Dagstuhl News

January - December 2007

Volume 10
2008

ISSN 1438-7581

Period: January - December 2007

Frequency: One per year

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.

Online version: http://www.dagstuhl.de/files/Reports/News/

Associates: Gesellschaft für Informatik e.V. (GI), Bonn
Technical University of Darmstadt
Frankfurt University
Technical University of Kaiserslautern
University of Karlsruhe
University of Stuttgart
University of Trier
Saarland University
Max Planck Society e.V. (MPG)
French National Institute for Research in Informatics and Automatic Control (INRIA)
Dutch National Research Institute for Mathematics and Informatics (CWI)

Membership: The Center is a member of the Leibniz Association.

Funded by: Federal funding and state funding of Saarland and Rhineland Palatinate

Information: Schloss Dagstuhl Office
Saarland University
P.O. Box 15 11 50
66041 Saarbrücken, Germany
Phone: +49-681-302-4396
Fax: +49-681-302-4397
E-mail: service@dagstuhl.de
http://www.dagstuhl.de/

Welcome

Here are the Dagstuhl News for 2007, the 10th 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 2007 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 2007 are on the Web under URL: `http://www.dagstuhl.de/Seminars/07/`

Our online-publication service, started to publish online proceedings of our Dagstuhl Seminars, is catching on as a service to the Computer Science community. The Dagstuhl Research Online Publication Server (DROPS) (`http://www.dagstuhl.de/publikationen/publikationsserver-drops/`) hosts the proceedings of a few external workshop and conference series. We are currently negotiating ways to make high-quality conferences hosted by us recognizably different from our own online proceedings and other workshop proceedings. We will also develop a business model using the open-access policy for the future.

The policy with the Dagstuhl online proceedings is that authors keep the copyrights to their contributions in order not to harm their rights to submit them to conferences or journals. We hope that the reputation of our Dagstuhl Seminars will make their proceedings a valuable source of information.

It is hard to believe, but Dagstuhl gets even more popular than it already is. We have received a record number of 57 proposals in the current round of applications for Dagstuhl Seminars and Perspective Workshops. Compare this number of applications for a half-year period to the roughly 50 applications we used to have for a full year! We try to accommodate more workshops by adding an extension building with 7 more rooms. It will allow us to run two Seminars in parallel once the building is finished.

The State and the Activities of the *Dagstuhl Foundation*

The foundation currently has 45 personal members and 7 institutional members.
We are experiencing a growing number of requests for travel support or a reduction of the seminar fees. In 2007, we have supported a number of guests in either of these ways.

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, November 2008

Contents

Chapter 1

Data Structures, Algorithms, Complexity

1.1 Exact, Approximative, Robust and Certifying Algorithms on Particular Graph Classes

Seminar No. **07211** Date **20.05.–25.05.2007**
Organizers: Andreas Brandstädt, Klaus Jansen, Dieter Kratsch, Jeremy P. Spinrad

The aim of this seminar was to bring together experts working on exact, approximative, robust and certifying algorithms on particular graph classes. Given the fast advances in various areas of graph algorithms on particular graph classes we have witnessed in the past few years, we believe that it was very important to offer researchers in these areas a forum for the exchange of ideas in the relaxed and inspiring workshop atmosphere that Dagstuhl always offers.

There was a strong interaction and healthy exchange of ideas which resulted in successful applications of exact, approximative, robust and certifying graph algorithms; in particular, the seminar dealt with the following topics and their interactions:

- Exact algorithms require that the algorithm provides exactly the result requested. The approach is interesting for NP-hard problems. Two different approaches are exponential-time algorithms and fixed-parameter algorithms. Exponential-time algorithms must solve the problem for all possible inputs exactly. The goal is to obtain an exponential running time being as small as possible as described in the important survey [G. Woeginger, Exact algorithms for NP-hard problems: A survey. In: Combinatorial Optimization - Eureka! You shrink!. M. Juenger, G. Reinelt and G. Rinaldi (eds.). LNCS 2570, Springer, 2003, pp 185-207.]

  Fixed-parameter algorithms are supposed to solve the problem exactly as long as the result is not larger than the given value of the parameter. In many cases fixed-parameter algorithms are tuned for "small parameters". Fixed-parameter algorithms have been studied extensively by Downey and Fellows [Fixed parameter complexity, Springer, 1999], and recent monographs by Niedermeier, and by Flum and Grohe.

- For approximative algorithms, two new concepts shall be discussed, which improve the running time considerably. The first one deals with parameterized complexity, where various parts of the input such as the number n of vertices or the size k of a maximum independent set play the role of a parameter and the running time of the algorithm is optimized with respect to the parameters. This approach is promising for polynomial approximation schemes.

  The second one concerns methods of (non-)linear programming for graph-theoretic problems. There are various optimization problems such as special network-flow problems or determining a maximum independent set in a perfect graph which have polynomial time algorithms but these are far from being really efficient since the algorithms have to solve large linear programming instances. The algorithms become much more efficient, however, if only approximative solutions (with good factors) are required and this is done using methods of (non-)linear programming.

- A robust algorithm for a graph class $\mathcal{C}$ and an algorithmic problem Π is always giving a correct answer: If the input graph G is in the class $\mathcal{C}$ then the problem Π will be correctly solved, and if G is not in $\mathcal{C}$ then either Π will be correctly solved or the algorithm finds out that G is not in $\mathcal{C}$. In both cases, the answer is correct, and the algorithm avoids recognizing $\mathcal{C}$. This can be of big advantage if recognizing $\mathcal{C}$ is NP-complete or even harder. There are various degrees of verification in the case that G is not in $\mathcal{C}$; a witness for this is desirable.

- Certifying recognition algorithms provide a proof respectively certificate for membership and non membership. Certifying algorithms are highly desirable in practice since implementations of correct algorithms may have bugs. Furthermore since the software producing the certificates may have bugs, the certificates have to be authenticated, and this should use a simple and efficient algorithm. A good example is the linear time certifying recognition algorithm for interval graphs [D. Kratsch, R. M. McConnell, K. Mehlhorn, J. Spinrad, Certifying algorithms for recognizing interval graphs and permutation graphs, SODA 2003: 158-167].

1.2 Structure Theory and FPT Algorithmics for Graphs, Digraphs and Hypergraphs

Seminar No. **07281** Date **08.07.–13.07.2007**
Organizers: Erik Demaine, Gregory Z. Gutin, Daniel Marx, Ulrike Stege

Fixed-parameter algorithmics (FPA) is a relatively new approach for dealing with NP-hard computational problems. In the framework of FPA we try to introduce a parameter k such that the problem in hand can be solved in time $O(f(k)n^c)$, where $f(k)$ is an arbitrary computable function, n is the size of the problem and c is a constant not dependent of n or k. When a parameterized problem P admits an algorithm of running time $O(f(k)n^c)$, P is called fixed-parameter tractable (FPT). The ultimate goal is to obtain such $f(k)$ and

c that for small or even moderate values of k the problem under consideration can be completely solved in a reasonable amount of time.

Many practical problems can now be tackled using FPA. The aim of the seminar, held from July 9, 2007 to July 14, 2008 was to bring together specialists of fixed-parameter tractability with researchers who could provide new theoretical tools for FPA and with practical computing practitioners who could benefit from FPA in their own application domains.

The possibility of deep and algorithmically useful combinatorial structure theory seems to be closely allied with FPT—in various combinatorial settings these two different aspects, the one mathematical and the other algorithmic, seem to go together. The parameterized problem Graph Minor Testing is FPT, and exposes in its allied structure theory, with such fundamental structural parameters as treewidth, the kinds of connections between parameterized structure theory and FPA that the workshop explored, encouraged and developed.

Beyond treewidth, which turned out to be a surprisingly universal structural parameter, there is a collection of newer related notions which are currently of intense research interest: cliquewidth of graphs, hypertreewidth of hypergraphs and various parameters measuring near-acyclicity of hypergraphs. The latter are of relevance to the natural input distributions in database and constraint satisfaction problems, and it is a major concern of the workshop to motivate and explore to what extent the successful structure theory and FPA of treewidth, etc. of graphs can be lifted to the setting of hypergraphs.

Although graphs have proven to be a hugely flexible computational modeling tool, and the structure theory and allied FPA of graphs has developed strongly, very little can yet be said for digraphs, even though in the grand scheme of things, digraphs are the more important modeling tool: the entire picture for digraphs in terms of structure theory and FPA has lagged far behind graphs. Some of the most important open problems in concrete FPA involve digraphs (e.g., the Directed Feedback Vertex Set problem that has a vast range of potential important applications, and was widely conjectured to be FPT).

During the 5 days of the conference, 23 talks were given by the participants. Two of these talks were 50-minute surveys given by founders of the field: Mike Fellows started the workshop by reviewing the latest technical and methodological developments and Mike Langston reported on recent algorithmic applications in computational biology.

As a highlight of the seminar, Jianer Chen and Igor Razgon presented their very recent work on the Directed Feedback Vertex Set problem.

The complexity status of this very important problem was open for 15 years or so, until two independent groups of researchers proved its fixed-parameter tractability earlier this year. The solution of the problem required a clever mix of old and new ideas. In recent years the field witnessed a more systematic identification, study, and dissemination of algorithmic ideas, leading to significant new results. There is no doubt that this progress was helped enormously by meetings such as the previous Dagstuhl seminars.

The talks left plenty of time for discussion in the afternoon. An open problem session was held on Monday. Problems raised there were discussed by different groups throughout the seminar.

1.3 Probabilistic Methods in the Design and Analysis of Algorithms

Seminar No. **07391** Date **23.09.–28.09.2007**
Organizers: Martin Dietzfelbinger, Shang-Hua Teng, Eli Upfal, Berthold Vöcking

It is difficult to overstate the importance of probabilistic methods in Theoretical Computer Science. They belong to the most powerful and widely used tools, for example in designing efficient randomized algorithms for tackling hard optimization problems; in establishing various lower bounds in complexity theory; in the proofs of many useful discrete properties in extremal combinatorics; in providing frameworks such as the average-case and smoothed analysis for measuring the performance of algorithms; in the theory of interactive proofs. The body of work using probabilistic methods has experienced an impressive growth in the recent years. The following topics attracted enormous attention both from theorists as well as practitioners during the recent years.

In the area of randomized algorithms, several new probabilistic techniques were developed. For example, there are several exciting recent developments in the probabilistic metric embedding with tree metrics. Because various optimization problems can be solved optimally on trees (e.g., by the dynamic programming approach), quality approximations of arbitrary metrics by tree metrics provide a systematic approach for designing approximation algorithms for general metrics. Further, new techniques for designing randomized data structures were developed that draw on methods from the theory of random graphs and random walks in graphs. A core issue here is the efficient simulation of high-degree randomness without the assumption of the inputs being random.

Impressive progress has also been obtained regarding the probabilistic analysis of algorithms. In practice, scientists and engineers often use heuristic algorithms for optimization problems ranging from network design to industrial optimizations. Most of these algorithms, after years of improvements, work well in practice. However, their worst-case complexity might still be very poor, for example, exponential in the input size. It is an old observation in quite a few application areas that the worst-case instances of an algorithm might not be "typical" and might never occur in practice. So worst-case analysis can improperly suggest that the performance of the algorithm is poor. Trying to rigorously understand and model the practical performance for such heuristic algorithms is a major challenge in Theoretical Computer Science.

Probabilistic methods have played an active role in developing analysis frameworks that provide "practical enough" measures, yet one can still conduct rigorous analyses using these frameworks. For example, the recently developed *smoothed analysis* uses small random perturbations for defining performance measures. This framework applies to algorithms whose inputs are subject to slight random noises. The smoothed complexity of an algorithm is then the maximum over its inputs of the expected running time of the algorithm under slight perturbations of that input. Smoothed complexity is measured in terms of the size of the input and the magnitude of the perturbation.

Another area in which random inputs play an important role is stochastic optimization.

Here uncertainty in the data is modeled by probability distributions. Stochastic optimization has a wide range of applications in various areas, including logistics, transportation, financial instruments, and network design. In recent years, there has been significant progress in analyzing important algorithms and heuristics used in this field. For example, the sample average approximation (SAA) method solves stochastic programs by sampling from the distribution of input scenarios. Recent theoretical results show that the SAA method has the properties of a fully randomized approximation scheme for a large class of multistage stochastic optimization problems.

The workshop covered recent progress in randomized algorithms and probabilistic measures of algorithms including the smoothed analysis, average-case analysis, semi-random analysis, and stochastic optimization. The presentations covered a large range of optimization problems such as linear programming, integer programming, random games, computational geometry, and scheduling. The most important contribution of the seminar is the exchange of new ideas between researchers using probabilistic methods in different contexts. In addition of providing an opportunity for information sharing and collaborations, the workshop exposed young researchers, students, and postdocs to recent developments and outstanding issues in probabilistic methods.

1.4 Algebraic Methods in Computational Complexity

Seminar No. **07411** Date **07.10.–12.10.2007**
Organizers: Manindra Agrawal, Harry Buhrman, Lance Fortnow, Thomas Thierauf

The seminar brought together almost 50 researchers covering a wide spectrum of complexity theory. The focus on algebraic methods showed once again the great importance of algebraic techniques for theoretical computer science. We had almost 30 talks of length between 15 and 45 minutes. This left enough room for discussions. We had an open problem session that was very much appreciated. In the following we describe the talks in more detail.

The construction of good extractors and expanders plays a crucial role in derandomization. Chris Umans explained how to construct highly unbalanced bipartite expander graphs with expansion arbitrarily close to the degree, essentially optimal. The construction is based on the ideas underlying the recent list-decodable error-correcting codes of Parvaresh and Vardy (FOCS '05). Anup Rao considered the model that the source, a family of distributions, gives a random point from some unknown low dimensional affine subspace with a low-weight basis. This model generalizes the well studied model of bit-fixing sources. He showed how to construct new extractors for this model that have exponentially small error, a parameter that is important for applications in cryptography.

Derandomization is strongly related to proving lower bounds. In 1998, Impagliazzo and Wigderson proved a hardness vs. randomness tradeoff for BPP: if one cannot derandomize BPP, then E needs exponential size circuits. Ronen Shaltiel considered the Artur-Merlin class AM instead of BPP. He showed uniform hardness vs. randomness tradeoffs for AM that are near-optimal for the full range of possible hardness assumptions.

From another point of view Eric Allender considered the question of how close we are to proving circuit lower bounds. For example any proof that NP is not equal to TC^0 will have to overcome the obstacles identified by Razborov and Rudich in their paper on "Natural Proofs". In his talk, he pointed to some plausible way to prove that TC^0 is properly contained in NC^1. Another obstacle in separating complexity classes like P and NP is *relativization*. Baker, Gill, and Solovay showed that no relativizable proof can separate P from NP. Since then we have seen some non-relativizing proofs like IP = PSPACE. Scott Aaronson, together with Avi Wigderson, extended the notion of relativization to *algebraization* and showed several results including that

1. All known relevant examples of non-relativizing proofs algebrize, and
2. Any proof that separates P from NP would require non-algebrizing techniques.

We had a series of talks on circuit complexity. Fred Green gave a complete characterization of the smallest circuits that compute parity that have a majority gate as output, a middle layer of MOD_3 gates and a bottom layer of AND gates of fan-in 2. Nitin Saxena presented a deterministic polynomial time algorithm for testing whether a *diagonal* depth-3 circuit C (i.e. C is a sum of powers of linear functions) is zero. Motivated by the problem of factoring integers, Pierre Mckenzie exhibited "gems", that is, arithmetic $\{+, -, x\}$-circuits that use only n multiplication gates to compute univariate integer polynomials having 2^n distinct integer zeros, for $n = 1, 2, 3, 4$. Rüdiger Reischuk talked on bit comparator sorting circuits that have a minimal average time complexity. Ingo Wegener talked about lower bounds for the multiplication function that can be obtained by the technique of Nechiporuk. Falk Unger considered circuits with noisy gates. He showed a negative result, that formulas built from gates with two inputs, in which each gate fails with probability at least epsilon, cannot compute any function with bounded error. Wim van Dam introduced a model of algebraic quantum circuit, for all finite fields GF(q). Farid Ablayev showed how bounded error syntactic quantum branching programs can be simulated by classical deterministic branching programs

We had a wide-range of talks on classical complexity. Rahul Santhanam showed that SAT is not instance compressible unless NP is contained in coNP/poly. A language L in NP is instance compressible if there is a polynomial-time computable function f and a set A such that f reduces L to A and for each $x \in L$, $f(x)$ is of size polynomial in the witness size of x. Harry Buhrman applied this result to show that there are no sub-exponential size complete sets for NP or coNP unless NP is contained in coNP/poly. John Hitchcock presented a connection between mistake-bound learning and polynomial-time dimension. As a consequence he showed that the class E does not reduce to sparse sets under certain reductions. N. Variyam Vinodchandran presented his new result that the directed planar reachability problem is in unambiguous logarithmic space (UL). Christian Glasser talked on the relation of autoreducibility and mitoticity for polylog-space many-one reductions and log-space many-one reductions. Rocco Servedio described recent results on approximating, testing, and learning halfspaces (also known as linear threshold functions, weighted majority functions, or threshold gates). Marius Zimand showed how to reduce the length of the advice given to heuristic algorithms to approximate problems in BP-TIME[sublinear].

Troy Lee gave a talk on communication complexity. He presented a direct product theorem for discrepancy, one of the most general techniques in communication complexity. Nikolai Vereshchagin studied the two party problem of randomly selecting a string among all the strings of length n. He presented protocols that have the property that the output distribution has high entropy, even when one of the two parties is dishonest and deviates from the protocol. Ben Toner considered the scenario that Alice and Bob share some bipartite d-dimensional quantum state. It is known that by performing two-outcome measurements, Alice and Bob can produce correlations that cannot be obtained classically. He showed that there is a classical protocol that can classically simulate any such correlation by using only two bits of communication. Julia Kempe considered multi-prover games where the provers share entanglement. She showed that it is NP-hard to determine, or even to approximate the entangled value of the game. In the same setting Oded Regev showed that when the constraints enforced by the verifier are 'unique' constraints (i.e., permutations), the value of the game can be well approximated by a semidefinite program for one-round games between a classical verifier and two provers who share entanglement.

Property testing deals with the question of distinguishing inputs that satisfy a given property from inputs that are far from satisfying it, using a number of queries that is as small as possible. Eldar Fischer suggested that seeking out properties that are by their nature "massively parameterized" is a worthy direction for property testing research. As an example, he considered the testability of the property of having a directed path from s to t in a graph.

Jack Lutz used connections between the theory of computing and the fine-scale geometry of Euclidean space to give a complete analysis of the dimensions of individual points in fractals that are computably self-similar.

As is evident from the list above, the talks ranged a wide area of subjects with the underlying of using algebraic techniques. It was very fruitful and has hopefully initiated new directions in research. We look forward to our next meeting!

1.5 Algorithmic-Logical Theory of Infinite Structures

Seminar No. **07441** Date **28.10.–02.11.2007**
Organizers: Rod Downey, Bakhadyr Khoussainov, Dietrich Kuske, Markus Lohrey, Moshe Y. Vardi

One of the important research fields of theoretical and applied computer science and mathematics is the study of algorithmic, logical and model theoretic properties of structures and their interactions. By a structure we mean typical objects that arise in computer science and mathematics such as data structures, programs, transition systems, graphs, large databases, XML documents, algebraic systems including groups, integers, fields, Boolean algebras and so on. From a mathematics point of view these are natural objects of study and the theory of computable structures initiated by Malcev and Rabin in the 60s witnesses it. This mathematical study has been one of the most active areas of research in the last decade. From a computer science point of view, there has been a growing

interest in understanding infinite structures. The need for this study comes from the fact that one cannot usually put bounds on typical objects of computer science such as the sizes of databases, programs, XML documents, stacks, communication buffers, and so on. Moreover, these objects are usually not seen as static components but are modified in a dynamic way, which leads to systems with infinite state spaces, i.e., infinite transition systems. These transition systems can be modeled by infinite graphs, where nodes correspond to states of the system and edges correspond to transitions between the states. Properties of such systems can be expressed in logical formalisms and it is an algorithmic task to determine the validity of such formulas or to calculate the set of states satisfying a certain property. In computer science, the following aspects for a logical theory of such systems are of interest.

- Infinite systems in computer science are usually represented by some finite description or abstract machine; system states are then configurations of that abstract machine.
- These system states have some internal structure that determines the global properties of the system. E.g., they may be represented by natural numbers, finite strings, trees, or graphs and transitions are defined by transducers on the data structures.
- The emphasis is put on efficient algorithms for the verification of system properties that are specified in a suitable logical language.

Several aspects of such a unified theory have already been addressed in the past. These include:

- Computable model theory, a branch of classical model theory and recursion theory, considers structures that are presented in some effective (and thus finite) way. An emphasis in this field is to study algorithmic properties of structures by comparing complexities of their undecidable features. Typical important structures in this context such as arithmetic of natural numbers do not have decidable first order theories. Essential tools of research here are the Turing degrees and methods of recursion theory.
- The theory of automatic structures, a newly formed direction of research in which structures are represented by finite state machines such as finite automata, tree automata, ω-automata, etc. In contrast to computable model theory, an emphasis here is given to understanding the algorithmic properties of structures by comparing complexities of their decidable features. All these structures have decidable first order theories. The development of this theory is based on methods of complexity theory, finite combinatorics, model theory and automata theory.
- Classes of infinite graphs that are generated by some kind of abstract machines and that lead to decidable monadic second-order theories became an active research topic in recent years. Currently, the most general class of graphs with decidable monadic second-order theories is the Caucal hierarchy. Graphs that are generated by ground tree rewriting systems constitute a class, where monadic second-order logic is undecidable in general, but first-order logic with reachability is still decidable.

