



DAGSTUHL REPORTS

Volume 6, Issue 1, January 2016

Evolution and Computing (Dagstuhl Seminar 16011) <i>Nick Barton, Per Kristian Lehre, and Nisheeth K. Vishnoi</i>	1
Global Measurements: Practice and Experience (Dagstuhl Seminar 16012) <i>Vaibhav Bajpai, Arthur W. Berger, Philip Eardley, Jörg Ott, and Jürgen Schönwälder</i>	15
Symmetric Cryptography (Dagstuhl Seminar 16021) <i>Frederik Armknecht, Tetsu Iwata, Kaisa Nyberg, and Bart Preneel</i>	34
Geometric and Graph-based Approaches to Collective Motion (Dagstuhl Seminar 16022) <i>Giuseppe F. Italiano, Marc van Kreveld, Bettina Speckmann, and Guy Theraulaz</i> .	55
Well Quasi-Orders in Computer Science (Dagstuhl Seminar 16031) <i>Jean Goubault-Larrecq, Monika Seisenberger, Victor Selivanov, and Andreas Weiermann</i>	69
Privacy and Security in Smart Energy Grids (Dagstuhl Seminar 16032) <i>George Danezis, Stefan Katzenbeisser, Christiane Peters, and Bart Preneel</i>	99
Reproducibility of Data-Oriented Experiments in e-Science (Dagstuhl Seminar 16041) <i>Juliana Freire, Norbert Fuhr, and Andreas Rauber</i>	108
Eyewear Computing – Augmenting the Human with Head-Mounted Wearable Assistants (Dagstuhl Seminar 16042) <i>Andreas Bulling, Ozan Cakmakci, Kai Kunze, and James M. Rehg</i>	160
Modern Cryptography and Security: An Inter-Community Dialogue (Dagstuhl Seminar 16051) <i>Kristin Lauter, Radu Sion, and Nigel P. Smart</i>	207
Dark Silicon: From Embedded to HPC Systems (Dagstuhl Seminar 16052) <i>Hans Michael Gerndt, Michael Glaß, Sri Parameswaran, and Barry L. Rountree</i> .	224

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, 2016

Bibliographic information published by the Deutsche Nationalbibliothek

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

License

This work is licensed under a Creative Commons Attribution 3.0 DE license (CC BY 3.0 DE).



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

- Attribution: The work must be attributed to its authors.

The copyright is retained by the corresponding authors.

Aims and Scope

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

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

- an executive summary of the seminar program and the fundamental results,
- an overview of the talks given during the seminar (summarized as talk abstracts), and
- summaries from working groups (if applicable).

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

Editorial Board

- Gilles Barthe
- Bernd Becker
- Stephan Diehl
- Hans Hagen
- Hannes Hartenstein
- Oliver Kohlbacher
- Stephan Merz
- Bernhard Mitschang
- Bernhard Nebel
- Bernt Schiele
- Nicole Schweikardt
- Raimund Seidel (*Editor-in-Chief*)
- Arjen P. de Vries
- Klaus Wehrle
- Reinhard Wilhelm

Editorial Office

Marc Herbstritt (*Managing Editor*)
Jutka Gasiorowski (*Editorial Assistance*)
Dagmar Glaser (*Editorial Assistance*)
Thomas Schillo (*Technical Assistance*)

Contact

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

Digital Object Identifier: 10.4230/DagRep.6.1.i

Evolution and Computing

Edited by

Nick Barton¹, Per Kristian Lehre², and Nisheeth K. Vishnoi³

1 IST Austria – Klosterneuburg, AT, nick.barton@ist.ac.at

2 University of Nottingham, GB, perkristian.lehre@nottingham.ac.uk

3 EPFL Lausanne, CH, nisheeth.vishnoi@epfl.ch

Abstract

This report documents the talks and discussions at the Dagstuhl seminar 16011 “Evolution and Computing”. The seminar brought together several research disciplines studying evolution, including population genetics and mathematical biology, theoretical computer science, and evolutionary computation.

Seminar January 4–8, 2016 – <http://www.dagstuhl.de/16011>

1998 ACM Subject Classification E.1 Data Structures, F.2 Analysis of Algorithms and Problem Complexity

Keywords and phrases Evolution, Evolutionary Computation, Natural Algorithms, Theory of Computation

Digital Object Identifier 10.4230/DagRep.6.1.1

Edited in cooperation with Jorge Perez Heredia

1 Executive Summary

Nick Barton

Per Kristian Lehre

Nisheeth K. Vishnoi

License  Creative Commons BY 3.0 Unported license
© Nick Barton, Per Kristian Lehre, and Nisheeth K. Vishnoi

Biological evolution has produced an extraordinary diversity of organisms, even the simplest of which is highly adapted, with multiple complex structures. Dynamic structures at even higher levels emerge from collective and social behaviour. These phenomena have traditionally been studied in population genetics, ecology and related disciplines.

However, theoretical computer scientists, endowed with a wide variety of tools, have recently made progress in describing and characterising the computational capabilities of evolution, analyzing natural algorithms, obtaining quantitative bounds for evolutionary models and understanding the role of sex in evolution. The field of evolutionary computation has found that many innovative solutions to optimisation and design problems can be achieved by simulating living processes, such as evolution via random variation and selection, or social behaviour in insects. Researchers in evolutionary computation have recently started applying techniques from theoretical computer science to analyze the optimization time of natural algorithms.

To further the connections and consolidate this burgeoning new discipline, this Dagstuhl seminar brought together participants from the population genetics, mathematical biology, theoretical computer science, and evolutionary computation communities. The seminar opened with a round of introductions, followed by five introductory talks presenting the



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

Evolution and Computing, *Dagstuhl Reports*, Vol. 6, Issue 1, pp. 1–14

Editors: Nick Barton, Per Kristian Lehre, and Nisheeth K. Vishnoi



Dagstuhl Reports

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

2 **16011 – Evolution and Computing**

perspectives of the disciplines attending. Benjamin Doerr introduced runtime analysis of evolutionary algorithms, Paul Valiant discussed evolution from the perspective of learning, Joachim Krug and Nick Barton introduced population genetics, and Nisheeth Vishnoi discussed evolutionary processes from the perspective of theoretical computer science. In addition to talks contributed by participants, there were several breakout sessions on topics identified during the seminar.

The organisers would like to thank the Dagstuhl team and all the participants for making the seminar a success.

2 Table of Contents

Executive Summary

<i>Nick Barton, Per Kristian Lehre, and Nisheeth K. Vishnoi</i>	1
---	---

Overview of Talks

Ancestral selection graph meets lookdown construction <i>Ellen Baake</i>	5
Population genetics and recombination <i>Nick Barton</i>	5
The effect of epistasis on the response to selection <i>Nick Barton</i>	5
Game Dynamics and Population Genetics <i>Erick Chastain</i>	6
Analysis of Evolutionary Algorithms <i>Benjamin Doerr</i>	6
Crossover as Repair Mechanism and the Usefulness Self-Adjusting Parameter Settings: The $(1 + (\lambda, \lambda))$ GA <i>Carola Doerr</i>	6
Eco-evolutionary dynamics Modeling evolution without defining fitness/pay-off <i>Paulien Hogeweg</i>	7
Evolution of mutation rates <i>Kavita Jain</i>	7
Basic concepts of population genetics: Adaptation in rugged fitness landscapes <i>Joachim Krug</i>	8
Genetic mechanisms for the advantage of recombination <i>Joachim Krug</i>	8
Negative Drift in Populations <i>Per Kristian Lehre</i>	8
Mutation as a computational event <i>Adi Livnat</i>	9
The Slime Mold Computer: Physarum can compute shortest paths <i>Kurt Mehlhorn</i>	9
The long-term response to directional selection <i>Tiago Paixao</i>	9
Natural Selection, Game Theory and Genetic Diversity <i>Georgios Piliouras</i>	10
Benefits of Crossover in Combinatorial Search <i>Adam Prugel-Bennett</i>	10
Mixing time of stochastic evolutionary dynamics <i>Piyush Srivastava</i>	11
Slime Mold Dynamics for Flows and Linear Programming <i>Damian Mateusz Straszak</i>	11

4 16011 – Evolution and Computing

The impact of genetic drift on the runtime of simple estimation-of-distribution algorithms <i>Dirk Sudholt and Carsten Witt</i>	11
Understanding Diversity and Recombination in Simple Evolutionary Algorithms <i>Dirk Sudholt</i>	12
The speed of adaptation of complex traits <i>Barbora Trubenova and Jorge Perez Heredia</i>	12
Evolutionary Dynamics <i>Nisheeth K. Vishnoi</i>	12
Coalescent trees, induced subtrees, their topology and site frequency spectrum <i>Thomas Wiehe</i>	13
Participants	14

3 Overview of Talks

3.1 Ancestral selection graph meets lookdown construction

Ellen Baake (*Universität Bielefeld, DE*)

License  Creative Commons BY 3.0 Unported license
© Ellen Baake

Population genetics today relies crucially on mathematical concepts such as ancestral lineages and random genealogies. The talk provided an overview. It started from an interacting particle system that describes the joint action of random reproduction, mutation, and selection; then explained the constructions used to trace back the ancestry of individuals and their genealogy, and presented some recent results. In particular, it presented a novel approach that unifies the two established concepts in the field, namely, the ancestral selection graph of Krone and Neuhauser (1997) and the lookdown construction of Donnelly and Kurtz (1999).

References

- 1 P. Donnelly and T. G. Kurtz. Genealogical processes for Fleming-Viot models with selection and recombination. *Ann. Appl. Probab.*, 9(4):1091–1148, 11 1999.
- 2 S. M. Krone and C. Neuhauser. Ancestral processes with selection. *Theoretical Population Biology*, 51(3):210–237, 1997.
- 3 U. Lenz, S. Kluth, E. Baake, and A. Wakolbinger. Looking down in the ancestral selection graph: A probabilistic approach to the common ancestor type distribution. *Theoretical Population Biology*, 103:27–37, 2015.

3.2 Population genetics and recombination

Nick Barton (*IST Austria – Klosterneuburg, AT*)

License  Creative Commons BY 3.0 Unported license
© Nick Barton

There has been a long-standing interest in evolutionary theory in how sex and recombination are maintained, despite their obvious costs. In an infinite population, one can find a general formula for the selection on a gene that slightly modifies recombination; this is derived assuming that interactions between genes (epistasis) are weak. The selection on recombination is expressed in terms of the effect of recombination in reducing mean fitness, and in increasing the additive variance in fitness. These results should be applicable to analogous problems concerning evolutionary algorithms.

3.3 The effect of epistasis on the response to selection

Nick Barton (*IST Austria – Klosterneuburg, AT*)

License  Creative Commons BY 3.0 Unported license
© Nick Barton

Quantitative genetics describes the evolution of complex traits, which depend on many genes with interacting effects. In 1960, Robertson derived a remarkably simple expression for the

total change in mean of a trait, that can be produced by selecting on a population: this equals the population size times the change in the first generation. Robertson’s result assumed an additive model, but it can be generalised to arbitrary gene interactions (i.e., epistasis). It is an application of the “infinitesimal model”, which approximates the evolution of traits that depend on large numbers of freely recombining genes. I review this and other results that attempt to set general limits on the possible response to selection, and discuss approaches to the same problem from computer science.

3.4 Game Dynamics and Population Genetics

Erick Chastain (Rutgers University – Piscataway, US)

License  Creative Commons BY 3.0 Unported license
© Erick Chastain

This is a review of recent work by C, Livnat, Papadimitriou & Vazirani on the connection between Game Theory, Algorithms, and Evolution. We also mention some interesting open problems and the progress we have made on them (including a partial extension of our results to Diploid organisms), indicating promising directions for those interested.

3.5 Analysis of Evolutionary Algorithms

Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR)

License  Creative Commons BY 3.0 Unported license
© Benjamin Doerr

In this first talk of the Dagstuhl seminar “Evolution and Computation”, I will give an easy introduction to the field of analyses of evolutionary algorithms, aimed at an audience with backgrounds in general algorithms or theoretical biology.

To get a quick start into the topic, I will present a particular, but typical result first, namely how simple evolutionary algorithms optimize pseudo-Boolean linear functions, and show-case how narrow occasionally our understanding is, namely by discussing that comparable results for monotonic functions are a famous open problem.

I will then give a broader introduction to this field, discuss the main research goals, the types of results targeted and the methods typically used. I will finally make some language precise that will help the audience to follow the other talks from this field in this seminar.

3.6 Crossover as Repair Mechanism and the Usefulness Self-Adjusting Parameter Settings: The $(1 + (\lambda, \lambda))$ GA

Carola Doerr (UPMC – Paris, FR)

License  Creative Commons BY 3.0 Unported license
© Carola Doerr

Joint work of Benjamin Doerr, Carola Doerr, Franziska Ebel

We present an evolutionary algorithm using crossover that performs better than any purely mutation-based algorithm on the Hamming distance problem. The main idea behind our

approach is a novel use of crossover as repair mechanism. We also discuss that for this algorithm self-adjusting parameter settings are provably superior to any static parameter choices.

3.7 Eco-evolutionary dynamics Modeling evolution without defining fitness/pay-off

Paulien Hogeweg (Utrecht University, NL)

License  Creative Commons BY 3.0 Unported license
© Paulien Hogeweg

In this talk I addressed the following questions earlier raised in this workshop:

- Can we model evolution without prior fitness (or pay-off) definition?
- How/when do functionally differentiated ecosystems evolve
- Self-adaptation in evolution: evolution of genotype-phenotype mapping

Using the RNA world as an example, I showed that, if we include structure evolution can be modelled without apriori fitness or payoff definition: the inherent structure in the RNA model provides the substrate for the evolution, in which evolution chooses its own fitness. Multi-species communities can evolve, in which the various lineages have different roles in the ecosystem (niche creation). Mutation rate plays a crucial role in the type of community which evolves, i.e. mutation rate has qualitative effect rather than just influencing speed and degree of adaptation. Important in these results is spatial pattern formation: higher order of selection emerges automatically. Moreover we see in this system the evolution of genotype-phenotype mapping in such a way that the mutant cloud (i.e. the quasispecies) is shaped by evolution such that it contains non-fit mutants which nevertheless play a crucial role in the evolved ecosystem, and can lead to fast adaptation to novel circumstances.

Finally I discussed the question: “can we derive general conclusion by studying specific (structured) examples”. My answer is yes we can. Features as discussed reoccur in many very differently structured models, and can be seen as generic properties of evolution *provided enough degrees of freedom are available to the evolutionary process*.

3.8 Evolution of mutation rates

Kavita Jain (JNCASR – Bangalore, IN)

License  Creative Commons BY 3.0 Unported license
© Kavita Jain

Because most mutations are deleterious, the mutation rate can not be too high and an upper bound is provided by error threshold. The mutation rate is not zero either. I will describe our results on the fixation probability of a nonmutator using a branching process and arrive at a lower bound on the mutation rate in a finite population.

3.9 Basic concepts of population genetics: Adaptation in rugged fitness landscapes

Joachim Krug (Universität Köln, DE)

License  Creative Commons BY 3.0 Unported license
© Joachim Krug

The purpose of the lecture was to introduce some basic concepts of evolutionary adaptation that appear in similar form in biological and computational contexts. Starting from the standard Wright-Fisher model of finite populations, the stochastic dynamics of fixation was introduced and used to identify the evolutionary regimes of periodic selection and clonal interference. The main part of the lecture was devoted to genotypic fitness landscapes, their empirical basis, probabilistic modeling, and exploration by random adaptive walks. Finally, the effects of recombination on adaptation in rugged fitness landscapes was briefly addressed.

