi, and Nigel P. Smart
04.–09. December, 2011 – www.dagstuhl.de/11491

Authors: Stefan Nürnberger, Benny Pinkas, Ahmad-Reza Sadeghi, and Nigel P. Smart

Introduction

Cloud computing offers IT resources, including storage, networking, and computing plat-forms, on an on-demand and pay-as-you-go basis. The high usability of today's cloud computing platforms makes this rapidly emerging paradigm very attractive for customers who want to instantly and easily provide web-services that are highly available and scal-able to the current demands. In the most flexible and general cloud computing model (*"Infrastructure as-a Service", IaaS*), customers are able to run entire Virtual Machines (VMs) inside the Cloud. VM images function as templates from which a virtually unlimited number of VM instances can be instantiated.

Problem Description

Due to virtualisation, limited physical resources are made available for masses. The sharing of these resources and the complex configuration and maintenance of the needed infrastructure is accompanied by security threats [2, 10]. According to the Cloud Security Alliance (CSA), the major inhibitor of a widespread adaptation of cloud computing is the protection of data [4], as data is no longer under the physical control of the owner (in this case the cloud customer). The cloud provider has access to data stored on disks and data transferred through the cloud network. The fact, that the physical hardware of the cloud is shared with other customers, potentially with adversaries, further stresses the need to protect data in order to thwart the lack of physical control over the own data. Moreover, the outsourced computations must be entrusted to the cloud service provider and face the risk of

Sloppy/Lazy provider: A provider that makes mistakes or simplifies computations. The sloppy and lazy provider might compromise the integrity of the result of computations. Verification of results would be a countermeasure here, for example by executing the computations on multiple, independent clouds.

Greedy provider: A provider which reduces security in order to save money. Greedy providers are willing to violate policies for economic reasons, thereby exposing the data to insider or outsider threats.

Malicious Tenant: A cloud customer (tenant) who is deliberately exploits security vul-nerabilities to gain access to data or intellectual insight of processes and computations.

The CSA recommends the use of encryption to protect data in transit and data at rest. However, cryptography in the cloud faces two problems:

1. cryptographic keys in a running VM instance are susceptible to run-time attacks like web server exploits, and
2. key provisioning to a VM is not feasible when we assume the cloud provider has access to data and VM images stored on disk.

Seminar Topics

The participants of this seminar were mainly concerned with the privacy of computation or data with respect to the cloud provider. From concrete examples like doctor-patient-confidentiality while processing genomic data at a third party [5, 6], to generic solutions that hide computations that are done at the cloud provider from the cloud provider itself [9]. Additionally, means to verify the result of an outsourced computation with significantly less computational effort than performing the calculation itself [1, 8, 7, 3]. And last, but not least, even outlooks to ad-hoc clouds that are formed by mobile devices on-demand.

References

[1] M. Atallah, K. Pantazopoulos, J. Rice, and E. Spafford. Secure outsourcing of scientific computations. *Advances in Computers*, 54:216–272, 2001.

[2] S. Bleikertz, M. Schunter, C.W. Probst, D. Pendarakis, and K. Eriksson. Security Audits of Multi-tier Virtual Infrastructures in Public Infrastructure Clouds. *Proceedings of the 2010 ACM workshop on Cloud computing security workshop*, 2010.

[3] K. Chung, Y. Kalai, and S. Vadhan. Improved delegation of computation using fully homomorphic encryption. In *CRYPTO'10*, volume 6223, pages 483–501. 2010.

[4] Cloud Security Alliance (CSA). Top threats to cloud computing, version 1.0. `http://www.cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf`, March 2010.

[5] M. Franz, B. Deiseroth, K. Hamacher, S. Jha, S. Katzenbeisser, and H. Schröder. Secure computations on real-valued signals. In *IEEE Workshop on Information Forensics and Security (WIFS'10). IEEE Press*, 2010.

[6] M. Franz, B. Deiseroth, K. Hamacher, S. Jha, S. Katzenbeisser, and H. Schroder. Towards secure bioinformatics services. In *Financial Cryptography and Data Security: 15th International Conference, FC 2011, Gros Islet, St. Lucia, February 28-March 4, 2011, Revised Selected Papers*, volume 7035, page 276. Springer-Verlag New York Inc, 2012.

[7] R. Gennaro, C. Gentry, and B. Parno. Non-interactive verifiable computing: outsourcing computation to untrusted workers. In *CRYPTO'10*, volume 6223, pages 465–482. 2010.

[8] S. Hohenberger and A. Lysyanskaya. How to securely outsource cryptographic computations. In *TCC'05*, volume 3378, pages 264–282, 2005.

[9] M.O. Rabin. How to exchange secrets by oblivious transfer. Technical report, Technical Report TR-81, Harvard Aiken Computation Laboratory, 1981.

[10] T. Ristenpart, E. Tromer, H. Shacham, and S. Savage. Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds. In *CCS'09*, pages 199–212, 2009.

10.8 Secure Architectures in the Cloud

Seminar No. **11492**
Organizers: Sabrina De Capitani di Vimercati, Wolter Pieters, Christian W. Probst, and Jean-Pierre Seifert
4.–9. December, 2011 – www.dagstuhl.de/11492

Authors: Wolter Pieters, Christian W. Probst and Sabrina de Capitani di Vimercati

Introduction

In cloud computing, data storage and processing are offered as a service, and the data resides outside the control of the owner. It is often argued that clouds improve security, as the providers have more security expertise than their (smaller) customers. However, despite theoretical breakthroughs in cryptography, there is little consensus on how we can provide architectural solutions guaranteeing that cloud data remains confidential, uncorrupted, and available. Also, it is unclear to what extent parties can be held accountable in case something goes wrong. In seminar *11492 Secure Architectures in the Cloud*, we searched for architectures, modelling approaches, and mechanisms that can help in providing guarantees for cloud security. The main question was which cloud-specific security architectures should and could be devised, and how they can be matched to security policies. The seminar was attended by researchers from different academic and industry communities, making it possible to propose integrated solutions and research directions that transcend disciplines. Four main topics have been the subject of this seminar (see also [14]):

1. *Data protection*. Data outside the data owner's control implies that privacy and even integrity can be put at risk. Guaranteeing the privacy and integrity of the data, whether stored in the system or communicated to external parties, becomes a primary requirement, and has raised the attention of both individuals and legislators. Cloud providers have to properly protect the privacy of (possible sensitive) information when storing, processing or sharing it with others [19], and have to adopt adequate access control solutions for enforcing selective access to the data. New approaches have emerged for identifying persons and roles and linking them to access privileges, such as identity-, attribute-, claims-, and data-based access control (e.g., [7, 15]). We discussed challenges of the cloud to the notions of privacy, accountability and user empowerment, their legal, ethical, and architectural implications, and possible solutions.

2. *Simulating physical constraints in the cloud*. In the cloud, we cannot easily enforce where data is stored and how long, and from where it is accessed. Location-based access control aims at limiting access to specific locations, thereby seemingly putting physical limitations back in place [23]. Measures proposed include use of GPS, trusted platform modules (TPMs), but also physically unclonable functions (PUFs) [21]. Also, data could be moved away from attacks [17]. With respect to time, mechanisms have

been proposed to assure deletion of data in the cloud [9, 12, 22]. We assessed to which extent these approaches are sufficient to simulate physical constraints, and which architectural solutions are needed to make such forms of assurance possible in practice.

3. *Misuse detection.* Many methods have been proposed for intrusion detection, penetration testing and digital forensics. Are these sufficient for cloud environments? The seminar identified necessary adaptations to system and threat models as well as security metrics, to adequately indicate which attacks are possible and which are actually happening, and thereby reduce cybercrime.

4. *Splitting the clouds.* Public clouds, containing data from different parties, are not deemed suitable for particularly sensitive information. This means that decisions will have to be made about which data to put in the cloud and which data not, which security properties to outsource and which not, and how to make sure that the entire system conforms to the security requirements (cf. [4]). The seminar investigated suitable architectures for "splitting the clouds". For example, in "security-as-a-service", not only IT infrastructure is rented, but also the security that is added to it. For authentication this seems to work pretty well, but how far can this concept be stretched to other security properties such as confidentiality and integrity?

Processing encrypted data was discussed in the parallel seminar *11491 Secure Computing in the Cloud*. This report covers the results of the seminar on Secure Architectures in the Cloud, abstracts of presentations, and proceedings of the working groups. The topics have been restructured during the seminar, and we will refer back to the topics originally proposed where appropriate. Several follow-up initiatives have been assigned to the participants.

Main Findings

As a general observation, we concluded that clouds require a different kind of architectural decisions than traditional information systems. In complex systems such as clouds, we cannot do lots of things manually anymore. For example, there is usually no way to inspect a cloud for evidence manually after an incident. This means that the architecture needs to allow for automation of such tasks, by providing not only functional services, but also meta-services to perform automated maintenance, recovery, etc. Moreover, the processes that make use of such meta-services need suitable architectures themselves. In particular, the following meta-services are needed:

- Automated policy checking,
- Automated configuration verification,
- Automated incident management,
- Automated auditing, and
- Automated forensics.

These processes could be deployed again in (different) clouds, but then the same security concerns apply to them as well.

In this sense, the cloud paradigm begs the question whether we can do everything as a service (XaaS). The participants came up with many different XaaS concepts. In particular, we proposed the concept of verification-as-a-service, which can refer to both the verification of the results of computations, as well as the verification of the (security) architecture and configuration in place at the cloud provider. The former is well-known in the field of electronic voting systems (cf. [20]); the latter resonates with the practice of security auditing. Verification-as-a-service is the main focus in relation to original topics 3 and 4. Specific challenges relate to the verification of negative properties (something is not the case in the architecture) and verification of the results of randomised algorithms. Also, testing-as-a-service could be employed to test functional and non-functional properties of cloud services.

As an instance of simulating physical constraints (topic 2) in relation to verification-as-a-service, we discussed the possibilities for verifying the location of data in the cloud (cf. [23]). One proposal is to integrate GPS with trusted hardware (such as TPM) to prove locations.

Verification-as-a-service provides a paradigm to organise accountability in the cloud. This could be realised by different techniques, for example by:

- Transparency of architecture/configuration (inspection/attestation),
- Forensics (e.g., watermarking),
- Regulation (precaution) and enforcement,
- Incident response (logging), or
- Creating incentives.

