DAGSTUHL

REPORTS

Volume 5, Issue 2, Feburary 2015

Dagstuhl Reports, Vol. 5, Issue 2 ISSN 2192-5283

ISSN 2192-5283

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at
http://www.dagstuhl.de/dagpub/2192-5283

Publication date
July, 2015

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

Digital Object Identifier: 10.4230/DagRep.5.2.i

Aims and Scope
The periodical *Dagstuhl Reports* documents the program and the results of Dagstuhl Seminars and Dagstuhl Perspectives Workshops.
In principal, for each Dagstuhl Seminar or Dagstuhl Perspectives Workshop a report is published that contains the following:
- an executive summary of the seminar program and the fundamental results,
- an overview of the talks given during the seminar (summarized as talk abstracts), and
- summaries from working groups (if applicable).

This basic framework can be extended by suitable contributions that are related to the program of the seminar, e. g. summaries from panel discussions or open problem sessions.

Contact
Schloss Dagstuhl – Leibniz-Zentrum für Informatik
Dagstuhl Reports, Editorial Office
Oktavie-Allee, 66687 Wadern, Germany
reports@dagstuhl.de
http://www.dagstuhl.de/dagrep

Report from Dagstuhl Seminar 15061

Non-Zero-Sum-Games and Control

Edited by

Krishnendu Chatterjee[1], Stéphane Lafortune[2], Nicolas Markey[3], and Wolfgang Thomas[4]

1 **IST Austria – Klosterneuburg, AT,** Krishnendu.Chatterjee@ist.ac.at
2 **University of Michigan – Ann Arbor, US,** stephane@umich.edu
3 **ENS – Cachan, FR,** markey@lsv.ens-cachan.fr
4 **RWTH Aachen, DE,** thomas@cs.rwth-aachen.de

Abstract

In this report, the program, research issues, and results of Dagstuhl Seminar 15061 "Non-Zero-Sum-Games and Control" are described. The area of non-zero-sum games is addressed in a wide range of topics: multi-player games, partial-observation games, quantitative game models, and – as a special focus – connections with control engineering (supervisory control).

Seminar February 1–6, 2015 – http://www.dagstuhl.de/15061

1998 ACM Subject Classification B.6.3 Design Aids, F.1.2 Modes of Computation, I.2.2 Automatic Programming, I.2.8 Problem Solving, Control Methods, and Search

Keywords and phrases non-zero-sum games, infinite games, multi-player games, partial-observation games, quantitative games, controller synthesis, supervisory control

Digital Object Identifier 10.4230/DagRep.5.2.1

Edited in cooperation with Benedikt Brütsch

1 Executive Summary

Krishnendu Chatterjee
Stéphane Lafortune
Nicolas Markey
Wolfgang Thomas

Games played on graphs provide the framework to study a wide range of problems that are central in computer science, for example, reactive synthesis, well-formedness of systems, checking compatibility of behavorial type systems, etc. The traditional study of games has been for two-player zero-sum perfect-information deterministic games with Boolean objectives (where a win of one player coincides with a loss by the other player). Fundamental results of this theory are contributions of automata theory that go back to the 1960's (Büchi, McNaughton, Rabin).

Significant progress has been achieved in the last few decades both in terms of theoretical results (understanding the complexity of such games, developing efficient algorithms) as well as their practical applicability (in reactive synthesis and controller synthesis). The current research directions explore several important extensions of the traditional study, namely, multi-player games, games with partial-observation, quantitative aspects in games, as well as application of game results in other domains. In this regard, the connection to control theory is important: The methodology of "supervisory control" has developed in parallel to the emergence of the game-theoretic approach, and a joint and integrating view of these closely related branches of research seemed overdue.

The Dagstuhl Seminar "Non-Zero Sum Games and Control" addressed these developments, with a particular emphasis on the connections to control theory. The response to the call for participation was very positive, and the 42 scientists joining the seminar represented the full range of topics mentioned above. There was a very good mixture between young and "established" researchers, and the participation of female researchers (making a quarter) was high (in the context of computer science). In order to support the understanding between the different research branches, it was decided to have on each half-day at least one survey talk, outlining a field and describing general challenges; some of these talks were contributed by young researchers. The speakers of the survey talks were Rüdiger Ehlers, Tom Henzinger, Barbara Jobstmann, Stéphane Lafortune, Kim Larsen, Peter Bro Miltersen, Jean-François Raskin, Armando Solar-Lezama, John Thistle, and Ufuk Topcu. Furthermore, a special evening session was organized on challenges in supervisory control. Besides the aim of joining and integrating different tracks of research in the area, an important objective of the seminar was to bring together (at least some) members of two large research communities in the area, namely the community of automata, logic, and games in Europe and the U.S. research network EXCAPE (Expeditions in Computer Augmented Program Engineering).

During the seminar, several small circles of participants started or continued joint work. As a general result of the seminar, confirmed by many positive and even enthusiastic comments of participants after the seminar, one may say that a much better understanding and appreciation between the various research branches was established. As one of the participants put it, the seminar was "eye-opening".

As an overview of the areas covered in the talks, we give a short description of the topics studied in the seminar.

Multi-player games

The study of multi-player games is an important extension of the two-player setting. In terms of theoretical study it gives rise to a rich class of questions related to different notions of equilibria, studying computability and complexity results for them, as well as different logics to express them. In terms of practical applicability, various notions of synthesis such as rational synthesis, secure equilibria, assume-guarantee synthesis, assume-admissible synthesis, etc. have been developed to apply the results of multi-player games for synthesis of component-based systems.

Partial-observation games

Partial-observation games extend perfect-information games where players do not have perfect knowledge about the game. This is particularly relevant in control theory, where the controller does not have access to private variables of the plant. The results of partial-observation games have been recently extended to the stochastic setting, as well as for finite-memory strategies, leading to a new framework which can potentially solve interesting applications from the control domain.

Quantitative game models

These are a prominent class of game models for applications to verification and synthesis. In particular, taking real-time constraints into account is especially important for such applications. Timed automata and timed games have already played an important role, as they are a convenient and expressive model enjoying efficient algorithms. Statistical model checking in particular offers a very effective technique for strategy optimization. Robustness analysis of timed models makes the verification process even more faithful.

- Weighted timed games extend timed games with the ability of modeling other quantitative aspects of cyber-physical systems. While the expressive power is greatly improved, the verification and synthesis problems get much more complex than for plain timed automata. Still, algorithms sometimes exist for *approximating* the optimal cost, which in most practical situations is sufficient.
- Timed automata have now reached maturity. Powerful data-structures and efficient symbolic algorithms have been designed to develop efficient algorithms. Statistical model checking is now also used in tools for efficiently optimizing strategies. These tools can now be applied on real-life scenarios, e.g. in home automation and motion planning.
- Probabilistic models form another important class of models of particular interest for representing and reasoning about e.g. systems involving stochastic behaviors. Efficient algorithms have recently been developed for diagnosing probabilistic automata, or for synthesizing strategies that guarantee good performance level with sufficient probability in Markov Decision Processes.

In the quantitative setting, the range of objectives is large; there are mean-payoff objectives, energy objectives, mean-payoff of energy objectives, their Boolean combinations, and combinations of quantitative (e.g., stochastic) semantics with adversarial semantics.

Other domains

Theoretical results developed for games have been generalized to problems in other settings as well. A prime example is that the lower bound for strategy improvement algorithms for parity games was modified to obtain lower bounds for linear-programming solutions, and recent results show that exploiting structures of Markov decision processes it can be established that several classical rules for linear-programming algorithms solve PSPACE-complete problems.

Control engineering

In the field of control engineering, research on supervisory control of discrete event systems and on formal methods in feedback control has recently emphasized distributed and decentralized control architectures that more accurately capture the physical constraints arising in cyber-physical and networked control systems. In these architectures, a set of controllers, possibly with different run-time information about the system, cooperate as a team in order to achieve a specification (either qualitative or quantitative objective) on the entire system behavior, in the presence of a reactive environment. For instance, costly sensors and actuators, as well as costly communication, lead to challenging synthesis problems, both conceptually (e.g., characterization of the information structure) and computationally (e.g., distributed synthesis of the controllers). Another important consideration is to ensure robustness of the synthesized implementation with respect to classes of disturbances on the controlled system.

Researchers are still trying to establish the precise boundary between decidable and undecidable problems in this research domain. It is known that synthesis for both safety and a form of liveness termed non-blockingness, well-understood in a centralized-information setting, becomes undecidable in a decentralized-information setting. But the decidability of special classes of this problem is still an open issue. Establishing concrete bridges between the theory of partial-observation games and such decentralized/distributed control problems for discrete abstractions of cyber-physical systems is an important research issue, both in terms of answering open undecidability questions and in terms of developing efficient synthesis procedures for decidable problems. Similarly, problems of intrusion by malicious agents into control architectures (e.g., taking over actuators or sensors) also lead to new classes of problems where the theory of games with quantitative objectives can be leveraged.

Recently-developed synthesis techniques for "correct-by-construction" controllers in engineering systems have exploited game formulations between the set of controllers on the one hand and the system/environment on the other hand. Related approaches have considered synthesis of the "complete" controller implementation from a "partial" implementation and a sample set of desired behaviors in the reactive environment under consideration.

2 Table of Contents

3 Overview of Talks

3.1 Diagnosis of probabilistic systems

Nathalie Bertrand (INRIA Rennes – Bretagne Atlantique, FR)

Fault diagnosis aims at detecting the occurrence of faulty events in partially observable systems. When it comes to probabilistic models, one can relax the notion of diagnosability by requiring a set of ambiguous sequences of null measure. In our talk, we answered the following questions: How to check for diagnosability of a probabilistic system? And in case it is not diagnosable, can one control a system so that it is?

References

1 N. Bertrand, S. Haddad and E. Lefaucheux. Foundation of Diagnosis and Predictability in Probabilistic Systems. In *Proc. of FSTTCS'14*, LIPIcs, Schloss Dagstuhl, 2014. DOI: 10.4230/LIPIcs.FSTTCS.2014.417

2 N. Bertrand, E. Fabre, S. Haar, S. Haddad and L. Helouet. Active diagnosis for probabilistic systems. In *Proc. of FoSSaCS'14*, LNCS, Springer, 2014. DOI: 10.1007/978-3-642-54830-7_2

3.2 Infinite games with finite knowledge gaps

Dietmar Berwanger (ENS – Cachan, FR)

Joint work of Berwanger, Dietmar; Mathew, Anup Basil
Main reference D. Berwanger, A. B. Mathew, "Infinite games with finite knowledge gaps," arXiv:1411.5820v1 [cs.GT], 2014.
URL http://arxiv.org/abs/1411.5820v1

Infinite games where several players seek to coordinate under imperfect information are known to be intractable, unless the information flow is severely restricted. Examples of undecidable cases typically feature a situation where players become uncertain about the current state of the game, and this uncertainty lasts forever.

In contrast, we consider games where the players attain certainty about the current state over and over again, along every play. This leads to a new class on which the distributed synthesis problem is solvable. We show that the question of whether a game belongs to the class is NLOGSPACE-complete, and in that case, it is NEXPTIME-complete to decide whether a joint winning strategy exists.

3.3 On the value problem in weighted timed games

Patricia Bouyer-Decitre (CNRS, FR)

Joint work of Bouyer, Patricia; Jaziri, Samy; Markey, Nicolas
Main reference P. Bouyer, S. Jaziri, N. Markey, "On the Value Problem in Weighted Timed Games," Research Report LSV-14-12, Laboratoire Spécification et Vérification, ENS Cachan, France, 2014.
URL http://www.lsv.ens-cachan.fr/~bouyer/mes-publis.php?l=fr&onlykey=rr-lsv-14-12

In this talk, I present the model of weighted timed games. I give an overview of rather old results about the value problem in such games, and then present our new results, that concern the approximability of the value in weighted timed games, under a slight restriction on the weight.

3.4 The Complexity of Admissibility in Omega-Regular Games

Romain Brenguier (Free University of Brussels, BE)

Joint work of Brenguier, Romain; Raskin, Jean-François; Sassolas, Mathieu
Main reference R. Brenguier, J. Raskin, M. Sassolas, "The complexity of admissibility in Omega-regular games," in Proc. of the Joint Meeting of the 23rd EACSL Annual Conf. on Computer Science Logic (CSL'14) and the 29th Annual ACM/IEEE Symp. on Logic in Computer Science (LICS'14), pp. 21:1–23:10, ACM, 2014; pre-print available as arXiv:1304.1682v3 [cs.GT].
URL http://dx.doi.org/10.1145/2603088.2603143
URL http://arxiv.org/abs/1304.1682v3

Iterated admissibility is a well-known and important concept in classical game theory, e.g. to determine *rational* behaviors in multi-player matrix games. As recently shown by Berwanger, this concept can be soundly extended to infinite games played on graphs with ω-regular objectives. In this paper, we study the algorithmic properties of this concept for such games. We settle the exact complexity of natural decision problems on the set of strategies that survive iterated elimination of dominated strategies. As a byproduct of our construction, we obtain automata which recognize all the possible outcomes of such strategies.

3.5 Weak subgame perfect equilibria in quantitative multiplayer games

Véronique Bruyère (University of Mons, BE)

We study subgame perfect equilibria (SPE), and variants (weak SPE and very weak SPE), in quantitative multiplayer turn-based games played on graphs. We give a characterization of the outcomes of such equilibria (as a Folk theorem for Nash equilibria). This characterization allows us to construct finite-memory SPE in quantitative reachability games.

3.6 Robustness and Cooperation

Rüdiger Ehlers (Universität Bremen, DE)

Joint work of Roderick Bloem; Rüdiger Ehlers; Gangyuan Jing; Robert Könighofer; Hadas Kress-Gazit; Ufuk Topcu; Kai Weng Wong

In reactive synthesis, a correct-by-construction system is computed from a specification and an environment model. The synthesized system is guaranteed to satisfy its specification if the environment's behavior is covered by the model. Unfortunately, when theory meets practice, synthesized controllers often exhibit unsuitable behavior. Environment assumptions may sometimes be violated temporarily in the field, and there is no requirement for the controller to behave correctly in such a case. Equally problematic are controllers that try to force the environment to violate its assumptions or needlessly rely on liveness assumptions to hold, which leads to bad performance of controllers in practice. Yet, such controllers are possible solutions to the standard synthesis problem.

The talk addresses these issues and discusses results on synthesizing controllers that are robust, eager, and cooperative. Games with quantitative objectives are mostly avoided along the way, so the approaches leave room for further combination with other quantitative or qualitative optimization objectives.

3.7 The Complexity of the Simplex Method

John Fearnley (University of Liverpool, GB)

Joint work of Fearnley, John; Savani, Rahul
Main reference J. Fearnley, R. Savani, "The Complexity of the Simplex Method," arXiv:1404.0605v2 [cs.DS], 2014.
URL http://arxiv.org/abs/1404.0605v2

The simplex method is a well-studied and widely-used pivoting method for solving linear programs. When Dantzig originally formulated the simplex method, he gave a natural pivot rule that pivots into the basis a variable with the most violated reduced cost. In their seminal work, Klee and Minty showed that this pivot rule takes exponential time in the worst case. We prove two main results on the simplex method. Firstly, we show that it is PSPACE-complete to find the solution that is computed by the simplex method using Dantzig's pivot rule. Secondly, we prove that deciding whether Dantzig's rule ever chooses a specific variable to enter the basis is PSPACE-complete. We use the known connection between Markov decision processes (MDPs) and linear programming, and an equivalence between Dantzig's pivot rule and a natural variant of policy iteration for average-reward MDPs. We construct MDPs and show PSPACE-completeness results for single-switch policy iteration, which in turn imply our main results for the simplex method.

3.8 Two-Player Perfect-Information Zero-Sum Shift-Invariant Submixing Stochastic Games Are Half-Positional

Hugo Gimbert (University of Bordeaux, FR)

Joint work of Gimbert, Hugo; Kelmendi, Edon
Main reference H. Gimbert, E. Kelmendi, "Two-Player Perfect-Information Shift-Invariant Submixing Stochastic Games Are Half-Positional," arXiv:1401.6575v1 [cs.GT], 2014.
URL http://arxiv.org/abs/1401.6575v1

We consider zero-sum stochastic games with perfect information and finitely many states and actions. The payoff is computed by a payoff function which associates to each infinite sequence of states and actions a real number. We prove that if the the payoff function is both shift-invariant and submixing, then the game is half-positional, i.e., the first player has an optimal strategy which is both deterministic and stationary. This result relies on the existence of epsilon-subgame-perfect equilibria in shift-invariant games, a second contribution of the paper.

3.9 Quantitative Sabotage Games

Axel Haddad (Free University of Brussels, BE)

Joint work of Brihaye, Thomas; Geeraerts, Gilles; Haddad, Axel; Monmege, Benjamin; Perez, Guillermo; Renault, Gabriel

We study a model of evolving environment systems called quantitative sabotage games. Sabotage games have been introduced in an essay by Johan van Benthem [1] as special games on evolving arenas: whereas one player (Walker) moves from vertex to vertex following the topology of a finite graph, trying to reach a special target vertex, the other player (Saboteur) takes out an edge after each step. Hence, this game is naturally finite, and stops after a bounded number of steps. It has been shown to be PSPACE-complete to solve these games [2] as well as a randomized variant of it [3]. Our aim is to extend the study of sabotage games to deal with behavior of infinite lengths. To that purpose, the process of sabotage has to be changed to incorporate possibly unbounded behaviors. Our choice is to handle this by introducing quantities. Whereas Walker continues to move along the edges, Saboteur has now a budget of tokens that he puts on vertices of the graphs, and is allowed to move one token at every step. The first player wants to minimize its average number of tokens he sees during the whole play.

If the budget is fixed a priori, the problem of knowing whether the first player has a strategy to obtain an average value below a given threshold seems to be solvable in polynomial time. The static version of this problem, where the second player simply drops his tokens on the graph and leaves them as they are forever, is coNP-complete. In the general case, we show that the problem is EXPTIME-complete (we did not have the EXPTIME-hardness at the moment of the talk).

An EXPTIME algorithm can be obtained by studying the configuration graph of this game and solving the induced Mean-Payoff game. The EXPTIME-hardness is obtained by studying a variant of those games where Walker wants to never see a token. We show that such safety games can be encoded into quantitative sabotage games, and furthermore we reduce an EXPTIME-complete problem, finding the winner in a two-player version of the SAT problem [4], into this variant.

References

1 Johan van Benthem. An essay on sabotage and obstruction. In *Mechanizing Mathematical Reasoning, Essays in Honor of Jörg H. Siekmann on the Occasion of His 60th Birthday*, vol. 2605 of *LNCS*, pp. 268–276, Springer, 2005. DOI: 10.1007/978-3-540-32254-2_16

2 Christof Löding and Philipp Rohde. Solving the Sabotage Game Is PSPACE-Hard. In *Proc. of MFCS'03*, vol. 2747 of *LNCS*, pp. 531–540, Springer, 2003. DOI: 10.1007/978-3-540-45138-9_47

3 Dominik Klein, Frank G. Radmacher, and Wolfgang Thomas. Moving in a network under random failures: A complexity analysis. *Science of Computer Programming*, 77(7-8):940–954, 2012. DOI: 10.1016/j.scico.2010.05.009

4 Larry J. Stockmeyer and Ashok K. Chandra. Provably Difficult Combinatorial Games. *SIAM Journal on Computing*, 8(2):151–174, 1979. DOI: 10.1137/0208013

3.10 Games Everywhere

Thomas Anton Henzinger (IST Austria – Klosterneuburg, AT)

We survey some uses of graph games with qualitative and quantitative, zero-sum and non-zero-sum objectives for specifying and verifying reactive systems (game logics), defining and checking the well-formedness and compatibility of reactive systems (receptiveness and interface theories), refining and synthesizing reactive systems (simulation games and supervisory control).

3.11 The Big Match in small space

Rasmus Ibsen-Jensen (IST Austria – Klosterneuburg, AT)

Joint work of Ibsen-Jensen, Rasmus; Koucky, Michal; Hansen, Kristoffer Arnsfelt

The talk will be on the well-studied game the Big Match, a concurrent, zero-sum, two-player, mean-payoff game. It is known that no finite memory strategy exists that achieves anything non-trivial. All previously known ε-optimal strategies uses space $\Omega(\log T)$ in round T. The talk will be about an ε-optimal strategy, in the limsup sense, that uses space $O(s(T))$ in round T whp., for any unbounded, nondecreasing function s. The talk will also describe an ε-optimal strategy, in the liminf sense, that uses space $O(\log \log T)$ in round T whp.

The talk will then consider the open problem of generalising the result to arbitrary concurrent, zero-sum, two-player, mean-payoff games and the problems encountered while trying to do so.

3.12 Quantitative Objectives In Reactive Synthesis

Barbara Jobstmann (EPFL Lausanne, CH)

Joint work of Bloem, Roderick; Chatterjee, Krishnendu; Henzinger, Thomas; Singh, Rohit; Greimel, Karin; von Essen, Christian

Reactive Synthesis aims to automatically construct a reactive system from a temporal specification that describes the desired functional behavior of the system. Graph-based games are at the heart of many synthesizers for reactive systems. In the classical approach a synthesizer can produce any system that satisfies the specification. However, in many settings users prefer one correct system over another, e.g., among two correct systems, one might prefer the more robust or more efficient system.

In this talk I will first give a short introduction to Reactive Synthesis. Then, I will present quantitative objectives to express default behavior, robustness, and efficiency, and discuss the games that result from these objectives. Finally, I will present an approach to select repairs for a faulty program, a problem that can be solved using reactive synthesis. The novelty of this approach is that it requires a repair to keep all the correct behaviors of the faulty program, which enables us to synthesize meaningful repairs, even for incomplete specifications.

3.13 Distributed Synthesis in Continuous Time

Jan Krčál (Universität des Saarlandes, DE)

Joint work of Hermanns, Holger; Krčál, Jan; Vester, Steen

Distributed synthesis has been studied for more than 25 years. Various forms of communication of distributed agents have been discussed. However, the focus of previous foundational approaches to distributed synthesis has abstracted the flow of time to a discrete setting. In this paper, we introduce a formalism modelling communication of distributed agents strictly in continuous-time. Within this framework, we study the problem of synthesizing local strategies for individual agents such that a specified set of goal states is reached, or reached with at least a given probability. The flow of time is modelled explicitly based on continuous-time randomness, with two natural implications: First, the non-determinism stemming from interleaving of communication steps disappears. Second, when we restrict to a subclass of non-urgent models, both the qualitative and quantitative reachability problem can be solved in EXPTIME. The crucial observation is that explicit continuous time enables the players to communicate their states by delaying synchronization, turning it into full-observation setting. In general, the quantitative problem is undecidable for two or more players and the qualitative problem is EXPTIME-hard for two players and undecidable for three or more players.

3.14 Fast learning of small strategies

Jan Křetínský (IST Austria – Klosterneuburg, AT)

Joint work of Brázdil, Tomáš; Chatterjee, Krishnendu; Chmelík, Martin; Fellner, Andreas; Forejt, Vojtěch; Křetínský, Jan; Kwiatkowska, Marta; Parker, David; Ujma, Mateusz

In verification, precise analysis is required, but the algorithms usually suffer from scalability issues. In machine learning, scalability is achieved, but with only very weak guarantees. We show how to merge the two philosophies and profit from both. In this talk, we focus on $1\frac{1}{2}$-player games (Markov decision processes). We show how to learn ε-optimal strategies fast and how to represent them concisely so that some understanding of the behaviour and debugging information can be extracted.

References

1 Tomáš Brázdil, Krishnendu Chatterjee, Martin Chmelík, Andreas Fellner, and Jan Křetínský. Counterexample Explanation by Learning Small Strategies in Markov Decision Processes. *CoRR*, 2015. http://arxiv.org/abs/1502.02834
2 Tomáš Brázdil, Krishnendu Chatterjee, Martin Chmelík, Vojtěch Forejt, Jan Křetínský, Marta Z. Kwiatkowska, David Parker, and Mateusz Ujma. Verification of Markov Decision Processes Using Learning Algorithms. In *Proc. of ATVA'14*, vol. 8837 of *LNCS*, pp. 98–114, Springer, 2014. DOI: 10.1007/978-3-319-11936-6_8

3.15 Open Problems in Supervisory Control of Partially-Observed Discrete Event Systems

Stéphane Lafortune (University of Michigan – Ann Arbor, US)

Joint work of Lafortune, Stéphane; Yin, Xiang

We present a set of unsolved problems in the area of supervisory control of discrete event systems in control engineering. We consider systems with partial observation and limited controllability, representing the limited sensing and actuation capabilities of the engineering system to be controlled. We wish to address safety and nonblocking specifications in a maximally permissive manner. We also wish to consider a specification on the minimum required behavior for the controlled system. We list open problems in centralized and decentralized control architectures in the context of the above requirements.

3.16 From Timed Games to Stochastic Hybrid Games

Kim Guldstrand Larsen (Aalborg University, DK)

In this talk the focus is on extensions of the well-established formalism of timed automata towards games. The formalism of (stochastic) timed automata has been extensively used for modelling several real-time controllers and communication protocols with efficient algorithms as found in the tool UPPAAL and its branch UPPAAL SMC allowing for automatic (statistical) model checking of a range of desired correctness and performance properties of these complex systems.

Moving to the setting of games marks a paradigm shift, where control programs that are correct-by-construction are automatically synthesized from the desired properties (objectives). The talk provides insight into the symbolic on-the-fly algorithm used in the tool UPPAAL TIGA for efficient construction of winning control strategies with respect to reachability, safety, time-bounded reachability as well as Büchi conditions. In fact the tool has previously been applied to the synthesis of control strategies for a number of industrial cases (including control for pig stables, and control for hydraulic pumps) that exhibit performance significantly better than the (at the time) current industrial solution.

The talk also presents the newest branch of the UPPAAL tool suit, namely UPPAAL STRATEGO. In STRATEGO strategies are first-class citizens that may be named and used to constrain any existing type of model checking query of UPPAAL and UPPAAL SMC; i.e. additional (performance) properties of a game under a synthesized strategy may be performed. In addition, given a (most permissive) strategy ensuring given (hard real-time) safety constraints – and obtained by symbolic methods from UPPAAL TIGA – the tool may synthesize sub-strategies that (near-) optimize the expectation of additional performance measures – e.g. total energy consumed, waiting time of given component. This optimization is made using reinforcement learning, a well-established technique from machine learning.

The methods of performance analysis (using UPPAAL SMC) and optimization (using UPPAAL STRATEGO) has been extended to stochastic hybrid automata/games. Future work includes strategy learning under partial observability, thus addressing the need for compact control programs. Also, the use of statistical methods for rare event estimation of stochastic timed automata, as well as data structures for more general classes of strategies are work to be dealt with in near future.

3.17 Average Energy in Weighted Games

Simon B. Laursen (Aalborg University, DK)

Joint work of Laursen, Simon; Bouyer, Patricia; Larsen, Kim G.; Markey, Nicolas; Randour, Mickael

We consider a two player game on a weighted graph, where the goal for Player 1 is to keep the average accumulated weight under a given threshold, we name this objective average energy. We describe how this objective relates to the mean-payoff and total-payoff objectives. Our results include complexity and memory analyses of problems related to average energy games combined with lower and upper bound energy constraints.

3.18 Temporal logics for non-zero-sum games

Nicolas Markey (ENS – Cachan, FR)

Joint work of Brihaye, Thomas; Da Costa-Lopes, Arnaud; Laroussinie, François; Markey, Nicolas

Several extensions of the temporal logics CTL have been proposed in the recent years to express properties of multi-player games. In this talk I develop one such extension, ATL with strategy contexts, and propose a generic technique for dealing with it. I then show how this technique is also applicable to Strategy Logic. I conclude with a list of directions for further research.

References

1 François Laroussinie and Nicolas Markey. Quantified CTL: Expressiveness and Complexity. *Logical Methods in Computer Science* 10(4:17), 2014.
2 François Laroussinie and Nicolas Markey. Augmenting ATL with strategy contexts. *Information and Computation*, 2015. To appear.

3.19 Computer Poker and Computational Game Theory

Peter Bro Miltersen (Aarhus University, DK)

We survey recent work by the computer poker community on computational game theory, focusing (we hope), on work relevant for the formal methods community in general and the non-zero-sum-games-for-controller-synthesis subcommunity in particular.

3.20 Why Negatively-Priced Timed Games Are Hard

Benjamin Monmege (Free University of Brussels, BE)

Joint work of Brihaye, Thomas; Geeraerts, Gilles; Krishna, Shankara Narayanan; Lefaucheux, Engel; Manasa, Lakshmi; Monmege, Benjamin; Trivedi, Ashutosh
Main reference T. Brihaye, G. Geeraerts, S. N. Krishna, L. Manasa, B. Monmege, A. Trivedi, "Adding negative prices to priced timed games," in Proc. of the 25th Int'l Conf. on Concurrency Theory (CONCUR'14), LNCS, Vol. 8704, pp. 560–575, Springer, 2014.
URL http://dx.doi.org/10.1007/978-3-662-44584-6_38

Priced timed games (PTGs) are two-player zero-sum games played on the infinite graph of configurations of priced timed automata where two players take turns to choose transitions in order to optimize cost to reach target states. Bouyer et al. [2] and Alur, Bernadsky, and Madhusudan [1] independently proposed algorithms to solve PTGs with non-negative prices under certain divergence restriction over prices. Brihaye, Bruyère, and Raskin [4] later provided a justification for such a restriction by showing the undecidability of the optimal strategy synthesis problem in the absence of this divergence restriction. This problem for PTGs with one clock has long been conjectured to be in polynomial time, however the current best known algorithms are exponential, like the one by Rutkowski [6] (adapted from a previous attempt from by Bouyer, Larsen, Markey, and Rasmussen [3]), or by Hansen, Ibsen- Jensen, and Miltersen [5].

In this talk, we study the extension of PTGs with both negative and positive prices. I will summarize the results we previously obtained: some new undecidability results, and the identification of a subclass (one-clock, bi-valued price-rates) for which we designed a pseudo-polynomial time algorithm (and even polynomial time if non-negative bi-valued price-rates) to partially answer the conjecture on the complexity of one-clock PTGs. The rest of my talk will show some examples of PTGs not in the latter subclass with interesting phenomena, showing some strong features of PTGs, and our hopes to obtain a decidability result for a large class of one-clock PTGs with negative prices.

References

1. Rajeev Alur, Mikhail Bernadsky, and P. Madhusudan. Optimal reachability for weighted timed games. In *Proc. of the 31st Int'l Colloquium on Automata, Languages and Programming (ICALP'04)*, vol. 3142 of *LNCS*, pp. 122–133, Springer, 2004. DOI: 10.1007/978-3-540-27836-8_13
2. Patricia Bouyer, Franck Cassez, Emmanuel Fleury, and Kim G. Larsen. Optimal strategies in priced timed game automata. In *Proc. of the 24th Conf. on Foundations of Software Technology and Theoretical Computer Science (FSTTCS'04)*, vol. 3328 of *LNCS*, pp. 148–160, Springer, 2005. DOI: 10.1007/978-3-540-30538-5_13
3. Patricia Bouyer, Kim G. Larsen, Nicolas Markey, and Jacob Illum Rasmussen. Almost optimal strategies in one-clock priced timed games. In *Proc. of the 26th Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS'06)*, vol. 4337 of *LNCS*, pp. 345–356, Springer, 2006. DOI: 10.1007/11944836_32
4. Thomas Brihaye, Véronique Bruyère, and Jean-François Raskin. On optimal timed strategies. In *Proc. of the 3rd Int'l Conf. on Formal Modeling and Analysis of Timed Systems (FORMATS'05)*, vol. 3829 of *LNCS*, pp. 49–64, Springer, 2005. DOI: 10.1007/11603009_5
5. Thomas Dueholm Hansen, Rasmus Ibsen-Jensen, and Peter Bro Miltersen. A faster algorithm for solving one-clock priced timed games. In *Proc. of the 24th Int'l Conf. on Concurrency Theory (CONCUR'13)*, vol. 8052 of *LNCS*, pp. 531–545, Springer, 2013. DOI: 10.1007/978-3-642-40184-8_37
6. Michał Rutkowski. Two-player reachability-price games on single-clock timed automata. In *Proc. of the 9th Workshop on Quantitative Aspects of Programming Languages (QAPL'11)*, vol. 57 of *EPTCS*, pp. 31–46, 2011. DOI: 10.4204/EPTCS.57.3

3.21 Correct-by-construction controller synthesis for highly dynamic systems: an application in automotive safety systems

Necmiye Ozay (University of Michigan – Ann Arbor, US)

Joint work of Ozay, Necmiye; Nilsson, Petter; Liu, Jun; Grizzle, Jessy; Chen, Yuxiao; Peng, Huei

A plethora of driver convenience and safety automation systems are being introduced into production vehicles, such as electronic stability control, adaptive cruise control, lane keeping, and obstacle avoidance. Assuring the seamless and safe integration of each new automation function with existing control functions is a major challenge for vehicle manufacturers. In this talk, I will present some preliminary results to address this problem through the use of formal methods and correct-by-construction controller synthesis techniques. Mechanisms for handling implementation, perception or model imperfections will be discussed. I will conclude the talk with some open problems and directions for future research.

3.22 Fun with funnels

Nicolas Perrin (UPMC – Paris, FR)

Related to the notion of Lyapunov stability, funnels are regions of finite time invariance that can help to reason about dynamical systems whose differential equations are too difficult to handle directly. We explain how finite libraries of funnels could be used to construct variants of timed automata with potentially interesting applications to switching controller synthesis with collision avoidance objectives.

3.23 Percentile Queries in Multi-Dimensional Markov Decision Processes

Mickael Randour (ENS – Cachan, FR)

Joint work of Randour, Mickael; Raskin, Jean-François; Sankur, Ocan
Main reference M. Randour, J.-F. Raskin, O. Sankur, "Percentile queries in multi-dimensional Markov decision processes," arXiv:1410.4801v1 [cs.LO], 2014.
URL http://arxiv.org/abs/1410.4801v1

Markov decision processes (MDPs) with multi-dimensional weights are useful to analyze systems with multiple objectives that may be conflicting and require the analysis of trade-offs. In this work, we study the complexity of percentile queries in such MDPs and give algorithms to synthesize strategies that enforce such constraints. Given a multi-dimensional weighted MDP and a quantitative payoff function f, thresholds v_i (one per dimension), and probability thresholds α_i, we show how to compute a single strategy to enforce that for all dimensions i, the probability of outcomes ρ satisfying $f_i(\rho) \geq v_i$ is at least α_i. We consider classical quantitative payoffs from the literature (sup, inf, lim sup, lim inf, mean-payoff, truncated sum, discounted sum). Our work extends to the quantitative case the multi-objective model checking problem studied by Etessami et al. in unweighted MDPs.

3.24 Beyond Two-Player Zero-sum Games. Motivations and Highlights of Recent Results

Jean-François Raskin (Free University of Brussels, BE)

We motivate and review three recent results that we have obtained about non-zero sum games:

1. *Doomsday Equilibria for Omega-Regular Games* (Krishnendu Chatterjee, Laurent Doyen, Emmanuel Filiot, Jean-François Raskin)

Two-player games on graphs provide the theoretical framework for many important problems such as reactive synthesis. While the traditional study of two-player zero-sum games has been extended to multi-player games with several notions of equilibria, they are decidable only for perfect-information games, whereas several applications require imperfect-information games. In this paper we propose a new notion of equilibria, called doomsday equilibria, which is a strategy profile such that all players satisfy their own objective, and if any coalition of players deviates and violates even one of the players objective, then the objective of every player is violated. We present algorithms and complexity results for deciding the existence of doomsday equilibria for various classes of omega-regular objectives, both for imperfect-information games, and for perfect-information games. We provide optimal complexity bounds for imperfect-information games, and in most cases for perfect-information games.

2. *The Complexity of Admissibility in Omega-Regular Games* (Romain Brenguier, Jean-François Raskin, Mathieu Sassolas)
Iterated admissibility is a well-known and important concept in classical game theory, e.g. to determine rational behaviors in multi-player matrix games. As recently shown by Berwanger, this concept can be soundly extended to infinite games played on graphs with omega-regular objectives. In this paper, we study the algorithmic properties of this concept for such games. We settle the exact complexity of natural decision problems on the set of strategies that survive iterated elimination of dominated strategies. As a byproduct of our construction, we obtain automata which recognize all the possible outcomes of such strategies.
3. *Meet Your Expectations With Guarantees: Beyond Worst-Case Synthesis in Quantitative Games* (Véronique Bruyère, Emmanuel Filiot, Mickael Randour, Jean-François Raskin)
We extend the quantitative synthesis framework by going beyond the worst-case. On the one hand, classical analysis of two-player games involves an adversary (modeling the environment of the system) which is purely antagonistic and asks for strict guarantees. On the other hand, stochastic models like Markov decision processes represent situations where the system is faced to a purely randomized environment: the aim is then to optimize the expected payoff, with no guarantee on individual outcomes. We introduce the beyond worst-case synthesis problem, which is to construct strategies that guarantee some quantitative requirement in the worst-case while providing an higher expected value against a particular stochastic model of the environment given as input. This problem is relevant to produce system controllers that provide nice expected performance in the everyday situation while ensuring a strict (but relaxed) performance threshold even in the event of very bad (while unlikely) circumstances. We study the beyond worst-case synthesis problem for two important quantitative settings: the mean-payoff and the shortest path. In both cases, we show how to decide the existence of finite-memory strategies satisfying the problem and how to synthesize one if one exists. We establish algorithms and we study complexity bounds and memory requirements.

3.25 Synthesis for Human-in-the-Loop Control Systems

Dorsa Sadigh (University of California – Berkeley, US)

Joint work of Sadigh, Dorsa; Li, Wenchao; Seshia, Sanjit; Sastry, Shankar
Main reference W. Li, D. Sadigh, S. A. Seshia, S. S. Sastry, "Synthesis for Human-in-the-Loop Control Systems," in Proc. of the 20th Int'l Conf. on Tools and Algorithms for the Construction and Analysis of Systems (TACAS'14), LNCS, Vol. 8413, pp. 470–484, 2014.
URL http://dx.doi.org/10.1007/978-3-642-54862-8_40

Several control systems in safety-critical applications involve the interaction of an autonomous controller with one or more human operators. Examples include pilots interacting with an autopilot system in an aircraft, and a driver interacting with automated driver-assistance features in an automobile. The correctness of such systems depends not only on the autonomous controller, but also on the actions of the human controller. In this paper, we present a formalism for human-in-the-loop (HuIL) control systems. Particularly, we focus on the problem of synthesizing a semi-autonomous controller from high-level temporal specifications that expect occasional human intervention for correct operation. We present an algorithm for this problem, and demonstrate its operation on problems related to driver assistance in automobiles.

3.26 Symbolic Quantitative Robustness Analysis of Timed Automata

Ocan Sankur (Free University of Brussels, BE)

Main reference O. Sankur,"Symbolic Quantitative Robustness Analysis of Timed Automata," in Proc. of the 21st Int'l Conf. on Tools and Algorithms for the Construction and Analysis of Systems (TACAS'15), LNCS, Vol. 9035, pp. 484–498, Springer, 2015.
URL http://dx.doi.org/10.1007/978-3-662-46681-0_48

We consider the robustness analysis of real-time systems modeled by timed automata, where the goal is to compute a bound on the timing imprecisions so that the model satisfies a given specification. We give a semi-algorithm for infinitesimal analysis, which consists in finding a safe bound on imprecisions (rather than the maximal bound). Our algorithm uses parameterized zones, and exact zone approximations, and performs LTL model checking. The implemented tool is shown to perform well on several standard benchmarks.

3.27 Quantitative Verification in Rational Environments

Sven Schewe (University of Liverpool, GB)

Joint work of Gupta, Anshul; Schewe, Sven
Main reference A. Gupta, S. Schewe, "Quantitative Verification in Rational Environments," in Proc. of the 21st Int'l Symp. on Temporal Representation and Reasoning (TIME'14), pp. 123–131, IEEE Computer Society, 2014.
URL http://dx.doi.org/10.1109/TIME.2014.9

We provide a motivation for and discuss the complexity of finding optimal leader / Stackelberg equilibria for non-zero-sum mean payoff games. In this setting, a clever leader trespasses on the Nash-y rationality of her followers, using their goals to increase her overall return. We show that this quantitative verification problem is NP-complete, and in PTIME with a (two player zero-sum) MPG oracle for a bounded number of players.

3.28 Supervisory Control Synthesis for Deterministic Context Free Specification Languages – Enforcing Controllability Least Restrictively

Anne-Kathrin Schmuck (TU Berlin, DE)

Joint work of Schmuck, Anne-Kathrin; Schneider, Sven; Raisch, Joerg; Nestmann, Uwe
Main reference A. K. Schmuck, S. Schneider, J. Raisch, U. Nestmann, "Extending Supervisory Controller Synthesis to Deterministic Pushdown Automata – Enforcing Controllability Least Restrictively," in Proc. of the 12th Int'l Workshop on Discrete Event Systems, pp. 286–293, 2014.
URL http://dx.doi.org/10.3182/20140514-3-FR-4046.00058

Supervisory Control Theory (SCT) was established by Ramadge and Wonham in the early 80's for the purpose of controller synthesis on formal languages using their finite automaton realizations. This implies that this synthesis can only be applied to regular plant and specification languages, as only regular languages can be realized by deterministic finite automata (DFA). While the plant can typically be realized by a DFA, requiring that the specification can also be realized by a DFA is rather restrictive.

After introducing the original synthesis for regular languages, we have discussed an extension of SCT to a larger class of specifications, namely specifications that can be written as deterministic context free languages.

3.29 Program Synthesis for Games and Control

Armando Solar-Lezama (MIT - Cambridge, US)

Joint work of Solar-Lezama, Armando; Chaudhuri, Swarat
Main reference A. Solar-Lezama, S. Chaudhuri, "Smooth Interpretation," in Proc. of the 2010 ACM SIGPLAN Conf. on Programming Language Design and Implementation (PLDI'10), pp. 279–291, ACM, 2010.
URL http://dx.doi.org/10.1145/1809028.1806629

This talk described how recent work on constraint based synthesis via sketches can be applied to solve for winning strategies for two player games. The talked used the example of the "Cinderella Game" to illustrate the approach and to show some of its limitations. The talk also described how the ideas of encoding a synthesis problem as a sketch can be leveraged in the context of controller synthesis. Specifically I described some joint work with Swarat Chaudhuri on the use of smooth interpretation together with sketches to facilitate the search of optimal controllers.

3.30 Home Automation Synthesis: From Toy Examples to Realistic Scenarios

Jiří Srba (Aalborg University, DK)

Joint work of Jiří Srba; Mathias Grund Sørensen; Kim G. Larsen; David Junker; Lasse E. Nielsen; Mads Mikkelsen; Martin Lykke; Martin Z. Kristensen; Søren B. Andersen

The synthesis of a controller for a home automation systems is a complex task. In this talk we shall present an if-this-then-that rule format for the description of intended behaviour of a smart house and discuss the different approaches for automatic controller synthesis. We discuss their limitations and demonstrate a prototype implementation on a scaled model of a real house.