3.10 Genetic mechanisms for the advantage of recombination

Joachim Krug (Universität Köln, DE)

License  Creative Commons BY 3.0 Unported license
© Joachim Krug
Joint work of Joachim Krug, Su-Chan Park

I will describe two results pertaining to the effect of recombination on the efficacy of evolutionary searches. First, I present a solvable model for the Fisher-Muller effect that predicts a twofold speedup of adaptation in a linear fitness gradient. Second, I discuss a minimal deterministic two-locus model which shows a phase transition as a function of recombination rate. Beyond the transition the model displays bistability in the dynamical systems sense and the escape time from a local fitness peak is infinite.

3.11 Negative Drift in Populations

Per Kristian Lehre (University of Nottingham, GB)

License  Creative Commons BY 3.0 Unported license
© Per Kristian Lehre

The expected running times of evolutionary algorithms are often analysed using so-called drift analysis where the current state of the algorithm is mapped to a real-valued potential. Bounds on the running time are derived from the expected change in the potential per generation. However, finding an appropriate potential function is non-trivial, particularly for population-based evolutionary algorithms.

In this talk, I presented an alternative drift theorem that provides tail bounds on the running time of population-based evolutionary algorithms given simple conditions on the variation operator and selection mechanism employed by the algorithm. I outlined the ideas behind the proof, which uses a combination of multi-type branching processes and classical drift analysis.

References

- 1 Per Kristian Lehre. Negative Drift in Populations. In *Proc. of Parallel Problem Solving from Nature (PPSN XI)*, LNCS, Vol. 6238, pp. 244–253, Springer, 2011.

3.12 Mutation as a computational event

Adi Livnat (University of Haifa, IL)

License  Creative Commons BY 3.0 Unported license
© Adi Livnat

In recent years it has become clear that germline mutation is affected by DNA sequence and structure and by complex biological mechanisms. Can this new empirical knowledge tell us something about how evolution works? I argue that opening up the black box of the nature of mutation affects fundamental concepts in our understanding of evolution, including the role of sex in evolution.

3.13 The Slime Mold Computer: Physarum can compute shortest paths

Kurt Mehlhorn (MPI für Informatik – Saarbrücken, DE)

License  Creative Commons BY 3.0 Unported license
© Kurt Mehlhorn

The slime mold *Physarum* can apparently compute shortest paths. Nakagaki, Yamada, and Tóth (Nature 2000) performed the following experiment: They prepared a maze, covered it with *Physarum*, and provided food at two locations. After a few hours, the slime mold had retracted to the shortest path connecting the two food sources. Tero et al (J. Theoretical Biology, 2007) provided a mathematical model for the dynamics of the slime mold and verified in computer simulations that the model converges to the shortest path. In the first part of the talk, I survey the experiments (a video is available at <http://people.mpi-inf.mpg.de/~mehlhorn/ftp/SlimeAusschnitt.webm>) and introduce the mathematical model. In the second part, I describe the path towards a proof of convergence. In the third part, I look into the future. The slime mold can also build beautiful networks. How can we understand its network building capabilities?

References

- 1 T. Nakagaki, H. Yamada, and A. Tóth. Maze-solving by an amoeboid organism. *Nature*, 407(6803):470, oct 2000.
- 2 A. Tero, R. Kobayashi, and T. Nakagaki. A mathematical model for adaptive transport network in path finding by true slime mold. *Journal of theoretical biology*, 244(4):553–64, feb 2007.

3.14 The long-term response to directional selection

Tiago Paixao (IST Austria – Klosterneuburg, AT)

License  Creative Commons BY 3.0 Unported license
© Tiago Paixao

The role of gene interactions in the response to selection has long been a controversial subject; while some have dismissed them as an important influence on adaptation, others have argued that their long-term effects are of high significance. Here, we derive simple and general predictions for the effect of gene interactions on the long-term response to selection from

standing variation (excluding new mutations). We show that when the dynamics of allele frequencies are dominated by genetic drift, the long-term response is surprisingly simple, depending only on the initial components of the trait variance, regardless of the detailed genetic architecture. Moreover, we show that this result applies when many gene contribute to fitness.

3.15 Natural Selection, Game Theory and Genetic Diversity

Georgios Piliouras (Singapore University of Technology and Design, SG)

License  Creative Commons BY 3.0 Unported license
© Georgios Piliouras

Joint work of Ruta Mehta, Ioannis Panageas, Georgios Piliouras

In a recent series of papers a strong connection has been established between standard models of sexual evolution in mathematical biology and Multiplicative Weights Updates Algorithm, a ubiquitous model of online learning and optimization. These papers show that mathematical models of biological evolution are tantamount to applying discrete replicator dynamics, a close variant of MWUA on (asymmetric) partnership games. We show that in the case of partnership games, under minimal genericity assumptions, discrete replicator dynamics converge to pure Nash equilibria for all but a zero measure of initial conditions. This result holds despite the fact that mixed Nash equilibria can be exponentially (or even uncountably) many, completely dominating in number the set of pure Nash equilibria. Thus, in haploid organisms the long term preservation of genetic diversity needs to be safeguarded by other evolutionary mechanisms, such as mutation and speciation.

3.16 Benefits of Crossover in Combinatorial Search

Adam Prugel-Bennett (University of Southampton, GB)

License  Creative Commons BY 3.0 Unported license
© Adam Prugel-Bennett

Crossover often proves to be a powerful research tool in finding good solutions to combinatorial optimisation problems. The landscape of such problems are very complex so that understanding why crossover is beneficial is difficult. To help elucidate possible mechanisms where population-based search, particularly using crossover, is beneficial we consider three ‘toy’ problems. The first demonstrates how a hybrid algorithm combining crossover with local search can solve a problem by combining building blocks. The second problem looks at how a population can efficiently explore a plateau region, while the third problem looks at how a population can solve a problem despite a very large of noise. In each case these problems require super-polynomial time for local search. The talk describes empirical work which has previously been published in [1]. We also describe a run time analysis carried out by Jonathan Shapiro, Jonathan Rowe and the author. Preliminary results have been published in [2].

References

- 1 A. Prügel-Bennett. Benefits of a population: Five mechanisms that advantage population-based algorithms. *Trans. Evol. Comp.*, 14(4):500–517, Aug. 2010.
- 2 A. Prugel-Bennett, J. Rowe, and J. Shapiro. Run-time analysis of population-based evolutionary algorithm in noisy environments. In *Proc. of the 2015 ACM Conf. on Foundations of Genetic Algorithms XIII (FOGA’15)*, pp. 69–75, New York, NY, USA, 2015. ACM.

3.17 Mixing time of stochastic evolutionary dynamics

Piyush Srivastava (California Institute of Technology – Pasadena, US)

License © Creative Commons BY 3.0 Unported license
© Piyush Srivastava

We study the time it takes for stochastic evolutionary dynamics to achieve a stationary steady state (i.e., its mixing time). We prove that the the mixing time of a wide class of such dynamics grows only logarithmically in the size of the state space. The class of dynamics we study includes as a special case a finite population stochastic version of the quasispecies model of molecular evolution.

Such dynamics, in particular the finite population quasispecies model, have been used to study the evolution of viral populations with applications to drug design strategies countering them. Here the time it takes for the population to reach a steady state is important both for the estimation of the steady-state structure of the population as well in the modeling of the treatment strength and duration. Our result, that such populations exhibit rapid mixing, may be seen as a theoretical justification for numerical simulations that use the above approach.

3.18 Slime Mold Dynamics for Flows and Linear Programming

Damian Mateusz Straszak (EPFL Lausanne, CH)

License © Creative Commons BY 3.0 Unported license
© Damian Mateusz Straszak

Joint work of Damian Mateusz Straszak, Nisheeth K. Vishnoi

We study dynamics inspired by *Physarum polycephalum* (a slime mold) for solving network flow problems and linear programs. These dynamics are arrived at by a local and mechanistic interpretation of the inner workings of the slime mold and a global optimization perspective has been lacking even in the simplest of instances. Our first result is an interpretation of the dynamics as an optimization process. We show that *Physarum* dynamics can be seen as a steepest-descent type algorithm on a certain Riemannian manifold. Moreover, we prove that the trajectories of *Physarum* are in fact paths of optimizers to a parametrized family of convex programs, in which the objective is a linear cost function regularized by an entropy barrier. Subsequently, we rigorously establish several important properties of solution curves of *Physarum*. We prove global existence of such solutions and show that they have limits, being optimal solutions of the underlying problems.

3.19 The impact of genetic drift on the runtime of simple estimation-of-distribution algorithms

Dirk Sudholt (University of Sheffield, GB) and Carsten Witt (Technical University of Denmark – Lyngby, DK)

License © Creative Commons BY 3.0 Unported license
© Dirk Sudholt and Carsten Witt

In this talk, we will consider simple estimation-distribution algorithms, including the so-called compact GA (cGA), on the classical OneMax benchmark problem. The perspective is runtime analysis. We will derive lower bounds on the runtime of the cGA, and discuss how genetic drift affects the optimal parameter setting of the algorithm.

3.20 Understanding Diversity and Recombination in Simple Evolutionary Algorithms

Dirk Sudholt (University of Sheffield, GB)

License  Creative Commons BY 3.0 Unported license
© Dirk Sudholt

I presented an open problem in runtime analysis: trying to understand the benefit of recombination in the context of a simple $(\mu+1)$ Genetic Algorithm on the OneMax problem. Experiments suggest that crossover is beneficial, especially for large populations. Yet despite the seeming simplicity of the setting, existing runtime analyses are restricted to small populations and often ignore the initial diversity in the population. Can we use techniques from Population Genetics or other fields to show that Genetic Algorithms exploit this diversity efficiently through recombination?

3.21 The speed of adaptation of complex traits

Barbora Trubenova (IST Austria – Klosterneuburg, AT) and Jorge Perez Heredia (University of Sheffield, GB)

License  Creative Commons BY 3.0 Unported license
© Barbora Trubenova and Jorge Perez Heredia

Many adaptations are complex; they involve large numbers of genes that may interact in non-additive ways. The efficacy of natural selection to produce these adaptations has been a long-standing question. In particular, how long does it take for natural selection to evolve such complex adaptations?

Here, we address this question by making use of tools from computer science to characterize the time it takes for a population to reach a particular genotype sequence. We focus on how this time scales with the complexity of the trait, and find the conditions on selection strength that enable efficient adaptation. We quantify the ‘cost of complexity’ on several classes of fitness landscapes and show that this cost depends strongly on the details of the genetic architecture. We distinguish between polynomial and exponential scalings as efficient and inefficient adaptation and show that there is sharp threshold between the two for populations in the weak mutation regime. Moreover, we show that even when the loci contributing to the trait interact in an extreme form of epistasis, the time required to reach the fitness peak scales polynomially.

3.22 Evolutionary Dynamics

Nisheeth K. Vishnoi (EPFL Lausanne, CH)

License  Creative Commons BY 3.0 Unported license
© Nisheeth K. Vishnoi

In this talk we survey two fundamental models in evolutionary dynamics: the infinite population Quasispecies model and the stochastic, finite population Wright-Fisher model. Subsequently, we discuss the notions of error threshold, time to convergence and mixing time rigorously.

3.23 Coalescent trees, induced subtrees, their topology and site frequency spectrum

Thomas Wiehe (Universität Köln, DE)

License  Creative Commons BY 3.0 Unported license
© Thomas Wiehe

Genealogies, when viewed forward-in-time, are equivalent to certain types of branching processes, for instance the Yule process. We are interested in measuring topological properties of binary trees which are generated by a Yule process. An easily accessible topological parameter is tree balance at (upper) internal nodes. Tree balance affects the mutation site frequency spectrum (SFS) and can introduce a bias in typical SFS-based statistics. For practical applications it is of interest to understand the dependence of topological properties – and hence of the SFS – in induced subtrees, when conditioned on the topology of a supertree. Furthermore, it is of interest to understand the impact of recombination in changing tree topologies. Last, I will shortly describe a ‘topological’ measure of linkage disequilibrium.

Participants

- Ellen Baake
Universität Bielefeld, DE
- Nick Barton
IST Austria –
Klosterneuburg, AT
- Arnab Bhattacharyya
Indian Institute of Science –
Bangalore, IN
- Erick Chastain
Rutgers Univ. – Piscataway, US
- Duc-Cuong Dang
University of Nottingham, GB
- Harold P. de Vlardar
Parmenides Center for the Study
of Thinking, DE
- Benjamin Doerr
Ecole Polytechnique –
Palaiseau, FR
- Carola Doerr
UPMC – Paris, FR
- Tobias Friedrich
Hasso-Plattner-Institut –
Potsdam, DE
- Paulien Hogeweg
Utrecht University, NL
- Kavita Jain
JNCASR – Bangalore, IN
- Timo Kötzing
Hasso-Plattner-Institut –
Potsdam, DE
- Joachim Krug
Universität Köln, DE
- Per Kristian Lehre
University of Nottingham, GB
- Adi Livnat
University of Haifa, IL
- Kurt Mehlhorn
MPI für Informatik –
Saarbrücken, DE
- Tiago Paixao
IST Austria –
Klosterneuburg, AT
- Ioannis Panageas
Georgia Institute of Technology –
Atlanta, US
- Jorge Perez Heredia
University of Sheffield, GB
- Georgios Piliouras
Singapore University of
Technology and Design, SG
- Adam Prugel-Bennett
University of Southampton, GB
- Jonathan L. Shapiro
University of Manchester, GB
- Piyush Srivastava
California Institute of Technology
– Pasadena, US
- Damian Mateusz Straszak
EPFL Lausanne, CH
- Dirk Sudholt
University of Sheffield, GB
- Andrew M. Sutton
Hasso-Plattner-Institut –
Potsdam, DE
- Barbora Trubenova
IST Austria –
Klosterneuburg, AT
- Paul Valiant
Brown Univ. – Providence, US
- Nisheeth K. Vishnoi
EPFL Lausanne, CH
- Thomas Wiehe
Universität Köln, DE
- Carsten Witt
Technical Univ. of Denmark –
Lyngby, DK
- Xin Yao
University of Birmingham, GB



Global Measurements: Practice and Experience

Edited by

Vaibhav Bajpai¹, Arthur W. Berger², Philip Eardley³, Jörg Ott⁴,
and Jürgen Schönwälder⁵

1 Jacobs University – Bremen, DE, v.bajpai@jacobs-university.de

2 Akamai Technologies – Cambridge, US, awberger@csail.mit.edu

3 BT Research – Ipswich, GB, philip.eardley@bt.com

4 TU München, DE, ott@in.tum.de

5 Jacobs University – Bremen, DE, j.schoenwaelder@jacobs-university.de

Abstract

This article summarises a 2.5 day long Dagstuhl Seminar on “Global Measurements: Practice and Experience” held in January 2016. This seminar was a followup of the seminar on “Global Measurement Frameworks” held in 2013, which focused on the development of global Internet measurement platforms and associated metrics. The second seminar aimed at discussing the practical experience gained with building these global Internet measurement platforms. It brought together people who are actively involved in the design and maintenance of global Internet measurement platforms and who do research on the data delivered by such platforms. Researchers in this seminar have used data derived from global Internet measurement platforms in order to manage networks or services or as input for regulatory decisions. The entire set of presentations delivered during the seminar is made publicly available at <http://materials.dagstuhl.de/index.php?semnr=16012>.