- Several approaches for model-checking systems with infinite state spaces were developed in the past by researchers working in verification. Typical examples in this context are unbounded communication buffers, stacks in procedural programming languages, or parameterized systems.

At this seminar, researchers from all these fields met. To give the non-specialists a general overview of the flavor, topics, methods, and open questions of the other fields, we had five keynotes given by Igor Walukiewicz, Wolfgang Thomas, Denis Hirschfeldt, Sasha Rubin, and Parosh Aziz Abdulla. These were complemented by contributed talks of the participants that span the whole spectrum laied out above.

1.6 Equilibrium Computation

Seminar No. **07471** Date **18.11.–23.11.2007**
Organizers: Jean-Jacques Herings, Marcin Jurdzinski, Peter Bro Miltersen, Eva Tardos, Bernhard von Stengel

The purpose of this Dagstuhl seminar was to bring together researchers from different disciplines working on algorithmic problems of finding equilibria. As explained in the Motivation for this seminar, the different topics were (with talks given by)

- *Parity Games,* Andersson (junior researcher), Gimbert (junior researcher), Grädel (survey talk), Svensson (junior researcher), Zwick (survey talk),
- *Complexity of Finding Equilibria,* Elkind (junior researcher), Etessami (survey talk), Goldberg (introductory survey talk on the class PPAD, not listed below), Fabrikant (junior researcher), Hoefer (junior researcher), Monien, Vöcking (survey talk),
- *Mathematical Programming,* Halman, Morris, Theobald,
- *Economic Equilibria,* Heydenreich (junior researcher), Jain, Peeters (survey talk),
- *Game Theory,* Balthasar (junior researcher), Goldberg, Jiang (junior researcher), Sørensen (junior researcher), Turocy (survey talk), Vermeulen, von Stengel.

In addition to his talk, Paul Goldberg gave an introduction to the complexity class PPAD which is central for defining the complexity of finding one equilibrium (for example, a Nash equilibrium in a game of n players). This talk is accessible through the following URL: `http://www.csc.liv.ac.uk/~pwg/PPADintro/PPADintro.html`.This colorful talk was prepared by Paul Goldberg in Dagstuhl in immediate response to requests for such an introduction, and is described as such on the mentioned webpage.

Furthermore, Mike Paterson gave a popular evening talk on the entertaining topic of piling bricks so that they can "stick out" as far as possible, which can be considered as an "equilibrium problem" in physics but not in the computational sense studied in the seminar.

Overall, the seminar talks represented a good balance of the topics, and were not too numerous so as to make listening tiresome. A significant time was spent in discussions, drawing on the expertise of experienced scholars in the field (for example, Nimrod Megiddo). We encouraged, with success, to let junior researchers present their work as much as possible. The survey talks, often given by senior researchers, gave introductions and overviews.

Many of the topics of the seminar are in areas with very hard and long-standing open questions. In particular, solving parity games, or the related mean-payoff and simple stochastic games, in polynomial time is an intriguing open problem. It is a plausible conjecture that this can be done, given that the problem is in the intersection of the complexity classes NP and co-NP. The most famous problem with that property is linear programming (LP), which is equivalent to finding an equilibrium in a zero-sum game, and thus closely related to the problems discussed in the workshop. The polynomial-time algorithms for linear programming, such as the ellipsoid method, were hailed as breakthroughs at the time. To this day, the understanding even of the linear programming problem is limited; for example, we do not have a "combinatorial", simplex-type algorithm that would solve this problem in polynomial time.

In this context, the results by Nir Halman on an abstract view of LP-type problems and their connection to the parity games and their relatives, provided one example of a "bridge" across several fields, Mathematical Programming, Parity Games, and the Complexity of Finding Equilibria. Some discussions started in how this could be extended to the computation of Nash equilibria of bimatrix games.

The equilibrium problems discussed in this workshop are much harder than linear programming, which by itself is an important interesting case. None of the main problems were solved – this would have been close to sensational –, but we could observe some (necessarily partial and incremental) progress.

The hard problems mentioned above are concerned with computing Nash equilibria. Another focus of the workshop was the computation of other, more refined solution concepts. Here, the interaction between the participants from the computer science community and the participants from the economics community was extremely fruitful. In particular, a number of computational problems were jointly formulated which together form an interesting research program.

A representative example is the following: Given a three-player game in normal form *and* a strategy profile of the game, can it be decided in polynomial time if the strategy profile is a (trembling hand) perfect equilibrium of the game? The corresponding computational problem for Nash equilibria is trivial. One can ask the same question for other refinement notions, and we believe it would be very interesting to classify refinement notions by hardness or easiness of their verification problem, for two reasons. First, an easiness result would be useful in practice for studying equilibria computationally. Secondly, a refinement notion where even the verification problem is computationally intractable may arguably be considered an inferior solution concept to one where it is tractable to determine if a given profile is in equilibrium.

The workshop demonstrated that the topic of "equilibrium computation" is of great current interest. The stimulating discussions showed that it was very well worth bringing researchers together who normally operate in different communities.

Chapter 2

Verification, Logic

2.1 Runtime Verification

Seminar No. **07011** Date **02.01.–06.01.2007**
Organizers: Bernd Finkbeiner, Klaus Havelund, Grigore Rosu, Oleg Sokolsky

The 2007 Dagstuhl Seminar 07011 on Runtime Verification was held from Tuesday January 2 to Saturday January 6, 2007. Thirty researchers participated and discussed their recent work and recent trends in runtime verification. Other terms for this subject are: program monitoring, dynamic program analysis, and runtime analysis. Over the past few years, this field has emerged as a focused subject in program analysis that bridges the gap between the complexity-haunted field of fully formal verification methods and the ad-hoc field of testing. Runtime verification supplements static analysis and formal verification with more lightweight dynamic techniques when the static techniques fail due to complexity issues. From the perspective of testing, runtime verification helps to formalize oracle specification. Runtime verification uses some form of instrumentation to extract, during test or in operation, a trace of observations from a program run and then applies formal verification to this trace. The focus on traces rather than on transition systems is of course what makes the approach more scalable but also less effective at the same time. However, applying rigor and advanced techniques in trace analysis may provide several practical advantages.

The seminar covered several areas, which we shall briefly touch upon. One of the corner stones of this field is the monitoring of program executions (theoretically thought of as traces) against formal specifications, for example represented in temporal logic, regular expressions or state machines. Specification logics can include real-time features enabling the monitoring of real-time properties. Some of the questions that arise in this context are the following: what expressive power is required of a monitoring logic; what characteristics should it have in order to make monitoring efficient with as little impact on the running program as possible; and what characteristics will make such a logic easy and attractive to use from a user's point of view. The two first questions are specific to runtime verification whereas the latter is of general interest to any formal method.

In order to monitor a program (or more generally: a system), the program (system) must be instrumented to feed the monitor. This can happen by instrumenting the program to generate a trace in a log-file, which can be analyzed off-line, or the program can be instrumented to drive the monitor directly during execution, in which case errors are detected immediately as they occur. Aspect oriented programming is an example of a technology for performing program instrumentation. An interesting trend is the concept of state-full aspects, which essentially extend the point-cut language of aspect oriented programming to temporal predicates over the execution trace. In this view an aspect advice consists of a temporal trace predicate and a statement to be executed when this predicate gets violated during a program execution. This approach can be seen as combining aspect oriented programming with runtime verification. The execution of repair code when a property gets violated is an example of a fault protection strategy. This leads into a paradigm for programming where programs are not expected to behave correctly and where a program is embedded in a protection armor, providing error diagnosis and recovery.

The formalization of properties in terms of specifications requires human effort, which is known to cause resistance. A branch of the field attempts to perform dynamic analysis in the absence of human-provided formal specifications. There are two variants of this work. In the first variant algorithms are pre-programmed that analyze for specific generic kinds of errors that are generally regarded as problems in any application. Examples are concurrency errors such as data races and deadlocks. The second variant, dynamic specification learning, consists of learning specifications from runs. Each run that is accepted by a user is regarded as contributing to a nominal behavior specification of the program. After a period such a nominal behavioral specification can be turned into an oracle used to detect deviations.

An important topic is the interaction between static and dynamic analysis. Static analysis can be used to minimize the impact of monitoring a program by for example reducing the number of program points where the program needs to interact with the monitor. A dual view of this interaction between static and dynamic analysis is to regard dynamic analysis as a rescue plan when static analysis cannot determine whether a program satisfies a particular property. It may for example be the case that a property can be proved about a program, but only under the assumption of a set of proof-obligations (lemmas), each of which can then be dynamically monitored during test runs or during operation.

The field of runtime verification overlaps with the field of testing from the perspective of test oracles. Often, a monitor for a formally specified property can be used to evaluate whether a test execution has been successful. However, runtime verification is less concerned with the test case generation aspect of testing, where the goal is to drive the program into all its corners. Runtime verification focuses on analyzing or collecting information from individual runs, independently of how they have been obtained. The Dagstuhl event included contributions from the testing field on topics such as test case generation, fault injection, and unit testing. These contributions explored the relationship between the fields of testing and runtime verification.

2.2 Deduction and Decision Procedures

Seminar No. **07401** Date **30.09.–05.10.2007**
Organizers: Franz Baader, Byron Cook, Jürgen Giesl, Robert Nieuwenhuis

Formal logic provides a mathematical foundation for many areas of computer science, including problem specification, program development, transformation and verification, hardware design and verification, relational databases, knowledge engineering, theorem proving, computer algebra, logic programming, and artificial intelligence.

Using computers for solving problems in these areas, therefore, requires the design and implementation of algorithms based on logical deduction. It remains one of the great challenges in informatics to make computers perform non-trivial logical reasoning. Significant progress, however, has been made in the past ten years. For example, real arithmetic algorithms in Intel processors were formally verified using an interactive theorem prover, interactive and automatic theorem provers are routinely used in formal methods tools, and automatic theorem provers and finite model building programs solved various open mathematical problems of combinatorial nature. Automated deduction, in particular for so called description logics, is widely assessed as a core enabling technology for the Semantic Web. Methods of interactive theorem proving have helped in formally verifying semantic (type) safety aspects of programming languages, such as Java. The "Schwerpunktprogramm Deduktion" funded by the Deutsche Forschungsgemeinschaft from 1992 to 1998 together with the seven previous Dagstuhl seminars on "Deduction" (held biennially since 1993) have been instrumental in obtaining these successes.

Because of the recent progress in applying automated deduction methods and tools in various application areas, like hardware and software verification, cryptographic protocols, programming languages, and the semantic web, the focus of the Deduction seminar in 2005 was on applications of deduction. The application areas best represented at that seminar (in terms of number of talks held) concerned various forms of "verification". From the talks on these applications, it became clear that the integration of theory specific reasoners, in particular decision procedures, into a core general purpose verification environment and the exible and semantically well-founded combination of such reasoners are extremely important in many applications of verification tools.

For this reason, the focus of the Deduction seminar in 2007 was on decision procedures and their integration into general purpose theorem provers. Our goal was to further the confluence between application-driven approaches for combining and using decision procedures in deduction and the strong foundational research on this topic.

In total we had 55 participants, mostly from Europe, but also from USA, Israel, and Australia. A good balance between more senior and junior participants was maintained. The program consisted of 37 relatively short talks, which gave ample time for discussion, both during and after the talks.

The talks and discussions showed that "Deduction and Decision Procedures" is currently a very important and vibrant research area, which creates both important new foundational

results and tools that are at the core of recent advances in various subareas of "verification". The approach most strongly represented at the seminar was the SMT (Satisfiability Modulo Theories) approach, a trend that was already discernible during the 2005 Seminar on "Deduction and Applications", but has become even stronger in the last two years.

The seminar showed that there are essentially three different approaches for using deduction in program verification: (1) extremely powerful interactive theorem provers which however lack the necessary automation, (2) automated theorem provers mainly based on first-order logic which however fail to cater for important aspects like arithmetic, and (3) specialized tools like SMT solvers based on powerful automatic procedures for particular logical theories. These three approaches have all led to impressive results in the last years. However, the communities working on these approaches are still unnecessarily disjoint. Therefore, the seminar revealed that a main goal for the future of deduction is to improve the collaboration and the combination of interactive and automated deduction (both for first-order logic and for SMT). For this reason, the next Deduction seminar (planned for 2009) will focus on "Interaction vs. Automation: The two Faces of Deduction".

The seminar consisted of a full, but not overly loaded program which left enough time for discussions.We felt the atmosphere to be very productive, characterized by many discussions, both on-line during the talks and off-line during the meals and in the evenings. Altogether, we perceived the seminar as a very successful one, which gave the participants an excellent opportunity to get an overview of the state-of-the-art concerning decision procedures in automated deduction and their applications in areas related to verification. The seminar has further enhanced the cross-fertilization between foundational research on this topic and application-driven approaches.

Chapter 3

Geometry, Image Processing, Graphics

3.1 Visualization and Processing of Tensor Fields

Seminar No. **07022** Date **09.01.–13.01.2007**
Organizers: David H. Laidlaw, Joachim Weickert

Motivation

While scalar- and vector-valued data sets are omnipresent as grayscale and color images in the fields of scientific visualization and image processing, also matrix-valued data sets (so-called **tensor fields**) have gained significant importance recently. This has been triggered by the following developments:

- Medical imaging techniques such as **diffusion tensor magnetic resonance imaging (DT-MRI)** are becoming more and more widespread. DT-MRI is a 3-D imaging method that yields a diffusion tensor in each voxel. This diffusion tensor describes the diffusive behavior of water molecules in the tissue. It can be represented by a positive semidefinite 3 x 3 matrix in each voxel.
- Tensors have shown their use as **feature descriptors** in image analysis, segmentation and grouping. This includes numerous applications of the structure tensor in fields ranging from motion analysis to texture segmentation, but also the tensor voting framework.
- **Tensor factorizations** have been proposed as compact multilinear models for multidimensional visual datasets. They successfully exploit spatial redundancy.
- A number of scientific applications require to visualize tensor fields. The tensor concept is a common **physical description of anisotropic behaviour**, for instance in geomechanics / earthquake simulation, satellite gradiometry, liquid crystals, and material science.

Problems

This has led to a number of challenging scientific questions, e.g.

- How should one visualize these high-dimensional data in an appropriate way?
- How can user interaction be coupled with visualization to serve scientific users' problems?
- How can structure of tensor fields be addressed? Topological methods have been used in visualization, but more research in this area is needed.
- What are the relevant features to be processed? Is it better to have component-wise processing, to introduce some coupling between the tensor channels or to decompose the tensor in their eigenvalues and eigenvectors and process these entities?
- How should one process these data such that essential properties of the tensor fields are not sacrificed? For instance, often one knows that the tensor field is positive semidefinite. In this case it would be very problematic if an image processing method would create matrices with negative eigenvalues.
- How should one adapt the processing to a task at hand, e.g. the tractography of fiber-like structures in brain imaging? This may be very important for a number of medical applications such as connectivity studies.
- How can one perform higher-level operations on these data, e.g. segment tensor fields? Current segmentation methods have been designed for scalar- or vector-valued data, and it is not obvious if and how they can be extended to tensor fields. Often this requires to introduce sophisticated novel metrics in the space of tensor data.
- How can one perform operations on tensor fields in an algorithmically efficient manner? Many tensor fields use 3 x 3 matrices as functions on a three-dimensional image domain. This may involve a very large amount of data such that a clear need for highly efficient algorithms arises.
- Is it possible to derive a generic visualization and processing paradigm for tensor fields that originate from different application areas?
- What are the scientific application areas that can be served by tensor field visualization and analysis? What are the fundamental relevant problems from those application areas?

Since tensor fields have been investigated in different application domains and the field is fairly young, not many systematic investigations have been carried out so far. It is thus not surprising that many results are scattered throughout the scientific literature, and often people are only aware of a small fraction of the relevant papers. In April 2004, a Dagstuhl Perspective Workshop organised by Hans Hagen and Joachim Weickert was the

first international forum where leading experts on visualization and processing of tensor fields had the opportunity to meet, sometimes for the first time. This workshop has identified several key issues and has triggered fruitful collaborations that have also led to the first book in its area. It contains a number of survey chapters that have been written in collaboration and that should enable also nonexperts to get access to the core ideas of this rapidly emerging field. Participants of the 2004 Perspective Workshop were very enthusiastic about the interdisciplinarity and the interaction possibilities, and they were very interested in pursuing this concept further in a second workshop. This is the goal of the current follow-up Dagstuhl seminar.

Goals

Similar to the first meeting, we want to gather people from scientific visualization, image processing, medical imaging and other tensor-oriented application fields in a real interdisciplinary atmosphere, and we intend to publish the scientific output of this meeting in a postproceedings volume. This time, however, the following innovations are planned:

- Since a number of fundamental issues has already been identified at the 2004 Perspective Workshop – and quite some progress has been achieved – we would also like to encourage younger participants to present their more recent findings. Many researchers have entered this field fairly recently. To give room for the latest advances, about 50 per cent new people who have not attended the first tensor meeting have been invited. Moreover, in order to gain better insights into the foundations of tensor fields, also more experts from applied mathematics have been invited.
- Directions that have not or hardly been addressed in the first workshop and will play a role in our current proposal include
  - tensor voting ideas
  - tensor approximations of high dimensional visual data
  - a more consequent use of differential geometry in image analysis of tensor fields
  - methods based on wavelets
  - clustering, labeling of clusters, and calculating quantitative measures from regions
  - a number of alternative applications beyond DT-MRI, such as tagged MRI for deformation analysis of the heart muscle, liquid crystals, geomechanical / earthquake data, satellite gradiometry, and material science.
- We explicitly encourage all participants to give stimulating, provocative and even controversal presentations that trigger discussions rather than polished, but less exciting technical talks.

Since this area is rather young, representatives of most relevant groups can meet within the framework of a relatively small seminar. We are confident that the unique atmosphere of Schloss Dagstuhl provides an ideal location to initiate a closer interaction within this emerging scientific field.

3.2 Computational Geometry

Seminar No. **07111** Date **11.03.–16.03.2007**
Organizers: Pankaj Kumar Agarwal, Helmut Alt, Franz Aurenhammer

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 theoretical questions remain an important topic 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 have been developed. Remarkably, these implementations emerged from the originally theoretically oriented computational geometry community itself. Consequently, many researchers are now concerned with theoretical foundations as well as implementation issues.

The seminar will focus on theoretical as well as practical issues in computational geometry. Some of the currently most important topics in computational geometry will be addressed in the seminar:

- 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, pseudo triangulations, generalization of triangulations and developed in connection with visibility and shortest-path problems, have turned out to be useful for many other applications and are being investigated intensively.
- Because of applications in molecular biology, computer vision, geometric databases, shape analysis has become an important topic.
- Another increasingly important application of computational geometry is modeling and reconstruction of surfaces. It brings about many interesting questions concerning fundamental structures like triangulations as well as new issues in computational topology.
- 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.

Dagstuhl seminars on computational geometry have been organized since 1990, in a two year rhythm in the recent years, have always been very successful in both disseminating the recent research and conceiving the new ideas.

Parallel to our seminar a second one on Cutting, Packing, Layout and Space Allocation will be organized in Dagstuhl. Because of the overlap of both topics there certainly will be talks in the one seminar which are of interest to the other group, as well. In addition, there should be a beneficial exchange of problems and ideas between both groups. Therefore, the groups of organizers of both seminars agreed to give participants the opportunity to move freely between the two seminars and to have at least one common session.

3.3 Cutting, Packing, Layout and Space Allocation

Seminar No. **07112** Date **13.03.–16.03.2007**
Organizers: Karen M. Daniels, Graham Kendall

Dagstuhl Seminar 07112 took place from Wednesday, 14th March, 2007 to Friday, 16th March. There were 17 participants from a total of 9 different countries. The seminar was led by Graham Kendall and Karen Daniels. Jan van der Veen was designated to collate the proceedings.

Seminar 07112's talks fostered productive discussions on problems of common interest to cutting, packing and space allocation researchers, with most of the presentations concerned with cutting and packing but it is apparent that "Cutting and Packing" and "Space Allocation" are closely related areas.

Some of the work addressed two-dimensional packing, with other presentations concentrating on the three-dimensional variant. Three-dimensional research is a largely unexplored research area, with scope for significant advances to be made. At this seminar we heard presentations which outlined a number of different approaches that are being utilised. For example:

- Packing boxes into containers using heuristics.
- Independent sets in conflict graphs were used for some box/container packing problems
- Convex and non-convex three-dimensional shapes were packed into minimal sized containers (with discrete rotations) using randomization for global search combined with gradient descent for local optimization.

The two-dimensional problem remains an active area of research, with heuristics and meta-heuristics being used for packing both rectangles and irregular shapes polygons. Despite its long history, there still remain many open problems and challenges in two-dimensional packing.

The space allocation talks dealt with two different topics. 1) The optimisation of commercial (retail) shelf space allocation. This was formulated as an integer programming problem which was optimised using simulated annealing combined with heuristic search. 2) Improving the utilisation of university teaching space. This talk outlined the many challenges that face university administrators and planners and emphasised the need for progress in this area to support them in tackling this important problem.