Verifying the *integrity* of data seems to be more intuitive than verifying its *confidentiality*. With integrity, it is possible, for example, to compare two different copies. With confidentiality, one would have to prove that only certain parties possess a copy. It only seems to be possible to falsify this after the fact, when it is indeed discovered that data has been leaked. Even in that case, one would need some kind of watermark to prove who leaked the data, for it might have been the user as well as the provider. How to develop a service that provides such watermarking in relation to confidentiality-as-a-service has been identified as an open problem, in relation to topic 4 (splitting the clouds).

Especially on the user side, accountability can be further enhanced by modifiability, or customisability, which allows the user to adapt services to his or her own policies. This requires negotiation on policies, not only between the user and the initial provider, but also between providers within the supply chain (cf. [24]). Customisability is the main topic discussed in relation to original topic 1 (data protection). Again, special services can be set up that allow the user to achieve this for multiple cloud services at the same time, which would amount to modifiability-as-a-service. Such services could be standardised to make sure that they really empower the user, by employing certain privacy policies themselves, and providing an understandable interface (cf. [6, 10]). We would then have achieved "standardised customisability".

We formulated several attacker models that lie behind these proposals. Many standard attacker models are problematic in the cloud. An evil/malicious cloud service provider implies that we cannot solve anything without advanced encryption methods, which are

costly or even infeasible in many scenarios. Assuming that computations are performed in the clear, we have to assume that the cloud service provider is *indifferent*, not curious. Thus, we trust the cloud provider on the issue of confidentiality, in the sense that we do not expect the provider to leak or misuse data intentionally. However, the provider may still be a:

- Sloppy provider (makes mistakes),
- Lazy provider (simplifies computations), or
- Greedy provider (reduces security to save money).

The sloppy and lazy provider might compromise the integrity of the result of computations. Verification of results would be a countermeasure here, for example by executing the computations on multiple, independent clouds.

Greedy providers are willing to violate policies for economic reasons, thereby exposing the data to insider or outsider threats. Although we do not assume malice on the side of the provider, we do assume malice on the side of other cloud users, who may or may not have specialised access (e.g., administrators). In relation to the greedy provider, one would want to have some means to verify the architecture in place.

Especially if services have been customised, one would want to have some kind of assurance that there is actually a change in configuration taking place based on the customisation. We proposed the development of a tool suite to support remote measurements of architectural variables, which would include existing proposals (cf. [2, 18, 25]). Care needs to be taken that acquiring such information does not violate customer privacy or company property rights [3]. Also, even if the architecture would be (partly) known, the user would then need meaningful support to choose among different providers (and thereby different architectures). This provides another incentive to develop quantitative models that can indeed calculate overall security risks from system architectures, based on existing qualitative approaches [1, 8, 11, 13, 16]. The user can then compare risks and costs to make decisions [5]. Such decisions could even be made in real-time based on information on the current security situation, leading to what has been called fluid information systems [17].

A remaining question is how to create incentives to invest in cloud security. If there is no immediate impact, investments may lag behind with respect to threat levels. Ironically, you can gain a competitive advantage by making your competitors invest in security. Do we really need big scandals to improve security? In any case, achieving more security by (self-)regulation, whether by law, seals, or otherwise, requires architectures such as proposed here, for it is impossible to impose constraints if they cannot be verified.

In conclusion, this seminar proposed architectures for verifying the results of cloud computations, verifying the configuration of cloud architectures, and supporting customisability of cloud services in terms of security. These were defined in relation to cloud-specific attacker models. Visual representations of the proposed architectures can be found under the results of the working groups. Open problems are defined at the end of the report.

References

[1] S. Bleikertz, T. Groß, and S. Mödersheim. Automated verification of virtualized infrastructures. In *CCSW'10: Proceedings of the 2010 ACM Cloud Computing Security Workshop*, pp. 47–58. ACM, 2010.

[2] S. Bleikertz, M. Schunter, C.W. Probst, D. Pendarakis, and K. Eriksson. Security audits of multi-tier virtual infrastructures in public infrastructure clouds. In *CCSW'10: Proceedings of the 2010 ACM Cloud Computing Security Workshop*, pp. 93–102. ACM, 2010.

[3] T.D. Breaux and C.B. Lotrionte. Towards a privacy management framework for distributed cybersecurity in the new data ecology. In *HST'11: Proceedings of the IEEE International Conference on Technologies for Homeland Security*, pp. 6–12. IEEE, November 2011.

[4] S. Bugiel, S. Nürnberger, A.-R. Sadeghi, and T. Schneider. Twin clouds: Secure cloud computing with low latency. In B. De Decker, J. Lapon, V. Naessens, and A. Uhl, editors, *Communications and Multimedia Security*, vol. 7025 of *Lecture Notes in Computer Science*, pp. 32–44. Springer Berlin / Heidelberg, 2011. DOI: 10.1007/978-3-642-24712-5_3.

[5] Y. Chen and R. Sion. To cloud or not to cloud?: musings on costs and viability. In *Proceedings of the 2nd ACM Symposium on Cloud Computing*, SOCC'11, pp. 29:1–29:7, New York, NY, USA, 2011. ACM.

[6] L. Coles-Kemp and E. Kani-Zabihi. On-line privacy and consent: a dialogue, not a monologue. In *Proceedings of the 2010 workshop on New security paradigms*, NSPW'10, pp. 95–106, New York, NY, USA, 2010. ACM.

[7] S. De Capitani di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi, G. Pelosi, and P. Samarati. Encryption-based policy enforcement for cloud storage. In *ICDCSW'10: IEEE 30th International Conf. on Distributed Computing Systems Workshops*, pp. 42–51. IEEE, 2010.

[8] T. Dimkov, W. Pieters, and P.H. Hartel. Portunes: representing attack scenarios spanning through the physical, digital and social domain. In *Proc. of the Joint Workshop on Automated Reasoning for Security Protocol Analysis and Issues in the Theory of Security (ARSPA-WITS'10). Revised Selected Papers, Paphos, Cyprus*, vol. 6186 of *LNCS*, pp. 112–129, Berlin, March 2010. Springer Verlag.

[9] R. Geambasu, T. Kohno, A.A. Levy, and H.M. Levy. Vanish: increasing data privacy with self-destructing data. In *Proceedings of the 18th conference on USENIX security symposium*, SSYM'09, pp. 299–316, Berkeley, CA, USA, 2009. USENIX Association.

[10] L. Jedrzejczyk, B.A. Price, A. Bandara, and B. Nuseibeh. "privacy-shake": a haptic interface for managing privacy settings in mobile location sharing applications. In *MobileHCI'10: Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services*, pp. 411–412, New York, NY, USA, September 2010. ACM.

[11] B. Kordy, S. Mauw, S. Radomirović, and P. Schweitzer. Foundations of attack–defense trees. In *Formal Aspects of Security and Trust, 7th International Workshop, FAST 2010*, vol. 6561 of *LNCS*, pp. 80–95. Springer, 2011.

[12] R. Perlman. The ephemerizer: making data disappear. Technical Report SMLI TR-2005-140, Sun Microsystems, Inc., Mountain View, CA, USA, 2005.

[13] W. Pieters. Representing humans in system security models: An actor-network approach. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 2(1):75–92, 2011.

[14] W. Pieters. Security and privacy in the clouds: a bird's eye view. In S. Gutwirth, Y. Poullet, P. De Hert, and R. Leenes, editors, *Computers, Privacy and Data Protection: an Element of Choice*, pp. 445–457. Springer, Dordrecht, 2011.

[15] W. Pieters and Q. Tang. Data is key: introducing the data-based access control paradigm. In E. Gudes and J. Vaidya, editors, *Data and Applications Security 2009, Montreal, Canada*, vol. 5645 of *Lecture Notes in Computer Science*, pp. 240–251. Springer, 2009.

[16] C.W. Probst and R.R. Hansen. An extensible analysable system model. *Information security technical report*, 13(4):235–246, 2008.

[17] C.W. Probst and R.R. Hansen. Fluid information systems. In *Proceedings of the 2009 workshop on New security paradigms workshop*, pp. 125–132. ACM, 2009.

[18] C.W. Probst, M.A. Sasse, W. Pieters, T. Dimkov, E. Luysterborg, and M. Arnaud. Privacy penetration testing: How to establish trust in your cloud provider. In S. Gutwirth, R. Leenes, P. De Hert, and Y. Poullet, editors, *European Data Protection: In Good Health?*. Springer, Dordrecht, 2012.

[19] J. Ruiter and M. Warnier. Privacy regulations for cloud computing: Compliance and implementation in theory and practice. In S. Gutwirth, Y. Poullet, P. De Hert, and R. Leenes, editors, *Computers, Privacy and Data Protection: an Element of Choice*, pp. 361–376. Springer Netherlands, 2011. DOI: 10.1007/978-94-007-0641-5_17.

[20] P.Y.A. Ryan, D. Bismark, J. Heather, S. Schneider, and Z. Xia. Prêt à voter: a voter-verifiable voting system. *IEEE Transactions on Information Forensics and Security*, 4:662–673, 2009.

[21] G.E. Suh and S. Devadas. Physical unclonable functions for device authentication and secret key generation. In *Proceedings of the 44th annual Design Automation Conference*, DAC'07, pp. 9–14, New York, NY, USA. ACM, 2007.

[22] Q. Tang. Timed-ephemerizer: Make assured data appear and disappear. In F. Martinelli and B. Preneel, editors, *Public Key Infrastructures, Services and Applications*, vol. 6391 of *LNCS*, pp. 195–208. Springer Berlin / Heidelberg, 2010. DOI: 10.1007/978-3-642-16441-5_13.

[23] A. van Cleeff, W. Pieters, and R. J. Wieringa. Benefits of location-based access control:a literature study. In *Proceedings of the 3rd IEEE/ACM International Conference on Cyber, Physical and Social Computing (CPSCom 2010), Hangzhou, China*, pp. 739–746, Los Alamitos, CA, November 2010. IEEE Computer Society.

[24] M. Winslett, T. Yu, K.E. Seamons, A. Hess, J. Jacobson, R. Jarvis, B. Smith, and L. Yu. Negotiating trust in the web. *IEEE Internet Computing*, 6(6):30–37, 2002.