Seminar January 4–7, 2016 – <http://www.dagstuhl.de/16012>

1998 ACM Subject Classification C.2.3 Network Operations, C.4 Performance of Systems

Keywords and phrases Internet measurements, Quality of experience, Traffic engineering

Digital Object Identifier 10.4230/DagRep.6.1.15

1 Executive summary

Vaibhav Bajpai

Arthur W. Berger

Philip Eardley

Jörg Ott

Jürgen Schönwälder

License  Creative Commons BY 3.0 Unported license

© Vaibhav Bajpai, Arthur W. Berger, Philip Eardley, Jörg Ott, Jürgen Schönwälder

Several large-scale Internet measurement platforms have been deployed during the last years in order to understand how the Internet is performing, to observe how it is evolving, and to determine where failures or degradations occur. Examples are the CAIDA Archipelago (Ark) platform [6] (used for Internet topology discovery and detecting congestion on interdomain links), the SamKnows platform [4] (used by regulators and network operators to study network performance), the RIPE Atlas platform [3, 5] (that provides measurement services to network operators and researchers), the Netradar system [8] (for performing wireless performance measurements), and the BISmark project [9]. European collaborative research



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

Global Measurements: Practice and Experience, *Dagstuhl Reports*, Vol. 6, Issue 1, pp. 15–33

Editors: Vaibhav Bajpai, Arthur W. Berger, Philip Eardley, Jörg Ott, and Jürgen Schönwälder



DAGSTUHL
REPORTS

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

projects lately have been working on a Measurement Plane (mPlane) [10] and how to incorporate measurement results into network management systems (e.g., Leone) [2]. Related projects (e.g., Flamingo) [1] are increasingly working with measurement data from these platforms. Large-scale measurements are meanwhile also used to drive network operations or to dynamically adjust how services are delivered to customers. Content Delivery Network (CDN) providers use measurement data to optimize content caches and to tune load balancing algorithms. One key challenge is that global Internet measurement systems can generate large amounts of data that need to be processed to derive relevant information.

This seminar (#16012) was a followup of the Dagstuhl seminar on Global Measurement Frameworks (#13472) [7]. The main focus of the first seminar was an exchange of ideas on the development of global measurement infrastructures, frameworks and associated metrics. Some of this work is now further pursued in standardization bodies [4] such as the IETF Large-Scale Measurement of Broadband Performance (LMAP) working group and the Broadband Forum. The goal of this followup seminar was to focus on the experience obtained with different metrics, tools, and data analysis techniques. It provided a forum for researchers to exchange their experience with different practices to conduct global measurements. The aim was to identify what works well in certain contexts, what has proven problematic in other contexts, and identify open issues that need further research. The seminar approached this by looking at three distinct dimensions: (a) Measurement metrics, (b) data processing technologies and (c) data analysis methodologies. Some key questions were:

1. Which metrics have been found useful for measuring Quality of Experience (QoE) of certain classes of services? Which metrics have been found problematic? Is it possible to find indicators for good metrics and problematic metrics?
2. Which technologies have been found useful for storing and processing large amounts of measurement data? Which technologies were found to be problematic? Are there new promising technologies that may be used in the future? What are the specific requirements for dealing with large-scale measurement data and how do they relate to or differ from other big data applications?
3. Which data analysis techniques have been found to be useful? Which data analysis techniques have been found to be problematic? Are there any novel promising techniques that need further research and development?

Although at the seminar the participants chose to organize the discussions on more general topics than these specific questions, during the discussions most of these questions were addressed to one degree or another.

References

- 1 Flamingo – Management of the Future Internet. <http://www.fp7-flamingo.eu>. [Online; accessed 18-January-2016].
- 2 Leone – From Global Measurements to Local Management. <http://www.leone-project.eu>. [Online; accessed 18-January-2016].
- 3 RIPE Atlas: A Global Internet Measurement Network. *Internet Protocol Journal*, September 2015. URL: <http://ipj.dreamhosters.com/wp-content/uploads/2015/10/ipj18.3.pdf>.
- 4 V. Bajpai and J. Schönwälder. A Survey on Internet Performance Measurement Platforms and Related Standardization Efforts. *Communications Surveys Tutorials, IEEE*, 17(3):1313–1341, thirdquarter 2015. doi:10.1109/COMST.2015.2418435.
- 5 Vaibhav Bajpai, Steffie Jacob Eravuchira, and Jürgen Schönwälder. Lessons Learned From Using the RIPE Atlas Platform for Measurement Research. *SIGCOMM Comput. Commun. Rev.*, 45(3):35–42, July 2015. doi:10.1145/2805789.2805796.

- 6 kc claffy. The 7th Workshop on Active Internet Measurements (AIMS7) Report. *SIGCOMM Comput. Commun. Rev.*, 46(1):50–57, January 2016. doi:10.1145/2875951.2875960.
- 7 Philip Eardley, Marco Mellia, Jörg Ott, Jürgen Schönwälder, and Henning Schulzrinne. Global Measurement Framework (Dagstuhl Seminar 13472). *Dagstuhl Reports*, 2014. doi:
<http://dx.doi.org/10.4230/DagRep.3.11.144>.
- 8 S. Sonntag, J. Manner, and L. Schulte. Netradar – Measuring the wireless world. WiOpt’13, May 2013.
- 9 Srikanth Sundaresan, Sam Burnett, Nick Feamster, and Walter De Donato. BISmark: A Testbed for Deploying Measurements and Applications in Broadband Access Networks. USENIX ATC’14, pages 383–394, 2014. URL: <http://dl.acm.org/citation.cfm?id=2643634.2643673>.
- 10 B. Trammell, P. Casas, D. Rossi, A. Bar, Z. Houidi, I. Leontiadis, T. Szemethy, and M. Mellia. mPlane: an Intelligent Measurement Plane for the Internet. *IEEE Communications Magazine*, 52(5):148–156, May 2014. doi:10.1109/MCOM.2014.6815906.

2 Table of Contents

Executive summary

Vaibhav Bajpai, Arthur W. Berger, Philip Eardley, Jörg Ott, Jürgen Schönwälder . 15

Invited Presentations

Experiences from Measuring Networks (Henning Schulzrinne) 19

Empirical Network Science (Daniel Karrenberg) 20

Global Measurements at Akamai (Arthur Berger) 20

Parallel Group Work

Measurement Platforms Integration 21

Doing it Wrong 22

Ethics 23

Reproducibility and Data Quality 24

Storage, Processing and Archival 25

Future Measurement Challenges 25

Lightning Talks

HostView: Measuring Internet quality of experience on end-hosts (Renata Teixeira) 28

A Path Transparency Observatory (Brian Trammell) 28

Virtual Measurement Accuracy (Al Morton) 29

From Local to Global Measurements (Georg Carle) 29

From Packet Counts to QoE (Markus Fiedler) 29

CheesePi – Swedish home network monitoring (Ian Marsh) 30

WebRTC Service Quality in the Wild (Varun Singh) 30

Haystack – Mobile traffic monitoring in user space (Srikanth Sundaresan) 31

Schengen Routing: A Compliance Analysis (Burkhard Stiller) 31

Conclusions and Next Steps 31

Participants 33

3 Invited Presentations

The invited presentations were intended as a basis for triggering discussions and identifying areas for group work.

3.1 Experiences from Measuring Networks (Henning Schulzrinne)

Henning Schulzrinne (Columbia University / FCC) began by sharing experiences gained through five iterations of the Federal Communications Commission (FCC) Measuring Broadband America (MBA) program [1], consisting of around 5,500 measurement hosts. The project is unique in that it is a collaboration between a regulator, a contractor (SamKnows [2]) developing and managing the infrastructure, about a dozen consumer ISPs and their trade associations, backbone ISPs, two third-party measurement facilities (M-Lab [3] and Level3) and university collaborators. Establishing a code of conduct and setting up a (light-weight) collaborative structure early on has helped work through conflicts and deal with data quality challenges. Since these measurements are used by competing providers, e.g., in TV commercials, the stakes are perceived to be higher than just scientific discovery. The project emphasizes long-term comparability of measurements, open data and reproducibility. For example, all scripts and spreadsheets used to produce the annual report are made available.

He described how the measurement report has changed, increasingly emphasizing variability in performance, across time and the user population, not just averages. One of the more contentious issues has been dealing with unexpected soft and subtle failures of measurement infrastructure, e.g., memory leaks and Ethernet port speed issues, as well as what to consider outliers. For example, a time period was excluded from the measurement month used for reporting since it coincided with the download traffic of iOS 8.0. Users may also delay upgrading their cable modem, causing performance to drop below the offered rate. In the long term, the current model of deploying hardware to end users does not scale well. It is hoped that building in-measurement functionality, e.g., through the IETF LMAP effort [2], rather than bolting it on later, will make measurement cheaper and more fine-grained.

He also emphasized that network diagnostics and network measurements can be highly complementary. For example, the ability to diagnose network problems may motivate end users to install network measurement devices and software. His recent research at Columbia University on measuring performance of YouTube streaming videos [4] finds a close correlation between QoE impairments and the abandonment of YouTube videos.

References

- 1 Measuring Broadband America – Federal Communications Commission. <https://www.fcc.gov/general/measuring-broadband-america>. [Online; accessed 18-January-2016].
- 2 V. Bajpai and J. Schönwälder. A Survey on Internet Performance Measurement Platforms and Related Standardization Efforts. *Communications Surveys Tutorials, IEEE*, 17(3):1313–1341, thirdquarter 2015. doi:10.1109/COMST.2015.2418435.
- 3 Constantine Dovrolis, Krishna Gummadi, Aleksandar Kuzmanovic, and Sascha D. Meinrath. Measurement Lab: Overview and an Invitation to the Research Community. *SIGCOMM Comput. Commun. Rev.*, 40(3):53–56, June 2010. doi:10.1145/1823844.1823853.
- 4 Hyunwoo Nam, Kyung-Hwa Kim, and Henning Schulzrinne. QoE Matters More Than QoS: Why People Stop Watching Cat Videos. In *INFOCOM, 2016 (to appear)*, April 2016.

3.2 Empirical Network Science (Daniel Karrenberg)

Daniel Karrenberg (RIPE NCC) shared experiences with doing empirical network science and derived some principles for good working practices from those experiences. He began by underlining the importance of reproducibility as a necessary condition for producing scientific work. In order to enable reproducibility, it requires one to archive everything during a scientific process. During an experiment, observations must be collected as close to the wire as possible. To avoid any mutation, the raw data derived from these observations must be archived with as little processing as possible. Virtual machines to build any software necessary should be encouraged. A good archive should also include documentation of the experiment such as immutable observations, metadata, experimental conditions, lab notes, calibration data, processed data, changelogs, comments, and analysis / publication backends. For instance, the experimental conditions must not only describe the state of the experiment but also document firmware/software versions to allow proper calibration. Since metadata (such as IP geolocations, IP reverse DNS records, IP prefix to origin AS mappings) is dynamic and volatile, it must also be archived in 'near observation' time. He encouraged the community to invest in storage not only for long term archival, but also to maximise headroom for future measurement results. He illustrated how around 2.1 PB of Hadoop Distributed File System (HDFS) [1] storage is currently (as of January 2016) allocated (with around 400 TB in use) for archiving measurement data produced by the RIPE Atlas project. He reasoned that a well organised and maintained archive not only makes analysis and publication easy, but also enables reuse of observations in the long run. Daniel related that data from some experiments in the RIPE Atlas public archive have indeed been re-used for different purposes. Furthermore providing basic controls to end-users enables unforeseen use of the measurement infrastructure as can be seen from creative usages of the RIPE Atlas measurement platform today.

References

- 1 Tom White. *Hadoop – The Definitive Guide: Storage and Analysis at Internet Scale (4. ed., revised & updated)*. O'Reilly, 2015. URL: <http://www.oreilly.de/catalog/9781491901632/index.html>.

3.3 Global Measurements at Akamai (Arthur Berger)

Based on experiences with the global measurement platform of Akamai Technologies, and relevant to the suggested topics for the seminar, Arthur Berger (Akamai Technologies) discussed three example performance metrics: (a) active measurements of latency and loss, which is used in Request Routing [1] to pick the best datacenter from which to serve a given client, (b) active measurements of latency and loss between Akamai servers, which is used to determine the via nodes in the Akamai routing overlay network [2] and (c) passive measurements of video downloads to clients, which shows the likelihood of abandonment by the end-user as a function of start-up time [3], indexed by the subject category of the video, such as sports, news or religion. He then discussed Akamai's processing of passive measurements recorded in log lines which consists of 1.2 PB data generated per day. He demonstrated where to find publicly available measurement results on Akamai's website [4]. Lastly, he suggested that an area that deserves more attention by the Internet measurement community is security and gave examples of some security measurements collected by Akamai.

References

- 1 Fangfei Chen, Ramesh K. Sitaraman, and Marcelo Torres. End-User Mapping: Next Generation Request Routing for Content Delivery. SIGCOMM'15. ACM, 2015. doi: 10.1145/2785956.2787500.
- 2 Ramesh K. Sitaraman, Mangesh Kasbekar, Woody Lichtenstein, and Manish Jain. *Overlay Networks: An Akamai Perspective*, pages 305–328. John Wiley & Sons, Inc., 2014. doi: 10.1002/9781118909690.ch16.
- 3 S. Shunmuga Krishnan and Ramesh K. Sitaraman. Video Stream Quality Impacts Viewer Behavior: Inferring Causality Using Quasi-experimental Designs. IMC'12, 2012. doi: 10.1145/2398776.2398799.
- 4 State of the Internet – Akamai. <https://www.stateoftheinternet.com>. [Online; accessed 18-January-2016].

4 Parallel Group Work

The afternoon sessions were used to discuss certain topics in more depth in smaller groups. This section summarises the discussions of each group.

4.1 Measurement Platforms Integration

Lately there has been a rise in new and upcoming active measurement platforms [4] on a more or less equivalent underlying substrate (Linux on small cheap boxes). It is unclear whether one can (or should) integrate the common parts of these platforms. There are legitimate reasons to have diversity. For instance, each platform is designed with a distinct goal and provides separate coverage of sources to measure from. However, there is a lot of hard but repetitive work to create a new platform and keep it working. It is also unclear whether all platforms measure the basic measurement primitives in the same way, or whether there are some differences. If we knew they all worked in the same way, then we would be able to compare their results and potentially perform combined studies for a more comprehensive study. Furthermore, if the (common) test code was publicly available, then future developers would not need to expend efforts towards developing yet another version of the same measurement primitive. As such, the idea is to start by building a common codebase / measurement OS distribution for building an integrated measurement platform. The ingredients of this common codebase can include basic measurement utilities and package management tools.

A large number of use cases are covered by a few measurement primitives: (a) loss and latency using `ping`, (b) data-plane topology using `traceroute`, and (c) HTTP GET for applications. It is assumed that these primitives work the same everywhere for comparability reasons. However, there is a need for cross-calibration studies to confirm this premise. A meta-API that glues APIs from multiple measurement platforms together would allow studies to include vantage points from multiple platforms. A design of a Domain Specific Language (DSL) over these primitives implemented by the common substrate would further reduce the barrier to entry. A literature survey is needed to determine how much work such a common platform can save or how it would allow a measurement study to scale up. Measurements also come with a prerequisite for data storage and archival. A volunteer cloud for storage of measurement results with data replication could help spread care and feeding labour and ensure cross-institutional continuity of measurement results.