Seminar 07112 was held during the same week as Seminar 07111, whose theme was Computational Geometry (CG). CG is a subfield of Computer Science that concerns the design, implementation, and analysis of efficient geometric algorithms on a computer. Analysis includes proofs of correctness and estimates of the amount of execution time and storage space required by the algorithms. Geometric problems are at the core of the problems in cutting, packing, and space allocation, so it was helpful to exchange ideas between the two seminars.

In addition to impromptu discussions between the two groups, there were four planned sessions that involved some, or all, of both groups.

1. This was an invited presentation by Sandor Fekete (Orthogonal Packing Problems and Benchmark Instances). This talk outlined some of his research as well as introducing the seminar to a web site (packlib) which provides access to various benchmark datasets, as well as a bibliography.

2. This was also an invited presentation by a member of the computational geometry community. Dan Halperin's talk (Arrangements and Their Applications: Recent Developments) included information on CGAL (the Computational Geometry Algorithm Library). This library, partly designed by Halperin, includes code to perform some basic geometric operations of importance to cutting and packing researchers.

3. As part of our seminar we held Open Problem Sessions, which was attended by several members from the computational geometry seminar. There was recognition of the need to understand the search landscape for cutting, packing and space layout problems. Rotational Minkowski sum operations are also needed because some packing problems allow arbitrary shape orientations. The Minkowski sum is the vector sum of two point sets and it is very useful in overlap questions that arise in cutting and packing. It was felt that the two communities could work more closely together in developing a set of library functions that the cutting and packing community could call upon (for example, no fit polygons etc.).

4. We also held a joint session (Thursday afternoon), where members of both seminars came together to hear presentations given by representatives from both seminars (see the schedule above).

As a result of all these interchanges, attendees of Seminar 071112 developed a better appreciation of the kinds of geometric problems that cutting and packing researchers need solutions for and Seminar 07112 members became more aware of the progress that has been made in recent years by CG researchers, particularly with respect to the Minkowski sum.

In summary, the seminar was felt to be a great success and there are many opportunities to hold related Dagstuhl seminars in the future.

3.4 Visual Computing – Convergence of Computer Graphics and Computer Vision

Seminar No. **07171** Date **22.04.–27.04.2007**
Organizers: Markus Gross, Heinrich Müller, Hans-Peter Seidel, Harry Shum

Introduction

Due to the importance of visual information for humans, visual computing is at the very core of the technologies enabling the modern information society. New and emerging technologies such as multimedia, digital television, telecommunication and telepresence, or virtual reality further indicate the tremendous potential of visual interaction with computers in the years to come. Typical for the field is the coincidence of very large data sets with the demand for fast, if possible interactive, user-adapted high quality visual display of the results. Furthermore, the user should be able to interact with the environment in a natural and intuitive way.

In order to address the challenges mentioned above, a new and more integrated scientific view of Visual Computing is required that unifies the previously separate "visual" disciplines of computer graphics and computer vision. Computer graphics is traditionally concerned with generating visual interfaces of computers and applications to the user. Computer vision focuses on enabling computers to understand and interpret visual information from static images and video sequences.

Summary of the Seminar

The seminar considered the whole pipeline from data acquisition over processing to rendering, including perceptional issues. This approach made it possible to uncover synergies between computer graphics and computer vision research.

The seminar had three types of sessions: research talks, keynotes, and break-out sessions. Apart from concrete research problems, several fundamental questions were addressed in particular in the latter two formats:

Generation of visual content

As the methods used to generate visual content become more and more complex,

and the data sets used in the modeling process grow, methods from computer vision become an integral part of the data acquisition and modeling pipeline. Here, automated methods are required that make it possible to handle huge amounts of data. Conversely, generative techniques developed in computer graphics can be used to generate auxiliary and intermediate data for computer vision tasks, where knowledge on how to create images helps in understanding images.

In computer graphics, data driven content generation has replaced model-driven techniques in many areas where the models become too complex to handle. Complex models are often acquired using learning techniques. This enables the use of more complex models without the need for designing such models from scratch. In even more complicated cases, tasks can be completed from data alone, either without an underlying model, or with only partial support from a simplified, coarse model. Such methods are being used in rendering and modeling of extremely complex scenes and materials (for instance human skin), and are also applied to physical simulations, where they lead to a reduction to a simpler model.

Analysis by synthesis

From the computer vision direction, computer graphics techniques are used in core vision tasks. In analysis by synthesis approaches, generative techniques produce hypotheses that can then be tested. In many problems, these approaches lead to more robust methods for optimization and learning tasks. Examples for such approaches include methods for face recognition, alignment tasks, tracking.

Level of resolution

Whenever we generate or analyze data in a visual form, the question arises at which level or resolution this should be done. Clearly, imposing reasonable limits on resolution is necessary, but it is unclear what level of detail and what resolution is needed for a realistic, convincing, or simply plausible result. Studying human perception can give hints, and limits in human vision and hearing can be exploited to save costs while delivering an equally convincing experience. This question is of immediate relevance in research concerning level-of-detail representations, not only considering the easier task of geometric simplification, but also model simplification and behavioral simplification.

Engineering versus science

Taking a step back, it is enlightening to ask whether future developments in the field will be due to engineering achievements or scientific insights. As computing power grows, models can become more complex, and more sophisticated numerical techniques can be applied to harder and larger problems. Such advances in engineering have contributed a great part to the rise of physical simulation in computer animation, for instance, and will continue to make important contributions to the progress of the field. However, physical simulations have evolved into a third fundamental approach to gaining scientific insight besides theory and experimentation. Thus, models created from real-world data, or created with the purpose of recreating real-world behavior, may well lead to scientific insights into the studied object, be it crowds, human behavior, or various materials, especially when the models can be

verified against real-world measurements. Such insights gained in visual computing research will have an impact not only in the field itself but also in other subject areas, such as for example biomechanics.

Modeling of human characters

There are several areas in ongoing research that cannot be tackled by computer graphics or computer vision alone. One such problem discussed in this seminar is the modeling of human characters. Specifically, in order to build a believable model of a human that can be used for content generation, automated techniques are needed. Model parameters should be inferred from video data, since manually creating the model is too complex to be feasible. This inference must be able to model subtleties such as the emotional state of the character. This ability will also lead to a deeper understanding of the principles of communication of emotions, which in turn can be used in related tasks in character animation.

In appendices, thoughts and grand challenges identified by two of the break-out sessions are compiled.

Conclusions of the Seminar

It became clear during the seminar that the fields of computer vision and computer graphics are not only closely related, but are mutually dependent. As techniques are exchanged between the fields, computer graphics and computer vision are converging into a discipline of visual computing. Using the knowledge about generative and analytic techniques that is available at both ends of the spectrum leads to the development of more robust and efficient tools able to handle the huge amounts of data that are typically dealt with. The understanding of both aspects of visual data, how to analyze it as well as how to generate it, helps in identifying fundamental principles that govern the processing of visual data in a computer. This knowledge leads to the development of better representations and primitive operations on a well-founded theoretical base, allowing use to replace heuristic and fragile approaches by robust and reliable methods in visual computing.

Appendix A.
Break-out session: Capturing reality

Summary by Leif Kobbelt and Wolfgang Heidrich

In science and engineering research, numerical simulations are quickly evolving as a third fundamental approach besides theory and experimentation. In this context it is becoming increasingly clear that efficient, robust, and mostly automatic techniques are required to capture all possible modes of information on real objects, including shape, material properties, and so forth. At the same time, an ever increasing level of detail for such digital models is also driven by the continued quest for increasingly realistic display of both real-world and synthetic environments.

Hence the problem of capturing reality and handling the resulting data sets is (at least) twofold: (1) how to acquire and merge all the different aspects of a real object of scene –

especially across different levels of detail/resolution and (2) how to efficiently handle the resulting huge amounts of data such that interactive response times become possible.

In the discussion, we considered three different questions:

Which modes/aspects are relevant for visual computing?

One working hypothesis would be to capture reality like it is perceived by humans (without technological support). From this principle, we could, e.g., derive the appropriate spatial resolution (no lightyears and no micrometers) and the kinds of modes. On the other hand one could argue that even not directly perceivable object properties are necessary to eventually be able to simulate its realistic behavior. This question turns out to be another instance of the more fundamental question whether visual computing should target at realistic output or rather at plausible/convincing output.

Data representation?

Here the central question is whether it is desirable to have one universal representation which can serve as a master model and from which more specialized representations can be derived. This would allow for representations which are adapted to the particular requirements in a certain application but at the same time guarantee a proper correspondence between the various modes. One the other hand, if eventually specialized representations are needed anyway, it might not be worth the extra effort to integrate all the partial information into one unified model.

More concretely, with respect to geometry representation, there are polygonal meshes as the today's established universal standard. However, what will be the representation of the future? Depending on whether flexibility or approximation power are the driving forces, polygon meshes might be replaced by unstructured point clouds or by higher order representations such as subdivision surfaces.

Besides data structures, another important question is how much individualized digital models have to be. For surgery planning, it is definitely necessary to have a model of the actual anatomy of a patient. However if one wants to model, e.g., a lawn it might be overdone to store the exact geometry of every single leaf. In this case it would be more appropriate to have one or more "typical" leafs and replicate them multiple times.

Science vs. Engineering?

There are many different technological as well as algorithmic approaches to capture and reconstruct the shape, material, and other physical properties of real objects and scenes. The question is, on which level do we need to improve reality capture systems in order to make significant progress in reliability, precision, and degree of automation. One standpoint is that the existing techniques are in principle sufficient and what is needed is a better implementation and integration ("engineering approach"). On the other hand, it could very well be that many of the open problems in shape and material acquisition that we still have today are due to the fact that the reconstruction principles known today have some intrinsic issues and new approaches have to be explored ("science approach").

Appendix B.
Break-out session: Extraction and retargetting of human physical properties to synthetic characters

Summary by Eugene Fiume

The solutions for modelling humans for character animation differ from similar solutions for biomechanics due to the workflow and usability needs for character generation and animation. Furthermore, virtual actors are almost invariably bad actors. That said, the extraction of physical, behavioural and morphological parameters from real people is a grand challenge for the field. These problems break down into subchallenges such as:

1. Automatically rigging a synthetic character from a video sequence of a human or other animal.
2. Mapping human performance to a synthetic character.
3. Inferring activation sequences and biomechanical properties from video.
4. Transfer of emotional state to characters.
5. Extraction and transfer of human behaviours from video sequences of crowds to synthetic crowds.

3.5 Information Visualization – Human-Centered Issues in Visual Representation, Interaction, and Evaluation

Seminar No. **07221** Date **28.05.–01.06.2007**
Organizers: Jean-Danial Fekete, Andreas Kerren, Chris North, John T. Stasko

Information Visualization (InfoVis) focuses on the use of visualization techniques to help people understand and analyze data. While related fields such as Scientific Visualization involve the presentation of data that has some physical or geometric correspondence, Information Visualization centers on abstract information without such correspondences.

One important aim of this seminar was to bring together theoreticians and practitioners from Information Visualization and related fields as well as from application areas. The seminar has allowed a critical reflection on actual research efforts, the state of field, evaluation challenges, etc. This document summarizes the event.

Information Visualization (InfoVis) is a relatively new research area that focuses on the use of visualization techniques to help people understand and analyze data. While related fields such as Scientific Visualization involve the presentation of data that has some

physical or geometric correspondence, Information Visualization centers on abstract information without such correspondences, i.e., there is no possibility to map this information into the physical world in most cases. Examples of such abstract data are symbolic, tabular, networked, hierarchical, or textual information sources. The ever increasing amount of data generated or made available every day confirms the urgent need of InfoVis tools. There are many possible visual representations but only a fraction are helpful for a given task or application domain. As prerequisite for building a successful visualization, InfoVis combines several aspects of different research areas, such as Scientific Visualization, Human-Computer Interaction, Data Mining, Information Design, Cognitive Psychology, Visual Perception, Cartography, Graph Drawing, and Computer Graphics. Also, aesthetic aspects play a more and more important role: the First Workshop on Computational Aesthetics 2005 in Girona, Spain, and the resulting Dagstuhl Seminar 06221 emphasize such aspects not only in the InfoVis area.

One main goal of this seminar was to bring together theoreticians and practitioners from the addressed research areas as well as from application areas, such as Bioinformatics, Finance, Geo Sciences, Software Engineering, Telecommunication, etc. There are several international conferences that include information visualization topics. Each of them has a slightly different high-level objective. In this context, a consolidation within one seminar appeared to be very beneficial.

Seminar Topics

- Human-Centered Aspects
- Human-Computer Interfaces
- Visualization Techniques and Models
- InfoVis Aesthetics
- User Interaction
- Multimodal Visualization
- Usability
- Scalability
- Quality Measures
- Perception and Cognition (Psychology Backgrounds)
- Prior Knowledge of the Users
- Education and Training
- Large or Mobile Displays
- Novel Visual Representations

- Visual Analytics
- Domain Specific Visualizations
- Evaluations and Empirical Studies

The seminar has allowed a critical reflection on actual research efforts, the state of field, evaluation challenges, etc. Participants also were encouraged to perform system demonstrations of prototypes and environments relevant to the seminar topics.

The organizers and participants decided to publish a book that should document and extend the findings and discussions of this Dagstuhl Seminar. Beforehand, the organizers gained the agreement of Springer Press to publish an LNCS State-of-the-Art issue on the seminar theme. The book will cover the problems discussed in the various sessions in detail. An extended seminar report also is planned to be sent to the Information Visualization Journal (IVS) published by Palgrave.

3.6 Scientific Visualization

Seminar No. **07291** Date **15.07.–20.07.2007**
Organizers: David S. Ebert, Hans Hagen, Kenneth I. Joy, Daniel A. Keim

Scientific visualization (SV) is concerned with the use of computer-generated images to aid the understanding, analysis and manipulation of data. Since its beginning in the early 90's, the techniques of SV have aided scientists, engineers, medical practitioners, and others in the study of a wide variety of data sets including, for example, high performance computing simulations, measured data from scanners (CAT, MR, confocal microscopy), internet traffic, and financial records. One of the important themes being nurtured under the aegis of Scientific Visualization is the utilization of the broad bandwidth of the human sensory system in steering and interpreting complex processes and simulations involving voluminous data sets across diverse scientific disciplines. Since vision dominates our sensory input, strong efforts have been made to bring the mathematical abstraction and modeling to our eyes through the mediation of computer graphics. This interplay between various application areas and their specific problem solving visualization techniques was emphasized in the proposed seminar.

Reflecting the heterogeous structure of Scientific Visualization, we will focus on the following:

Visual Analytics:
The fields of information analysis and visualization are rapidly merging to create a new approach to extracting meaning from massive, complex, evolving data sources and stream. Visual analytics is the science of analytical reasoning facilitated by interactive, visual interfaces. The goal of visual analytics is to obtain insight into massive, dynamic and often conflicting pieces and formats of information; to detect the expected and to discover the unexpected; and to yield timely assessments with evidence and confidence levels.

Quality Measures:
It is vital for the visualization field to establish quality metrics. An intrinsic quality metric will tremendously simplify the development and evaluation of various algorithms. The establishment of quality metrics will also advance the acceptance and use of visualization in industrial and medical applications.

Ubiquitous Visualization:
As ubiquitous computing is getting increased attention, also visual display of everywhere available data is necessary. Challenges include: heterogeneous output devices, novel interaction metaphors, network bandwidth (availability, reliability), graceful degradation of algorithms with respect to largely varying resources, invivo visualization (real time, no pre-processing, robust).

Multifield and multiscale visualization:
The output of the majority of computational science and engineering simulations is typically a combination of fields, so called multifield data, involving a number of scalar fields, vector fields, or tensor fields. Similarly, data collected experimentally is often multifield in nature (and from multiple sources). The ability to effectively visualize multiple fields simultaneously, for both computational and experimental data, can greatly enhance scientific analysis and understanding. Multiscale problems with scale differences of several orders of magnitude in CFD, nanotechnology, biomedical engineering and proteomics pose challenging problems for data analysis. The state of the art in multiscale visualization considerably lags behind that of multiscale simulation. Novel solutions to multiscale and multifield visualization problems have the potential for a large impact on scientific endeavors.

Our Dagstuhl workshop was arranged into three general types of sessions: Senior Short Talks, In-Depth Research Talks, and Break-out Sessions. The senior short talks were designed to pose research challenges and approaches for the future and had a very short presentation followed by long, lively discussions. The in-depth research talks allowed for detailed presentation of research approaches and projects, as well as a special session on education challenges/approaches within scientific visualization. The break-out sessions were used to stimulate group focused discussions on important topics and actions for the future.

Chapter 4

Artificial Intelligence, Computer Linguistic

4.1 Normative Multi-agent Systems

Seminar No. **07122** Date **18.03.–23.03.2007**
Organizers: Guido Boella, Leon van der Torre, Harko Verhagen

Norms like obligations, permissions and prohibitions have been proposed in multi-agent systems to deal with coordination and security issues of multi-agent systems, to model legal issues in electronic institutions and electronic commerce, to model multi-agent organizations, and so on. In the context of this workshop we use the following definition: A normative multiagent system is a multiagent system together with normative systems in which agents on the one hand can decide whether to follow the explicitly represented norms, and on the other the normative systems specify how and in which extent the agents can modify the norms.

Since norms are explicitly represented, the question should be raised how norms are represented. Norms can be interpreted as a special kind of constraint, and represented depending on the domain in which they occur. Since not all agents behave according to the norm, norms are not hard constraints, but soft constraints. The question will be raised how the system monitors the behavior of agents and enforce sanctions in case of violations, or reward good behavior. Also, the question will be raised how to represent permissive norms, and how they relate to obligations. For example, the permission to access a resource under an access control system cannot be represented as a constraint. Finally, the question will be raised how norms evolve.

In electronic commerce research, for example, cognitive foundations of social norms and contracts are studied. Protocols and social mechanisms are now being developed to support such creations of norms in multiagent systems. When norms are created, the question how they are enforced will be raised. For example, when a contract is violated, the violator may have to pay a penalty. But then there has to be a monitoring and sanctioning system, for example police agents in an electronic institution. Such protocols or roles in

a multiagent system are part of the construction of social reality, and such social realities are constructed by constitutive norms. This again raises the question how to represent such constitutive or counts-as norms, and how they are related to regulative norms like obligations and permissions.

Norms should therefore be represented as a domain independent theory, for example in deontic logic. Deontic logic studies logical relations among obligations, permissions, prohibitions and counts-as conditionals, and more in particular violations and contrary-to-duty obligations, permissions and their relation to obligations, and the dynamics of obligations over time. Therefore, insights from deontic logic can be used to represent and reason with norms. Deontic logic also offers representations of norms as rules or conditionals. However, there are several aspects of norms which are not covered by constraints nor by deontic logic, such as the questions where do norms come from, how are they created by a single legislator, how do they emerge spontaneously, or how are they negotiated among the agents. Moreover, what is the relation between the cognitive abilities of agents and the global properties of norms, how can agents acquire norms, how can agents violate norms, how can an agent be autonomous, how are group obligations distributed over the members of the group, and so on.

Not only the relation between norms and agents must be studied, but also the relation between norms and other social and legal concepts. How do norms structure organizations? How do norms coordinate groups and societies? How about the contract frames in which contracts live, and the legal contexts in which contract frames live? Though in some normative multiagent systems there is only a single normative system, there can also be several of them, raising the question how normative systems interact. For example, in a virtual community of resource providers each provider may have its own normative system, which raises the question how one system can authorize access in another system, or how global policies can be defined to regulate these local policies.

Summarizing, normative multiagent systems study general and domain independent properties of norms. It builds on results obtained in deontic logic for the representation of norms as rules, the application of such rules, contrary-to-duty reasoning and the relation to permissions. However, it goes beyond logical relations among obligations and permissions by explaining the relation among social norms and obligations, relating regulative norms to constitutive norms, explaining the evolution of normative systems, how agents interact with norms, and much more.

Information for the invited scientists:

During the past years we observe a rising interest in computer science and in social theory in multi-agent systems, which is moving more and more from the individual, cognitive focussed agent models to models of socially situated agents. In particular attention is given to normative multi-agent systems, because the use of norms is the key of human social intelligence. If artificial agents are to display behavior equal to human intelligent behavior or collaborate with humans, norms are essential. Norms have been mentioned in agent research for quite some time, but lately we can see that the research field has matured.

The NorMAS05 symposium at the 2005 AISB conference has raised the questions of which problems must be solved in norms, and how to approach these problems. The need for models, theories and tools in multi-agent systems has also been observed in the related area of "deontic logic of computer science", where its biannual workshops were interested in applications in multi-agent systems (DEON 2004) and artificial social systems (DEON 2006). However, the gap between the DEON community and the multi-agent systems community, due to the fact that the DEON community restricts itself to one formal language, a branch of modal logic called deontic logic, whereas in multi-agent systems also other models and languages are used, ranging from game theory to Z specifications.

Norms are also considered in various workshops in multi-agent systems concerned with for example legal issues, organizations, institutions, and so on. We have observed the need for an occasion to discuss and compare the various proposals now under development. The traditional agent workshops are not sufficient, because they are not just concerned with the norms in multi-agent systems but also with some other issues (coordination, security, etc). However, there is no common theory of normative multi-agent systems, due to the lack of a universal theory in the social sciences. Therefore, presently many multi-agent system researchers are developing their own ad hoc theories and applications.