3.31 Games among arbitrarily many players

John G. Thistle (University of Waterloo, CA)

Joint work of M. Hadi Zibaeenejad; Thistle, John G.
Main reference M. H. Zibaeenejad, J. G. Thistle, "Weak invariant simulation and its application to analysis of parameterized networks," IEEE Transactions on Automatic Control, 59(8):2024–2037, 2014.
URL http://dx.doi.org/10.1109/TAC.2014.2315311

Distributed control of manufacturing networks, transportation networks, groups of mobile robots and the like, is a form of strategy design for nonzero-sum games. In many of these applications, the number of players may vary with time; it may be large, making explicit modelling cumbersome; or indeed it may be unknown. For such reasons it may be of interest to study games involving arbitrary numbers of players. Control design – or strategy design – can be seen as involving alternating phases of synthesis and verification: for example, disabling of a state transition or a move may give rise to livelocks or deadlocks, requiring further design iterations. This talk will focus on verification of the existence of reachable deadlock states of a game.

The literature – specifically the literature on "parameterized systems" – contains a number of decidability results for verification of relatively simple games among arbitrary numbers of players (where the allowed sequences of moves of each player are defined by an isomorphism to a given template process, and where the players' strategies are likewise isomorphic). For instance, players may be assumed to form a ring network, in which they interact directly only with their immediate neighbours, and only through the passing of tokens that carry no data. This talk asks whether there exist more complex games that still admit decision procedures for deadlock. It submits as an answer a somewhat tentative "yes". Our "existence proof" is based on a novel, and unconventional, process-algebraic simulation relation that allows the formulation of less restrictive assumptions about modes of interaction. In the case of ring networks, these assumptions allow the set of reachable deadlocked states to be characterized as those involving circular waits – and such states can be represented as the language accepted by a finite automaton over the alphabet of state symbols of the template process. For a more general class of network topologies featuring branching, we can similarly characterize the set of suitably generalized circular waits as the language accepted by a finite automaton on finite trees. Under some additional assumptions, we can show that the deadlocked network states are exactly those containing generalized circular waits.

3.32 Correct-by-construction control protocol synthesis for autonomous systems

Ufuk Topcu (University of Pennsylvania, US)

How can we affordably build trustworthy autonomous, networked systems? Partly motivated by this question, I describe a shift from the traditional "design+verify" approach to "specify+synthesize" in model-based development of autonomous systems. I then discuss our recent results on automated synthesis of correct-by-construction, hierarchical control protocols. These results account for hybrid dynamics that are subject to rich temporal logic specifications and heterogenous uncertainties, and that operate in adversarial environments. They combine ideas from control theory with those from computer science, and exploit underlying system-theoretic interpretations to suppress the inherent computational complexity. The expressivity of the resulting design methodology enables us to formally investigate a number of emerging issues in autonomous, networked systems. I conclude my talk with a brief overview of several such issues from my ongoing projects: (i) compositional synthesis for the so-called fractionated systems; (ii) effects of perception imperfections on protocol synthesis; (iii) interfaces between learning modules and reactive controllers with provable guarantees of correctness; and (iv) human-embedded autonomy.

3.33 Refinement Calculus for Reactive Systems

Stavros Tripakis (University of California – Berkeley, US)

Joint work of Preoteasa, Viorel; Tripakis, Stavros
Main reference V. Preoteasa, S. Tripakis, "Refinement Calculus of Reactive Systems," in Proc. of the 14th Int'l Conf. on Embedded Software (EMSOFT'14), pp. 2:1–2:10, ACM, 2014; pre-print available as arXiv:1406.6035v1 [cs.SE].
URL http://dx.doi.org/10.1145/2656045.2656068
URL http://arxiv.org/abs/1406.6035v1

Refinement calculus is a powerful and expressive tool for reasoning about sequential programs in a compositional manner. In this talk I present an extension of refinement calculus for reactive systems. Refinement calculus is based on monotonic predicate transformers, which transform sets of post-states into sets of pre-states. To model reactive systems, we introduce monotonic property transformers, which transform sets of output traces into sets of input traces. This semantics can model refinement, sequential composition, demonic choice, and other semantic operations on reactive systems. Syntactically, monotonic property transformers can be described in higher order logic, but also using other formalisms more amenable to automation, such as linear temporal logic (suitable for specifications) and symbolic transition systems (suitable for implementations). This framework generalizes previous work on relational interfaces so as to be able to express systems with infinite behaviors and liveness properties.

3.34 Synthesizing Finite-state Protocols from Scenarios and Requirements

Stavros Tripakis (University of California – Berkeley, US)

Joint work of Alur, Rajeev; Martin, Milo; Raghothaman, Mukund; Stergiou, Christos; Tripakis, Stavros; Udupa, Abhishek
Main reference R. Alur, M. M. K. Martin, M. Raghothaman, C. Stergiou, S. Tripakis, A. Udupa, "Synthesizing Finite-state Protocols from Scenarios and Requirements," in Proc. of the 10th Int'l Haifa Verification Conf. (HVC'14), LNCS, Vol. 8855, pp. 75–91, Springer, 2014; pre-print available as arXiv:1402.7150v1 [cs.FL].
URL http://dx.doi.org/10.1007/978-3-319-13338-6_7
URL http://arxiv.org/abs/1402.7150v1

Scenarios, or Message Sequence Charts, offer an intuitive way of describing the desired behaviors of a distributed protocol. In this paper we propose a new way of specifying finite-state protocols using scenarios: we show that it is possible to automatically derive a distributed implementation from a set of scenarios augmented with a set of safety and liveness requirements, provided the given scenarios adequately *cover* all the states of the desired implementation. We first derive incomplete state machines from the given scenarios, and then synthesis corresponds to completing the transition relation of individual processes so that the global product meets the specified requirements. This completion problem, in general, has the same complexity, PSPACE, as the verification problem, but unlike the verification problem, is NP-complete for a constant number of processes. We present two algorithms for solving the completion problem, one based on a heuristic search in the space of possible completions and one based on OBDD-based symbolic fixpoint computation. We evaluate the proposed methodology for protocol specification and the effectiveness of the synthesis algorithms using the classical alternating-bit protocol.

3.35 Omega-regular and Max-regular Delay Games

Martin Zimmermann (Universität des Saarlandes, DE)

Joint work of Zimmermann, Martin; Klein, Felix
Main reference F. Klein, M. Zimmermann, "How Much Lookahead is Needed to Win Infinite Games?" arXiv:1412.3701v1 [cs.GT], 2014.
URL http://arxiv.org/abs/1412.3701v1

In delay games, one of the players is able to delay her moves to obtain a lookahead on her opponent's moves. Recently, we showed that exponential lookahead is sufficient for omega-regular games with delay, gave a matching lower bound, and showed solving such games to be EXPTIME-complete.

Similar techniques are also applicable to quantitative extensions of the omega-regular languages, e.g., max-regular languages. These are the languages definable in weak MSO with the unbounding quantifier. We showed that max-regular delay games with respect to bounded lookahead are decidable and gave a doubly-exponential upper bound on the necessary lookahead. On the other hand, we showed that bounded delay is not always sufficient to win such a game.

I will present these results and discuss open problems, mainly on quantitative winning conditions, e.g., there is no matching lower bound in the max-regular case and the exact complexity is open.

This is (partially) joint work with Felix Klein (Saarland University).

4 Working Groups

4.1 Discussion on Supervisory Control and Reactive Synthesis

Stéphane Lafortune (University of Michigan – Ann Arbor, US)

We discussed how to approach open problems in supervisory control of discrete event systems using reactive synthesis techniques. We considered in particular the range problem in supervisory control under partial observations, where in addition to the safety and non-blocking specifications, a minimum required behavior is imposed on the controlled system. We also discussed how to establish the decidability, or lack thereof, of special cases of the decentralized supervisory control problem where non-blockingness (for which undecidability is known) is replaced by the weaker condition of deadlock-freeness or by a condition on local maximality of the solution. The participants exchanged several ideas on these problems that led to conjectures that will be pursued in future investigations.

Participants

- Nathalie Bertrand
INRIA Rennes – Bretagne Atlantique, FR
- Dietmar Berwanger
ENS – Cachan, FR
- Patricia Bouyer-Decitre
CNRS, FR
- Romain Brenguier
Free University of Brussels, BE
- Benedikt Brütsch
RWTH Aachen, DE
- Véronique Bruyère
University of Mons, BE
- Krishnendu Chatterjee
IST Austria – Klosterneuburg, AT
- Laurent Doyen
ENS – Cachan, FR
- Rüdiger Ehlers
Universität Bremen, DE
- John Fearnley
University of Liverpool, GB
- Gilles Geeraerts
Free University of Brussels, BE
- Hugo Gimbert
University of Bordeaux, FR
- Alessandro Giua
University of Aix-Marseille, FR & University of Cagliari, IT
- Axel Haddad
Free University of Brussels, BE
- Thomas Anton Henzinger
IST Austria – Klosterneuburg, AT
- Rasmus Ibsen-Jensen
IST Austria – Klosterneuburg, AT
- Barbara Jobstmann
EPFL Lausanne, CH
- Jan Krčál
Universität des Saarlandes, DE
- Jan Křetínský
IST Austria – Klosterneuburg, AT
- Stéphane Lafortune
University of Michigan – Ann Arbor, US
- Kim Guldstrand Larsen
Aalborg University, DK
- Simon B. Laursen
Aalborg University, DK
- Christof Löding
RWTH Aachen, DE
- Nicolas Markey
ENS – Cachan, FR
- Peter Bro Miltersen
Aarhus University, DK
- Benjamin Monmege
Free University of Brussels, BE
- Necmiye Ozay
University of Michigan – Ann Arbor, US
- Nicolas Perrin
UPMC – Paris, FR
- Sophie Pinchinat
IRISA – Rennes, FR
- Mickael Randour
ENS – Cachan, FR
- Jean-François Raskin
Free University of Brussels, BE
- Dorsa Sadigh
University of California – Berkeley, US
- Ocan Sankur
Free University of Brussels, BE
- Sven Schewe
University of Liverpool, GB
- Anne-Kathrin Schmuck
TU Berlin, DE
- Armando Solar-Lezama
MIT – Cambridge, US
- Jiří Srba
Aalborg University, DK
- John G. Thistle
University of Waterloo, CA
- Wolfgang Thomas
RWTH Aachen, DE
- Ufuk Topcu
University of Pennsylvania, US
- Stavros Tripakis
University of California – Berkeley, US
- Martin Zimmermann
Universität des Saarlandes, DE

Report from Dagstuhl Seminar 15062

Domain-Specific Languages

Edited by

Sebastian Erdweg[1], Martin Erwig[2], Richard F. Paige[3], and Eelco Visser[4]

1 **TU Darmstadt, DE,** erdweg@informatik.tu-darmstadt.de
2 **Oregon State University, US,** erwig@eecs.oregonstate.edu
3 **University of York, GB,** richard.paige@york.ac.uk
4 **TU Delft, NL,** e.visser@tudelft.nl

Abstract

This report documents the program and outcomes of Dagstuhl Seminar 15062 "Domain-Specific Languages", which took place February 1–6, 2015. The seminar was motivated on the one hand by the high interest in domain-specific languages in academia and industry and on the other hand by the observation that the community is divided into largely disconnected subdisciplines (e.g., internal, external, visual, model-driven). The seminar included participants across these subdisciplines and included overview talks, technical talks, demos, discussion groups, and an industrial panel. This report collects the abstracts of talks and other activities at the seminar and summarizes the outcomes of the seminar.

Seminar February 1–6, 2015 – http://www.dagstuhl.de/15062

1998 ACM Subject Classification D.2 Software Engineering, D.3 [Programming Languages] Compilers, D.3.1 [Formal Definitions and Theory] Semantics, F.3.2 Semantics of Programming Languages

Keywords and phrases Internal DSLs, External DSLs, Domain-specific modeling, Extensible languages, Language workbenches, Textual/graph-based/visual languages, Language design, Language implementation techniques

Digital Object Identifier 10.4230/DagRep.5.2.26

Edited in cooperation with Daco Harkes

1 Executive Summary

Sebastian Erdweg
Martin Erwig
Richard F. Paige
Eelco Visser

Software systems are the engines of modern information society. Our ability to cope with the increasing complexity of software systems is limited by the programming languages we use to build them. Domain-specific languages (DSLs) successfully address this challenge through linguistic abstraction by providing notation, analysis, verification, optimization, and tooling that are specialized to an application domain. DSLs are already ubiquitous in industrial software development with prominent examples such as HTML, SQL, Make, AppleScript, Matlab, or Simulink.

There is a wide range of methods and techniques for the development of DSLs. Each of these makes different trade-offs that enable different usage scenarios. After the initial design of a DSL, switching to another approach can be very expensive or even impossible. Therefore, the trade-offs and implications of different approaches must be well understood by practitioners from the beginning. However, there is no clear account of what exactly these trade-offs are; neither in industry nor in academia.

The goal of the proposed seminar was to bring together key representatives from the communities that address DSLs from different perspectives: (1) internal DSLs, (2) external DSLs, (3) domain-specific modeling, (4) extensible languages, (5) graph-based languages, and (6) formal semantics. To enable constructive exchange between seminar participants from different communities, the seminar started with one introductory talk per community by a representative. These introductory talks were essential for raising awareness for each other's discipline, the challenges involved, and the problems already solved.

The first day of the seminar was concluded with a poster session. Before the seminar, the organizers invited each participant to prepare and bring a poster that describes their position with respect to the seminar topic. Many participants followed this invitation or used a flip chart for an impromptu presentation. During the poster session, the participants alternated between presenting their own poster and receiving introductions by others. While the seminar did not feature a separate round of introductions at the beginning of the first day, this did not at all hinder discussion and interaction during the talks prior to the poster session. The organizers of this seminar would like to encourage other organizers to consider a poster session as replacement for an introduction round.

After the community and personal introductions on the first day, the second day featured four talks about the "design history" of four existing DSLs. The presenters reported on how the design of their DSLs began, what features turned out to be good, what features turned out to require revision, and how modifications of the design were formed, decided, and implemented. Beyond reporting on their experience, the four talks provided concrete examples of DSLs that could be referred to by all participants during the remainder of the seminar. Subsequently to the design histories, the seminar featured a session on DSL evaluation followed by an industrial panel on industrial DSL requirements.

In the morning of the third day, the participants had the chance to present their latest research results in lightning talks. These were the only talks during the seminar without precise instructions by the seminar organizers. In total, there were eight lightning talks. We observed a high degree of interaction across communities. In the afternoon most participants joined for the excursion: A hike around Schloss Dagstuhl.

Thursday morning was reserved for four talks on DSL type systems. The four talks illustrated different ways of addressing DSL type systems. From a distinguished metalanguage and to automated mechanization to type-system embedding and attribute grammars. The presented work was not mature enough to allow for a meaningful discussion of benefits and disadvantages of the individual approaches. On Thursday afternoon the participants split into two breakout groups on Language Design Patterns and Name Binding. Some participants of the breakout groups decided to continue exchange and discussion after the seminar. The breakout groups were followed by tool demonstrations, where participants could freely move between demos.

Finally, on Friday morning the seminar ended with a session on establishing a research agenda, that is, relevant research questions that should be addressed by the DSL community. Moreover, the participants found that no new dedicated venue for DSLs needs to be established, because there are sufficiently many venues for DSL research available already.

This report collects the abstracts of the talks, and summarises other activities (including a panel and a discussion on a research agenda). The summaries and abstracts suggest outcomes and potential directions for future scientific research.

2 Table of Contents

3 Overview of Talks

3.1 A status update on Ensō

William Cook (University of Texas at Austin, US)

Joint work of Cook, William; van der Storm, Tijs

I gave an update on Ensō, which is a experimental workbench for building and integrating interpreted domain-specific specification languages. Ensō is also implemented in itself.

3.2 Type Systems for the Masses: Deriving Soundness Proofs and Efficient Checkers

Sebastian Erdweg (TU Darmstadt, DE)

Joint work of Erdweg, Sebastian; Grewe, Sylvia; Mezini, Mira

The correct definition and implementation of non-trivial type systems is difficult and requires expert knowledge, which is not available to developers of domain-specific languages and specialized APIs in practice. We propose an approach that automatically derives soundness proofs and efficient and correct algorithms from a single, high-level specification of a type system. Our approach supports the modularization of the specification and the composition of derived proofs and algorithms in order to scale the underlying verification procedure up to real-world languages and to enable specification reuse for common language features.

3.3 Semantics-Driven Language Design

Martin Erwig (Oregon State University, US)

Joint work of Erwig, Martin; Walkingshaw, Eric
Main reference M. Erwig, E. Walkingshaw, "Semantics First! – Rethinking the Language Design Process," in Proc. of the 4th Int'l Conf. on Software Language Engineering (SLE'11), LNCS, Vol. 6940, pp. 243–262, Springer, 2011; pre-print available from author's webpage.
URL http://dx.doi.org/10.1007/978-3-642-28830-2_14
URL http://web.engr.oregonstate.edu/~erwig/papers/abstracts.html#SLE11

The design of languages is still more of an art than an engineering discipline. Although recently tools have been put forward to support the language design process, such as language workbenches, these have mostly focused on a syntactic view of languages. While these tools are quite helpful for the development of parsers and editors, they provide little support for the underlying design of the languages.

Convention dictates that the design of a language begins with its syntax. We argue that early emphasis should be placed instead on the identification of general, compositional semantic domains, and that grounding the design process in semantics leads to languages with more consistent and more extensible syntax. We demonstrate this semantics-driven design process through the design and implementation of a DSL for defining and manipulating calendars, using Haskell as a metalanguage to support this discussion. We emphasize the importance of compositionality in semantics-driven language design.

3.4 Macros and Extensible Languages

Matthew Flatt (University of Utah – Salt Lake City, US)

Macros and extensible languages naturally implement a DSL as an "internal" or "embedded" DSL. Using the QL survey language as an example, we illustrate its implementation in Racket. We start with an S-expression notation, demonstrate techniques for improving syntax errors and adding type checking, show how to make the language look more "external" by putting its in its own module, and show how non-S-expression syntax can be supported with editing support in DrRacket.

3.5 DSL type systems: experiences from using reference attribute grammars

Görel Hedin (Lund University, SE)

Reference attribute grammars support the building of very modular compilers. The technique has been used for implementing complex domain-specific languages with user-defined types and inheritance, like Modelica, as well as general-purpose languages like Java. In this talk, I discuss a number of design strategies that have emerged for using RAGs to implement type systems.

3.6 Graphical DSLs

Steven Kelly (MetaCase – Jyväskylä, FI)

Main reference S. Kelly, J.-P. Tolvanen, "Domain-Specific Modeling: Enabling Full Code Generation," 448 pages, ISBN 978-0-470-03666-2, Wiley, 2008.
URL http://dsmbook.com/

The idea of graphical languages is anathema to many working with textual languages – either from lack of familiarity, familiarity with the all-too-common bad examples, or being burned earlier by a failure in their own attempt to use them. This talk will briefly examine the fundamental and empirical evidence for the utility of graphical languages, and in particular how they differ from textual languages for the language user, the language engineer, and the maker of a language workbench.

3.7 Model-Driven Grant Proposal Engineering

Dimitris Kolovos (University of York, GB)

Joint work of Kolovos, Dimitris; Matragkas Nicholas; Williams, James; Paige, Richard
Main reference D. S. Kolovos, N. Matragkas, J. R. Williams, R. F. Paige, "Model-Driven Grant Proposal Engineering," in Proc. of the 17th Int'l Conf. on Model-Driven Engineering Languages and Systems (MODELS'14), LNCS, Vol. 8767, pp. 420–432, Springer, 2014.
URL http://dx.doi.org/10.1007/978-3-319-11653-2_26

During this talk, we will demonstrate the application of Model-Driven Engineering techniques to support the development of research grant proposals. In particular, we will report on using model-to-text transformation and model validation to enhance productivity and consistency in research proposal writing, and present unanticipated opportunities that were revealed after establishing an MDE infrastructure. We will discuss the types of models and the technologies used, reflect on our experiences, and assess the productivity benefits of our MDE solution through automated analysis of data extracted from the version control repository of a successful grant proposal; our evaluation indicates that the use of MDE techniques improved productivity by at least 58%.

3.8 Resugaring: Lifting Evaluation Sequences through Syntactic Sugar

Shriram Krishnamurthi (Brown University – Providence, US)

Joint work of Pombrio, Justin; Krishnamurthi, Shriram
Main reference J. Pombrio, S. Krishnamurthi, "Resugaring: Lifting Evaluation Sequences through Syntactic Sugar," in Proc. of the 35th ACM SIGPLAN Conf. on Programming Language Design and Implementation (PLDI'14), pp. 361–371, ACM, 2014.
URL http://dx.doi.org/10.1145/2594291.2594319

Syntactic sugar is pervasive in language technology. It is used to shrink the size of a core language; to define domain-specific languages; and even to let programmers extend their language. Unfortunately, syntactic sugar is eliminated by transformation, so the resulting programs become unfamiliar to authors. Thus, it comes at a price: it obscures the relationship between the user's source program and the program being evaluated.

This presentation motivates the problem, explains it through working examples, and outlines the solution.

3.9 Policy languages

Shriram Krishnamurthi (Brown University – Providence, US)

A brief survey of policy languages, especially ones used in security and networking.

3.10 An operational semantics for QL

Peter D. Mosses (Swansea University, GB)

Joint work of Churchill, Martin; Mosses, Peter D.; Sculthorpe, Neil; Torrini, Paolo
Main reference M. Churchill, P. D. Mosses, N. Sculthorpe, P. Torrini, "Reusable components of semantic specifications," Transactions on Aspect-Oriented Software Development Vol. XII, LNCS, Vol. 8989, pp. 132–179, Springer, 2015; pre-print available from author's webpage.
URL http://dx.doi.org/10.1007/978-3-662-46734-3_4
URL http://www.plancomps.org/taosd2015

We present and discuss an abstract syntax and operational semantics for QL using I-MSOS, which is a highly modular variant of structural operational semantics. We also raise some questions about the intentions of the language designer as expressed in the informal language specification available at http://www.languageworkbenches.net/wp-content/uploads/2013/11/Ql.pdf.

3.11 Semantic Modularity for DSLs

Bruno C. d. S. Oliveira (University of Hong Kong, HK)

This talk discusses how to implement modular components for DSLs. The idea is to have collections of language components (abstract syntax and semantics) that can be easily combined and reused in language implementations. Semantic modularity means that such components can, not only, be separately defined, but also be given precise interfaces, be type-checkable and separately compiled. Using Object Algebras, a recently introduced designed pattern, this talk shows how semantic modularity can be achieved in common OO languages such as Java, Scala or C#.

3.12 Evaluating DSLs

Richard F. Paige (University of York, GB)

Introduction to the session on evaluating DSLs. Includes an introduction to a tagging/coding exercise on DSL qualities, and an overview of evaluating DSLs in the context of a course.

3.13 Domain Specific Languages in Practice – An Example

Julia Rubin (MIT – Cambridge, US)

In this talk, we present an example of a real-life domain-specific language (DSL). We use this example to highlight problems faced by organizations that use DSLs in practice:

- difficulties to identify a desirable syntax for the language;
- difficulties to modify, extend and evolve the language;
- difficulties to educate new developers in the use of the language;
- the lack of an ecosystem around DSL management operations, e.g., refactoring and dead-code elimination.

3.14 Through the Looking Glass: A Design History of DSLs for Self-Reconfigurable Robots

Ulrik Pagh Schultz (University of Southern Denmark – Odense, DK)

Joint work of Bordignon, Mirko; Stoy, Kasper; Christensen, Johan
Main reference U. Schultz, M. Bordignon, K. Stoy, "Robust and reversible execution of self-reconfiguration sequences," Robotica, 29:35–57, 2011.
URL http://dx.doi.org/10.1017/S0263574710000664

An overview of my experience in developing domain-specific languages for self-reconfigurable robots, focusing on the main path of development. An analysis of how the development took place is presented, based in part (1) on relating different steps to different sources of inspiration, (2) classifying the kinds of steps that took place and the forces that affected the development, and (3) proposing a list of so-called language design patterns that influenced the development.

3.15 Streams a la Carte: Extensible Pipelines with Object Algebras

Yannis Smaragdakis (University of Athens, GR)

Joint work of Biboudis, Aggelos; Palladinos, Nick; Fourtounis, George; Smaragdakis, Yannis
Main reference A. Biboudis, N. Palladinos, G. Fourtounis, Y. Smaragdakis, "Streams à la carte: Extensible Pipelines with Object Algebras," to appear in Proc. of the 29th Europ. Conf. on Object-Oriented Programming (ECOOP'15); pre-print available from author's webpage.
URL http://cgi.di.uoa.gr/~biboudis/streamalg.pdf

Streaming libraries have become ubiquitous in object-oriented languages, with recent offerings in Java, C#, and Scala. All such libraries, however, suffer in terms of extensibility: there is no way to change the semantics of a streaming pipeline (e.g., to fuse filter operators, to perform computations lazily, to log operations) without changes to the library code. Furthermore, in some languages it is not even possible to add new operators (e.g., a zip operator, in addition to the standard map, filter, etc.) without changing the library.

We address such extensibility shortcomings with a new design for streaming libraries. The architecture underlying this design borrows heavily from Oliveira and Cook's object algebra solution to the expression problem, extended with a design that exposes the push/pull character of the iteration, and an encoding of higher-kinded polymorphism. We apply our design to Java and show that the addition of full extensibility is accompanied by high performance, matching or exceeding that of the original, highly-optimized Java streams library.

3.16 Constraint based language tools

Friedrich Steimann (Fernuniversität in Hagen, DE)

Main reference F. Steimann, "From well-formedness to meaning preservation: model refactoring for almost free," Software & Systems Modeling, 14(1):307–320, 2015.
URL http://dx.doi.org/10.1007/s10270-013-0314-z

It is shown how based a single constraint-based specification of static semantics: well-formedness can be checked; bindings can be computed; errors can be fixed; and refactorings can be performed using a constraint solver.

References

1 Friedrich Steimann: From well-formedness to meaning preservation: model refactoring for almost free. *Software & Systems Modeling* 14:1 (2015) 307–320.
2 Friedrich Steimann, Bastian Ulke: Generic Model Assist. In *Proc. of MoDELS* (2013) 18–34.

3.17 Language workbenches: textual and projectional

Tijs van der Storm (CWI – Amsterdam, NL)

Main reference S. Erdweg, T. van der Storm, M. Völter, M. Boersma, R. Bosman, W. R. Cook, A. Gerritsen, A. Hulshout, S. Kelly, A. Loh, G. D. P. Konat, P. J. Molina, M. Palatnik, R. Pohjonen, E. Schindler, K. Schindler, R. Solmi, V. A. Vergu, E. Visser, K. van der Vlist, G. H. Wachsmuth, J. van der Woning, "The State of the Art in Language Workbenches," in Proc. of the 6th Int'l Conf. on Software Language Engineering (SLE'13), LNCS, Vol. 8225, pp. 197–217, Springer, 2013.
URL http://dx.doi.org/10.1007/978-3-319-02654-1_11

Language workbenches are IDEs consisting of meta languages to build languages and IDEs. In this talk I will sketch a brief history of Language Workbenches, summarize current tools and illustrate how they work internally. In particular I'll distinguish textual language workbenches (based on parsing) from projectional ones (based on structure editing). Concretely, I'll show the questionnaire DSL in the Rascal workbench. Although language workbenches can greatly improve productivity in developing domain-specific languages (DSLs), there are trade-offs in comparison to internal approaches (embedding, macros, etc.) to DSL implementation.

3.18 Interpreter composition

Laurence Tratt (King's College London, GB)

Joint work of Tratt, Laurence; Barrett, Edd; Bolz, Carl Friedrich; Vasudevan, Naveneetha

Language composition is typically crude, and synonymous with the concept of "foreign function interface". In this talk, I show that fine-grained language composition is possible, even down to the level of cross-language variable scoping. By using meta-tracing, we are able to make such compositions perform close to their mono-language peers. This opens up entirely new possibilities for language composition.

3.19 EMF/Ecore-based DSL engineering

Dániel Varró (Budapest University of Technology & Economics, HU)

The Eclipse Modeling Framework (EMF) is an industrial technology to define the abstract syntax of domain-specific modeling languages (DSMLs). After defining the metamodel of the language (in a so-called Ecore model) EMF generates a domain-specific libraries to support model manipulation (interfaces and implementations), model persistency, notifications and commands, undo and redo support as well as a tree based model editor and JUnit test cases demonstrating the use of the API. The generated EMF tool forms the basis of supporting more advanced language and editor features with graphical or textual concrete syntax, model validators, incremental graphical views, model transformations and code generators etc.

3.20 Incremental queries for DSMLs

Dániel Varró (Budapest University of Technology & Economics, HU)

Main reference Z. Ujhelyi, G. Bergmann, Á. Hegedüs, Á. Horváth, B. Izsó, I. Ráth, Z. Szatmári, D. Varró, "EMF-IncQuery: An integrated development environment for live model queries," Science of Computer Programming, 98(1):80–99, 2105.
URL http://dx.doi.org/10.1016/j.scico.2014.01.004

EMF-IncQuery is a framework for defining declarative graph queries over EMF models, and executing them efficiently without manual coding in an imperative programming language such as Java. EMF-IncQuery enables to (1) define model queries using a high level query language (2) execute the queries efficiently and incrementally, with proven scalability for complex queries over large instance models and (3) integrate queries into domain-specific modeling environments to support incremental graphical views, constraint validator, derived features or data binding.

3.21 A Short History of Name Biding in Stratego/XT and Spoofax

Eelco Visser (TU Delft, NL)

Highlights of the development of methods for the implementation and specification of name binding in the Stratego/XT transformation framework and the Spoofax Language Workbench.

3.22 A Theory of Name Resolution

Eelco Visser (TU Delft, NL)

Joint work of Pierre Neron, Andrew Tolmach, Guido Wachsmuth
Main reference P. Neron, A. Tolmach, E. Visser, G. Wachsmuth, "A Theory of Name Resolution," in Proc. of the 24th Europ. Symp. on Programming (ESOP'15), LNCS, Vol. 9032, pp. 205–231, Springer, 2015.
URL http://dx.doi.org/10.1007/978-3-662-46669-8_9

We describe a language-independent theory for name binding and resolution, suitable for programming languages with complex scoping rules including both lexical scoping and modules. We formulate name resolution as a two-stage problem. First a language-independent scope graph is constructed using language-specific rules from an abstract syntax tree. Then references in the scope graph are resolved to corresponding declarations using a language-independent resolution process. We introduce a resolution calculus as a concise, declarative, and language-independent specification of name resolution. We develop a resolution algorithm that is sound and complete with respect to the calculus. Based on the resolution calculus we develop language-independent definitions of α-equivalence and rename refactoring. We illustrate the approach using a small example language with modules. In addition, we show how our approach provides a model for a range of name binding patterns in existing languages.

3.23 Domain-specific type systems

Guido Wachsmuth (TU Delft, NL)

Joint work of Gabriël Konat, Vlad Vergu, Eelco Visser

This talk starts with a short history of specification techniques for typing rules in the Spoofax language workbench. First, we show how typing rules used to be specified with Stratego rewrite rules. In this approach, declarations are first stored in scoped dynamic rules, before declarations and references are renamed using globally unique identifiers, and types are calculated and checked. Next, we introduce new meta-languages for the declarative specification of name binding rules (NaBL) and type systems (TS). We discuss their core concepts, the underlying incremental analysis framework, and show example specifications for language constructs in domain-specific languages such as QL (questionnaires), SDF3 (syntax definition), and Green-Marl (graph processing).

3.24 Cognitive Dimensions for DSL Designers

Eric Walkingshaw (Oregon State University, US)

Joint work of Green, Thomas; Petre, Marian
Main reference T. R. G. Green, M. Petre, "Usability Analysis of Visual Programming Environments: A 'Cognitive Dimensions' Framework," Journal of Visual Languages & Computing, 7(2):131–174, 1996.
URL http://dx.doi.org/10.1006/jvlc.1996.0009

The cognitive dimensions provide a terminology for discussing human factors in programming languages and the tools we use to work with them. This work from the psychology of programming has many applications for DSL designers. They can improve communication by helping us state claims about usability more precisely and to effectively motivate design rationale. They support the design process by providing a way to enumerate common usability issues and by making the tradeoffs between these issues clear. Finally, they support the qualitative evaluation of languages and tools.

4 Demo Session

4.1 MetaEdit+: Industrial Strength Graphical DSLs

Steven Kelly (MetaCase – Jyväskylä, FI)

http://www.metacase.com/

4.2 Epsilon

Dimitris Kolovos (University of York, GB)

http://eclipse.org/epsilon/

4.3 Diagram Editors – Layout and Change Tracking

Sonja Maier (Universität der Bundeswehr – München, DE)

http://www.unibw.de/sonja.maier

In the first part of the demo, we presented a layout framework, which is specifically designed for diagram editors. The key idea is that so-called layout patterns encapsulate certain layout behavior. Its main strengths are that several layout patterns may be applied to a diagram simultaneously, even to diagram parts that overlap and that the layout is continuously maintained during diagram modification.

In the second part of the demo, we presented a framework whose purpose it is to keep track of diagram changes. The key ideas are that all diagram changes are recorded, and that these low-level changes are filtered, aggregated and visualized. The goal is to enable new functionality and to improve usability.

4.4 Incremental model queried in EMF-IncQuery

Dániel Varró (Budapest University of Technology & Economics, HU)

http://eclipse.org/incquery

4.5 Composition of Languages and Notations with MPS

Markus Völter (Völter Ingenieurbüro – Stuttgart, DE)

http://www.jetbrains.com/mps/
http://mbeddr.com/

4.6 The Spoofax Language Workbench

Guido Wachsmuth and Daco Harkes (TU Delft, NL)

http://metaborg.org/spoofax/

5 Working Groups

5.1 Language Design Patterns

We met in a breakout group of ca. 15 participants to discuss whether systematic approaches to (domain-specific or general-purpose) programming language design could be communicated in the form of language design patterns. The breakout group was organized by Tillmann Rendel, Ulrik Pagh Schultz, and Eric Walkingshaw. Main topics of discussion included:

- Is a pattern language appropriate to talk about language design?
- Which granularity of language design should the patterns address?
- How would domain-specific patterns differ from general-purpose patterns?
- What are good examples of language design patterns?
- How should language design patterns be evaluated?
- How can the community work together to find and describe language design patterns?

The breakout session resulted in the creation of a mailing list to further discuss these and other questions related to language design patterns or the communication of systematic approaches to language design in general:

- https://listserv.uni-tuebingen.de/mailman/listinfo/language-design

Some participants consider to work more on language design patterns, maybe co-author a book on the topic and/or organize a workshop to collaboratively work on language design patterns.

6 Open Problems

6.1 Research Agenda Discussion

There was consensus not to have (or revive) a conference dedicated to DSLs specifically, because existing conferences, such as SLE or GPCE, provide a suitable platform for presenting DSL research. It was suggested to create a repository of DSL success stories that researchers could point to (in papers) to emphasize the importance and impact of DSLs. (The DSM community has this already. The DSM Forum site (dsmforum.org) has 50 published case studies and about 100 publications)

As for research questions that deserved to be studied in the future, the following topics were mentioned:

- Languages with visual syntax, heterogeneous syntax
- Domain-specific type systems
- Composition of DSLs (specific to DSLs)
- Debugging of DSLs and DSL programs
- End-user/domain experts as DSL developers, scientists who can program
- Customizability: evolving one DSL into another
- Integration: opening of tools to provide access for others and hooking tools into other tools without standardization
- User studies: formative and summative evaluation
- Study (repositories of) DSL artefacts: like what language constructs are used

7 Panel Discussions

7.1 Industrial Panel

Ralf Lämmel (University of Koblenz-Landau – DE)

Summary compiled by the panel moderator.

7.1.1 Acknowledgment

I am grateful to the panelists (listed below) for contributing to the discussion. The design of the panel is mainly due to the Dagstuhl organizers and specifically Martin Erwig, Sebastian Erdweg, Eelco Visser, and Richard Paige.

7.1.2 Objective of the panel

The overall objective of the panel was to discuss the match between academic research on DSL and the priorities in industry. To this end, the panelists were asked about their thoughts on these questions:

- Do the academic approaches to DSL development work for industry?
- What does industry need to get useful DSLs?
- What is the actual practice for DSL development in your industrial setting?

7.1.3 Format of the panel

Each of the five panelists presented a two-minutes opening remark with the help of one slide. The idea was that each panelist would focus on the single most important issue or problem that should be addressed. Possibly, that issue or problem could be related to the questions listed above. The majority of the 90 minutes of panel time were dedicated to discussions while giving equal time to contributions from the audience and responses from the panelists.

As a follow-up to the panel, if possible, the most important issue or problem should be synthesized. However, such a synthesis turned out to be infeasible, as the panelists had somewhat complementary, albeit related priorities.

7.1.4 The panelists' positions

Five panelists were invited on the grounds of their current or recent involvement in DSL-related projects in industry. The panel was moderated by Ralf Lämmel, Software Languages Team, University of Koblenz-Landau. Here is the list of panelists and a short indication of their 'single most important issue or problem', as communicated to and edited by the moderator:

Hassan Chafi, Oracle The important issue is *arriving at successful lifecycles for DSLs and their development.* More specifically, how can DSLs be gradually introduced into existing projects (lots of valuable code already exists and would benefit from DSLs)? What are interoperability models between DSLs and general purpose languages that either host them or glue them to other parts of the system?

Steven Kelly, MetaCase The important issue is *moving from characters to objects.* This entails, for example: i) Don't rebuild understanding from scratch, store the understood structure. ii) Manipulate, co-operate, link on that level, not as characters. iii) Allow free representation and persistent layout, not forced automatic layout iv) Together, these allow scalability in model and team size and complexity, and seem to be significant factors in cases where good graphical DSLs have been shown to be 5-10x faster than programming.

Oleg Kiselyov, Tohoku University The important issue is this: *What exactly is a DSL?* To quote from an email by the panelist: The issue came up at the Shonan seminar which we organized past May. We were talking about some DSL for HPC, and then someone asked: why do we call it DSL? What exactly is 'domain-specific' here? What is the domain? HPC is so broad that the classification is almost useless. The issue is especially blurred for embedded DSLs. How is that different from a library?

Julia Rubin, MIT The important issue is *maintainability.* This entails specific questions like these: How to keep the language current when new requirements emerge or when new technologies are introduced? Also, how to combine domain specific languages and how to keep the language attractive, e.g., to new employees that are unfamiliar with the language? How to build an ecosystem of solutions around DSLs, e.g., for code analysis, refactoring, debugging, etc.?

Markus Völter, Völter Ingenieurbüro The important issue is *scalability.* There are several dimensions. i) Scaling languages and language ecosystems in terms of complexity: more declarative and analyzable language specifications, fewer specs for more language aspects. ii) Scaling Languages, Domains and Audiences: wider ranges of notations and 'language experiences', more stripped down, user-friendly tools. iii) Scaling some aspects of tools: incremental generation for big systems, incremental type checking for non-trivial type system rules, and others.

Participants

- Lennart Augustsson
Standard Chartered Bank – London, GB
- Hassan Chafi
Oracle Labs – Belmont, US
- William R. Cook
University of Texas – Austin, US
- Sebastian Erdweg
TU Darmstadt, DE
- Martin Erwig
Oregon State University, US
- Matthew Flatt
University of Utah – Salt LakeCity, US
- Andrew Gill
University of Kansas, US
- Daco Harkes
TU Delft, NL
- Görel Hedin
Lund University, SE
- Steven Kelly
MetaCase – Jyväskylä, FI
- Oleg Kiselyov
Tohoku University – Sendai, JP
- Dimitris Kolovos
University of York, GB
- Shriram Krishnamurthi
Brown University, US
- Ralf Lämmel
Universität Koblenz-Landau, DE
- Sonja Maier
Universität der Bundeswehr – München, DE
- Peter D. Mosses
Swansea University, GB
- Bruno C. d. S. Oliveira
University of Hong Kong, HK
- Klaus Ostermann
Universität Tübingen, DE
- Richard F. Paige
University of York, GB
- Tillmann Rendel
Universität Tübingen, DE
- Julia Rubin
MIT – Cambridge, US
- Ulrik Pagh Schultz
University of Southern Denmark – Odense, DK
- Yannis Smaragdakis
National Kapodistrian University of Athens, GR
- Friedrich Steimann
Fernuniversität in Hagen, DE
- Laurence Tratt
King's College London, GB
- Tijs van der Storm
CWI – Amsterdam, NL
- Dániel Varró
Budapest University of Technology & Economics, HU
- Eelco Visser
TU Delft, NL
- Markus Völter
Völter Ingenieurbüro – Stuttgart, DE
- Guido Wachsmuth
TU Delft, NL
- Eric Walkingshaw
Oregon State University, US

Report from Dagstuhl Seminar 15071

Formal Foundations for Networking

Edited by

Nikolaj Bjørner[1], Nate Foster[2], Philip Brighten Godfrey[3], and Pamela Zave[4]

1 **Microsoft Research – Redmond, US,** nbjorner@microsoft.com
2 **Cornell University – Ithaca, US,** jnfoster@cs.cornell.edu
3 **University of Illinois – Urbana-Champaign, US,** pbg@illinois.edu
4 **AT&T Labs Research – Bedminster, US,** pamela@research.att.com

Abstract

This report documents the program and outcomes of Dagstuhl Seminar 15071 "Formal Foundations for Networking." Networking is in the midst of a revolution being driven by rapidly expanding infrastructures and emerging software-defined networking architectures. There is a growing need for tools and methodologies that provide rigorous guarantees about performance, reliability, and security. This seminar brought together leading researchers and practitioners from the fields of formal methods, networking, programming languages, and security, to investigate the task of developing formal foundations for networks.

Seminar Feburary 8–13, 2015 – http://www.dagstuhl.de/15071
1998 ACM Subject Classification C.2.3 Computer-Communication Networks – Network Protocols, D.2.4 Software/Program Verification – Formal Methods, F.3.1 Logics and Meaning of Programs – Specifying and Verifying and Reasoning about Programs
Keywords and phrases Formal methods, logic, middleboxes, model checking, networking, program synthesis, security, software-defined networking, verification
Digital Object Identifier 10.4230/DagRep.5.2.44

1 Executive Summary

Nikolaj Bjørner
Nate Foster
Philip Brighten Godfrey
Pamela Zave

The scale and complexity of computer networks has increased dramatically in recent years, driven by the growth of mobile devices, data centers, and cloud computing; increased concerns about security; and generally more widespread and diverse uses of the Internet. Building and operating a network has become a difficult task, even for the most technologically sophisticated organizations.

To address these needs, the research community has started to develop tools for managing this complexity using programming language and formal methods techniques. These tools use domain-specific languages, temporal logics, satisfiability modulo theories solvers, model checkers, proof assistants, software synthesis, etc. to specify and verify network programs.