There are a number of challenges when integrating multiple measurement platforms. For one, reconciling design philosophies (simple vs. complex) is tricky. Seattle [6] is such an integrated measurement platform, albeit designed with a different goal to foster educational cloud computing but with a similar idea. However, the platform turns out to be an overkill for some simple use case scenarios. As such, a requirements description on necessary items for a minimal viable prototype is needed. It is also unclear how to form a community around this goal. It certainly helps to make participants feel good about doing something beneficial for the Internet but having venues to disseminate experience helps bring more people on board.

The management of the system is a first step towards integration. A vanilla OS distribution with stripped down packages with additional cross-compiled packages provided as an overlay similar to the BISmark platform [19] would be ideal. The goal should also be to allow the measurement suite to run inside a virtual machine. Virtual environments help keep dependency issues to a minimum. Given probes are remotely managed, another challenge is to avoid an update that renders them permanently unusable. BISmark uses a manual firmware update process with a possibility to fall-back to a trusted image to help mitigate this risk. Access control is another issue. It is unclear how much control the probe host must receive for hosting the probe.

The second step is to identify the measurement primitives that must be supported. Some candidate primitives may include: `dig`, `ping`, `curl`, `iperf` and a constant-bitrate packet generation tool. Including multiple variations of one primitive (such as `traceroute`) that are designed with slightly different goals (such as `scamper` [14] and `tracebox` [8]) adds value. The possibility of hosting multiple versions of each primitives must be supported. The primitives themselves must also support a common machine-readable output format. The ability of the primitives to write to a database (such as `sqlite`) increases the possibility of reuse since the results can simply be queried. It is a challenge to provide an exhaustive list of primitives that satisfies all measurement studies. As such requirements gathering and a survey to scope the problem is needed. For instance, `tcpdump` may be useful, but it has privacy implications and shipping data produced by this primitive may be a sensitive issue. A survey to identify incremental benefits of each measurement primitive is needed. It is also unclear if exceptions must be made for certain primitives to run in root privilege mode.

Furthermore, a number of future challenges were identified. For instance, a bootstrapping mechanism to get clients registered, an authentication mechanism to identify the clients, a server mechanism to be a destination for primitives, a communication channel to describe this client and server communication, an API to interface with measurement data, encryption and handling of key distribution are few identified areas that require work.

4.2 Doing it Wrong

A basic understanding of the strengths and weaknesses of a measurement method is useful since some weaknesses may inhibit interpretation of certain data and may lead to wrong conclusions. As such it is better to enumerate all the ways to collect the data one needs to answer a research question and then document their pros and cons. Continuous validation [13] is also important to produce data reliably, in particular if there is a dependency on 3rd party components (that may change in unanticipated ways). The key is to ask the question: Why do I have outliers or unexpected results? In order to be able to answer this question it is essential to have a world model [18] and some expectations of the data. At the same time one

also needs to be prepared for the world model to be wrong. Unexpected data requires careful analysis in order to determine whether there is a measurement error, a data analysis error, or a world model error. It is vital to be extensive in the description of the metadata [17] and the documentation of the experiment. As such, it is best to try to gather as much metadata (or context) as possible of the data, but at the same time also being honest about the limits of the data. One also needs to think in multiple timescales since time itself is a complicated thing to get right. Dealing with time can be notoriously hard due to accuracy and precision issues, clock drift issues, synchronization issues, issues caused by non-monotonic clocks and issues with time interpretation (such as ordering, timezone knowledge). Keeping raw data is important and so is the ability to reproduce the analysis.

4.3 Ethics

There can be a tension between scientific principles (measurements and meta-data should be public) and consistency with ethical principles. For instance, we have recently witnessed controversial papers [5, 10] that, although published, raised ethical concerns within the program committee.

There has already been activity by the community on ethical practice. For instance, a dedicated SIGCOMM workshop on Ethics in Networked Systems Research [1] was recently organized in 2015. Moreover, the call for papers for Internet Measurement Conference (IMC) encourages authors when appropriate to include a subsection describing ethical considerations and provides appropriate links for further information on ethical principles [11] and guidance [2] on ethical data sharing. As a followup to the Dagstuhl seminar on Ethics in Data Sharing [7, 9], SURFnet is preparing a document [21] on Data Sharing policy. The final policy will most likely come into effect in the first quarter of 2016.

The issue of what can be considered ethical is often a grey area since opinions can dramatically vary by different parties. For example, a study [3] that analyses causes of collateral damage of censorship by identifying DNS injection activities of the Great Firewall of China could potentially be viewed as unethical by the government of the People's Republic of China. Several intriguing questions from the ethical standpoint deserve discussion. For instance, in a measurement study of cyber crime, is it appropriate to buy products from criminals? and is it appropriate to crawl a website to obtain all of the information even when the site explicitly states that one should not do this?

Ethical issues pertain to more than just privacy infringement. For instance, disrupting the service of an end-user and possibly even endangering an end-user without the user's consent. There is a fine line between legal and ethical issues. Moreover, ethical issues encompass the entire measurement chain starting from the design of an experiment, conducting measurements, data storage, data processing, and data sharing. The security research community is increasingly sensitive to this issue. For instance, some IT departments avoid collecting data just so that they have no data if asked by a law-enforcement agency. By nature, research tends to push the boundaries, however risk analysis can be hard. If it is known in advance how the data will be used, collect just what is needed.

The Internet measurement community needs to publish further guidelines on ethical practice. A key target audience is researchers that are not aware of the issues, but would want to do the right thing. For the Internet measurement community, continued discussion to gain more clarity in the aforementioned grey areas is needed. Regardless of whether there is consensus in the community as a whole, a conference program committee should have the

discretion to reject a paper on ethical grounds. The authors of the rejected paper could be asked for permission to make known the aspect of the work that was considered unethical, so as to provide guidance to the wider community. Furthermore, since a paper rejection occurs after the unethical practice has already occurred, the goal must be to avoid the unethical practice to happen in the first place. To address this part, an interesting question is how to have curricula embrace ethics educations. An ethics background is not just only needed for measurement studies, but in general for people working in computer science, both in academia and industry.

4.4 Reproducibility and Data Quality

Repeatability and reproducibility are often misinterpreted in practice. Repeatability is the notion of re-running the same experiment with a change in time. Reproducibility on the other hand is being able to derive the same conclusions with a change in both space and time. Reproducibility imbibes the flexibility of using different measurement methods to arrive at the same conclusion. In order to foster reproducibility, a number of aspects need to be documented: (a) measurement method, (b) metric, (c) vantage points and (d) implementation. This requires a characterisation plan of statistical tests to imply significance of data analysis.

There are a number of difficulties in reproducing an experiment. For one, statistical analysis is hard. There is a danger of confirmation bias with a tendency to abandon experiments if results are boring. It is often not possible to measure the metric directly (or only as a one-off calibration) since generally there is a lack of stable ground truth. Moreover, it is difficult to publish a study that reproduces an experiment. Worse, documenting the limitations of an experiment is often (wrongly) seen as a weakness. Particularly, sharing datasets of an experimental study with others is hard. For one, legal rules vary by jurisdiction but more so one wants to have a first mover advantage with the associated data collection activity.

There is a need to add rigour in statistical analysis to enable reproducibility. This starts with hypothesis testing and concrete research questions. Factor analysis during the experimental design to cover the design space of variables is often ignored. Outliers that fail hypothesis need treatment. As such, a prospective journal that invites reproducibility would help reduce probability of early abandonment of experiments that confirm previous results. Calibration and quality checks must be encouraged. Conferences can be encouraged to dedicate special sessions devoted to papers that reproduce results. Researchers on the other hand must be encouraged to write a technical report that describes the dataset used in a publication. Such a report must document the measurement method, dataset fields, limitations and scope of the dataset. There are QoE standards that provide test conditions and advice on where the measurement method is applicable. In cases where raw dataset cannot be shared for some reason, researchers must still be encouraged to share the dataset in at least some restricted form by either removing some data columns, obfuscating some fields, or by allowing limited access to the dataset using SQL queries.

4.5 Storage, Processing and Archival

There is currently lack of a best current practice guide on how to store, process and archive measurement data. It seems that academics generally tend to rely on a Network-attached Storage (NAS) coupled with a few highly performant data crunching machines for data analysis operations.

One clear recommendation is to transition away from NAS because they cannot provide local computing power. Apache Hadoop [23] is a better alternative since it provides a tight coupling of storage and compute power, scales gradually over time and in the process turns out to be cost effective. Cloudera Distribution Including Apache Hadoop (CDH) packages [23] provide a simple head start into the Hadoop ecosystem. They can be used to deploy the Hadoop cluster and packages provide tools to make management easy. A transition to Hadoop can be done in multiple iterations. A first step is to get it functional by storing already existing measurement data in CSV or JSON format. However, in the long run a good serialization format such as Apache Avro [23] (for row-oriented datasets) or Apache Parquet [23] (for columnar storage) can help future proof storage in a structured format. Naturally, it is better to make a choice at the very outset of data collection. HBase [23], a non-relational database that can run on top of HDFS, can be used for high performance analysis for specialised applications, although it tends to have a steeper learning curve especially for SQL users. Cloudera Impala adds an SQL engine on top of HDFS. GraphQL can be used to decouple presentation from querying on the data. Message queues such as RabbitMQ can be used for stream-based processing requirements. A recently developed large-scale active measurement platform for DNS [22] uses Hadoop for storage and analysis of data. Experiences with this platform show that there is potential for using Hadoop for Internet measurements.

Certainly the Hadoop tool chain is not the answer to all problems. It must be viewed as HDFS for storage with optional powerful processing on top. Although at times, a powerful machine with lots of computing cores can also serve the same task at hand. As such, at the end the size of the data matters. Hadoop distributes I/O and processing and thus it can crunch large volumes of data in short time. A single fast machine has I/O limits and may have CPU / memory limits that are difficult to scale (but it is often the I/O limit that is difficult or expensive to change).

4.6 Future Measurement Challenges

Measurement methods will evolve beyond traditional active and passive techniques. Al Morton in [15] describes hybrid measurement methods which are subset of both active and passive methods. For instance, Type I hybrid measurements employ methods that augment or modify the stream of interest, while Type II hybrid measurements employ methods that utilize two or more streams of interest with some degree of mutual coordination to collect multiple metrics.

A number of Internet measurement tools are designed with inherent assumptions (about layer-2 networks) [20] that are not true for underlying wireless links. In particular WiFi home networks and cellular networks are impacted by bitrates, retransmission rates, and signal strengths as wireless channel conditions change. As such, we need to design measurement approaches and tools that are also suitable for measuring wireless links.

There are also challenges with metrics that measure available bandwidth. In the view of mostly elastic traffic, partly in combination with wireless links, it is not clear whether a

convincing solution can be expected. Moreover, with existing tools, probing for capacity does not work well with tools that assume that the link is work-conserving.

For many web-based applications, end-to-end traffic is split [16] into a transport session from end system to the front-end servers, and another transport session to the backend infrastructure. In the transport session to the front-end servers, many short-term TCP flows may be observed (in contrast to long-lived TCP flows in the transport session to the backend infrastructure). In such a scenario, protocols used to establish the transport session to the front-end servers can be changed quickly. For instance, Quick UDP Internet Connections (QUIC) [12], which is increasingly used to establish a transport session to the front-end servers, may behave more aggressively than TCP.

There is an increasing demand for low-latency communication. Many technological advances reduce latency significantly. For instance, compared with 4G, 5G claims that it will reduce latency by a factor of 100. However, it is unclear how one can measure latency in these new environments with the required level of accuracy. The security aspects of Internet of Things (IoT) devices are becoming critical. It is unclear whether there is a need for specialized measurement tools and methods in this space. An analysis of measurement challenges with respect to IoT security is needed.

A large number of network functions are being virtualised today. It remains unclear how to measure in such virtualised scenarios. Additional measurement objectives and metrics need to be identified particularly due to the resource sharing effects of such virtual network functions.

References

- 1 Workshop on Ethics in Networked Systems Research – ACM SIGCOMM 2015. <http://conferences.sigcomm.org/sigcomm/2015/netethics.php>. [Online; accessed 18-January-2016].
- 2 Mark Allman and Vern Paxson. Issues and Etiquette Concerning Use of Shared Measurement Data. IMC’07, pages 135–140. ACM, 2007. doi:10.1145/1298306.1298327.
- 3 Anonymous. The Collateral Damage of Internet Censorship by DNS Injection. *SIGCOMM Comput. Commun. Rev.*, 42(3):21–27, June 2012. doi:10.1145/2317307.2317311.
- 4 V. Bajpai and J. Schönwälder. A Survey on Internet Performance Measurement Platforms and Related Standardization Efforts. *Communications Surveys Tutorials, IEEE*, 17(3):1313–1341, thirdquarter 2015. doi:10.1109/COMST.2015.2418435.
- 5 Sam Burnett and Nick Feamster. Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests. SIGCOMM’15. ACM, 2015. doi:10.1145/2785956.2787485.
- 6 Justin Cappos, Ivan Beschastnikh, Arvind Krishnamurthy, and Tom Anderson. Seattle: A Platform for Educational Cloud Computing. SIGCSE’09. ACM, 2009. doi:10.1145/1508865.1508905.
- 7 Julie E. Cohen, Sven Dietrich, Aiko Pras, Lenore D. Zuck, and Hildebrand Mireille. Ethics in Data Sharing (Dagstuhl Seminar 14052). *Dagstuhl Reports*, 4(1):170–183, 2014. doi: <http://dx.doi.org/10.4230/DagRep.4.1.170>.
- 8 Gregory Detal, Benjamin Hesmans, Olivier Bonaventure, Yves Vanaubel, and Benoit Donnet. Revealing Middlebox Interference with Tracebox. IMC’13. ACM, 2013. doi:10.1145/2504730.2504757.
- 9 Sven Dietrich, Jeroen van der Ham, Aiko Pras, Roland van Rijswijk-Deij, Darren Shou, Anna Sperotto, Aimee van Wynsberghe, and Lenore Zuck. Ethics in Data Sharing: developing a model for best practice. CREDs II, 2014.
- 10 Marcel Dischinger, Andreas Haeberlen, Krishna P. Gummadi, and Stefan Saroiu. Characterizing Residential Broadband Networks. IMC’07. ACM, 2007. doi:10.1145/1298306.1298313.