The expected results of the seminar are to have a clear view of the ontological similarities and differences between the use of the different concepts connected to norms in the research disciplines. The goal of the seminar is to gather specialists from different areas such as computer science, logic, sociology, and cognitive science to discuss the fundamental concepts and ontologies connected to the use of norms in human and artificial systems, more in particular the use of norms as a mechanism in multi-agent systems and the use of multi-agent systems to study the concept and theories of norms and normative behavior.

4.2 Formal Models of Belief Change in Rational Agents

Seminar No. **07351** Date **26.08.–30.08.2007**
Organizers: Giacomo Bonanno, James Delgrande, Jerome Lang, Hans Rott

The theory of belief revision studies how a rational agent should change its beliefs when receiving or perceiving new information about the environment. This new information could include objective properties of the actual world, occurrences of events, and, in the case of multiple agents, public or private communications among agents (possibly concerning their beliefs and preferences) as well as actions taken by other agents. Not surprisingly, this area has been of interest to researchers in different communities.

The initial research in belief change came from the philosophical community, wherein belief change was studied generally from a normative point of view (that is, providing axiomatic foundations about how rational agents should behave with respect to the information flux). Subsequently, computer scientists, especially in the artificial intelligence (AI) and the database (DB) communities, have been building on these results. Belief change, as studied by computer scientists, not only pays attention to behavioral properties characterizing evolving databases or knowledge bases, but must also address computational issues such

as how to represent beliefs states in a concise way and how to efficiently compute the revision of a belief state.

The most important question in Game Theory is how to rationally form a belief about other players' behavior and how to rationally revise those beliefs in light of observed actions. Traditionally Game Theory has relied mostly on probabilistic models of beliefs, although recent research has focused on qualitative aspects of belief change. A new branch of logic, called Dynamic Epistemic Logic, has emerged that investigates the epistemic foundations of game theory from the point of view of formal logic. Another, related, new field of research, called Social Software, maintains that mathematical models developed to reason about the knowledge and beliefs of a group of agents can be used to deepen our understanding of social interaction and aid in the design of successful social institutions. Social Software is the formal study of social procedures focusing on three aspects: (1) the logical and algorithmic structure of social procedures (the main contributors to this area are computer scientists), (2) knowledge and information (the main contributors to this area are logicians and philosophers), and (3) incentives (the main contributors are game theorists and economists).

There are various newly emerging links between the research areas mentioned above. The purpose of the Workshop was to bring together researchers from all these different areas; these researchers normally do not meet together. Workshops such as this one promote an exchange of ideas and cross-fertilization across different fields.

We found the Workshop successful, especially on the following two achievements: first, the seminar made participants aware of a commonality of interests across different disciplines; second, it suggested new directions for research that will probably be taken up by researchers in the next couple of years.

Where is the field going? We can mention at least two emerging issues:

- the field is broadening with respect to theoretical underpinnings and is beginning to incorporate notions from game theory and social choice theory. It is also broadening with respect to application areas, moving beyond traditional areas in AI and database systems, to include areas in description logics, the semantic web and economics.

- there is an emerging focus on epistemic notions having to do with communicating, negotiating, competing, and collaborating agents. Dynamic epistemic logic seems to have an important role to play here.

Moreover, it looks like belief merging and iterated belief revision are still hot topics and will remain so for the next few years.

For the future, we plan further workshops to encourage continued interdisciplinary interactions.

4.3 Mobile Interfaces Meet Cognitive Technologies

Seminar No. **07371** Date **09.09.–14.09.2007**
Organizers: Jan-Olof Eklundh, Ales Leonardis, Lucas Paletta, Bernt Schiele

The ubiquity and miniaturization of mobile sensing devices as well as the increase of computational power of mobile and handheld devices has led to a large increase in research, algorithms and applications in the area. In the near future mobile imaging technology will become ubiquitous such as in camera phones, vision enhanced handhelds, and wearable cameras. Beyond passive data display and transmission, future information technologies will provide smart mobile services being capable of real-time analysis, purposeful selection and interpretation of enormous quantities of sensed and retrieved data. Image understanding for mobile multimodal interfaces would make new approaches possible in object recognition, context awareness, and augmented reality, aiming towards application scenarios of personal assistance, mobile work, and assistive services. The research challenges are not only efficiency, speed and low-complexity algorithms, but also additional demands on robustness of interpretation because of the mobility of the devices as well as the impact of dynamically changing and noisy conditions within urban environments. Since multi-sensor information analysis affords cueing and indexing into databases with nowadays huge information spaces, the sensed data have to be processed in an intelligent way to provide "in time delivery" of the requested relevant information. Knowledge has to be applied in an intelligent way about what needs to be attended to, and when, and what to do in a meaningful sequence, in correspondence with multi-sensor feedback.

Mobile interfaces relate these non-trivial system aspects to the dimensions of human presence, coupling interaction patterns in human behaviours with the system requirements of the multimodal interface. This will bind future mobile technologies with the emerging science on artificial cognitive systems that focuses research towards a technology that can interpret information, act purposefully and autonomously towards achieving goals. The development of this science already borrows insights from the bio-sciences and artificial intelligence, and provides a new framework with revolutionizing insights about perception, understanding, interaction, learning and knowledge representation. As we rely more and more on complex systems in mobile interaction with both real and virtual environments, we will need cognitive technologies to cope with in a focused and structured way.

Goals and Content of the Seminar

The goal of this seminar is to provide an interdisciplinary forum for researchers from Ubiquitous Computing, Artificial Intelligence, Artificial Cognitive Systems, and Cognitive Sciences to communicate and discuss the scientific challenges in designing mobile cognitive technologies for urban scenarios.

The key questions to be discussed are:

- What are efficient methodologies to represent contextual knowledge from multi-sensor information?

- How is contextual knowledge applied in mobile interface technology?
- What design strategies are successful in mobile attentive interfaces?
- What is the contribution of machine learning to prediction based mobile services?
- Will existing models of spatial cognition enhance mobile interface technologies?
- What kind of knowledge representation do we use to support dynamic user interaction?
- Which level of abstraction should be selected for the filtering of the incoming stream of multimodal information?
- What is the human style of interaction in typical urban scenarios involving mobile interfaces?
- What are the specific challenges for computer vision on mobile imagery and how can AI contribute to corresponding system solutions?

With respect to the highly interdisciplinary background of the theme, the structure of the program should on the one hand enable the participants to be fully informed about the key scientific viewpoints via overview lectures. We think of daily key talks (50') that will be followed by shorter presentations about individual, complementary views (30'), together making up a lecture session on a specific theme. On the other hand, small discussion and working groups will be involved to exchange their views on the presented concepts and work in an interactive way. Finally, poster sessions will be offered in order to reinforce the understanding of the individual viewpoints in an even more personal way.

Chapter 5

Software Technology

5.1 Software Dependability Engineering

Seminar No. **07031** Date **14.01.–19.01.2007**
Organizers: Rance Cleaveland , H. Dieter Rombach, Mary Shaw

During the last few years, functionality and complexity of software products has been increasing dramatically. Customers require more and more functionality and especially high quality products tailored to their particular environment. Therefore, software development faces the challenge to reduce cost, effort, and time-to-market of the software products, but simultaneously ensuring the delivery of the required software product that fulfills the customer's functional and especially quality expectations.

As a direct consequence, acceptable products may not satisfy all quality requirements perfectly, which is, however, rarely communicated explicitly and clearly – in particular among software practitioners. As a consequence, many software approaches are applied with the implicit intention of achieving "best" quality with respect to all kinds of product characteristics. This is not a problem as long as the typical process configurations always fully satisfy all product and project requirements. This can, however, only rarely be maintained in practice over a long period of time. Changes to functionality, technology, quality, or organization force development approaches to be continuously adapted or optimized.

Therefore, organizations need to understand how to explicitly model their various dependability requirements and how to define and use strategies to meet these requirements. Additionally they must understand the associated costs and trade-offs of different strategies.

The seminar addresses these needs by discussing various needs for dependability especially emphasizing dependability of end products over qualities of intermediate development artifacts. The workshop facilitates the discussion on how to integrate dependability requirements with organizational and project constraints.

During the workshop various groups will focus on selected aspects of software dependability engineering or approach it in various ways. Participants are invited to propose topics.

5.2 Programming Paradigms for the Web: Web Programming and Web Services

Seminar No. **07051** Date **28.01.–02.02.2007**
Organizers: Richard Hull, Peter Thiemann, Philip Wadler

The web raises a variety of new programming challenges. To name a few: programming user interfaces at the level of the web browser, data-centric approaches, and attempts to automatically discover and compose web services. This seminar brought together researchers from the web programming and web services communities. Both groups had much to learn from each other, and the focus on programming paradigms was a useful perspective on the diverse web community.

"Web (application) programming" describes writing software ("web applications") that relies on a web browser as its user interface. Typical tasks for web programming include the generation of dynamic web pages, accessing databases, querying web services, and dealing with concurrency. Web applications often involve many "tiers", each of which is a homogenous level of software which interacts over a network with other tiers; a typical application might involve, for example, a client (a web browser programmed with JavaScript, VBScript, or Flash), a server (programmed with Java, Ruby, PHP, Python, or Perl), and a database (programmed with SQL or XQuery); connecting these tiers can be a key challenge for web programmers. As another challenge, web applications do not presently support modes of user interaction as rich as their desktop counterparts, and providing anything other than the simplest form-based interaction can be a great difficulty for programmers—thus, we asked, how can web programming paradigms support coding rich user interfaces?

"Web services", by contrast, are programs that interact primarily with other software systems using web technologies. The web-services paradigm, which might be viewed as an instance of the Service-Oriented Architecture (SOA), provides both a framework and specific interfaces (e.g. SOAP, WSDL) for a new generation of distributed software. The paradigm provides for rich flexibility in creating services that use other web services. To date, programming of web services has focused largely on adaptations of workflow approaches to a peer-to-peer framework, and the Business Process Execution Language (BPEL) has emerged as the industrial programming language of choice. There has also been significant research on "semantic web services", which provide explicit mechanisms to represent and reason about the impact of services on the world, as well as their messaging and internal behavior. Frameworks such as WSDL-S, OWL-S, SWSO, and WSMO may provide the basis for programming paradigms to work effectively in this context.

Prime discussion topics were: the application of these techniques to web applications, browser-based programs, and web services, programming languages for the web, scripting, authoring, type checking, databases, web service semantics, service composition, process and data flow, XML and other data manipulation, concurrency, sessions and transactions, performance, and scalability.

To maintain a focus on programming, speakers were asked to center their talk on actual

code that illustrates their research. Here 'code' was broadly interpreted to include a program in a programming language, a formula of logic, a specification, or a query.

As an outcome of the seminar we expected to understand better the interplay between the various styles of programming for the web, along with proposals towards a more unified approach to such programming. Elsewhere, we have started to compose a list of the key scientific challenges (or at least discussion items leading in that direction) in this domain.

The meeting was very productive; it provoked many new ideas and provided new perspectives on the topic. The participants learned a lot from each other—in particular, there was a lively exchange of knowledge between the programming languages and the database communities. We hope to further consolidate the results in an article that provides research directions for web programming and related areas.

5.3 Autonomous and Adaptive Web Services

Seminar No. **07061** Date **04.02.–09.02.2007**
Organizers: Jana Koehler, Marco Pistore, Amit P. Sheth, Paolo Traverso, Martin Wirsing

The Dagstuhl Seminar on Autonomous and Adaptive Web Services brought researchers together whose current research interests are centered around web services composition and adaption.

Web Services provide the universal basis for the integration of networked applications and business processes that are distributed among the most disparate entities, both within and across organizational borders. The fundamental idea of Web Services is that applications are built by interacting with and composing external components – services – that are available on the Web, and that are not under the control of a single party or stakeholder. A new challenge arises from this idea: the success of service oriented applications is unavoidably depending on the capability of a service to autonomously adapt to an environment that is not fully under control. Therefore a need exists for techniques that enable the flexible composition and adaption of web services.

The participants of the seminar discussed to which extent a fully automatic composition and self-adaptation of web services is possible and which prerequisites have to be fulfilled in order to enable such a high degree of autonomy. Two main technologies were reflected in the presentations: First, approaches to the semantic web that improve the precise semantic descriptions of web services played an important role in the discussion. The resulting semantic web services and their role in service-oriented architectures were discussed by several participants.

Second, formal analysis and verification techniques that provide the foundation for composition and adaptation algorithms that sometimes made use of semantic web services were presented. Petri net techniques, model checking and process calculi were discussed in detail and their opportunities and limitations explored.

A tools session offered interesting insights into the capabilities of various tools and allowed participants to compare and position their approaches in detail.

The seminar showed that web services, semantic web, and service composition and adaption are quickly moving areas at the moments. Progress in the various areas enables novel solutions to be built that enable fascinating applications.

5.4 End-User Software Engineering

Seminar No. **07081** Date **18.02.–23.02.2007**
Organizers: Margaret M. Burnett, Gregor Engels, Brad A. Myers, Gregg Rothermel

The number of end users creating software is far larger than the number of professional programmers. These end users are using various languages and programming systems to create software in forms such as spreadsheets, dynamic web applications, and scientific simulations. This software needs to be sufficiently dependable, but substantial evidence suggests that it is not.

Solving these problems involves not just software engineering issues, but also several challenges related to the users that the end user software engineering intends to benefit. End users have very different training and background, and face different motivations and work constraints, than professional programmers. They are not likely to know about such things as quality control mechanisms, formal development processes, system models, language design characteristics, or test adequacy criteria, and are not likely to invest time learning about them.

It is important to find ways to help these users pursue their goals, while also alerting them to dependability problems, and assist them with their explorations into those problems. Further, it is important to work within the contexts with which these users are familiar, which can include programming environments that have not been directly considered by software engineering or programming languages researchers.

These challenges require collaborations by teams of researchers from various computer science subfields, including specialists in end-user-programming (EUP) and end-user development (EUD), researchers expert in software engineering methodologies and programming language design, human-computer interaction experts focusing on end-user programming, and empiricists who can evaluate emerging results and help understand fundamental issues in supporting end-user problem solving. Collaborations with industrial partners must also be established, to help ensure that the real needs of end-user programming environments in industry are met.

This Dagstuhl seminar was organized in order to bring together researchers from these various groups and with the various appropriate backgrounds, along with an appropriate selection of industrial participants. The seminar allowed the participants to work together on the challenges faced in helping end-user programmers create dependable software, and on the opportunities for research addressing these challenges. Our goals were to help these researchers better understand (1) the problems that exist for end-user programmers, (2) the environments, domains and languages in which those programmers create software, (3) the types of computing methodologies (especially in the areas of software engineering

and programming language design) that can be brought to bear on these problems and in these domains, and (4) the issues that impact the success of research in this area. In addition, an overarching goal was to build awareness of the interdisciplinary connections and opportunities that exist for researchers working in the area.

The seminar included several tutorial-style presentations by experts on software engineering, programming languages, human-computer interaction, and empirical studies in relation to end-user software engineering. The program was complemented with brief presentations by some participants on topics of a more specialized nature, grouped into sessions on related topics. We also incorporated system demonstrations of prototypes and environments relevant to the topics. Ample time was allowed for interactive discussion sessions.

Most of the seminar participants provided white papers summarizing their primary interests in the area, including work that they are doing and open problems. These white papers are compiled into the seminar proceedings. Additional contributions to the seminar were provided as slides, and are available on the Dagstuhl website for the seminar.

5.5 Tools for the Model-based Development of Certifiable, Dependable Systems

Seminar No. **07241** Date **10.06.–15.06.2007**
Organizers: Michaela Huhn, Hardi Hungar, Doron A. Peled

Introduction

The development of software for dependable systems on which the safety or security of individuals, organizations, and property may rely has become an important application and research field. In many cases, law enforces certification as a prerequisite for the introduction of a technical, dependable system. The certification formally assures that the systems and its development process meet the technical standards and all efforts have been made to reduce the risks. The complexity and discrete nature of software makes an assurance of the required properties of the software components of a dependable system extremely difficult. This applies even more as the software often constitutes the control in a so-called embedded system where coupling electronic and mechanical components is still a challenge. The relevant national and international standards (e.g. IEC 61508) recommend formal methods for the development of systems that shall be certified for the higher safety integrity levels. However, deriving constructive design guidelines and formally verifiable software constraints from safety requirements on the system level is still a challenging problem.

In many industries developing high-assurance products, model-based design is already well established and a number of tools used in safety-critical software design are founded on formal semantics and support in parts automated code generation, formal analysis, verification or error detection. But certification requires more than formal semantics of a

modelling notation or the formal verification of the artefacts of a particular development step. Formal methods and tools have to be embedded in a seamless design process which covers all development phases and which includes an efficient construction of a safety case for the product. Moreover, whereas most (semi-)formal modelling approaches focus on functional issues additional concepts are required for dependable systems like fault tolerance, timing or security. Even if these concepts are addressed – as several UML profiles do – they are supported at most rudimentarily by the tools.

Some Statistics

The workshop aimed at bringing together people actually working on the certification of dependable systems with researchers who develop and validate (semi-) formal methods and tools for modelling and verification. About 30 researchers and practitioners followed the invitation. The program featured about 25 presentations, 14 research presentations from academia, 4 reports on experiences in the certification processes of particular products and 6 about practical issues from industrial participants. In addition, a discussion session on the status quo of formal methods in the certification of dependable systems was arranged.

The Railway Level Crossing Case Study

To focus discussions and to directly compare different approaches a case study on the design of the software control for a simple railway level crossing was given to the participants beforehand. Several participants addressed the case study in their presentation. So it was shown that some immediate shortcomings in the requirement specification could be detected with simple static analysis techniques applied on an abstract design model. This observation was independent of formal modelling notations actually used. A more subtle ambiguity was detected by those approaches that employed formal verification on a behavioural model of the level crossing. Another participant used a generalized version of the case study to prove scalability of a verification approach. Last but not least, the case study has set up the discussion to what extent implicit assumptions are revealed when transferring an informal requirements specification into a formal model and how to deal with different categories of implicit assumptions methodologically.

Discussion on the Status Quo of Formal Methods and Certification of Dependable Systems

On each of the issues raised during the discussion, there was a wide spectrum of partly divergent opinions and observations. A summary of each of the main themes follows below.

Mathematical Rigour

On the one hand, it was reported, that all necessary techniques were available to treat the subject satisfactorily. By a relevant number of pilot projects it has been proven that

systems of impressive size can be verified working with theorem proving or model checking on the basis of a cleanly and well defined semantics for systems and problem statements. A decomposition into well-built layers, each verified itself, and dedicated construction of relations between layers, was considered to be a successful approach. This goes very well with the model-based design paradigm, because most model-based development processes provide modelling guidelines for different levels of abstraction. The proponents of a rigorous formal approach were convinced that even the popular counter-arguments – as ongoing problems with scalability of formal verification, the high investments needed to develop the initial foundation of a formal process (e.g. verified layers), and the severe shortage of experts who have profound knowledge on formal methods and the application domain – would merely be surmountable obstacles.

This positive view was not shared by everybody. First of all, a number of participants considered the mentioned obstacles to be substantial. Additionally, the maturity of formal methods and tools was challenged: So far, a relevant portion of the success stories was achieved with prototype tools in a research environment. These are not easily transferred into the fields of practice, not only because matters like tool support, robustness and scalability, and domain specific adaptations have to be solved for industrial development. The most important barrier for introducing academic tools in dependable system industries is that certification imposes high requirements not only on the products and the process but also on each tool used in the development and in particular for quality assurance.

Verification and Validation Tools

There was a general agreement that tool support for verification and validation (V&V) is indispensable and will grow in its importance in the following years. There were of course, again, differing viewpoints.

One observation concerned the introduction of new analysis methods like hybrid model checking into the field of (control) engineering. Up to now, these model checkers have not yet shown much promise to be able to handle even medium size designs. Are, henceforth, the "classical" engineering techniques in this field going to be replaced in the foreseeable future? And how are they, then, going to be integrated with formal specifications and reasoning?

A similar integration concern has been raised with respect to the growing usage of automatic simulation or test generation tools. Though considerable progress has been made in automation techniques, it is difficult to profit from a tool which does not guarantee complete coverage. Related to this issue is the question of how to find the "interesting" points via simulation or testing? Hence, these techniques have to be equipped with means to direct the search to those parts of the state space where complex and error-prone behaviour occurs, i.e. interactions between components from different suppliers, or mode transitions in the presence of exceptional behaviour. Moreover, the technique should be able to identify dead code or unreachable parts of the state space or other anomalies indicating potential errors.

To summarize, it shows more promising to specifically tailor a tool to a particular purpose explicitly required by regulations than (naively) introducing an unspecific test genera-

tion tool into the V&V process. For example the avionics standard DO-178B requires MC/DC coverage, which can be addressed by a dedicated procedure relying on abstract interpretation.

If requirements are not that strict, as for instance it is often the case in the automotive domain, it is somewhat easier to profit from the availability of high-level models. This argument gets detailed in the section on processes.

Specification

The question of whether specification has to be formal is controversial. Clearly, formality helps to uncover errors and to avoid misinterpretation, and eases subsequent automation and application of formal techniques. Additionally, the ubiquitous UML carries the promise of an emerging de-facto standard of becoming a lingua franca, understood across domains.