[25] W. Xu, X. Zhang, H.n Hu, G.-J. Ahn, and J.-P. Seifert. Remote attestation with domain-based integrity model and policy analysis. *IEEE Transactions on Dependable and Secure Computing*, 99(PrePrints), 2011.

10.9 Privacy and Security in Smart Energy Grids

Seminar No. **11511**
Organizers: Stefan Katzenbeisser, Klaus Kursawe, Bart Preneel, and Ahmad-Reza Sadeghi
18.–21. December, 2011 – www.dagstuhl.de/11511

Authors: Klaus Kursawe, Stefan Katzenbeisser, Bart Preneel, Ahmad-Reza Sadeghi

The smart grid initiative is an attempt to improve reliability and efficiency of the electricity grid by adding communication and intelligence to its components all the way from end-user devices to the utilities. On the end user side, detailed usage information will be transferred to both home systems and the utilities; the utility can provide load- and pricing information to the meters and end-devices in real time. On the grid side, intelligent systems will allow for a more flexible energy distribution. Naturally, adding smartness to a critical and sizeable infrastructure system such as the electricity grid imposes extreme requirements on security and privacy, while facing numerous conflicting requirements from the different players. In addition, legislation is pushing hard to implement a large scale smart grid in a very short time: In Europe, the commission plans to achieve 80% smart grid coverage by 2020, with some countries starting to roll out meters at a large scale in 2012; in the US, the rollout has already started.

In such a setting, security and privacy are vital. A security breach of a smart energy grid can have severe consequences for power availability. With respect to privacy, the information gathered by the utility reveals a wealth of information about individual customers: examples are the day rhythm (power consumption data may reveal that a customer always comes home after the bars close, and has too little time between getting up and leaving the house to have breakfast), religious patterns (a devout Muslim may turn on the light for a morning prayer, or a catholic family may always leave home during the Sunday sermon), relationship patterns (energy usage may identify the days on which a group of people stayed in a house and the time when they went to bed), and even TV schedules (by combining electricity and water consumption measurements).

While it is not clear yet to which extent this data is going to be exploited, the potential privacy implications are substantial and have already been identified (after interoperability) as the second most important issue with the smart grid by NIST.

It is thus essential to build security and privacy protection into smart energy grids right from the start. The goal of this seminar was thus to raise awareness of this critical problem that may affect every European citizen within a couple of years and to bring together academic researchers as well as utility experts in order to start an open dialogue on smart grid privacy and security problems and potential solutions.

Topics covered during the seminar were:

- **Communication Security.** For the smart grid to work efficiently, end-user devices will need to communicate with the utility. The main challenge is that the end devices may be extremely limited in their capacity, and that commissioning—i.e., integration

of a new device into a home- or office network—has to be simple and efficient. This will require new ways of secure communication between power consuming devices and smart meters as well as new ways to set up communication networks covering extremely small devices (such as light bulbs).

- **Privacy.** The amount of data collected about individual users in a smart grid setting is unprecedented, and leads to massive concerns about user's privacy. The setting is rather unique for privacy research: the data is not gathered for the profit of some company, but for the more noble cause of global energy savings, and the nature of the system makes it hard to temporarily opt out. Flexible Privacy-Enhancing Technologies are required to balance the conflicting requirements of privacy and data usage.

- **Implementation Security.** Already now, the first attacks on implementations of smart meters have been published. With a huge number of small embedded devices suddenly getting connected, implementation security becomes critical. Unfortunately, vendors of those devices are usually not experienced in protecting against network-based attacks, and resource constraints on such devices do not allow implementation of many standard security solutions designed to protect larger computer systems. Thus, new hardware security mechanisms are required.

- **Grid Architectures.** The smart grid combines architectural requirements that are inherently contradictory. On one side, control networks for critical systems should always put safety first, i.e., rather risk a data loss than a disruption in functionality. On the other side, this particular network deals with a huge amount of privacy related and security critical data, requiring adequate protection from data theft. New architectures need to be designed to accommodate both privacy and dependability at the same time.

Subject Area 11

Data Bases, Information Retrieval, Data Mining

11.1 Challenges in Document Mining

Seminar No. **11171**
Organizers: Hamish Cunningham, Oren Etzioni, Norbert Fuhr, and Benno Stein
25.–29. May, 2011 – www.dagstuhl.de/11171

Authors: Hamish Cunningham, Norbert Fuhr, Benno Stein

About Document Mining

Document mining is the process of deriving high-quality information from large collections of documents like news feeds, databases, or the Web. Document mining tasks include cluster analysis, classification, generation of taxonomies, information extraction, trend identification, sentiment analysis, and the like. Although some of these tasks have a long research history, it is clear that the potential of document mining is still to be fully realised.

Part of the problem is that relevant document mining techniques are often applied in an isolated manner, addressing – from a user perspective – only a part of a task. For example, an intelligent cluster analysis requires adequate document models (from information retrieval) that are combined with sensible merging algorithms (from unsupervised learning), complemented by an intuitive labelling (from information extraction, natural language processing).

The deficit that we observe may also be understood as a lack of application and user orientation in research. For example, given a result set clustering task, users expect:

1. as many clusters as they identify topics in the result set,
2. that the documents within each cluster are semantically similar to each other, and
3. that each cluster is labeled intuitively.

In order to achieve such a satisfying solution, the state-of-art of concepts and algorithms from information retrieval, unsupervised learning, information extraction, and natural language processing have to be combined in a user-focussed manner.

Goals of the Seminar

The general idea was to to take an overview of the state of the art in document mining research and to define a research agenda for further work. Since document mining tasks are not tackled by a single technology, we wanted to bring a sample of the leading teams together and look at the area from a multidisciplinary point of view. In particular, the seminar should focus on the following questions:

- What are the relevant document mining tasks? The expectations and the potential for document mining changed significantly over time. Influential in this connection is

the discovery of the enormous contributions of users to the Web, among others in the form of blogs, comments and reviews, as highly valuable information source.

- What are the options and limitations of cluster analysis? A major deal of cluster analysis research has been spent to merging principles and algorithms; today, and especially in document mining, the focus is on tailored document models, user integration, topic identification and cluster labelling, on the combination with retrieval technology (e.g. as result set clustering). Especially non-topical classification tasks attracted interest in this connection, such as genre classification, sentiment analysis, or authorship grouping. Moreover, theoretical foundations of cluster analysis performance in document mining as well as commonly accepted optimality measures are open questions.

- What are the document mining challenges from a machine learning perspective? A crucial constraint is the lack of sufficient amounts of labelled data. This situation will become even more unbalanced in the future, and current research—to mention domain transfer learning and transductive learning—aim at the development of technology to exploit the huge amount of unlabelled data to improve supervised classification.

- How will NLP and IE affect the development of the field? The use of NLP and IE in document mining is a success factor of increasing importance for document mining. NLP contributes technology for document modelling, style quantification, document segmentation, topic identification, and various information extraction and semantic annotation tasks. In this regard authorship and writing style modelling is still coming of age; this area forms the heart for high-level document mining tasks such as plagiarism analysis, authorship attribution, and information quality assessment.

- Are new interaction paradigms on the rise? Interface design and visualization are very important for effective user access to the output of the document mining process. Moreover, interactive document mining approaches like e.g. scatter-gather clustering pose new challenges for both the interface and the backend.

- How to evaluate and compare the different research efforts? Evaluation is essential for developing any kind of data mining method. So far, mainly system-oriented evaluation approaches have been used, where the data mining output is compared to some "gold standard". There is a lack of user-oriented evaluations (e.g. observing users browsing a cluster hierarchy), that also take into account the tasks the users want to perform—e.g. using Borlund's concept of simulated work tasks.

Seminar Organization

To stimulate debate and cross-pollination we scheduled a mixture of of talks, working groups and demos. Following Dagstuhl tradition, the talks were characterized by interactive discussions and provided a platform for presenting and discussing new ideas. The working group topics were arrived at by a brainstorming session. Due to Easter Monday we had only four days for our seminar and shifted parts of the program to the evening.

Selected Results

This week showed that there is a number of recurring themes that are addressed by different researchers:

1. The processing hierarchy: Classic methods in document mining deal with document clustering and classification, thus regarding documents as a whole (or an "atomic" unit). Recently, researchers have become interested in deeper analyses of texts, such as sentiment analysis and the extraction of entities and relations.
2. Unsupervised vs. supervised methods: The former can be applied easily, but often lead only to modest results. Supervised methods produce more valuable results, but require large training sets for generating high-quality output. However the two approaches are not real alternatives: there are various attempts for their combination, like e.g. using prior knowledge for improving clustering, or using unclassified data with clustering for classification.
3. Whereas most supervised methods are strongly domain-dependent, there are now attempts for developing more domain-independent or cross-domain methods that can be applied more universally.
4. User feedback and user interaction has become an important component of document mining: There are many approaches aiming at better visualizations of the mining results. More recently, visual analytics methods have become popular, which aim at supporting the user during the mining process itself, thus incorporating the user's knowledge in the actual analysis (and not only during the training stage or for result presentation).

The week also showed (or confirmed) deficiencies in document mining and pointed to future research directions:

1. The multiplicity of retrieval questions, mining solutions, test corpora and evaluation measures (to mention only a few determinants) emerges naturally when satisfying individual information needs in our information-flooded society. However, this multiplicity hinders the comparison of solutions and hence the consequent improvement of the most promising technology. What is the ideal research infrastructure to exploit synergies?
2. That we suffer from an information overload is a commonplace. We ask: A single person has no chance to process a substantial part of the information at our disposal—but a machine can. How can be we benefit from this fact?
3. Current retrieval and mining technology is text-centered. The question is if and how the respective machinery can be applied to complex objects and artificially generated data: Which elements of the state-of-the-art retrieval technology is of generic type, and, can we develop a retrieval theory for complex structures?

First answers and arguments related to these questions can be found in the working groups section of the report.

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

Seminar No. **11361**
Organizers: Markus Schneider, Gottfried Vossen, and Esteban Zimányi
4.–9. September, 2011 – www.dagstuhl.de/11361

Authors: 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 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.