- 11 D. Dittrich and E. Kenneally. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. Technical report, U.S. Department of Homeland Security, 2012.
- 12 Ryan Hamilton, Jana Iyengar, Ian Swett, and Alyssa Wilk. QUIC: A UDP-Based Secure and Reliable Transport for HTTP/2. Internet-Draft draft-tsvwg-quic-protocol-02, Internet Engineering Task Force, January 2016. URL: <http://tools.ietf.org/html/draft-tsvwg-quic-protocol-02>.
- 13 Balachander Krishnamurthy, Walter Willinger, Phillipa Gill, and Martin Arlitt. A Socratic method for validation of measurement-based networking research. *Computer Communications*, 34(1):43–53, 2011. doi:<http://dx.doi.org/10.1016/j.comcom.2010.09.014>.
- 14 Matthew Luckie. Scamper: A Scalable and Extensible Packet Prober for Active Measurement of the Internet. IMC'10. ACM, 2010. doi:10.1145/1879141.1879171.
- 15 Al Morton. Active and Passive Metrics and Methods (and everything in-between, or Hybrid). Internet-Draft draft-ietf-ippm-active-passive-06, Internet Engineering Task Force, January 2016. URL: <http://tools.ietf.org/html/draft-ietf-ippm-active-passive-06>.
- 16 Abhinav Pathak, Y. Angela Wang, Cheng Huang, Albert Greenberg, Y. Charlie Hu, Randy Kern, Jin Li, and Keith W. Ross. Measuring and Evaluating TCP Splitting for Cloud Services. PAM'10, pages 41–50. Springer-Verlag, 2010. URL: <http://dl.acm.org/citation.cfm?id=1889324.1889329>.
- 17 Vern Paxson. Strategies for Sound Internet Measurement. IMC'04, pages 263–271. ACM, 2004. doi:10.1145/1028788.1028824.
- 18 M. Roughan, W. Willinger, O. Maennel, D. Perouli, and R. Bush. 10 Lessons from 10 Years of Measuring and Modeling the Internet's Autonomous Systems. *JSAC*, 29(9):1810–1821, October 2011. doi:10.1109/JSAC.2011.111006.
- 19 Srikanth Sundaresan, Sam Burnett, Nick Feamster, and Walter De Donato. BISmark: A Testbed for Deploying Measurements and Applications in Broadband Access Networks. USENIX ATC'14, pages 383–394, 2014. URL: <http://dl.acm.org/citation.cfm?id=2643634.2643673>.
- 20 Srikanth Sundaresan, Nick Feamster, and Renata Teixeira. Measuring the Performance of User Traffic in Home Wireless Networks. PAM'15, pages 305–317, 2015. doi:10.1007/978-3-319-15509-8_23.
- 21 Roland van Rijswijk-Deij. Ethics in Data Sharing: a best practice for NRENs. TNC'15. GÉANT, 2015. URL: <https://tnc15.terena.org/getfile/1869>.
- 22 Roland van Rijswijk-Deij, Mattijs Jonker, Anna Sperotto, and Aiko Pras. The Internet of Names: A DNS Big Dataset. SIGCOMM'15, pages 91–92. ACM, 2015. doi:10.1145/2785956.2789996.
- 23 Tom White. *Hadoop – The Definitive Guide: Storage and Analysis at Internet Scale (4. ed., revised & updated)*. O'Reilly, 2015. URL: <http://www.oreilly.de/catalog/9781491901632/index.html>.

5 Lightning Talks

Participants were also encouraged to volunteer for a lightning talk to provide a perspective into their recent measurement research work.

5.1 HostView: Measuring Internet quality of experience on end-hosts (Renata Teixeira)

There is interest in automated performance diagnosis on user laptops or desktops. One interesting aspect that has received little attention is the user perspective on performance. To conduct research on both end-host performance diagnosis and user perception of network and application performance, Renata Teixeira (INRIA) presented an end-host data collection tool, called HostView. HostView [1] not only collects network, application and machine level data, but also gathers feedback directly from users. User feedback is obtained via two mechanisms: a system-triggered questionnaire and a user-triggered feedback form. In her talk, she described experiences with the first deployment of HostView. Using data from 40 users, she articulated challenges in this line of research, and reported initial findings in correlating user data to system-level data. She then described more recent efforts in conducting an in-depth study with 12 users in France to guide the design of the next version of HostView and of methods to infer user context and activities.

References

- 1 Ahlem Reggani, Fabian Schneider, and Renata Teixeira. An End-Host View on Local Traffic at Home and Work. PAM'12, pages 21–31, 2012. doi:10.1007/978-3-642-28537-0_3.

5.2 A Path Transparency Observatory (Brian Trammell)

The growing deployment of middle boxes in the Internet has reduced the degree to which the Internet is still an end-to-end network in accordance with its original design. This lack of end-to-endness leads to ossification of the transport layer [1]: new protocols are difficult or impossible to deploy as they must be designed around middle boxes, either those which have been observed, or conjectured to exist. It is necessary to guide protocol engineering for transport protocol innovation on a basis of observations of the Internet as it is, but these observations are hard to come by. Brian Trammell (ETH Zürich) proposed a Path Transparency Observatory, which can take observations of path transparency (the likelihood a packet stream that arrives at the end of the path is the one that was sent, with certain properties) and impairment (something that keeps a path from being transparent for a certain kind of traffic) from multiple sources, with multiple resolutions of condition definition and information about the endpoints and path involved. An observatory collects single observations of a path and a condition on that path at some point in time, with references to the code that created the observations so they can be repeated, and a set of equivalence functions so that equivalent conditions and paths can be compared. He explained that this work is ongoing, and a public observatory will become available within the scope of the Measurement and Architecture for a Middleboxed Internet (MAMI) project [2] over next two and a half years.

References

- 1 Michio Honda, Yoshifumi Nishida, Costin Raiciu, Adam Greenhalgh, Mark Handley, and Hideyuki Tokuda. Is It Still Possible to Extend TCP? IMC'11. ACM, 2011. doi:10.1145/2068816.2068834.
- 2 MAMI – Measurement and Architecture for a Middleboxed Internet. <https://mami-project.eu>. [Online; accessed 18-January-2016].

5.3 Virtual Measurement Accuracy (Al Morton)

The movement towards Network Function Virtualization (NFV) means that measurement system virtualisation will take place for active, passive, and hybrid methods of measurement. This evolution will allow on-demand deployment of measurement systems in general purpose servers. The designs must be cost-effective, but there is tension between cost of physical resources and accuracy. Al Morton (AT&T) presented this trade-off and associated challenges.

5.4 From Local to Global Measurements (Georg Carle)

Georg Carle (TU München) provided a summary of measurement based research work conducted within his group. He explained that understanding Internet phenomena requires both local and global measurements. Local measurements such as on the MEMPHIS test bed allows for reproducible experiments. As part of this project, the MoonGen Traffic generator [1] is an example that allows for high precision by directly accessing hardware features such as precise time stamping from the application space, while bypassing the operating system. Furthermore, Georg reasoned how Software-defined Network (SDN) mechanisms can be used for performing very high-speed flow monitoring using Commercial Off-the-shelf (COTS) components and adaptive load-balancing. One objective of security-related global measurements is to identify prefix hijacking. An innovative approach [2] to identify benign anomalies is to use information with business relations and ownership information from a publicly accessible Internet Routing Registry (IRR) to combine it with collected TLS certificates, and using these certificates as fixed points to be checked in time intervals in which routing anomalies are observed. For performing measurements with wireless links, he presented the MearDroid Android app, which allows to perform wireless measurements from many vantage points.

References

- 1 Paul Emmerich, Sebastian Gallenmüller, Daniel Raumer, Florian Wohlfart, and Georg Carle. MoonGen: A Scriptable High-Speed Packet Generator. IMC'15. ACM, 2015. doi: 10.1145/2815675.2815692.
- 2 Johann Schlamp, Ralph Holz, Oliver Gasser, Andreas Korsten, Quentin Jacquemart, Georg Carle, and Ernst W. Biersack. Investigating the Nature of Routing Anomalies: Closing in on Subprefix Hijacking Attacks. TMA'15. doi:10.1007/978-3-319-17172-2_12.

5.5 From Packet Counts to QoE (Markus Fiedler)

Markus Fiedler (BTH) discussed challenges in measuring QoE. He stressed that the main challenge for the interpretation of network measurements in light of QoE is that user perception happens far up the network stack, far away from where Quality of Service (QoS) problems (such as latency and packet loss) arise and where monitoring takes place. QoS may be transformed significantly throughout the stack. The recently proposed QoE Hourglass Model [1] is one way to formalize such transformations, capturing impacts of transport protocols, display devices and other factors. Using an example from a project with a major European telecommunications provider, he proposed a method that enables the exploitation of information from packet counters for QoE assessment. It starts with the definition of the user-perceived problem to be attacked, followed by the determination of parameters that

reflect those problems far down in the network stack, and of the critical timescale for the user, and finally the use of appropriate comparative summary statistics.

References

- 1 Tahir Nawaz Minhas and Markus Fiedler. Quality of experience hourglass model. In *ComManTel 2013*, Jan 2013. doi:10.1109/ComManTel.2013.6482371.

5.6 CheesePi – Swedish home network monitoring (Ian Marsh)

Ian Marsh (SICS) described the architecture of a distributed measurement system, CheesePi. He utilized the IETF LMAP framework [1] terminology to describe this system. CheesePi uses Raspberry Pi hardware devices to allow always-on, simple and reliable monitoring of users' home Internet connections. By running CheesePi on a Raspberry Pi (termed a Measurement Agent (MA)) connected to their home network, a non-expert user can continuously monitor their connection quality. He argued that a common hardware platform for all MAs gives greater consistency between the collected measurements and it also simplifies the codebase. The result is a common software platform for measurement tasks that can host, execute and record arbitrary network behaviour. Ian explained the reason for deploying dedicated monitoring devices. The project is tailored towards capturing the network connectivity that devices are able to achieve. This can significantly depend on the last hop technology (e.g. Ethernet or WiFi), which would be missed by passive monitoring of user traffic at the home gateway. This work is performed in collaboration with the Swedish regulator Post and Telecom Authority (PTS), who are particularly concerned with expanding connection performance metrics from naive throughput measurements of a particular location and time to something more instructive. He argued that an easily comparable and widely understood metric (e.g., download/upload rates) does not necessarily indicate the QoE of a user.

References

- 1 P. Eardley, A. Morton, M. Bagnulo, T. Burbridge, P. Aitken, and A. Akhter. A Framework for Large-Scale Measurement of Broadband Performance (LMAP). RFC 7594 (Informational), September 2015. URL: <http://www.ietf.org/rfc/rfc7594.txt>.

5.7 WebRTC Service Quality in the Wild (Varun Singh)

Varun Singh (Aalto University) introduced callstats.io, a Web Real-Time Communication (WebRTC) analytics and diagnostics service. It measures service- and conference-level metrics for a WebRTC application service. At the service-level, annoyances (such as, how often do conferences fail, what are the reasons for failure and what is the typical network latency?) are measured. Varun described how callstats.io will share the aggregate quality metrics measured across tens of WebRTC services (big and small, local and global) with the measurement community at large.

5.8 Haystack – Mobile traffic monitoring in user space (Srikanth Sundaresan)

Despite our growing reliance on mobile phones for a wide range of daily tasks, we remain largely in the dark about the operation and performance of our devices, including how (or whether) they protect the information we entrust to them, and with whom they share it. The absence of easy, device-local access to the traffic of our mobile phones presents a fundamental impediment to improving this state of affairs. To develop detailed visibility, Srikanth Sundaresan (ICSI) presented Haystack [1], a system for unobtrusive and comprehensive monitoring of network communications on mobile phones, entirely from user-space. Haystack correlates disparate contextual information such as app identifiers and radio state with specific traffic flows destined to remote services, even if encrypted. Haystack facilitates user-friendly, large-scale deployment of mobile traffic measurements and services to illuminate mobile app performance, privacy and security. Srikanth described the design of Haystack and demonstrated its feasibility with an implementation that provides 26-55 Mbps throughput with less than 5% CPU overhead. He stressed that the system and results highlight the potential for client-side traffic analysis to help understand the mobile ecosystem at scale.

References

- 1 Abbas Razaghpanah, Narseo Vallina-Rodriguez, Srikanth Sundaresan, Christian Kreibich, Phillipa Gill, Mark Allman, and Vern Paxson. Haystack: In Situ Mobile Traffic Analysis in User Space. *CoRR*, abs/1510.01419, 2015. URL: <http://arxiv.org/abs/1510.01419>.

5.9 Schengen Routing: A Compliance Analysis (Burkhard Stiller)

Burkhard Stiller (UZH) described Schengen routing as a strategy to keep traffic originating from sources located in the Schengen area (an area comprising of 26 European countries that have abolished passport and any other type of border control at their common borders) and targeted to destinations located in the Schengen area within the Schengen area. He summarised results of a larger-scale measurement effort [1] performed to quantify Schengen routing compliance in parts of today's Internet. Based on a few thousand TCP, UDP, and ICMP `traceroute` measurements executed from RIPE Atlas probes located in over 1100 different Autonomous Systems (AS) in the Schengen area, it was observed that 34.5% to 39.7% of these routes are Schengen-compliant, while compliance levels vary from 0% to 80% among countries.

References

- 1 Daniel Dönni, Guilherme Sperb Machado, Christos Tsiaras, and Burkhard Stiller. Schengen Routing: A Compliance Analysis. In *AIMS 2015, Ghent, Belgium, June 22-25, 2015.*, 2015. doi:10.1007/978-3-319-20034-7_11.

6 Conclusions and Next Steps

Participants with a mix of senior and junior researchers hailing from both academia and industry encouraged fruitful dialogue. A number of future research agendas were recognized. Brian Trammell volunteered to initiate further discussion on the seminar mailing list towards measurement platform integration. An action item to create a code repository to hold basic primitives that can output results in a machine readable manner was created. Furthermore,

discussion on an Internet measurement cloud for not only storing measurement results but also facilitate its reliable distribution will begin. The organizing team also received valuable feedback. An interest to identify a specific problem to try to tackle it during a prospective future seminar was identified.

Acknowledgements

The organisers would like to thank the participants for their contributions: Special thanks to Olivier Bonaventure, Henning Schulzrinne, Ian Robin Marsh, and Roland van Rijswijk-Deij for reviewing the manuscript. This work was funded by Flamingo, a Network of Excellence project (ICT-318488) supported by the European Commission under its Seventh Framework Programme.

Participants

- Vaibhav Bajpai
Jacobs University – Bremen, DE
- Arthur W. Berger
Akamai Technologies –
Cambridge, US
- Georg Carle
TU München, DE
- Renata Cruz Teixeira
INRIA – Le Chesnay, FR
- Philip Eardley
BT Research – Ipswich, GB
- Markus Fiedler
Blekinge Institute of Technology –
Karlskrona, SE
- Phillipa Gill
Stony Brook University, US
- Oliver Hohlfeld
RWTH Aachen, DE
- Steffie Jacob Eravuchira
SamKnows Ltd. – London, GB
- Daniel Karrenberg
RIPE NCC – Amsterdam, NL
- Mirja Kühlewind
ETH Zürich, CH
- Andri Lareida
Universität Zürich, CH
- Jukka Manner
Aalto University, FI
- Ian Robin Marsh
Swedish Institute of Computer
Science – Kista, SE
- Al Morton
AT&T – Middletown, US
- Jörg Ott
TU München, DE
- Colin Perkins
University of Glasgow, GB
- Philipp Richter
TU Berlin, DE
- Jair Santanna
University of Twente, NL
- Jürgen Schönwälder
Jacobs University – Bremen, DE
- Henning Schulzrinne
Columbia Univ. – New York, US
- Varun Singh
Callstats.IO, FI
- Burkhard Stiller
Universität Zürich, CH
- Srikanth Sundaresan
ICSI – Berkeley, US
- Brian Trammell
ETH Zürich, CH
- Roland van Rijswijk-Deij
University of Twente, NL



Symmetric Cryptography

Edited by

Frederik Armknecht¹, Tetsu Iwata², Kaisa Nyberg³, and
Bart Preneel⁴

1 Universität Mannheim, DE, armknecht@uni-mannheim.de

2 Nagoya University, JP, iwata@cse.nagoya-u.ac.jp

3 Aalto University, FI, kaisa.nyberg@aalto.fi

4 KU Leuven, BE, bart.preneel@esat.kuleuven.be

Abstract

From January 10–15, 2016, the seminar 16021 in Symmetric Cryptography was held in Schloss Dagstuhl – Leibniz Center for Informatics. It was the fifth in the series of the Dagstuhl seminars “Symmetric Cryptography” held in 2007, 2009, 2012, and 2014.