There are obvious objections concerning the communicability of formalisms and the difficulty of checking consistency and completeness. These are more severe deficiencies in a formal than in an informal specification, since the latter appeals more to the developer's experience and competence. Also, it was said that specification is "inherently an iterative process", where some aspects cannot be fixed before at least a prototypical exploration of the design is applied. That specification cannot be fixed up front is experienced even in cases where standards require this explicitly, like the EN 50128. Often, customers require late changes to a product. To incorporate such changes into a specification is usually more difficult when the specification is formal.

Finally, UML is not naturally suited for every kind of application: Parts of the UML are not unambiguously defined, others presume a particular model of execution as for instance statecharts with event pools and run-to-completion semantics, that does not match the semantic world of every application. For modelling the hardware architecture appropriately, which is an integral part of most design and analysis tasks in dependable system development, the UML adoption SysML plus profiles like MARTE are needed.

There was wide agreement that there is no one language for formal specification, nor could there be one. Application-specific languages certainly have their place, as well as weak formalisms like the UML for communication purposes. Whether one should strive for full formality depends at least today on the means available and on the requirements.

Process

The specification is only one of the artefacts to be produced in a design process. The seminar's focus was on dependability and safety, so their role was prominently discussed. In several domains standards prescribe a process skeleton; and a number of well established industrial processes for developing dependable and certifiable systems instantiate and refine these skeletons.

Typically, certification and thus development processes for dependable systems necessitates construction of an argument of safety in a formalised document – a so-called *safety*

case. It is considered advisable to take specific care so that the safety case construction can be done hand-in-hand with the development, i.e. integrate safety arguments with specification refinement and implementation. Today, this is not yet done satisfactorily efficient. Improvements will most likely be possible on a basis of formal development artefacts. Some safety concerns give rise to additional functional requirements, e.g., when redundancy or watchdog constructs are introduced as safety measures. Here, a tight integration of the design steps and the construction of the safety case is particularly recommended.

Another issue is the structure in the line of arguments in the safety case itself. Approaches like *Goal Structuring Notation* improve the conciseness and understandability of reasoning. Moreover, context information and constraints, limiting the applicability of a particular argument or reasoning method, can be made explicit. However, full integration with model-based design methods or formal V&V methods is missing so far.

In the automotive area, a domain-specific standard for the development of safety related electronic programmable systems is still pending (e.g. ISO WD 26262). Here, the development of automotive systems is performed according to an iterative, approximative process. First a core control model is developed on the basis of an ideal plant model. In the next step the normal behaviour of the control model is explored and validated in the real plant. Then exceptional behaviour and specific case treatment is added to the control model. Then a fine-tuning and optimization phase follows. Finally, the control model is transformed into target specific code, which is validated against the runs of the ideal model. In all phases, design proceeds based on the division of labour between OEM and suppliers.

Conclusion

It was commonly agreed that formal specification has already reached the stage of being an effective support in the development of software-intensive dependable systems and that its role will increase in the future.

Technical progress in verification, for instance in component-oriented reasoning (assume-guarantee proofs) and (semi-)automated abstraction techniques, significantly expand the potential for applying formal methods on complex systems.

A tighter integration of models for design and models for verification has already begun and proven to be a key factor for the introduction of formal methods into the industrial practice. A number of verification approaches directly start with design models from UML or Matlab/Simulink as analysis input and offer seamless tool integration. However, these methods are often restricted to one particular verification goal that is considered relevant in one design phase or to a single concern like functional correctness or timing. Thus, a major challenge for the future will be to integrate formal approaches dealing with the different concerns that contribute to safety like functional correctness, reliability, timing, et cetera.

5.6 Mining Programs and Processes

Seminar No. **07491** Date **02.12.–07.12.2007**
Organizers: Abraham Bernstein, Harald Gall, Tao Xie, Andreas Zeller

Each software system has a history: a history of *changes* during the software development process, a history of *executions* during testing and production, and a history of *successes and failures* that occurred during these executions. Most of this history is recorded in *software archives*. In the last years, researchers have begun to analyze these archives, exploring a huge mass of data that can be mined, abstracted, and leveraged:

- Discovering patterns in program runs can help establish *abstractions* that characterize similar runs.
- Analyzing the history of a product can tell how changes in software are *related* to software features -notably successes and failures.
- Changes, executions, and failures are all *intercorrelated*: understanding how they influence each other helps in understanding how software should be built.

One issue that researchers must cope with is the sheer *volume of data*: programs experience thousands of changes, are tested in millions of runs, and execute in billions of cycles. This volume of data calls for advanced data mining techniques that assist in extracting suitable abstractions from programs and software development processes.

For data miners, the analysis of software as data offers a number of challenges:

Software code as data has logical structure.
The software structure can be used as background knowledge in the data mining process.

Software-related data is multi-relational.
Code segments might be related to other code segments (e.g., inheritance, method calls, and co-changes)

Software-related data evolves over time.
Changes may have consequences later (e.g., changing a location may trigger a bug report).

Each of these challenges is already subjects of active research in the data mining or machine learning community. Combining these challenges, paired with the massive volume of data, might warrant new approaches.

In this seminar, we therefore want to bring together researchers in data mining with researchers extracting information from programs and processes. Our main concern is to exploit the *synergy* of these communities and to provide a platform to forge new collaborations.

Participants are invited to present a few plenary talks and demos of new tools, beside which the seminar will provide ample opportunities for small working groups on themes suggested by the participants. We expect the seminar to result in ample crossfertilization between the different research areas and to show up exciting directions for improving the understanding of real-life programs and their history.

Chapter 6

Applications, Multi-Domain Work

6.1 Power-aware Computing Systems

Seminar No. **07041** Date **21.01.–26.01.2007**
Organizers: Luca Benini, Naehyuck Chang, Ulrich Kremer, Christian W. Probst

The program of the Dagstuhl seminar 07041 on Power-aware Computing Systems featured presentations of about 25 participating researchers from academia and industry. They were chosen to represent major areas in targeting the energy consumption of a computing system—Applications, Compilers, Virtual-execution Environments, Operating Systems, and Hardware.

In order to continue the work of the predecessor Dagstuhl seminar held in 2005, the results of that seminar were discussed, with the aim of developing a vision of challenges, problems, and research activities in some of the key areas identified in 2005. The first part of the seminar was dedicated to lively discussions that led to the identification of three areas that were considered being most interesting. As a result, three groups were formed to further identify challenges and opportunities. The results of these groups are presented in the report. In addition, abstracts of the presentations as well as work-in-progress papers are published in the proceedings.

The second Dagstuhl seminar on Power-aware Computing Systems picked up the discussion results of its predecessor, and continued the discussion of challenges in the area. We think that the results, partly described in the report, partly described in the papers published as part of the seminar proceedings, are suited to give the involved communities ideas for future challenges. We would like to thank all participants of the seminar for making it a fruitful and inspiring event—and especially Dagstuhl's wonderful staff, for their support both before and during the seminar.

6.2 Experimental Fluid Mechanics, Computer Vision & Pattern Recognition

Seminar No. **07121** Date **18.03.–23.03.2007**
Organizers: Jean-Paul Bonnet, Etienne Mémin, Christoph Schnörr, Cam Tropea

1 Overview

Recent advances in imaging and measurement techniques have enabled the generation of huge amounts of data in the field of Experimental Fluid Mechanics, providing a unique basis for the understanding of spatial structures in unsteady flows. Progress in this field has an immediate impact on different areas ranging from aerodynamics to biology, and related industrial applications.

The design of computational approaches for the evaluation of image data of fluids raises a range of unsolved problems. A non exhaustive list of these problems includes multiscale image representation, image motion computation, image sequence segmentation, representation of physical prior knowledge, probabilistic inference, spatiotemporal event recognition, and stochastic models for tracking. These fundamental issues are investigated in Computer Science (Computer Vision, Pattern Recognition) as well, although in connection with different applications areas. Despite a confluence of methodological interests, there has been surprisingly only little interaction between Computer Science and Experimental Fluid Mechanics, so far.

In view of the scientific importance, two research programmes have been established quite recently, bringing together researchers from Fluid Mechanics and Computer Science: the national German priority programme on Image Measurements in Fluid Mechanics (DFG-SPP 1147), and the European project Fluid Image Analysis and Description.

Objective of the Dagstuhl seminar is to address these issues in a larger international context with leading experts from Fluid Mechanics and Computer Science.

2 Topics to be discussed are:

1. Image Measurements, Image Processing
2. Computer Vision and Spatio-Temporal Pattern Recognition
3. Synergy of Experiments and Computer Simulations in Fluid Mechanics
4. Applications

3 Goals

The Dagstuhl seminar will provide an ideal platform for communicating approaches and ideas between researchers of two scientific communities that pursue similar overall goals:

the design of computational systems for the understanding of complex spatiotemporal phenomena from image measurements.
Attendees will be asked specifically to report not only on past experience and present state-of-the-art but to devote a significant portion of their presentation to future needs and capabilities. A goal of the Seminar is to develop a realistic vision of the future, including identifiable initial steps. Novel contacts and cooperation may be expected that will lead to joint research work with a high scientific impact.

4 Publications

Experiments in Fluids will publish a Special Issue consisting of papers selected from those contributed to the Seminar. To maximise the possibility of being selected, authors are encouraged to prepare their papers with a quality level consistent with being submitted directly to a journal for review.

6.3 Similarity-based Clustering and its Application to Medicine and Biology

Seminar No. **07131** Date **25.03.–30.03.2007**
Organizers: Michael Biehl, Barbara Hammer, Michel Verleysen, Thomas Villmann

In medicine, biology, and medical bioinformatics, more and more data arise from clinical measurements such as EEG or fMRI studies for monitoring brain activity, mass spectrometry data for the detection of proteins, peptides and composites, or microarray profiles for the analysis of gene expressions. Typically, data are high dimensional, noisy, and very hard to inspect using classical (e.g. symbolic or linear) methods. At the same time, new technologies ranging from the possibility of a very high resolution of spectra to high throughput screening for microarray data are rapidly developing and carry the promise of an efficient, cheap, and automatic gathering of tons of high quality data with large information potential. Thus, there is a need for appropriate machine learning methods which help to automatically extract and interprete the relevant parts of this information and which, eventually, help to enable understanding of biological systems, reliable diagnosis of faults, and therapy of diseases such as cancer based on this information.

The seminar centered around developments, understanding, and application of similarity-based clustering in complex domains related to the life sciences. These methods have a great potential as an intuitive and flexible toolbox for mining, visualization, and inspection of large data sets since they combine simple and human-understandable principles with a large variety of different, problem adapted design choices. The goal of the seminar was to bring together researchers from Computer Science and Biology to explore recent algorithmic developments, discuss theoretical background and problems, and to identify important applications and challenges of the methods.

Results

A variety of open problems and challenges came up during the week. Before the seminar, the main challenge of similarity-based clustering in medicine and biology was seen as the problem to adapt similarity-based learning for complex, high-dimensional, and possibly non-euclidean data structures as they occur in these domains. During the discussions a much more widepread and subtle picture emerged, identifying the following topics as central issues for clustering:

- Feature extraction
- Cluster evaluation
- Comparison/Benchmarks
- Good sampling

Overall, the presentations and discussions revealed that similarity-based clustering constitutes a highly evolving field which seems particularly suitable for problems in medicine or biology and which still waits with quite a few open problems from researchers, a central problem being a formalization of goals and implicit regularization of clustering in the context of medicine and biology.

6.4 Towards Interoperability of Biomedical Ontologies

Seminar No. **07132** Date **27.03.–30.03.2007**
Organizers: Mark A. Musen, Michael Schroeder, Barry Smith

With the advent of high-throughput experiments and the need to make sense of massive quantities of data, the computer has become an indispensable part of the biomedical investigator's toolkit. But computers can work effectively only to the degree that data from myriad sources can be brought together within a single framework. Ontology research in biomedical research is however still far removed from the creation of such a unifying framework. The different kinds of biological phenomena studied by the different life science disciplines exist at many different levels of granularity, from molecule to species, and there are significant theoretical and practical obstacles standing in the way of unification of data across this granular divide. The present seminar is devoted to the goal of unification of the life science ontologies thus far developed.

Considerable efforts are being undertaken to overcome these problems and work in biomedical ontologies is burgeoning. Unfortunately, current ontology development efforts are being undertaken in an uncoordinated manner, the results are not interoperable, and so scientists in biology and medicine are presented with a conflicting array of terms and representation formats to choose from when annotating their experimental and clinical data.

These conditions create major barriers to accessing and using expanding data repositories, in ways which effectively hinder new scientific and medical advances. Our seminar is designed to break down these barriers both theoretically and practically. We will bring together the leading researchers in the field of theoretical ontology with core members of all the principal groups engaged in biomedical ontology development in order to establish the conditions which will enable the integration of development efforts which have been hitherto largely uncoordinated.

Our goal is to forge a common set of principles and a common methodology for those actively engaged in ontology building in the life sciences. General methodological problems to be addressed include:

- developing and enforcing best practices in ontology design
- developing tools for ontology unification/integration across granularities
- developing tools for reconciling ontologies with overlapping domains
- developing methods for the quantitative evaluation (e.g. of usability and usefulness) of ontologies (enabling peer review, creating standard training sets allowing objective comparisons of competing ontologies)

A variety of different types of papers will be presented:

- on specific ontological topics/problem areas manifested in current work in the life sciences, for example: the proper representation of instance data in the Electronic Patient Record; the proper treatment of biological functions in ontologies;
- on the relative merits of the principal tools of ontology construction in the life science field (Protégé various Description Logic frameworks; OBO-Edit, ...);
- on the analysis and cross-comparison of the specific ontologies being developed by participants in the meetings. These papers will analyze critical design decisions and include proposals as to how the ontologies in question might incorporate best practices of a type designed to ensure cross-ontology integration.

The seminar is designed to enable all participants to obtain direct feedback from specialists on the ontologies currently being developed, with the goal of enhancing interoperability, encouraging common design choices and facilitating the learning of lessons from others' mistakes.

6.5 Event Processing

Seminar No. **07191** Date **06.05.–11.05.2007**
Organizers: Mani Chandy, Opher Etzion, Rainer von Ammon

During the week of May 6-11, the event processing Dagstuhl seminar took place; In this seminar there were 43 participants from the following countries: Canada, Germany, Holland, Ireland, Israel, Korea, New-Zeeland, Portugal, Spain, Sweden, Switzerland, UK, and USA. The seminar had unusual proportion of industrial participants and included participants from: CITT, Cordys, Gartner, IBM, IDS Scheer, Microsoft, SoftwareAG, Oracle, RuleCore, and WestGlobal.

The seminar also consisted of people with several core disciplines such as: distributing computing, databases, software engineering, business process management, sensor networks, simulation and verification.

One of the participants commented that it seems that both academic and industrial people are interested in languages and implementation issues of complex event processing (either by rules or queries), in addition the academic people only were interested in the fundamental middleware issues (maybe since the industry people view it as engineering topics and not as research topic), while the industry people were interested on locating event processing in the buzzword-oriented universe (SOA, BI, BAM), a discussion that lacks research content, in the view of the academic people. Everybody though that the participants of the industry people had major contribution to this seminar. At the concluding session, the industry people compiled a list of research topics that the industry wishes to see, and the research people produced their wish list from the industry.

The seminar held deliberations (with some evening sessions in the wine cellar) on: what is event processing – use cases and classifications, is it a paradigm shift? Positioning EP relative to industry buzzwords (SOA, BAM, BI...), Semantics issues, modeling issues, and implementation issues.

6.6 Fair Division

Seminar No. **07261** Date **24.06.–29.06.2007**
Organizers: Steven Brams, Kirk Pruhs, Gerhard Woeginger

The problem of fair division—dividing goods or "bads" (e.g., costs) among entities in an impartial and equitable way—is one of the most important problems that society faces. A Google search on the phrase "fair allocation" returns over 100K links, referring to the division of sports tickets, health resources, computer networking resources, voting power, intellectual property licenses, costs of environmental improvements, etc.

There is an enormous but scattered literature on fair division in the fields of economics, political science, mathematics, operations research, and computer science, among others. In the recent years, there have been several academic books, and one popular book, on the subject.

Predictably, researchers in different disciplines study different aspects of fair division. They publish in different journals, attend different conferences, and even use different terminology. Thus, the impact of a development in one field may take years to be felt in another field.

Many problems that arise in fair division demand formal protocols, in part because of the many actors or the numerous activities they undertake that must be processed, and in part because of the need for consistency and transparency. For example, the 1982 Convention of the Law of the Sea, which was signed by 159 countries, specifies a simple cut-and-choose protocol for dividing seabed mining tracts. As more business and society interactions migrate to the web, it will become even more critical to have formal, well-studied protocols for fair division.

The general setting for most academic research is simple: There is a collection of goods or bads that need to be divided among a set of entities, but there are conditions on feasible allocations. For example, if the goods to be divided are divisible, like money or land, the situation is very different from that in which the goods are indivisible, such as most marital property in a divorce.

There are many ways to formalize "fairness", including max-min fairness, proportional fairness, envy-free fairness, etc. These variations may or may not lead to stable allocations, resulting in so-called Nash equilibria in a game.

Recognizing the problem created by different definitions of, and approaches to, fair division, we invited top researchers and promising young scientists—including a few advanced graduate students—to the seminar. We encouraged the top researchers, several of whom had authored books or done pioneering work in their fields, to outline major research approaches and discuss important open problems. Most of the young scientists reported on their research, which tended to reflect the latest trends and innovative tools that have been applied in a variety of areas. All speakers were asked to avoid highly technical or specialized vocabulary so that people outside their disciplines could better understand the questions and issues they were raising.

To conclude, we believe the seminar opened up the eyes of many participants to aspects of fair division not normally studied within their own disciplines. The lively intellectual interchange may well spawn cross-disciplinary research collaborations. In fact, we know of three participants from different disciplines who met at the seminar and are now collaborating on a joint paper.

6.7 Computational Social Systems and the Internet

Seminar No. **07271** Date **01.07.–06.07.2007**
Organizers: Peter Cramton, Rudolf Müller, Eva Tardos, Moshe Tennenholz

The Internet has an increasing influence on the functioning of traditional social systems, in particular if those systems are related to economic transactions. The Internet also enables the formation of new social systems. Social systems enabled or supported by the Internet are by definition computational as they can make use of intense computational support. Search engines that are based on page ranking, sponsored links, recommender systems, reputation systems and massive auctions, are prominent examples. They are also mechanisms, in the sense that the implementation of the system pre-defines the actions

that can be taken by participants, and the strategic behavior of the participants is what defines the actual performance of the system.

Due to these developments, the performance of a computational system is not anymore solely a question of its technical characteristics, the design of the underlying algorithms, but is heavily influenced by the behavior of its users and other computational systems to which it is connected. In recent years, Computer Science has responded to this development by incorporating more and more Game Theory and Economic Theory into its tools and models.

The interaction between the social sciences (and in particular economics and game theory) and computer science may lead to influence in both directions. In the particular case of computational social systems we see, for example, that traditional models in economic theory, such as the concept of Bayesian Equilibrium in games with incomplete information, are complemented by models that have been developed in Computer Science. For example, issues such as approximation and worst case / competitive analysis are suggested as natural alternatives to Bayesian analysis. Furthermore, the Computer Science approach questions assumptions made in many economic models in terms of decision capabilities of participating agents. Economic theory often neglects the bottlenecks due to exponential computation and communication in a mechanism on one hand, and the powerful capabilities of computer programs on the other.

Still, the adaptation of Game Theory and Economic Theory within Computer Science is at an early stage. In particular, this is true for experimental studies of the newly generated social systems. Behavioral economics is an area in economics that successfully incorporates behavioral sciences by use of laboratory experiments with human participants. As Computer Science has suggested new tools, their verification is still pending. This is further underlined by the fact that many of the social systems on the Internet are based on non-monetary incentives.

The seminar on computational social systems brought together leading researchers from theoretical computer science, artificial intelligence, economic theory, and behavioral economics to discuss computational social systems on the Internet from the viewpoint of their disciplines. The participants discussed theories which can support the emerging markets in the Internet, and suggest insight into future markets. Points of departure have been social and economic mechanisms suggested and inspired by the Internet, such as reputation systems, ranking systems, recommender systems, and online auctions and other markets. In 47 excellent presentations, models and analysis tools inspired by social systems on the Internet were presented and critically evaluated by the audience, based on the tradition of each of the disciplines. An important role in the seminar was devoted to the study of combinatorial auctions and to the study of congestion settings, as these areas have already a tradition of interdisciplinary research.

By far the most important contribution of the seminar is the research network that is established through the exchange of ideas among the scholars. This is especially beneficial for interdisciplinary seminars like this one. The mix of economics, computer science, and operations researchers fostered an exchange of methods, and problems that is likely to lead to path-breaking research in the essential area of social networks and the Internet.

6.8 Computational Issues in Social Choice

Seminar No. **07431** Date **21.10.–26.10.2007**
Organizers: Ulle Endriss, Jerome Lang, Francesca Rossi, Tuomas Sandholm

Motivation

For a few years now, computer scientists (especially in Artificial Intelligence) have been taking more and more of an interest in collective decision making. There are two main reasons for that, leading to two different lines of research. Roughly speaking, the first one is concerned with importing concepts and procedures from social choice theory for solving questions that arise in computer science and AI application domains. This is typically the case for managing societies of autonomous agents, which calls for negotiation and voting procedures. The second line of research goes the other way round: it is concerned with importing notions and methods from computer science for solving questions originally stemming from social choice.