11.3 Foundations of distributed data management

Seminar No. **11421**
Organizers: Serge Abiteboul, Alin Deutsch, Thomas Schwentick, and Luc Segoufin
16.–21. October, 2011 – www.dagstuhl.de/11421

Authors: Serge Abiteboul, Alin Deutsch, Thomas Schwentick, and Luc Segoufin

Description of the Seminar's Topic

The Web has brought fundamentally new challenges to data management. Web data management differs from traditional database management in a number of ways. First, Web data differ in their structure: trees with links (usually described by mark-up languages such as XML) instead of tables. Also, Web data are by nature distributed, often on a large number of autonomous servers. Finally, Web data are typically very dynamic and imprecise.

Unlike for the classical relational database model, there is still no commonly accepted model for data management over the Web. The lack of a clean, simple, mathematical model further prevents us from designing general solutions to typical data management problems, such as building indexes, optimizing queries, and guaranteeing certain properties of applications.

As witnessed by the two seminars that previously occurred in Dagstuhl on this topic (Seminar 01361 in 2001 and Seminar 05061 in 2005, both entitled "Foundations of Semistructured Data"), most of the recent research efforts have concentrated on adapting traditional database techniques to the XML setting. In particular, foundational research on XML focused on the tree structure of XML documents, applying well-developed techniques based on logic and automata for trees. These lines of research have been very successful. However, they do not address all the facets of Web data. In particular distribution, dynamicity, incompleteness and reliability had received limited attention in past work, but play a central role in a Web setting. The aim of Seminar 11421 was to bring together researchers covering this spectrum of relevant areas, to report on recent progress in terms of both results as well as new, relevant research questions. It was organized at the initiative of members of the EU funded research projects FoX (fox7.eu) and Webdam (webdam.inria.fr) that are acknowledged for their support.

The seminar focused on the following key aspects of Web data management.

Semistructured data/query languages, with particular emphasis on XML/XPath and RDF/SPARQL. Semistructured data is the preferred way to organize information on the Web, and de facto and de jure standards are emerging. The study of XML data management and XML query languages remains a constant in the entire line of seminars culminating with 11421. Four notable additions over the predecessor seminars deserve mentioning. First is the emphasis on XML with Data values (and in its simplified version, on Data Words), where the data labeling of an XML tree (a word) is drawn from an infinite domain.

Query evaluation, and static analysis tasks in this setting are considerably harder, often undecidable, and finding the right limitations of the logics used for querying and of the data models is still an open research challenge. Second is the RDF data model, with its associated query language SPARQL. They are used for modeling/querying semantic Web data (ontologies), but also classical semistructured data. Since the standardization process is still ongoing, the work performed by researchers in our extended community has significant potential for impact. Indeed, one of the seminar participants, Marcelo Arenas, sits on the standard working group and is a leader in studying the semantics, query evaluation complexity, as well as optimization potential of the SPARQL language. Third is the emphasis on static typing. This is applied to XML data, schema inference, and the experimental evaluation of large collections of XML schemas found in real life (the work conducted in the framework of the above mentioned project FoX is relevant). In addition, the new SPARQL language is in need of foundational contributions towards type inference. Fourth is the work on related languages which, while not being under consideration as official standards, have established themselves as quasi-standards for querying graph databases in the database theory community. These include variations on the language of regular path queries, in which reachability queries in the graph are expressed using various classes of regular expressions over the alphabet of edge labels.

Incomplete and Probabilistic Databases. Information found on the Web is often incomplete, or uncertain due to contradictory facts across distinct data sources. Blindly applying classical query evaluation techniques to such databases leads to inconsistent answers. In the past, the database community has proposed a revolutionary way to view such information, namely as a set of possible databases, sometimes with an associated probability distribution. Query evaluation becomes a more refined task, in which query results are classified as *possible*, i.e. they belong to the answer over some possible database, or *certain*, i.e. they belong to the query answer over all possible databases. When the set of possible databases is accompanied by a probability distribution, the likelihood of possible answers can be derived. Not surprisingly, query evaluation in this setting is harder than in the standard relational setting, and work on finding the trade-offs between evaluation complexity and query language expressivity is always challenging. For Web data management, with its in-flux design for the data models and query languages, answering these questions is particularly timely.

Data Exchange is concerned with the (materialized or virtual) migration of data between data sources. Since in a Web setting such data sources are likely autonomous and have distinct schemas even when modeling similar real-life concepts, it is proposed to specify declaratively how data from the source database relates to the data published into the target database. These specifications are known as *schema mappings*, and they are exploited for various tasks, ranging from actually migrating the data from source to target, to leaving the data at the source but migrating queries from the target schema to the source schema. Seminar 11421 gave particular attention to the case of data exchange for XML data (prior work confines itself mostly to relational data sources), and for incomplete data (prior work focuses solely on complete data). It also addressed the problem of inferring schema mappings from examples given by less sophisticated users, who simply associate source/target data pairs and expect a tool to automatically generate the mappings. The seminar included a tutorial by Phokion Kolaitis, co-founder of the data exchange field.

Distribution of data across sources (typically within peer-to-peer networks), as well as of the computation performed by queries, and more generally, processes on top of such data, is another prominent topic of Web data management. The seminar explored recent answers to the long-standing challenge of coming up with models of computation that enable expressive languages that are semantically clean, efficiently executable and nevertheless admit automatic optimization. The above mentioned, highly visible, European research project Webdam proposes a vision inspired by the quintessentially declarative Datalog language from classical relational database research. A notable related approach is motivated by the area of declarative networking, which has gained the attention of the systems community in past years, and more recently of the theory community, which is now carrying out foundational research to complement and enhance the existing systems contributions. Such models as the relational transducer networks are being proposed to formalize famous (but so far informally stated) conjectures about expressivity and evaluation complexity of declarative networking programs. The seminar was also interested in general questions on Peer-to-Peer networks.

Static verification of temporal properties is key to increasing the reliability and facilitating the design of various classes of processes powered by an underlying (collection of) databases. Notable examples include electronic commerce Web sites, declarative networking programs, and general business processes. In all these cases, the underlying data is dynamic, its evolution in time governed by large collections of declarative rules, whose interference with each other and global effect are impossible to predict without automatic verification tools. Of particular interest is the verification of properties pertaining to the temporal evolution of the system, which are naturally expressed in various temporal logic flavors.

Crowdsourcing is another highly relevant recent development in the Web data management arena, one in which practice has pressed ahead of foundational work, which is now attracting the interest of the theory community. The seminar dedicated particular attention to this topic, reserving a long talk slot for a survey.

Organization of the Seminar and Activities

The workshop brought together 51 researchers from complementary areas of database theory, logic, and theoretical computer science in general, all with an established record of excellence in Web data management. The participant pool comprised both senior and junior researchers, including several advanced PhD students.

Participants were invited to present their own work, and/or survey state-of-the-art advances and challenges in the field. Thirty-four talks were given, which included four (60–90 minute) tutorials and thirty regular (30 minute) talks. All presentations were scheduled prior to the workshop, and due to the flood of volunteered talks, the organizers had to cap the number of slots. Talks were chosen so as to represent well the aspects of Web data management described above. To the organizers' pleasant surprise, some of the results established surprising bridges between fields previously seen as unrelated (such as Machine Learning and Data Exchange), and brought in techniques from novel areas (such as Nominal Sets).

Concluding Remarks and Future Plans

Due to the rich coverage of the area of foundations of Web data management, as achieved by both the presentations and the informal interactions, the organizers regard the seminar as a great success.

The weeklong format was well-suited to such an ambitious topic. The topic was well-received, as witnessed by the high rate of accepted invitations, and the exemplary degree of involvement by the participants. These volunteered such a high number of exceptional-quality talks that the organizers were faced with not being able to accommodate demand.

Bringing together researchers from different areas of data management, programming languages, theoretical computer science and logic fostered valuable interactions and led to fruitful collaborations, as reflected also by the very positive feedback from the audience.

The organizers wish to express their gratitude toward the Scientific Directorate of the Center for its support of this seminar, and hope to continue this seminar series on Web data management.

11.4 Data Mining, Networks and Dynamics

Seminar No. **11451**
Organizers: Lars Eldén and Andreas Frommer
06.–11. November, 2011 – www.dagstuhl.de/11451

Authors: Lars Eldén and Andreas Frommer

The main focus of the Dagstuhl Seminar 11451 "Data Mining, Networks and Dynamics" was the theory and the computational aspects of methods for the extraction of information in evolving networks and recent advances in algorithms for related linear algebra problems. The ever increasing size of data sets and its influence on algorithmic progress appeared as a recurrent general theme of the seminar. Some of the participants are experts in the modeling aspects, some are focusing on the theoretical analysis, and some are more directed toward software development and concrete applications. There was a healthy mix of participants of different age and academic status, from several PhD students and post-docs to senior researchers. As the subject has immediate and important applications, the seminar had some attendants with an industrial background (Yahoo and Google). Those participants also contributed greatly by introducing the academic researchers to new applications. The seminar also had several academic participants from application areas, who presented recent advances and new problems. Apart from well-known applications in social networks, search engines and biology approached from a different angle (Alter, Bast, Groh, Harb, Stumme), new applications were presented such as network methods in epidemiology (Poletto), human contact networks (Yoneki), credibility analysis of Twitter postings (Castillo), structure determination in cryo-electron microscopy (Boumal).

On the methodological side, new methods from graph theory and corresponding numerical linear algebra were presented (Bolten, Brannick, Delvenne, Dhillon, Gleich, Ishteva, Kahl, Savas, Szyld). Of particular interest is the extraction of information from extremely large graphs. Given that the class of multigrid methods is standard for solving large sparse matrix problems derived from partial differential equations, it is very natural that these methods should be tried for graph problems. In this direction, new adaptive algebraic multigrid methods for obtaining the stationary distribution of Markov chains were presented.

It has recently been recognized that optimization on manifolds (Boumal, Sepulchre) is a powerful tool for solving problems that occur in information sciences. As real world data are often organized in more than two categories, tensor methods (Alter, De Lathauwer, Eldén, Khoromskij, Sorber) are becoming a hot topic, and the talks showed that the techniques are developing so that now large problems can be treated. Tensor methods have been used for a long time for extremely large problems arising in physics. It has been conjectured that those techniques can be used also for problems in information science. Preliminary discussions along those lines took place. It is also interesting that some tensor computations are based on manifold optimization. Thus there were interesting discussions on the interplay between these areas.

The atmosphere of the meeting was very informal and friendly. During and after the talks lively discussions took place that also continued after dinner. Although it is too early to tell whether the seminar lead to new collaborations between the participants, some preliminary contacts were made. An open problem in spectral partitioning was raised (Eldén) and a preliminary solution was suggested (Gleich, Kahl).