During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations were given during the seminar. The first section describes the seminar topics and goals in general.

Seminar January 10–15, 2016 – <http://www.dagstuhl.de/16021>

1998 ACM Subject Classification E.3 Data Encryption, H.2.0 General – Security, Integrity, and Protection, K.6.5 Security and Protection

Keywords and phrases authenticity, block ciphers, confidentiality, cryptanalysis, hash functions, integrity, lightweight cryptography, provable security, stream ciphers

Digital Object Identifier 10.4230/DagRep.6.1.34

1 Executive Summary

Frederik Armknecht

Tetsu Iwata

Kaisa Nyberg

Bart Preneel

License © Creative Commons BY 3.0 Unported license
© Frederik Armknecht, Tetsu Iwata, Kaisa Nyberg, and Bart Preneel

One lesson learned from the Snowden leaks is that digital systems can never be fully trusted and hence the security awareness of citizens has increased substantially. Whenever digital data is communicated or stored, it is subject to various attacks. One of the few working countermeasures are the use of cryptography. As Edward Snowden puts it: “*Encryption works. Properly implemented strong crypto systems are one of the few things that you can rely on.*”¹

Consequently it holds that although modern cryptography addresses a variety of security challenges, efficiently protecting the enormous amount of daily electronic communication represents a major challenge. Here, symmetric cryptography is especially highly relevant not only for academia, but also for industrial research and applications.

¹ See <http://techcrunch.com/2013/06/17/encrypting-your-email-works-says-nsa-whistleblower-edward-snowden/>.



Although symmetric cryptography has made enormous progress in the last couple of decades, for several reasons regularly new insights and challenges are evolving. In the past, the AES competition was led by US NIST to standardize a next generation block cipher to replace DES. Similar competitions, such as the eSTREAM and the SHA-3 competition, resulted in new standard algorithms that meet public demands. The outcome of the projects are practically used in our daily lives, and the fundamental understanding of the cryptographic research community of these primitives has been increased significantly.

While this seminar concentrates in general on the design and analysis of symmetric cryptographic primitives, special focus has been put on the following two topics that we explain in more detail below:

1. Authenticated encryption
2. Even-Mansour designs

Authenticated Encryption. Today the central research question is the construction of schemes for *authenticated* encryption. This symmetric primitive efficiently integrates the protection of secrecy and integrity in a single construction. The first wave of solutions resulted in several widely used standards, including CCM and GCM standardized by NIST, and the EAX-prime standardized by ANSI. However, it turns out that these constructions are far from optimum in terms of performance, security, usability, and functionality. For instance a stream of data cannot be protected with CCM, as the length of the entire input has to be known in advance. The security of GCM heavily relies on the existence of data called a nonce, which is supposed to never be repeated. Indeed, the security of GCM is completely lost once the nonce is repeated. While it is easy to state such a mathematical assumption, experience shows that there are many practical cases where realizing this condition is very hard. For instance the nonce may repeat if a crypto device is reset with malice aforethought, or as a consequence of physical attacks on the device. Furthermore, weak keys were identified in GCM, and the security of EAX-prime is questionable.

Thus there is a strong demand for secure and efficient authenticating encryption scheme. As a consequence, the CAESAR project (Competition for Authenticated Encryption: Security, Applicability, and Robustness) has been initiated.² The goal of the project is to identify a portfolio of authenticated encryption schemes that (1) offer advantages over GCM/CCM and (2) are suitable for widespread adoption. The deadline of the submission was March 15, 2014, and the project attracted a total of 56 algorithms from 136 designers from all over the world. There are plenty of innovative designs with attractive features, and the final portfolio is planned to be announced at the end of 2017.

This seminar took place in the middle of the CAESAR competition; it is two years from the submission deadline and we have about two years until the announcement of the final portfolio. Therefore, it was a perfect point in time to sum up the research done so far, to exchange ideas and to discuss future directions.

Even-Mansour Designs. Another strong trend in the current symmetric key cryptography is related to the so-called *Even-Mansour designs*. This design paradigm was proposed in 1991 and can be seen as the abstraction of the framework adopted in the design of AES. This general design framework iterates r times the xor of a key and a public permutation. The design framework is highly relevant in practice, and it has been adopted in a variety of recent hash functions, block ciphers, and even in the underlying primitive of several CAESAR submissions. Despite its long history of practical use, the community has so far failed to

² See <http://competitions.cr.yo.to/caesar.html> for details.

develop a complete understanding of its security. From a theoretical viewpoint, the original proposal was accompanied with a proof of security, dealing with the case of $r = 1$ iteration.

Only 20 years after the initial proposal, in 2012, a bound was proven for the security of $r = 2$ iterations. In 2014, the question was solved to cover the general case of r iterations. However, these results only deal with the simple case of distinguishing attack on a single, unknown key setting. Its security in more advanced, yet practically relevant security models, such as the related-key setting or the chosen/known-key setting, is largely unexplored.

Another problem here is that the theoretical analysis assumes that the permutation used therein is ideal and the keys are ideally random, which is not the case for practical constructions. This implies that the theoretical results do not directly translate into the practical constructions, and the security analysis has to be repeated for each constructions.

Summing up, Evan-Mansour designs represent a fruitful and challenging area of research, that hopefully will lead to a fundamental understanding of iterated constructions and ultimately to more efficient and more secure ciphers.

Seminar Program. The seminar program consists of the presentations about the above topics, and relevant areas of symmetric cryptography, including new cryptanalytic techniques and new designs. Furthermore, there were three discussion sessions. In “discussion on attacks,” we discussed what constitutes a valid cryptographic attack in light of weak key classes, “discussion on secret agency crypto standards” was about cryptography developed by secret agencies, and there was a discussion session about the ongoing CAESAR project.

2 Table of Contents

Executive Summary

Frederik Armknecht, Tetsu Iwata, Kaisa Nyberg, and Bart Preneel 34

Overview of Talks

On Ciphers that Continuously Access the Non-Volatile Key <i>Frederik Armknecht</i>	39
Another view of the division property <i>Anne Canteaut</i>	39
How to Tweak Even-Mansour Ciphers <i>Benoît Cogliati</i>	40
On modes and primitives in the CAESAR competition <i>Joan Daemen</i>	40
New Attacks on Hash function Combiners <i>Itai Dinur</i>	41
Second Preimage Attacks against Dithered Hash Functions with Practical Online Memory Complexity <i>Orr Dunkelman</i>	41
Some Results on the GOST block ciphers <i>Orr Dunkelman, Ashur Tomer, Bar-On Achiya, and Keller Nathan</i>	42
Provable Security Evaluation of Structures against Impossible Differential and Zero Correlation Linear Cryptanalysis <i>Jian Guo</i>	42
On GCM-SIV <i>Tetsu Iwata</i>	43
Key Alternating PRFs and provable security of stream ciphers against time-memory-data tradeoff attacks <i>Matthias Krause</i>	43
Even-Mansour Type Block Ciphers Based on Involutions <i>Jooyoung Lee</i>	43
Dynamic Cube Attacks Revisited, with Applications to Grain-128a <i>Willi Meier</i>	44
Improved Masking for Tweakable Blockciphers with Applications to Authenticated Encryption <i>Bart Mennink</i>	44
Parallel MAC with Low Overhead <i>Kazuhiko Minematsu</i>	45
Simpira: A Family of Efficient Permutations Using the AES Round Function <i>Nicky Mouha</i>	46
Revisiting Structure Graph and Its Applications to CBC-MAC and EMAC <i>Mridul Nandi</i>	46

Even-Mansour cipher analysis reduced to the generalized birthday problem <i>Ivica Nikolic</i>	47
The Problem of Estimating the Variance of the Linear Cryptanalysis Test Statistic <i>Kaisa Nyberg</i>	47
Mirror Theory and Cryptography <i>Jacques Patarin</i>	49
S-Box Reverse-Engineering: Recovering Design Criteria, Hidden Structures and New Boolean Function Results <i>Léo Paul Perrin and Alex Biryukov</i>	50
Invariant Subspace Attack Against Full Midori64 <i>Yu Sasaki</i>	50
Transitivity aspects of the (iterated) Even-Mansour cipher <i>Ernst Schulte-Geers</i>	51
Polytopic cryptanalysis <i>Tyge Tiessen</i>	52
Universal Multidimensional and Multiple Zero-Correlation Cryptanalysis <i>Meiqin Wang</i>	52
Bit Cryptanalysis on Symmetric Ciphers <i>Xianyun Wang</i>	53
Panel discussions	
Discussion on Secret Agency Crypto Standards <i>Orr Dunkelman</i>	53
Participants	54

3 Overview of Talks

3.1 On Ciphers that Continuously Access the Non-Volatile Key

Frederik Armknecht (Universität Mannheim, DE)

License © Creative Commons BY 3.0 Unported license
© Frederik Armknecht

Joint work of Frederik Armknecht, Christian Müller, Vasily Mikhalev

Due to the increased use of devices with restricted resources, the community has developed various techniques for designing lightweight ciphers. One approach that is increasingly discussed is to use the key that is stored on the device in non-volatile memory not only for initialization but during the encryption/decryption process as well. This may on the one hand help to save area size, but also may allow for a stronger key involvement and hence higher security.

However, only little is known so far if and to what extent this approach is indeed practical. In this work, we investigate this question. After a discussion on reasonable approaches for storing a key in non-volatile memory, motivated by several commercial products we focus on the case that the key is stored in EEPROM. Here, we highlight existing constraints and derive that some designs are better suited for reducing the area size than others. Based on these findings, we improve an existing design for proposing a new lightweight stream cipher that (i) has a significantly smaller area size than almost all other stream ciphers and (ii) can be efficiently realized using common non-volatile memory techniques. Hence, we see our work as an important step towards putting such designs on a more solid ground and to initiate further discussions on realistic designs.

3.2 Another view of the division property

Anne Canteaut (INRIA – Paris, FR)

License © Creative Commons BY 3.0 Unported license
© Anne Canteaut

Joint work of Anne Canteaut, Christina Boura

A new distinguishing property against block ciphers, called the division property, was introduced by Todo at Eurocrypt 2015. Our work gives a new approach to it by the introduction of the notion of parity sets. First of all, this new notion permits us to formulate and characterize in a simple way the division property of any order. At a second step, we are interested in the way of building distinguishers on a block cipher by considering some further properties of parity sets, generalising the division property. We detail in particular this approach for substitution-permutation networks. To illustrate our method, we provide low-data distinguishers against reduced-round Present. These distinguishers reach a much higher number of rounds than generic distinguishers based on the division property and demonstrate, amongst others, how the distinguishers can be improved when the properties of the linear and the Sbox layer are taken into account.

3.3 How to Tweak Even-Mansour Ciphers

Benoît Cogliati (University of Versailles, FR)

License  Creative Commons BY 3.0 Unported license
© Benoît Cogliati

Joint work of Benoît Cogliati, Rodolphe Lampe, Yannick Seurin

Tweakable block ciphers are a generalization of traditional block ciphers which take an extra input for variability called a tweak. This primitive has proved to be useful to construct various higher level cryptographic schemes such as length-preserving encryption modes, online ciphers, message authentication codes and authenticated encryption modes.

In this talk, we focus on the state of the art about the construction of efficient tweakable block ciphers in the Random Permutation model, where all parties have access to public random permutation oracles, using generalizations of the standard Even-Mansour construction. We present the most recent constructions (Mennink’s XPX construction [1], the TEM construction introduced by Cogliati et al [2]. and the MEM construction introduced by Granger et al [3]) and their best known security results. We also explain the proof techniques behind those results, which are all based on Patarin’s H coefficient technique, and discuss some related open problems.

References

- 1 Mennink, B. *XPX: Generalized Tweakable Even-Mansour with Improved Security Guarantees*. IACR Cryptology ePrint Archive 2015:476 (2015).
- 2 Cogliati, B., Lampe, R., Seurin, Y. *Tweaking Even-Mansour Ciphers*. Advances in Cryptology – CRYPTO 2015 – Proceedings, Part I, volume 9215 of LNCS, pages 189–208. Springer Berlin Heidelberg (2015).
- 3 Granger, R., Jovanovic, P., Mennink, B., Neves, S. *Improved Masking for Tweakable Block-ciphers with Applications to Authenticated Encryption*. Advances in Cryptology – EURO-CRYPT 2016, to appear. Springer Berlin Heidelberg (2016).

3.4 On modes and primitives in the CAESAR competition

Joan Daemen (STMicroelectronics – Diegem, BE)

License  Creative Commons BY 3.0 Unported license
© Joan Daemen

I have made a proposal for the evaluation of 2nd round candidates in the CAESAR competition for authenticated encryption schemes. This proposal mainly consists in separately evaluating primitives (block ciphers, tweakable block ciphers, permutations, . . .) from modes (sponge, OCB, . . .). In many candidates there is a clear distinction between the two and across candidates very similar modes or primitives are used. In many candidates the novelty is concentrated in either the mode or the primitive. These typically take as primitive a standard block cipher like AES or as mode a (close variant) of a published mode such as OCB. There are a few 2nd round candidates for which this split does not apply and that will have to be evaluated as a whole. I illustrated the proposal with a preliminary classification of the modes and primitives in the 2nd round CAESAR candidates.

The presentation gave rise to some discussion and finally a meeting of the CAESAR committee was held at Dagstuhl. The evaluation of the 2nd round candidates will use some of the presented ideas.

3.5 New Attacks on Hash function Combiners

Itai Dinur (Ben Gurion University – Beer Sheva, IL)

License © Creative Commons BY 3.0 Unported license
© Itai Dinur

Main reference I. Dinur, “New Attacks on the Concatenation and XOR Hash Combiners”, IACR Cryptology ePrint Archive, Report 2016/131, 2016.

URL <http://eprint.iacr.org/2016/131>

We study the security of the concatenation combiner $H_1(M)\|H_2(M)$ for two independent iterated hash functions with n -bit outputs that are built using the Merkle-Damgård construction. In 2004 Joux showed that the concatenation combiner of hash functions with an n -bit internal state does not offer better collision and preimage resistance compared to a single strong n -bit hash function. On the other hand, the problem of devising second preimage attacks faster than 2^n against this combiner has remained open since 2005 when Kelsey and Schneier showed that a single Merkle-Damgård hash function does not offer optimal second preimage resistance for long messages.

In this paper, we develop new algorithms for cryptanalysis of hash combiners and use them to devise the first second preimage attack on the concatenation combiner. The attack finds second preimages faster than 2^n for messages longer than $2^{2n/7}$ and has optimal complexity of $2^{3n/4}$. This shows that the concatenation of two Merkle-Damgård hash functions is not as strong as a single ideal hash function.

Our methods are also applicable to other well-studied combiners, and we use them to devise a new preimage attack with complexity of $2^{2n/3}$ on the XOR combiner $H_1(M)\oplus H_2(M)$ of two Merkle-Damgård hash functions. This improves upon the attack by Leurent and Wang (presented at Eurocrypt 2015) whose complexity is $2^{5n/6}$ (but unlike our attack is also applicable to HAIFA hash functions).

Our algorithms exploit properties of random mappings generated by fixing the message block input to the compression functions of H_1 and H_2 . Such random mappings have been widely used in cryptanalysis, but we exploit them in new ways to attack hash function combiners.

3.6 Second Preimage Attacks against Dithered Hash Functions with Practical Online Memory Complexity

Orr Dunkelman (University of Haifa, IL)

License © Creative Commons BY 3.0 Unported license
© Orr Dunkelman

Joint work of Orr Dunkelman, Barham Muhammad

In this work we show how to reduce the online memory complexity of second preimage attacks against dithered hash functions to less than 1 GB.