Social choice theory is concerned with designing and evaluating methods of collective decision making. Solving a problem in social choice theory often amounts to proving the existence (or non-existence) of a procedure for collective decision making meeting certain requirements. However, to date computational issues have not received enough attention in social choice. This is one place where computer science comes into play. For instance, the classical impossibility result stating that any non-dictatorial voting procedure is manipulable can be bypassed by ensuring that manipulation is sufficiently hard to compute, which has called for complexity studies of manipulation problems. Another example is the study of elicitation, compact representation, and aggregation of preferences in combinatorial domains. This line of research has emerged from work in the AI community on logical and graphical representation languages. These languages are designed to allow for representing, in as little space as possible, a preference structure whose size would be prohibitively large if it were to be represented explicitly. Yet another example is the field of social software originating in the computational logic community, which is concerned with the application of logic-based techniques for specification and verification to the analysis of social choice procedures, to prove, for instance, the correctness of a fair division algorithm. As a final example of how the application of methods from computer science to problems of social choice can open up new research areas we mention here the idea of automated mechanism design, which aims at developing algorithms to generate social mechanisms for specific problem instances automatically. This approach can sometimes circumvent classical impossibility results from economics, which apply to the design of mechanisms for entire classes of problems, rather than to the specific problem instances at hand.

The aim of this Dagstuhl Seminar was to address all lines of work concerning computational issues in social choice, to a broad extent: this includes algorithmic and complexity-theoretic studies of social choice problems (both at the level of the agent and at that of the mechanism), but also, more generally, research that aims at importing concepts or methods from computer science to study social choice mechanisms.

Scientific Programme

The Dagstuhl Seminar on *Computational Issues in Social Choice* brought together leading researchers from several areas, including Social Choice, Artificial Intelligence, and Theoretical Computer Science. It gathered 44 participants from 13 countries, namely: Australia, 1; Austria, 1; France, 11; Germany, 5; Israel, 1; Italy, 3; Luxembourg, 2; The Netherlands, 5; New Zealand, 1; Switzerland, 1; UK, 5; USA, 7.

The technical programme of the seminar included regular talks as well as an open discussion and a "rump session". There were 30 regular talks, most of them on the following topics:

- complexity of voting;
- strategic behaviour in voting;
- computational barriers against strategic behaviour;
- voting games;
- coalitional games and coalition formation;
- fair division and barter exchange markets;
- preference representation and voting on multi-issue domains;
- mechanism design.

The topic of the open discussion was "Complexity of Voting". It was lead by Lane Hemaspaandra (University of Rochester, USA). It lead to a stimulating two-hour discussion involving many participants (including several students). The aim of the rump session was to allow for participants to expose, in five minutes, a new idea on any subject.

Conclusion

We found the seminar successful, especially in terms of the following achievements: First, the seminar made participants aware of a commonality of interests across different disciplines. Second, it suggested new directions for research that will probably be taken up by researchers in the next couple of years; in particular, some new areas of social choice emerged from interaction with computer scientists (such as the computational barriers to strategic behaviour, false-name proofness, or voting on very large domains). Last, student participants got very involved in the discussions.

The field is more promising than ever, and we expect the community to broaden up in the next couple of years. The next event aiming at gathering the community will be the 2nd International Workshop on Computational Social Choice (Liverpool, September 2008).

6.9 Assisted Living Systems – Models, Architectures and Engineering Approaches

Seminar No. **07462** Date **14.11.–17.11.2007**
Organizers: Arthur I. Karshmer, Jürgen Nehmer, Hartmut Raffler, Gerhard Tröster

All countries of the western hemisphere are more or less facing the fact that their population is continually growing older. In aging societies, an increasing proportion of people are suffering from a general loss of their motor, sensor and cognitive capabilities as well as from age-related diseases such as Parkinson's and dementia. Usually, several capability losses and diseases occur together (multimorbidity). As a consequence, the amount of elderly people who are unable to live an independent, self-determined life in their preferred environment has dramatically increased in recent years, with a tendency to grow even further.

Governmental bodies, hospitals, healthcare and social care institutions have expressed their concern about this development, which

- marks a deep cut in the quality of people's life, frequently ending in isolation and depression, and
- creates enormous costs for society caused by the need for intensive care or rehabilitation at home or in nursing homes.

Can assistive technologies based on computer-based Ambience Intelligence Technology help to substantially extend the period of self-determined life for elderly people? This was the key question addressed in the seminar on Assisted Living Systems, which attracted 40 specialists from 14 nations and 5 continents who discussed assisted living systems from three different viewpoints:

- the medical/psychologists viewpoint
- the outside viewpoint (users and industry)
- the inside viewpoint (sensor and software technology)

In the closing session, all participants agreed on keeping this group together by establishing an international competence network on assisted living systems. The next meeting of the group is planned for 2009 at Carnegie Mellon University.

Chapter 7

Distributed Computation, Networks, VLSI, Architecture

7.1 Geometry in Sensor Networks

Seminar No. **07151** Date **09.04.–13.04.2007**
Organizers: Subhash Suri, Roger Wattenhofer, Peter Widmayer

Abstract Networks of smart sensors offer exciting new possibilities for achieving sensory omnipresence: tiny, inexpensive, untethered sensor devices can measure and observe various environmental parameters, thereby allowing real-time and fine-grained monitoring of physical spaces around us. In order to realize this vision, however, several challenging research problem must be solved, many of which involve geometry due to the embedded nature of sensor devices. The aim of this seminar, held from April 10 to April 13, 2007, was to bring together experts from several areas of computer science and mathematics for discussions and exchange of ideas on the role of geometry in the evolution of sensor networks.

Enabled by recent advances in micro-electronics and fabrication, a new generation of integrated embedded devices, called *smart sensors*, has emerged that seems capable of realizing the long-cherished vision of *sensory omnipresence* or *ubiquitous awareness*. Through collaboration and *ad hoc* wireless networking, a collection of such devices can provide real-time, fine-grained sensing, monitoring, and actuation across large geographical areas. A key fact distinguishing sensor networks from other networked systems is that sensor nodes are *embedded* in the physical environment in which they function. For instance, unlike more traditional networks such as the Internet or the phone network, communication in sensor networks is dictated less by the desires of the end nodes and more by the geography of the sensor field and the associated signal landscapes, as well as the overall network task. As a result, geometry plays a fundamental and crucial role in all aspects of the sensor network, including their design and operation. In particular, the network must discover its own geometry through self-localization of nodes, construct a lightweight and self-organizing naming and routing structure using virtual or physical coordinates, exploit physical embedding to perform information aggregation and dissemination etc. Motivated

by these observations, the discussion during the workshop focused largely on techniques of a geometric or topological character that are particularly relevant to sensor networks.

We give a brief roundup of the excellent presentations: Pilu Crescenzi presented Bluetooth connectivity problems stemming from unavailable neighbor information. Leszek Gasieniec discussed how to escape the quadratic lower bound in geo-routing by including pre-processed information. Erik Jan van Leeuwen exhibited better approximation algorithms for disk graphs with bounded ply. Bastian Katz presented a new sensor network localization heuristic based on recursively applying rigidity theory, and discussed noisy measurements. Andrzej Pelc surveyed results on broadcasting in radio networks. Zvi Lotker discussed the MST of random geometric graphs, and its connection to the upper box dimension. Leonidas Guibas discussed how to aggregate data from sparse events by double rulings. Anish Arora continued presenting results about data aggregation with data of nearby nodes being fresher. Jie Gao presented new sensor network localization heuristics by means of rigidity theory, similarly as Bastian Katz. Jung-Geon Park discussed localization in a mobile environment. Alex Kroeller and Sandor Fekete first showed their video on sensor networks, and then presented new results on energy constrained flow problems. Michael Elkin gave new insights into distributed sparse spanner constructions in a dynamic model. Lata Narayanan presented an impossibility result for geo-routing in 3D, and a complementing possibility result for "$2\frac{1}{2}$D". Andrea Richa and Christian Scheideler presented SIT, a new model beyond the unit disk graph, and a new algorithm for dominating sets in this model, using only constant storage. Li Erran Li investigated how much it helps to communicate with two power levels. Paolo Santi discussed new theoretical and practical insights in topology control. Shakhar Smorodinsky surveyed various results in conflict-free coloring. Alon Efrat gave a talk on sensor coverage, with a survey on the related art gallery problem.

One common and recurring theme in many talks and discussions was the lack of an appropriate model for sensor networks. For instance, many theoretically elegant results for routing in ad hoc wireless networks have been derived using the idealized unit-disk model, which fails to capture the intricate reality of radio transmission.

An open problem session was held on April 10, 2007, the first day of the workshop. Several participants (Roger Wattenhofer, Evangelos Kranakis, Li Erran Li, Paolo Penna, Zvi Lotker, Leonidas Guibas) posed specific technical problems that have defied progress. Many of these problems elicited significant discussion among the participants, and in some cases participants even pointed to known or related results in their fields.

A general discussion forum was held on the last evening of the seminar, April 12, 2007, to speculate about the promising future directions of research in this young and emerging field. Many people felt that sensor networks are significantly different from general-purpose networks (such as the Internet) and a close coupling of applications and the networking will be important, unlike the Internet that advocates a clean separation of different layers.

In conclusion, the seminar offered a great opportunity for researchers with different, but overlapping, interests to share their expertise, and engage in intellectually stimulating discussions about the important future directions. The format of the workshop provided an ideal environment to question various assumptions in each others' work, find ideas and inspiration in their results, and get a better, more holistic, sense of how different areas

of expertise can contribute to this emerging technology. Geometric approaches, through concepts and techniques, offer a number of opportunities in sensor networks to address problems at structural, functional and application levels. It is our belief that the exchange of ideas among the participants of this workshop will impact how they approach their future research in sensor networks and influence the field.

7.2 Resilient and Survivable Networks, Infrastructure and Services

Seminar No. **07301** Date **22.07.–25.07.2007**
Organizers: Hermann de Meer, David Hutchison, Bernhard Plattner, James P. G. Sterbenz

As we become increasingly reliant on networked applications in the consumer, commercial, government, and military sectors, it is essential that they are resilient and survivable to a number of challenges, including:

- unusual but legitimate traffic load (e.g. flash crowds)
- high-mobility of nodes and sub-networks
- weak and episodic connectivity of wireless channels
- unpredictably long delay paths either due to length (e.g. satellite) or as a result of episodic connectivity
- attacks against the network hardware, software, or protocol infrastructure
- large-scale natural disasters
- failures due to mal-configuration or operational errors
- natural faults of network components

The challenge is to provide acceptable service to applications, in particular the ability for users and applications to access information and for the maintenance of critical end-to-end communication associations. Furthermore, resilient network services must remain accessible whenever possible, ensure correctness of operation when performance is degraded, and automatically recover. Finally, resilient and survivable networks need to be engineered to have emergent behaviour so that they can resist challenges to their operation, recognize when challenges occur and autonomically limit their effects, recover rapidly to normal operation, and refine future behaviour. We propose to exploit techniques in programmable, active, and cognitive networking to achieve these goals.

Only a few aspects of these challenges have been explored by the research community. Fault tolerance is a mature discipline from which we can learn, but considers only a very small subset of the problem. Recent work on DDOS attacks and early work in disruption-tolerant

networking are also a piece of this puzzle. We aim to consider all aspects of the provision so application, service, and network resilience and survivability. We view resilience and survivability as an emerging hot topic that is in need of a systematic framework to guide future research.

The goal of this seminar is to bring together researchers and engineers who have explored parts of this space, and who can contribute to the overall goals of helping to create a research agenda in resilience and survivability. Common research challenges should be identified and potential solutions discussed. Typical conferences and workshops do not allow such an interdisciplinary approach because of their tendency to focus on particular technologies or aspects of the overall framework. As an outcome of this seminar, gaps among different research communities will be bridged, and as result their research agendas may be better aligned.

Areas of interest include, but are not limited to the following:

- Mobile, wireless, and sensor networking in challenged environments
- Resilient backbone and access network architecture
- Cross-layer optimizations (knobs and dials in vertical control loops)
- Resource tradeoffs (processing, memory, bandwidth, energy, latency)
- Resilience as a QoS property and metrics for resilience
- Resilience and survivability as an extension of conventional security
- Lessons learned from past failures and disasters (e.g. Hurricane Katrina, 2005 SBB rail network outage, 2003 Northeast US power failure, 9/11, 1988 Chicago PSTN central office fire)
- Managing complexity and feature interactions
- Contributions from fault tolerant distributed systems community
- Interaction between network resilience and other infrastructures (e.g. the power grid)
- Resilient service creation, composition, and deployment
- Resilient and survivable auto-configuration, self-organisation, and self-management

7.3 Autonomic Management of Networks and Services

Seminar No. **07302** Date **25.07.–28.07.2007**
Organizers: Raouf Boutaba, Marcus Brunner, Jürgen Schönwälder, Rolf Stadler

During the past years, numerous novel architectures were developed for communication networks, the networked services or applications running of the core infrastructure, and its

related protocol specifications. There are several activities undertaken for designing and developing novel network architectures that enable flexible, dynamic, and fully autonomic formation of network nodes as well as whole networks. It will allow dynamic adaptation and re-organization of the network according to the working, economical and social needs of the users. This is expected to be especially challenging in a mobile context where new resources become available dynamically, administrative domains change frequently, and the economic models may vary.

Assuming these new types of network architectures will emerge, the operation and management of those types of networks must undergo dramatic change compared to today's way of operating networks. Additionally, those networks will change the way services are deployed and delivered to users in a much more dynamic, personalized, location-aware way.

The goal of the seminar is to bring together researchers and engineers experienced in different emerging networking technologies, service technologies, and management technologies to discuss the future direction of those networks and specifically the operation and management issues concerned with these networks and services. The seminar will allow participants to advance issues associated with autonomic behavior in networking and services.

Another interesting question in this regard is the differences, synergies, and possible interaction between autonomic management and autonomic networking, or: how much management functionality should be built into the network, and how much should be implemented outside.

The seminar will consist of individual presentations and a set of group work, consisting of break-out sessions and plenary sessions in which results from individual groups are presented and discussed. Prior to the seminar, the participants are requested to provide an extended abstract on their talk. These abstracts will be used to structure the sessions with individual talks. The following aspects are of particular interest: pointing out current problems and promising directions, describing synergies, competitions and weaknesses of the different research areas mentioned in the description of the seminar topic, and identifying development perspectives, potential and recommendations for further research. As one outcome of the seminar, it is planned to edit a whitepaper on the definition(s) and research area and future research directions. This report would combine the most important issues with relevance for the future that will be identified in the seminar, and potentially will be published in a magazine or journal.

This seminar will be co-sponsored by EMANICS, the European Network of Excellence for the Management of Internet Technologies and Complex Services.

7.4 Code Instrumentation and Modeling for Parallel Performance Analysis

Seminar No. **07341** Date **19.08.–24.08.2007**
Organizers: Adolfy Hoisie, Barton P. Miller, Bernd Mohr

Given the exponential increase in the complexity of nowadays parallel systems, parallel applications often fail to exploit the full power of the underlying hardware. At scale, it is not uncommon for applications to run at parallel efficiencies in the low single digits. Moreover, their optimization is extremely difficult due to the inherent complexity of the systems and of the applications themselves. Therefore, a variety of projects aim to develop tools for the measurement, analysis, and visualization of parallel program performance in order to help and guide users in the optimization process.

This meeting was the third in a series of seminars related to the topic "Performance Analysis of Parallel and Distributed Programs", with previous meetings being the Dagstuhl Seminar 02341 on "Performance Analysis and Distributed Computing" held in August 2002 and Seminar 05501 on "Automatic Performance Analysis" in December 2005. While these seminars concentrated on the "analysis" part of performance analysis, at the most recent seminar the focus was on the building blocks of program instrumentation and modeling that are prerequisites for the analysis phase. As a result, the presentations of the participants concentrated on several fundamental issues related to instrumentation for generating high quality performance data, methodologies for performance modeling leading to accurate predictions for the performance, and on the ways in which these techniques are combined for the performance analysis of applications and systems.

The program consisted of 28 presentations and practical tool demonstrations as well as two "open mic" sessions where time was set aside for spontaneous discussions and "brain storming". The seminar brought together a total of 48 researchers and developers working in the area of performance from universities, national research laboratories and, especially important, from three major computer vendors. The goals were to increase the exchange of ideas, knowledge transfer, foster a multidisciplinary approach to attacking this very important research problem with direct impact on the way in which we design and utilize parallel systems to achieve high application performance.

The presentations can be grouped thematically as follows:

- Session "Performance Analysis in General"
- Session "Instrumentation"
- Session "Modeling"
- Session "Scalability"
- Session "Tools"
- "Open Mic" Sessions
- Session "Short Announcements"

Despite the larger than normal number of participants, the seminar was very successful due to the dedicated professionalism and discipline of the participants on one side and the very helpful and professional staff of Dagstuhl on the other side. Lively discussions

and spontaneous computer demonstrations continued every day well beyond midnight. It is important to note that the group meeting and residential aspects of Dagstuhl and the five-day format provide a continuity of thought and discussion unavailable in other conference, workshop, or meeting settings. At Dagstuhl, we have time for considered (and reconsidered!) dialogs whose impact last well beyond the meeting week.

7.5 Programming Models for Ubiquitous Parallelism

Seminar No. **07361** Date **02.09.–07.09.2007**
Organizers: Albert Cohen, Maria J. Garzaran, Christian Lengauer, Samuel P. Midkiff, Chi-Leung Wong

The Internet has an increasing influence on the functioning of traditional social systems, in particular if those systems are related to economic transactions. The Internet also enables the formation of new social systems. Social systems enabled or supported by the Internet are by definition computational as they can make use of intense computational support. Search engines that are based on page ranking, sponsored links, recommender systems, reputation systems and massive auctions, are prominent examples. They are also mechanisms, in the sense that the implementation of the system pre-defines the actions that can be taken by participants, and the strategic behavior of the participants is what defines the actual performance of the system.

Due to these developments, the performance of a computational system is not anymore solely a question of its technical characteristics, the design of the underlying algorithms, but is heavily influenced by the behavior of its users and other computational systems to which it is connected. In recent years, Computer Science has responded to this development by incorporating more and more Game Theory and Economic Theory into its tools and models.

The interaction between the social sciences (and in particular economics and game theory) and computer science may lead to influence in both directions. In the particular case of computational social systems we see, for example, that traditional models in economic theory, such as the concept of Bayesian Equilibrium in games with incomplete information, are complemented by models that have been developed in Computer Science. For example, issues such as approximation and worst case / competitive analysis are suggested as natural alternatives to Bayesian analysis. Furthermore, the Computer Science approach questions assumptions made in many economic models in terms of decision capabilities of participating agents. Economic theory often neglects the bottlenecks due to exponential computation and communication in a mechanism on one hand, and the powerful capabilities of computer programs on the other.

Still, the adaptation of Game Theory and Economic Theory within Computer Science is at an early stage. In particular, this is true for experimental studies of the newly generated social systems. Behavioral economics is an area in economics that successfully incorporates behavioral sciences by use of laboratory experiments with human participants. As Computer Science has suggested new tools, their verification is still pending. This is

further underlined by the fact that many of the social systems on the Internet are based on non-monetary incentives.

The seminar on computational social systems brought together leading researchers from theoretical computer science, artificial intelligence, economic theory, and behavioral economics to discuss computational social systems on the Internet from the viewpoint of their disciplines. The participants discussed theories which can support the emerging markets in the Internet, and suggest insight into future markets. Points of departure have been social and economic mechanisms suggested and inspired by the Internet, such as reputation systems, ranking systems, recommender systems, and online auctions and other markets. In 47 excellent presentations, models and analysis tools inspired by social systems on the Internet were presented and critically evaluated by the audience, based on the tradition of each of the disciplines. An important role in the seminar was devoted to the study of combinatorial auctions and to the study of congestion settings, as these areas have already a tradition of interdisciplinary research.

By far the most important contribution of the seminar is the research network that is established through the exchange of ideas among the scholars. This is especially beneficial for interdisciplinary seminars like this one. The mix of economics, computer science, and operations researchers fostered an exchange of methods, and problems that is likely to lead to path-breaking research in the essential area of social networks and the Internet.

Chapter 8

Embedded Systems

8.1 Quantitative Aspects of Embedded Systems

Seminar No. **07101** Date **04.03.–09.03.2007**
Organizers: Boudewijn Haverkort, Joost-Pieter Katoen, Lothar Thiele

1 Summary

1.1 Current status

Embedded software controls the core functionality of many systems. Embedded software is omnipresent: it controls telephone switches and satellites, drives the climate control in our offices and cars, runs pacemakers, is at the heart of our power plants, and makes our cars and TVs work properly. As such systems are massively encroaching on daily life, our reliance on embedded software is growing rapidly. But, how justifiable is this reliance?

Whereas traditional software has a rather transformational nature mapping input data onto output data, embedded software is different in many respects. Most importantly, embedded software is subject to complex and permanent interactions with their—mostly physical—environment via sensors and actuators. Typically software in embedded systems does not terminate and interaction usually takes place with multiple concurrent processes at the same time. Reactions to the stimuli provided by the environment should be prompt (timeliness or responsiveness), i.e., the software has to "keep up" with the speed of the processes with which it interacts.