Yet despite their importance, tools for programming and reasoning about networks are still in a state of infancy. The programming models supported by major hardware vendors require configurations to be encoded in terms of low-level constructs – e.g., hardware forwarding rules and IP address prefixes. To express richer policies, network operators must incorporate "tribal knowledge" capturing requirements that cut across different customers, service-level agreements, and protocols and can easily lead to contradictions. In addition, networks are rarely static, so operators must deal with updates to configurations and the complications that arise during periods of transition or when unexpected failures occur.

The goal of this seminar was to bring together leading practitioners from the areas of formal methods, networking, programming languages, and security, to exchange ideas about problems and solutions, and begin the task of developing formal foundations for networks. The seminar program was grouped into broad categories addressing the following issues:

- Networking Applications (Akella, Gember-Jacobson, Jayaraman, Rexford). What are the key concerns in enterprise, data center, and wide-area networks today? What kinds of modeling, verification, and property-checking tools are operators deploying? What kinds of scalability challenges are they facing?
- Emerging Areas (Papadimitriou, Rozier). What are the key issues in emerging areas such as crowd-sourced networks and aerospace engineering? Can existing tools be easily adapted to these areas? How can new researchers get involved?
- Distributed Systems (Canini, Cerny). What are some techniques for handling the distributed systems issues that arise in modeling and reasoning about networks? How can we exploit these insights to build practical tools for verifying properties in the presence of replicated state, asynchronous communication, and unexpected failures?
- Domain-Specific Tools (Chemeritskiy, Mahajan, Panda, Rybalchenko, Sagiv). What are the best approaches for verifying properties of real-world networks? How can we incorporate features such as dynamic control programs and mutable state? How can we make these tools scale to networks of realistic size?
- General Tools (Brucker, Ganesh, Guha, Jia, Nelson, Rosenblum, Rybalchenko). There is a rich literature on temporal logics, satisfiability modulo theories checkers, model checkers, proof assistants, Datalog, etc. What are the key differences between these tools and how can they be applied to networks?
- Platforms and Models (Guha, Schlesinger, Reitblatt, Walker, Zave). What is the state-of-the-art in network programming? How can we build compilers and hypervisors that correctly translate from high-level models to low-level implementations?
- Program Synthesis (Buchbinder, Chaudhuri, Cerny, Yuan). Synthesis is a promising approach to building correct software, since programs are generated automatically using a verification tool. What are the best current techniques for using model checkers and satisfiability-modulo theories solvers to generate network configurations, update protocols, and policies?

The seminar comprised four and a half days of presentations, interspersed with discussions, tool demonstrations, and working groups. This report collects the abstracts of the presentations, gives summaries of discussions and working groups, and lists open questions.

2 Table of Contents

3 Overview of Talks

3.1 Abstractions for Network Functions

Aditya Akella (University of Wisconsin – Madison, US)

Joint work of Akella, Aditya; Gember-Jacobson, Aaron
URL http://opennf.cs.wisc.edu

Network functions (NFs), also called middleboxes, are devices that perform custom packet processing functions. With the advent of network functions virtualization (NFV), where NFs are deployed within VMs or as software processes, it has become markedly simpler to bring up of tear down NFs within networks. Furthermore, software defined networking (SDN) is being used to flexibly steer traffic through specific NF instances. The confluence of NFV and SDN opens up the door to two exciting sets of scenarios, namely distributed processing and service chaining. We argue that key attributes of NFs – which differentiate them from traditional routers/switches – impede our ability to realize the full potential of distributed processing and service chaining. We present abstractions and systems that help overcome these impediments.

3.2 Modeling and Conformance Testing of Network Access Policies in HOL-TestGen

Achim D. Brucker (SAP – Karlsruhe, DE)

Modern systems need to comply to large and complex security policies that need to enforced at runtime. This runtime enforcement needs to happen on different levels, e.g., ranging from high level access control models to firewall rules. We present an approach for the modular specification of security policies (e.g., access control policies, firewall policies). Based on this formal model, i.e, the specification, we discuss a model-based test case generation approach that can be used for both testing the correctness of the security infrastructure as well as the conformance of its configuration to a high-level security policy.

3.3 Distributed SDN controllers

Marco Canini (Université catholique de Louvain – Louvain-la-Neuve, BE)

This talk motivates and illustrates the challenges to correctly design and reason about distributed controllers. It then presents Software Transactional Networking, a distributed SDN control plane based on software transactional memory principles that supports concurrent policy updates while ensuring consistent policy composition and high availability.

3.4 Program Synthesis for Network Updates

Pavol Cerny (University of Colorado at Boulder – Boulder, US)

Joint work of McClurg, Jedidiah; Hojjat, Hossein; Cerny, Pavol; Foster, Nate
Main reference J. McClurg, H. Hojjat, P. Cerny, N. Foster, "Efficient Synthesis of Network Updates," arXiv:1403.5843v3 [cs.PL], 2015.
URL http://arxiv.org/abs/1403.5843v3

Software-defined networking (SDN) is revolutionizing the networking industry, but current SDN programming platforms do not provide automated mechanisms for updating global configurations on the fly. Implementing updates by hand is challenging for SDN programmers because networks are distributed systems with hundreds or thousands of interacting nodes. Even if initial and final configurations are correct, naively updating individual nodes can lead to incorrect transient behaviors, including loops, black holes, access control violations, and others. This talk presents an approach for automatically synthesizing updates that are guaranteed to preserve specified properties. We formalize network updates as a distributed programming problem and develop a synthesis algorithm that uses counterexample-guided search and incremental model checking to dramatically improve performance. We describe our prototype implementation, and present results from experiments on real-world topologies and properties demonstrating that our tool scales to updates involving thousands of nodes in a few seconds.

3.5 Program Synthesis

Swarat Chaudhuri (Rice University – Houston, US)

The field of program synthesis envisions a software design process where programmers write partial, nondeterministic specifications of programming tasks, and powerful algorithms are used to find correct implementations of these specifications. In this talk, I will describe some of my recent work in this area and its potential application in the domain of networks. Specific topics will include:

1. Lambda^2, a new algorithm for example-driven synthesis of higher-order functional programs.
2. Pliny, a new project on program synthesis and repair that utilizes knowledge implicit in large pre-existing corpora of programs.

3.6 VERMONT (Verifying Network Monitor)

Evgeny Chemeritskiy (Applied Research Center for Computer Networks – Moscow, RU)

VERifying MONiTor (VERMONT) is a software toolset for checking the consistency of network configurations with formally specified invariants of Packet Forwarding Policies (PFP).

Correct and safe management of networks is a very hard task. Every time the current load of flow tables should satisfy certain requirements. Some packets have to reach their destination, whereas some other packets have to be dropped. Certain switches are forbidden for some packets, whereas some other switches have to be obligatorily traversed. Loops are not allowed. These and some other requirements constitute a PFP. One of the aims of network engineering is to provide such a loading of switches with forwarding rules as to guarantee compliance with the PFP. VERMONT provides some automation to the solution of this task.

VERMONT can be installed in line with the control plane. It observes state changes of a network by intercepting messages sent by switches to the controller and command sent by the controller to switches. It builds an adequate formal model of a whole network and checks every event, such as installation, deletion, or modification of rules, port and switch up and down events, against a set formal requirements of PFP. Before a network update command is sent to a switch VERMONT anticipates the result of its execution and checks whether a new state of network satisfies all requirements of PFP. If this is the case then the command is delivered to the corresponding switch. Upon detecting a violation of PFP VERMONT blocks the change, alerts a network administrator, and gives some additional information to elucidate a possible source of an error.

VERMONT has a wide area of applications. It can be attached to a SDN controller just to check basic safety properties (the absence of loops, black-holes, etc) of the network managed by the controller. VERMONT may be also cooperated with software units (like FlowVisor) that aggregate several controllers. In this case VERMONT checks the compatibility of PFPs implemented by these controllers. This toolset can be used as a fully automatic safeguard for every software application which implements certain PFP on a SDN controller.

3.7 The Impact of Community Structure on SAT Solver Performance

Vijay Ganesh (University of Waterloo – Waterloo, CA)

Modern CDCL SAT solvers routinely solve very large industrial SAT instances in relatively short periods of time. It is clear that these solvers somehow exploit the structure of real-world instances. However, to-date there have been few results that precisely characterise this structure. In this paper, we provide evidence that the community structure of real-world SAT instances is correlated with the running time of CDCL SAT solvers. It has been known for some time that real-world SAT instances, viewed as graphs, have natural communities in them. A community is a sub-graph of the graph of a SAT instance, such that this sub-graph has more internal edges than outgoing to the rest of the graph. The community structure of a graph is often characterised by a quality metric called Q. Intuitively, a graph with high-quality community structure (high Q) is easily separable into smaller communities, while the one with low Q is not. We provide three results based on empirical data which show that community structure of real-world industrial instances is a better predictor of the running time of CDCL solvers than other commonly considered factors such as variables and clauses. First, we show that there is a strong correlation between the Q value and Literal Block Distance metric of quality of conflict clauses used in clause-deletion policies in Glucose-like solvers. Second, using regression analysis, we show that the the number of communities and the Q value of the graph of real-world SAT instances is more predictive of the running time of CDCL solvers than traditional metrics like number of variables or clauses. Finally, we show that randomly-generated SAT instances with $0.05 \leq Q \leq 0.13$ are dramatically harder to solve for CDCL solvers than otherwise.

3.8 Machine-Verified Network Controllers

Arjun Guha (University of Massachusetts at Amherst – Amherst, US)

Joint work of Guha, Arjun; Reitblatt, Mark; Foster, Nate

In many areas of computing, techniques ranging from testing to formal modeling to full-blown verification have been successfully used to help programmers build reliable systems. But although networks are critical infrastructure, they have largely resisted analysis using formal techniques. Software-defined networking (SDN) is a new network architecture that has the potential to provide a foundation for network reasoning, by standardizing the interfaces used to express network programs and giving them a precise semantics.

This talk describes the design and implementation of the first machine-verified SDN controller. Starting from the foundations, we develop a detailed operational model for OpenFlow (the most popular SDN platform) and formalize it in the Coq proof assistant. We then use this model to develop a verified compiler and runtime system for a high-level network programming language. We identify bugs in existing languages and tools built without formal foundations, and prove that these bugs are absent from our system. Finally, we describe our prototype implementation and our experiences using it to build practical applications.

3.9 Management Plane Analytics

Aaron Gember-Jacobson (University of Wisconsin at Madison – Madison, US)

While it is generally held that network management is tedious and error-prone, it is not well understood which specific management practices increase the risk of failures. To address this gap, we propose a management plane analytics framework that an organization can use to: (1) infer which management practices can impact network health, and (2) develop a predictive model of health, based on observed practices, to improve network management. We overcome the challenges of noisy data and insufficient samples by adopting a “big data” approach, in which we synthesize data from many networks and build predictive models over aggregates. Our current models can predict network health with an accuracy of 76–89%.

3.10 Network Verification in Microsoft Azure

Karthick Jayaraman (Microsoft Research – Redmond, US)

Network verification is problem of critical importance for managing large-scale data centers. In this talk, we will describe a system that we built to perform run-time monitoring of the data plane of Azure data center network, and assure its correctness and consistency. Our system uses a tool called SecGuru that leverages Z3, a SMT solver, to prove properties about access-control lists and routing tables. Our objective is to prove local correctness of devices, and most of the global properties can be decomposed to local properties. Our system is currently deployed, and its use has led to a positive measurable impact in assuring reliability of the network.

3.11 Verifying Network Protocols using Declarative Networking

Limin Jia (Carnegie Mellon University – Pittsburgh, US)

In this talk, I will first give a short tutorial on Declarative Networks. I will explain the syntax and semantics of NDlog, a network declarative language, and how to program in NDlog. In the second part of my talk, I will present our recent work on verifying secure network protocols using declarative networks. NDlog serves as a unified specification both for generating low-level implementations for empirical evaluation and for verifying formal properties. I will present a program logic that we developed for deriving invariant properties of NDlog programs that execute in an adversarial environment. These invariant properties are safety properties on traces that are expressive enough to express origin and path authenticity properties of networks.

3.12 Configuration verification: A missing link toward fully verified networks

Ratul Mahajan (Microsoft Research – Redmond, US)

What does it mean to be running a fully verified network? One perspective is that a verified network is one that is running verified 1) hardware; 2) software; 3) data plane; and 4) configuration. Given prior work in first three domains, the missing link is configuration verification. I'll describe our recent work on a logic-based approach for verifying the configuration of large, traditional (non-SDN) networks. I'll also describe how our approach can help bridge traditional and SDN networks.

3.13 SAT Applications Tutorial (plus a pinch of Margrave)

Tim Nelson (Brown University, Providence, US)

This talk provides an overview of using SAT-solvers for network analysis. For perspective, it includes a discussion of our prior work on configuration analysis in Margrave, and shows how rich questions about configuration behavior can be answered via SAT-solvers. It also shows examples of how to use Margrave-style analyses to resolve questions about the stateful behavior a configuration describes.

3.14 Verifying Mutable Datapaths

Aurojit Panda (University of California, Berkeley – Berkeley, US)

Recent work has made great progress in verifying the correctness of forwarding tables in networks. However, these approaches cannot be used to verify networks containing middleboxes such as caches and firewalls whose forwarding behavior depends on previously observed traffic. We explore how to verify reachability properties for networks that include such "mutable datapath" elements. Our work leverages recent advances in SMT solvers, and the main challenge lies in scaling the approach to handle large and complicated networks. While the straightforward application of model checking to this problem can only handle very small networks (if at all), our approach can verify invariants on networks containing 30,000 middleboxes in a few minutes.

3.15 Software-Defined Crowd-Shared Networks

Panagiotis Papadimitriou (Leibniz Universität Hannover – Hannover, DE)

Recently there has been an increasing interest in providing wider access to Internet. One opportunity for sharing Internet access in residential areas is to exploit the spare capacity in home broadband connections. Since such crowd-shared networks entail considerable configuration overhead both for home network users and ISPs, we leverage on SDN to outsource their configuration and management to third parties. Enabling a third party to federate wireless home networks can reduce the expenditure for network operators and enable new economic models for generating revenue from currently underutilized infrastructures.

3.16 Proof-Carrying Network Code

Mark Reitblatt (Cornell University – Ithaca, US)

Joint work of Reitblatt, Mark; Foster, Nate; Kozen, Dexter; Mamouras, Konstantinos; Silva, Alexandra

In many network settings, multiple parties interact to form a network program. For example, in PANE, different network users and applications are delegated with "shares" of the network which they can in turn control or delegate. To protect parties from one another, these shared networks usually enforce a rigid mode of interaction. Instead, we propose a mechanism (Proof Carrying Network Code) for a more flexible, policy-based interaction. Principals are able to specify requirements for other parties to meet when handling their traffic, and the other parties are required to show that their program meets these requirements. Analogous to Proof Carrying Code, in which binaries carry verifiable proofs of their own safety, Proof Carrying Network Code comes with a verifiable certificate that a certain policy is satisfied. In this talk I'll present our initial design and implementation of the framework, and discuss future directions for exploration.

3.17 SDN Applications

Jennifer Rexford (Princeton University – Princeton, US)

Software-Defined Networking (SDN) is changing how networks are designed and managed. By offering a simple, open interface to the data plane, SDN enables data-plane verification techniques to check that a network-wide snapshot of the forwarding rules satisfies key network invariants. In this presentation, we discuss several example SDN applications, with an eye toward new research opportunities in verification and programming languages. We consider toy applications like MAC learning, stateful firewalls, and server load balancing to illustrate subtle bugs that can be hard to prevent or detect. We also survey several prominent commercial SDN applications (e.g., wide-area traffic engineering, network virtualization for multi-tenant data centers, and traffic steering through middleboxes) to illustrate further opportunities for research in language abstractions and verification techniques. Example challenges include (i) performing multiple tasks simultaneously with a single set of rules, (ii) policies (and network invariants) that change over time, (iii) uncertainty in the ordering of events, (iv) limitations on rule-table space, (v) non-deterministic applications, and (vi) interactions with other network protocols (such as TCP and BGP).

3.18 A Brief Look at Probabilistic Model Checking

David Rosenblum (National University of Singapore – Singapore, SG)

In this talk I will give an overview of the motivation and main ideas behind probabilistic model checking, and I will discuss recent research on the use of perturbation theory as a way of dealing with uncertainty about probability parameters in stochastic models and the effect of perturbations on verification results. I will conclude with a discussion of some of the challenges in applying probabilistic model checking to problems in networking.

3.19 Formal Methods Challenge: *E*fficient Reconfigurable *C*ockpit Design and Fleet *O*perations using Software Intensive, *N*etworked, and Wireless-Enabled Architecture (ECON)

Kristin Yvonne Rozier (University of Cincinnati – Cincinnati, US)

We are at the dawn of a new generation of aircraft! Current aircraft are consistently built over-weight and over-budget and suffer from limitations that cannot carry forward into our increasingly digital age. In order to meet the demands placed on future aircraft, we need to design a new type of *net-enabled* aircraft by reducing aircraft weight, increasing automation and modularity, moving from hardware to software systems, moving from on-board/aircraft-based systems to cloud/fleet-based systems, and thereby facilitating easier and more responsive maintenance and system health management procedures.

For example, the wiring alone on the A-380 weighs approximately six tons, while each ton of weight in the aircraft infrastructure costs an estimated $7.2 billion in fuel across a fleet in the U.S. each year. We ask the question: can we replace some of these wires with wireless systems? Can we add wireless backup systems for wired systems that don't currently have backups due to weight constraints? But then how do we design these new hybrid systems in a way that allows us to rigorously reason about their safety, reliability, availability, and security? For another example, can we replace heavy, customized, and therefore hard-to-maintain cockpit systems by lighter, more modular, alternatives using wireless, software, or cloud-based technologies? We have not solved the formal verification problem for current cockpits; how will we scale to net-enabled cockpits?

We must reason about an aircraft as a network of avionics sub-systems, about a fleet as a network of aircraft, and about both of these being located in the cloud. What restrictions do we need to make to enable formal verification, from design-time to runtime? Join this NASA-lead team of government, academia, and industry experts as we attempt to answer the ultimate question: how can we design and verify a new class of net-enabled aircraft that are *lighter*, *cheaper*, and *safer* than ever before?

3.20 Constrained Horn Clauses for Software, Hardware, Network Verification and Synthesis

Andrey Rybalchenko (Microsoft Research UK – Cambridge, GB)

We show how models of software, hardware, and networks can be represented using logical formulas in a way that facilitates deductive reasoning about them using verification conditions. These verification conditions appear in form of constrained Horn clauses and can be solved efficiently using state-of-the-art automated deduction techniques.

3.21 Reasoning about Stateful Networks

Mooly Sagiv (Tel Aviv University – Tel Aviv, IL)

We describe techniques for verifying properties of networks of middleboxes.

3.22 Protocol-independent Packet Processing

Cole Schlesinger (Princeton University – Princeton, US)

The OpenFlow protocol initially provided a compelling but simple abstraction for network hardware: A switch is a match-action table configured by a controller. Later versions recognized that switch architectures comprise pipelines of tables, each table capable of operating on a fixed set of header fields, and fine-tuned the interface to allow for control of the pipeline.

But in some switches, the packet parser and pipeline are not fixed; rather, they can be configured in advance to extract and operate on arbitrary header fields. Architectures like Intel's FlexPipe and Cisco's Doppler already have this functionality under the hood – after all, it would be impractical to fabricate a new ASIC each time a new protocol is adopted. And new hardware has been proposed to expose parser and pipeline configuration directly to the controller in an SDN setting.

This talk explores the capabilities of reconfigurable switches as well as emerging languages and compilation techniques for guiding reconfiguration.

3.23 Decentralizing SDN Policies

Sharon Shoham Buchbinder (Academic College of Tel Aviv-Yaffo – Tel Aviv, IL)

Joint work of Padon, Oded; Immerman, Neil; Karbyshev, Aleksandr; Lahav, Ori; Sagiv, Mooly

Software-defined networking (SDN) is a new paradigm for operating and managing computer networks. SDN enables logically-centralized control over network devices through a "controller" – software that operates independently of the network hardware. Network operators can run both in-house and third-party SDN programs on top of the controller, e.g., to specify routing and access control policies.

In practice, having the controller handle events limits the network scalability. Therefore, the feasibility of SDN depends on the ability to efficiently decentralize network event-handling by installing forwarding rules on the switches. However, installing a rule too early or too late may lead to incorrect behavior, e.g., (1) packets may be forwarded to the wrong destination or incorrectly dropped; (2) packets handled by the switch may hide vital information from the controller, leading to incorrect forwarding behavior. The second issue is subtle and sometimes missed even by experienced programmers.

The contributions of this paper are two fold. First, we formalize the correctness and optimality requirements for decentralizing network policies. Second, we identify a useful class of network policies which permits automatic synthesis of a controller which performs optimal forwarding rule installation.

3.24 Online Data Center Modeling

Robert Soulé (Università della Svizzera italiana – Lugano, CH)

Modern enterprise data centers are crucial infrastructure, but are also complex, dynamic, highly networked systems. As such, their capacity, performance, behavior, and failure modes are difficult to predict, understand, and plan for. A major reason for this complexity is that the many conceptual layers involved in an enterprise data center (physical network connectivity, physical and virtual machines, link layers, VLANs, routing, service-oriented architectures, application deployment, etc.) are managed today by tools and techniques which focus on only one or a few layers.

Rather than focusing on mechanisms to control and manage subsets of a data center, we will create a common data model and representation for the state of an operational data center which can be populated and driven by logs, traces, and configuration information; queried by operators to determine global properties of the system (such as traffic matrices), and drive online workload-driven simulations to explore the effects of configuration changes. Our goal is to provide a shared substrate for diverse data center management functionality, analogous to the way that the relational model of databases provided a common substrate for tabular data.

3.25 Implementing Path Queries

David Walker (Princeton University – Princeton, US)

Joint work of Narayana, Srinivas; Arashloo, Mina; Rexford, Jennifer; Walker, David

Decades of experience suggests that complex programming systems should be implemented using a stack of compiler intermediate languages. Indeed, virtually all modern compilers use such an architecture, and we see no reason that network programming systems should deviate from this well-established trend. In this talk, we illustrate this idea via a case study involving the implementation of an expressive new query language for software-defined networks. This query language allows users to measure the flow of packets along user-specified paths in a network. It can help diagnose link congestion, implement traffic engineering algorithms, or mitigate DDoS attacks. More specifically, the language allows network operators or control software to issue queries specified as regular expressions over predicates on packet locations and header values. It also uses SQL-like "groupby" constructs to aggregate results anywhere along a path. A run-time system compiles the high-level queries into a deterministic finite automaton, which is encoded in the NetKAT intermediate language. The NetKAT program is then compiled into a set of OpenFlow rules. Finally those OpenFlow rules are distributed to a set of switches. As packets flow through the network, switches stamp packets with automaton states, which tracks the packets' progress towards fulfilling a query. Only when a query is satisfied are packets counted, sampled, or sent to collectors for further analysis. By processing queries inline in the data plane, users "pay as they go" as data-collection overhead is limited to only those packets that satisfy the query. We have implemented our system on top of Pyretic, an open source SDN platform, and evaluated its performance on a campus topology.

3.26 Programming Network Policies by Examples

Yifei Yuan (University of Pennsylvania – Philadelphia, PA)

The emergence of programmable interfaces to network controllers offers network operators the flexibility to implement a variety of policies. We propose NetEgg, a programming framework that allows a network operator to specify the desired functionality using example behaviors. Our synthesis algorithm automatically infers the state that needs to be maintained to exhibit the desired behaviors along with the rules for processing network packets and updating the state. We report on an initial prototype of NetEgg. Our experiments evaluate the proposed framework based on the number of examples needed to specify a variety of policies considered in the literature, the computational requirements of the synthesis algorithm to translate these examples to programs, and the overhead introduced by the generated implementation for processing packets. Our results show that NetEgg can generate implementations that are consistent with the example behaviors, and have performance comparable to equivalent imperative implementations.

3.27 A formal model of a cloud, featuring modular, expressive, re-usable abstractions

Pamela Zave (AT&T Labs Research – Bedminster, US)

Today's Internet has many layers at different scopes and levels, instead of the fixed 5 or 7 prescribed by classic reference models. This talk presents a new architectural model suitable for today's Internet, in which each layer instantiates and constraints the same template of mutable state to carry out its particular purposes. The model is illustrated with four formally-specified layers implementing a realistic large-scale cloud with rich functionality including middlebox policies, live VM migration, tenant isolation, and multiple data centers. The model supports formal, modular verification of a wide variety of properties including reachability, security, consistent updates, and safe optimizations.

4 Working Groups

4.1 Beyond Verification

This working group was motivated by the prevalence of discussions focused on verification during the early days of the seminar. Many people felt that the goal of users such as network operators is not verification, but rather proper, predictable functioning of a network. To achieve the real goal, it is necessary to reason about less well-understood "ility" properties, as well as well-understood logical properties. Reasoning about "ility" properties may require different frameworks from logical properties. Even for logical properties, verification is just one type of formal reasoning, and one of the most expensive. Consequently, we decided that "before and besides verification" describes our concerns more accurately.

The most important research questions concern formal descriptions: What is a basic set of formal descriptions that will cover all needs? What are the relationships among these descriptions? Given such a set of descriptions, the group produced the following list of (non-verification) goals or ways of reasoning about them:

- safe optimization
- semantic-difference analysis
- change analysis
- root-cause analysis
- test-case generation
- simulation
- performance analysis
- understanding design principles
- understanding trade-offs in a design space

We agreed that the distinction between networks and distributed systems is institutional only. That being the case, we wondered about what part of networking is missing from distributed systems, to the disadvantage of the work? We also talked about the well-known "design principle" papers, namely "End-to-end arguments in system design" (Saltzer, Reed, and Clark) and "The design philosophy of the DARPA Internet protocols" (Clark). If we had up-to-date formal descriptions as a basis for reasoning, how would these principles hold up? Or would they be replaced by others?

4.2 Abstractions for Capturing Intent

This working group discussed the design of higher-level abstractions that capture operator intent. The discussion was motivated by the observation that many problems in networking stem from the fact that network software is expressed at a low-level of abstraction – one that does not necessarily match intent.

The first part of the discussion enumerated aspects of network behavior that are important to operators. Starting with existing configuration languages, the group noted a focus on two features: (i) forwarding paths and (ii) security policies. However, precisely capturing security policies often requires describing both what is allowed and what is not allowed, which is difficult using mechanisms such as access control lists. Similarly, working with concrete forwarding paths can create a mismatch with intent – the operator may specify concrete paths when actually any of a much large set of paths would suffice. Performance is another important aspect of intent, but is not typically captured in low-level configuration languages. Operators often want to optimize for objectives such as low latency or congestion. Finally, the group noted that service-level agreements often state resilience to failures and incorporate temporal notions.

The second part of the discussion investigated concrete mechanisms for expressing intent. Some participants suggested reverse engineering intent from legacy deployments. The idea here would be to extract information from low-level configurations and put them into higher-level analytical models that are designed to capture intent. Other participants highlighted languages based on path formalisms such as NetKAT and FlowLog, languages that incorporate bandwidth guarantees such as Merlin, and platforms such as Click and P4. The group closed with a more speculative discussion on whether networks will be built using domain-specific software stacks or Turing-complete languages.

4.3 Network Verification: What's New?

Most current network verification tools are either based on domain-specific decision procedures (e.g., Header Space Analysis, VeriFlow, etc.) or encode networking problems as inputs for existing tools (e.g., Alloy, Z3, etc.). This working group explored the question: what, if anything, is new in network verification?

Many participants felt that although it might be possible in principle to encode networking problems as inputs for other solvers, there may be value in developing domain-specific decision procedures – typically there is domain knowledge that can be exploited to improve precision and performance. However, others noted that domain knowledge can sometimes be incorporated into the encoding itself, although this makes the encoding more complicated and susceptible to bugs. The group also discussed properties and noted that building on existing tools may make it difficult or impossible to express certain features such as quantitative constraints or probabilistic bounds. Moreover, there may be symmetries based on topology that can be exploited to obtain parametric and compositional reasoning, which is quite powerful. Another pragmatic challenge stems from the fact that network behavior is defined both by a program and a configuration. Almost all participants agreed that having open repositories of benchmarks and challenge problems would help focus community effort on the right problems. The group closed with a discussion of testing tools and security properties such availability, confidentiality, and anonymity.

4.4 Building Decomposable Control Abstractions

The question posed to this working group was: What abstractions can allow us to decompose a logically-centralized controller into modules? The discussion identified that controller modularity might come in several flavors: a "horizontal decomposition" into a collection of distributed controllers, or a "vertical decomposition" pushing certain elements of control downward into the data plane. These decompositions might be combined. For example, a logically-centralized control program could be fragmented across a central controller, end-hosts, and software in the data plane of switches, taking advantage of different information and reaction times available to each type of component.

Why would one want to decompose a controller? The participants saw many potential opportunities: eliminating reactivity at the controller, avoiding latency to the controller, offloading work from the hypervisor (e.g. via vertical decomposition), measurement and monitoring, real-time congestion avoidance inside network switches or end-hosts (as in several recent papers), and flexible placement and relocation of middlebox functions. Some application-level processing such as MapReduce aggregation could even be offloaded to network devices for greater efficiency.[1]

Discussion participants pointed out a growing diversity of hardware and software providing substrates for network control. For example, Facebook's "6-pack" is a modular switch that runs with split control, using a local controller on the switch that communicates with a centralized controller. Several SDN and traditional switch vendors have released whitebox hardware enabling computation directly on switches, and Open Network Linux provides an operating system for such gear. The rise of "disaggregation" and rack-sized machines will provide new opportunities for control of their networks. Even the OpenFlow protocol, typically thought of as programming dumb switches with simple instructions, can encode perhaps surprisingly complex behavior in the forwarding plane.[2] Bianchi et al.[3] propose to further enrich OpenFlow with a minimal amount of stateful processing – taking a step in a finite state machine – to build functionality such as firewalls and load balancers. All these provide interesting new options for factoring a control application.

But hardware and protocols by themselves are not enough. The question originally posed to the working group – that of the right abstractions for decomposition – is an open problem. Several recent works provide some steps: Kandoo (Yeganeh and Ganjali, HotSDN 2012) decomposes control into two layers, global centralized control and local control at each switch, with applications such as network measurement. Recursive SDN (work by McCauley, Panda, Liu, Kazemkhani, Koponen, Raghavan, and Shenker) builds a hierarchy of SDN controllers, tackling problems of network repair and traffic engineering. Beehive (also Yeganeh and Ganjali, HotNets 2014) provides a programming abstraction similar to a centralized controller, and automatically decomposes the application into distributed execution across multiple controllers. The participants in the working group speculated about further sources of inspiration for solutions: OSPF areas, FlowLog, and work on automated parallelization of programs. From the latter research area, the key take-away is that the data dependencies are the trickiest aspect of the problem to grapple with.

5 Open Problems

Several open problems emerged from discussions throughout the seminar:

- What are the most important verification challenges in networking from the perspective of practitioners? Can the community assemble repositories of benchmarks and challenge problems to focus attention on the right issues?
- What are the right high-level language abstractions for programming networks, and what guarantees could we expect a compiler to provide – reachability, security, or even properties

[1] Luo Mai, Lukas Rupprecht, Abdul Alim, Paolo Costa, Matteo Migliavacca, Peter Pietzuch, and Alexander Wolf, "NetAgg: Using Middleboxes for Application-specific On-path Aggregation in Data Centres", in ACM CoNEXT 2014.

[2] Michael Borokhovich, Liron Schiff, and Stefan Schmid, "Reclaiming the Brain: Useful OpenFlow Functions in the Data Plane", in ACM HotNets 2014.

[3] Giuseppe Bianchi, Marco Bonola, Antonio Capone, and Carmelo Cascone, "OpenState: Programming Platform-independent Stateful OpenFlow Applications Inside the Switch", in ACM SIGCOMM Computer Communication Review, April 2014.

as detailed as the correct use of cryptography? Can these abstractions streamline some of the distributed aspects of network programming?

- What are appropriate formalisms for expressing and automatically verifying network properties? Reachability properties are ubiquitous in networking, being useful for characterizing connectivity, routing, and access control policies. But operators also care about quantitative properties such as latency, congestion, and resilience.
- What is the right division of labor between static and dynamic verification? Static tools find errors earlier in the development process and do not impose a run-time overhead. But dynamic techniques can be simpler and are able to detect bugs in control software and hardware errors.

Participants

- Aditya Akella
University of Wisconsin – Madison, US
- Nikolaj Bjørner
Microsoft Res. – Redmond, US
- Achim D. Brucker
SAP Research – Karlsruhe, DE
- Marco Canini
University of Louvain, BE
- Pavol Cerny
Univ. of Colorado – Boulder, US
- Swarat Chaudhuri
Rice University, US
- Evgeny Chemeritskiy
ARCCN – Moscow Region, RU
- Shiu-Kai Chin
Syracuse University, US
- Bryan Ford
Yale University, US
- Nate Foster
Cornell University – Ithaca, US
- Vijay Ganesh
University of Waterloo, CA
- Aaron Gember-Jacobson
University of Wisconsin – Madison, US
- Philip Brighten Godfrey
Univ. of Illinois at Urbana, US
- Arjun Guha
University of Massachusetts – Amherst, US
- Arie Gurfinkel
Carnegie Mellon University – Pittsburgh, US
- Karthick Jayaraman
Microsoft Res. – Redmond, US
- Limin Jia
Carnegie Mellon University – Pittsburgh, US
- Ethan Katz-Bassett
Univ. of Southern California – Los Angeles, US
- Shriram Krishnamurthi
Brown University, US
- Ori Lahav
MPI-SWS – Kaiserslautern, DE
- Nuno Lopes
Microsoft Research UK – Cambridge, GB
- Ratul Mahajan
Microsoft Res. – Redmond, US
- Tim Nelson
Brown University, US
- Aurojit Panda
University of California – Berkeley, US
- Panagiotis Papadimitriou
Leibniz Univ. Hannover, DE
- Mark Reitblatt
Cornell University – Ithaca, US
- Jennifer Rexford
Princeton University, US
- Timothy Roscoe
ETH Zürich, CH
- David Rosenblum
National Univ. of Singapore, SG
- Kristin Yvonne Rozier
University of Cincinnati, US
- Andrey Rybalchenko
Microsoft Research UK – Cambridge, GB
- Mooly Sagiv
Tel Aviv University, IL
- Cole Schlesinger
Princeton University, US
- Stefan Schmid
TU Berlin, DE
- Sharon Shoham Buchbinder
Academic College of Tel Aviv, IL
- Robert Soulé
University of Lugano, CH
- David Walker
Princeton University, US
- Alexander L. Wolf
Imperial College London, GB
- Burkhart Wolff
University Paris South, FR
- Yifei Yuan
University of Pennsylvania, US
- Vladimir Zakharov
Moscow State University, RU
- Pamela Zave
AT&T Labs Research – Bedminster, US

Report from Dagstuhl Seminar 15072

Distributed Cloud Computing

Edited by

Yvonne Coady[1], James Kempf[2], Rick McGeer[3], and Stefan Schmid[4]

1 **University of Victoria, CA,** `ycoady@cs.ubc.ca`
2 **Ericsson – San Jose, US,** `james.kempf@ericsson.com`
3 **HP Labs – Palo Alto, US,** `rick@mcgeer.com`
4 **TU Berlin, DE,** `stefan.schmid@tu-berlin.de`

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 15072 "Distributed Cloud Computing". A distributed cloud connecting multiple, geographically distributed and smaller datacenters, can be an attractive alternative to today's massive, centralized datacenters. A distributed cloud can reduce communication overheads, costs, and latency's by offering nearby computation and storage resources. Better data locality can also improve privacy. In this seminar, we revisit the vision of distributed cloud computing, and identify different use cases as well as research challenges.

Seminar February 8–11, 2015 – http://www.dagstuhl.de/15072
1998 ACM Subject Classification C.2.4 [Computer-Communication Networks] Distributed Systems
Keywords and phrases Distributed Cloud Computing, Distributed Systems
Digital Object Identifier 10.4230/DagRep.5.2.64
Edited in cooperation with Oliver Hohlfeld

1 Executive Summary

Yvonne Coady
James Kempf
Rick McGeer
Stefan Schmid

The Dagstuhl Seminar on Distributed Cloud Computing was held Feb. 8–11, 2015. 22 researchers attended the multidisciplinary seminar from the areas of networking, cloud computing, distributed systems, operations research, security, and system administration. In contrast with the centralized cloud deployment model where applications are restricted to a single mega-data center at some network distance from the customers, in the distributed cloud deployment model, many smaller data centers are deployed closer to customers to supplement or augment the larger mega-data centers, and the smaller data centers are managed as one pooled resource. Two administrative models of a distributed cloud are common today: the integrated model where a single administrative entity controls all the data centers and the federated model where multiple administrative entities control the data centers and users authenticate for resource access using a federated identity management system. Over the course of the 3 day seminar, 15 presentations were given on various aspects of distributed cloud or the disciplinary areas relevant to distributed cloud. The seminar shared two talks with the concurrent seminar on Foundations of Networking and attended one of the Foundations of Networking Talks. Taking the presentations as input, the seminar then broke into three groups to discuss a research agenda for distributed cloud. The groups were requested to come up with 3 questions in their particular area (distributed systems, programming models, and cloud) and two for the other two groups. At the end of the seminar, the group discussed forming a research community around distributed cloud with an annual conference. Currently, a workshop on distributed cloud is held annually, called DCC (for Distributed Cloud Computing). This year's workshop will be held in conjunction with SIGMETRICS in Portland, Oregon in June. Slides, abstracts of the talks and reports from the breakout groups are available in the Dagstuhl content management web site. An extended version of this report appeared in the April 2015 issue of ACM SIGCOMM Computer Communication Review [1].

References

1 Yvonne Coady, Oliver Hohlfeld, James Kempf, Rick McGeer, and Stefan Schmid. Distributed cloud computing: Applications, status quo, and challenges. *SIGCOMM Comput. Commun. Rev.*, 45(2):38–43, April 2015. DOI: 10.1145/2766330.2766337

2 Table of Contents

3 Overview of Talks

3.1 Scalable consistency – all the way to the edge!

Annette Bieniusa (TU Kaiserslautern, DE)

Joint work of Zawirski, Marek; Bieniusa, Annette; Preguica, Nuno; Duarte, Sergio; Balegas, Valter; Shapiro, Marc
Main reference M. Zawirski, A. Bieniusa, V. Balegas, S. Duarte, C. Baquero, M. Shapiro, N. Pregiuça, "SwiftCloud: Fault-Tolerant Geo-Replication Integrated all the Way to the Client Machine," Research Report RR-8347, HAL ID hal-00870225, 2013.
URL https://hal.inria.fr/hal-00870225

Distributed cloud computing allows to move the execution of distributed applications towards client machines. Current data management solutions for cloud infrastructures replicate data among several geographically distributed data centres but lack support for managing data maintained by clients. This talk presents SwiftCloud, a storage infrastructure for cloud environments that covers this gap. SwiftCloud addresses two main issues: maintaining replicas consistent and maintaining client replicas up-to-date. SwiftCloud pushes the scalability and concurrency envelope, ensuring transactional causal consistency using Conflict-Free Replicated Data Types (CRDTs). CRDTs provide higher-level object semantics, such as sets, maps, graphs and sequences, support unsynchronised concurrent updates, while provably ensuring consistency, and eschewing rollbacks. Client-side replicas are kept up to date by notifications, allowing client transactions to execute locally, both for queries and for updates.

3.2 Seattle Testbed (overview)

Justin Cappos (New York University, US)

URL https://seattle.poly.edu

Traditional distributed computational models, such as client-serverand cloud computing, involve moving computation from geographically distributed devices with little computational power to well-provisioned centralized servers. In this work, we explore the idea of harnessing computational resources on end user devices in an on-demand, cross application manner. Using this paradigm, we have constructed the Seattle testbed. Seattle makes it practical for arbitrary Internet users to securely participate in our testbed without compromising the security or performance of their laptop, desktop, phone, tablet or other device. Seattle has been deployed for six years across tens of thousands of end user devices. Seattle has wide spread practical use as a testbed for researchers and educators, including use in more than 50 classes at two dozen universities. The talk will include a demo of the system described.

3.3 A few Words on Data Management

Lars Eggert (NetApp Deutschland GmbH – Kirchheim, DE)

A brief overview of what I mean by data management, as well as some storage trends.

3.4 IoT Meets Cloud – Now we have to work out the details

Johan Eker (Lund University / Ericsson Research, SE)

In a not too distant future we expect to have tens of billions of devices connected to cloud and some of them will be providing services that require predictable latency and high availability. The cloud of tomorrow will move beyond today's IT services and into mission critical areas such as automation and health-care. To explore the full potential of such a scenario we must provide a programming platform that exposes cloud services and network functionalities in a simple and straightforward matter. This talk will present the Calvin project that aims at developing a programming framework for applications that spans over a heterogeneous hardware platform consisting of mobile sensors and cloud. This is work-in-progress and it will likely raise more questions than it answers.

3.5 The Rise of Software Defined Infrastructure

Chip Elliott (BBB Technologies – Cambridge, US)

We observe the interesting convergence between multi-tenant clouds, distributed clouds, network functions virtualization, and software defined networking, and discuss the emergence of software defined infrastructure.

3.6 Towards Federated Big Data Processing

Patrick Eugster (Purdue University – West Lafayette, US)

The "cloud" has a strong potential for efficiently performing analyses overlarge datasets. However, several hurdles need to be overcome to fulfill the vision of federated big data analyses across say patient records hosted by different parties and in different datacenters. First, copying all data to a single datacenter for subsequent analysis is inefficient, if feasible at all under sharing regulations. Second, especially when leveraging public clouds, processing confidential data in the cloud is all but desirable with the security dangers implied by multi-tenancy underlying cloud platforms. In this talk I will survey some practical first steps we made towards addressing these challenges. This includes our work on (a) geo-distributed big data analysis and (b) assured cloud-based big data analysis. In short, the former consists in moving computation towards data rather than only the other way around, and the latter consists in leveraging a combination of replication and partially homomorphic encryption to ensure integrity/correctness and privacy of big data analyzed in the cloud.

3.7 Distributed Clouds: Opportunities and Challenges for IT security

Hannes Hartenstein (KIT – Karlsruher Institut für Technologie, DE)

A security objective does typically not exist in isolation, but in combination with other objectives: confidentiality *and* performance, confidentiality *and* availability etc. In this talk/discussion we look at two use cases that both show benefits and challenges of distributed clouds, namely confidential data outsourcing and secret sharing schemes. We quickly scan through existing work that makes use of fragmentation techniques and of confidentiality preserving indexes. We show how confidentiality can be achieved based on non-colluding cloud providers and how the resulting trade-offs with performance and availability can be tuned for the cases of outsourcing databases and outsourcing strong cryptographic keys based on secret sharing schemes. The presentation should serve as a starting point for a discussion on appropriate security-related assumptions on properties of the architectures of distributed clouds.