The participants of this seminar had a chance to interact with the Dagstuhl seminar 11542 "Analysis of Dynamic Social and Technological Networks" held at the same time. Indeed, the Thursday morning session was arranged as a common session between both seminars, focusing at introducing the different methodological approaches to all particpants.

11.5 Analysis of Dynamic Social and Technological Networks

Authors: Vito Latora, Cecilia Mascolo, Mirco Musolesi

In the recent years, we have witnessed an increasing interest in the analysis of complex networks, i.e., networks composed of many interacting entities that show emergent behavior at global level. Usually, the key features of these networks can be captured by means of a statistical characterisation of their properties at local level such as their degree distribution and clustering coefficient. Models can then be built in order to describe the system in its entirety and to study the processes taking place on it, also over time. These models can be used to understand several phenomena dependent on the structure and dynamics of these networked systems, such as the spreading of computer viruses.

One of the main motivations of the explosion of interest in social networks that we are seeing is the availability of large data sets, e.g., the "snapshots" of the Internet structure or the maps of online social networks obtained by crawling extremely popular Web sites like Facebook and Twitter. Moreover, many data collections exercises of data sets related to human interactions by means of Bluetooth radio or GPS receivers have been carried in the recent years. These large data sets provide information that is not limited to a particular instant of time, but cover a very large time interval as well as fine grained space information. These data sets can be used to study the evolution of the network and dynamic processes happening over time such as simulated epidemics. Other large data sets that have attracted considerable interest include biological, commodity and economic networks. The initial research efforts have been focussed on the analysis of the static properties of these networks, including the presence of clusters, hubs, and community structure. More recently, researchers became interested in studying dynamic processes taking place on these networks such as information diffusion in the Internet, disease epidemics and malware propagation.

The goal of this seminar was broad, including both mathematical aspects and practical applications of theoretical models and techniques. The seminar focused on two key classes of networks that are of fundamental importance not only in computer science but also in the everyday life of millions of people, namely technological networks and online social networks.

The main contributions of this seminar were:

- Presenting a wide range of recent research results on the dynamics of processes and structure of technological and social networks.
- Exchanging solutions and practices in the different areas of computer science and other disciplines in order to find novel solutions and start fruitful long-term collaborations among the seminar attendants.

- Exploring the new challenges and opportunities arising from the analysis of data from mobile devices and social network tools, which offer the change to collect very rich data sets of information about the everyday life of people including their movements, their contacts and their social network.
- Discussing how the social networks extracted from mobile device interactions are time and location dependent, requiring new models and techniques to study them.
- Examining the application of machine learning and data mining techniques to the analysis of technological and social networks, bringing together researchers and practitioners working on the massive available networking data sets and machine learning experts interested in real-world problems.
- Studying and considering the computational challenges presented by the scale of these data sets, which impose the design of novel algorithms and the rethinking of existing techniques.
- Discussing the ethical problems arising from the treatment of privacy-sensitive user data and potential technical and legal solutions to overcome them.

Subject Area 12

Machine Learning

12.1 Constraint Programming meets Machine Learning and Data Mining

Seminar No. **11201**
Organizers: Luc De Raedt, Heikki Mannila, Barry O'Sullivan, and Pascal Van Hentenryck
15.–20. May, 2011 – www.dagstuhl.de/11201

Authors: Luc De Raedt, Siegfried Nijssen, Barry O'Sullivan, Pascal Van Hentenryck

Goal of the Seminar

Over the past two decades the fields of constraint programming, machine learning and data mining have become well-established research fields within computer science. They have contributed many foundational techniques that are routinely applied in real-life scientific and industrial applications. At the same time, awareness has grown that constraints can be very useful during mining and learning, and also that machine learning and data mining may allow one to automatically acquire constraints from data.

Both the data mining and machine learning communities have been interested in *constraint-based mining and learning*, that is, the use of constraints to formalize mining and learning problems. Examples are the specification of desirable properties of patterns to be mined, or clusters to be found. The task of the data mining or machine learning system is to generate all patterns or to compute the optimal clustering satisfying the constraints. A wide variety of constraints for local pattern mining, clustering and other machine learning problems exist and they have been implemented in an even wider range of specific data mining and machine learning systems for supporting such constraints. Some of these methods are based on mathematical programming techniques, such as linear programming or quadratic programming; other problems, however, cannot be modeled using these techniques. So far, the machine learning and data mining communities have been unable to develop general solvers that are applicable to a wide range of machine learning and data mining problems.

On the other hand, the artificial intelligence community has studied several types of constraint-satisfaction solvers. The most general systems are now gathered in the area of *constraint programming*. In constraint programming, the user specifies the model, that is, the set of constraints to be satisfied and constraint solvers generate solutions. Thus, the goals of constraint programming and constraint based mining and learning are similar; it is only that constraint programming targets *any* type of constraint satisfaction problem, whereas constraint-based mining and learning *specifically* targets data mining and machine learning applications. Therefore, it is surprising that despite the similarities between these two endeavours, the two fields have evolved independently of one another, and also, that – with a few recent exceptions – constraint programming tools and techniques are not yet applied to data mining and machine learning, and, vice versa, that problems and challenges from data mining and machine learning have not yet been taken up by the constraint

programming community. Exploring the possibilities for exploiting constraint programming in data mining and machine learning was one goal of this seminar.

The second goal was to study the use of machine learning and data mining in constraint programming. Practitioners of constraint programming have to formulate explicitly the constraints that underly their application. This is often a difficult task. Even when the right constraints are known, it can be challenging to formalize them in such a way that the constraint programming system can use them efficiently. This raises the question as to whether it is possible to (semi)- automatically learn such constraints or their formulations from data and experience. Again, some initial results in this direction exist, but we are away from a complete understanding of the potential of this approach.

In this seminar, we aimed at bridging the gap between these two fields by investigating, on the one hand, how standard constraint-programming techniques can be used in data mining and machine learning, and on the other hand, how machine learning and data mining can contribute to constraint programming. Therefore, this workshop brought together researchers in the areas of constraint programming, machine learning and data mining to discuss these issues, to identify interesting opportunities and challenges for research, and to consolidate and strengthen a promising line of research.

Seminar Organization

Given the various backgrounds of the participants, the seminar started with several introductory presentations. The focus was first on constraint programming and the use of machine learning and data mining in constraint programming; Helmut Simonis provided an introduction of constraint programming, while Barry O'Sullivan presented the possibilities for using machine learning and data mining in constraint programming.

Subsequently, the focus moved to data mining and machine learning and the possibilities for using constraint solvers in data mining and machine learning. Bart Goethals provided an introduction to data mining; subsequently, Ian Davidson discussed uses of constraint solving in clustering, Dan Roth uses of solvers in inference problems, Siegfried Nijssen uses of solvers in pattern mining, and Tijl De Bie uses of solvers in statistical machine learning.

The remainder of the program consisted of a mix of technical presentations as well as meetings of discussion groups, each of which focused in more detail on the various possibilities for combining machine learning, data mining and constraint programming.

The topics of the discussion groups were determined after discussion at the first day of the seminar and were the following:

- declarative data mining, to discuss questions regarding the use of constraint programming solvers and declarative modelling to solve data mining problems;
- programming languages for machine learning, to discuss questions regarding the development of languages specifically for machine learning;
- applications, to discuss applications in which machine learning, data mining and constraint programming can be used;

- challenges, to discuss the possibilities for setting up benchmark problems –both real-life and artificial– that can be used to determine the success of combining machine learning, data mining and constraint programming;
- learning to solve, to discuss the use of machine learning and data mining to improve the efficiency and quality of constraint solving;
- learning constraints, to discuss the use of machine learning and data mining to learn appropriate models from examples.

The seminar was concluded with plenary presentations in which the results of the discussion groups were summarized.

The intention is to publish the results of the seminar in an edited book.

12.2 Mathematical and Computational Foundations of Learning Theory

Authors: Matthias Hein, Gabor Lugosi, Lorenzo Rosasco, Steve Smale

The study of learning is at the very core of the problem of intelligence both in humans and machines. We have witnessed an exciting success story of machine learning in recent years. Among other examples, we now have cars that detect pedestrians, and smart-phones that can be controlled simply by our voices. Indeed, aside from the increase in computational power and availability of large amount of data, the key to these successes has been the development of efficient learning algorithms based on solid theoretical foundations. As the science and engineering of learning move forward to understand and solve richer and more articulated classes of problems, broadening the mathematical and computational foundations of learning becomes essential for future achievements.

The main goal of our seminar was to account for the newest developments in the field of learning theory and machine learning as well as to indicate challenges for the future. This seminar was in the same spirit of two very successful conferences titled "Mathematical Foundations of Learning Theory", organized in 2004 in Barcelona and 2006 in Paris. The seminar brought together leading researchers from computer science and mathematics to discuss the state of the art in learning and generate synergy effects between the different usually disconnected communities. This Dagstuhl seminar has been the first to cover the full range of facets of modern learning theory.

The seminar has focused on three main topics, while trying to keep a broader view on all recent advances. The three main topics were: 1) the role of sparsity in learning, 2) the role of geometry in learning, and 3) sequential learning and game theory. Experts in each field gave tutorials on each topic, covering basic concepts as well as recent results. The meeting was hold in a very informal and stimulating atmosphere. The participants all agreed that such a seminar should be come a regular meeting.

12.3 Learning in the context of very high dimensional data

Seminar No. **11341**
Organizers: Michael Biehl, Barbara Hammer, Erzsébet Merényi, Alessandro Sperduti, and Thomas Villmann
22.–26. August, 2011 – www.dagstuhl.de/11341

Authors: Michael Biehl, Barbara Hammer, Erzsébet Merényi, Alessandro Sperduti, Thomas Villmann

Goals of the seminar

Rapidly increasing sensor technology, greatly enhanced storage capabilities, and dedicated data formats have lead to a dramatic growth of the size of electronic data available today and, even more so, its dimensionality. Examples include diverse formats such as spectral data, micro- and macroarrays, biotechnological sequence data, or high resolution digital images. Due to its dimensionality and complexity, these data sets can hardly be addressed by classical statistical methods; nor do standard presentation and visualization tools allow an adequate direct inspection by humans. Thus, the need for efficient and reliable automatic processing and analysis tools for very high dimensional data arises in different areas such as bioinformatics, medicine, multi-band image analysis, robotics, astrophysics, geophysics, etc.