Furthermore, characteristic for embedded systems is that they have to meet a multitude of quantitative constraints. These constraints involve the resources that a system may use (computation resources, power consumption, memory usage, communication bandwidth, costs, etc.), assumptions about the environment in which it operates (task arrival rates, task sizes), and requirements on the services that the system has to provide (timing constraints, performance, response time) and requirements of the continuity with which these services are delivered (availability, dependability, fault tolerancy, etc.).

The observed difference between traditional software and embedded software has recently led to initiatives for dedicated international conferences such as EMSOFT (since 2001), and

ACM SIGPLAN Workshop on Languages, Compilers, and Tools for Embedded Systems (since 1998), as well as journals such as ACM Transactions on Embedded Computing Systems, (since 2002) and Design Automation for Embedded Systems Journal (since 2003). In various countries research institutes on embedded systems have been set up with a strong industrial cooperation, for instance, in Denmark (CISS), the Netherlands (ESI), and the US (CHESS).

1.2 A lack of quantitative assessment

Despite the importance of the quantitative constraints for the well-operation of embedded systems, the proper assessment of cost, resources, performance, dependability, robustness, etc., often comes as an afterthought. It is rather common for embedded software to be fully designed and functionally tested before any attempt is undertaken to determine its performance, dependability or resource-usage characteristics. One of the main reasons for this situation is that well-developed and rigorous evaluation techniques for non-functional, i.e., quantitative system aspects have not become an integral part of standard software engineering practice. This undesirable situation has led to the increased interest by embedded software researchers to extend the usual functional specification and properties with a set of "performance indices", e.g., stated in terms of costs, timeliness, speed and the like, and constraints on these indices. Also in industry, a growing interest in assessing non-functional aspects of embedded systems as early as possible in the system design life cycle can be witnessed.

2 Where are we going?

Model-Driven Development (MDD) is a new software development technique in which the primary software artifact is a model. Ideally, the MDD technique allows engineers to (graphically) model the requirements, behaviour and functionality of computer-based systems. The design is iteratively analysed, validated, and tested throughout the development process, and automatically generated production-quality code can be output in a variety of languages.

Existing MDD tools for embedded systems are rather sophisticated in handling functional requirements but their treatment of quantitative constraints is still in development. Although methods for verification of real-time system designs, using for instance timed automata, are being developed, these methods are not yet mature enough for dealing with larger industrial embedded systems. Hence, MDD will not realise its full potential in the embedded systems area unless the ability to handle quantitative properties is drastically improved.

In contrast to the situation in the design of embedded software systems, in the design of computer-communication systems, quantitative methods to determine the quality of the system, expressed in terms of throughput or response time have been used for a long time. Next to methods from classical queueing theory and discrete-event simulation, recently the use of analytical/numerical methods for evaluating complex systems has become more widespread. This has lead to methods and techniques to specify complex system behaviour

using some formal method (Petri nets, process algebra) enhanced with time and probabilities. Subsequently, appropriate (Markovian) models are generated from these high-level models, which, after numerical analysis, provide detailed insight in performance and dependability measures of interest. More recently, also constraints related to power usage have been taken up.

Over the last, say, 5 years, very good progress has been made in pairing the above quantitative techniques to techniques known from the verification area, esp. model checking of properties specified in logics like CSL (an extension of CTL with stochastic time). This has lead to model checking algorithms and tools for Markovian models of system. The state-of-the-art, however, is still such that expert knowledge is required to use these techniques, hence, large-scale application in embedded software system design and implementation is still a dream rather than a reality.

Of course, also known quantitative techniques from the area of real-time systems (classical ones, such as EDF, or more advanced compositional ones), or methods known from network calculus (originally developed for dimensioning communication networks at a high level of abstraction), data flow graphs, and so on, can, and probably should be used as part of the embedded system design.

What is clear, though, is that all of the above techniques can only be used during the design of embedded systems after appropriate adaptation and embedding in a design trajectory, e.g., based on MMD.

Furthermore, where each of the above mentioned approaches has its strengths and weaknesses, an important first task is to map these strength and weaknesses (applicability, scope, modelling power, costs of evaluation, etc.). A second and more challenging question is then how to combine or integrate these methods. Such questions can only be answered when key researchers for these various approaches come together and exchange and discuss their ideas.

3 Seminar goal

Given the above considerations, the goal of this Dagstuhl seminar has been to bring together experts in the areas of embedded software design and implementation, model-based analysis of quantitative system aspects, and researchers working on extending all kinds of formal (design and analysis) methods with quantitative system aspects. These three areas are clearly well- related in the context of embedded systems, but have not been addressed as such in the past, as they have been worked upon in different communities. Thus, the seminar will lay bridges between these three areas, so that knowledge and experience can be shared, transferred and, ultimately, be generated.

8.2 Model-Based Engineering of Embedded Real-Time Systems

Seminar No. **07451** Date **04.11.–09.11.2007**
Organizers: Holger Giese, Gabor Karsai, Edward Lee, Bernhard Rumpe, Bernhard Schätz

Today, embedded software plays a central role in most advanced technical systems such as airplanes, cell phones, and cars, and has become the main driver for innovation. Development, evolution, configuration and maintenance of embedded and distributed software nowadays often are serious challenges as a drastic increase of the software complexity can be observed in practice. The application of model-based engineering technologies to embedded real-time systems seems to be a good candidate to tackle some of the resulting problems.

Model-based development strategies and automatic code generation are becoming established technologies on the functional level. However, they are mainly applied in monolithic systems. The use of similar modeling strategies on the system, technical, and configuration levels remains challenging, especially with the increasing shift to networks of systems, deepened interaction between control-engineering and reaction-oriented parts of a system, and the growing number of variants introduced by product lines. Specific domain constraints such as real-time requirements, resource limitations and specific hardware dependencies often impede the acceptance of standard high-level oriented modeling techniques and their model-based application. Much effort in industry and academia therefore goes into the adaptation and improvement of object-oriented and component-based methods and model-based engineering that promise to facilitate the development, deployment, and reuse of software components embedded in real-time environments. The model-based development approach for embedded systems and their software proposes application specific modeling techniques using domain specific concepts (e.g., time-triggered execution or synchronous data flow) to abstract from the details of the implementation such as interrupts or method calls. Furthermore, analytical techniques (like, e.g., verification of the completeness of function deployment and consistency of dynamic interface descriptions) and generative techniques (e.g., automatic schedule generation, default behavior generation) can then be applied to the resulting more abstract models to enable the efficient development of high quality software.

Our Dagstuhl seminar brought together researchers and practitioners from the field of model-based engineering of embedded real-time systems. The topics covered included: frameworks and methods, validation, model-based integration technology, formal modeling of semantics, fault management, concurrency models and models of computation, requirements modeling, formal derivation of designs from requirements, test modeling and model-based test generation, quality assurance, design management, abstractions and extensions, and development techniques and problems of application domains. The broad spectrum of presentations has clearly illustrated the prevalence of model-based techniques in the embedded systems area, as well as progress in the field.

The seminar included mostly conference-like presentations followed by short discussions,

and three group discussion sessions with panels selected from the attendees. In all cases, the emphasis was on fostering interaction among the participants and on gaining new insights and better understanding. Most of the seminar participants provided abstracts and the slides for their presentations that are available on the Dagstuhl website for the seminar. We plan to compile a state-of-the-art survey on model-based development of embedded real-time systems addressing foundational issues, language engineering, domain-specific issues, and life-cycle issues. The survey will be based on contributions of the participants of the seminar and will be published in the Springer LNCS series.

Chapter 9

Modelling, Simulation, Scheduling

9.1 Numerical Methods for Structured Markov Chains

Seminar No. **07461** Date **11.11.–14.11.2007**
Organizers: Dario Bini, Beatrice Meini, Vaidyanathan Ramaswami, Marie-Ange Remiche, Peter Taylor

Markov chain models are of paramount importance in many applications, including performance evaluation of telecommunications and computer systems, information retrieval, page ranking and queueing models. Whilst retaining algorithmic tractability, Markov chains offer flexibility in choosing the parameters one may incorporate into a model.

Systems such as the wireless standard IEEE 802.11, Peer-to-Peer (P2P) communication, wireless video transmission and congestion control algorithms in public telecommunications have been successfully modeled by means of Markov chains exhibiting particular structures. For example, Markov fluid models can be used to mimic IP traffic or to analyse the performance of a token bucket model, and Markov chains of M/G/1, G/M/1 and QBD-type have been used to solve a wide variety of queueing problems.

It is of note that the transition matrices resulting from such models often exhibit particular structures that allow for development of particularly efficient algorithms for their analysis.

Besides their importance in applications, structured Markov chains are interesting for the richness of the mathematical tools needed for their treatment. The analysis and development of efficient numerical methods for these Markov Chains constitutes one of the major incumbent challenges in this field. Conversely, the existence of such powerful methods actually incites engineers to model complex systems via Markov chains. Matrix analytic methods and structured matrix technology are important tools for the design of effective algorithms.

The analysis and development of efficient numerical methods for Markov Chains constitutes one of the major incumbent challenges in this field. Conversely, the existence of such powerful methods actually incites engineers to model complex systems via Markov chains. Matrix analytic methods and structured matrix technology are important tools for the design of effective algorithms.

The seminar was attended by 26 scholars, mostly from the academic world, including 8 PhD students and postdoctoral fellows. The participants came from North America, Europe, and Australia.

One session was devoted to celebrate the 60th birthday of Guy Latouche, one of the major experts in this multidisciplinary field of applied probability, numerical analysis and modeling. In this session the most recent advances on numerical methods for structured Markov chains were presented.

Specific subjects of interest can be grouped in the following areas:

- theory of phase-type and matrix-exponential distributions,
- matrix analytic methods
- design and analysis of algorithms
- model analysis and inference procedures in the telecommunications

The seminar was closed by an open discussion on the state of the art of research and on the future research directions.

This Dagstuhl seminar has brought together leaders and young researchers in the fields of analysis of numerical algorithms, applied stochastic modeling and statistical inference, with the result of stimulating exchange of methodologies and experiences and generating synergetic collaborations.

This has favored a better communication between these worlds where problems from the applications feed the theoretical research and where advanced numerical tools can be utilized in applications with reciprocal advantages.

Chapter 10

Cryptography, Security

10.1 Symmetric Cryptography

Seminar No. **07021** Date **07.01.–12.01.2007**
Organizers: Eli Biham, Helena Handschuh, Stefan Lucks, Vincent Rijmen

Cryptography provides techniques for secure communication in adversarial environments. Cryptographic primitives are symmetric, if both the sender and the receiver of a message are using the same secret key, as in the case of block and stream ciphers and message authentication codes. Another type of symmetric primitives are cryptographic hash functions, where neither sender nor receiver need to know a secret key at all. In contrast to this, cryptographic primitives are asymmetric, if sender and receiver are using different keys, typically a "public" and a "private" one.

Symmetric Cryptography deals with designing and analysing

- symmetric primitives (block and stream ciphers, message authentication codes and hash functions), and
- cryptographic protocols employing these primitives.

Since symmetric cryptosystems are much more efficient in practice than asymmetric systems, most security applications use symmetric cryptography to ensure the privacy, the authenticity and the integrity of sensitive data. Even most applications of public-key cryptography are actually working in a hybrid way by transmitting a cipher key with asymmetric techniques while symmetrically encrypting the payload data under the cipher key.

Participation and Program

The Seminar brought together about 35 researchers from industry and academia. Most of the participants came from different European countries, but quite a few also came

from America and Asia. Almost all the participants gave a presentation. Most of them gave a "regular" talk of 30 to 50 minutes (including discussion time), some gave a "rump session" talk, and a few even gave two presentations, a regular one and another at the rump session.

The institution of a "rump session" for short talks on recent results, fresh ideas and open problems has a long tradition at cryptographic workshops and conferences. At the Seminar, the "rump session" was on Thursday evening. Each "rump session" talk was limited to at most ten minutes.

Topics and Focus Areas

The Seminar topics (stream ciphers, message authentication, hash functions, provable security, algebraic attacks, lightweight cryptography,...) were various, but closely related and interleaved. All these topics received their share of interest, but two areas caught more attention than others:

1. The design and analysis of hash functions.
2. The security of stream ciphers against nonstandard "repeated initial value" attacks.

The participant's interest in the first area is rather unsurprising. In 2004 and 2005, the cryptanalysis of hash functions has made a big leap forward. Attacks against hash functions in wide practical use, such as MD5 and SHA-1, have been published. There is an urgent need for new practical hash functions. Quite a few talks and many discussions dealt with advancing the theory and practice of hash function design, including the study of hash function attacks.

The excitement for the second area mirrors very recent research advances in research in Symmetric Cryptography. At the Seminar, further progress was made.

Advances and Outlook

Most presentations at the seminar dealt with very recent results on Symmetric Cryptography – unpublished research which either had been submitted to one of the leading conferences in the area, or is designated to be submitted soon. Some participants also presented their research in progress, promising but not mature enough for publication. We anticipate that most of the presentations at the Seminar will ultimatively lead to peer-reviewed publications.

The atmosphere at the Seminar was very inspiring and stimulating. Participants reacted on other participants' open problems, and collaborations were initiated. Some progress made by our participants during the course of the Seminar and *already presented at the Seminar*:

- As a reaction on Greg Rose's presentation of a new stream cipher called "Shannon", *Alexander Maximov* presented some "repeated IV" attacks at the rump session.

- Following some discussions (during the days of the Seminar) with Alexander Maximov and others, *Greg Rose* confirmed the attack at the rump session and explained which design choices lead to the weakness.
- Inspired by Bart Preneel's talk on a "repeated IV" attack against the stream cipher "Phelix", *Doug Whiting* (one of the authors of Phelix), presented a tweak for Phelix at the rump session. The tweak defends against the weakness exploited by Preneel.
- After Elena Andreeva's talk on the RMC hash function design and its generalised security properties, it was observed that the HAIFA hash iteration mode can be instantiated with compression functions that satisfy the extra conditions required for RMC. If one does so, the RMC proof of security by Andreeva and her co-authors is applicable to the HAIFA mode as well, i.e., HAIFA satisfies the generalised RMC security properties. *Orr Dunkelman* (one of the authors of HAIFA) presented this observation at the rump session.
- In a quickly-scheduled regular talk on Friday morning, *Ralph-Philipp Weinmann* and *Ulrich Kühn* presented the idea of using algebraic attack techniques for a rather unusal kind of block cipher analysis: The adversary is allowed to control plaintexts *and keys*. The adversary's goal is to find out unknown parts of the block cipher specification (namely, a description of the secret S-box). This collaboration was initiated by a discussion at the Seminar.

Again, we anticipate that some – and perhaps all – these presentations will eventually lead to peer-reviewed publications.

10.2 Mobility, Ubiquity and Security

Seminar No. **07091** Date **25.02.–02.03.2007**
Organizers: Gilles Barthe, Heiko Mantel, Peter Müller, Andrew C. Myers, Andrei Sabelfeld

Increasing code mobility and ubiquity raises serious concerns about the security of modern computing infrastructures. The focus of this seminar was on securing computing systems by design and construction.

The seminar covered a wide span of application areas, including:

- telecommunications
- automotive industry
- web browsers
- electronic voting
- web services
- distributed systems
- media distribution
- data mining

The need for security in these applications is critical. The seminar structure reflected the general categories of security properties that are required in scenarios as above. Each category served as a theme for presentations on each of the first four days of the seminar. Each of these days was kicked off by a tutorial talk. These categories were:

Confidentiality David Sands' tutorial showed that (partial) equivalence relations were ubiquitous in security modeling. The tutorial was followed by these talks:
Anindya Banerjee: Information flow, modularity and declassification
Alejandro Russo: Closing internal timing channels by transformation
Gregor Snelting: Information flow control for Java based on path conditions in dependency graphs
Peeter Laud: Dependency-graph-based protocol analysis
Henning Sudbrock: A probabilistic justification of the combining calculus
Mads Dam: A complete logic of knowledge and one-way computable terms
Alexander Reinhard: Controlling the what and where in language-based security
David Pichardie: A certified lightweight non-interference java bytecode verifier
Richard Bubel: Integration of a security type system into a program logic

Integrity Joshua Guttman's tutorial illuminated the interaction between two aspects of integrity: invariants vs. causality. The tutorial was followed by these talks:
Steve Zdancewic: Combining access control and information flow in DCC
Cédric Fournet: Secure implementations for typed session abstractions
Fausto Spoto: Optimality and condensing of information flow through linear refinement
Brendan Eich: JavaScript: Mobility and ubiquity—two out of three ain't bad
Peter Ryan: Trustworthy elections
Flemming Nielson: Static analysis for DRM
Amy Felty: Program verification, noninterference, and declassification applied to privacy in data mining
Dieter Hutter: Preserving privacy in service composition using information flow control
Brigitte Pientka: Contextual modal logic

Availability Thomas Jensen's tutorial emphasized that even simple availability were hard to enforce. The tutorial was followed by these talks:
Pierpaolo Degano: A static approach to secure service composition
Andrew Myers: Ensuring confidentiality, integrity, and availability by construction

Foundations of cryptography Cédric Fournet's tutorial demonstrated that computational and semantic views of security can be reconciled, although more progress is needed. The tutorial was followed by these talks:
Tamara Rezk: Computational noninterference
Hanne Riis Nielson: Flow sensitive analysis of security properties
Aslan Askarov: Gradual release: unifying declassification, encryption, and key release policies
Santiago Zanella Béguelin: Towards code-based cryptographic proofs
Daniel Hedin: A framework for parameterizing type systems with relational information
Ian Stark: Resource type checking in database queries
Joshua Guttman: Programming cryptographic protocols

The seminar was concluded with the following talks:
Peter Müller: Generic universe types

Arnd Poetzsch-Heffter: A behavioral semantics of object-oriented components
Gilles Barthe: Certificate translation
Marieke Huisman: BML
Fabio Martinelli: Modeling and enforcing security & trust management policies (on JVM)

Thanks to Dagstuhl's stimulating environment, many insightful discussions, planned and unplanned, took place. There were two large organized discussions, where all participants were involved: a panel on electronic voting (e-voting) and a general discussion.

Panel on e-voting A panel on e-voting was moderated by Peter Ryan and featured Jorge Cuellar, Joe Kiniry, and Carsten Schürman. This panel generated a lively discussion on the role of formal methods in e-voting. E-voting includes both supervised and remote scenarios where the results are processed electronically. While several concerns were raised about trust involved in various e-voting schemes much evidence was brought up for benefits of e-voting and the need for formal methods for its support.

General discussion The general discussion reiterated the need for building the security in. It arrived at the following important directions for future research:

- There is potential in combining advanced static analyses, program logics, type systems, and program transformation for security.
- Integrated approaches to enforcing multiple security properties are much desired.
- The web page is the new operating system. Language-based techniques may help securing it.
- Formal methods are needed for e-voting protocol design and implementation.

With a top-of-the-line collection of invitees placed in Dagstuhl's productive environment, it may seem that little could have gone wrong with the seminar. Still, we are fully satisfied that our efforts on organizing the meeting have been rewarded by a seminar with a clear focus; good balance between talks, panels, and discussions; and rich cross-fertilization that have already resulted in new collaborations.

10.3 Frontiers of Electronic Voting

Seminar No. **07311** Date **29.07.–03.08.2007**
Organizers: David Chaum, Miroslaw Kutylowski, Ronald L. Rivest, Peter Y. A. Ryan

Democracy and voting systems have received considerable attention of late, with the validity of many elections around the world being called into question. The US experience demonstrates that simply deploying technological "solutions" does not solve the problem

and can easily exacerbate it. Nevertheless, many other countries are either deploying e-voting and e-counting systems or planning to do it.

The aim of the seminar was to present and discuss promising technologies, schemes, and cryptographic protocols to achieve high assurance of accuracy and privacy in the casting and counting of votes. Special attention was given to attacks and dangers that emerge for electronic voting systems.

The challenge is highly socio-technical in nature: requires an excellent understanding of the potentialities and dangers of technological approaches as well as an appreciation of the social, legal and political impact. The seminar thus aimed to bring together researchers and practitioners from academia and industry, whose work relates to electronic voting systems, to evaluate the state of the art, to share practical experiences, and to look for possible enhancements. The overall aim then was to stimulate discourse between the various stakeholders and enhance the understanding of voting technologies and practices.

Dagstuhl Accord on Electronic Voting

Participants of the 2007 Dagstuhl Conference on Frontiers of E-Voting agree that:

Taking advantage of technology to improve large-scale elections has recently captured the interest of researchers coming from a number of disciplines. The basic requirements pose an apparently irreconcilable challenge: while voter confidence hinges on transparently ensuring integrity of the outcome, ballot secrecy must also be ensured. Current systems can only address these essential requirements by relying on trust in those conducting the election or by trust in the machines and software they use. Some promising new systems dramatically reduce the need for such trust. What are called "end-to-end" voting systems, for example, allow each voter to ensure that his or her vote cast in the booth is recorded correctly. They then allow anyone to verify that all such recorded votes are included in the final tally correctly. Surprisingly, through use of encryption typically, these systems can also provide privacy of votes. They do this without introducing any danger of "improper influence" of voters, as in vote buying and coercion. Moreover, such systems offer all these properties without relying on trust in particular persons, manual processes, devices, or software.