3.8 User Perception of Distributed Clouds

Oliver Hohlfeld (RWTH Aachen University, DE)

A missing understanding of user requirements can challenge system engineering by its potential to cause over-engineered system architectures and sub optimal user experience. To foster research on improved distributed cloud architectures, this talk motivates the study of user requirements. Concrete requirements are dependent on several factors including applications, data, or different user types: humans (e.g., cloud gaming) vs. machines (e.g., virtualized network functions in carrier clouds). The talk first discusses restrictions on data storage locations as one example of potential user requirements. In this example, policy languages restrict data processing and storage in distributed clouds. It is further shown how user studies can be conducted to assess the impact of system design / network artifacts on end-user experience.

3.9 Cloud/WAN Networking

James Kempf (Ericsson – San Jose, US)

Cloud/WAN networking has become easier in recent years, but still needs work. I briefly discussed a few motivating use cases, bandwidth calendering, dynamic on the fly branch office VPN gateway deployment, telcom applications, then presented the Cloud Atlas architecture. Cloud Atlas supports orchestration of cloud/WAN network connections on any sort of wide area network virtualization technology by stitching the WAN connections into the Neutron tenant networks in an OpenStack cloud. A few simple abstractions provide the programmer with easy access to the WAN VPN. We implemented Cloud Atlas on a few different WAN VPN substrates, all supporting the MPLS L2VPN service VPLS. Dynamic on the fly branch office VPN deployment was also implemented under Cloud Atlas, with VLAN tags inside IPSec/GRE tunnels providing the WAN virtualization. Our experience with these prototypes led us to simplify the architecture down to a single primitive, the Gateway API, with bridges an L2 Neutron network to a Provider network. The Gateway API has been submitted to the Neutron working group for incorporation into the OpenStack release.

3.10 Mobile Distributed Cloud

James Kempf (Ericsson – San Jose, US)

I discussed a few representative use cases for the distributed cloud. One involves applications that have tight latency constraints (less than 50 ms). Another involves applications that have lots of hyper-local data that needs to be downloaded quickly or applications where a large amount of data must be processed prior to a deadline and the WAN connection to a central data center isn't sufficiently large enough to transfer the data to a central data center. The third is local deployment of Virtualized Network Functions.v

3.11 Model Checking of Threshold-based Fault-tolerant Distributed Algorithms

Igor Konnov (TU Wien, AT)

Joint work of Konnov, Igor; Gmeiner, Annu; Schmid, Ulrich; Veith, Helmut; Widder, Josef
Main reference A. John, I. Konnov, U. Schmid, H. Veith, J. Widder, "Parameterized model checking of fault-tolerant distributed algorithms by abstraction," in Proc. of the 13th Conf. on Formal Methods in Computer-Aided Design (FMCAD'13), pp. 201–209, IEEE, 2013.
URL http://dx.doi.org/10.1109/FMCAD.2013.6679411

Model checking of fault-tolerant distributed algorithms is challenging: the algorithms have multiple parameters that are restricted by arithmetic conditions, the number of processes and faults is parameterized, and the algorithm code is parameterized due to conditions involving counting the number of received messages (thresholds). We present our framework that allows us to model threshold-based fault-tolerant distributed algorithms and then efficiently model check them. To address parameterization, we introduced several model checking techniques for verification of such algorithms for all system sizes and all possible numbers of faults. We give an overview of these techniques and of our recent verification results.

3.12 Computation Distribution Networks and Distributed Clouds

Rick McGeer (HP Enterprise Services – Palo Alto, US)

The continuing performance increase of computation over communication has spurred the rise of many new network services over the past two decades: people are using the ubiquity of cheap, powerful computation to offer collaborative content distribution, transport proxies that offer high quality of service, in-network transcoding, adaptive routing, and so on. The efficacy of all of these services pales in comparison to the ability to arbitrarily site computation, particularly at the network edge: easily the most effective thing that one can do with a network is send a program over it. Specific use cases of computation distribution include in-situ data reduction, particularly for data with high-bandwidth sensors such as cameras, and high-bandwidth or latency-sensitive user interactions (real-time interactive simulation and distributed collaborative visualizations, for example). In this talk, I'll give some examples of computation distribution, and discuss a prototype of a Computation Distribution Network.

3.13 The Discovery initiative

Jonathan Pastor (INRIA Rennes – Bretagne Atlantique, FR)

Although the concept of Micro and Nano Data Centers (DCs) has been proposed to deliver more efficient as well as sustainable Utility Computing (UC) resources,the questions of where deploying and how federating thousands of such facilities are still far from being solved and the current trend of building larger and larger DCs in few strategic locations still prevails.The DISCOVERY initiative proposes to directly deploy the concept of Micro/NanoDCs upon the network backbones in order to benefit from existing network centers, starting from the core nodes of the backbone to the different network access points in charge of interconnecting public and private institutions.By such a mean, network and UC providers would be able to mutuality resources that are mandatory to operate network/data centers while delivering widely distributed UC platforms being able to better match the geographical dispersal of users.

3.14 Decomposing consistency

Marc Shapiro (UPMC – Paris, FR)

Joint work of Shapiro, Marc; Saeida-Ardekani, Masoud; Zawirski, Marek; Balegas, Valter; Najafzadeh, Mahsa; Gotsman, Alexey

There are many competing consistency models (serialisability, snapshot isolation, eventual consistency, etc.), all subtly different and hard to understand. This variety reflects a fundamental trade-off between fault tolerance, performance, and programmability. Furthermore, most papers describe consistency models in terms of acceptable histories, which is not very informative. The design choices are particularly vexing at large scale and in the presence of failure, for instance in geo-replicated or edge clouds. We believe that what programmers really care about is a consistency model's *properties*. We study two classes of properties: guarantees (i.e., what kind of application invariants are ensured automatically by a model) and scalability properties (i.e., opportunities for parallelism and implementation freedoms in a model). The properties are duals between the two classes; they are (mostly) orthogonal within a class. A particular composition of properties will characterize a model. We also study some abstract classes of guarantees (e.g., partial-order-type invariants, equivalance-type invariants, identical-observer guarantee) and opportunities (e.g. genuine partial replication) and their impact on the protocols and the models.

3.15 How to distribute cloud computing to the edge?

Hagen Woesner (BISDN GmbH – Berlin, DE)

URL http://www.fp7-unify.eu/

The talk introduces the *fat CPE* as opposed to other approaches of stripping down *virtualized CPEs* to a single tunnel endpoint. Some use cases motivate this choice. In the following, a universal node architecture is introduced (coming from EU project UNIFY). We end asking questions of how resources should be exposed: As network function forwarding graphs (NF-FG)?

3.16 Is the Cloud Your Data Center, Your Network, or Both?

Tim Wood (George Washington University – Washington, US)

Main reference J. Hwang, K. K. Ramakrishnan, T. Wood, "NetVM: High Performance and Flexible Networking using Virtualization on Commodity Platforms," in Proc. of the 11th USENIX Symp. on Networked System Design and Implementation (NSDI'14), pp. 445-458, USENIX Association, 2014.
URL https://www.usenix.org/conference/nsdi14/technical-sessions/presentation/hwang

Today's trends have been towards a handful of major cloud platforms composed of a small number of massive data centers. While this gives benefits from economy of scale, it may not be able to achieve the resiliency and performance of a more distributed model. This talk will discuss how software-based networks using SDN and NFV may help move us back towards a distributed cloud model by allowing software services to easily and efficiently run in the network itself. By pushing storage and computational capabilities into the network, we can perform computation precisely when and where it is needed, reducing latency and increasing fault tolerance.

4 Panel Discussions

Over the course of the seminar, we organized three panel discussions on the topics of (i) general aspects of distributed cloud computing, (ii) networking aspects of distributed cloud computing, and (iii) distributed systems and programmability aspects of distributed cloud computing. The panel discussions aimed to stimulate breakout sessions that identify research challenges.

4.1 Panel Discussion on Distributed Cloud Computing

Justin Cappos, Lars Eggert, Chip Elliott, Oliver Hohlfeld, and Rick McGeer

- Seattle: distributed cloud that is larger than Planet Lab
- What is a distributed cloud? It is an infrastructure with substantially distributed computation and storage resources. A CDN would not count as DC since they only serve traffic and provide no to limited possibilities of running compute jobs / programmability.
- It is all about latency. Latency makes nodes different from each other.
- What is the differences between traditional means of computation and distributed cloud computing? The cloud should simplify approaches.
- A lot of research is needed to meet performance criteria / SLAs in practice. There is no one button to be clicked to instantiate a cloud app that meets specified criteria.
- The motivation for cloud computing is driven by costs. Distributed cloud computing offers redundancy and reliability. In addition it provides instant fail-overs by having remote replicas that can be booted up in immediately increase of failures.
- Use Case 1: Educational games. There are reasons for not sharing game code with the end users: 1) prevent cheating / maintain control and 2) simplify portability: by only streaming video rendered by the game over the network, less platform need to be supported by game vendors.
- Use Case 2: Programs should be sent over the network rather than data.Programs are much smaller but typically generate large amounts of data. For example, detectors at the LHC at Cern can generate 3M events in a short amount of time. Since not all of these events are of interest, the network utilization can be reduced by moving a filter program next to the sensor.
- Argument challenged: Big data applications like LHC will always rely on dedicated infrastructures. Ordinary users will not run compute job on LHC cloud nodes and will demand different infrastructures.
- Sending VMs to storage systems is challenging since the CPU capacity is limited and will be occupied by compute jobs. It is better to move the data to dedicated compute

clusters for processing when the data is hot. The processing pipeline should focus on hot data rather than cold data.
- Privacy vs. performance
- Privacy concerns drive migrations to private clouds running Open Stack. Another approach relies on moving storage devices owned by the data owners next to big data centers (e.g., Amazon) for fast-path access to compute clusters.
- What are the programming models for distributed cloud computing? (other than simply placing a VM)
- Infrastructure management / DHT centralized lookup systems
- Support for desktop orchestration. It should be possible to add a distributed file system as fuse model to be run at desktops.

Identified discussion topics:
1. What is a distributed cloud?
2. Programmability models
3. Scalability and repeatability of experiments
4. How can academia compute with big companies?
5. Service composition / Decomposition – what are the primitives for infrastructure management?
6. What are the applications using distributed cloud computing?
7. Is there an educational use for distributed cloud computing?

4.2 Panel Discussion on Networking Aspects of Distributed Cloud Computing

Mark Berman, Lars Eggert, Oliver Hohlfeld, James Kempf, and Hagen Wösner

- Discussion on the Open Stack networking architecture
- Open Stack runs the VM placement algorithm first and then performs the wiring of the placed VMs. Emulab uses simulated annealing for its optimization. Can such an approach be applied? How do VM images get copied over the network?
- Shipping VM images in principle is a bad idea. They should be kept local / cached.
- We use dedicated fibers are used to ship images. It can be used but does not represent the general case.

Use cases:
- Future robot control software generates massive amounts of data that need to be processed in a local cloud since the robots are not powerful enough.
- Service providers would like to move VMs to users' homes for debugging. Such a use case was once built on top of Seattle and is also discussed in the ETSI NFV use cases.
- Home gateways should have higher compute powers in order to run VMs.

Identified discussion questions:
1. Relationship between clouds and CPE's? Does the cloud land in my house or is my house outsourced to some cloud?
2. Joint placement of CFNet
3. Clouds in mobile networks
4. Cloud to WAN: next steps

4.3 Panel Discussion on Aspects of Distributed Systems and Programmability

Annette Bieniusa, Patrick Eugster, Oliver Hohlfeld, Marc Shapiro, and Tim Wood

Which programming primitives are needed for distributed cloud computing?
- Virtual machine is an execution model but not a programming model.
- We need a more fine grained understanding of what a programmer intends to do with resources provided by distributed clouds. E.g., it might be worth to migrate single users of a VM rather than the entire VM.
- Which primitives are needed for distributed state
- There can be different kinds of state that require different consistency models. Thus, a library providing support for a large variety of such models is needed.
- At which level is such a library envisioned? At a service composition or at a bit level? A generic model will not provide good performance at all layers, i.e., there should also be specialized models be available.
- In distributed systems, we didn't end up needing as much consistency as we initially thought.
- Programming languages failed in the distributed world as there is no one size fits it all in distributed systems. Different requirements need to be supported. Why is distributed cloud computing different from classical work in distributed systems?
- Different application demands
- Trend towards domain specific languages Orthogonal to these 3 areas, there is *latency* and *faults*.
- Variations in performance will be a big challenge in distributed cloud computing.

What is the appeal for the educational use of a distributed cloud platform such as Seattle?
- Seattle simplifies experimentation with a distributed networked systems. It is very appealing to school students to test their approaches by running at remote computers, e.g., phones located in China.

The following questions were identified by the panel:
1. Fault tolerance
2. Latency requirements
3. If one size doesn't fits it all, how many sizes do we need?
4. Configuration management
5. Data quality
6. Requirements and models of different applications run on distributed clouds
7. Automated decision support

5 Working Groups Identifying Research Challenges

We used questions identified in the panel discussions to form and stimulate working groups. The formed working groups then identified potential research questions in their respective areas.

5.1 Research Challenges in Distributed Clouds

This breakout session concerned identifying research challenges that arise in distributed clouds. We identified five research questions out of which one is dedicated to programming language techniques and one to networking aspects.

- In multi-domain distributed clouds, different parts of the cloud will be owned and operated by different organizations. How will resources be described and obtained? Will there be brokers? How does multi-domain management works?
- If parts of the cloud will be battery powered, what impact will this have on the architecture in terms of reliability, consistency, etc.?
- What are the 3 to 4 driving use cases for distributed cloud computing?
- Programming Languages: When is consistency needed and to what degree? When is BFT needed and to what degree? How frequently will apps need these techniques and how does one know they are worth the complexity?
- Distributed Systems: How does SDN relate to distributed cloud computing,if at all?

5.2 Distributed Systems Research Challenges for Distributed Cloud

This breakout session concerned identifying research challenges that arise in distributed systems as a driver for distributed clouds. We identified three research questions for distributed systems, and two for the other two breakout groups (clouds and programming languages).

- (Distributed Systems) How about performing a threat analysis on some consistency mechanisms to determine possible security issues?
- (Distributed Systems) How do you write control algorithms for controlling the distributed cloud that are salable?
- (Distributed Systems)In causal consistency, what would the effect be of replacing the deterministic causal graph with a probabilistic causal graph?
- (Programming Languages) How to provide good abstractions in the face of heterogeneous underlying resources (i.e. not making the abstractions least common denominator and not providing so many parameters that the abstractions become too complex)?
- (Clouds) What are the trust models for distributed cloud and does distributed cloud make security easier or harder?

5.3 Programmability Group

- How do we deploy and monitor apps running on DCC?
- How is programmability constrained by the infrastructure?
- Which primitives do we need to express what we want?
  - Primitives for describing the placement of components- Primitives for configuring the non-functional requirements
  - Primitives for configuring the functional requirements- Low-level vs. High-level primitives
  - Dynamic environment discovery
  - Primitives for SLA
- What are the relevant applications? Identify major classes of applications. Suggestions:
  - Big data analytics
  - Services (control loop, drive my car, do face matching)

    - Games MMOG- Latency-sensitive applications
    - Web servers

  What are their requirements in terms of programmability? Note: all these kinds of applications define what a DCC is.
- How do we educate developers?
    - Focus on testing
    - Focus on performance evaluation
- Question for the cloud group: How do you build an infrastructure allowing the deployment of computations and data wherever needed?
- Question for the distributed system group: How do we address the placement issue? → we might want these two components to be far away. Or to be on the same computer. Or on one particular resource.

Participants

- Mark Berman
BBN Technologies – Cambridge, US
- Annette Bieniusa
TU Kaiserslautern, DE
- Justin Cappos
New York University, US
- Yvonne Coady
University of Victoria, CA
- Lars Eggert
NetApp Deutschland GmbH – Kirchheim, DE
- Johan Eker
Lund University / Ericsson Research, SE
- Chip Elliott
BBN Technologies – Cambridge, US
- Erik Elmroth
University of Umeå, SE
- Patrick Eugster
Purdue University, US
- Hannes Hartenstein
KIT – Karlsruher Institut für Technologie, DE
- Oliver Hohlfeld
RWTH Aachen, DE
- James Kempf
Ericsson – San Jose, US
- Igor Konnov
TU Wien, AT
- Rick McGeer
HP Enterprise Services – Palo Alto, US
- Christine Morin
INRIA Rennes – Bretagne Atlantique, FR
- Jörg Ott
Aalto University, FI
- Jonathan Pastor
INRIA Rennes – Bretagne Atlantique, FR
- Vivien Quéma
INP – Grenoble, FR
- Stefan Schmid
TU Berlin, DE
- Marc Shapiro
UPMC – Paris, FR
- Hagen Wösner
BISDN GmbH – Berlin, DE
- Tim Wood
George Washington University – Washington, US

Report from Dagstuhl Seminar 15081

Holistic Scene Understanding

Edited by

Jiří Matas[1], Vittorio Murino[2], Laura Leal-Taixé[3], and Bodo Rosenhahn[5]

1 **Czech Technical University, CZ,** matas@cmp.felk.cvut.cz
2 **Italian Institute of Technology – Genova, IT,** vittorio.murino@iit.it
3 **ETH Zürich, CH,** leal@geod.baug.ethz.ch
5 **Leibniz Universität Hannover, DE,** rosenhahn@tnt.uni-hannover.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 15081 "Holistic Scene Understanding". During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper. The first section describes the seminar topics and goals in general. Overall, the seminar was a great success, which is also reflected in the very positive feedback we received from the evaluation.

Seminar February 15–20, 2015 – http://www.dagstuhl.de/15081
1998 ACM Subject Classification I.2.6 Learning, I.2.10 Vision
Keywords and phrases Scene Analysis, Image Understanding, Crowd Analysis, People and Object Recognition
Digital Object Identifier 10.4230/DagRep.5.2.80
Edited in cooperation with Roberto Henschel

1 Executive Summary

Jiří Matas
Vittorio Murino
Laura Leal-Taixé
Bodo Rosenhahn

Motivations

To *understand* a scene in a given image or video is much more than to *simply* record and store it, extract some features and eventually recognize an object. The overall goal is to find a mapping to derive semantic information from sensor data. Purposive Scene understanding may require a different representation for different specific tasks. The task itself can be used as prior but we still require an in-depth understanding and balancing between local, global and dynamic aspects which can occur within a scene. For example, an observer might be interested to understand from an image if there is a person present or not, and beyond that, if it is possible to look for more information, e.g. if the person is sitting, walking or raising a hand, etc.

When people move in a scene, the specific time (e.g. 7:30 in the morning, workdays, weekend), the weather (e.g. rain), objects (cars, a bus approaching a bus stop, crossing bikes, etc.) or surrounding people (crowded, fast moving people) yield to a mixture of low-level and high-level, as well as abstract cues, which need to be jointly analyzed to get an in-depth understanding of a scene. In other words, generally speaking, the so-called *context* is to be considered for a comprehensive scene understanding, but this information, while it is easily captured by human beings, is still difficult to obtain from a machine.

Holistic scene interpretation is crucial to design the next generation of recognition systems, which are important for several applications, e.g. driver assistance, city modeling and reconstruction, outdoor motion capture and surveillance.

With such topics in mind, the aim of this workshop was to discuss which are the sufficient and necessary elements for a complete scene understanding, i.e. what it really means to *understand* a scene. Specifically, in this workshop, we wanted to explore methods that are capable of representing a scene at different level of semantic granularity and modeling various degrees of interactions between objects, humans and 3D space. For instance, a scene-object interaction describes the way a scene type (e.g. a dining room or a bedroom) influences objects' presence, and vice versa. An object-3D-layout or human-3D-layout interaction describes the way the 3D layout (e.g. the 3D configuration of walls, floor and observer's pose) biases the placement of objects or humans in the image, and vice versa. An object-object or object-human interaction describes the way objects, humans and their pose affect each other (e.g. a dining table suggests that a set of chairs are to be found around it). In other words, the 3D configuration of the environment and the relative placements and poses of the objects and humans therein, the associated dynamics (relative distance, human body posture and gesture, gazing, etc.), as well as other contextual information (e.g., weather, temperature, etc.) support the holistic understanding of the observed scene.

As part of a larger system, understanding a scene semantically and functionally allows to make predictions about the presence and locations of unseen objects within the space, and thus predict behaviors and activities that are yet to be observed. Combining predictions at multiple levels into a global estimate can improve each individual prediction.

Since most scenes involve humans, we were also interested in discussing novel methods for analyzing group activities and human interactions at different levels of spatial and semantic resolution. As advocated in recent literature, it is beneficial to solve the problem of tracking individuals and understand their activities in a joint fashion by combining bottom-up evidence with top-down reasoning as opposed to attack these two problems in isolation.

Top-down constraints can provide critical contextual information for establishing accurate associations between detections across frames and, thus, for obtaining more robust tracking results. Bottom-up evidence can percolate upwards so as to automatically infer action labels for determining activities of individual actors, interactions among individuals and complex group activities. But of course there is more than this, it is indeed the cooperation of both data flows that makes the inference more manageable and reliable in order to improve the comprehension of a scene.

We gathered researchers which are not only well-known in Computer Vision areas such as object detection, classification, motion segmentation, crowd and group behavior analysis or 3D scene reconstruction, but also Computer Vision affiliated people from other communities in order to share each others point of view on the common topic of scene understanding.

Goals

Our main goals of the seminar can be summarized as follows:

- Address holistic scene understanding, a topic that has not been discussed before in detail at previous seminars, with special focus on a multidisciplinary perspective for sharing or competing the different views.
- Gather well-known researchers from the Computer Vision, Machine Learning, Social Sciences (e.g. Cognitive Psychology), Neuroscience, Robotics and Computer Graphics communities to compare approaches to representing scene geometry, dynamics, constraints as well as problems and task formulations adopted in these fields. The interdisciplinary scientific exchange is likely to enrich the communities involved.
- Create a platform for discussing and bridging topics like perception, detection, tracking, activity recognition, multi-people multi-object interaction and human motion analysis, which are surprisingly treated independently in the communities.
- Publication of an LNCS post-proceedings as previously done for the 2006, 2008 and 2010 seminars. These will include the scientific contributions of participants of the Seminar, focusing specially on the discussed topics presented at the Seminar.

Organization of the seminar

During the workshop we discussed different modeling techniques and experiences researchers have collected. We discussed sensitivity, time performance and e.g. numbers of parameters required for special algorithms and the possibilities for context-aware adaptive and interacting algorithms. Furthermore, we had extensive discussions on open questions in these fields.

On the first day, the organizers provided general information about Dagstuhl seminars, the philosophy behind Dagstuhl and the expectations to the participants. We also clarified the kitchen-rules and organized a running-group for the early mornings (5 people participated frequently!).

Social event. On Wednesday afternoon we organized two afternoon event: One group made a trip to Trier, and another group went on a 3h hike in the environment.

Working Groups. To strongly encourage discussions during the seminar, we organized a set of working groups on the first day (with size between 8–12 people). As topics we selected
- What does “Scene Understanding” mean ?
- Dynamic Scene: Humans.
- Recognition in static scenes (in 3D).

There were two afternoon slots reserved for these working groups and the outcome of the working groups has been presented in the Friday morning session.

LNCS Post-Proceedings. We will edit a Post-Proceeding and invite participants to submit articles. In contrast to standard conference articles, we allow for more space (typically 25 single-column pages) and allow to integrate open questions or preliminary results, ideas, etc. from the seminar into the proceedings. Additionally, we will enforce joint publications of participants who started to collaborate after the seminar. All articles will be reviewed by at least two reviewers and based on the evaluation, accepted papers will be published. We will publish the proceeding at the Lecture Notes in Computer Science (LNCS-Series) by Springer. The papers will be collected during the summer months.

Overall, it was a great seminar and we received very positive feedback from the participants. We would like to thank castle Dagstuhl for hosting the event and are looking forward to revisit Dagstuhl whenever possible.

2 Table of Contents

3 Overview of Talks

3.1 Superhuman-in-the-Loop Computer Vision

Gabriel Brostow (University College London, GB)

Joint work of Mac Aodha, Oisin; Stathopoulos, Vassilios; Terry, Michael; Jones, Kate E.; Brostow, Gabriel J.; Girolami, Mark
Main reference O. Mac Aodha, V. Stathopoulos, G. J. Brostov, M. Terry, M. A. Girolami, K. E. Jones, "Putting the Scientist in the Loop – Accelerating Scientific Progress with Interactive Machine Learning," in Proc. of the 22nd Int'l Conf. on Pattern Recognition, pp. 9–17, IEEE, 2014; pre-print available from author's webpage.
URL http://dx.doi.org/10.1109/ICPR.2014.12
URL http://web4.cs.ucl.ac.uk/staff/g.brostow/papers/engage_icpr_2014.pdf

Despite ubiquitous computing, most normal people are not benefiting from advancements in computer vision research. Equally, most vision systems do not improve with time or learn from their users' experience. This is a terrible waste, but is understandable: there are plenty of specific vision problems where progress a) can be made "offline" in labs trying to beat a recognized benchmark score, and b) the specific problem affects a big industry, like scene-flow for cars, or image-retrieval for search engines.

In this talk, I advocate that we should be aiming for responsive algorithms, and that these should be measured in terms of accuracy-improvement, and the user's ability to perform their specific tasks. This means we will need new benchmarks, and that we need to engage with real users for our models and experiments to be meaningful. While my group has started making software that adapts to specialist users, e.g. biologists/zoologists, the ageing population is just one mass-scale cohort that will require new computer vision models and interfaces.

3.2 From groups to crowds: a social signal processing perspective

Marco Cristani (University of Verona, IT)

After years of research on automated analysis of individuals, the computer vision community has shifted its attention towards the new issues of modeling groups and crowds. Within the scope of computer vision, groups are generally defined simply as two or more people moving at a similar velocity, spatially and temporally close to one another. However, things are a bit more complex: there are many kinds of groups, that differ in dimension, durability (ephemeral, ad hoc or stable groups), in/formality of organization, degree of "sense of belonging", level of physical dispersion etc.. Along the same lines, crowds are usually intended as a large number of persons gathered closely together; but, even in this case, the notion of crowd is much more complex and requires a more detailed account, which is basically missing in the computer vision community. In this talk, we build on concepts inherited from the sociological analysis and we offer a detailed taxonomy of groups and crowds. As we will see, this analysis individuates many typologies of social gatherings, each with its own characteristics and behavior dynamics. These differences are not only useful for a mere classification purpose, but are crucial when the need of automatic modeling comes into play, eliging particular computer vision techniques and models as the most appropriate to account for such differences. In this talk, in particular, we will focus on a specific kind of group, i.e. free-standing conversational group, and one kind of crowd, i.e. spectator crowd, showing recent advancements in their automatic modeling.

3.3 Semantic Motion Segmentation and 3D Reconstruction

Alessio Del Bue (Italian Institute of Technology–Genova, IT)

Joint work of Del Bue, Alessio; Crocco, Marco; Rubino, Cosimo
Main reference C. Rubino, M. Crocco, A. Perina, V. Murino, A. Del Bue, "3D Structure from Detections," arXiv:1502.04754v2 [cs.CV], 2015.
URL http://arxiv.org/abs/1502.04754v2

In this talk I will present how to embed semantic information in classical multi-view geometry problems such as multi-body motion segmentation. The key feature is the explicit inclusion of geometrical priors given by general purpose object detectors that boost the segmentation of the moving objects. In the classical formulation of the problem, only 2D matched points between views are used to identify independently moving objects leveraging the principle that a set of points belonging to a moving object would satisfy some given multi- view relations (e.g. multi-body epipolar constraints). We improve and speedup motion segmentation by including the information that a set of 2D matches may belong to the same object given the output of a detector. As such, instead of sampling points uniformly with a RANSAC based strategy, the selection of the matches is driven by the position and score confidence of the object detectors. After some experimental evidence, conclusions will show that other problems, (e.g. 3D reconstruction) can be supported by the inclusion of semantic information extracted from a generic scene.

3.4 Feature Regression-based Pose Estimation for Object Categories

Michele Fenzi (Leibniz Universität Hannover, DE)

Joint work of Fenzi, Michele; Leal-Taixé, Laura; Rosenhahn, Bodo; Ostermann, Jörn
Main reference M. Fenzi, L. Leal-Taixé, B. Rosenhahn, J. Ostermann, "Class Generative Models based on Feature Regression for Pose Estimation of Object Categories," in 2013 IEEE Conf. on ", IEEE Conference on Computer Vision and Pattern Recognition (CVPR'13), pp. 755-762, IEEE, 2013.
URL http://dx.doi.org/10.1109/CVPR.2013.103

I present an approach to pose estimation for object categories based on feature regression [1]. Pose estimation for object categories is becoming increasingly important and of interest, both as a fundamental part of larger tasks or as a standalone challenge. Among the many approaches proposed in literature, those based on local features have shown to work effectively. While some use explicit 3D information, others have shown that coupling feature regression and view labelling is enough to solve this task. I present a method for learning a class representation and a pose estimation algorithm that returns a continuous value for the pose of an unknown class instance using only 2D data and weak labelling information. Our method is based on generative feature models, i.e., regression functions learnt from local descriptors of the same patch collected under different viewpoints. The individual generative models are then clustered in order to create class generative models which form the class representation.

At run-time, geometric consistency is introduced in the matching step by means of a graph matching strategy [2]. Finally, the pose of the query image is estimated in a probabilistic fashion by combining the regression functions belonging to the matching clusters.

References

1 Michele Fenzi, Laura Leal-Taixé, Bodo Rosenhahn, Jörn Ostermann, "Class Generative Models based on Feature Regression for Pose Estimation of Object Categories", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, Oregon, USA, June 2013

2 Michele Fenzi, Jörn Ostermann, "Embedding Geometry in Generative Models for Pose Estimation of Object Categories", British Machine Vision Conference (BMVC), Nottingham, United Kingdom, September 2014

3.5 Fish detection, tracking, recognition and analysis with the Fish4Knowledge Dataset

Bob Fisher (University of Edinburgh, GB)

Main reference B. J. Boom, J. He, S. Palazzo, P. X. Huang, C. Beyan, H.-M. Chou, F.-P. Lin, C. Spampinato, R. B. Fisher, "A research tool for long-term and continuous analysis of fish assemblage in coral-reefs using underwater camera footage," Ecological Informatics, Vol. 23, pp. 83–97, 2014.
URL http://dx.doi.org/10.1016/j.ecoinf.2013.10.006

The research presented here was based on the data collected by the EU funded Fish4Knowledge project. Altogether, 80 Tb of video was recorded from 9 cameras, resulting in 90,000 hours of video. From this data, 1.4 billion fish observations were detected and tracked, resulting in 145 million trajectories. The individuals from these trajectories were then classified into 23 species. Some of the interesting results observed in this scene data was:

1. Much of the data was degraded, due to water quality after storms, algae growing on the lenses and compression artifacts.
2. The distribution of fish species was greatly unbalanced, resulting in e.g. 50 times greater observation frequencies for the most common species relative to the 10th most commonly observed fish.
3. Because of similarities in appearance, a hierarchical classifier gave the best species recognition performance, with 97% recognition accuracy over all fish and 75% when averaged over the top 15 classes.
4. Although each observation varied by the individual behaviour and noise in the data, having millions of observations allowed conclusion to be made – eg. that there appear to be more fish observed at the ends of the day than in the middle, and the swimming speed of D. reticulatus increases as the water temperature increases.
5. Another interesting result presented concerned the ability to recognise individual fish. Using a collection of colour and texture properties, it was possible to strongly cluster A. clarkii (clownfish) observations. Although no ground truth was available, the fish in the clusters were highly similar in appearance and different from the fish in other clusters.

3.6 Human Pose: A Cue for Activities and Objects

Jürgen Gall (Universität Bonn, DE)

Joint work of Gall, Jürgen; Jhuang, Hueihan; Zuffi, Silvia; Schmid, Cordelia; Black, Michael J.; Srikantha, Abhilash
URL http://ps.is.tuebingen.mpg.de/project/JHMDB

In this talk, I discuss human pose as a cue for action recognition and object discovery. In order to allow a systematic performance evaluation of an action recognition pipeline, we annotated human joints for the HMDB dataset (J-HMDB). The annotation can be used to systematically replace the output of various algorithms in an existing pipeline with ground truth data to analyze the components with the highest potential for improving the recognition accuracy. For example, is it worth to invest more time on improving low-level algorithms like optical flow, is the image location of the human performing the action important, or would knowledge about human pose be helpful? Given pose and activities, we can also reason about objects. For example, small objects can be discovered from videos.

References

1 Hueihan Jhuang, Jürgen Gall, Silvia Zuffi, Cordelia Schmid and Michael J. Black. *Towards Understanding Action Recognition*. ICCV, 2013
2 Abhilash Srikantha and Jürgen Gall. *Discovering Object Classes from Activities*. ECCV, 2014

3.7 Explore Multi-source Information for Holistic Visual Search

Shaogang Gong (Queen Mary University of London, GB)

Joint work of Gong, Shaogang; Xiang, Tao; Hospedales, Tim; Loy, Chen Change; Zheng, Wei-Shi; Fu, Yanwei

For making sense of big visual data captured by large scale distributed multi-cameras in urban environments, understanding human activities and behaviour, detecting and searching their whereabouts in crowded spaces are required. In this talk, I will present some recent progress on exploring multiple information sources for modelling holistic context-aware object detection and activity profiling in large volumes of surveillance video data, addressing the problems of abnormal behaviour/action/event detection in public spaces and person re-identification in large scale distributed CCTV camera networks. In particular, I will discuss the needs and open-questions on a number of model learning challenges, including: learning visual context of activity for abnormal behaviour discovery and object association context for increasing detection robustness; exploring human-in-the-loop active learning for overcoming sparse labelling information in person re-identification and minimising false detection in anomaly detection; exploring crowd information for learning space-time camera network topology in disjoint multi-camera person tracking; and exploring semantic structure (e.g. the WordNet semantic space) for Zero-Shot-Learning in object recognition.

References

1 Gong, Xiang. *Visual Analysis of Behaviour: From Pixels to Semantics*, Springer, May 2011 http://dx.doi.org/10.1007/978-0-85729-670-2

2 Gong, Cristani, Yan, Loy. *Person Re-Identification*, Springer, January 2014 http://dx.doi.org/10.1007/978-1-4471-6296-4
3 Fu, Hospedales, Xiang, Gong. *Transductive Multi-View Zero-Shot Learning*. IEEE TPAMI, in press, 2015. http://dx.doi.org/10.1109/TPAMI.2015.2408354

3.8 VocMatch: Efficient Multiview Correspondence for Structure from Motion

Michal Havlena (ETH Zürich, CH)

Joint work of Havlena, Michal; Schindler, Konrad
Main reference M. Havlena, K. Schindler, "VocMatch: Efficient Multiview Correspondence for Structure from Motion," in Proc. of the 13th Europ. Conf. on Computer Vision (ECCV'14), LNCS, Vol. 8691, pp. 46–60, Springer, 2014.
URL http://dx.doi.org/10.1007/978-3-319-10578-9_4

Feature matching between pairs of images is a main bottleneck of structure-from-motion computation from large, unordered image sets. We propose an efficient way to establish point correspondences between all pairs of images in a dataset, without having to test each individual pair. The principal message is that, given a sufficiently large visual vocabulary, feature matching can be cast as image indexing, subject to the additional constraints that index words must be rare in the database and unique in each image. We demonstrate that the proposed matching method, in conjunction with a standard inverted file, is 2-3 orders of magnitude faster than conventional pairwise matching. The proposed vocabulary-based matching has been integrated into a standard SfM pipeline, and delivers results similar to those of the conventional method in much less time.

3.9 Solving Multiple People Tracking Using Minimum Cost Arborescences

Roberto Henschel (Leibniz Universität Hannover, DE)

Joint work of Henschel, Roberto; Leal-Taixé, Laura; Rosenhahn, Bodo
Main reference R. Henschel, L. Leal-Taixé, B. Rosenhahn, "Efficient Multiple People Tracking Using Minimum Cost Arborescences," in Proc. of the 36th German Conf. on Pattern Recognition (GCPR'14), LNCS, Vol. 8753, pp. 265–276, Springer, 2014.
URL http://dx.doi.org/10.1007/978-3-319-11752-2_21

Current state-of-the-art multiple people tracker that use a hierarchical tracklet framework are prone to error propagation due to wrong decisions in ambiguous situations.

I will thus present a new type of tracklets, which are called tree tracklets, that helps to solve this issue. A tree tracklet contains bifurcations to naturally model ambiguous situations. The optimal data association is derived from a minimum cost arborescence in an acyclic directed graph. Thereby a solution of the association problem is obtained in linear time. I will present experiments on six well-known multiple people tracking datasets showing the good performance compared to state-of-the art tracking algorithms.

3.10 Generic Object Detection in Video using Saliency and Tracking

Esther Horbert (RWTH Aachen, DE)

Joint work of Horbert, Esther; M. Garcia, German; Frintrop, Simone; Leibe, Bastian
Main reference E. Horbert, G. M. García, S. Frintrop, B. Leibe, "Sequence-Level Object Candidates Based on Saliency for Generic Object Recognition on Mobile Systems," in Proc. of the 2015 IEEE Int'l Conf. on Robotics and Automation (ICRA'15), to appear; pre-print available from author's webpage.
URL http://www.vision.rwth-aachen.de/projects/kod/horbert-icra15-preprint

We propose a novel approach for generating generic object candidates for object discovery and recognition in continuous monocular video. Such candidates have recently become a popular alternative to exhaustive window-based search as basis for classification. Contrary to previous approaches, we address the candidate generation problem at the level of entire video sequences instead of at the single image level. We propose a processing pipeline that starts from individual region candidates and tracks them over time. This enables us to group candidates for similar objects and to automatically filter out inconsistent regions. For generating the per-frame candidates, we introduce a novel multi-scale saliency approach that achieves a higher per-frame recall with fewer candidates than current state-of-the-art methods. Taken together, those two components result in a significant reduction of the number of object candidates compared to frame level methods, while keeping a consistently high recall.

3.11 A Spectral Perspective on Invariant Measures of Shapes

Ron Kimmel (Technion–Haifa, IL)

Joint work of Aflalo, Yonathan; Brezis, Haim; Kimmel, Ron
Main reference On the optimality of shape and data representation in the spectral domain
URL http://www.cs.technion.ac.il/~ron/publications.html

We explore the power of the Laplace Beltrami Operator (LBO) in processing and analyzing visual and geometric information. The decomposition of the LBO at one end, and the heat operator at the other end provide us with efficient tools for dealing with images and shapes. Denoising, segmenting, filtering, exaggerating are just few of the problems for which the LBO provides a solution. We will review the optimality of a truncated basis provided by the LBO, and a selection of relevant metrics by which such optimal bases are constructed. Specific example is the scale invariant metric for surfaces that we argue to be a natural selection for the study of articulated shapes and forms.

3.12 Visual Odometry and 3D Roadside Reconstruction

Reinhard Klette (University of Auckland, NZ)

Joint work of Klette, Reinhard; Johnny Chien; Haokun Geng; Simon Hermann; Waqar Khan; Sandino Morales; Radu Nicolescu

Future cars might possibly contribute to some kind of incremental 3D roadside reconstruction, supporting a spatio-temporal model of road environments. Besides accurate car trajectory calculation and provision of accurate depth data, it is also necessary to unify multiple runs in a uniform world coordinate system, to ignore data caused by dynamic or transient static objects, and to be forgiving when reconstructing secondary surfaces such as of plants.

The talk informed about data recording in a test vehicle (trinocular, 16 bit per pixel, 2046 x 1080, 30 Hz), what results into 126 GB in just about 5.5 minutes of recording. The generated 3D roadside data can be used to enhance 3D city scene reconstructions obtained by aerial mapping (illustrated by a 2007 example of reconstructing the Sony Centre in Berlin with 7 cm ground-sample accuracy; courtesy by H. Hirschmueller and K. Scheibe). Current iOS Maps do not yet have the accuracy required for roadside models. Stereo matchers such as iSGM (S. Hermann and R. Klette, winner of Robust Vision Challenge at ECCV 2012) or linBPM (W. Khan and R. Klette, IEEE IV 2013) may be considered to be satisfactory tools for providing 3D roadside depth data.

The talk suggested a definition for "robustness", considering the sum of accuracies for "challenging" scenarios times the probabilities of such scenarios. For defining the accuracy for one scenario, the third-eye technology (S. Morales and R. Klette, CAIP 2009) might be considered as an option, using an NCC measure defined by pixels closed to image discontinuities.

Examples (four short sequences of just 400 input stereo frames) illustrated that the performance of stereo matchers depends on the input data, characterised by complexity of scene geometry, weather or lighting conditions, traffic density, and so forth.

Generating visually satisfying roadside geometry requires dense depth maps with "visually accurate" occlusion edges; methods designed in (D. Liu and R. Klette, The Visual Computer, 2015) might be considered for depth map corrections.

The baseline algorithm (SfM: apply visual odometry for mapping clouds of points into a uniform world coordinate system) can be enhanced by various considerations. First, depth is more accurate closer to cameras, and road geometry further away can be considered later, when the car arrives there. However, far-away depth values already provide a useful approximation of the expected scene geometry. NCC values (third-eye technology) also provide weights for obtained depth data, to be considered when integrating into an already existing surface model.

Two options have been considered for visual odometry, sparse bundle adjustment (with simplifying 3D surface representations by depth-based representations only; see J. Chien, H. Geng, R. Klette, 2015, submitted) or feature-matching also using GPS data (with iconic Kalman filters, and an extended Kalman filter for the overall movement; see H. Geng, J. Chien, R. Nicolescu, and R. Klette, 2015, submitted). Tracked features have been studied for invariance properties, showing invariance issues for all the studied features (including SIFT and SURF).

The required accuracy for ego-motion detection, and the automated unification of noisy multi-run 3D data into one spatio-temporal roadside model appear to be the two most challenging subjects in this area.

3.13 Learning an image-based motion context for multiple people tracking

Laura Leal-Taixé (ETH Zürich, CH)

Joint work of Leal-Taixé, Laura; Fenzi, Michele; Kuznetsova, Alina; Savarese, Silvio; Rosenhahn, Bodo
Main reference L. Leal-Taixé, M. Fenzi, A. Kuznetsova, B. Rosenhahn, S. Savarese, "Learning an image-based motion context for multiple people tracking," in Proc. of the 2014 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'14), pp. 3542–3549, IEEE, 2014.
URL http://dx.doi.org/10.1109/CVPR.2014.453

In this talk, I presented a novel method for multiple people tracking that leverages a generalized model for capturing interactions among individuals. At the core of our model lies a learned dictionary of interaction feature strings which capture relationships between the motions of targets. These feature strings, created from low-level image features, lead to a much richer representation of the physical interactions between targets compared to hand-specified social force models that previous works have introduced for tracking. One disadvantage of using social forces is that all pedestrians must be detected in order for the forces to be applied, while our method is able to encode the effect of undetected targets, making the tracker more robust to partial occlusions. The interaction feature strings are used in a Random Forest framework to track targets according to the features surrounding them. Results on six publicly available sequences show that our method outperforms state-of-the-art approaches in multiple people tracking.