3.7 Some Results on the GOST block ciphers

Orr Dunkelman (University of Haifa, IL)

License © Creative Commons BY 3.0 Unported license

© Orr Dunkelman, Ashur Tomer, Bar-On Achiya, and Keller Nathan

Joint work of Orr Dunkelman, Ashur Tomer, Bar-On Achiya, Keller Nathan

The talk covered several new attacks reported against the GOST family of block ciphers:

1. Attacking GOST2 using a reflection property (for a weak key class of 2^{224} keys),
2. New improved cycle finding attack on GOST's original key schedule – attacking the weak key class of $K1K2K3K4K4K3K2K1$ in 2^{36} data and 2^{40} time.

3.8 Provable Security Evaluation of Structures against Impossible Differential and Zero Correlation Linear Cryptanalysis

Jian Guo (Nanyang TU – Singapore, SG)

License © Creative Commons BY 3.0 Unported license

© Jian Guo

Joint work of Bing Sun, Meicheng Liu, Jian Guo, Vincent Rijmen, Ruilin Li

Main reference B. Sun, Z. Liu, V. Rijmen, R. Li, L. Cheng, Q. Wang, H. AlKhzaimi, C. Li, “Links Among Impossible Differential, Integral and Zero Correlation Linear Cryptanalysis,” in Proc. of the 35th Annual Cryptology Conference – Advances in Cryptology (CRYPTO'15), LNCS, Vol. 9215, pp. 95–115, Springer, 2015.

URL http://dx.doi.org/10.1007/978-3-662-47989-6_5

Impossible differential and zero correlation linear cryptanalysis are two of the most important cryptanalytic vectors. To characterize the impossible differentials and zero correlation linear hulls which are independent of the choices of the non-linear components, Sun *et al.* proposed the structure deduced by a block cipher at CRYPTO 2015. Based on that, we concentrate in this paper on the security of the SPN structure and Feistel structure with SP-type round functions. Firstly, we prove that for an SPN structure, if $\alpha_1 \rightarrow \beta_1$ and $\alpha_2 \rightarrow \beta_2$ are possible differentials, $\alpha_1|\alpha_2 \rightarrow \beta_1|\beta_2$ is also a possible differential, i.e., the OR “|” operation preserves differentials. Secondly, we show that for an SPN structure, there exists an r -round impossible differential if and only if there exists an r -round impossible differential $\alpha \not\rightarrow \beta$ where the Hamming weights of both α and β are 1. Thus for an SPN structure operating on m bytes, the computation complexity for deciding whether there exists an impossible differential can be reduced from $\mathcal{O}(2^{2m})$ to $\mathcal{O}(m^2)$. Thirdly, we associate a primitive index with the linear layers of SPN structures. Based on the matrices theory over integer rings, we prove that the length of impossible differentials of an SPN structure is upper bounded by the primitive index of the linear layers. As a result we show that, unless the details of the S-boxes are considered, there do not exist 5-round impossible differentials for the AES and ARIA. Lastly, based on the links between impossible differential and zero correlation linear hull, we projected these results on impossible differentials to zero correlation linear hulls. It is interesting to note some of our results also apply to the Feistel structures with SP-type round functions.

3.9 On GCM-SIV

Tetsu Iwata (Nagoya University, JP)

License © Creative Commons BY 3.0 Unported license
© Tetsu Iwata

Joint work of Tetsu Iwata, Kazuhiko Minematsu

At CCS 2015, Gueron and Lindell proposed GCM-SIV, a provably secure authenticated encryption scheme that remains secure even if the nonce is repeated. We first point out that GCM-SIV allows a trivial distinguishing attack with about $2^{(n-32)/2}$ attack complexity, where n is the block length of the underlying blockcipher and $n = 128$ for GCM-SIV. This shows the tightness of the security claim and does not contradict the provable security result. We present a minor variant of GCM-SIV, which we call GCM-SIV1, that is secure up to the standard birthday-bound-security, in the total number of input blocks, of about $2^{n/2}$ attack complexity. We then explore constructions of a scheme with a stronger security guarantee. We present GCM-SIV2 that is obtained by running two instances of GCM-SIV1 in parallel and mixing them in a simple way. We show that it is secure up to about $2^{2n/3}$ attack complexity. Finally, we generalize this to show GCM-SIV r by running r instances of GCM-SIV1 in parallel, where $r \geq 3$, and show that the scheme is secure up to about $2^{nr/(r+1)}$ attack complexity.

3.10 Key Alternating PRFs and provable security of stream ciphers against time-memory-data tradeoff attacks

Matthias Krause (Universität Mannheim, DE)

License © Creative Commons BY 3.0 Unported license
© Matthias Krause

We consider keystream generator based stream ciphers which generate the keystream packet-wise, like the Bluetooth cipher E_0 . We show a method how to design such ciphers in such a way that beyond-the-birthday-bound security against generic time-memory-tradeoff attacks can be proved. This allows, in principle, for designing practical stream ciphers with a significantly smaller inner state length. One further consequence is that only a small change in the state initialization algorithm of the E_0 -cipher suffices for raising the security level from $n/2$ to $(2/3)n$. We obtain our results by modelling the state initialization – and keystream generation process by Even-Mansour like constructions, and analyzing them in a generalized random oracle model.

3.11 Even-Mansour Type Block Ciphers Based on Involutions

Jooyoung Lee (Sejong University – Seoul, KR)

License © Creative Commons BY 3.0 Unported license
© Jooyoung Lee

In this work, we study the security of Even-Mansour type ciphers whose encryption and decryption are (almost) the same. Such ciphers, called involutorial, possibly allow efficient hardware implementation with a same circuit shared for encryption and decryption, expected

to be suitable for lightweight environment where low power consumption and implementation costs are desirable.

With this motivation, we consider a single-round Even-Mansour cipher using an involution as its basing primitive. Then the decryption of such a cipher is the same as encryption with the order of the round keys reversed. It is known that such a cipher permits an attack using only construction queries below the birthday bound, while it has been open how it provides provable security within the range below the birthday bound. We prove that the Even-Mansour cipher based on a random involution is as secure as the permutation-based one when the number of construction queries is limited by the birthday bound.

In order to achieve security beyond the birthday bound, we propose a two-round Even-Mansour-like construction that makes a single call to each of the basing permutation P and its inverse using a fixed permutation in the middle layer. The security of this construction is proved beyond the birthday bound. As an open problem, we ask for the block cipher construction that uses only a single involution and provides security beyond the birthday bound at the same time.

3.12 Dynamic Cube Attacks Revisited, with Applications to Grain-128a

Willi Meier (FH Nordwestschweiz – Windisch, CH)

License  Creative Commons BY 3.0 Unported license
© Willi Meier

Joint work of Willi Meier, Yonglin Hao

Dynamic cube attacks are revisited, and a probabilistic model of their success is given. This model identifies the main factors influencing the success probability of dynamic cube attacks. Based on this model, a new strategy for constructing the necessary cube testers is provided so that a higher success probability can be acquired. The correctness of our deductions are verified experimentally on the round-reduced stream cipher Grain-128a. Similar methods enable dynamic cube key recovery attacks on up to 177 of the 256 initialization rounds of this cipher. These are the first practical results on key recovery of (round-reduced) Grain-128a in the single key model.

3.13 Improved Masking for Tweakable Blockciphers with Applications to Authenticated Encryption

Bart Mennink (KU Leuven, BE)

License  Creative Commons BY 3.0 Unported license
© Bart Mennink

Joint work of Robert Granger, Philipp Jovanovic, Bart Mennink, Samuel Neves

Main reference R. Granger, P. Jovanovic, B. Mennink, S. Neves, “Improved Masking for Tweakable Blockciphers with Applications to Authenticated Encryption,” IACR Cryptology ePrint Archive, Report 2015/999, 2015.

URL <https://eprint.iacr.org/2015/999>

A popular approach to tweakable blockcipher design is via masking, where a certain primitive (a blockcipher or a permutation) is preceded and followed by an easy-to-compute tweak-dependent mask. In this work, we revisit the principle of masking. We do so alongside the introduction of the tweakable Even-Mansour construction MEM. Its masking function combines the advantages of word-oriented LFSR- and powering-up-based methods. We show

in particular how recent advancements in computing discrete logarithms over finite fields of characteristic 2 can be exploited in a constructive way to realize highly efficient, constant-time masking functions. If the masking satisfies a set of simple conditions, then MEM is a secure tweakable blockcipher up to the birthday bound. The strengths of MEM are exhibited by the design of fully parallelizable authenticated encryption schemes OPP (nonce-respecting) and MRO (misuse-resistant). If instantiated with a reduced-round BLAKE2b permutation, OPP and MRO achieve speeds up to 0.55 and 1.06 cycles per byte on the Intel Haswell microarchitecture, and are able to significantly outperform their closest competitors.

3.14 Parallel MAC with Low Overhead

Kazuhiko Minematsu (NEC – Kawasaki, JP)

License © Creative Commons BY 3.0 Unported license
© Kazuhiko Minematsu

Joint work of Tetsu Iwata, Kazuhiko Minematsu

In this talk we propose a new message authentication code (MAC) mode of operation based on blockcipher. We first survey popular MAC modes, such as CMAC and PMAC. Our survey reveals that there is no known scheme to achieve all the following four properties.

1. Optimal efficiency with pre-computation: m blockcipher calls to process m -block message, for any $m \geq 1$, with one precomputed encrypted block (typically $L = E_K(0^n)$).
2. Quasi-optimal efficiency w/o pre-computation: m BC calls for $m > 1$ and 2 calls for $m = 1$. It does not need a precomputation of L .
3. One-key (key is a BC key)
4. Well parallelizable

Here, CMAC (a.k.a. OMAC [1]) achieves Properties 1 and 3, and PMAC[4] achieves Properties 1 and 3 and 4. A variant of CMAC called GCBC [3] achieves Properties 2 and 3. In other words, what is lacked here is a parallelizable MAC without precomputation of L . It means computation overhead is low, which is important when memory is precious or low-latency operation is required. Based on the work on a MAC proposal by Minematsu[2], we provide a scheme which enables all four properties, in particular, parallelizable up to around n blocks in case n -bit blockcipher is used. The security proof is work in progress but we expect standard birthday-type bound for the forgery probability.

References

- 1 Iwata, T., Kurosawa, K.: OMAC: one-key CBC MAC. In: FSE. Lecture Notes in Computer Science, vol. 2887, pp. 129–153. Springer (2003)
- 2 Minematsu, K.: A short universal hash function from bit rotation, and applications to blockcipher modes. In: ProvSec. Lecture Notes in Computer Science, vol. 8209, pp. 221–238. Springer (2013)
- 3 Nandi, M.: Fast and Secure CBC-Type MAC Algorithms. In: FSE. Lecture Notes in Computer Science, vol. 5665, pp. 375–393. Springer (2009)
- 4 Rogaway, P.: Efficient Instantiations of Tweakable Blockciphers and Refinements to Modes OCB and PMAC. In: Lee, P.J. (ed.) ASIACRYPT. Lecture Notes in Computer Science, vol. 3329, pp. 16–31. Springer (2004)

3.15 Simpira: A Family of Efficient Permutations Using the AES Round Function

Nicky Mouha (KU Leuven, BE)

License © Creative Commons BY 3.0 Unported license
© Nicky Mouha

Joint work of Shay Gueron, Nicky Mouha

Main reference S. Gueron, N. Mouha, “Simpira v2: A Family of Efficient Permutations Using the AES Round Function,” IACR Cryptology ePrint Archive, Report 2016/122, 2016.

URL <https://eprint.iacr.org/2016/122>

This talk introduces Simpira, a family of cryptographic permutations that supports inputs of $128 * b$ bits, where b is a positive integer. Its design goal is to achieve high throughput on virtually all modern 64-bit processor architectures, that nowadays already have native instructions to support AES computations. To achieve this goal, Simpira uses only one building block: the AES round function. For $b = 1$, Simpira corresponds to 12-round AES with fixed round keys, whereas for $b \geq 2$, Simpira is a Generalized Feistel Structure (GFS) with an F-function that consists of two rounds of AES. From the security viewpoint, we claim that there are no structural distinguishers for Simpira with a complexity below 2^{128} , and analyze its security against a variety of attacks in this setting. From the efficiency viewpoint, we show that the throughput of Simpira is close to the theoretical optimum, namely, the number of AES rounds in the construction. For example, on the latest Intel Skylake processor, Simpira has throughput below 1 cycle per byte for $b \leq 4$ and $b = 6$. For larger permutations, where moving data in memory has a more pronounced effect, Simpira with $b = 32$ (512 byte inputs) evaluates 732 AES rounds, and performs at 802 cycles (1.56 cycles per byte), i.e., less than 10% off the theoretical optimum. The Simpira family offers an efficient solution for multiple usages where operating on wide blocks, larger than 128 bits, is desired.

3.16 Revisiting Structure Graph and Its Applications to CBC-MAC and EMAC

Mridul Nandi (Indian Statistical Institute – Kolkata, IN)

License © Creative Commons BY 3.0 Unported license
© Mridul Nandi

Joint work of Mridul Nandi, Ashwin Jha

Main reference A. Jha, M. Nandi, “Revisiting Structure Graph and Its Applications to CBC-MAC and EMAC,” IACR Cryptology ePrint Archive, Report 2016/161, 2016.

URL <http://eprint.iacr.org/2016/161>

In CRYPTO’05, Bellare et al. proved $O(\ell q^2/2^n)$ bound for the PRF (pseudorandom function) security of the CBC-MAC based on an n -bit random permutation Π , provided $\ell < 2^{n/3}$. Here an adversary can make at most q prefix-free queries each having at most ℓ “blocks” (elements of $\{0, 1\}^n$). In the same paper $O(\ell^{o(1)} q^2/2^n)$ bound for EMAC (or encrypted CBC-MAC) was proved, provided $\ell < 2^{n/4}$. Both proofs are based on **structure graphs** representing all collisions among “intermediate inputs” to Π during the computation of CBC. The problem of bounding PRF-advantage is shown to be reduced to bounding the number of structure graphs satisfying certain collision patterns. Unfortunately, we have shown here that *the Lemma 10 in the Crypto’05 paper, stating an important result on structure graphs, is incorrect*. This is due to the fact that the authors **overlooked certain structure graphs**. This invalidates the proofs of the PRF bounds. In ICALP’06, Pietrzak improved the bound for EMAC by

showing a *tight bound* $O(q^2/2^n)$ under the restriction that $\ell < 2^{n/8}$. As he used the same flawed lemma, this proof also becomes invalid. In this paper, we have revised and sometimes simplified these proofs. We revisit structure graphs in a slightly different mathematical language and provide a complete characterization of certain types of structure graphs. Using this characterization, we show that PRF security of CBC-MAC is about $\sigma q/2^n$ provided $\ell < 2^{n/3}$ where σ is the total number of blocks in all queries. We also recovered the tight bound of EMAC with a much relaxed constraint $\ell < 2^{n/4}$ than the original.

3.17 Even-Mansour cipher analysis reduced to the generalized birthday problem

Ivica Nikolic (Nanyang TU – Singapore, SG)

License  Creative Commons BY 3.0 Unported license
© Ivica Nikolic

We show that full subkey recovery of iterated Even-Mansour ciphers can be reduced to the generalized birthday problem.