Care must still be taken to ensure proper implementation and education of voters in order to avoid misuse or incorrect perceptions. Some are also concerned that the level of understandability and observability of hand-counting of paper ballots in polling places will not be matched by electronic systems. The challenge for governments and civil society should be to find ways to foster development and testing of new election paradigms in general and to allow them to be assessed and expeditiously rise to meet their potential to improve elections.

The challenges for the technical research community now forming around election technology includes further exploration and refinement of these new types of systems. Particularly promising and important areas include analysis, formal modeling, and rigorous proofs regarding systems and potential threats. Initial deployments of these systems are starting to provide valuable real-world experience, but effective ways to communicate and expose

their workings may also be important. The goal is systems that increase transparency regarding the correctness of the election results and yet maintain secrecy of individual votes. Improved voter confidence may follow.

Voting over electronic networks has various attractions, is starting to be deployed, and is regarded by some as inevitable. No solution, however, has been proposed that provides safeguards adequate against various known threats. Problems include attacks against the security of the computers used as well as attacks that impede communication over the network. Improper influence of remote voters is also a significant problem, although it is tolerated with vote by mail in numerous jurisdictions. Securing network voting is clearly an important research challenge. We cannot, however, prudently recommend any but unavoidable use of online voting systems in elections of significant consequence until effective means are developed to address these vulnerabilities.

10.4 Cryptography

Seminar No. **07381** Date **16.09.–21.09.2007**
Organizers: Johannes Blömer, Dan Boneh, Ronald Cramer, Ueli Maurer

Cryptography is of paramount importance for information security. Cryptographic primitives are the core building blocks for constructing secure systems. The last three decades have seen tremendous progress in cryptography and the field has substantially matured. Major achievements include the proposal of adequate security definitions, of new cryptographic schemes, and of security proofs for these schemes, relative to the security definition. As a consequence, cryptography has shifted from an ad-hoc discipline with many interesting tricks and ideas to a mathematically rigorous science. Despite this progress many essential problems in cryptography still remain open and new areas and topics arise constantly. The field is more lively than ever before.

While the number of scientific conferences focusing on cryptography is increasing, most of these meetings have a broad focus, and due to a growing interest by practitioners, the number of non-expert attendees has increased. As a result, it becomes more difficult to discuss the details of the advancement of the field, as well as to identify promising innovative trends. Therefore, the aim of the seminar was to provide an opportunity for key cryptographers to meet, to interact, to focus on the scientific foundation of cryptography, to spot the emerging new areas, and to work on them. Applications were also covered but the emphasis was on the conceptual framework that allows the use of appropriate models, amenable to mathematical reasoning.

The seminar brought together about 40 leading cryptographers from all over the world. Almost all participants gave a presentation about their recent research and also about future research plans they have, encouraging others to join in. In many cases the choice of the subject for the talk was targeted to the unique list of participants. The presentations were highly interactive and led to lively discussions, well into the evenings and nights. A number of new collaborations were initiated at the seminar. Overall, the seminar was

a great success, as is also documented by the feedback given by the participants on the questionnaires.

The topics covered in the seminar spanned most areas of cryptography, in one way or another, both in terms of the types of schemes (public-key cryptography, symmetric cryptography, hash functions and other cryptographic functions, multi-party protocols, etc.) and in terms of the mathematical methods and techniques used (algebra, number theory, elliptic curves, probability theory, information theory, combinatorics, quantum theory, etc.). The range of applications addressed in the various talks was broad, ranging from secure communication, key management, authentication, digital signatures and payment systems to e-voting and Internet security.

While the initial plan had been to focus more exclusively on public-key cryptography, it turned out that this sub-topic branches out into many other areas of cryptography and therefore the organizers decided to expand the scope, emphasizing quality rather than close adherence to public-key cryptography. This decision turned out to be a wise one.

What was common to almost all the talks is that rigorous mathematical proofs for the security of the presented schemes were given. In fact, a central topic of many of the talks were proof methodologies for various contexts.

10.5 Formal Protocol Verification Applied

Seminar No. **07421** Date **14.10.–19.10.2007**
Organizers: Liqun Chen, Steve Kremer, Mark D. Ryan

Introduction

Security protocols are a core part of distributed computing systems, and are part of our everyday life since they are used in web servers, email, mobile phones, bank transactions, etc. However, security protocols are notoriously difficult to get right. There are many cases of protocols which are proposed and considered secure for many years, but later found to have security flaws. Formal methods offer a promising way for automated security analysis of protocols. While there have been considerable advances in this area, most techniques have only been applied to academic case studies and security properties such as secrecy and authentication. The seminar brought together researchers deploying security protocols in new application areas, cryptographers, and researchers from formal methods who analyse security protocols. The interaction between researchers from these different communities aims to open new research topics, e.g., identify new security properties that need verification and refine abstractions of the abstract models of crytpographic primitives.

The seminar

Because of the multi-disciplinary nature of the workshop, not all of the participants knew each other in advance. We devoted the first morning to five-minute introductions of

ourselves and our areas of research, given by each participant (including those who did not later give a full talk). Additionally, we scheduled some tutorial talks on the first day in order to enable all of the participants to understand the relevant foundations. We had four tutorials from internationally renouned speakers, as follows:

- Kenny Paterson: Introduction to Provable Security
- Hubert Comon-Lundh: Introduction to Formal Methods Approach to Protocol Verification
- Catuscia Palamidessi: Overview of Formal Approaches to Information-hiding
- Ahmad-Reza Sadeghi: Tutorial on Security Protocols on Trusted Platforms

In addition, we had 24 technical talks, each of which brought together two or more of the themes of the workshop. The following table attempts to give a flavour for how the talks cut across and brought together the themes of the workshop. Naturally, most of the talks involved several themes so the categorisation represented by the table should not be taken too seriously.

	Analysis aspect		
Protocol aspect	Design	Provable	Formal meth.
Application			
Key exchange		Armknecht	Etalle
		Tsay	
Identity management			Bhargavan
Denial of service	Chadha		
Trusted computing		Chen	Maffei
			Rudolph
Payment			Klay
Password-based prot	Kremer		
Contract signing		Küsters	
Coupons		Löhr	
Voting			Ryan
Web services			Vigneron
Theory		Blanchet	Chatzikokolakis
			Comon-Lundh
			Corin
			Cremers
			Fournet
			Gordon
			Mödersheim
			Ritter
			Smyth

Microsoft sponsorship The seminar was sponsored by Microsoft Research Cambridge. A special dinner was held on Thursday evening to note this contribution.

Conclusion The seminar has led to much extensive discussion among the participants during and after the event. Quite a few of the papers presented have now been published.

Chapter 11

Data Bases, Information Retrieval

11.1 Web Information Retrieval and Linear Algebra Algorithms

Seminar No. **07071** Date **11.02.–16.02.2007**
Organizers: Andreas Frommer, Michael W. Mahoney, Daniel B. Szyld

A seminar concentrating on the intersection of the fields of information retrieval and other web-related aspects with numerical and applied linear algebra techniques was held with the attendance of scientists from industry and academia.

The scientific community has witnessed the increasing importance of linear algebra algorithms and of Markov chain modeling in several applications from computer science. Of particular importance is linear algebra algorithms to study the structure of the Web and information retrieval (IR) on the Web. The main focus of the seminar was the evolving theory and computational aspects of methods for web information retrieval, including search engines, that are inspired by traditional and recent advances in algorithms for linear algebra problems. To this end, the seminar brought together scientists from academia with background in computer science or numerical mathematics and scientists working in industry, mostly from Yahoo Research (both from the US and Europe).

Structure of the seminar

The seminar was attended by forty-seven participants coming from thirteen different countries. We had a good mixture of graduate students, young researchers, scientists in mid-career, and senior investigators from academia and industry. There was a total of thirty-one talks. Due to the diverse backgrounds of the attendees it was decided to have five longer expository talks which included introductions to the subjects and methods of the respective fields.

Outcome of the seminar

We want to highlight that the seminar really fostered interaction between people from academia and industry. Many participants observed that they benefited greatly from the contributions presented from researchers working in other fields or other settings.

Among the findings of this seminar, we mention the following: While it became clear from the scientists working in web retrieval that Pagerank now is just a minor ingredient in web ranking algorithms, it turns out that Pagerank like approaches continue to play an important role in other areas such as social science or community behavior. In this area, but also in more advanced, semantic models, the properties of eigenvalues and eigenvectors of huge sparse matrices and their computation continue to be at the heart of current research. Similarly, other classical matrix factorization techniques like the singular value decomposition have new applications, for example, in cluster analysis.

Techniques using low rank (and thus data efficient) approximations to huge matrices become increasingly important for data analysis and representation. For example, recent work has focused on employing randomization to improve low-rank computations and also large statistical regression problems. A particularly difficult issue is that traditional methods such as the SVD and QR decomposition destroy sparsity. Thus, low-rank approximations that respect sparsity are important. A second issue is that in many applications, one is not interested in the results of low-rank computations per se, but instead one wants to use it to learn from the data. Thus, studying matrix decompositions with good learning or generalization properties is important. Relatedly, in many cases an important question has to do with the best way to represent the data, i.e., which vector space is most appropriate to model the data in order to perform efficient computations.

Asynchronous iterative approaches, as they arise naturally in loosely coupled networks of processors have been analyzed from the theoretical side and are being used in practice. One challenging problem discussed, was that of data streams which cannot be stored, so that standard numerical techniques have to be enhanced, for example, with statistical analyses or using novel algorithmic methods. Another point of intersection between the disciplines were novel graph partitioning approaches using iterative methods from numerical linear algebra. This represents a particularly challenging direction since the local geometry of the data that arise in Web IR applications is very different from the geometry that arises in traditional applications.

11.2 Constraint Databases, Geometric Elimination and Geographic Information Systems

Seminar No. **07212** Date **20.05.–25.05.2007**
Organizers: Bernd Bank, Max J. Egenhofer, Bart Kuijpers

During the past 15 years the topic of constraint databases [1] has evolved into a mature area of computer science with sound mathematical foundations and with a profound

theoretical understanding of the expressive power of a variety of query languages. Constraint databases are especially suited for applications in which possibly infinite sets of continuous data, that have a geometric interpretation, need to be stored in a computer. Today, the most important application domains of constraint databases are geographic information systems (GIS), spatial databases and spatio-temporal databases [2,1]. In these applications infinite geometrical sets of continuous data are finitely represented by means of finite combinations of polynomial equality and inequality constraints that describe these data sets (in mathematical terms these geometrical data sets are known as semi-algebraic sets and they have been extensively studied in real algebraic geometry). On the other hand, constraint databases provide us with a new view on classic (linear and nonlinear) optimization theory.

A variety of languages, mostly extensions of first-order logic over the reals, has been proposed and studied for querying constraint databases in various applications. The expressive power of these query languages has been analyzed in many aspects, especially with applications in GIS and spatial databases in mind. On the other hand, beyond general complexity results of real algebraic geometry, little is known about the specific complexity of query evaluation in constraint database systems. Consequently the propagation of theoretical research results into database practice is hindered by the inefficiency of general purpose algorithms from real algebraic geometry used up to now for the implementation of query evaluation. These implementations are mostly based on quantifier-elimination and only query languages for linear constraint databases have been implemented in practice. The need for efficient algorithms is most visible for the basic query language FO, first-order logic over the reals. Also extensions of FO by for instance the "sum (of a finite set)", "topological connectivity","path connectivity" or other topological operators have received much attention in recent years and are considered to be of great importance for practical applications, specifically in GIS. Both for FO and for these extensions query evaluation is implemented through the standard general purpose algorithms from real algebraic geometry. The sequential time complexity of these algorithms depends intrinsically (and in worst case exponentially) on the arrangement size of the data and (superexponentially) on the number of quantifier alternations of the query under consideration. On the other hand this complexity is polynomial for fixed arrangement size of the data and fixed number of quantifier alternations of the query.

From the above it should be clear that researchers from the areas of constraint databases, geometric elimination algorithms and geographic information systems should work together to address the feasibility of the constraint database approach to deal with application demands in geographic information systems.

GIS researchers find in the constraint database model a powerful and elegant tool for application in spatial databases and GIS. Its clean mathematical formulation allows the study of the expressive power of query languages in much more rigorous way than is the case for most other, often ad-hoc, approaches in GIS. From the users side, GIS researchers can describe the requirements of applications and specify which fragments of the constraint database query languages are useful and needed in GIS practice.

Researchers in geometric elimination theory find a practical application par excellence of their algorithms in constraint databases. Efficient elimination algorithms form a bottleneck

for the development of practical development of constraint database systems that have potential for commercial use in GIS.

The aim of this seminar is to bring together researchers from the areas of constraint databases, geometric elimination algorithms and geographic information systems to address the feasibility of the constraint database in the area of geographic information systems. This seminar also has the explicit purpose of identifying an appropriate forum for presenting and discussing future advances and exploring cross-fertilization to related topics, possibly in the form of setting up a joint conference (or series of conferences) on this topic.

References:

[1] G. Kuper, L. Libkin and J. Paredaens (eds.), Constraint Databases, Springer-Verlag, 2000.

[2] P. Rigaux, M. Scholl, and A. Voisard, Spatial Databases—With Application to GIS. Morgan Kaufmann, 2001.

Chapter 12

Machine Learning

12.1 Probabilistic, Logical and Relational Learning – A Further Synthesis

Seminar No. **07161** Date **15.04.–20.04.2007**
Organizers: Luc De Raedt, Thomas Dietterich, Lise Getoor, Kristian Kersting, Stephen H. Muggleton

Data Mining and Machine Learning are in the midst of a "structured revolution". After many decades of focusing on independent and identically-distributed (iid) examples, many researchers are now studying problems in which examples consist of collections of inter-related entities or are linked together. A major driving force is the explosive growth in the amount of heterogeneous data that is being collected in the business and scientific world. Example domains include bioinformatics, chemoinformatics, transportation systems, communication networks, social network analysis, link analysis, robotics, among others. The structures encountered can be as simple as sequences and trees (such as those arising in protein secondary structure prediction and natural language parsing) or as complex as citation graphs, the World Wide Web, and logical knowledge bases. In all these cases, structured representations can give a more informative view of the problem at hand, which is often crucial for the development of successful mining and learning algorithms.

The field of *probabilistic, logical and relational learning* (aka. *statistical relational learning*, *probabilistic inductive logic programming*) tackles the structured input-output problem sketched above by combining expressive knowledge representation formalisms such as relational and first-order logic with principled probabilistic and statistical approaches to inference and learning, and hence lies at the heart of artificial intelligence. It is a relatively young and all the more active field of research offering a lot of research opportunities. This was already witnessed by a previous seminar on "Probabilistic, Logical and Relational Learning - Towards a Synthesis" that took place from January 30 to February 04, 2005, which succeeded in bringing together a significant number of researchers from all over the world that are working on all aspects of probabilistic, logical and relational learning. The result was a better understanding of the common grounds of this newly emerging field and the identification of a number of key research challenges.

The goal of the 2007 seminar was to provide answers to some of this key research challenges in the area, including:

1. What is the relationship among the many different probabilistic, logical and relational representations that are being used?
2. What are suitable settings for learning such representations? And, what are the challenges raised by the different learning settings? Also, can one arrive at a learning theory focused on probabilistic, logical and relational representations?
3. What are the application areas for which probabilistic, logical and relational learning is well-suited? What does it take to develop show-case applications in these areas? Can we identify common and concrete application challenges on which progress can be measured and techniques? Providing answers to these questions should — ultimately — provide the field with a commonly agreed upon framework as well as provide an application focus, which together could form the basis for further developments in the area.

Not all of the questions could have been answered yet but significantly progress has been made as shown by the great collection of abstracts. They have been collected from 45 seminar attendees from 11 different countries. The presentations at the seminar, varying in length, covered a large variety of topics, including novel results on lifted inference within first-order probabilistic languages, learning infinite relational models, statistical predicate invention, and applications within citation analysis, robotics, and life sciences. Talks were spread over the week to allow for plenty of time for discussions. Breakout sessions on special interest topics were organized on the fly using the Seminar's Wiki page. The breakout sessions gave the participants a chance to exchange problems and discuss ideas and challenges lying ahead in depth. We are positive that many of the breakout sessions will lead to new results, collaborations, and publications. Within the talks and the breakout sessions, we saw very lively debates showing the growing demand and opportunities for statistical relational learning within theory and practice of machine learning. We were also very pleased to see the significant progress made between the present seminar and the previous one. This was very clear in the demonstration session, where a number of academic prototypes of probabilistic, logical and relational learning systems were presented.

12.2 Parallel Universes and Local Patterns

Seminar No. **07181** Date **01.05.–04.05.2007**
Organizers: Michael R. Berthold, Katharina Morik, Arno Siebes

Introduction

Learning in parallel universes and the mining for local patterns are both relatively new fields of research. Local pattern detection addresses the problem of identifying (small)

deviations from an overall distribution of some underlying data in some feature space. Learning in parallel universes on the other hand, deals with the analysis of objects, which are given in different feature spaces, i.e. parallel universes; and the aim is on finding groups of objects, which show "interesting" behavior in some of these universes. So, while local patterns describe interesting properties of a subset of the overall space or set of objects, learning in parallel universes also aims at finding interesting patterns across different feature spaces or object descriptions. Dagstuhl Seminar 07181 on Parallel Universes and Local Patterns held in May 2007 brought together researchers with different backgrounds to discuss latest advances in both fields and to draw connections between the two.

Local Patterns

Research on local pattern detection emerged from the fact that most traditional methods in knowledge discovery and databases (KDD) seek to find global models, which describe the overall structure of a dataset and hence explain most of the objects contained in it, but tend to miss local deviations from a background model. The insights learned from such global models are often limited to observations which the domain expert is mostly already aware of and which are therefore not of special interest. In 2002, Hand organized a workshop on pattern detection and discovery and proposed the field of local pattern detection. Since then researchers with different backgrounds (e.g. statistics, machine learning, multi-relational data mining) have come together to establish and unify the field. Following the 2002 workshop, a second workshop took place in spring 2004 with the goal to find a definition for local patterns. The discussions brought up a number of – often only slightly different – definitions. This seminar continued the inspection of outliers (e.g., Neill Adams, Ira Assent).

Parallel Universes

The field of learning in parallel universes originated from the observation that the true objective of *data analysis* is not about mining the data but about mining their underlying objects. These objects are, for instance molecules, images or processes, which (by lack of a better representation) are described based on measurable features (e.g. molecular weight). Such a (set of) features is usually referred to as *data* but obviously there are manifold ways to derive features (or data) while focusing on different aspects of the underlying objects.The notion of learning in parallel universes has first been introduced in ?. The different feature spaces are regarded as parallel universes and the analysis in parallel is called learning in parallel universes. The aim is to identify configurations in the data, which are shared among different – not necessarily all – universes but also those which are typical to individual universes only. Communities of users which share a certain view of a collection of objects also share a set of features describing the objects, where other communities constitute another view of the same collection. The notion of parallel universes has obvious connections to the research field of Multi-View learning; however, multi-view learning requires all universes (or views) to contain the same information, i.e. there are no patterns specific to individual universes only.

How Local Patterns link to Parallel Universes

Throughout the seminar, there were lively discussion as to where to draw a connection between local pattern detection and learning in parallel universes. One obvious link is locality. In terms of local pattern detection it addresses the identification of small deviations from a background model. Similarly, in terms of parallel universes it means the identification of certain patterns that are typical to few (in its extreme one) universes only. Both a single local pattern as well as a pattern which occurs in one/few universes, can give valuable insights to the expert.

Locality in subgroup discovery (Stefan Rüping, Martin Scholz), in term sets (Francesco Bonchi, Jean-Francois Boulicaut, Bruno Cremilleux, Elisa Fromont), in clustering (Michael Berthold, Frank Höppner, Katharina Morik, Bernd Wiswedel), and over time (Bart Goethals, Frank Höppner) was investigated with respect to its link to parallel universes.

A link between both concepts can be drawn by reducing the learning in parallel universes to, without loss of generalizability, learning with different similarity measures. A local pattern induces also a (simple) similarity measure: two tuples are either equal or they are not (Arno Siebes). Similarly, a link was drawn between parallel universes and multi-objective learning (Ingo Mierswa, Claus Weihs): a universe is constituted by a criterion of success.

Organization

The goal of the proposed workshop was threefold. Firstly, we wanted to bring together researchers from the different disciplines to agree on a unifying framework for local pattern mining in parallel universes. So far, only algorithms that find clusters as local patterns have been proposed, for example for the grouping of active molecular compounds or the modeling of user preference clusters in different musical genres. It is not straight forward to extend this scenario to other types of pattern mining algorithms, which requires a careful study of the state of the art and a combination of existing approaches. Secondly, the interaction between different local patterns is an aspect that hinders existing algorithms. If a pattern belongs to two or more local patterns or, inversely, if two local patterns in different universes describe overlapping subsets of the data it becomes more complicated to algorithmically derive the entire set of local patterns that may exist in the data. Thirdly, the workshop aimed to produce a series of white papers describing the state of the art in local pattern mining in application areas where related problems have appeared in the past.

In order to achieve this, we invited researchers from different communities: local pattern mining, statistical data analysis, machine learning, and data mining. In addition we also invited participants from the Visual Data Mining community, since local pattern detection -especially in several descriptor spaces in parallel- is a method that inherently requires user feedback to be successful (Rudolf Kruse, Matthias Steinbrecher). For this, it is crucial to be able to present the user with a variety of -preferably interactive- views on the data (Arno Knobbe), each showing summaries of the discovered patterns in each universe together.