3.14 Image Annotation – From Machine Learning to Machine Teaching

Oisin Mac Aodha (University College London, GB)

Joint work of Mac Aodha, Oisin; Campbell, Neill DF; Johns, Edward; Kautz, Jan; Brostow, Gabriel J
Main reference O. Mac Aodha, N. D. F. Campbell, J. Kautz, G. J. Brostow, "Hierarchical subquery evaluation for active learning on a graph," in Proc. of the 2014 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'14), pp. 564–571, IEEE, 2014
URL http://visual.cs.ucl.ac.uk/pubs/graphActiveLearning/index.html

Current state-of-the classification and detection algorithms used in scene understanding require large quantities of labeled training data. Typically, these datasets are manually annotated–at a great expense, both in terms of time and money. The goal of Active Learning is to reduce this effort by only requesting labels for the most informative data. In practice however, many Active Learning strategies are computationally inefficient, rendering them useless for interactive labeling. Another problem arises if the human annotator does not possess the knowledge or experience required to correctly label the data. This problem exists in many domains, such as species classification and image geolocalization, where fine-grained visual discrimination is essential for effective image understanding.

In this talk I will present an algorithm for efficient Active Learning in the context of semi-supervised graph based learning. Inspired by this, I will also present very recent work that aims to improve the ability of annotators by turning the human into the learner, and the algorithm into the teacher.

3.15 Structured Models for Recognition: Towards Sub-Category and Interaction Discovery

Greg Mori (Simon Fraser University–Burnaby, CA)

Joint work of Mori, Greg; Khodabandeh, Mehran; Lan, Tian; Vahdat, Arash; Zhou, Guang-Tong; Kim, Ilseo; Oh, Sangmin; Sigal, Leonid

Visual recognition involves reasoning about structured relations between objects at multiple levels of detail. For example, human behaviour analysis requires a comprehensive labeling covering individual low-level actions to pair-wise interactions through to high-level events. Scene understanding can benefit from considering visual sub-categories and their relations. In this talk I will present structured models for scenes and group activity recognition, with holistic analysis of people interacting and taking different social roles. I will describe our work on latent max-margin clustering as a means to discover sub-categories of objects and types of interactions between people.

3.16 Towards Holistic Scene Understanding: low-level cue extraction, collective activity classification, and behavior recognition

Vittorio Murino (Italian Institute of Technology–Genova, IT)

Joint work of Murino, Vittorio; Cristani, Marco; Del Bue, Alessio; Nabi, Moin; Crocco, Marco; Zanotto, Matteo; et al, et al;

In this talk I'll address the scene understanding problem from the perspective of analyzing and to figure out the human behavior. Humans are essentially a social species and interactions among conspecifics is one essential characteristics of the life: understanding such interplays constitutes an interesting yet important issue which can lead to effective applications and to discover diverse basic insights. Considering the "classic" computer vision paradigm, understanding behavior implies the design of a data processing pipeline composed by a high-level reasoning process which needs of low-level cues or "features". Such cues are, in this cases, detection, tracking and other algorithms able to identify a person and track over time while extracting other important features like, for instance, face orientation or gaze. Actually, tackling the understanding problem from a "social" perspective, means to look for the so called "social signals" which are formed by a set of nonverbal cues which have a social meaning. They can be conveyed in many way, primarily by means of the voice, gesture, posture, face and gazing, in which also the environment (e.g., the spatial location of people wrt the scene geometry) has a role.

In this context, I will briefly describe a couple of algorithms developed for low level cue extraction, namely 1) a face detection and orientation technique based on a powerful statistical descriptor (covariance, entropy and mutual information) associated to a boosting-based classification method, and 2) a group tracking algorithm based on the decentralized particle filtering and dirichlet process mixture model able to manage any number of groups and (group) split and merge events.

I will also show a couple of methods for high-level reasoning process. First, I will show an algorithm based on a new time-based descriptor, the temporal poselet, which is based on poselet detectors extended to time, and able to detect people assemblies and classify collective activities. Second, I will address the problem of classifying mouse behavior. Starting from the outcome of an automatic algorithm able to track mice in a cage and classify atomic action frame by frame, we proposed a technique based on a Bayesian Nonparametric approach able to spot higher-level patterns of behaviors, that is a latent sequence of atomic actions which are not visible to neuroscientists but potentially useful to "understand" the genetically modified mouse behavior wrt control mice.

Finally, I will conclude the talk presenting a different analysis of a scene based on the acoustics. We build an acoustic camera (coupled with an optical one) able to sense "acoustically" the environment and identify relevant acoustic sources, classify the sound received and track targets of interest. This constitutes a complementary approach wrt to vision, indeed necessary and useful to reach a full understanding of a scene.

3.17 Interactions Between Scene Elements

Caroline Pantofaru (Google Inc.–Mountain View, US)

The physical world consists of interrelated objects, spaces, scenes and activities, whose positions in 3D all influence each other. When we see a scene we naturally understand which object is in front of another, which objects are grouped together to form a functional unit, how these objects relate to their space and what kind of space it is. At a glance we also understand people's spatial interactions and use this information to determine groups, movement patterns and activities. In visual media we understand where the cameras are positioned and how it affects their representation of the world. In this talk I discussed methods for using 3D reasoning and the relationships between semantic scene elements to improve our understanding of the visual world.

References

1 W. Choi, Y. Chao, C. Pantofaru, and S. Savarese. *Indoor Scene Understanding with Geometric and Semantic Contexts.* International Journal of Computer Vision, 2014.
2 W. Choi, Y. Chao, C. Pantofaru, and S. Savarese. *Discovering Groups of People in Images.* European Conference on Computer Vision, 2014.

3.18 Temporally Consistent Superpixels

Matthias Reso (Leibniz Universität Hannover, DE)

Joint work of Reso, Matthias;Jachalsky, Jörn; Rosenhahn, Bodo; Ostermann, Jörn
Main reference M. Reso, J. Jachalsky, B. Rosenhahn, J. Ostermann, "Temporally Consistent Superpixels," in Proc. of the 2013 IEEE Int'l Conf. on Computer Vision (ICCV'13), pp. 385–392, IEEE, 2013.
URL http://dx.doi.org/10.1109/ICCV.2013.55

Superpixel algorithms represent a very useful and in- creasingly popular preprocessing step for a wide range of computer vision applications, as they offer the potential to boost efficiency and effectiveness. This talk presents a highly competitive approach for temporally consistent superpixels for video content. The approach is based on energy-minimizing clustering utilizing a novel hybrid clustering strategy for a multi-dimensional feature space working in a global color space and local spatial spaces. For a thorough evaluation the proposed approach is compared to state of the art supervoxel and video superpixel algorithms using established benchmarks and shows a superior performance.

References

1 Matthias Reso, Jörn Jachalsky, Bodo Rosenhahn and Jörn Ostermann. *Superpixels for Video Content Using a Contour-based EM Optimization*. ACCV, 2014

3.19 Describing Videos with Natural Language

Anna Rohrbach (MPI für Informatik–Saarbrücken, DE)

Joint work of Rohrbach, Anna; Rohrbach, Marcus; Qiu, Wei; Friedrich, Annemarie; Pinkal, Manfred; Schiele, Bernt
Main reference A. Rohrbach, M. Rohrbach, W. Qiu, A. Friedrich, M. Pinkal, B. Schiele, "Coherent Multi-sentence Video Description with Variable Level of Detail," In Proc. of the German Conference on Pattern Recognition (GCPR'14), LNCS, Vol. 8753, pp. 184–195, Springer, 2014.
URL http://dx.doi.org/10.1007/978-3-319-11752-2_15

Generating descriptions for video is an interesting task requiring core techniques of computer vision and computational linguistics. Existing approaches for automatic video description focus on generating single sentences at a single level of detail. We address both of these limitations: for a variable level of detail we produce coherent multi-sentence descriptions of complex videos featuring cooking activities [1]. We follow a two-step approach where we first learn to predict a semantic representation (SR) from video and then generate natural language descriptions from it. For our multi-sentence descriptions we model across-sentence consistency at the level of the SR by enforcing a consistent topic. To understand the difference between detailed and short descriptions, we collect and analyze a video description corpus with three levels of detail.

To foster the research on automatic video description we propose a new MPII Movie Description Dataset [2], featuring movie snippets aligned to scripts and DVS (Descriptive video service). DVS is a linguistic description that allows visually impaired people to follow a movie. We benchmark state-of-the-art computer vision algorithms to recognize scenes, human activities, and participating objects and achieve encouraging results in video description on this new challenging dataset.

References

1 A. Rohrbach, M. Rohrbach, W. Qiu, A. Friedrich, M. Pinkal, and B. Schiele. *Coherent multi-sentence video description with variable level of detail*. In Proceedings of the German Conference on Pattern Recognition (GCPR), 2014.
2 A. Rohrbach, M. Rohrbach, N. Tandon, and B. Schiele. *A dataset for movie description*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

3.20 Towards 3D scene understanding: how to assemble the pieces?

Konrad Schindler (ETH Zürich, CH)

Computer vision has made great progress. Individual low- and mid-level components like object detection, 3D reconstruction etc. are working well– yet, we have not managed to connect them into convincing "scene understanding" systems, and it is unclear how to tackle that next challenge. I argue that this might be in part due to the fact that we do not know what lower-level input we need for scene understanding. I will look at different basic visual functions and ask what information they should deliver in order to be useful for high-level semantic understanding. The talk will not provide answers, but will hopefully trigger a discussion about what individual pieces of the vision pipeline need to deliver to enable "scene understanding".

3.21 "At-a-glance" Visualization of Highlights in Raw Personal Videos

Min Sun (National Tsing Hua University – Hsinchu, TW)

Joint work of Sun, Min; Farhadi, Ali; Seitz, Steve Seitz
Main reference M. Sun, A. Farhadi, S. Seitz, "Ranking Domain-specific Highlights by Analyzing Edited Videos," in 13th Europ. Conf. on Computer Vision (ECCV'14), LNCS, Vol. 8689, pp. 787–802, Springer, 2014.
URL http://dx.doi.org/10.1007/978-3-319-10590-1_51
Main reference M. Sun, A. Farhadi, B. Taskar, S. Seitz, "Salient Montages from Unconstrained Videos," in 13th Europ. Conf. on Computer Vision (ECCV'14), LNCS, Vol. 8689, pp. 472–488, Springer, 2014.
URL http://dx.doi.org/10.1007/978-3-319-10584-0_31

Nowadays, people share tons of images and videos online. However, raw personal videos typically will rarely be watched again, since they are long and boring most of the time. Our research focuses on generating an "at-a-glance" visualization of highlights in raw person videos. We have two related work for this project. Our first work, "Ranking Domain-specific Highlights by Analyzing Edited Videos" focuses on automatically finding highlights in raw personal videos. Our second work, "Salient Montages from Unconstrained Videos" focuses on automatically generating an at-a-glance visualization (referred to as salient montages) of highlights in raw personal videos.

3.22 Detecting conversational groups: a game-theoretic approach with sociological and biological constraints

Sebastiano Vascon (Italian Institute of Technology–Genova, IT)

Joint work of Vascon, Sebastiano; Mequanint Z., Eyasu; Cristani, Marco ; Hung, Hayley; Pelillo, Marcello; Murino, Vittorio
Main reference S. Vascon, E. Zemene Mequanint, M. Cristani, H. Hung, M. Pelillo, V. Murino, "A Game-Theoretic Probabilistic Approach for Detecting Conversational Groups," in Proc. of the 12th Asian Conf. on Computer Vision (ACCV'14), LNCS, Vol. 9007, pp. 658–675, Springer, 2014.
URL http://dx.doi.org/10.1007/978-3-319-16814-2_43

In the last decade the problem of detecting groups of people in a scene is gaining increasing interest due to its importance in many fields like video surveillance, social signal processing, scene understanding and social robotics to cite a few. In this talk, I have presented a recently published work to detect groups of persons that are conversing. The modelling of such groups is not a trivial task because persons create groups respecting certain biological and sociological constraints. We proposed a game-theoretic framework to detect groups in which these constraints are both satisfied, outperforming the state of the art.

References

1 Sebastiano Vascon, Eyasu Zemene Mequanint, Marco Cristani, Hayley Hung, Marcello Pelillo, Vittorio Murino. *A Game-Theoretic Probabilistic Approach for Detecting Conversational Groups*. The 12th Asian Conference on Computer Vision (ACCV 2014), Singapore, 2014

3.23 Understanding scenes on mobile devices

Stefan Walk (Qualcomm Austria Research Center GmbH, AT)

Understanding a scene is an important step towards being able to interact with the real world on mobile devices. Seeing recognized objects in their scene context is a prominent example for this. In this talk I will show how problems like this are solved at Qualcomm. Solutions include being able to interact with unknown objects on a plane initialized from a known object, geometrically understanding the planes that a scene is composed of, and being able to track a scene when the object of interest temporarily leaves the view. I will also highlight problems that occur in real-world settings when bringing computer vision algorithms from research to commercial products.

3.24 Structured prediction in Remote Sensing: The key to automated cartographic mapping?

Jan Dirk Wegner (ETH Zürich, CH)

Joint work of Wegner, Jan Dirk; Montoya, Javier; L'ubor Ladický, Konrad Schindler
Main reference J. A. Montoya, J. D. Wegner, L. Ladický, K. Schindler, "Mind the gap: modeling local and global context in (road) networks," in Proc. of the 36th German Conf. on Pattern Recognition (GCPR), LNCS, Vol. 8753, pp. 212–223, Springer, 2014; pre-print available from author's webpage.
URL http://dx.doi.org/10.1007/978-3-319-11752-2_17
URL http://www.igp.ethz.ch/photogrammetry/publications/pdf_folder/montoya-etal-roads-gcpr14.pdf

The interpretation of remote sensing images is still done manually today, which is very costly in terms of time and money. Our long term goal is to completely automate this task. In a first step towards automated cartographic mapping we focus on the extraction of road networks that we view as the "structural backbone" of cities.

We propose a method to label roads in aerial images and extract a topologically correct road network. Three factors make road extraction difficult: (i) high intra-class variability due to clutter like cars, markings, shadows on the roads; (ii) low inter-class variability, because some non-road structures are made of similar materials; and (iii) most importantly, a complex structural prior: roads form a connected network of thin segments, with slowly changing width and curvature, often bordered by buildings, etc. We model this rich, but complicated contextual information at two levels. Locally, the context and layout of roads is learned implicitly, by including multi-scale appearance information from a large neighborhood in the per-pixel classifier. Globally, the network structure is enforced explicitly: we first detect promising stretches of road via shortest-path search on the per-pixel evidence, and then select pixels on an optimal subset of these paths by energy minimization in a CRF, where each putative path forms a higher-order clique. The model outperforms several baselines on two challenging data sets, both in terms of precision/recall and w.r.t. topological correctness.

3.25 Sparse Optimization for Motion Segmentation

Michael Yang (Leibniz Universität Hannover, DE)

Motion segmentation aims to decompose a video sequence into different moving objects that move throughout the sequence. In this talk, I will show some state-of-the-art motion segmentation methods and our new framework based on subspace clustering with sparse optimization. We combine two sparse representations to optimize both the global and local estimation. Sparse PCA is applied for the global optimization. The local subspace separation is achieved via automatically selecting the sparse nearest neighbours. In the end of this talk, I will show some experimental results on the Hopkins 155 Dataset and Freiburg- Berkeley Dataset.

3.26 Gesture Recognition Portfolios for Personalization

Angela Yao (Universität Bonn, DE)

Joint work of Yao, Angela; Van Gool, Luc; Kohli, Pushmeet
Main reference A. Yao, L. Van Gool, P. Kohli, "Gesture Recognition Portfolios for Personalization," in Proc. of the 2014 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'14), pp. 1923–1930, IEEE, 2014; pre-print available from author's webpage.
URL http://dx.doi.org/10.1109/CVPR.2014.247
URL http://www.vision.ee.ethz.ch/~yaoa/pdfs/yao_cvpr2014.pdf

Human gestures, similar to speech and handwriting, are often unique to the individual. Training a generic classifier applicable to everyone can be very difficult and as such, it has become a standard to use personalized classifiers in speech and handwriting recognition. In this paper, we address the problem of personalization in the context of gesture recognition, and propose a novel and extremely efficient way of doing personalization. Unlike conventional personalization methods which learn a single classifier that later gets adapted, our approach learns a set (portfolio) of classifiers during training, one of which is selected for each test subject based on the personalization data. We formulate classifier personalization as a selection problem and propose several algorithms to compute the set of candidate classifiers. Our experiments show that such an approach is much more efficient than adapting the classifier parameters but can still achieve comparable or better results.

4 Working Groups

4.1 Working Group Summary on *What does "Scene Understanding" mean ?*

Bob Fisher

We had quite a wide-ranging discussion from what does a Rumba understand, to would a system need to be "Conscious" to do proper scene understanding. The discussion went around a lot, but could be categorised as belonging to the following 14 topics:

- Historical
  - Scene understanding has been around for c. 40 years at least, with the outdoor scene analysis of Hanson & Riseman's Visions system, and the SRI indoor scene labeling. Plus Brook's ACRONYM system in the early 1980s.
  - How much general progress have we made? What has limited the progress?
- What does "Scene Understanding" mean?
  - Is it general? Or maybe different tasks restrict their domain of focus/labels?
  - There seem to be many different tasks that require different sorts of SU.
  - There are different levels of SU: 1) A meeting, 2) a collection of interacting people, 3) a set of people and objects in a spatial configuration.
- Are there key classes of SU?
  - Static SU:
    - What type of scene is this?
    - What are the objects in this scene?

    * What are the relationships between the objects in the scene?
    * Are the objects and relationships consistent?
    * What are the subtleties of the scene (eg. aesthetics, tidiness, etc)?
  - Dynamic SU:
    * What type of action is this?
    * What/who is doing the action?
    * What is the sequence of sub-actions?
    * Is the set of actors, actions, and sequence consistent?
    * What are the subtleties of the actions (eg. for emotion/expression analysis, lie detection)?
- What quality of sensors do we need?
  - Is it task dependent, or do we want as much as we can get? Would this get us better data, but not better understanding?
    * Faster sampling, more pixels, better depth resolution, more spectral channels, more bits to the data?
  - Could we do everything we want if we had a great low-level 2.5D+RGB video sensor? As an alternative to struggling with impoverished data, letting us focus on just the SU task?
  - Is it cheating to use specialised sensors? Eg. hyperspectral sensors for material classification? Or is it sensible to use the sensor that allow optimal SU performance?
- What is the role of temporal data?
  - Do we need it for static scene analysis?
- What representations do we need for SU?
  - Do we need all of high-level vision?
  - Should the representations be for more general-purpose SU, or should there be many different reps, each tuned for more focussed competences?
  - Does static SU require only 2D models, or 2.5/3D models?
  - Do these models need to be generative, or will discriminative do?
- Learning the content in the representations needed to do SU
  - How much "prior" or "common sense" knowledge should be innate?
  - Should we expect to learn the variety of the world from many examples, or from only a few?
  - Do we need to be able to act to learn, or can we be passive?
  - Do we need an oracle/teacher/human in the loop?
- What databases do we need to underpin SU?
  - By analogy to WordNet and ImageNet, do we need an ActivityNet and SpatialRelationNet and ObjectContextNet?
- What technologies are needed for SU?
  - Would a CNN do it all? What would be the outputs? What would be the inputs?
  - Would we need an excessive amount of hand-labeled training data (images and videos)?
  - Should we be using simulators to generate training data? At least at the initial stages? Or at just the higher levels, symbolic rather than signal level? A lot of controversy here.
- What is the Roadmap to competent SU?
  - We can do a few objects, relations and activities now.
  - There should be some serious work on a Roadmap.
- What are some of the key problems that limit better SU?
  - Imprecision of language for describing scenes

  - Many truths/many equally valid descriptions of the same scene/image.
  - Descriptions can lie at many levels.
  - Are we biased by the closed-world assumption?
  - The natural variation of the real-world. And how much of this do we need to model (which depends on the task).
  - The breadth of possible applications for SU of the same scene.
- How can we measure progress in SU?
  - Research moves around from topic to topic, but there are not good benchmarks.
  - Is the goal binary (can we do it at all?) or graded (can we do it some of the time, or be partly accurate)? How do we assess this?
  - There seems to be a large number of results that we can expect from SU. How can we tell if/what is making progress? Some results are focussed and easy to assess (How many people are in this scene?), others are open-ended (What is happening in this scene?).
  - It's a bit like a Turing Test (or Watson).
  - Maybe some assessment of whether SU is happening could be by question-oriented tests:
    * What happens next?
    * What must have happened between these 2 events?
    * Is there anything out-of-place in this scene?
    * Did anyone do something unusual?
  - Would have to avoid Eliza-like behaviour.
- Benchmarks to assess and drive progress?
  - What should they be?
  - Should they be wide ranging for general SU? Or focussed to assess specific competences?
  - The risk is developing programs to solve benchmark, not the real problem.
- Do we really need SU?
  - Yes, otherwise we might lose our jobs.
  - More seriously: Is it useless because it does not solve a task?
  - Can we hope to have a sufficiently competent general SU capable of answering any question about a scene/video?
  - Or do we instead need a set of highly competent specialised capabilities, for the 100 (?) general tasks that most people have to do.
  - Is it SU only as long as we cannot get a machine to do it, and thereafter it's simply automation?

4.2 Working Group Summary on *Dynamic Scene: Humans*

Greg Mori

Summary of topics/open questions:

- We need a big collection of data with good sampling. People are a unique object, will have lots of data but not enough labeled.
- We need to interpret groups of people from crowds as a texture down to individuals.
- We need to think about occlusion when detecting these people.
- You need to decide what you want to recognize, then you can build it?

- Can we build a semantic space of actions to rival WordNet? Bottom-up clustering will be hard, we don't have the semantic information from something like WordNet.
- Ask people to provide descriptions at multiple levels of detail when viewing an action.
- Noun and verb only? Probably not, need longer-term annotations to go to the story-level.

Data
- Granularity of representation for people
- From individuals to groups and a crowd of people
- 2d boxes
- joint angles and positions
- Annotate with everything or not
  - What can we do with lots of annotated data ?

Detection
- Rescue robotics with unusual poses, non-moving people
- Thermal cameras, multi-sensor input
- Face detection is almost solved, but not non-pedestrian people
- Dataset
  - Imagenet for humans?
  - Should we have a dataset of varied poses? Is this the main thing needed?
- What about representation?
  - Parts are good, lots of people in a crowd, just detect heads?
  - Poselets?
  - Do we need a 3d representation at all?
- Motion?
  - Range of pose in videos is larger than images (collected)
  - Should we segment the people from the videos
  - Maybe motion is not so important, we can focus on just detecting people in images?
  - If we have the video can we try prediction at any time scale
  - Subtle motion is very important in surveillance, this is hard to deal with
  - Temporal information is often just used for smoothing, not enough modeling of time
  - Motion is dependent on scale, at a far distance, it is hard to see the motion, but close-up you can segment
  - Partial occlusion of people in a dataset would be very useful
    - A few % improvement by using templates for 2 people, perhaps a little more for 3 people (Schiele et al.)
  - Label occlusion for people in a dataset
  - Interactions with objects for detection (Fua et al. ECCV14)
  - How many images do you think we need for a "good" human detector in a city scene (any photo taken from a hand-held camera)
    - Millions?
    - Stratified sampling is important
    - Rarity in the benchmark is important, need to analyze different categories
    - How do we collect such a dataset, there are lots of different cases and there is not a Wordnet equivalent to organize it
    - Maybe active learning / clustering / exemplar SVM is needed to build this
    - Could do this based on 3d body configuration representation
      - Sample possible configurations
    - PeopleNet

    * How do you label bounding boxes on these people, e.g. people on stilts
    * Large diversity is important, videos are not so good for diversity
    * Active learning approach for collecting a dataset, pay people on AMT who come up with images which are false negatives for the current algorithm
- Detecting individuals is not all there is
  - Some applications don't need person-centric or individual representations
  - Crowds of people moving
    * Is this dynamic texture?
    * Emergence of individuals from the flock, flows from the crowd
    * Phase changes
  - Do you want to find abnormal behaviour of one person in the crowd?
  - The scale of the people seems to dictate this
    * Once the people are too small we just model as a crowd
- PeopleNet could have gestures and videos
  - Go to actions of individuals

Tracking
- Evaluation of tracking for the purpose of activity analysis
  - Measuring tracklet importance using cameraman's gaze
- Hands and gesture vs. human bounding box tracking
  - Landmarks on body or a generic tracker that works on any objects

Actions
- What are the important actions to recognize
- Jump, wave hands, walk
  - Are these important?
  - Labeling crowds in hockey stadiums
  - Lists of postures and poses that are important to psychologists
  - Need help from psychologists
  - We don't have the equivalent of WordNet for actions
    * Derive it by looking at verbs in language
    * Some are not actionable
  - Given a frame, label it dangerous / not dangerous
  - Psychologists' representations should be latent variables
- These 4 problems are intertwined, but not all tracklets of all people are equally important
- Starting and ending point of an action

Group activities
- The scale of inference
  - When is motion necessary, treat the group as a texture or a whole
  - When is it necessary to reason about individuals
- A project on combining holistic crowd representations with the actions of individuals and multi-scale representations
- With images of a crowd vs. videos of a crowd, some things are not possible
  - But is this a natural task?
- Let's say you have unreliable tracking, and just get some people tracked but lots of incomplete tracklets, some false positives
  - Can you infer group activities?
  - Can we cluster or find dominant patterns?
  - Clusters of optical flow?

- How do we build systems that are robust across density and scale of people/vehicles
  - When you have rows of cars on a busy street, too hard to detect individuals, just represent optical flow
- Does looking at a group help understand occlusions / tracking for an individual?
- Should we learn these together or separately

Long-term story

- People in the event give knowledge about what is important (record gaze)
- Rich data includes video meta-data which give high-level information about the purpose of a video
  - Sports videos are probably good for this
  - Surveillance videos are bad
- Video summarization?
  - No, it's interpolate beyond what you can see
  - No, it's about in textual domain what happened
  - Yes, show me the beginning, middle, end of a story
- The longer activities, bigger pieces
  - What is happening beyond just one image or a short video
  - Describe and infer things about what is happening in a scene
  - Not everything is actually observed
  - How do you expand to a larger view
- Cooking videos, what was being prepared, exhaustively detecting all the details
- Inferring intent from few observations
- Temporal order is not fixed, events can take place in different orders
- What are the key, dominant phases in an activity
- Problems in the editing of the domain, e.g. cooking videos are edited to show only interesting clips
- How do we judge success for this?
- Summarize large video collections
  - Captions and titles are very useful
- Text is a better representation than video
  - But dogs can't cook omelettes
- Task is to ignore the mundane bits but focus on the parts of the video that people haven't seen often
- Look for 2 goals in 90 minutes of soccer
  - Should you include the boring stuff in the summary
  - E.g. include the boring parts of a game in the summary or not?
  - Storytelling constrained by the low-level detections we can do now
- How do I go from chopping onions to making soup
- How do we get datasets that contain more than just a clip-let
- Produce as much information in a summary of 4000 characters so that a cook can produce the desired output
- Use clips with (noisy) meta-data to generate a summary
- It's about how to give an instruction not just summarize a video
- Figure out what a person did over a whole day
- Build datasets with long clips that will eventually contain the events you want
  - Lots of sports videos
  - Are egocentric videos good for this?
- Analyze what people are doing over a longer time span, not just what someone is doing right now

4.3 Working Group Summary on *Recognition in static scenes (in 3D)*

Alessio Del Bue

Scene understanding has the goal of building machines that can see like humans to infer general principles and current situations from imagery. More detailed, the ultimate goal is to let a machine pass a "visual Turing challenge", e.g. to allow arbitrary questions from a human and to obtain useful answers. Sitting in a music room, there was a longer discussion about the height of a plant standing on a window ceil with a radiator below.

Starting from a single image, pointing to a single pixel, asking a straight (single sentence) question ("How tall is this?"), in a few years, the system will identify and segment the object, estimate the context and reach some conclusion that the height of the plant is xy-cm +/- a variance. Due to ambiguities, occlusions or missing information, the system should be allowed to infer to provide uncertain answers and to recheck with the human to increase the certainty. Questions can be roughly arranged into several categories, e.g. extrapolation, interpolation or straight) and answers such as category (discrete) or continuous.

- Data/Example collection: Use a robot in a room and an open internet to collect questions
- Alternative : Show random images and let people ask random questions and get people to answer them ...
- → Classify questions and learn the manifold of questions (Counting, Measuring, stupid q.)
- Parsing methods for questions (Grammars), Analysis/Rephrasing to canonical from.
- E.g. question + recognition → Text description + query expansion (cloud / big data).
  - Generative / discrimative learning, priors for regression ?
  - ...

Still problematic: Current accuracies of naïve methods are in the order of 7–12 %, the Human Baseline is 60%. Are we humans really "that good" ?

There is still the Question of metrics unsolved. For quantitative evaluation should be in relation to the quality humans can reach.

5 Schedule

Monday, February 15th, 2015

09:15–09:30	Bodo Rosenhahn: Opening
09:30–10:00	1 minute self-presentations
Chair:	Michael Yang
10:45–11:15	Anna Rohrbach: *Describing Videos with Natural Language*
11:15–11:45	Sebastiano Vascon: *Detecting conversational groups: a game-theoretic approach with sociological and biological constraints*
11:45–12:00	Discussions
Chair:	Jürgen Gall
14:00–14:30	Gabriel Brostow: *Superhuman-in-the-Loop Computer Vision*
14:30–15:00	Sanja Fidler: *Understanding Complex Scenes and People That Talk about Them*
15:00–15:30	Oisin Mac Aodha: *Image Annotation–From Machine Learning to Machine Teaching*
Chair:	Anna Rohrbach
16:00–16:30	Marco Cristani: *From groups to crowds: a social signal processing perspective*
16:30–17:00	Konrad Schindler: *Towards 3D scene understanding: how to assemble the pieces?*
17:00–17:30	Caroline Pantofaru: *Interactions Between Scene Elements*

Tuesday, February 16th, 2015

Chair:	Stefan Walk
09:00–09:30	Laura Leal-Taixé: *Learning an image-based motion context for multiple people tracking*
09:30–10:00	Michael Yang: *Sparse Optimization for Motion Segmentation*
Chair:	Jan Wegner
10:15–10:45	Angela Yao: *Gesture Recognition Portfolios for Personalization*
10:45–11:15	Roberto Henschel: *Solving Multiple People Tracking Using Minimum Cost Arborescences*
11:15–11:45	Greg Mori: *Structured Models for Recognition: Towards Sub-Category and Interaction Discovery*
11:45–12:00	Discussions
14:00–15:00	Jiří Matas, Vittorio Murino: Working Group Definition
15:30–18:00	Working Group Meeting

Wednesday, February 17th, 2015

Chair: Roberto Henschel

09:00–09:30 Vittorio Murino: *Towards Holistic Scene Understanding: low-level cue extraction, collective activity classification, and behavior recognition*

09:30–10:00 Bob Fisher: *Fish detection, tracking, recognition and analysis with the Fish4Knowledge Dataset*

Chair: Oisin Mac Aodha

10:15–10:45 Michele Fenzi: *Feature Regression-based Pose Estimation for Object Categories*

10:45–11:15 Jürgen Gall: *Human Pose: A Cue for Activities and Objects*

11:15–11:45 Abhinav Gupta: *Geometry, Function and Common Sense*

11:45–12:00 Discussions

14:00–18:00 Social Event: One group made a rip to Trier, and anothr group went on a 3h hike in the environment.

Thursday, February 18th, 2015

Chair: Jörn Jochalsky

09:00–09:30 Michal Havlena: *VocMatch: Efficient Multiview Correspondence for Structure from Motion*

09:30–10:00 Jan-Michael Frahm: *Understanding Scene Dynamics from Crowd Sourced Imagery*

Chair: Michele Fenzi

10:15–10:45 Ron Kimmel: *A Spectral Perspective on Invariant Measures of Shapes*

10:45–11:15 Matthias Reso: *Temporally Consistent Superpixels*

11:15–11:45 Jiří Matas: *History of Object Recognition: to be learned for Scene Understanding or a Scrapyard?*

11:45–12:00 Discussions

Chair: Michal Havlena

14:00–14:30 Reinhard Klette: *Visual Odometry and 3D Roadside Reconstruction*

14:30–15:00 Jan Wegner: *Structured prediction in Remote Sensing: The key to automated cartographic mapping?*

Chair: Matthias Reso

16:00–16:30 Min Sun: *"At-a-glance" Visualization of Highlights in Raw Personal Videos*

16:30–17:00 Alessio del Bue: *Semantic Motion Segmentation and 3D Reconstruction*

17:00–17:30 Raquel Urtasun: *Towards autonomous driving*

Friday, February 19th, 2015

Chair: Markus Enzweiler

09:00–09:30 Esther Hobert: *Generic Object Detection in Video using Saliency and Tracking*

09:30–10:00 Stefan Walk: *Understanding scenes on mobile devices*

10:15–10:45 Shaogang Gong: *Explore Multi-source Information for Holistic Visual Search*

11:00–11:55 Working Group Meeting and Summary

Participants

- Gabriel Brostow
University College London, GB
- Marco Cristani
University of Verona, IT
- Alessio Del Bue
Italian Institute of Technology – Genova, IT
- Markus Enzweiler
Daimler AG – Böblingen, DE
- Michele Fenzi
Leibniz Univ. Hannover, DE
- Sanja Fidler
University of Toronto, CA
- Bob Fisher
University of Edinburgh, GB
- Jan-Michael Frahm
University of North Carolina – Chapel Hill, US
- Jürgen Gall
Universität Bonn, DE
- Shaogang Gong
Queen Mary University of London, GB
- Abhinav Gupta
Carnegie Mellon University – Pittsburgh, US
- Michal Havlena
ETH Zürich, CH
- Roberto Henschel
Leibniz Univ. Hannover, DE
- Esther Horbert
RWTH Aachen, DE
- Jörn Jachalsky
Technicolor – Hannover, DE
- Ron Kimmel
Technion – Haifa, IL
- Reinhard Klette
University of Auckland, NZ
- Laura Leal-Taixé
ETH Zürich, CH
- Oisin Mac Aodha
University College London, GB
- Jiri Matas
Czech Technical University, CZ
- Greg Mori
Simon Fraser University – Burnaby, CA
- Vittorio Murino
Italian Institute of Technology – Genova, IT
- Caroline Pantofaru
Google Inc. – Mountain View, US
- Matthias Reso
Leibniz Univ. Hannover, DE
- Anna Rohrbach
MPI für Informatik – Saarbrücken, DE
- Bodo Rosenhahn
Leibniz Univ. Hannover, DE
- Bernt Schiele
MPI für Informatik – Saarbrücken, DE
- Konrad Schindler
ETH Zürich, CH
- Min Sun
National Tsing Hua University – Hsinchu, TW
- Raquel Urtasun
University of Toronto, CA
- Sebastiano Vascon
Italian Institute of Technology – Genova, IT
- Stefan Walk
Qualcomm Austria Research Center GmbH, AT
- Jan Dirk Wegner
ETH Zürich, CH
- Michael Yang
Leibniz Univ. Hannover, DE
- Angela Yao
Universität Bonn, DE

Report from Dagstuhl Seminar 15082

Limitations of Convex Programming: Lower Bounds on Extended Formulations and Factorization Ranks

Edited by

Hartmut Klauck[1], Troy Lee[2], Dirk Oliver Theis[3], and Rekha R. Thomas[4]

1 **Nanyang Technological University, SG,** cqthk@nus.edu.sg
2 **Nanyang Technological University, SG,** troyjlee@gmail.com
3 **University of Tartu, EE,** dirk.oliver.theis@ut.ee
4 **University of Washington, US,** rrthomas@uw.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 15082 "Limitations of convex programming: lower bounds on extended formulations and factorization ranks" held in February 2015. Summaries of a selection of talks are given in addition to a list of open problems raised during the seminar.

Seminar February 15–20, 2015 – http://www.dagstuhl.de/15082
1998 ACM Subject Classification G.1.6 Convex Optimization, F.2 Analysis of Algorithms and Problem Complexity
Keywords and phrases Convex optimization, extended formulations, cone rank, positive semidefinite rank, nonnegative rank, quantum communication complexity, real algebraic geometry
Digital Object Identifier 10.4230/DagRep.5.2.109

1 Executive Summary

Hartmut Klauck
Troy Lee
Dirk Oliver Theis
Rekha R. Thomas

The topic of this seminar was the rapidly developing notion of *cone rank* of a matrix/polytope that is an important invariant controlling several properties of the matrix/polytope with connections to optimization, communication complexity and theoretical computer science. This meeting was a follow-up to the 2013 Dagstuhl seminar *13082: "Communication Complexity, Linear Optimization, and lower bounds for the nonnegative rank of matrices"* organized by Leroy Beasley, Hartmut Klauck, Troy Lee and Dirk Oliver Theis.

The cone rank of a nonnegative matrix is an ordered notion of matrix rank with emerging applications in several fields. A well-known example is the nonnegative rank of a nonnegative matrix that appears in areas ranging from communication complexity, to statistics, to combinatorial optimization, and algebraic complexity theory. A related notion arising as an invariant in representations of convex sets as projections of affine slices of the positive semidefinite (psd) cone is the positive semidefinite (psd) rank. The psd rank is a very new quantity and is still relatively poorly understood.

The purpose of this seminar was to bring together researchers from optimization, computer science, real/convex/tropical algebraic geometry, and matrix theory, to discuss relevant techniques from each area that can contribute to the development of both the theory and computation of cone ranks and cone factorizations of nonnegative matrices, as well as their many emerging applications.

In optimization and computer science, a common approach to finding approximate solutions to NP-hard problems is to look at tractable convex relaxations of the problem as either a linear or semidefinite program. An optimal solution to such a relaxation gives a bound on the objective value of the original problem. While much previous work has focused on specific relaxations of a problem, or a family of relaxations coming from a hierarchy, cone ranks allow the study of the best possible linear, semidefinite, or other convex formulations of a NP-hard problem independent of specific construction methods. These formulations all write the underlying feasible set as the projection of an affine slice of a closed convex cone and is commonly refereed to as an extended formulation of the underlying feasible region. The nonnegative rank of a polytope is the smallest size of a linear extended formulation of the polytope while psd rank of the polytope is the size of the smallest possible semidefinite extended formulation of the polytope. Linear extended formulations are the best understood so far and an exciting development in this area is the recent breakthrough by Rothvoß showing that the matching polytope does not admit a polynomial sized linear extended formulation, settling a notorious open problem in combinatorial optimization. Very recently, there has also been exciting new developments in the area of psd rank such as the result of Lee, Raghavendra, and Steurer that shows that the psd rank of certain polytopes such as the traveling salesman polytope of a graph with n vertices must be exponential in n.

In the field of communication complexity, nonnegative and psd ranks are exactly characterized by a model of randomized and quantum communication complexity, respectively. This connection has allowed tools from communication and information theory to help create lower bounds for these ranks.

A central question in the field of real algebraic geometry is the semidefinite representability of convex sets. While polytopes only project to polytopes, affine slices of psd cones have much greater expressive power as their projections are convex sets, which allows the definition of psd rank for semi algebraic convex sets. Psd rank has inherent semi algebraic structure and its study crosses over into real and convex algebraic geometry, algebraic complexity, and semidefinite programming.

2 Table of Contents

3 Organization of the Seminar

One of the key goals of this seminar was to bring experts from all the different communities that intersect at the notion of cone ranks to inform the participants on the key tools and questions from each area that pertain to this topic. To achieve our goals of making the seminar part tutorial, we invited nine prominent researchers from a variety of areas to give expository lectures. These lectures were given by the following researchers listed along with the general topic of their talks:

1. Alexander Barvinok (University of Michigan): *Approximations of convex bodies by polytopes*
2. Greg Blekherman (Georgia Institute of Technology): *Sums of squares on the hypercube*
3. Hamza Fawzi (MIT): *Equivariant lifts of polytopes*
4. João Gouveia (University of Coimbra): *Survey on psd rank*
5. Volker Kaibel (University of Magdeburg): *Survey on nonnegative rank*
6. James Lee (University of Washington): *Lower bounds for SDP relaxations I*
7. Markus Schweighofer (University of Konstanz): *Real algebraic geometry*
8. David Steurer (Cornell University): *Lower bounds for SDP relaxations II*
9. Ronald de Wolf (CWI, Amsterdam): *Quantum communication complexity and psd rank*

Roughly two of these speakers were scheduled each day. In addition, there were short scientific talks by almost all the remaining participants as well as an open problems session on Wednesday evening.

4 Summaries of selected talks

In this section, we highlight some selected talks from the seminar. The first five are tutorial talks while the last is a sample research talk.

4.1 João Gouveia

João Gouveia gave the opening talk of the seminar which was a *survey on positive semidefinite rank*. The purpose of this talk was to get the audience primed for this topic which was a main feature at the seminar. The talk was based on the recent survey (with this title) that was written by Hamza Fawzi, João Gouveia, Pablo Parrilo, Richard Robinson and Rekha Thomas that is available on the arXiv.

The talk began with the definition of psd rank of a nonnegative matrix, basic properties and relationships to other common notions of rank. There was an instructive running example using circulant matrices that illustrated the main ranks that were introduced. Then he moved on to more sophisticated properties of psd rank such as the guarantee of factorizations of controlled norm. This result in particular played an important role in the recent work of Briët, Dadush and Pokutta who proved that 0/1 polytopes in a fixed dimension cannot all have small psd rank. Other features were also examined such as the space of factorizations, its connectivity properties, and symmetric factorizations. Several open questions were stated during the course of the talk which set the stage for the rest of the week. In particular, there was some discussion of the many open problems on the computational complexity of psd rank.

4.2 Greg Blekherman

Greg Blekherman gave another of our expository lectures on the topic *Symmetry Reduction for Sums of Squares on the Hypercube*.

The polytopes whose psd rank has attracted the most attention come from combinatorial optimization. In this setting, the polytope under consideration is the convex hull of a collection of vectors from $\{0,1\}^n$, where each vector denotes a subset of some ground set, and the problem is to minimize a polynomial function over it. This model covers many NP-hard problems and hence their psd rank is believed to be high. Yet, we have so far been able to only prove this for specific classes of polytopes such as cut polytopes (c.f. talks of James Lee and David Steurer).

The max cut problem which is NP-hard can be formulated as optimizing a quadratic polynomial over the entire n-dimensional hypercube. Thus polynomial optimization even over this simple-seeming 0/1-polytope is already very interesting and complicated. In max cut the polynomial to be optimized does not only have low degree, but is also symmetric under permutation of variables. One would expect that symmetric polynomial optimization has nice properties since one can bring to bear the power of the group action on the problem.

Greg's talk addressed precisely this situation. Given a symmetric polynomial $p(x)$ of low degree, he discussed the problem of writing it as a sum of squares polynomial modulo the equations of $\{0,1\}^n$. He explained a remarkably simple method to reduce this question to a univariate sum of squares problem on the "levels" of the hypercube which are the numbers $0,\ldots,n$. This is based on the observation that if we define $t := x_1 + \ldots + x_n$, then a symmetric polynomial p in the variables $x_1,\ldots,x_n$ is a polynomial in t.