The aim of the seminar was to bring together researchers who develop, investigate, or apply machine learning methods for very high dimensional data to advance this important field of research. The focus was put on broadly applicable methods and processing pipelines which offer efficient solutions for high-dimensional data analysis appropriate for a wide range of application scenarios.

Questions tackled in the seminar included the following areas:

1. **Sparse representation and regularization:**
   (a) Which general principles (such as information theory, preservation of inherent data structures) offer suitable frameworks in which to achieve a compact representation of high dimensional data? Is it possible to turn these general principles into efficient algorithmic form as mathematical regularization conditions?
   (b) Which models are suitable to represent high dimensional data in a dense form (such as prototype based methods, functional data representation, dedicated algebraic structures, decomposition methods)? What are their adaptive parameters and how can they be adapted?
   (c) How can the number of free model parameters be restricted by regularization such that the available data provides a sufficient statistics for the resulting model? Is it possible to derive explicit mathematical bounds on the generalization ability?

2. **Dedicated metrics and kernels for high-dimensional data:**

   (a) How can the inherent non-Euclidean structure be inferred from the data in the presence of high dimensionality? Which aspects of particular relevance for the application should be emphasized by the corresponding similarity structure and how can this information be estimated with robust statistical tools?

   (b) How can this information be embedded into metrics or kernels? Do there exist particularly suited approaches for high dimensional data of specific form such as kernels which make use of sparsity or functional dependencies of the data? How can this be realized algorithmically in an efficient way regarding the high dimensionality?

   (c) Is it possible to partially automate the detection of a suitable similarity structure for the analysis tool and accompany this with guarantees such as consistency, or bounds on the generalization ability in the context of high-dimensionality?

3. **Efficient realizations:**

   (a) How can robust learning algorithms be designed for the model parameters, ones that can deal with noise and uncertain data, missing values, etc., particularly pronounced in high dimensional data? Are these techniques insensitive with respect to the choice of the metaparameters (such as learning rate, degree of regularization), such that generic methods suitable for non-experts in the field result?

   (b) How can the methods be realized efficiently in the context of very high dimensionality? Methods which are linear in the number of dimensions are probably already too slow in this context.

   (c) Can the learning algorithms be realized in such a way that adaptation to new data and life-long learning become possible? What are characteristic time scales required for learning certain parameters in dependence of the data dimensionality?

4. **Evaluation of methods:**

   (a) What are inherent evaluation criteria for the reliability of the models, also when the number of data points is small compared to the dimensionality? What are stable conformal predictors for model adequacy and accuracy?

   (b) How can the results be presented to experts such that humans can judge the reliability and quality of the model? How can, in turn, user feedback be integrated into the models?

   (c) Can simplifying models of learning scenarios give insight into the performance of practical algorithms? Which information visualization methods are suitable in this context?

Structure

39 experts from 12 different countries joined the seminar, including a good mixture of established scientists and promising young researchers working in the field. Thereby, the special interests of the researchers ranged from dedicated algorithmic design connected

to diverse areas such as dimensionality reduction, data visualization, metric learning, functional data analysis, to various application scenarios including diverse areas such as the biomedical domain, hyperspectral image analysis, and natural language processing.

This setup allowed us to discuss salient issues in a way that integrated perspectives from several points of views and scientific approaches, thereby providing valuable new insights and research contacts for the participants. Correspondingly, a wide range of topics was covered during discussions and presentations in the seminar.

During the week, 29 talks were presented which addressed different aspects of how to deal with high dimensional data and which can be grouped according to the following topics:

- Dimensionality reduction techniques and evaluation measures
- Biomedical applications
- Distances, metric learning, and non-standard data
- Functional data processing
- Probabilistic models for dimensionality reduction
- Feature selection and sparse representation of data

The talks were divided into a variety of tutorials which gave introductory overviews and opinions on important research directions and shorter talks which focused on specific recent (partially yet unpublished) scientific developments. The talks were supplemented by vivid discussions based on the presented topics as well as the traditional social event on Wednesday afternoon in the form of a visit to the beautiful town of Trier.

Results

A variety of open problems and challenges came up during the week. The following topics were identified as central issues in the context of the seminar:

- **Desired properties of dimensionality reduction techniques, possibilities of their formal evaluation:** The topic of dimensionality reduction and data visualization has been addressed in several presentations including several tutorial talks.

  It became apparent that the topic is currently a very rapidly emerging field in machine learning, with a manifold of advanced algorithms being published in the recent literature.

  Key issues, however, remain a challenge: to use advanced methods in applications, there is the need for widely parameterless techniques, clear interpretability of the results, and comfortable usage e.g. regarding processing speed or uniqueness and robustness of the results.

  For these reasons, advanced methods are often not used in practice.

  It has been discussed that the desired properties and results of dimensionality reduction techniques depend on the given task at hand and cannot universally be formalized. Nevertheless, there is a need for formal evaluation methods of dimensionality reduction to compare techniques, and to guide parameter choice and optimization.

Promising general evaluation schemes have been proposed in recent years as presented in the seminar, but an extensive evaluation of their suitability is so far lacking.

- **Good scientific benchmarks and evaluation criteria:** In this context, is has been raised that good, accessible benchmarks are rare. Albeit high dimensional electronic data are ubiquitous, these data often require complex preprocessing, they do not allow evaluation of formal methods due to the lack of objective evaluation information, or they are even sometimes subject to restrictions. For that reason, real life data are partially not accessible, and there is the risk that methods are over-adapted according to the available benchmarks which do not necessarily mirror the demands in practical applications. During the seminar, however, it has been raised that quite a few benchmarks have become available in the context of contest data.

- **Where to use complex models as compared to well-established linear techniques:** During the seminar, it became apparent that there is a gap in between advanced techniques proposed in the context of machine learning for high dimensional data and methods which are actually used in application domains such as biomedical data analysis.

  Often, in practical applications simple linear techniques seem sufficient to reliably detect important and interpretable information in high dimensional data collections. A variety of reasons has been discussed in this context: in particular in bioinformatics, the technology to gather data and large data collections are often comparably novel such that information still lies 'at the surface' of the data.

  For very high dimensional data, linear techniques sometimes seem the only methods for which sufficient reliability can be guaranteed, more complex nonlinear methods likely focusing on noise in the data due to the lack of appropriate regularizers which suite the given setting. In this context, it has been raised that the embedding of data into high-dimensions where linear methods often suffice constitutes one of the most prominent approaches to actually solve standard nonlinear problems, popular examples in this context being the support vector machine, the extreme learning machine, or reservoir computing.

  Further, linear techniques seem to focus on 'universally important' issues which are relevant independent of the context due to universal statistical properties. For more advanced techniques, domain knowledge is required to set up the models or to interpret the results in a reliable way. This argument has been substantiated in the seminar by several presentations. For example, knowledge about biological networks and metabolic pathways can be integrated into biological data processing and it can greatly enhance the performance.

- **How to deal with complex structures:** It has been raised that an intelligent preprocessing of the data is often more important than the choice of the model. Alternatively, data can be tackled with appropriate problem adapted metrics, followed by rather simple machine learning techniques. In the seminar, quite a number of complex data formats have been presented in applications, such as e.g. data with inherent functional form connected to spectrometry data. Besides the necessity to come up with suitable metrics, this also leads to very interesting theoretical problems. For example, it can be proved that it is not possible to learn in the space of functions

at all, unless very strong requirements are fulfilled.

Nevertheless, impressive applications have been achieved in this context. Thus, it seems worthwhile to investigate which constraints are fulfilled in practical applications such that learnability is guaranteed. On the other side, a variety of powerful metric adaptation schemes exist ranging from fast and efficient convex techniques to nonlinear cost functions. In most cases, however, a simple Mahalanobis distance is used and, often, only basic machine learnings are integrated such as simple k nearest neighbor. Further, a unifying theory and guidelines, which technique is best suited in which scenarios, remain unsolved problems.

- **How to model in the correct way for very high dimensions:** Statistics being the universal background for almost all machine learning techniques, it has been raised that statistical models are often rather uniform in their principled design, not taking advantage of the rather flexible way to model dependencies of data and dimensions. While it is common in the context of simple PCA to swap the role of input dimensions and data points in case of very high dimensions, more complex nonlinear models stick to the classical setting as used in the case of comparably low dimensionality. It could be worthwhile to put different principled modeling paradigms into general patterns which allow to reformulate established techniques such that they become suitable for very high dimensional data.

Altogether, the seminar opened quite a few perspectives pointing into important research directions in the context of very high dimensional data.

Subject Area 13

Bioinformatics

13.1 Formal Methods in Molecular Biology

Authors: Rainer Breitling, Adelinde M. Uhrmacher

The second Dagstuhl Seminar on Formal Methods in Molecular Biology took place from 10–15 April, 2011. 35 participants from 8 countries gathered to discuss the most recent advances in Systems Biology and the contribution of computational formalisms to the successful modeling of biological systems. Major recurrent themes were the description of stochastic phenomena in biology, the modeling of spatial aspects of cellular behavior, and the robustness of cellular switches in the face of molecular noise and uncertainty of parameter inference. The computational modeling approaches applied to these challenges were particularly diverse, ranging from differential equation-based models to various flavors of rule-based languages, Petri Nets and process algebras.

A central component of the seminar was the Second International Biomodeling Competition. Teams formed during the first day and worked on biological case studies using a variety of modeling formalisms and analysis methods; the results were presented on Thursday afternoon and the winner determined by a joint vote of the audience.

The 1st prize went to the team of Kirill Batmanov, Antje Beyer, Matthias Jeschke and Carsten Maus, for their work on 'Synchronization of cell populations'.

The 2nd prize went to the team of Andrea Bracciali, Mostafa Herajy, Pietro Lió, Chris Myers, Brett Olivier, and Natal van Riel for their work on 'A bistable gene switch'.

Special prizes were awarded to the team of Chiara Bodei, Luca Bortolussi, Davide Chiarugi, Maria Luisa Guerriero, Jane Hillston Ivan Mura, Alberto Policriti, and Alessandro Romanel (for 'Critical Analysis'), the team of Mary Ann Blätke, Qian Gao, David Gilbert, Simon Hardy, Monika Heiner, Andrzej Kierzek, Fei Liu and Wolfgang Marwan (for 'Innovative use of Petri Nets'), and the team of Maciej Dobrzyński, Mathias John, Céline Kuttler, Bartek Wilczyński and Verena Wolf (for 'A pure stochastical approach').