Using this reduction he was able to derive simple proofs of previously known results such as certain quadratic polynomials can only be equal to sum of squares polynomials of high degree. The method promises to have more applications and Greg gave an excellent exposition of his methods on the black board at the meeting.

4.3 Markus Schweighofer

Markus Schweighofer was the main representative at this seminar from the field of *real algebraic geometry* which is the home of the theory of sums of squares polynomials. The organizers had requested from him a survey talk that would explain the real algebraic methods (via semidefinite programming) for solving polynomial optimization problems. Markus did precisely that. He developed the popular Lasserre hierarchy for solving polynomial optimization problems in an elementary way starting with systems of polynomial inequalities. He showed how the method is a natural consequence of introducing the right kind of combinations of existing inequalities followed by linearizations. This was instructive even to the experts. In the process, Markus introduced the important Positivestellensatz which underlies real algebraic geometry. This is a certificate for the infeasibility of a system of polynomial inequalities analogous to Hilbert's Nullstellensatz that provides a certificate for the infeasibility of systems of polynomial equations. Most of the audience was far from real algebraic geometry and this talk was both a friendly and natural introduction to the theorems in real algebraic geometry that contribute to polynomial optimization and the underlying convex sets. Markus went on to describe recent results which were more advanced, but the talk served well the purpose of providing the audience a taste of real algebraic geometry and its connections to psd rank.

4.4 Alexander Barvinok

In contrast to the setting of extended formulations, Alexander Barvinok talked about approximating a convex body $B \subset \mathbb{R}^d$ by a polytope $P \subset \mathbb{R}^d$ in the same dimensional space. In the work he discussed, based on the paper "Thrifty approximations of convex bodies by polytopes", the goal is to find a polytope P satisfying $P \subset B \subset \tau P$ with as few vertices as possible. He showed that if B is centrally symmetric, then for $\tau = 1+\epsilon$ for a small constant ϵ one can find a $1+\epsilon$ approximating polytope with $(\frac{1}{2\sqrt{\epsilon}}\ln(\frac{1}{\epsilon}))^d$ many vertices. This improves by about a square-root factor of ϵ over the standard volumetric argument for constructing an ϵ-net which gives $(1+2/\epsilon)^d$ many points. The proof has essentially two steps. The first step relies on John's theorem which says that for a centrally symmetric convex body B, the maximal volume ellipsoid E contained in B satisfies $E \subseteq B \subseteq \sqrt{d}E$. This is used to find a polytope with few vertices (only $O(d)$) that gives a weak $O(\sqrt{d})$ approximation to B. The second step amplifies the quality of the approximation, at the expense of adding more vertices to the approximating polytope. This is done by mapping B to a convex body $\widehat{B}$ in a higher dimensional space where it is argued using Chebyshev polynomials that a weak approximation to $\widehat{B}$ implies a very good approximation of B. Then the previous argument using John's theorem can be applied to $\widehat{B}$.

4.5 James Lee and David Steurer

One of the major motivating questions in the organization of this Dagstuhl seminar was to show superpolynomial lower bounds on the positive semidefinite extension complexity of an explicit polytope. This was achieved by James Lee, David Steurer, and Prasad Raghavendra in November 2014. James and David were in attendance at the seminar and treated us to 3 hours of lectures going through the proof in detail. At a high level, the proof "lifts" lower bounds against the Lasserre/sum of squares hierarchy to lower bounds for positive semidefinite extension complexity. In more detail, for a function $f : \{0,1\}^m \to \mathbb{R}_+$ they consider a matrix with rows indexed by subsets $S \subset [n], |S| = m$ and columns indexed by $x \in \{0,1\}^n$, where n is polynomial in m. The matrix $M_f(S,x) = f(x|_S)$ is shown to have psd rank that is exponential in the sum of squares degree of f.

The basic idea of the proof is the following. If f has sum of squares degree d, then there is a certificate of this fact known as a pseudoexpectation. This is a function that has negative correlation with f, but nonnegative correlation with any sum of squares polynomial of degree $< d$. The proof leverages this pseudodistribution into a functional that separates M_f from the set $\mathcal{S}$ of all matrices with small psd rank and that have the same ℓ_1 and ℓ_∞ norms as M_f. On M_f this functional is negative ($\leq -\epsilon$). However, it is shown that the functional is at least $-\epsilon/2$ on any matrix in $\mathcal{S}$. This is done in two steps. The first step, quantum learning, shows that if N has a small size factorization then the value of the functional on N is close to that of a matrix $\tilde{N}$ that has a factorization $N(S,x) = \mathrm{Tr}(A_S B_x)$ where B_x has bounded sums of squares degree as a matrix polynomial in x. The second step, based on random restrictions, shows that the value of the functional on a matrix with a bounded degree factorization can only be slightly negative $> -\epsilon/2$. Taken together this gives the separating functional as desired.

4.6 Samuel Fiorini

Samuel Fiorini talked about 2-level polytopes. A 2-level polytope is a polytope P where for each facet F there is a parallel facet F' such that $F \cup F'$ contains all vertices of P. Thus the slack matrix of P, after scaling the columns, is a matrix with all entries 0 or 1. As such 2-level polytopes give nice examples of nontrivial boolean matrices of low rank. Fiorini raised the very interesting question of the "log rank conjecture" for 2-level polytopes. This is the question: for the slack matrix of a 2-level polytope is the deterministic communication complexity at most a polynomial in the logarithm of the rank of the matrix? A relaxation of this question mentioned by Fiorini is to show that the extension complexity of any d-dimensional 2-level polytope is at most $2^{\mathrm{polylog(d)}}$. So far the best result known follows from Lovett's results on the log rank conjecture that shows a bound of $2^{\sqrt{d}\log(d)}$.

5 Overview of Talks

5.1 Thrifty approximations of convex bodies by polytopes

Alexander Barvinok (University of Michigan – Ann Arbor, US)

Main reference A. Barvinok, "Thrifty Approximations of Convex Bodies by Polytopes," International Mathematical Research Notices, Vol. 2014(16):4341–4356, 2013; pre-print available as arXiv:1206.3993v2 [math.MG].
URL http://dx.doi.org/10.1093/imrn/rnt078
URL http://arxiv.org/abs/1206.3993v2

This is a survey talk on how well a convex body can be approximated by a polytope with a given number of vertices. We measure the quality of approximation with respect to the Banach-Mazur distance and its versions (that is, given a convex body and an inscribed polytope, by what factor the polytope should be dilated to contain the body) and consider both fine (factors close to 1) and coarse (large factors) approximations.

5.2 Preservers of Completely Positive Matrix Rank

LeRoy B. Beasley (Utah State University, US)

Let $M_{m\times n}(R)$ denote the set of m-by-n matrices with entries in R. We write $M_{m\times n}(R_+)$ to denote the subsets of matrices, all of whose entries are nonnegative. Let $S_n(R)$ denote the set of all n-by-n real symmetric matrices. A matrix $A \in S_n(R)$ is said to be completely positive if there is some matrix $B \in M_{n\times k}(R+)$ such that $A = BB^t$. The CP-rank of the matrix A is the smallest k such that $A = BB^t$ for some $B \in M_{n\times k}(R+)$. In this article we shall investigate the linear operatros on $S_n(R)$ that preserve sets of matrices defined by the CP- rank. We classify those that preserve the CP-rank function, those that preserve the set of CP-rank-1 matrices , those that preserve the sets of CP-rank-1 matrices and the set of CP-rank-2 matrices, and those that strongly preserve the set of CP-rank-1 matrices.

5.3 Symmetry reduction for sums of squares polynomials on the hypercube

G.eg Blekherman (Georgia Institute of Technology – Atlanta, US)

Main reference G. Blekherman, J. Gouveia, J. Pfeiffer, "Sums of Squares on the Hypercube," arXiv:1402.4199v1 [math.AG], 2014.
URL http://arxiv.org/abs/1402.4199v1

Let p be a symmetric polynomial, i.e. a polynomial fixed under permutations of variables, and let H be the discrete hypercube $\{0,1\}^n$. The question of whether p is a sum of squares of polynomials of low degree on H can be reduced to a univariate sum of squares problem. I will present this reduction and explain how some known results on the Lasserre sum of squares hierarchy on H easily follow from it.

5.4 Rescaling PSD factorizations

Daniel Dadush (CWI – Amsterdam, NL)

I will present a short proof that any PSD factorization of a matrix M can be rescaled so that the operator norm of each matrix in the factorization can be bounded by a function of the maximum entry size and rank of the factorization.

5.5 2-level polytopes

Samuel Fiorini (University of Brussels, BE)

A polytope is called 2-level if for each facet-defining hyperplane H, the vertices that are not on H all lie in a hyperplane that is parallel to H. A polytope is 2-level iff it has a binary slack matrix. These polytopes appeared e.g. as a solution of a problem of Lovász in Gouveia, Parrilo and Thomas (2010), and earlier under the name of compressed polytopes in works of Stanley (1980) and Sullivant (2006). In this talk I will give motivations to study 2-level polytopes, survey some results about their structure and state many open problems about them.

5.6 Numerical Computation of Nonnegative and PSD Factorizations

Nicolas Gillis (University of Mons, BE)

Main reference A. Vandaele, N. Gillis, F. Glineur, D. Tuyttens,"Heuristics for Exact Nonnegative Matrix Factorization," arXiv:1411.7245v1 [math.OC], 2014.
URL http://arxiv.org/abs/1411.7245v1

Nonnegative (resp. positive semidefinite–PSD) factorizations allow to compute linear (resp. semidefinite) extended formulations of polytopes. In this talk, we present several numerical algorithms to compute such factorizations using standard low-rank matrix approximation formulations. These algorithms (sometimes) allow to provide explicit extended formulations and hence upper bounds for the nonnegative and PSD ranks (that is, the sizes of the smallest extended formulations). We illustrate this on regular n-gons, and show how our algorithms can give insight on their smallest extended formulations.

Matlab code to compute nonnegative factorizations can be downloaded from https://sites.google.com/site/exactnmf/

5.7 Extension complexity bounds for polygons – Numerical factorizations and conjectures

Francois Glineur (University of Louvain, BE)

Joint work of Glineur, Francois; Nicolas Gillis; Arnaud Vandaele and Julien Dewez

In this talk we describe an explicit nonnegative factorization of the slack matrix of the regular n-gon. We conjecture its rank to be optimal based on extensive numerical computations (cf. Gillis' talk on heuristic exact matrix factorization). In particular we describe the behaviour of the nonnegative tank for values of n lying between powers of two. We also compare this nonnegative rank to available computable lower bounds. Finally we describe our attempts at obtaining a (smaller) explicit positive semidefinite factorization for this slack matrix.

5.8 Positive Semidefinite Rank

João Gouveia (University of Coimbra, PT)

Main reference H. Fawzi, J. Gouveia, P. A. Parrilo, R. Z. Robinson, R. R. Thomas, "Positive semidefinite rank," arXiv:1407.4095v1 [math.OC], 2014.
URL http://arxiv.org/abs/1407.4095v1

In this talk we will cover basic properties of the positive semidefinite rank. A special emphasis will be put on its geometry and complexity, its relation to the square root rank and its space of factorizations. Based on joint work with Hamza Fawzi, Pablo Parrilo, Richard Robinson and Rekha Thomas.

5.9 Linear algebra invariants over semirings

Alexander Guterman (Moscow State University, RU)

Semiring matrix invariants are useful in different problems of linear algebra and its applications. In the talk we will discuss the properties of different matrix invariants which replace rank, determinant, etc., over semirings, in particular, over tropical semirings. The preferences will be given to the interrelations between different invariants and their applications. The talk is based on several works joint with Marianne Akian, LeRoy Beasley, Stephane Gaubert, and Yaroslav Shitov.

5.10 Extended Formulations: Constructions and Obstructions

Volker Kaibel (Universität Magdeburg, DE)

Joint work of Kaibel, Volker; Lee, Jon; Walter, Matthias; Weltge, Stefan

We are going to demonstrate both the power of extended formulations as well as the limitations of the concept. In the first part (based on joint work with Jon Lee, Matthias Walter and Stefan Weltge), we present polynomial size extended formulations for the independence polytopes of regular matroids. On the way, we review beautiful extended formulations for the spanning tree polytopes of planar graphs due to Williams (2001) and encounter some very simple, but stunning occurences of representations in higher dimensional spaces. In the second part (based on joint work with Stefan Weltge), we give an elementary combinatorial proof of an exponential lower bound on the rectangle covering numbers (and hence on the nonnegative ranks) of unique disjointness matrices, and we review the implications for the minimal sizes of extended formulations of the correlation polytopes.

5.11 Semialgebraic geometry of nonnegative and psd ranks

Kaie Kubjas (Aalto Science Institute, FI)

Cohen and Rothblum asked in 1993 whether every rational matrix of nonnegative rank r has a size r rational nonnegative factorization. In joint work with Elina Robeva and Bernd Sturmfels, we answer this question positively for matrices of nonnegative rank 3. I will explain how looking for a semialgebraic description of the set of matrices of nonnegative rank at most 3 helped us to derive this result, and talk about ongoing research with Elina Robeva and Richard Z. Robinson on semialgebraic geometry of the set of matrices of psd rank at most k.

5.12 Lower bounds on semidefinite programming relaxations: Part I

James R. Lee (University of Washington – Seattle, US)

Main reference James R. Lee, Prasad Raghavendra, David Steurer, "Lower bounds on the size of semidefinite programming relaxations," arXiv:1411.6317v1 [cs.CC], 2014.
URL http://arxiv.org/abs/1411.6317v1

We introduce a method for proving lower bounds on the size of SDP relaxations for combinatorial problems. In particular, we show that the cut, TSP, and stable set polytopes on n-vertex graphs are not the affine image of the feasible region of any spectrahedron of dimension less than 2^{n^c} for some constant $c > 0$. A spectrahedron is the feasible region of an SDP: The intersection of the positive semidefinite cone and an affine subspace. This yields the first super-polynomial lower bound on the semidefinite extension complexity of any explicit family of polytopes.

Our results follow from a general technique for proving lower bounds on the positive semidefinite rank of a nonnegative matrix. To this end, we establish a close connection between arbitrary SDPs and those arising from the sum-of-squares hierarchy. For approximating maximum constraint satisfaction problems, we prove that SDPs of polynomial-size are equivalent in power to those arising from degree-O(1) sum-of-squares relaxations. This result implies, for instance, that no family of polynomial-size SDP relaxations can achieve better than a 7/8-approximation for max 3-sat. Part I: After a brief review of the SDP extended formulation model, I will recast the SDP model as a proof system for certifying the nonnegativity of a family of statements. The size of the SDP corresponds to the number of axioms in the system. Then we will attempt to prove that for the family of nonnegative quadratic functions on the discrete cube, the smallest family of axioms is the subspace of low-degree multilinear polynomials. This will be reduced the tak of approximating an arbitrary low-rank PSD factorization by a "low-degree" PSD factorization. In Part II, David will discuss and prove the existence of an approximate low-degree factorization.

5.13 Extension complexity of polytopes with few vertices (or facets)

Arnau Padrol Sureda (FU Berlin, DE)

The number of combinatorial types of d-polytopes with up to d+4 vertices grows superexponentially with d. However, only quadratically many can have realizations with extension complexity smaller than d+4. These are easy to classify into finitely many families and the exact extension complexity of each realization is easy to decide.

5.14 Polytopes of minimum positive semidefinite rank in dimension four

Kanstantsin Pashkovich (University of Waterloo, CA)

For a given polytope the smallest size of a semidefinite extended formulation can be bounded from below by the dimension of the polytope plus one. This talk is about polytopes for which this bound is tight, i.e. polytopes with positive semidefinite (psd) rank equal to their dimension plus one. I will speak about a classification of psd minimum polytopes in dimension four.

Joint work with Gouveia, Robinson and Thomas.

5.15 Completely positive semidefinite cone

Teresa Piovesan (CWI – Amsterdam, NL)

Main reference S. Burgdorf, M. Laurent, T. Piovesan, "On the closure of the completely positive semidefinite cone and linear approximations to quantum colorings," arXiv:1502.02842v1 [math.OC], 2015.
URL http://arxiv.org/abs/1502.02842v1

We investigate structural properties of the completely positive semidefinite cone, consisting of all the n-by-n symmetric matrices that admit a Gram representation by positive semidefinite matrices of any size. This cone has been introduced to model quantum graph parameters as conic optimization problems. Recently it has also been used to characterize the set Q of bipartite quantum correlations, as projection of an affine section of it. We have two main results concerning the structure of the completely positive semidefinite cone, namely about its interior and about its closure. On the one hand we construct a hierarchy of polyhedral cones which covers the interior of the completely positive semidefinite cone, which we use for computing some variants of the quantum chromatic number by way of a linear program. On the other hand we give an explicit description of the closure of the completely positive semidefinite cone, by showing that it consists of all matrices admitting a Gram representation in the tracial ultraproduct of matrix algebras.

Join work with Sabine Burgdorf and Monique Laurent

5.16 Small Linear Programs Cannot Approximate Vertex Cover Within a Factor of $2 - \epsilon$;

Sebastian Pokutta (Georgia Institute of Technology, US)

We prove that every linear programming (LP) relaxation that approximates vertex cover within a factor of $2 - \epsilon$ has super-polynomially many inequalities. As a direct consequence of our methods, we also establish that LP relaxations that approximate independent set within any constant factor have super-polynomially many inequalities.

5.17 Geometric properties, matroids, and forbidden minors

Raman Sanyal (FU Berlin, DE)

Nonegative rank, PSD-rank, and Theta rank are just some geometric properties that are interesting but difficult to compute. In this talk I will discuss such geometric properties for matroid base polytopes. In this setup, the class of matroids with bounded rank is closed under taking minors. I will give some excluded-minor descriptions and connections to classical results in matroid theory. This is joint work with Francesco Grande.

5.18 On the exactness of moment relaxations

Markus Schweighofer (Universität Konstanz, DE)

Semidefinite programming is optimizing a linear function over a spectrahedron, i.e., the solution set of a linear matrix inequality. It becomes increasingly important in discrete and continuous optimization since it serves to solve the hierarchy of moment relaxations for polynomial optimization problems introduced by Lasserre around the turn of the millennium. In our days, algorithms like the Goemans-Williamson relaxation for the maximum cut problem can be viewed as the first level of this hierarchy. Recent work of Lee, Raghavendra and Steurer seems to suggest that moment relaxations might be universal semidefinite programming relaxations. The semidefinite programs dual to moment relaxations are sum-of-squares programs. In this talk, we will try to analyze the moment relaxations for writing a convex semi-algebraic set as a projected spectrahedron. We have a new positive result building upon the work of Helton and Nie.

5.19 Sublinear extensions of polygons

Yaroslav Shitov (NRU Higher School of Economics – Moscow, RU)

Main reference Y. Shitov, "Sublinear extensions of polygons," arXiv:1412.0728v1 [math.CO], 2014.
URL http://arxiv.org/abs/1412.0728v1

One of the central questions of extended formulations theory is the existence of strong lower bounds for polytopes naturally arising in optimization. The study of lower bounds for linear formulations dates back to the 1991 paper by Yannakakis, and, since then, many important polytopes have been proven to have exponential extension complexity.

This talk will be devoted to the case of generic polytopes, that is, polytopes whose vertices chosen randomly in space. How strong are linear formulations in this case? Beasley and Laffey conjectured that, even for a convex n-gon on the plane, the extension complexity can be worst possible (equal to n). However, it turns out that linear formulations are not that weak: we show that the extension complexity of any convex n-gon is o-small of n. Moreover, we can provide examples of generic n-gons whose complexities do not exceed $c \cdot \sqrt{n}$; this upper bound coincides with the known lower bound up to a constant factor. Our results allow us to make a number of conjectures concerning higher-dimensional generic polytopes.

5.20 Lower bounds for semidefinite programming relaxations: part II

David Steurer (Cornell University – Ithaca, US)

Main reference J. R. Lee, P. Raghavendra, D. Steurer, "Lower bounds on the size of semidefinite programming relaxations," arXiv:1411.6317v1 [cs.CC], 2014.
URL http://arxiv.org/abs/1411.6317v1

We introduce a method for proving lower bounds on the size of SDP relaxations for combinatorial problems. In particular, we show that the cut, TSP, and stable set polytopes on n-vertex graphs are not the affine image of the feasible region of any spectrahedron of dimension less than 2^{n^c} for some constant $c > 0$. A spectrahedron is the feasible region of an SDP: The intersection of the positive semidefinite cone and an affine subspace. This yields the first super-polynomial lower bound on the semidefinite extension complexity of any explicit family of polytopes.

Our results follow from a general technique for proving lower bounds on the positive semidefinite rank of a nonnegative matrix. To this end, we establish a close connection between arbitrary SDPs and those arising from the sum-of-squares hierarchy. For approximating maximum constraint satisfaction problems, we prove that SDPs of polynomial-size are equivalent in power to those arising from degree-O(1) sum-of-squares relaxations. This result implies, for instance, that no family of polynomial-size SDP relaxations can achieve better than a 7/8-approximation for max 3-sat.

Part II: We will show that for certain families of matrices, every low-rank PSD factorization can be approximated by a low-degree PSD factorization. Following Part I, this will reduce lower bounds for PSD rank to lower bounds for sum-of-squares degree.

5.21 Subgraph Polytopes and Independence Polytopes of Count Matroids

Stefan Weltge (Universität Magdeburg, DE)

Given an undirected graph, the non-empty subgraph polytope is the convex hull of the characteristic vectors of pairs (F,S) where S is a non-empty subset of nodes and F is a subset of the edges with both endnodes in S. We obtain a strong relationship between the non-empty subgraph polytope and the spanning forest polytope. We further show that these polytopes provide polynomial size extended formulations for independence polytopes of count matroids, which generalizes recent results obtained by Iwata et al. referring to sparsity matroids. As a byproduct, we obtain new lower bounds on the extension complexity of the spanning forest polytope in terms of extension complexities of independence polytopes of these matroids.

5.22 Quantum Communication Complexity as a Tool to Analyze PSD Rank

Ronald de Wolf (CWI – Amsterdam, NL)

Main reference Jedrzej Kaniewski, Troy Lee, Ronald de Wolf, "Query complexity in expectation," arXiv:1411.7280v1 [quant-ph], 2014.
URL http://arxiv.org/abs/1411.7280v1

We start with a brief introduction to the general area of quantum communication complexity, and then describe the connection with the positive-semidefinite rank of matrices (due to Fiorini et al. '12): the logarithm of the psd rank of a matrix M equals the minimal quantum communication needed by protocols that compute M in expectation. This means that results about quantum communication complexity can be used to obtain results about psd rank, both lower and upper bounds. As an example of the latter we present an efficient quantum communication protocol (due to Kaniewski, Lee, and de Wolf '14) that induces an exponentially-close approximation for the slack matrix for the perfect matching polytope, of psd rank only roughly $2^{\sqrt{n}}$. In contrast, Braun and Pokutta'14 showed that such approximating matrices need nonnegative rank $2^{\Omega(n)}$.

6 Problems resolved from Dagstuhl Seminar 13082

Progress has been made on a number of open problems discussed at the preceding seminar in February 2013. We briefly highlight some of this work here.

1. Is there a nonnegative n-by-n matrix of rank 3 and nonnegative rank n (for $n \geq 7$)? Shortly after the last seminar Yaroslav Shitov showed that this is not the case. He has since improved these results to show that every nonnegative matrix of rank 3 has nonnegative rank $o(n)$ in this paper http://arxiv.org/abs/1412.0728.
2. Does $\text{rank}_+(A \otimes B) = \text{rank}_+(A) \cdot \text{rank}_+(B)$? This was disproven by Hamza Fawzi using the software to compute nonnegative rank of Nicolas Gillis. The analogous statement for positive semidefinite rank is also false as shown by Lee, Wei, and de Wolf http://arxiv.org/abs/1407.4308.
3. Does positive semidefinite rank depend on the underlying field? Say that a matrix M has rational entries, and we require the matrices A_i, B_j in the factorization $M_{ij} = \text{Tr}(A_i B_j)$ to have rational entries as well. Can this require larger matrices than if we allowed A_i, B_j to have real entries. Gouveia, Fawzi, and Robinson indeed give an example of a rational matrix M where the rational psd rank is larger than the real psd rank http://arxiv.org/abs/1404.4864. Lee, Wei, and de Wolf give an example of a family of matrices where the real psd rank is asymptotically larger than the complex psd rank (where the factors are allowed to be complex Hermitian psd matrices) by a factor of $\sqrt{2}$ http://arxiv.org/abs/1407.4308.
4. What is the largest possible gap between the approximate rank and the approximate nonnegative rank for boolean matrices? Let $\text{rank}_{1/10}(M)$ denote the minimum rank of a matrix S such that $\|S - M\|_\infty \leq 1/10$. Define $\text{rank}^+_{1/10}(M)$ in the same way for the nonnegative rank. Kol, Moran, Shpilka, and Yehudayoff (Approximate Nonnegative Rank is Equivalent to the Smooth Rectangle Bound, ICALP 2014) made progress on

this question by showing that the approximate nonnegative rank of the set intersection matrix of subsets of $[n]$ is $2^{\Omega(n)}$ while it is known that its approximate rank is $2^{O(\sqrt{n})}$.

5. For a n-by-n boolean matrix A satisfying $A \circ A^T = I_n$ how small can the rank be? Here $\circ$ denotes the entrywise product of matrices A^T is the transpose of A. This question is motivated by the fooling set method in communication complexity. It is known that the rank must be at least $\sqrt{n}$. Friesen, Hamed, Lee, and Theis http://arxiv.org/abs/1208.2920 constructed a growing family of matrices with rank $\sqrt{(}n) + O(1)$ over a finite field, and another construction of matrices with integer entries with rank $O(\sqrt{n})$ over the reals. Shigeta and Amano (http://arxiv.org/abs/1311.6192) have now essentially answered the original question, constructing a growing family of *boolean* matrices with rank $n^{1/2+o(1)}$.

7 Open Problems

Many open problems were posed during the talks at the seminar. There was also a session on Wednesday evening for presenting open problems at the board and discussing them with the participants. Jon Swenson from the University of Washington kindly recorded several of these problems on behalf of the organizers. The main questions that came up are listed below. Some of these problems are classical—the name indicates the promotor of the problem, not necessarily the originator.

1. (R. de Wolf) For a regular n-gon it is known that the nonnegative rank of the slack matrix is $\Theta(\log(n)$. For the psd rank, the best lower bound is $\Omega(\sqrt{\frac{\log n}{\log \log n}})$ and the best upper bound is still $O(\log n)$. What is the true psd rank of the slack matrix of the n-gon?
2. (T. Lee) Let the *rational* psd rank of a matrix $A \in \mathbb{R}_+^{m \times n}$ be the minimum r such that $A(i,j) = \frac{B(i,j)}{C(i,j)}$ for all $i \in [m], j \in [n]$ where B, C are nonnegative matrices with $r = \text{rank}_{psd}(B) + \text{rank}_{psd}(C)$. Does the slack matrix of the correlation polytope have exponential rational psd rank? Note for contrast that the slack matrix of the perfect matching polytope has rational psd rank $O(n^4)$.
3. (J. Gouveia) Let $\mathcal{F}_k(M) = \mathcal{SF}(M)/\text{GL}(k)$ be the space of factorizations of a matrix M with psd rank k. Under what conditions is $\mathcal{F}_k(M)$ connected? Perhaps one can come up with a condition involving psd rank, ordinary rank, or something else. (For notation, see http://www.mat.uc.pt/~jgouveia/dagstuhl.pdf.)
4. (K. Pashkovich) Give a geometric characterization of the polytopes in $\mathbb{R}^n$ for which the psd rank of the slack matrix of P is equal to $\dim(P) + 1$. This is the smallest that the psd rank of a polytope can be. Such a characterization exists now for $n \leq 4$.
5. (S. Fiorini) Is there a constant k such that for every d and 2-level polytope P of dimension d the linear extension complexity is $2^{O(\log(d)^k)}$?
6. (V. Kaibel) From results of Rothvoß we know that there is a matroid polytope with exponential linear extension complexity. Give an explicit example of an explicit matroid polytope with exponential linear extension complexity.
7. (J. Lee) What is the positive semidefinite rank of the matching polytope?
8. (J. Lee) We currently know that the slack matrix of the correlation polytope $\text{CORR}_n(x, y) =$ convex hull($x^T y$ for $x, y \in \{0,1\}^n$) satisfies $N^{2/13} \leq \log_2(\text{rank}_{\text{psd}}(\text{CORR}_n)) \leq N$ where $N = 2^n$. What is the right answer? This is the same as the rank of M(f,x)=f(x) where f ranges over all quadratic nonlinear functions that are nonnegative on the hypercube.

9. (J. Lee) Consider the matrix $M(G,x) = 0.99 - val_G(x)$ where G ranges over n-vertex graphs, x ranges over cuts, and $val_G(x)$ is the fraction of edges in G that cross the cut x. What is the non-negative rank of M? The best lower bound we have is $n^{\Omega(logn)}$, and there is a barrier of sorts because the n^{-10}-approximate nonnegative rank (additive error) is $n^{O(\log n)}$.
10. (J. Lee) Can one exhibit an explicit function $f : \{0,1\}^n \to \mathbb{R}_+$ so that f cannot be written as a sum of squares of s-sparse polynomials with $s = n^{O(1)}$? Here s-sparse means that the polynomial contains at most s monomials.

8 Problems resolved from Dagstuhl Seminar 15082

Cohen and Rothblum raised the question in 1993 if the nonnegative rank of a rational matrix over the reals is the same as its nonnnegative rank over the rationals. More generally they asked: if A has entries in a sub-semiring S of the semiring R of the nonnegative reals, how large can the gap be between $\text{rank}_S(A)$ and $\text{rank}_R(A)$? For the latter question, Yaroslav Shitov (http://arxiv.org/abs/1505.01893) has recently provided the first example where $\text{rank}_S(A) \neq \text{rank}_R(A)$.

Participants

- Alexander Barvinok
University of Michigan – Ann Arbor, US
- LeRoy B. Beasley
Utah State University, US
- Greg Blekherman
Georgia Inst. of Technology, US
- Daniel Dadush
CWI – Amsterdam, NL
- Ronald de Wolf
CWI – Amsterdam, NL
- Samuel Fiorini
University of Brussels, BE
- Nicolas Gillis
University of Mons, BE
- Francois Glineur
University of Louvain, BE
- João Gouveia
University of Coimbra, PT
- Sebastian Gruler
Universität Konstanz, DE
- Alexander Guterman
Moscow State University, RU
- Volker Kaibel
Universität Magdeburg, DE
- Hartmut Klauck
Nanyang TU – Singapore, SG
- Kaie Kubjas
Aalto Science Institute, FI
- James R. Lee
University of Washington – Seattle, US
- Troy Lee
National Univ. of Singapore, SG
- Arnau Padrol Sureda
FU Berlin, DE
- Kanstantsin Pashkovich
University of Waterloo, CA
- Teresa Piovesan
CWI – Amsterdam, NL
- Sebastian Pokutta
Georgia Inst. of Technology, US
- Raman Sanyal
FU Berlin, DE
- Markus Schweighofer
Universität Konstanz, DE
- Yaroslav Shitov
NRU Higher School of Economics – Moscow, RU
- David Steurer
Cornell University – Ithaca, US
- Jonathan Swenson
University of Washington – Seattle, US
- Dirk Oliver Theis
University of Tartu, EE
- Rekha R. Thomas
University of Washington – Seattle, US
- Hans Raj Tiwary
Charles University – Prague, CZ
- Antonios Varvitsiotis
National Univ. of Singapore, SG
- Stefan Weltge
Universität Magdeburg, DE

Report from Dagstuhl Seminar 15091

Smart Buildings and Smart Grids

Edited by

Hans-Arno Jacobsen[1], Randy H. Katz[2], Hartmut Schmeck[3], and Christoph Goebel[4]

1 **TU München, DE,** arno.jacobsen@tum.de
2 **University of California – Berkeley, US,** randy@cs.berkeley.edu
3 **KIT – Karlsruher Institut für Technologie, DE,** hartmut.schmeck@kit.edu
4 **TU München, DE,** christoph.goebel@tum.de

Abstract

This report provides an overview of the program, discussions, and outcomes of Dagstuhl Seminar 15091 "Smart Buildings and Smart Grids", which took place from 22–27 February 2015 at Schloss Dagstuhl – Leibniz Center for Informatics. The main goal of the seminar was to provide a forum for leading Energy Informatics (EI) researchers to discuss their recent research on Smart Buildings and Smart Grids, to further elaborate EI research agenda and methods, and to kick-start new research projects with industry. The report contains abstracts of talks that were held by the participants and the outcomes of several discussion sessions on the focal topics of the seminar (e.g., information technology driven developments in building and power system management, as well as cross-cutting topics, such as computer networks, data management, and system design.

Seminar February 22–27, 2015 – http://www.dagstuhl.de/15091

1998 ACM Subject Classification D.0 [Software] Software, G.0 [Mathematics of Computing] General, H.0 [Information Systems] General, J.2 [Computer Applications] Physical Sciences and Engineering – Engineering

Keywords and phrases Energy Informatics, Smart Grids, Smart Buildings, Cyber-Physical Systems

Digital Object Identifier 10.4230/DagRep.5.2.128

1 Executive Summary

Hans-Arno Jacobsen
Randy H. Katz
Hartmut Schmeck
Christoph Goebel

Motivation

Motivated by the increasing importance of producing and consuming energy more sustainably, a new and highly dynamic research community within computer science has evolved: Energy Informatics (EI). Researchers active in the EI field investigate information age solutions for monitoring and controlling large cyber-physical infrastructures with a focus on the following goals: (i) an overall reduction of the energy consumption of these infrastructures, and (ii) the integration of distributed renewable energy sources into these infrastructures. This seminar focused on two use cases of existing cyber-physical systems, buildings and power grids. These use cases were chosen due to their relevance in terms of energy footprint. The seminar has three major goals: (i) to provide a forum for leading EI researchers to discuss their recent research on Smart Buildings and Smart Grids, (ii) to further elaborate EI research agenda and methods, and (iii) to kick-start new research projects with industry.

Smart Buildings. Modern buildings already incorporate increasingly sophisticated Building Management Systems (BMS) that integrate building control with improved sensors and better data collection and presentation capabilities. However, these systems currently only enable simple, decoupled control of building services like lighting, ventilation, heating and cooling. Their architectures and Application Programming Interfaces (APIs) are not standardized, and often proprietary: only the BMS vendor can add functionality. This slows the pace of innovation, buildings remain rigid in the functions and services they provide, and their quality and effectiveness remain difficult to quantify. Contemporary BMS attempt to achieve global service levels based on local control instead of meeting individual occupant requirements based on global control. Standardized building management APIs and scalable middleware solutions that enable reliable communication between building sensors, users, control systems, and machinery could accelerate energy innovation in the building sector.

Smart Grids. Contemporary electricity grids and markets were designed for a scenario in which large and mostly fossil-fueled power plants are dispatched to meet an almost inflexible demand. Achieving sustainable energy supply, however, requires moving towards a scenario where the variable power supplied by distributed renewable resources like wind and solar has to be absorbed by supply-following loads and energy storage whenever it is available. Thus, instead of dispatching a relatively small number of large generators, the large-scale integration of new types of generators and loads into electric grids requires new types of information systems for monitoring and controlling them, while making efficient use of existing assets. The task of controlling large numbers of flexible loads, e.g., air conditioning systems in buildings, electric vehicles, and small-scale energy storage systems, while guaranteeing overall system stability, is highly demanding in terms of computational complexity, required data communication and data storage. In the Smart Grid space, the challenge faced by EI researchers is to develop and carefully evaluate new ideas and actual system components enabling Smart Grid systems that are scalable, efficient, reliable, and secure.

Organization of the Seminar

The week-long workshop plan was as follows. Day 1 introduced the attendees to each other, and set the stage through invited tutorial presentations and brainstorming sessions. Day 2 was spent in breakouts focused on identifying the research challenges and opportunities, organized by application area such as Smart Buildings or Energy Grids, based on attendee interest and expertise. On Day 2, we also held the first out of two presentation session, where participants could give a short overview about their current research. Day 3 was used to assess the workshop at mid-stream, conduct group discussions, and make necessary corrections. Initial writing assignments, to document the discussions of the breakout sessions, were made on this day, as well. Day 4 consisted of a second round of breakouts focusing on enablers and crosscutting issues (e.g., data management, system design patterns, and human machine interaction) and the second participants' presentation session. Work on completing the report draft continued on that day. The last day consisted of the reviewing of the report draft, and through group discussion, identify the summary findings and recommendations.

2 Table of Contents

3 Overview of Keynote Talks

3.1 A Composable Method for the Real-Time Control of Active Distribution Networks with Explicit Power Setpoints

Jean-Yves Le Boudec and Mario Paolone (EPFL, CH)

The classic approach to the control of medium and low voltage distribution networks involves a combination of both frequency and voltage controls at different time scales. With the increased penetration of stochastic resources, distributed generation and demand response, it shows severe limitations both in the optimal/safe operation of these networks, as well as in aggregating the network resources for upper-layer power systems.

To overcome this difficulty, we propose a radically different control philosophy, which enables low and medium voltage distribution networks as well as and their resources to directly communicate with each other in order to define explicit real-time setpoints for active/reactive power absorptions or injections. We discuss a protocol for the explicit control of power flows and voltage, combined with a recursive abstraction framework. The method is composable, i.e., subsystems can be aggregated into abstract models that hide most of their internal complexity.

Within this control framework we specifically analyze the case of a low-overhead decentralized Demand Response (DR) control mechanism, henceforth called Grid Explicit Congestion Notification (GECN), intended for deployment by distribution network operators (DNOs) to provide ancillary services to the grid by a seamless control of a large population of elastic appliances.

Contrary to classic DR approaches, the proposed scheme aims to continuously support the grid needs in terms of voltage control by broadcasting low-bit rate control signals on a fast time scale (i.e., every few seconds). Overall, the proposed DR mechanism is designed to i) indirectly reveal storage capabilities of end-customers and ii) have a negligible impact on the end-customer.

3.2 Bringing Distributed Energy Resources to Market

Christoph Goebel and Hans-Arno Jacobsen (Technische Universität München, DE)

The first part of this talk was meant to be a tutorial on current challenges in the operation of power systems induced by renewable integration. The second part provided an overview of our current research in this area.

In the first part, we introduced the different stakeholders involved in electricity generation, transmission, and consumption and how coping with the individual challenges they face requires the use of innovative information technology. We argued that better monitoring and control of renewables, grids, and the demand side competes with more traditional ways to deal with renewable integration challenges, e.g., aggressive grid expansion or construction of highly flexible power plants. In the long-term, with several developments happening at the same time, e.g., declining prices of batteries and cheap control infrastructures, the intelligent dispatch of distributed energy resources by aggregators could prevail in this competition. The task of such aggregators is to "make the most" out of the potential of various distributed energy resources while respecting the individual resource and wholesale market constraints.

In the second part, we presented several research challenges in the area aggregator systems. Among other things, these challenges include predictive capabilities, multi-resource/multi-purpose dispatch, profit maximization of aggregators, new optimization techniques, distributed system designs to achieve scalability and fault tolerance, and optimal data storage schemes for representing resource schedules and system states. We presented recent work in the area multi-purpose dispatch and new optimization techniques in more detail. We closed by motivating new research in the Smart Grid field, but also repeated the necessity for putting more common effort into the collection and consolidation of public data for research purposes and the development of open source tools, e.g., test beds and simulation environments. As starting points, we presented two such efforts we recently initiated at TUM.

3.3 The Software-Defined Building: A Technical Approach for Smart Buildings

Randy H. Katz (University of California – Berkeley, US)

The built environment is the human-made surroundings that provide the setting for human activity, ranging from buildings to cities and including their supporting infrastructures, such as for water or energy. Buildings are the starting point for Smart Cities. They represent a large component of modern economies. They are where we spend 90% of our time, consume 70% of our electricity (40% of our primary energy consumption) and generate 40% of our GHG emissions. The aggregate energy expenditure for the U.S. buildings stock exceeds $400 billion and construction represents more than 10% of the entire U.S. GDP. Over the last several decades, information-processing abilities have grown enormously; yet thus far this has had little effect on the built environment.

Given the widespread deployment of digital controls in commercial buildings, we can consider them to be Cyber-Physical Systems, and we can enhance their functionality through software programmability. The analogy is how phones are (almost) infinitely extended through network-connected applications. Traditional building facilities are stovepipes; e.g., HVAC is separated from lighting controls, even though there is a correlation between lit and conditioned spaces. Awareness of occupant location and activities offers new opportunities for the building to become aware of and respond to occupant needs, faster and with more efficiency. Occupant devices integrated with the building frees it from the limitations of mechanical and rigidly placed sensors and controls, and awareness of environmental factors like weather and sun orientation can be exploited to better control interior spaces. The building's information processing capabilities can become cloud integrated, providing unlimited capacity and the ability to extend control to fleets of buildings, neighborhoods or complete cities.

The key research need is the development of a new category of operational software, a Building Operating System, a City Operating System, etc., to provide a common foundation for abstracting and managing the resources of the Built Environment, providing integration with user devices and external information sources, data analytical processing, and enabling new kinds of control, information presentation, and planning applications. The advantage will be demonstrated by quantifiable improvements in such figures of merit as energy efficiency (including agile and intelligent interaction with the electric grid for energy flex), reduced cost of ownership (maintenance and management) and importantly, improved occupant satisfaction (comfort, indoor air quality, aesthetics, information transparency, e.g., how activities translate into energy consumption or savings; this can increase productivity of the workforce and sales for retail spaces), as well as enhanced controllability, agility, and extensibility via an open platform and technology ecosystem to achieve these goals.

3.4 Industrial Perspectives on Smart Buildings and IoT Impact

Milan Milenkovic (Intel – Santa Clara, US)

Buildings are one of the primary users of energy (40% of all energy and 70% of electricity according to US Department of Energy and other international agencies), but some 20-40% of that is wasted due to inefficiencies. We advocate an Internet of Things (IoT) based approach to solving this problem that reduces costs and provides numerous benefits. Commercial buildings today have a variety of automated systems to monitor and control different aspects of their behavior, including: HVAC and lighting (usually BMS), energy consumption, lifts, security and access, fire alarm, water management, parking, landscaping and irrigation, audio visual, digital signage. Coordinated behaviors of such systems can result in increased efficiencies, safety, and occupant comfort, such as monitoring of occupancy to dynamically adjust heating/cooling and lighting, or automatically moving elevator cars to ground floor for safety in case of imminent power failures. Most of the currently deployed legacy building control systems are isolated proprietary systems that do not interoperate because that would require prohibitively expensive custom interfacing.

Use of IoT in building control systems brings standards, lower costs and well known benefits of Internet – connectivity and interoperability, scalability, tested tools and design practices, faster/cheaper development and interoperability. This tutorial presents an end-to-end IoT system architecture and argues for interoperable sensor/control data and meta-data definition to facilitate collaboration among domains and building control systems in particular. We also introduce examples of building deployments with IoT gateways in commercial buildings (usually BMS) and residential (retrofit example) and rooftop HVAC units.

Smart Grid needs to interface with building control systems. They are a key load and, when instrumented, can provide detailed feedback on current and projected electricity usage, so production can be more adaptive and the two systems can balance consumption and production using detailed usage information, estimates (based on accurate building models and past behavior), and interact/balance in real-time using techniques like demand response.

3.5 Energy Management in Smart Homes

Florian Allerding, Birger Becker, and Hartmut Schmeck (KIT – Karlsruher Institut für Technologie, DE)

Motivated by the challenges of the energy transition ("Energiewende") like fluctuating power generation, uncertainty of supply and inherent decentralization, this talk presents approaches to smart home management systems with a focus on energy management. While some of these (commercially available) systems address monitoring and remote control only, others include learning user habits.

As a particular example of a research prototype running in some test locations, the Organic Smart Home of KIT is presented, which includes an Energy Management Panel for visualizing the current energy situation and for discovering or specifying user preference for the operation times of appliances. An observer/controller-based architecture then optimizes the operation times of smart home appliances and other energy relevant devices, complying with the degrees of freedom specified by the residents of the home and considering also external information about time varying power tariffs and potential power limits.

Furthermore, an outlook is given on extensions to regional energy management by organizing a large number of devices (like appliances, CHPs, electric vehicles, heat pumps) into a pool in order to provide a cascaded form of responses for coping with spontaneous deviations from power schedules. Concluding, a number of questions is presented which are relevant for the design of Energy Management Systems in Smart Homes.

4 Overview of Participant's Talks

4.1 Organic Smart Home Energy Management and Building "Operating System"

Florian Allerding (KIT – Karlsruher Institut für Technologie, DE)

The Organic Smart Home is a flexible, generic "operating system" for Smart Buildings in real-world applications, which is already in use in households and office buildings. The major contribution is the design of a "plug & play"-type Evolutionary Algorithm for optimizing and management distributed generation, storage and consumption using a sub-problem based approach. Relevant power consuming or producing components identify themselves as sub-problems by providing an abstract specification of their genotype, an evaluation function and a back transformation from an optimized genotype to specific control commands. The generic optimization respects technical constraints as well as external signals like variable energy tariffs.

4.2 Direct Control of Demand Flexibility: Applicability of Batch Reinforcement Learning

Bert Claessens (Vito – Antwerp, BE)

In this talk, I presented recent work that contributes to the application of Batch Reinforcement Learning (RL) to demand response. In contrast to conventional model-based approaches, batch RL techniques do not require a system identification step, which makes them more suitable for a large-scale implementation. This talk discussed how fitted Q-iteration, a standard batch RL technique, can be extended to the situation where a forecast of the exogenous data is provided. In general, batch RL techniques do not rely on expert knowledge on the system dynamics or the solution. However, if some expert knowledge is provided, it can be incorporated by using our novel policy adjustment method. Finally, we tackled the challenge of finding an open-loop schedule required to participate in the day-ahead market. We proposed a model-free Monte-Carlo estimator method that uses a metric to construct artificial trajectories and we illustrate this method by finding the day-ahead schedule of a heat-pump thermostat. Our experiments showed that batch RL techniques provide a valuable alternative to model-based controllers and that they can be used to construct both closed-loop and open-loop policies.

4.3 Load Prediction of Non-Controllable Household Devices

Christoph Doblander (TU München, DE)

While many devices in the household could be controlled and the energy usage is known, the majority of the energy is consumed by devices which are not controllable. To increase self-sufficiency, we propose an architecture where the energy consumption of non-controllable loads is predicted and taken as an input for a control algorithm which controls the other devices to maximize for self sufficiency. One benefit of such a system is financial, since the incentive to feed rooftop solar energy back into the grid are declining.

We evaluate multiple machine learning algorithms, support vector machines, naive bayes and benchmark against persistence. We also evaluate the prediction error reduction when additionally features are extracted from the time series. The predictions are done on a sliding window, e.g., every minute, the load of the next 15 minutes is predicted based on the last 15 minutes, hence supporting a scenario in which a control algorithm continuously optimizes the actuation of the controllable devices.

The data was collected within the field trial of the "PeerEnergyCloud" project. Roughly 30 households were equipped with up to 7 plug meters. The results suggest that additional feature extraction reduces the prediction error and the benchmark persistence can be beaten. The evaluation of the prediction algorithms was done on 20 different households which allows to derive significant conclusions. The prediction error was measured by the Mean Absolute Percentage Error (MAPE) to compare it to existing literature

4.4 Are Energy Markets Efficient? The Case of Real and Virtual Storage

Nicolas Gast (INRIA – Grenoble, FR)

The electrical grid of the future will require more storage to compensate for the intermittency of distributed generators (such as solar, wind, combined heat and power). Storage will be real (batteries, water reservoirs) or virtual (demand response). In this talk, we analyzed the impact of storage of the real electricity markets, using several stochastic models. We showed that there exists a market price such that selfish users are provided with incentives to control their appliances in a socially optimal way. However, by setting these prices, users have an incentive to install a sub-optimal quantity of storage.

4.5 Loose Coupling Approach to Demand Response for Distribution Networks

Kai Heussen (Technical University of Denmark – Lyngby, DK)

In this talk, I reviewed our recent research in the area of congestion control in distribution grids. In particular, I focused on the necessary coordination between aggregators and distribution grid operators. The insights of our research have led to the development of FLECH, a market place for flexibility.

4.6 Information Systems and Science for Energy (ISS4E) at the University of Waterloo

Srinivasan Keshav (University of Waterloo, CA)

The current grid suffers from several problems, ranging from a high carbon footprint to very coarse control of loads. These can be solved using three key technologies: solar energy, energy storage, and the Internet of Things. In this talk, I discussed the difficult challenges that need to be solved using these technologies, such as meeting stochastic loads using stochastic generation and the need for control over multiple time scales. I then outlined three approaches being taken by the ISS4E group at Waterloo (http://iss4e.ca) to meet these challenges: a) using the Internet as an inspiration for Smart Grid architecture b) analysis of Smart Grid data sets and c) using Internet technologies for smart sensing and control.

4.7 Integrated Simulation of Power and ICT Systems

Johanna Myrzik (TU Dortmund, DE)

Co-simulation of ICT and power systems becomes increasingly important to develop and test Wide-Area Monitoring, Protection, and Control (WAMPAC) applications. In this talk, I presented INSPIRE, an integrated simulation of power and ICT systems for real-time evaluation. I provided insights into the architecture of INSPIRE and presented selected simulation results obtained in different control scenarios.

4.8 Introduction of the Grid4EU Project

Peter Noglik (ABB AG Forschungszentrum Deutschland – Ladenburg, DE)

In this talk, I gave a short overview of a large EU project, Grid4EU, which focuses on the large scale demonstration of advanced Smart Grid solutions with wide replication and scalability potential for Europe. It focuses on how distribution system operators can dynamically manage electricity supply and demand, which is crucial for integration of large amounts of renewable energy.

4.9 A Business Model for Scalable Demand Response

Anthony Papavasiliou (University of Louvain, BE)

I introduced ColorPower, a business model for scalable residential demand response (DR). The ColorPower software enables precision control of the consumer's demand flexibility, which it dispatches prioritized by customer preference. Demand management impact is divided into green (any time), yellow (peak periods), and red (emergencies). I discussed results from several experiments applying ColorPower, including automatic emergency DR and precision DR shaping.

4.10 Agent-Based Smart * Management Platform with Plug & Play

Yvonne-Anne Pignolet (ABB Corporate Research – Baden-Dättwil, CH)

Smart * Management System (SMS) to control and coordinate residential, commercial and industrial sites need to be flexible and support devices entering and leaving the network due to malfunctions, mobility or when new components are added. In these environments the burden of administration quickly overwhelms any potential benefit if the devices require explicit configuration in order to work together. Moreover, the embedded systems used in these scenarios have very severe requirements in terms of costs and sizes, offering very limited resources to any application running on them.

In this talk, we present a platform that requires no human intervention for the configuration and simplifies the addition of new devices. The platform enables the interaction of independent (distributed) agents with a publish subscribe architecture; different agents do different things in different places. This permits the system to grow as needed. In addition heterogeneous technologies are supported with an abstraction layer hiding the specific requirements of each appliance and offers a uniform interface to the agents above it. The user benefits from two technologies: SmartScript and SmartEnvironment. The first allows the user to write powerful building automation rules in a simple language, at a really high abstraction level. Users don't need to know how the appliances work, they only specify what they want the system to do. The second technology needs even less interaction with the user because it learns from the habits of the user and subsequently controls the appliances to increase the comfort and to reduce the energy consumption automatically according to the user behavior it observed. The whole architecture has been proven to work with inexpensive devices with low power consumption and very constrained HW resources.

4.11 Distributed Optimization in Smart Grids

Jose Rivera (TU München, DE)

The introduction of the Smart Grid will allow the active control of several devices: Electric vehicles, distributed storage units, distributed generation units and smart home appliances. This poses new challenges for operators of large power systems: How can they actively control large numbers of devices in a scalable, reliable and efficient way? This talk explored the contributions that distributed optimization can make towards answering these questions.

4.12 Low-Voltage Grid Control over Heterogeneous Communication Networks

Hans-Peter Schwefel (FTW Forschungszentrum Telekommunikation Wien GmbH, AT)

Voltage control in the Low-Voltage (LV) distribution grid can be performed by a hierarchical control architecture, in which a controller placed on the secondary substations, called Low-Voltage Grid Controller LVGC, communicates setpoints to the local controllers on assets in the corresponding parts of the LV grid.

In order to minimize the communication overhead and to maximize asset utilization, such LVGC is designed to be passive while the voltage is within certain boundaries, and only becomes activated when a sensor measures and communicates an exceedance of such voltage band. Example results from co-simulation of an example grid with photo-voltaic assets show that such voltage control is effective to reduce duration and extent of voltage events in the low-voltage grid. The design of the information exchange between assets and LVGC, however, has a strong impact on the performance of the controller: When using an adaptive monitoring framework to optimize the time instances at which asset information is requested by the controller, control performance can be significantly improved and robustness to communication network delays can be achieved. The adaptive monitoring scheme thereby uses as optimization target an information quality metric, which can be efficiently calculated based on asset dynamics and the (measured) communication network delays.

In the final part of the talk, additional challenges of control use-cases in the low-voltage and medium voltage grid were outlined and solution approaches followed up by the FP7 Research Project SmartC2Net were introduced: adaptive control approaches, adaptive grid and network monitoring, ICT capability analysis, communication network reconfiguration, and assessment approaches via analytic models, co-simulation models, and coupling of different lab test beds.

4.13 Revealing Household Characteristics from Smart Meter Data

Thorsten Staake (Universität Bamberg, DE)

Utilities are currently deploying smart electricity meters in millions of households worldwide to collect fine-grained electricity consumption data. We present an approach to automatically analyzing this data to enable personalized and scalable energy efficiency programs for private households. In particular, we develop and evaluate a system that uses supervised machine learning techniques to automatically estimate specific "characteristics" of a household from its electricity consumption. The characteristics are related to a household's socio-economic status, its dwelling, or its appliance stock.

We evaluate our approach by analyzing smart meter data collected from 4232 households in Ireland at a 30-min granularity over a period of 1.5 years. Our analysis shows that revealing characteristics from smart meter data is feasible, as our method achieves an accuracy of more than 70% over all households for many of the characteristics and even exceeds 80% for some of the characteristics.

The findings are applicable to all smart metering systems without making changes to the measurement infrastructure. The inferred knowledge paves the way for targeted energy efficiency programs and other services that benefit from improved customer insights. On the basis of these promising results, the paper discusses the potential for utilities as well as policy and privacy implications.

4.14 Smart Metering: What Drives the Impact of Behavior-specific Feedback

Verena Tiefenbeck (ETH Zürich, CH)

Transparency of consumption is supposed to foster energy efficient behavior, but the conservation effect in smart metering trials has been smaller than expected. I presented the results of a study involving 697 Swiss households on the impact of real-time hot water consumption feedback using a new metering device. We observed a stable average reduction of 23% of both energy and water use in the shower.

4.15 EV Fast Charging on German Highways

Victor del Razo (TU München, DE)

The limited driving range of electric vehicles makes them a sub-optimal alternative for long distance trips, particularly on highways, where higher driving speed has a negative effect on power consumption. In this talk, we discussed alternatives for using ICT for reducing the overall driving time, while keeping the required additional infrastructure at a minimum. Through a route optimization and charging time reservation system we reduce the trip duration and make the energy demand from power stations more predictable.

5 Reports from the Breakout Groups

5.1 Smart Grid Data Analytics

Bert Claessens, Nicolas Gast, Christoph Goebel, Mario Paolone, Anthony Papavasiliou, Jose Rivera, Joachim Sokol, Andreas Veit, and Holger Ziekow

Ecosystem

The smart (electric) grid ecosystem encompasses many stakeholders with well-defined tasks and objectives. These stakeholders include conventional and renewable power generation companies (gencos), electricity retailers, industry, aggregators, con- and prosumers (end customers on low voltage side), transmission system operators (TSOs), distribution system operators (DSOs), and market operators, technology providers, as well as organizations that play several stakeholder roles at the same time. The integration of renewable power generation into the electric grid leads to new challenges and opportunities in operational control, long-term capacity planning, and business strategy (including new business models). Information technology will play a major role in coping with these challenges as well as taking advantage of new business opportunities. This report focuses in particular on the role of measurement data and how current and future ICT applications of different stakeholders may take advantage of more and more of this data becoming available from different sources. The contents of this report are structured along the stakeholder axis. Applications including state-of-the-art and opportunities are therefore described on the stakeholder level.

Renewable Power Generation Companies

Renewable power generation companies include, e.g., wind farms, solar PV plants, and hydro power plants. In general, renewable gencos are interested in maximizing their power output, which in contrast to conventional generation depends on variable environmental conditions. This application is called maximum power point tracking. For instance, the blades of wind turbines can be adjusted to extract the maximum a. c. power from the wind based on the characteristics of the turbine. The mechanisms for maximum power point tracking could be further improved by data analytics, e.g., to coordinate the control of single generation units interconnected in larger wind farms or solar PV installations (advanced maximum power point tracking applications).

Wind and solar power generation is highly variable. Thus, if renewable gencos have to fully participate in electricity markets, which penalize deviations from scheduled market bids, more accurate predictions of their power output will become economically advantageous. While renewable gencos, similar to other gencos, monitor their total power production over time (market participation requirement), they often don't correlate this data with other potentially available data, e.g., weather data and forecasts. Moreover, many of them do not forecast output since they do not have economic incentives to do so if the receive a guarantee that their entire production is fed into the grid (which is the case under the current subsidy scheme in Germany, for instance). Output prediction is therefore an important data-based application that will become more important as soon as renewable gencos are forced to become regular market participants. Another important application based on the data collected from sensors in wind and solar power generation units is predictive maintenance, which allows gencos to predict the possible failure of generation equipment and therefore opens up cost savings opportunities in the maintenance area.

Current data sources:
- Metering and monitoring data from generators
- Generator configuration data

Current applications:
- Operational control
- Maximum power point tracking
- Health monitoring

Future data sources:
- Local weather data and forecasts
- Additional monitoring data from generators
- Historical configuration and power output data on the generation unit level

Future applications:
- Power output predictions on multiple time scales for market participation
- Advanced maximum power point tracking
- Predictive maintenance

Data volume estimate: from small to medium

Data velocity estimate: from small to medium

Transmission Grid Operators

Transmission grid operators are in charge of operating transmission grids (high voltage) within the secure region to prevent instability that could lead to blackouts. They own sophisticated models of the transmission grid which they use to infer the current grid state (e.g., current on all lines, voltage angles, etc.) using the available real-time data from generators, substations, high-voltage transmission lines, etc. This data is transferred to the

TSO's back-end system, which can happen at very high data rates based on the type of sensors deployed on the transmission grid level. For instance, phasor measurement units can transmit updated measurements at rates of up to 10 kHz.

To extrapolate the system state into the future and react to potential threats, TSOs access additional data sources, in particular weather data and market-cleared schedules of dispatchable generators. Their portfolio of countermeasures includes reconfiguration of the grid structure (e.g., by disconnecting transmission lines), activation of reserves, re-dispatch of generators, and load shedding. The more data can be accessed (higher granularity, special resolution, etc.), the higher the probability that the accuracy of statistical models will improve and enable more accurate extrapolations of the system state. Apart from operational control, TSOs are involved in long-term capacity planning. They are responsible for extending the transmission grid's capacity and provide reserves in close coordination with regulatory institutions. This capacity planning process is based on historical operation data as well as longer term models describing the evolution of the underlying variables.

Current data sources:
- Metering and monitoring data on the transmission level
- Weather data and forecasts
- Historical system states (currently produced at up to 30Hz)

Current applications:
- Real-time state estimation and visualization
- Contingency analysis (n-1)
- Decision support for operational control including reconfiguration, activation of reserves, generator re-dispatch, load shedding
- Reserve provision (usually via auctions)
- Decision support for long-term capacity planning

Future data sources:
- More detailed weather data and forecasts
- Additional monitoring data from generators, substations, and transmission lines using PMUs

Future applications:
- Advanced state estimation (probabilistic, multiple time scales, etc.)
- Advanced decision support
- Advanced long-term planning based on models and historical data

Data volume estimate: from low to high

Data velocity estimate: constantly high

Distribution Grid Operators

Distribution system operators are responsible for assuring the power quality and supply security in power distribution systems. They are currently able to monitor relevant metrics at substations, but have little visibility downstream, i.e., about the conditions at the end consumers. The can perform voltage regulation at the substation level by switching, transformer tap changes, or reactive power injections, but usually cannot remote-control any of the elements further downstream (e.g., protection or voltage regulators), which mostly operate independently. Distributed generation, load flexibility, and distributed energy storage will in the long term complicate the traditional tasks of DSOs described above.

Current data sources:
- Substation monitoring data

Current applications:
- Voltage regulation using tap transformers
- Switching at substations
- Long-term capacity planning (transformers, power lines, etc.)

Future data sources:
- Smart meter data
- Data from RTUs and PMUs deployed on the distribution level
- Local weather data and forecasts
- Detailed power output data from distributed generation, especially solar PV

Future applications:
- State estimation for distribution grids (probabilistic, multiple time scales, etc.)
- Advanced decision support from distribution system operation, in particular voltage regulation
- Advanced long-term planning based on models and historical data

Data volume estimate: from low to high

Data velocity estimate: from low to high

Aggregators

Aggregators bring the flexibility of their customers (industrial loads, thermal loads, bio gas plants, etc.) to market by buying the right to control them within certain limits. Due to the ongoing integration of variable renewable power generation into the grid, more flexibility is needed, which will eventually be reflected in more short-term trading and higher prices for flexibility, e.g., reserves. Once the large and obvious sources of flexibility (e.g., cooling houses, large commercial buildings, industrial loads with high flexibility) have been brought to market, smaller and less obvious resources may be accessed (e.g., smaller buildings, EVs, solar-attached storage). The data analytics requirements will therefore increase: The better aggregators can predict the capability of resources over time, the more efficiently they can dispatch these resources to maximize profits.

Current data sources:
- Resource monitoring data (e.g., temperature, power consumption, etc.)
- Resource meta data and models (e.g., battery capacity, charging/discharging rates, etc.)
- Market data (prices, bids, etc.)

Current applications:
- Market participation (reserve and spot markets) via resource dispatch

Future data sources:
- More detailed and accurate resource monitoring data, data from new types of resources (e.g., EVs, residential batteries, HVAC systems, etc.)
- Historical resource monitoring data and controls
- Weather data and forecasts

Future applications:
- Use of more accurate resource models for more profitable control / more efficient use of resource pool
- Advanced resource scheduling techniques applicable to large pools of heterogeneous resources (stochastic optimization, etc.)

Data volume estimate: from low to high

Data velocity estimate: constantly medium

Retailers

The business model of electricity retailers is to buy energy on the markets and sell it end consumers. While electricity prices on the wholesale markets vary over time, end consumers usually pay a fixed price. They can only make a profit if they manage to buy their electricity in a smart way and negotiate sufficiently low transmission and distribution prices with TSOs and DSOs. Imbalances of supply and demand decrease their profit since they result in penalties. Therefore retailers are interested in accurate predictions of electricity prices and demand. Electricity prices and demand will in the future depend more heavily on weather conditions, thus retailers may be interested in corresponding prediction services.

Current data sources:
- Meter data (nowadays usually measured 1-2 times a year)
- Customer data (address, payments, contract, etc.)
- Wholesale market prices

Current applications:
- Data analytics for market price and demand forecasting (OTC, spot market)
- Marketing, sales, billing

Future data sources:
- Smart meter data
- Weather data and forecasts

Future applications:
- Data analytics for market price and demand forecasting
- Data analytics for marketing and sales based on smart meter data
- More profitable market participation based on better supply and demand forecasting

Data volume estimate: from low to medium

Data velocity estimate: constantly low

Prosumers

Prosumers are end consumers of electricity that actively manage their consumption and may generate electricity on their own. As more accurate and detailed data on their own electricity usage becomes available to them, they can take advantage of innovative systems for consumption feedback and automatic energy saving mechanisms (NEST being a good example for such a system). If retailers offer time of use tariffs, such systems can also be used to shift load and save some money on the electricity bill. With more and more distributed generation and storage being deployed in the future, advanced energy management systems will enable even lower energy provision costs for prosumers. The higher the amount of relevant data that these systems can access gets, the more efficiently they can fulfill their purpose. Relevant data includes in particular data on the bounds of user comfort (e.g., temperature), user behavior (e.g., user at home or not), and environmental conditions (weather data and forecasts). It will be interesting to see if services can be established that transfer knowledge extracted from the pooled data of a given population of prosumers to new prosumers installing energy management systems in their homes.

Current data sources:
- Smart meter data
- Smart home sensors (dis-aggregated electricity demand, temperature sensors, etc.)

Current applications:
- Consumption feedback, e.g., to achieve higher energy efficiency
- Savings on electricity bill via load shifting (if time of use tariffs are available)
- Smart home control and automation (e.g., NEST)

Future data sources:
- Data from "personal sensors" – (e.g., smartphones, smart wristbands, etc.)
- Local weather data and predictions
- Derived preference data (rules), possibly based on populations of prosumers
- Derived activity data (rules), possibly based on populations of prosumers

Future applications:
- Advanced building energy management systems to achieve energy efficiency, load flexibility, and higher degree of autarky

Data volume estimate: from low to medium

Data velocity estimate: constantly medium

5.2 Smart Grid Communications

Jean-Yves Le Boudec, Srinivasan Keshav, Hermann de Meer, Florian Michahelles, Peter Noglik, Victor del Razo, and Kai Strunz

Summary

Communications are at the heart of the Smart Grid: sensing the status of grid assets and loads and controlling them depends critically on the availability of an underlying communication infrastructure. This infrastructure needs to span a wide geographic scale across multiple continents, yet provide a high density of communication endpoints, such as tens or hundreds of sensors in a single room.

Smart Grid communications are diverse in many dimensions: Smart Grid communications takes place over both private networks and public networks. Most communications today is over private networks, for security and reliability, but we expect that this will shift eventually to VPN communications over a public network. Communications use both wired and wireless links. The links are attached to a diverse set of endpoints – from tiny sensors to multi-million-dollar PMUs. Endpoints have a diverse set of performance requirements technologies used for communication links are also diverse, and depend on the geographic scale of communication.

Recommendations

Communication network technology is mature. Nevertheless, we have identified several avenues for research, that we discuss next. In the short term, we believe that there is the need for increasing the scale of communications at an economically-affordable cost per carried bit. This could be used, for instance, to carry sensor values over IP multicast or to obtain smart meter readings from millions of endpoints. There is also a need for reliable, very low-latency communication between controllers and sensors and controllers and actuators. In the mid-term, we see a strong need for integration of legacy wired/wireless technologies into IP for both WAN and LAN applications. We also see the need to upgrade and maintain the communication infrastructure and achieve reliability in communications, such as by authenticated sensor readings and tamper-proof communication. An interesting challenge is to assure communication despite loss of electrical power, so that the communication network can be used to restart the grid.

There is a need for research into achieving low-latency, low jitter, reliable, tamper-proof communication for sensing and control. In the long term, we believe that the research focus should be on infrastructure security and dependability. This would require solutions to problems such as securing devices despite physical access by attackers, and ensuring that emergency "back doors" to communication equipment does not compromise security. A more ambitious goal is to add intelligence to the communication network so that we evolve from a Smart Grid to a Semantic Smart Grid.

Appendix

5.2.1 Geographic Scale and Endpoints

Table 1 Taxonomy of Smart Grid communication.

<table>
<tr><th></th><th>WAN/MAN – Public</th><th>WAN/MAN – Private</th><th>LAN (Private)</th></tr>
<tr><td>Wired</td><td rowspan="2">RTUs
HEMS (Gateway)
EV charger
PV inverters</td><td>PMU
Protection devices
RTUs
STN
Sensors
PV inverters</td><td rowspan="2">HEMS (Gateway)
EV chargers
PV inverters
Sensors
Storage management</td></tr>
<tr><td>Wireless</td><td>Sensors
PV inverters</td></tr>
</table>

A taxonomy of Smart Grid communication can be structured in different ways. The way chosen here is according to the geographical scale of communication and the connection endpoints. In addition, we distinguish between public/private and wired/wireless communication.

On geographical level we have have two primary scales:

- Wide area / Metropolitan area networks
- Local area networks

WAN/MAN are usually used to connect to endpoints which already have collected and/or aggregated data. Typical endpoints are, for example, RTUs (Remote Terminal Units) in primary and secondary substations and HEMS (Home Energy Management Systems), EV

Chargers, PMU (Phasor measurement units) and so forth. The communication media can be wireless as well as wired with different flavors. The choice of network technology depends on the requirements of the application and will be discussed later. As of today, a considerable part of the communication is done over private networks which are built for dedicated usage. This closed communication is increasingly becoming public. To secure it against cyber attacks, usually a VPN with encryption is used. There is a wide range of well-defined and accepted protocols which are used for the purpose.

For local area networks, all networks are private. The connection endpoints are here the data concentrators, like HEMS, as well the sensors and actuators. Actuators are not only, for example, blind controls, but also EV-charging systems. Even in a private network, encryption has become more and more advanced to meet cyber security requirements.

Many communication protocols are established in the market. Some of them are well defined and accepted like IEC61850 or KNX, but there are also many proprietary protocols, which makes it very difficult to mix endpoints from different vendors.

5.2.1.1 Requirements

This section summarizes the communication requirements of future Smart Grid networks.

Interoperability. The Smart Grid will be composed of multiple grid systems which have to share and exchange information. Thus, a common understanding of exchanged information and interfaces has to be established, even between equipment bought from multiple vendors.

Ability to upgrade/maintain. In order to cope with future needs and requirements the Smart Grid has to be flexible to incorporate evolving technologies. Development and adoption of standards could help to avoid customized efforts but to maximize utilization for all users.

Reliability (despite grid failure). Adding information technologies to the power-grid should improve the reliability of the grid, limit the extent of breakdowns, accelerate recovery from failures, and establish self-healing of nodes. Additionally, as the grid goes down the communication system has to remain active in order to take control of the grid and manage the recovery.

"Low" cost. Despite the merits and expectations of an evolving Smart Grid, its implementation and operation have still to be cost-efficient. While there is an understanding that today's energy prices are too low, it's unclear how consumers and commercial sector depending on power would react on significant price increases. Introductions of Smart Grid have to be incremental in order to balance costs, collect experience, minimize failures and develop business models, incentive models and education of customers correspondingly.

Delay bounds. Total delays of data in a Smart Grid's control must not exceed certain delay limits in order to the requirement of a reliably Smart Grid. Thus, power grid control may need its own dedicated private networks or should be prioritized on public networks.

Throughput needed. A Smart Grid's control commands must have a high rate of successful delivery across communication channels being used. Communication channels, public or private, have to be designed in order to provide enough bandwidth for transmitting also the maximum of control communication (e.g., in emergencies) successfully.

Error rate bounds. In the case of bandwidth requirements exceeding the network availability, error rates should be minimized in order not to exceed the delay for a successful grid control.

Clock/time distribution (PTP) (synchronicity). Clocks of all components involved in the control of the Smart Grid have to synchronized in order to allow for an efficient control in the distributed network.

Privacy. Privacy has to be protected of both the Smart Grid operators and its users connected by public networks. The operators' managing information has to be protected from the users as well as the individual user-specific characteristics of using the grid from the operators.

Security. The smart power grid will be controlled by an information communication system whose confidentiality, integrity and availability has to be protected. The security measures have to be applied against system failures, use control errors and external events.

5.2.1.2 Technologies

Communication network technologies are mature and well understood. For completeness, these are summarized in the table below.

Table 2 Communication network technologies.

	WAN/MAN	LAN (Private)
Wired	Fiber Optics PLC A number of proprietary technologies	Ethernet BACNet and Others Powerline PRP
Wireless	3G/4G Proprietary	WiFi Zigbee Bluetooth WiFi-Direct Many proprietary technologies

5.2.1.3 Research Opportunities

Near-term research goals (2 years out):
- Scalable IP multicast that works for wide area (short-to-mid-term)
- Scalable and interoperable communication paths (for smart meters)
- Reliable (multi-path), practical and cost-effective communication
- Guaranteed very low latencies in WAN/MAN for real-time control applications

Mid-term research goals (5 years out):
- Integration of legacy wired/wireless technologies into IP for both WAN and LAN
- Ability to upgrade/maintain communication infrastructure
- Authenticated sensor readings
- Making communication tamper proof
- Reliability of communication despite power network failures

Long-term research goals (10 years out):
- Securing devices despite physical access by attackers (booting, etc.)
- Going from communication to semantic SG
- Emergency back doors that are safe
- Localization and integration of pervasive sensors in secure way (long-term)

5.3 Smart Grid Control

Bert Claessens , Frank Eliassen, Christoph Goebel, Kai Heussen, S. Keshav, Hermann de Meer, Johanna Myrzik, and Mario Paolone

Summary

Control is pervasive in the existing grid and will play an even more critical role in the future Smart Grid. Control actions, the players taking these actions, their objectives, and the control mechanisms themselves are diverse, complex, and sometimes mutually conflicting. Control actions today span continents (such as with HVDC interconnects) and 12 orders of magnitude in terms of time-scales of control; from milliseconds to decades (if we may interpret planning as a form of control). They are taken by entities such as government bodies and market regulators, as well as by transmission system operators (TSOs) and distribution system operators (DSOs). The elements being controlled include equipment such as load tap changers and PV inverters but also some of the entities themselves (for example: the establishment of a grid code by a government is one way for them to control TSOs; a TSO requests demand response via an aggregator who is responsible for the actual unit control).

Not surprisingly, the objectives of control are also diverse, ranging from supply security and greenhouse gas mitigation to technical frequency and voltage regulation. These objectives are achieved by a number of mechanisms, including day-ahead and hourly markets, establishment of regulatory legislation, changing transformer taps, and topology reconfiguration through sectioning. One outcome of our work is a comprehensive analysis of the numerous control mechanisms in common use today, which can be found in the appendix.

Recommendations

Based on our analysis, we make a number of recommendations for research directions in the area of Smart Grid control. In the short term, we suggest the study of novel control policies for decentralized, policy-based control of voltage & frequency, and coming up with better models for demand-response capabilities of loads (i.e., characterize their flexibility). It would also be interesting to pin-point and eliminate inefficiencies in market design, to adapt to characteristics of changing energy resources and uncertainty.

In the short-to-medium term, we suggest a number of research areas. A few are discussed here, details of the rest can be found in the Appendix. We suggest research into better prediction of loads and stochastic generation, especially at short time-scales. We would also like to predict not just loads and generation but also characterize the uncertainty and variability in this forecast and integrating improved forecasts into state estimation, decision support and control systems. We suggest studying innovative bid types in market that take into account energy constraints as well as load flexibility. We also advocate research into optimal rules for storage operation and better state estimation. In the medium term, we believe that it is critical to allow SCADA to scale to much larger sets of inputs and processing frame-rate (a "SCADA on steroids").

In the medium-to-long term, we believe that we need to study the supervision and design of dependable self-reconfiguring control systems that can act semi-autonomously on behalf of the system operators. We also believe that it is important to study how uncertainty can be quantified in market clearing to represent actual cost of uncertainty and promote the value of flexibility resources.

In the long term, we recommend studying the merging of control-based & data-driven control in distribution sub-systems. Another (more radical) idea would be to design and operate fully-local and autonomous micro-grids that can operate entirely independent from the grid. It may perhaps be possible to eliminate the grid altogether, if resource availability in each microgrid exceeds the benefits from grid connection. This may be the best way to allow renewables integration at low cost and without impact on system stability. Even with grid connection, such a solution offers higher availability than the present grid when faced with systemic failure.

Appendix

The following points briefly expand on the summary, adding some important details.

5.3.1 Geographic Scale

- Inter-continental interconnection
- Super grid (HVDC)
- Country level
- Regional
- Primary substation
- MV feeder
- LV grid
- Single building
- Single room

5.3.2 Time Scales

- Decades
- Investment horizon (1-2 years)
- Seasonal (6 months)
- Day ahead
- Intra-day (hourly)
- Tertiary (balancing)
- Secondary (60 s)
- Primary (0-10 s, droop and inertia)
- Protection (100 ms)

5.3.3 Players

- P0. International bodies (IEC, IEEE), equipment suppliers
- P1. Continent wide regulators (ENTSOE, FERC, . . .)
- P2. SuperTSOs[1], market operators, electricity authorities (government)
- P3a. Gencos, balancing operators, aggregators
- P3b. Pure TSOs, DSOs
- P4. Retailers, aggregators, [micro-grid operators]
- P5. Industrial, commercial, and residential consumers, prosumers, aggregators
- P6. Individuals, plant operators

[1] By SuperTSO, we mean an entity that provides both transmission and market making services, such as those found in Germany. In contrast a Pure TSO provides only high voltage transmission.

5.3.4 Control Objectives

- O1. Supply security
- O2. Greenhouse gas mitigation
- O3. Energy affordability
- O4. Risk assessment and management
- O5. Minimizing COE / max. profit
- O6. Congestion management (transmission line planning)
- O7. Supply security – reserve contracts
- O8. Balance (day-ahead and faster)
- O9. Frequency stabilization
- O10. Voltage stabilization
- O11. Rotor angle stabilization
- O12. Protection
- O13. Intra-day portfolio balancing
- O14. Management of performance requirements for ancillary services

5.3.5 Control Elements

- E1. Power plants / generators
- E2. Transmission line switches and topology
- E3. Grid inverters
- E4. Reactive compensators
- E5. FACTS and universal power flow controllers
- E6. HVDC point-to-point
- E7. OLTC
- E8. Controllable load
- E9. DG elements
- E10. Inverters
- E11. Energy storage devices
- E12. Asynchronous generators
- E13. Reclosers/switches
- E14. Protections
- E15. Power conditioners

5.3.6 Mechanisms

A mechanism is characterized as: "PLAYER meets (CONTROL) OBJECTIVE by controlling ELEMENT/PLAYER using MECHANISM"

5.3.7 Detailed Recommendations

Time-scales: S = Short, M = Medium, L = Long

- Improve market design (S-L)
- Fully local micro-grid, total decoupling (L)
- Novel control policies for decentralized policy based control of voltage and frequency (S)
- Innovative bid types in markets: energy constraint, flexibility (S-M)
- Quantifying uncertainty as part of clearing strategy (M-L)
- Performance quantification and service requirements of new ancillary services (S-M)

Table 3 Smart Grid Control Mechanisms.

Player	Objective	Element/ Players	Mechanism
P0	O1, O2, O9, O10, O11, O12	E1, P0	Standards, making GHG measurable
P1, P2	O1-O8	P2, P3, P4, P5, P6	Grid code, FIT and subsidies, market regulation
P2	O4-O12	P3, E2-7, E10, E11	Bid types (product types), dispatching
P3a	O5, O13, O14	E8, E9	DR, droop and compound control, dispatch
P3b	O4, O9-12	E2-7, E9-11, P3a	Real-time monitoring through SCADA, droop/comp control, ancillary dispatch, reconfiguration, state estimation, prediction, dispatch aggregators
P4	O5	E8, E9	Monitoring, prediction, arbitrage
P5, P6	O1, O2, O5, O10	E9, E12	Turn on/off, set preference for DR, choose tariffs, choose aggregators, install PV, install energy storage

- High-frame rate optimal control (decisions per second) (S-M)
- Scalable SCADA systems – SCADA on steroids (S-M)
- Supervision of autonomous control systems (M-L)
- Better state estimation (real-time) (S-M)
- OPF and reconfiguration – optimal solution (S-M)
- Better prediction
- Ultra-short-term (S-M)
- Predict uncertainty
- Reconfigurable control system (M-L)
- Self-organization
- Model DR capabilities
- Unified modeling framework (S-M)
- Merging of control-based and data-driven control in DSS (Distributed Storage Systems)
- Controller conflict detection at all levels (S-M)
- Optimal rules for storage operation (S-M)
- Minimizing number of sensors in Smart Grid (S-M)
- Synchronicity of control structures and asynchronous control architectures (M-L)

5.4 Smart Commercial Buildings

Florian Allerding, Birger Becker, Christoph Doblander, Frank Eliassen, Manuel Götz, Hanno Hildmann, Hans-Arno Jacobsen, Peter Nogilik, Jiri Rojicek, Mischa Schmidt, Verena Tiefenbeck, Anwar Ul Haq, and Holger Ziekow

Today's Problems

Many of today's commercial buildings are equipped with some level of instrumentation and automation, addressable through more or less sophisticated installed Building Management

Systems (BMS). However, for advanced analytics and optimized control, information about the location of devices and systems is needed, too. While ontologies and standards like Industry Foundation Classes (IFC) exist in some buildings there is often no consolidated mapping of a building's structure, its systems and the devices. The process of creating and maintaining this Building Information model (BIM) is labor intensive.

A further problem in current building operation constitutes the often encountered contradiction of control actions, e.g., cooling and heating being active simultaneously. This is a symptom caused by a deeper rooted lack of orchestration of different building systems. Often individual energy systems are optimized for their purpose, but building control and management does not take complex interactions into account.

Also, over the lifetime of a building, there may be multiple retrofits, additions, upgrades of the building structure, its use and/or its instrumentation. Often equipment from multiple vendors is installed, giving rise to potential compatibility issues and proper documentation of installed systems is often found to be lacking. Very often this puts the BMS which integrates the different pieces of equipment into a powerful position where in essence the owner is in a BMS vendor lock-in situation.

Another challenge is the high degree of heterogeneity of commercial buildings; many of them are customized to individual functional or specific geographic requirements. In combination with a high degree of complexity of the supporting building infrastructure (e.g., HVAC or security system) and a variety of standards and communication protocols used by different vendors, solutions are difficult to transfer from one building to another. In addition to that, continuous training of the building operation staff is required to ensure that technical upgrades and new functionalities of the building are understood and not tampered with.

State-of-the-Art

There are available a number of commercial services offering energy audits, fault detection, diagnostics, and other information necessary to make energy-efficient and money-saving business decisions. Those services create value for building owners, but based on proprietary platforms and, as such, not open for further improvements by the community.

5.4.1 Automating Energy Audits, Fault Detection, and Diagnostics

Today, energy audits and diagnostics largely rely on experts that analyze the captured data with limited tool support. Inefficiencies and system faults are detected, e.g., through manual analysis of descriptive statistics and visualizations of sensor data. The rather high degree of human participation in this process limits the scalability of energy services.

5.4.2 Plug & Play Configuration

Setting up building sensors/actuators still requires a large degree of manual configuration. This includes the definition of communication endpoints as well as capturing the context and semantics of each sensor/actuator. An additional challenge is, that the configuration and context of sensors may change over time, and has to be reflected in the system.

5.4.3 Building Commodity Interactions

While building simulation models take interactions among different building systems into account, e.g., the additional warmth created by lighting when switched on, so that additional heating may not be required to reach comfort parameters, BMS currently do not consider this kind of interplay.

Research Challenges

5.4.4 Automating Energy Audits, Fault Detection, and Diagnostics

A challenge for future research is to increase the degree of automation in audit, fault detection, and diagnostic processes. New analytics methods need to be developed that automatically detect inefficiencies and faults as well as assist in defining corresponding actions. The prospect to increase the degree of automation can make corresponding energy services available at a larger scale and to more users. Energy experts may still be part of the process but may be able to server more installations though better analytics support. The automation of this process may benefit from better integrating BIM data with information about the building instrumentation.

5.4.5 Plug & Play Configuration

Future research should address solutions that support the process configuring sensors and the maintenance of this configuration. This goes beyond automatically establishing communication channels but includes mechanisms to capture as well as maintain the context and semantics of an installation. Ideally, adding/changing or removing a sensor in an application should be limited to executing the physical deployment and not require any further human interaction with the system. In this context also the location within the building of the new added sensor should be automatically identified.

5.4.6 Multi-tenancy

In commercial buildings it is common to find multiple tenants (organizations). BMS should support, e.g., via a virtualization concept, a separation of concerns and allow tenants to individually access and control systems of their individual concern, while resolving situations where tenants have conflicting interests causing inefficiencies. On a more individual level, individuals have different preferences, e.g., with respect to thermal comfort or lighting. As a result, to the extent that these preferences are compatible with other individuals' preferences, building occupants should have the possibility to adjust settings for sub-spaces (e.g., individual offices). Ideally, the operation system learns individual preferences and automatically balances local system states between conflicting individual preferences.

5.4.7 IT/Cyber Security

In any case, the integrity of current and future BMS must be ensured. It must be possible to identify a sensor/actor and make sure that this device is exactly the device that the system thinks it is. Furthermore encryption is must. Any physical access to sensors/actors must be detected by the system and not lead into unrecognized exchange of device with unpredictable system behavior.

5.4.8 Building Commodity Interactions

BMS need to be enhanced to optimize overall energy efficiency within the operational constraints by exploiting building commodity interaction effects. Moreover, buildings should be enabled to adapt their needs of the different energy forms to the utility requirements including power, gas, district heating etc. by exploiting these interactions.

5.4.9 Research Opportunities

Measuring energy demand on a plug-level requires today the usage of smart plugs. These are currently in the price-range of EUR 50-70. Even if there price will drop in the following years they will not scale economically. Hence, an alternative way to determine the energy use of individual devices needs to be found. Disaggregation approaches have come up recently promising to identify household appliances based on their characteristic fingerprint. Analyzing the aggregated overall consumption (e.g., at the Smart Meter) will allow the assignment of energy usage to each consumer. These methods need to become reliable and accurate.

5.4.10 Research Goals

Near-term research goals (2 years out): Anomaly detection today relies on a number of data points when many buildings are not well instrumented. To introduce methods utilizing virtual data points or working on less data points is a near term task. Also, enabling BMS to leverage building commodity interactions to optimize overall energy efficiency can and should be near-term research goal.