Subject Area 14

Applications, Interdisciplinary Work

14.1 Multimodal Music Processing

Seminar No. **11041**
Organizers: Meinard Müller, Masataka Goto, and Simon Dixon
23.–28. January, 2011 – www.dagstuhl.de/11041

Authors: Meinard Müller. Masataka Goto, Simon Dixon

Music can be described, represented, and experienced in various ways and forms. For example, music can be described in textual form not only supplying information on composers, musicians, specific performances, or song lyrics, but also offering detailed descriptions of structural, harmonic, melodic, and rhythmic aspects. Furthermore, music notation can be encoded in text-based formats such as MusicXML, or symbolic formats such as MIDI. Music annotations, metadata, and social tags are also widely available. Beside textual data, increasingly more types of music-related multimedia data such as audio, image or video data are widely available. For example, there are numerous MP3 and CD audio recordings, digitized images of scanned sheet music, and an increasing number of video clips of music performances. In this seminar we discussed and studied various aspects of the automated processing of music-related documents that vary in their modalities and formats, i.e., text, symbolic data, audio, image and video. Among others, the topics dealt with aspects of data analysis, retrieval, navigation, classification, and generation of music documents while considering music-specific characteristics, exploiting multiple information sources, and accounting for user-specific needs.

In this executive summary, we give a brief overview of the main topics addressed in this seminar. We start by briefly describing the background of the seminar participants and the overall organization. We then give an overview of the presentations and the results from working groups and panels. Finally, we reflect on the most important aspects of this seminar and conclude with future implications.

Participants, Interaction, Activities

In our seminar, we had 35 participants, who came from various countries around the world including North America (7 participants), Japan (4 participants), New Zealand, Singapore, and Europe (Austria, France, Germany, Netherlands, Portugal, Spain, United Kingdom). Most of the participants came to Dagstuhl for the first time and expressed enthusiasm about the open and retreat-like atmosphere. Besides its international character, the seminar was also highly interdisciplinary. While most of the participating researchers are working in computer science and its neighboring fields, a large number of participants also have a strong background in music and musicology. This made the seminar very special in having not only interactive and provoking scientific discussions, but also numerous social activities including common music making. One particular highlight of such social activities was a three-hour spontaneous concert on Thursday evening, where various participants performed in changing ensembles a wide variety of music including popular music, jazz, and classical music.

Overall Organization and Schedule

Dagstuhl seminars are known for having a high degree of flexibility and interactivity, which allow participants to discuss ideas and to raise questions rather than to present research results. Following this tradition, we fixed the schedule during the seminar asking for spontaneous contributions with future-oriented content, thus avoiding a conference-like atmosphere, where the focus is on past research achievements. The first day was used to let people introduce themselves and express their expectations and wishes for the seminar. We then had a brainstorming session on central topics covering the participants' interests while discussing the overall schedule and format of our seminar. In particular, we identified a total of six topics for discussion. For four of these topics, we divided into four groups, each group discussing one of the topics in greater depth in parallel sessions on Tuesday. The results and conclusions of these group meetings were then presented to the plenum on Wednesday. For the remaining two topics, we decided on having panel-like discussions within the plenum with introductory stimulus talks (Thursday). Finally, group and panel discussions were interleaved with regular sessions that allowed participants to present their personal research to the plenum. This mixture of presentation elements gave all participants the opportunity for presenting their ideas to the plenum while avoiding a monotonous conference-like presentation format.

Main Topics and Results

We discussed various topics that address the challenges of organizing, understanding, and searching music-related information in a robust, efficient, and intelligent manner. Here, a particular focus was put on music-specific aspects, the fusion of multiple information sources, as well as the consideration of user-specific needs. After a joint brainstorming session, we agreed on discussing six central topics which fitted in the overall theme of the seminar and reflected the participants' interests. We now give a brief summary of these topics, which were discussed within four parallel group meetings and two panels. Then, we give an overview of additional contributions made by the participants in the regular sessions of the seminar.

1. The *"Model"* group discussed the issue of how signal models can be developed that exploit multimodal information. Here, one main goal was to review strategies for combining different sources of information to support music analysis. In particular, various early and late fusion approaches were identified and advantages and weaknesses of the respective approaches were discussed. Particular attention was paid to the aspect of data uncertainty and its propagation in the fusion processes.

2. The *"User"* group addressed the topic of user-aware music information retrieval. Here, a central question was how contextual information can be integrated into the retrieval process in order to account for short-term user interests and long-term user behavior. Additionally, it was discussed how search engines may yield satisfying results in terms of novelty, popularity, and serendipity of the retrieved items.

3. The *"Symbol"* group discussed the question of how to bridge the gap between visual, symbolic, and acoustic representations of music. Here, particular attention was

given to the problem referred to as *Optical Music Recognition* (OMR) with the goal of converting an image-based sheet music representation into a symbolic music representation where note events are encoded explicitly. In this context, user interfaces were reviewed that allow for a synchronized presentation of visual and acoustic music content.

4. The *"Meaning"* group addressed the subject of how musical meaning can be derived from musical data and, in particular, from musical sound. Here, the path from the given low-level (acoustic) raw data to high-level musical models was traced from a human-based as well as from a machine-based perspective.

5. In the *"Ground Truth"* panel, fundamental issues related to the interpretation, usage, and generation of ground truth annotations were discussed. Some of the questions raised during the panel were: What is ground truth in music? How can one handle inter- and intra-annotator variations? How can the quality of ground truth be evaluated? Are there alternatives to manually generated ground truth annotation?

6. Finally, in the *"Grand Challenges"* panel we discussed in which way music information research may and should impact our daily lives and our society in the future. Here fundamental questions were how to provide the best music for each person, how to predict specific effects of music on our society, and how to enrich human relationships by music.

Beside the extensive discussion of these six topics, we had a number of additional contributions where participants presented more specific research results. These contributions covered a number of different topics such as audio parameterization, music alignment and synchronization, singing voice processing, crowd music listening, music tagging, music indexing, interfaces for music exercises, personalization issues in music search, analysis of ethnic music, and many more.

These topics were complemented by some more interdisciplinary contributions relating the field of music processing to neighboring fields such as speech processing, musicology, music perception, and information retrieval. For example, we discussed the ways in which the field of music processing has benefitted from older fields such as speech processing and how music processing might give something back to these fields. Furthermore, a musicologist reported on the difficulties and resistance experienced when introducing novel computer-based methods into traditional humanistic sciences such as musicology. Another highlight of our seminar was a keynote presentation given by Hannah Bast on her CompleteSearch Engine that allows for very fast processing of complex queries on large text collections.

Conclusions

In our seminar, we addressed central and groundbreaking issues on how to process music material given in various forms corresponding to different musical aspects and modalities. In view of the richness and complexity of music, there will be no single strategy that can cope with all facets equally well. Therefore unifying frameworks and fusion approaches are needed which allow for combining, integrating, and fusing the various types of information sources to support music analysis and retrieval applications. Also, to further enhance our field, one needs to understand better the complex relationships within music as well as

the complex effects of music on the human mind, thus requiring interdisciplinary research efforts. The Dagstuhl seminar gave us the opportunity for discussing such issues in an inspiring and retreat-like atmosphere. The generation of novel, technically oriented scientific contributions was not the focus of the seminar. Naturally, many of the contributions and discussions were on a rather abstract level, laying the groundwork for future projects and collaborations. Thus the main impact of the seminar is likely to take place in the medium to long term. Some more immediate results, such as plans to share research data and software, also arose from the discussions. As measurable outputs from the seminar, we expect to see several joint papers and applications for funding (e.g. to the European Union) proceeding from the discussions held at Dagstuhl.

Beside the scientific aspect, the social aspect of our seminar was just as important. We had an interdisciplinary, international, and very interactive group of researchers, consisting of leaders and future leaders in our field. Most of our participants visited Dagstuhl for the first time and enthusiastically praised the open and inspiring atmosphere. The group dynamics were excellent with many personal exchanges and common activities. Younger scientists mentioned their appreciation of the opportunity for prolonged discussions with senior researchers—something which is often impossible during conference-like events.

In conclusion, our expectations of the seminar were not only met but exceeded, in particular with respect to networking and community building. Last but not least, we heartily thank the Dagstuhl board for allowing us to organize this seminar, the Dagstuhl office for their great support in the organization process, and the entire Dagstuhl staff for their excellent services during the seminar.

14.2 Sparse Representations and Efficient Sensing of Data

Seminar No. **11051**
Organizers: Stephan Dahlke, Michael Elad, Yonina Eldar, Gitta Kutyniok, and Gerd Teschke
30. January–04. February, 2011 – www.dagstuhl.de/11051

Authors: Stephan Dahlke, Michael Elad, Yonina Eldar, Gitta Kutyniok, and Gerd Teschke

Modeling of data is the crucial point in enabling various processing of it. This modeling can take many forms and shapes: it can be done in a low-level way that ties the data samples directly or in higher levels that search for structures and constellations. The task of modeling data is so fundamental that it is underlying most of the major achievements in the fields of signal and image processing. This is true also for processing of more general data sources. Indeed, the field of machine learning that addresses this general problem also recognizes the importance of such modeling. In this realm of models, there is one that stands out as quite simple yet very important: this is a model based on sparse description of the data. The core idea is to consider the data as a sparse linear combination of core elements, referred to as atoms. This model has attracted huge interest in the past decade, with many mathematicians, computer scientists, engineers, and scientists from various disciplines working on its different facets, and building a set of tools that lead all the way from pure mathematical concepts to practical tools to be used in other computational sciences as well as applications. Using this model, researchers have shown in recent years a wide battery of computational research disciplines and applications that directly benefit from it, leading to state-of-the- art results. Various reconstruction problems, data compression, sampling and sensing, separation of signals, cleaning and purifying data, adaptive numerical schemes, and more, all require the utilization of sparse representations to succeed in their tasks.