Mid-term research goals (5 years out): A medium-term research goal should be to enable the use of building commodity interactions to adapt building energy needs to utility DR requirements. Furthermore, the logic of current building systems usually are scoped for one tenant and hence can only optimize for this tenant. We therefore expect that future research will create system that optimize the building under considerations of all its inhabitants, which could be another mid-term research goal. For instance, the energy consumption will optimized under consideration of the building as a whole, and not, e.g., apartment by apartment. However, this optimization comes with challenges like mediating between conflicting goals and the increased complexity of the optimization problem. While supporting multi-tenancy, privacy and security are required to be maintained.

Long-term research goals (10 years out): In the long run (by 2025–2030), smart buildings will evolve to become semi-autonomous smart entities (as compared to today's rigid and inflexible shells which merely contain increasing numbers of smart elements) with varying degrees of autonomy. Such entities will be able to interact with the outside world in the same way an organism (comprised of a host of interconnected elements, all of which may have requirements and restrictions of their own and which are strongly dependent on each other and the organism) interacts with its environment. Abstractly speaking, the whole will become more than the sum of its parts. This may initially be restricted in the sense that the building manages and assigns resources between its parts, however the ability to do so in an independent manner and driven by objectives that are not necessarily shared with its internal elements will already constitute an improved level of "smartness" over the current state of affairs.

We envision buildings which effectively act as a layer between the different internal entities, resources and capabilities; and, in addition, between their internal entities and virtually all aspects of the outside world. The building-wide management and control of resources, ranging from parking slots to renewable energies, bandwidth, storage and supplies up to human know how and manpower will enable a dramatic increase in the efficiency of usage of these resources and as well as transparency, reliability, and accountability.

Shifting the jurisdiction over resources from the stakeholders to a smart semi-independent entity will enable the dynamic usage of resources, which today are still assigned permanently to one owner. This will both lower the cost for the resources incurred by the respective user, as well as increase the utilization thereof. As this evolution of buildings progresses across cityscapes, the ability to share resources and capacities between buildings and or their stakeholders will play an integral part in the evolution of our urban environments from relatively static institutions to truly smart cities.

5.5 Smart Residential Buildings

Florian Allerding, Birger Becker, Manuel Görtz, Jiri Rojicek, Hartmut Schmeck, Thorsten Staake, Kai Strunz, Verena Tiefenbeck, Anwar Ul Haq, and Andreas Veit

Problem Statement

Smart residential buildings share a number of properties with commercial buildings, but they also have distinct special characteristics. They refer to private living spaces in houses or apartment buildings, owned or rented by private persons, commonly referred to as smart homes. Investment decisions on infrastructure in residential buildings are quite often based on emotional or convenience arguments in contrast to being rational or business case driven. A typical motivation for investing into smart infrastructure may be the intention to demonstrate a "green" attitude. The users interacting with the smart residential building are the residents themselves and will thus usually be technically rather illiterate, hence needing out-of-the-box solutions which are mostly self-explaining. Nevertheless, in particular with respect to trustworthiness, installed services should provide interfaces for optional control on demand and they should inform transparently about their functionality and operational state. A private resident usually will not invest into completely integrated smart home solutions, but rather, over time, get a collection of verticals providing functionality for a special use case. The size of components in a residential building usually is much smaller than in a commercial scenario. The same is true for the expected lifetime, ranging from 3 to 5 years in a home to at least a decade in commercial buildings. Finally, a critical aspect with respect to the acceptability of smart home infrastructure is the aspect of security and privacy protection. In particular, security should be guaranteed by the systems while privacy aspects have to be under complete user control. This also needs guarantees from the verticals for compliance with privacy protection preferences. In simple words, a private resident wants to get perfect service support without being bothered by technical details, but having the option of complete control.

The objectives connected to smart home services depend very much on the resident's individual attitude. Most users are interested to increase comfort by building automation, with less priority on cost-effectiveness. To maximize the benefit of smart homes for users and operators, the smart home verticals are expected to provide value-added services, e.g., the maximization of self-supply or an improvement of maintenance aspects. An essential objective of a private resident may be to achieve independence from public infrastructure.

Nevertheless, users will even be able to provide services such as demand response themselves, potentially for making money with it or just for the feeling of doing something good for society. In any case, the integration of smart home technology into a building must not compromise the whole system's reliability.

State-of-the-Art

Currently, there are several players on the market offering specialized smart home solutions, ranging from home automation over ambient assisted living to energy monitoring and control. Examples include Dropcam, Nest, Netatmo, or Plugwise. They are "verticals", only rarely capable to interact with each other. Commonly, smart phones or existing home gateways are used as their hub to stream data into the vendor's cloud and to interact with devices in the home. Some new players appear on the market delivering "smart home" platforms for (mostly wireless) connection of household components, sensors, and actuators. These platforms offer the potential for interaction between different devices and their data, respectively. Quite often, they use a single open or proprietary near-field communication protocol (such as ZigBee, Z-Wave or Homematic) for connecting the devices. This heterogeneity of communication protocols is an essential barrier for home users, as the individual devices (household components, sensors, and actuators) are usually not compatible with each other. However, there are a few commercial approaches to combining several technologies into an integrated approach (like EE-Bus, Qivicon, or Homee). In connection with smart metering, some utilities (e.g., EnBW, RWE, Vattenfall) are offering metering data analytics and recommendation services. On a research level, several management platforms or "operating systems" for smart homes are emerging. Some examples are Organic Smart Home (OSH), OGEMA, FPAI, Eclipse Smart Home, ABASG, or Open Energy Monitor, offering platforms for monitoring, management and control functions for energy scenarios. A widely missing feature is plug-and-play as known from computers today. One approach to realize this feature for a management platform uses ZeroConf. Furthermore, security and privacy aspects are often neglected which can generate additional resistance of the customer to buy into the systems. Since private users are widely sensitive to energy efficiency labeling (like A++), additional "smart home readiness" – labels might positively influence buying decisions.

Outlook

Over the next 15 years, substantial progress in both the adoption as well as the capabilities of smart homes is to be expected. Future applications for smart homes require two key enablers: A sufficient level of infrastructure/instrumentation and well-established standards. In the light of EU directive 2009/72/EC, we expect a substantial increase in the deployment of electricity (and gas) smart meters in most EU countries in the short- to mid-term. In the mid-term (around 2022), we expect different home automation providers to establish themselves. Rather than classical energy providers, likely candidates include telecom providers and entertainment companies as they already have physical presence within residential buildings. While these home automation systems will be able to integrate some parts of the smart infrastructure, we expect to still see a broad range of independent verticals for surveillance, HVAC, load monitoring as well as entertainment and content.

In the long run (until 2030), we expect two key developments for smart homes. Firstly, we see a strong integration and advances of the smart capabilities of residential buildings, and secondly, we see an increased independence of smart residential areas from external infrastructure, like the electricity transmission grid. In-home infrastructure will be created which is able to accommodate various elements like PV panels, batteries, thermal storage, or EVs. This includes monitoring, analytics, control and a coordinated interplay of all elements involved. Regarding smart capabilities, we expect that the technology enabling intelligent features of smart homes will be almost invisible to the user. A key aspect will be the ease of use for the residents. Smart home systems will be able to observe the users' feelings and preferences with non-intrusive sensors in a variety of devices such as phones and wearables. The invisible intelligence of the smart home will be able to autonomously understand the user's needs and adapt its services accordingly. Although the residents will not be required to actively tell the home automation about their preferences, they will be able to overwrite autonomous control, if necessary. In fact, these systems will not only be able to adapt to the needs of single users, but also to understand and negotiate between the needs of groups of people in a room. Further, the home automation system will be able to automatically detect and integrate new devices that enter the building and know or learn their best utilization. Regarding the increased independence from external infrastructure, energy supply will decrease in importance. This is due to much more efficient building insulation and usage of distance heat from, say, factories and compute centers. Further, there will be significantly increased energy "harvesting" from within residential areas.

This increased independence of smart residential areas with a focus on decentralized systems could make high voltage power transmission obsolete, which would be a strongly disruptive development. Nevertheless, while this outlook refers to highly industrialized countries, heavily increasing energy demand in other parts of the world might lead to different scenarios, where the need for smart residential building automation is even stronger, in particular with a focus on energy conservation and demand optimization.

Research Challenges

The gap between state of the art and prospects of future smart home technologies translates into several research challenges at the intersection between ICT, energy systems, human behavior, and the policy framework. These challenges include the collection of data from the user and its environment as well as robust mechanisms to translate the data into stimulating and insightful information. Furthermore, the data has to serve as actionable input for automatic control systems in order to meet the highly individual user preferences including comfort, safety, and sustainability. Working towards these objectives requires considerable progress in the following fields:

To collect the data, a multitude of sensors and other data sources need to cooperate. These sensors need to be non-intrusive, lightweight, and energy-autarkic to neither burden the user nor cause high costs for installation and operation. A particular challenge is that the systems will often need to become part of existing infrastructures, and have to adhere to a multitude of domain-specific standards and characteristics.

Domain-specific machine learning techniques will be the second corner stone of smart buildings. The algorithms need to be capable of dealing with data time series of varying depth and quality. One major concern is the large variety of data characteristics influencing the accuracy of control and predictions.

Human Machine Interaction is another vital aspect. Looking at the state of the art of both, current products and research prototypes, it becomes apparent that the energy informatics community has to still go a long way to build systems that are easy to use for the general public and at the same time effectively motivate a desired behavior. In this domain, a closer cooperation with psychologists and behavioral economists might help to establish the tools and methods that trigger behavioral change.

Since integration of a large variety of verticals is essential, a core challenge relates to the design and dissemination of an adequate infrastructure operating system. It will have to provide typical services known from computer operating systems like resource scheduling and allocation, software updating, resilience and access control to name a few.

The multitude of highly personal data collected in smart homes imposes a challenge for data management and privacy. A framework is needed that provides the analysis of sensor data in a way that preserves users' privacy and maintains security. Similarly, the smart infrastructure needs to be well-protected against attempts to compromise the user's safety. This requires robust methods of authentication for users who want to access the building system as well as control processes for providers of software services for the building's infrastructure.

Life cycle analysis with respect to resource efficiency is another important aspect. It touches all fields from producing, shipping, operating, and recycling the growing number of smart devices.

Last but not least, whether residential building automation systems become a success will be largely determined by the underlying business cases and their attractiveness for service providers. Since much depends on the availability of consumption data (e.g., from smart metering), policy makers will have to find a delicate balance between limiting the use of data and privacy protection. The energy informatics community can contribute to these considerations by providing methods for effective data usage control.

5.6 Smart Transportation

Longbo Huang, Randy Katz, Victor del Razo, and Hans-Peter Schwefel

Breakout Focus

This breakout session focused on the research issues in Smart Transportation. In addition to commercial and residential buildings, transportation represents the largest source of energy consumption in industrialized societies (approximately 40%) and the largest source of greenhouse gas emissions, with critical implications for the earth's climate. The breakout considered both personal vehicles (cars) and public transportation (including bikes). There are many incentives for smarter transportation systems, including reduced energy consumption and GHG production, improved air quality with implications for health and quality of life, reduced societal externalities associated with the very high cost of car ownership, and the promise of reducing accidents and road deaths.

State-of-the-Art

The current state-of-the-art is largely dedicated user applications for assessing and reporting traffic congestion and estimated time of arrival to their cell phone users. Accuracy and timeliness can be an issue, resulting in the driver finding herself in congested traffic before it can be avoided. Such applications maintain proprietary data silos, with no sharing of trip data even though this is in principle owned by the users. Rerouting recommendations are not coordinated among such applications, which can lead to failure to improve congestion through rerouting suggestions. Generally the tools to help users out of private vehicles and on to public transport are limited in their effectiveness and usability.

Potential Game Changers

There are several potential game changers that are likely to challenge the current state of transportation systems. The first is the more IT-centered entrants in the vehicle sector, such as Tesla, Google, and Apple. The second is the rise of the so-called Sharing Economy, characterized by such firms as BikeShare, ZipCar, Uber, and Lyft. The third is the pervasive ability to track the location of virtually every vehicle on the road, whether it is publically available such as for public transit, or implicitly collected by such end user applications as Google Maps and Waze. A finally is disruption is the electrification of the vehicle fleet and the implications this has for the Grid and ready access to charging services.

Computer Science Challenges

The overarching computer science challenges in transportation can be stated as follows. The goal is to turn data about trips and transportation usage, collected from many sources and across many time scales, into actionable information for infrastructure and vehicle operators as well as passengers. The time scales vary from seconds, for safety and accident avoidance, to minutes to days for route options, including behavioral and economic incentives for alternative routing, load balancing, and road congestion avoidance, to months or years, for infrastructure planning and provisioning (e.g., charging stations, bike stations, bus routing, etc.). This will require a distributed and decentralized operational architecture to collect data, process it, and infer and decide at scale across a region and at the appropriate time scale. Such an architecture must be sensitive to concerns of information ownership, relevant business models, and privacy/security considerations.

Example Research Challenges

One identified challenge was EV range extension via route planning and charging station reservation, with capacity planning to inform charging station placement. Another was shared transportation resource planning and placement via demand and trip awareness (e.g., how many bikes and where to place them). This raises the important question of who owns mobility data and what is the business model for how it is collected and used. Privacy issues must be understood in such circumstances. A final example challenge was the definition and implementation of vehicle-to-vehicle and vehicle-to-infrastructure communications systems, to enable safety considerations, accident avoidance, and which span vehicles and railways, across dedicated or shared infrastructures.

Research Opportunities

The near-term research opportunity is to explore overlay architectures that allow the combination of multiple traffic sources for more accurate and timely congestion detection.

The medium-term research opportunities are to investigate how new vehicle sharing models impact the transportation system. Open for investigation is extensions of the overlay architecture to allow for intelligent rerouting with load balancing of traffic. Within such an architecture, introduce incentives to change operator and/or passenger behavior, such as migrate towards using higher density, more energy efficient transit modes such as public transit or shift modify travel times to avoid congestion. Large user studies should be undertaken. Finally, the study of the effect of new technologies, such as autonomous vehicles like cars or UAVs, on logistics systems should be undertaken.

The long-term research opportunities include investigations of the implications of high-penetration electrification of the vehicle fleet, such as, charging station infrastructure provisioning and placement. Also the value of vehicle trip data for managing the Smart Grid for EV charging should be investigated. Also ripe for inquiry is the impact of self-driving vehicles on the overall transportation system, from the perspective of fundamental changes to the existing car ownership model. These include effects migration to a largely shared vehicle fleet, with implications for road occupancy and parking, and the avoidance of new road and parking construction. Other implications to be studied include accident avoidance and road safety; ride sharing, trip scheduling, and road congestion avoidance; dynamic locating of shared vehicles to where they are likely to be needed within a city; and so on. The interface to Smart Cities should also be explored, including incorporating vehicles as full participants in the Internet of Things.

5.7 Data Crosscut

Manuel Görtz, Randy Katz, Srinivasan Keshav, Jiri Rojicek, Hans-Peter Schwefel, and Anwar Ul Haq

Breakout Focus

This breakout session focused on data issues within and across Smart Infrastructures. A particular focus of discussion was the Smart Grid, and the exchange of data between (aggregated) loads and entities that supply that load, across time scales and geographic regions. Discussion included the processes for collection, cleaning, processing, and curation of data, and how it forms inputs to decision making. Furthermore, there was an awareness that other sources, such as weather, events (e.g., World Cup), transportation and mobility, social networks, and so on, provide potentially useful indicators of human activity that influence energy demand and ultimately could be an input to the energy system.

State-of-the-Art

The current state-of-the-art was assessed starting with smart metering of end loads. While in some jurisdictions considerable effort has been dedicated to deploying smart meters, generally smart meter data is unused and not particularly useful. This is due in part to consumer resistance to dynamic pricing (the original motivation for smart meters) and a general mistrust of the utility operator. Generally, transmission and distribution operators have good quality forecasting tools to predict demand and provision the grid for the current state of grid architecture. Fine time grain data is not needed, since agile control has not yet been deployed. Since consumer level load data is not actually used, there is in fact no real privacy concerns. At the level of large-scale aggregation, operators are deploying better measurement infrastructure, in the form of PMUs (Phase Management Units), to better manage their networks in the face of increasingly complex load and supply dynamics, and grid interconnections. Nevertheless, these devices are expensive and generally limited to the transmission grid.

Potential Game Changers

There are several potential game changers that are likely to challenge the current state of data collection, processing, and use in the Grid. First is the emerging disaggregation of the Energy Network. The Energy Network is evolving from an integrated end-to-end system, to one that features looser connections between TSOs and DSOs, to one that may eventually become made up of semi-independent microgrids. The implication is the end of the "Law of Large Numbers" that allows statistical multiplexing to hide time variations in individual generation supplies and loads. The second game changer is greater penetration of edge PV generation and EV charging, yielding even higher time variation in supplies and loads. This will drive the requirement for information about instantaneous energy supply and consumption on finer time scales and smaller regions. The third trend is migration of intelligence towards the edge, with a greater prevalence of microgrids (e.g., office parks, campuses, industrial parks, shopping centers, etc.). We foresee a future Grid in which large aggregates of supplies and loads are replaced with smaller aggregates, managed with real-time intelligence for more localized control. We believe that natural unit of aggregation is likely to be at the level low voltage distribution (e.g., secondary transformers) supplying loads to something on the order of 10–50 homes.

Research Opportunities

The near-term research opportunities are to lower measurement and analysis costs by developing more rugged technologies for monitoring and performance analysis of the Grid. There is also an immediate need to improve customer awareness of energy usage, provide a better user experience, and deliver an improved perceived value of collecting edge energy usage data. This will require better tools and visualizations for consumers and other edge customers (e.g., building and campus facilities managers) to understand their detailed energy usage data. Finally, there is an opportunity to collect existing and to create new datasets of energy usage that can be made available to the research community for analyses at larger scales and over greater diversity.

The medium-term research opportunities are to use these new tools and data sets to better understand individual load profiles, e.g., to level of individual appliances usage, from house level load data (NILM: non-intrusive load monitoring). The data architecture that links the monitoring capability at the secondary substation level to the home load should be designed, prototyped, and evaluated in the context of dynamic Grid control. This architecture must be developed in a way that is sensitive to privacy issues, in part by using the appropriate level of aggregation and data partitioning to avoid tracking of individual behavior. A data processing architecture needs to be developed for characterizing and classifying of loads (clustering), tracking of data transformations and its long term archiving (curation), data placement and storage (collection), processing, and dissemination (up-sharing of aggregated and sampled data). Larger scale building and home data sets, including metadata, should be collected and made available for further study. Further analysis of the effects of human activity indicators on aggregated energy loads and microgrid coupled supply and load behavior should be undertaken, and control architectures and algorithms developed.

The long-term research opportunities is to design and demonstrate effective data-informed control of highly dynamic disaggregated loads and generation assets in a disaggregated Grid environment while understanding how automated exchange of data exchange across societal infrastructures can lead to better, more agile control algorithms.

5.8 Design Patterns and Paradigms for Smart Infrastructure

Frank Eliassen, Kai Heussen, Hanno Hildmann, Peter Noglik, Jose Rivera, and Hartmut Schmeck

Breakout Focus

The group discussed design patterns for the areas of Smart Buildings and Smart Grids. Buildings and electricity networks are primarily a built physical infrastructure with historically little embedded ICT, functionality of conventionally engineered buildings and electricity infrastructure has been conceived the same (obviously) inflexible structures as the physical structures that enable them: i.e., shelter, heat, or reliable electricity supply. The purpose of a deeper embedding and higher sophistication of software technology within these infrastructures would be to offer enhanced functionality and reconfigurability that has been achieved with software defined systems in other domains.

Design patterns inform high level choices that have to be made at an early stage in a system development process. A stereotypical choice is whether to take a centralized versus a decentralized approach or whether to use hierarchical structures – and if so, whether to do so strictly or loosely. A central management structure may be replaced by (or at least supplemented for some aspects) by self-* approaches so as to push some of the management overhead closer to the device layer. Monolithic architectures are receiving increasing competition from modular systems. These are of interest as composability allows for wider application of systems due to enhanced customization and facilitates the compartmentalization of problems.

The area of application is vast, but two aspects take a central role: the definition of suitable software platforms and the re-formulation of system control architecture. Engineering processes that consider jointly physical structure and interdependencies, control structures and software systems have to be developed. Engineering methods and tools are needed to support these considerations for off-line design work, run-time updates, and, particularly in context of electricity grids, also for re-engineering of the operational system (running systems).

Regarding the evaluation criteria, performance requirements and time constraints are as relevant as considerations for feasibility, resilience, reliability and stability, robustness and, last but not least, simplicity. Criteria that seem to have a high impact on the acceptance of a system range from disruptability (i.e., whether a platform or architecture can be amended while in use or whether parts of can be taken down without bringing the system to a halt), over versatility and vulnerability to trustworthiness. In addition, explainability and the identification and/or assignment of liability are of high importance to ensure wide acceptance of an approach.

A larger number of fields contributes to the state-of-the-art (see below), but common properties that seem to span across most of these fields are reconfigurability, adaptiveness and robustness, the ability to predict models and the ability to implement parts of the system in a distributed manner.

State-of-the-Art

- Control theory
- Trade-off matrix for different approaches

- Self *
- Autonomic computing
- Organic computing
- Multi-Agent Systems (MAS)
- Machine learning (learning systems)
- Emergence and emergent effects of local actions and interactions
- Power system control
- Software requirements engineering
- Cyber-physical simulations

Research Challenges and Opportunities

One of the main challenges we have identified is the need to consolidate in a coherent picture of the vast amount of knowledge available on this topic. It is often the case that several disciplines have developed the subject independently of each other, and many times unaware of the results achieved in other disciplines. For instance, we need to couple control engineering requirements with software/systems engineering principles, we need requirements engineering. Merging mathematical control theory with the self* is another important component, failsafe engineering. For this we will need to develop engineering process for self* systems and create the theoretical basis to understand them. The consolidation of this knowledge represents a key research challenge and will have a major impact in the design of future energy systems. We consider that a common benchmark platform that addresses the needs of different disciplines provides an opportunity to consolidate the different research efforts.

Several aspects like resilience, reliability versatility, and liability are also a major challenges and will play a key role in acceptance of these systems. There is a multitude of options we need to develop, e.g., we need to design self-reconfigurable cyber-physical architectures, we need architectures for real-time distributed MAS, we need to be able to do runtime deployment and, last but not least, we need to factor in the aspect of IT security. We consider that there will be a multitude of options, such that the community of researchers and practitioners will need a trade-off matrix of the different approaches.

Near-term research goals (2 years out):

- Trade-off matrix for different approaches
- Combine control engineering and software/systems engineering
- Engineering process for self* systems
- IT-security
- Runtime deployment
- Requirements engineering

Mid- and long-term research goals (5–10 years out):

- Merge self* with mathematical control theory
- Failsafe engineering
- Design rules

5.9 Human Machine Interaction in Energy Informatics

Birger Becker, Christoph Doblander, Johanna Myrzik, Thorsten Staake, and Verena Tiefenbeck

Problem Statement

Human Computer and Human Machine Interaction (HMI) concepts without doubt account for a large share of the success of many ICT products and services. Based on research results and learnings from practical deployments, industry has come a long way to better reflect the needs of the individuals using their systems. However, HMI in the energy domain faces specific challenges that are rooted in particular in the dynamic and multifaceted constraints and requirements, including the need to deal with time-critical changes on the supply side, the large number of energy-consuming applications and actions, and the often difficult to predict and highly situation-specific user intentions. Both, balancing energy supply and demand as well as energy efficiency increasingly has to follow supply without putting a too large burden on the end-user. Hence the degree of automation and prediction has to increase while preserving the users' freedom to intervene and overrule control systems.

Additional, interrelated challenges of HMI in the energy domain include:

- The relatively low average level of consumer interest in energy topics. In fact, energy itself is often neither visible nor of primary concern for the user since energy related expenses represent only a small share of a consumer's wallet: Moreover, energy conservation might develop into a pronounced social norm, but today's limited visibility of related actions limits the effort the user is willing to invest.
- Unclear benefits of smart energy system from many user's perspective. This is mainly the case due to the high level of comfort and reliability of the existing systems as well as due to the complexity of relationships of cause, action, and outcome of changes in energy systems.
- Short lived interest in energy dashboards. Closely related to the aforementioned issues, motivational "cues" are needed achieve a sustained system usage. Such "cues" are often not present in today's, engineering-oriented designs.
- Limited trust in utility companies. Systems (and the concepts behind marketing them) face the additional challenge that users do often not understand the motives of their energy providers behind offering energy efficiency products. This leads to distrust, which needs to be mitigated.
- Little tolerance for wrong system assumptions (e.g., cold shower water in the morning). Related to the reliable but inefficient existing system, user expectations are high. In most cases, developers will simply have to work to meet the requirements.
- Last but not least, user preference are inherently dynamic and situation specific (e.g., temperature preferences are not stable over time). Thus, prediction algorithms must be very accurate, or the system must offer a convenient way to allow for adjustments by the user.

State-of-the-Art

HMI concepts are used in different domains such as safety, security, and control. Energy-specific HMI concepts are needed to provide energy management systems with interfaces for user interaction. Systems that are available on the market are often domain- or application-specific, only vertically integrated (e.g., from a temperature and a dedicated motion sensor down to the heating control), and cannot be integrated with other systems. An example for an innovative system could be the NEST thermostat: The HMI has an innovative design, it is structured very well, and it is easy to use even for non IT-affine users. However, its "self-learning" features currently receive much criticism.

HMI approaches on a prototype level are able to provide flexible and configurable visualization and the capability for "complexity only on demand". This means that users can choose between individual levels of detail for visualization. Additionally, the status of self-adaptable systems will become visible for the user through the HMI.

In current research approaches for energy specific HMI, three different types of interaction are often distinguished: (i) visualization, (ii) parametrization, and (iii) configuration. In a first step, the state of the connected system will be visualized by current HMI solutions on a different degree of detail, so that the user is able to get information about the operation of single appliances or even of the overall system. Selective visualization may cause a change of user behavior. Additionally, parametrization of HMI allows the user to interact with the system during its operation concerning the individual demand. In this way, the individual parameterization enables the HMI to communicate personal preferences (e.g., the degree of freedom regarding the on-times of a washing machine) to the system. Configuration is the third interactive step of HMI, which allows the initial adaptation of the system to the real environment.

Vision

In the medium term, we expect an increased interest in energy management system, in particular among prosumers who are able to put the information in context with the energy production of their own systems. The increased interest is fueled by running out feed-in subsidies from the government, shifting goals towards self-supply. This requires HMI solutions which adapts to the behavior of the user. For this purpose, high reliability of the predictions and understanding of the user preferences are needed.

Information from these systems can be embedded into object-specific displays or mobile devices e.g., smart watches or phones. However, many other application compete for user attention, and particular attention has to be paid not to overwhelm the user. Through meaningful information display, the complexity can be decreased.

Motivated by the increasing share of intermittent generation, the time criticality and flexibility requires highly adaptive systems. Most of the decisions should be done by the system automatically, however the user should be always able to intervene. Cross domain interpretation of sensor data can be brought into future HMI applications. In general, HMI has the potential to increase trust in complex automation systems.

Research Challenges

The vision formulated before translates into a number of challenges for HMI research. This is especially true since several specific characteristics of energy supply and demand need to be considered: Energy – despite its enormous value for our society – is relatively cheap given the value it provides. It is a low-involvement good, with consumers not per se requiring energy but the services it enables (e.g., heat, light, telecommunication, etc.). At the same time, energy use is spread over a very large number of activities, and many devices need to be activated long before they can provide their service. These aspects add to the supply-side challenges and make the control – and with it the HMI – difficult. Among the many research challenges in HMI, the following bear a special reference to energy-related aspects:

- A large number of energy-consuming devices and activities will require many sensors and other data sources to arrive at a complete picture of energy demand, the state of the environment, and the user's objectives. This in turn translates into the need to effectively deploy and maintain many sensors, and to retrieve and combine the data from multiple sources. Many of the data sources will serve multiple purposes – they may been originally installed to increase comfort of safety, not energy efficiency – and the multi-faceted use will add to the complexity of their integration in energy systems. Energy informatics is thus probably one of the most advanced application domains for interoperability concepts.
- Raw data on energy supply, demand, the environment, and last but not least the user needs to be processed in order to reveal their underlying patterns. The insights from machine learning which range from electricity prices to the mood of the inhabitant are often required to trigger a target behavior (i.e., enable behavioral interventions) and render possible adaptive control systems that do not require any user interaction. Predictions in the energy domain are especially challenging as influencing factors are numerous, dynamic, and related to many application-specific characteristics. Yet predictions are important for many processes with high latencies. Even problems that may at first sight appear to be not hard – such as, for example, predicting the time of the following day an apartment is empty – are indeed very challenging, yet solving them could help to considerably conserve energy on heating.
- Energy systems often need to consider a large number of constraints that are highly user specific (internal: desired temperature, time, mood; external: weather, prices, etc.). Yet the expectations regarding the level of comfort are high, so states achieved by "smart systems" that are perceived as further away from the optimum than the outcome of conventional techniques (that are often comfortable yet energy intense) are hardly accepted by consumers. Optimization problems are hard in many energy applications, with extensive future work needed to arrive at suitable approaches for both the building and the transportation sector.
- A challenge for HMI that especially holds in the energy context is that complex information must be boiled down to a few easy-to-understand key figures in order to make the interaction between user and system feasible. Other than in health or nutrition, energy information is less tangible for the user. Nevertheless, complex or hidden relationships between behavior, energy use, and the consequences thereof need to be conveyed to the vast majority of non-energy literates including those who are not interested in becoming energy experts.Yet, a complete picture must be available upon demand (e.g., in case of failure), and the visualization interfaces must thus be configurable. Covering the balance between simplicity and in-depth information poses new challenges to the interface design.
- In order to trigger a target behavior regarding a non-visible, low-cost, and low-involvement good requires a very solid understanding of the behavioral concepts underlying human behavior. Unlike in health, fitness, or ostentatious consumption where the (perceived) benefits mighty be immediately felt, motivation to conserve energy is often smaller. An important research challenge certainly is to further develop interventions from time-invariant problems to tackle time-variant challenges (e.g., load shifting). Both visualization and interventions are more difficult to realize since many actions are time-dependent and inherently dynamic: There might be more than enough electricity available one day at noon and a shortage the next day at the same time. Timing is important. Interruptions need to be placed (of an automated service or a user action), without being annoying.

Moreover, the use of concepts from psychology and behavioral economy (e.g., social norms, goals, competition, etc.) should be given due consideration when working on HMI. Relying on rational choice models and monetary incentives – as mostly done today – is simply not sufficient.

The ultimate challenge probably is to make HMI disappear while making the residential building a place that – without consuming resources of its inhabitants – balances the requirements of all stakeholders. It will be a long and interesting way to work towards this goal.

5.10 Smart Cities

Kai Heussen, Hanno Hildmann, Longbo Huang, Milan Milenkovic, Johanna Myrzik, Jose Rivera, Misha Schmidt, Hans-Peter Schwefel, Joachim Sokol, and Randy Katz

Breakout Focus

Motto: "The battle for sustainability will be won or lost in (smart) cities."

Rather than attempting to (re)define the term "smart city", the team opted to outline its salient characteristic as an informed, data-driven, management of city and its operation for the benefit of citizens, and in accordance with their expressed consensus/preferences. The informed part comes from a combination of technologies, predominantly ICT and IOT based, including new and existing sensors, data crowd sourcing via social networks (existing and purpose built) for collecting citizen inputs in terms of preferences and observations, and opening of data from existing legacy city systems for integrated services. The vision is to create a city platform that integrates all of this information into a holistic view and allows creation of new applications and services to improve quality of life in the city and to streamline and increase efficiency of its operations. Over time, the city platform is expected to evolve to real-time sensing and reaction to events, to perform optimizations in accordance with policies derived from citizen preferences and to alleviate emergency situations when they occur. The city platform itself is supposed to provide mechanisms for realization of desired operational objectives, rather than assuming or imposing specific policies for doing so. Objectives and concerns of a city are generally driven by the desire to improve its attractiveness and quality of life, including: implementation of suitable infrastructure, access to affordable healthcare, good educational resources, transportation, supply of food and goods, social equity, mobility offerings, safety etc. The smart-city platform is expected to aid in achieving many of these objectives by providing metrics and data-driven basis for real-time management in accordance with objectives and priorities established by its citizens and government.

With regard to the city's inhabitants' objectives it is going to be a multitude of quality-of-life enhancing aspects rather than a single "killer app" that fuels the continued support and investment in the smart city platforms. The process is likely to be a long-term (decade(s) rather than years) process that will move forward with great momentum.

Today's Problems

The problem space of today of cities is different in different regions and vary from city to city. Nevertheless, major trends and drivers can be identified, which create pressure for actions and solutions:

- Population growth
- Increasing urbanization
- Increasing life expectancy
- Aging society
- Water scarcity
- Traffic congestion
- Pollution (air, soil, water, noise, waste)
- Energy supply
- Global warming
- Social divide
- Lack of funds
- Increased safety and security constraints

The list is not comprehensive, but it illustrates the complexity, magnitude, and dependencies involved in the underlying economic and societal pressures that challenge the rapid transition and change of city KPIs.

State-of-the-Art

The state-of-the-art is characterized by disjointed legacy management systems that will require an interoperability layer and addition of sensing fabric to evolve into a city platform that we envision. While there is a flurry of activity in smart city pilots and engagements worldwide, there are no major deployments that would validate the benefits of a smart city platform that we envision or clarify the business case.

- Disparate, uncoordinated systems for administration and management
- Individual siloed systems and administration, e.g., energy, buildings, water, transportation, energy
- Legacy infrastructure
- Many smart city pilot blueprints for implementation and testing possibilities, but difficulties in converting to real impact on cities (that a deployment may have)
- Business case still unclear: new city funding or funding from existing budget categories (do something that is being done better/faster, more efficiently)

Innovation Areas

- Energy efficiency/conservation in buildings: commercial and residential
- Water usage/conservation
- Air quality monitoring
- Transportation: public, multimodal travel, commuting
- Shared transportation: bicycles, electric cars, short-term sharing cars
- Smart parking
- Smart lighting
- Crowd control, including prediction of human (group) behavior
- Security, crime (prevention)

- Emergency (response)
- Maintenance, repair (potholes, water breaks)
- Citizen participation: communicate with government and each other, crowdsourcing of relevant information
- (Business) attractiveness

Research Challenges, Barriers and Drivers

- IoT/ICT infrastructure cost and complexity
- Administrative and organizational "silos" in city administration
- Privacy concerns
- Security
- Digital divide
- Ownership of (existing) infrastructure
- Existing (age of) infrastructure
- Local culture (affinity: sharing, computers)
- Level of education
- Social and economic divide
- Demographics, age distribution
- Legislation
- Communications – mobile/cellular data communication for sensors
- Data ownership

Research Opportunities

Near-term research goals (2 years out):

- Concepts for integrated or federated (evolution of) smart city computing infrastructure/-platform
- Standards: data and meta-data interoperable formats
- Common set of (generic) use cases
- Cross-silo integration at service level for (collaboration) infrastructure
- Big data: storage and analytics, access to multi-domain data sources
- Open standards for IoT sensing
- Simulations / planning studies – demonstrate impact of changes (to be done in advance before engagement, multi-dimensional)

Mid-term research goals (5 years out):

- Demonstration of smart city platform and service-level interoperability
- Standards: data and meta-data interoperable formats understood, tested and (widely?) deployed
- Linkage between ICT and energy/water/transport sensing infrastructure
- Big data: holistic data processing (including social media and consumer) and big-data driven control and operator support
- (Common) reference architecture for end-to-end system (including interoperable data and metadata formats)
- Real-time stream processing vs. batch processing
- Distributed vs. centralized processing
- Legislation and security, privacy

- Energy footprint for IoT and data
- "Smart city emergency/disaster response training center"

Long-term research goals (10 years out):

- Smart-city platform that supports full interoperability and (third party) cross-domain application and service deployment
- Big data based real-time action/control (closed loop, policy driven)
- (Common) reference architecture for end-to-end system with distributed data processing, storage, and analytics

5.11 Infrastructure Operating System, Application Platforms, Stakeholder Interoperation, and Plug & Play Resource Management

Florian Allerding, Florian Michahelles, Milan Milenkovic, Victor del Razo, Joachim Sokol, and Holger Ziekow

Breakout Focus

The group interpreted the scope of this breakout session as an investigation of the benefits of an integrated platform that would support development of cross-domain services and applications. Such an integrated platform would combine Smart Grid and Smart Cities (with smart building and citizenry involved through social networking), to allow global optimizations and benefits. It was group consensus that integrated platforms provide sufficient benefit and are likely to be developed, despite the complexity inherent in such complex multidisciplinary endeavors. A first-order simplifying approximation is to achieve integration at the level of common application (data retrieval) APIs with interoperable data and meta-data formats. The second-level, deeper integration would include a modular implementation based on a common reference architecture, but that is a longer-term research challenge. In this discussion we focused on the data-plane aspect of the platform (data formats, flows, and application APIs). We recognized that a deeply integrated platform would also need to include a control plane definition (configuration, management, security) but decided to defer this discussion, given that data-plane integration is complex enough and it will likely come first.

Motivation

Current and expected advances in ICT have triggered the smartification of basic and commodity services, opening a door of opportunities and presenting us with a number of challenges. Traditional facility management evolves towards smart building management. The increase in renewable energy generation require a more actively managed grid, particularly at the distribution level. Cities begin a transformation towards integrated services. Smart transportation and smart infrastructure concepts emerge as means of providing users with added value.

The need for interaction between services, equipment, data sources etc. becomes a must. The level of complexity of such level of integration can only be solved by the active participation of all stakeholders. An ecosystem that facilitates interaction between these stakeholders and allows for competition and entry to new stakeholders becomes a requirement. ICT paradigms and technologies that would enable integration and service composition, like Service Oriented Architecture (SOA) already exist. The question is which of, and whether, these technologies meet the requirements of the future smart services.

This report aims at identifying the domains where potential smartification would benefit from massive integration. We take a look into the different stakeholders, potential new comers and their requirements and potential interactions. Finally, we discuss challenges and opportunities towards enabling the level of integration that we believe is required.

State-of-the-Art

In current practice, cross-domain interactions are rare and difficult since they require expensive custom interfacing between closed systems. As a first step towards development of an interoperable platform and data formats, our team started by identifying key stakeholders and domains and mapping possible useful interactions between them. We believe that the table and matrix below, albeit incomplete at present, provide a useful framework and structure for reasoning about such interactions. The table below summarizes the domains where we see a potential evolution towards smart* and horizontal integration as well as the identified stakeholders for energy and buildings domain:

Domains:
- Building
- Power and energy
- Cities
- Transportation
- Industrial consumption/production

Stakeholders:
- City management
- Utilities
- Operators
- Industrial customers
- Commercial customers
- Residential customers
- Equipment providers
- Service providers

Most of these domains are characterized by vertically integrated businesses. Resulting in limited interoperability and cooperation. Additionally stakeholders tend to intensively protect their technology, intellectual property and data. Therefore openness to integration is commonly faces resistance. Each stakeholder has different set of requirements in terms of what they need from others and what they are willing to share. These requirements are an important building block for any interaction. Existing sensor and devices involve relatively high integration efforts and its operation and management is not trivial. A number of standardization, open interfaces and common reference architecture efforts have already been started (e.g., open ADR, open SCADA, SGAM).

Challenges and Opportunities

Given the interest of the stakeholders of protecting their data and intellectual property, it is important to understand and define strategies for evaluating what can be shared, the potential benefits, and the methodological and technological alternatives for doing so. We don't believe that such systems will evolve to fully open systems. Rather, we see them interacting on a common interfaces and communication rules. This means that the problem is more of a system composition than a system design one. The group identified the following challenges:

- Interoperable platform, but not necessarily public data: privacy, security, intellectual property subject to business arrangements between stakeholders
- Definition of data and meta-data formats, interoperable across domains
- Definition of basic service-level APIs for data and meta-data retrieval, subject to security and privacy constraints
- Understanding security and privacy risks, level of required functionality
- Design of a common reference architecture
- Separation of control (security, management) and data planes (data and meta-data)
- Strategies and technologies for device and sensor management and configuration
- Plug & play, sensors and apps
- In the particular case of Smart Grid devices, availability of ICT resources and devices despite power grid failure
- Matching legacy and evolution
- Matching and tracking requirements of heterogeneous target groups (also in terms of SLA and KPIs)
- Different expectations by stakeholders in terms of product/technology life-cycle

While the team did not have enough time to characterize research topics into short-, medium- and long-term categories, we do believe that a reasonable proxy for those may be found in the recommendations from the Smart Cities breakout session. They describe cross-domain interactions a more limited set of domains and stakeholders, but can serve as an illustration for a broader scope covered in this section.

Participants

- Florian Allerding
KIT – Karlsruher Institut für Technologie, DE
- Birger Becker
FZI – Forschungszentrum Informatik, DE
- Bert Claessens
Vito – Antwerp, BE
- Hermann de Meer
Universität Passau, DE
- Victor del Razo
TU München, DE
- Christoph Doblander
TU München, DE
- Frank Eliassen
University of Oslo, NO
- Nicolas Gast
INRIA – Grenoble, FR
- Christoph Goebel
TU München, DE
- Manuel Görtz
AGT International – Darmstadt, DE
- Kai Heussen
Technical Univ. of Denmark – Lyngby, DK
- Hanno Hildmann
NEC Laboratories Europe – Heidelberg, DE
- Longbo Huang
Tsinghua University – Beijing, CN
- Hans-Arno Jacobsen
TU München, DE
- Randy H. Katz
University of California – Berkeley, US
- Srinivasan Keshav
University of Waterloo, CA
- Jean-Yves Le Boudec
EPFL – Lausanne, CH
- Florian Michahelles
Siemens Corporation – Berkeley, US
- Milan Milenkovic
Intel – Santa Clara, US
- Johanna Myrzik
TU Dortmund, DE
- Peter Noglik
ABB AG Forschungszentrum Deutschland – Ladenburg, DE
- Mario Paolone
EPFL – Lausanne, CH
- Anthony Papavasiliou
University of Louvain, BE
- Yvonne-Anne Pignolet
ABB Corporate Research – Baden-Dättwil, CH
- Jose Rivera
TU München, DE
- Jiri Rojicek
Honeywell Prague Laboratories, CZ
- Hartmut Schmeck
KIT – Karlsruher Institut für Technologie, DE
- Mischa Schmidt
NEC Laboratories Europe – Heidelberg, DE
- Hans-Peter Schwefel
FTW Forschungszentrum Telekommunikation Wien GmbH, AT
- Joachim Sokol
Siemens AG – München, DE
- Thorsten Staake
Universität Bamberg, DE
- Kai Strunz
TU Berlin, DE
- Verena Tiefenbeck
ETH Zürich, CH
- Anwar Ul Haq
TU München, DE
- Andreas Veit
Cornell University – Ithaca, US
- Holger Ziekow
AGT International – Darmstadt, DE