The goals of the seminar can be summarized as follows:

- Establish communication between different focusses of research
- Open new areas of applications
- Manifest the future direction of the field
- Introduce young scientists

To reach these seminar goals, the organizers identified in advance the most relevant fields of research:

- Sampling and Compressed Sensing
- Frames, Adaptivity and Stability
- Algorithms and Applications

The seminar was mainly centered around these topics, and the talks and discussion groups were clustered accordingly. During the seminar, it has turned out that in particular 'generalized sensing', 'data modeling', and corresponding 'algorithms' are currently the

most important topics. Indeed, most of the proposed talks were concerned with these three issues. This finding was also manifested by the discussion groups.

The course of the seminar gave the impression that sparsity with all its facets is definitely one of the most important techniques in applied mathematics and computer sciences. Also of great importance are associated sampling issues. We have seen many different view points ranging from classical linear and nonlinear to compressive sensing. In particular, new results on generalized sampling show how to design effective sampling strategies for recovering sparse signals. The impact of these techniques became clear as they allow an extension of the classical finite dimensional theory of compressive sensing to infinite dimensional data models. Moreover, it was fascinating to see how sampling and sparsity concepts are by now influencing many different fields of applications ranging from image processing / compression / resolution to adaptive numerical schemes and the treatment of operator equations/inverse problems. It seems that the duality between sparse sampling and sparse recovery is a common fundamental structure behind many different applications. However, the mathematical technicalities remain quite challenging. As algorithmic issues were also discussed quite intensively, we could figure out that we are now essentially at some point where ℓ_1-optimization is competitive speed-wise with classical linear methods such as conjugate gradient.

Summarizing our findings during the seminar, we believe that the research agenda can be more focused on the actual bottlenecks, being in problem/signal modeling, design of sampling and recovery methods adapted to specific problems, and algorithmic improvements including performance bounds and guarantees.

14.3 Artificial Immune Systems

Seminar No. **11172**
Organizers: Emma Hart, Thomas Jansen, and Jon Timmis
26.–29. April, 2011 – www.dagstuhl.de/11172

Authors: Emma Hart, Thomas Jansen, Jon Timmis

Artificial immune systems (AISs) are inspired by biological immune systems and mimic these by means of computer simulations. They are seen with interest from immunologists as well as engineers. Immunologists hope to gain a deeper understanding of the mechanisms at work in biological immune systems. Engineers hope that these nature-inspired systems prove useful in very difficult computational tasks, ranging from applications in intrusion-detection systems to general optimization. Moreover, computer scientists identified artificial immune systems as another example of a nature-inspired randomized search heuristic (like evolutionary algorithms, ant colony optimization, particle swarm optimization, simulated annealing, and others) and aim at understanding their potential and limitations. While the relatively new field has its successful applications and much potential its theoretical foundation is still in its infancy. Currently there are several not well connected strands within AIS theory, not even a general agreement on what the central open problems are, and only a weak connection between AIS theory and AIS applications. The main goals of the proposed seminar include bringing together computer scientists and engineers to strengthen the connections within AIS theory, connections to other researchers working on the theory of randomized search heuristics, and to improve connectivity between AIS theory and applications.

Biological immune systems show great resilience in harsh environments and demonstrate the ability to cope with large amounts of sensory data as well as the unpredictability of the natural world. Indeed, a great deal of attention is now being paid to these aspects of the immune system by the wider computing research community.

Given the practical success of AIS, there is a serious lack of theoretical work in the area. Many AIS algorithms are based purely on clonal selection mechanisms, without any interaction between the different members of the cell populations. The dynamics of cell populations in the immune system have been modeled extensively using nonlinear dynamical systems. At present, however, there is no centrally agreed approach on how to tackle important theoretical issues in AIS. All too often theory is undertaken without the due attention to the practical implications. For theory to have a serious impact, collaboration between theoreticians and engineers is needed to identify key engineering issues, relevant theoretical issues and crucially how the theory can help support the engineering process. While starting point of the seminar and its driving force are deficits in the theoretical foundation of AIS its main goals are clearly beyond theory. At the heart of the seminar's motivation is the conviction that there is nothing more practical than a good theory.

The seminar took place from April 26th to April 29th 2011. It started with a series of talks aimed at providing a suitable level of introduction to the main areas of discussion to provide

a levelling ground for all participants. The format of the seminar was then a series of short presentations by researchers on topics that ranged from swarm robotics to immunology and theoretical frameworks for algorithm analysis. These were then followed by a series of *breakout* group sessions which focussed discussion on the issues raised by the speakers with results from those discussions being reported back to the main group at regular intervals. Towards the end of the week, a convergence into four key topics emerges: (1) The principled development of bio-inspired algorithms and how the translation from computational models into usable algorithms is managed, (2) the relationship between evolution and immunity and how it might be possible to evolve an artificial immune system in complex engineering problems, specifically swarm robotic systems, (3) the development of a definitive clonal selection algorithm with appropriate theoretical analysis and (4) the development of novel immune algorithms and the use of models from computational immunology for both the understanding of immunological processes and the development of new algorithms. These four topics are to be taken forward as journal papers by participants from the seminar.

As a result of the seminar there will be a special issue published in Natural Computing a leading journal in the area that will not only publish papers outlined above, but provide a roadmap for the future direction of AIS and serve as, it is hoped, an authoritative guide to the area of artificial immune systems.

14.4 Computer Science in Sport – Special emphasis: Football

Seminar No. **11271**
Organizers: Martin Lames, Tim McGarry, Bernhard Nebel, and Karen Roemer
03.–06. July, 2011 – www.dagstuhl.de/11271

Authors: Martin Lames

The Dagstuhl Seminar 11271 "Computer Science in Sport – Special emphasis: Football" stands in a row of 3 seminars introducing the field of computer science in sports. The general aim is to bring experts from computer science together with experts from sports science to explore the options of interdisciplinary work in this exciting field.

An additional aspect was in the focus of this seminar in July, 2011. We invited not only scientists from the field of football research but also practitioners like Max Reckers (NL) who was responsible for computer science at FC Bayern München under coach Van Gaal. This measure led to livelier discussions because the aspects "Does this work in practice?" or "Does practice really need that?" were not only discussed but also given answers from the view of practice.

Another focus was on the pros and cons of technological aids in football. Here, the discussion brought about many issues, being far apart from a totally affirmative standpoint. The reason for the outstanding position of football in European societies is basically founded in its value for entertainment. Each game broadcasted live can be seen as a drama, a ritualized conflict that will produce a result on that very evening that cannot be foreseen in any way. It is this kind of authenticity that gives football its importance. Concerning technical aids we have to be careful that they do not endanger the dramatic properties of the game. So, as in other fields also, it is not wise to do everything we can do.

Finally, the seminar proved again the benefits of the Dagstuhl seminar concept. Experts from different fields that would hardly meet in their normal business had the opportunity to exchange their ideas in many informal meetings. There was positive resonance from most of the participants stressing especially this fact. Several ideas for new projects among the participants were produced and meanwhile initiated. An application for a next seminar on computer science in sports again with an exciting focus will be prepared soon. A "Dagstuhl Manifesto" is going to be published explaining the interdisciplinary perspectives between sports science and computer science in depth.

14.5 Decision Procedures in Soft, Hard and Bio-ware (Follow Up)

Seminar No. **11272**
Organizers: Nikolaj Bjørner, Robert Nieuwenhuis, Helmut Veith, and Andrei Voronkov
03.–06. July, 2011 – www.dagstuhl.de/11272

Authors: Nikolaj Bjørner

The main goal of the seminar *Decision Procedures in Soft, Hard and Bio-ware (Follow Up)* was to bring together renowned as well as young aspiring researchers from two groups. The first group formed by researchers who develop both theory and efficient implementations of decision procedures. The second group comprising of researchers from application areas such as program analysis and testing, crypto-analysis, hardware verification, industrial planning and scheduling, and bio-informatics, who have worked with, and contributed to, high quality decision procedures. The purpose of the seminar was to heighten awareness between tool and theory developers for decision procedures with the array of applications found in software, hardware and biological systems analysis.

The seminar fell on two and a half days in the week of July 4–6, 2011. 25 researchers from 12 countries (Germany, Austria, Italy, France, USA, United Kingdom, China, Hungary, Spain, Sweden, Czech Republic, Ireland) participated.

14.6 The Future of Research Communication

Authors: Philip E. Bourne, Tim Clark, Robert Dale, Anita de Waard, Ivan Herman, Eduard Hovy, and David Shotton

Research and scholarship lead to the generation of new knowledge. The dissemination of this knowledge has a fundamental impact on the ways in which society develops and progresses, and at the same time it feeds back to improve subsequent research and scholarship. Here, as in so many other areas of human activity, the internet is changing the way things work: it opens up opportunities for new processes that can accelerate the growth of knowledge, including the creation of new means of communicating that knowledge among researchers and within the wider community. Two decades of emergent and increasingly pervasive information technology have demonstrated the potential for far more effective scholarly communication. However, the use of this technology remains limited; research processes and the dissemination of research results have yet to fully assimilate the capabilities of the web and other digital media. Producers and consumers remain wedded to formats developed in the era of print publication, and the reward systems for researchers remain tied to those delivery mechanisms.

Force11 (the Future of Research Communication and e-Scholarship) is a community of scholars, librarians, archivists, publishers and research funders that has arisen organically to help facilitate the change toward improved knowledge creation and sharing. Individually and collectively, we aim to bring about a change in scholarly communication through the effective use of information technology. Force11 has grown from a small group of like-minded individuals into an open movement with clearly identified stakeholders associated with emerging technologies, policies, funding mechanisms and business models. While not disputing the expressive power of the written word to communicate complex ideas, our foundational assumption is that scholarly communication by means of semantically-enhanced media-rich digital publishing is likely to have a greater impact than communication in traditional print media or electronic facsimiles of printed works. However, to date, online versions of 'scholarly outputs' have tended to replicate print forms, rather than exploit the additional functionalities afforded by the digital terrain. We believe that digital publishing of enhanced papers will enable more effective scholarly communication, which will also broaden to include, for example, better links to data, the publication of software tools, mathematical models, protocols and workflows, and research communication by means of social media channels.

The document reports on the presentations and working groups that took place during the Force11 workshop on the Future of Research Communication and e-Scholarship held at Schloss Dagstuhl, Germany, in August 2011. More about Force11 can be found at `http://www.force11.org`.[1]

---

[1](Editorial) The manifesto of 11331 is available at `http://dx.doi.org/10.4230/DagMan.1.1.41`.