3.18 The Problem of Estimating the Variance of the Linear Cryptanalysis Test Statistic

Kaisa Nyberg (Aalto University, FI)

License  Creative Commons BY 3.0 Unported license
© Kaisa Nyberg

Joint work of Celine Blondeau, Kaisa Nyberg

Main reference C. Blondeau, K. Nyberg, “Joint Data and Key Distribution of the Linear Cryptanalysis Test Statistic and Its Impact to Data Complexity Estimates of Multiple/Multidimensional Linear and Truncated Differential Attacks,” IACR Cryptology ePrint Archive, Report 2015/935, 2015.

URL <http://eprint.iacr.org/2015/935>

Until recently, most statistical models of linear key-recovery attacks determine and analyze the attack statistic with fixed keys and taking only the data as a random variable. When using such models in practice, it is assumed that for all cipher keys, all wrong key candidates draw the value of the test statistic from the same (uniform) distribution, and similarly, all correct key candidates draw the value of the test statistic from the same (non-uniform) distribution. Previously, in [4, 5] experiments were provided to demonstrate that the probability distributions of the test statistic vary significantly over the key. In [3], the simple linear attack using one linear approximation with a single dominant trail was considered and the wrong-key randomization hypothesis revised accordingly. As the result, the estimate of the data complexity was improved as demonstrated in experiments. In [6], the variation in the probability distribution of capacity over the right key was studied in the context of multiple and multidimensional linear attacks. In particular, the authors determined weak-key quantiles, that is, lower bounds of capacity that are satisfied by a given proportion, say one half, or 30% of the keys. Such approach was previously taken also in [7] in the case of single linear hull.

In [1] we presented the first complete treatment on the probability distributions of linear attack test statistics, that is, the empirical correlations and capacities, by considering both the data sample and the key as random variables. We analyzed and combined the different

previously presented models and went beyond by studying the joint probability distribution of the test statistic, where in addition to the data, also both the wrong and right keys are taken as random variables. From this model one can derive formulas for success probability and data complexity in multiple and multidimensional linear key-recovery attacks.

We also mention in [1] how to apply the same approach to the simplest case of linear key-recovery attack, that is, Matsui's Algorithm 2, which uses one linear approximation with a single dominant trail. In this case, the probability distribution of the empirical correlation observed from data with the correct key can be approximated by a union of two normal distributions. For details, we refer to [9]. One of the main benefits of integrating the key as a random variable in the model is that the data complexity of the attack can be expressed as a function of the *ELP* of the linear approximation. Until now, the data complexity was determined from a fixed-key statistical model and assuming that the expected capacity of the probability distribution of the test statistic is equal for all keys. The new integrated statistical model gives the data complexity estimate for a random key. As a consequence, the issue raised in [8] is resolved. In particular, the fact that multiple strong characteristics cancel each other for many keys is not a problem for linear cryptanalysis in general. Indeed, it is very likely that the average correlation is equal to zero. The situation is as stated in [5]: "The average correlation of a hull gives no indication about the complexity of a linear attack. Therefore, we only talk about the *ELP* of a hull." While it has been known by most authors that *ELP* is the right quantity to consider in the context of linear attacks, no satisfactory presentation of exactly how it determines the data complexity of the attack for a random encryption key has not been given in the literature until now.

Two major problems remained in the treatment given in [1]. First, it was observed that the formula of variance of capacity gives serious underestimates in the experiments on SmallPRESENT. This formula originated from the work of [6] and was obtained under the assumption of independently and identically distributed correlations of the involved linear approximations.

Secondly, using the results of our analysis we ended up with somewhat pessimistic results about the success of previous attacks on PRESENT. In particular, we had estimated the capacity based on the enumerated characteristics of the strongest linear approximations, and concluded that if this capacity estimate is less than the capacity of random noise, then distinguishing of the wrong-key and right-key distributions becomes impossible. Fortunately, this problem turned out to be easy to solve. In this Dagstuhl seminar, it was pointed out to us by Bogdanov [2] that also many weaker linear approximations contribute to the total capacity at least as much as random noise. Indeed, if their impact is taken into account (similarly as we had done in our analysis of Matsui's Algorithm 2) and even if not more than random noise, the capacity estimate will never be less than the capacity of random noise. It follows that distinguishing may be possible depending now crucially on the variances of the distributions of the test statistics. Then it is even more important to get the variances correct.

In this talk, we focused on the non-trivial problem of obtaining an accurate estimate of the variance of the capacity of the value distribution of the test statistic in the multidimensional linear key-recovery attack. In this context, the set of linear approximations involved in the online attack is typically not the same as the one used in the offline analysis of the capacity. In the offline analysis the cryptanalyst usually identifies only the strongest linear approximations, which form a small subset of all linear approximations involved in the multidimensional linear attack. Moreover, it is often possible to get accurate estimates of their *ELPs*, which in turn allow a more realistic estimate of the variance of the test statistic.

References

- 1 Céline Blondeau and Kaisa Nyberg. Joint data and key distribution of the linear cryptanalysis test statistic and its impact to data complexity estimates of multiple/multidimensional linear and truncated differential attacks. *IACR Cryptology ePrint Archive*, 2015:935, 2015.
- 2 Andrey Bogdanov. Private communication. Dagstuhl seminar 16021“Symmetric Cryptography”, 2016.
- 3 Andrey Bogdanov and Elmar Tischhauser. On the wrong key randomisation and key equivalence hypotheses in Matsui’s Algorithm 2. In Shiho Moriai, editor, *Fast Software Encryption – 20th International Workshop, FSE 2013, Singapore, March 11-13, 2013. Revised Selected Papers*, volume 8424 of *Lecture Notes in Computer Science*, pages 19–38. Springer, 2013.
- 4 Joan Daemen and Vincent Rijmen. Probability distributions of correlation and differentials in block ciphers. *IACR Cryptology ePrint Archive*, 2005:212, 2006.
- 5 Joan Daemen and Vincent Rijmen. Probability distributions of correlation and differentials in block ciphers. *J. Mathematical Cryptology*, 1(3):221–242, 2007.
- 6 Jialin Huang, Serge Vaudenay, Xuejia Lai, and Kaisa Nyberg. Capacity and data complexity in multidimensional linear attack. In Rosario Gennaro and Matthew Robshaw, editors, *Advances in Cryptology – CRYPTO 2015 – 35th Annual Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2015, Proceedings, Part I*, volume 9215 of *Lecture Notes in Computer Science*, pages 141–160. Springer, 2015.
- 7 Gregor Leander. On Linear Hulls, Statistical Saturation Attacks, PRESENT and a Cryptanalysis of PUFFIN. In K. G. Paterson, editor, *EUROCRYPT*, volume 6632 of *LNCS*, pages 303–322. Springer, 2011.
- 8 Sean Murphy. The effectiveness of the linear hull effect. Technical report, Royal Holloway College London, 2009.
- 9 Kaisa Nyberg. Linear cryptanalysis. *SAC Summer School, Sackville, New Brunswick*, 2015.

3.19 Mirror Theory and Cryptography

Jacques Patarin (University of Versailles, FR)

License  Creative Commons BY 3.0 Unported license
© Jacques Patarin

“Mirror Theory” is the theory that evaluates the number of solutions of affine systems of equalities ($=$) and non equalities (\neq) in finite groups. It is deeply related to the security and attacks of many generic cryptographic secret key schemes, for example random Feistel schemes (balanced or unbalanced), Misty schemes, Xor of two pseudo-random bijections to generate a pseudo-random function etc. We will present here general definitions, some theorems, and many examples and computer simulations.

3.20 S-Box Reverse-Engineering: Recovering Design Criteria, Hidden Structures and New Boolean Function Results

Léo Paul Perrin (University of Luxembourg, LU) and Alex Biryukov (University of Luxembourg, LU)

License © Creative Commons BY 3.0 Unported license

© Léo Paul Perrin and Alex Biryukov

Joint work of Léo Paul Perrin, Alex Biryukov, Aleksei Udovenko

Main reference A. Biryukov, L. Perrin, A. Udovenko, “Reverse-Engineering the S-Box of Streebog, Kuznyechik and STRIBOBr1,” *Advances in Cryptology – EUROCRYPT 2016*, to appear 2016.

S-Boxes are key components of many symmetric primitives. Their properties can be used to provide convincing security arguments. However, they may be specified using only a look-up table without providing any rationale. Skipjack, designed by the American NSA, and Kuznyechik, designed by the Russian FSB, are two block ciphers with S-Boxes designed in an unknown fashion.

In this talk, we described how to analyse S-Boxes with secret design criteria or structure. First, a statistical test based on the differential and linear properties of the S-Box can be used to rule out randomness [1]. Second, visual patterns in the Linear Approximation Table can provide useful informations. In fact, we described how these were used in the first step of our reverse-engineering of the S-Box of the last Russian standards [2].

We also presented new results on the 6-bit APN permutation published by Dillon et. al. Using the same methods, we found a decomposition of this function which leads to a more efficient implementation. The structure found can also be generalized to larger dimensions and, while not APN, remains differentially 4-uniform.

References

- 1 Biryukov, A., Perrin, L. *On Reverse-Engineering S-Boxes with Hidden Design Criteria or Structure*. *Advances in Cryptology – CRYPTO 2015. Lecture Notes in Computer Science*, pp. 116–140. Springer Berlin Heidelberg (2015).
- 2 Biryukov, A., Perrin, L., Udovenko, A. *Reverse-Engineering the S-Box of Streebog, Kuznyechik and STRIBOBr1*. *Advances in Cryptology – EUROCRYPT 2016*, to appear.

3.21 Invariant Subspace Attack Against Full Midori64

Yu Sasaki (NTT Labs – Tokyo, JP)

License © Creative Commons BY 3.0 Unported license

© Yu Sasaki

Joint work of Jian Guo, Jérémy Jean, Ivica Nikolić, Kexin Qiao, Yu Sasaki, Siang Meng Sim

We show that the block cipher Midori64 allows a class of invariant subspace. With 2^{32} fractions of the key, the cipher can be distinguished from random permutation with 1 chosen plaintext query. In addition, the key can be recovered with 2 chosen plaintext queries and 2^{18} computations. We the investigate further research directions. The first approach is extending the class of invariant subspaces, which reveals weaker keys. The second approach is designing S-boxes that resist the invariant subspace no matter how the other components of the cipher is chosen. The last approach is a probabilistic transition, which can be applied to reduced-round versions of Midori128.

3.22 Transitivity aspects of the (iterated) Even-Mansour cipher

Ernst Schulte-Geers (BSI – Bonn, DE)

License  Creative Commons BY 3.0 Unported license
© Ernst Schulte-Geers

As a consequence of the CSFG the highly transitive permutation groups have also been classified in the past century.

In particular, the following is true:

► **Theorem.** *Let G be a permutation group on a set X with $|X| \geq 25$. If G is neither the alternating group $A(X)$ nor the symmetric group $S(X)$, then G is at most 3-transitive.*

We interpret the implications for iterated Even-Mansour constructions with non-ideal public permutations P_i (e.g. round functions), and with $X = \{0, 1\}^n$ ($n \geq 5$) as the set plain-/ciphertext blocks.

For non-ideal P_i it may be desirable to strengthen the encryption by iterating several rounds (with independent keys). Under the assumption that the permutation group G generated by the keyed encryption functions is (at least) the alternating group $A(X)$, this generating process should be as fast as possible, i.e. the key additions should interact with the public permutations in such a way such that the r -round encryptions are “totally unrelated, diverse” permutations. From the permutation group viewpoint this is interpreted here as the requirement that no large fraction of the keyed round functions should lie (in large part) in the same “small” permutation group (otherwise the r -round encryptions would “leave” this group only slowly).

Interpreting “small” as “small transitivity”, in our view the theorem above then suggests the following aim: the keyed encryption functions should “look” 4-transitive after as few iterations as possible (since by the theorem above (and recalling that $A(X)$ resp. $S(X)$ are $(|X| - 2)$ - resp. $|X|$ -transitive) a 4-transitive permutation group on X is either $A(X)$ or $S(X)$).

This aim seems only loosely related to conventional cryptographic quality criteria (consider the case where each P_i is the inversion in $\text{GF}(2^n)$).

Ideally only 4 independent keys (i.e. 3 rounds E-M) could suffice to reach the aim.

The orbit counting lemma gives the possibility to estimate statistically the “transitivity look” of r -round encryption functions: the first four factorial moments of the empirical fixed point distribution should all be (approximately) 1, in this respect the empirical fixed point distribution should resemble the *Poiss*(1) distribution. (Recall $\text{Prob}(\text{Poiss}(1) = k) = e^{-1}/k!$.) This gives also a theoretical means to determine a round number: take (say) the smallest no. of rounds (with independent keys) for which the first four factorial moments are close enough to 1 (of course, such a decision would have to be further backed up by cryptanalysis).

However, for block sizes of practical interest this method is impractical.

A better understanding of the mechanisms which lead to maximal diversity, and a practical diversity measure would be desirable.

3.23 Polytopic cryptanalysis

Tyge Tiessen (Technical University of Denmark – Lyngby, DK)

License © Creative Commons BY 3.0 Unported license
© Tyge Tiessen

Main reference T. Tiessen, “Polytopic Cryptanalysis,” IACR Cryptology ePrint Archive, Report 2016/160, 2016.

URL <https://eprint.iacr.org/2016/160.pdf>

Standard differential cryptanalysis uses statistical dependencies between the difference of two plaintexts and the difference of the respective two ciphertexts to attack a cipher. Here we introduce polytopic cryptanalysis which considers interdependencies between larger sets of texts as they traverse through the cipher. We prove that the methodology of standard differential cryptanalysis can unambiguously be extended and transferred to the polytopic case including impossible differentials. We show that impossible polytopic transitions have generic advantages over impossible differentials. To demonstrate the practical relevance of the generalization, we present new low-data attacks on round-reduced DES and AES using impossible polytopic transitions that are able to compete with existing attacks, partially outperforming these.

3.24 Universal Multidimensional and Multiple Zero-Correlation Cryptanalysis

Meiqin Wang (Shandong University – Jinan, CN)

License © Creative Commons BY 3.0 Unported license
© Meiqin Wang

Joint work of Ling Sun, Huaifeng Chen, Meiqin Wang

Multidimensional zero-correlation linear attack and multiple zero-correlation linear attack have been two of the most powerful cryptanalytic techniques for block ciphers. Nevertheless, questions remain regarding how these attacks can be universal without any limitations and can be used to accurately estimate data complexity and success probability. More concretely, the current models for multidimensional and multiple zero-correlation cryptanalysis are not valid in the setting with limited number of zero-correlation linear approximations and the accuracy of the estimation for data complexity can not be guaranteed under that setting. However, in a lot of cases, using too many zero-correlation linear approximations may cause an unacceptable time complexity which leads the attack unfeasible. In order to construct the generalization of the original models built by Bogdanov *et al.* using normal approximation of χ^2 -distribution, we provide new models to estimate data complexity and success probability for multidimensional and multiple zero-correlation attacks without such approximation. As a result, our new models are valid in every setting of multidimensional and multiple linear attacks, which release the limitation on the number of zero-correlation linear approximations, so we name them as universal multidimensional and multiple zero-correlation linear distinguishers.

As an illustration, we apply the universal multiple zero-correlation linear attack on TEA and XTEA. These new attacks can cover more rounds of TEA and XTEA than the previous multiple zero-correlation attacks. Moreover, we reevaluate almost all existing multidimensional and multiple zero-correlation cryptanalysis for various block ciphers, such as CLEFIA, Camellia, LBlock, TWINE, E2, and so on.

3.25 Bit Cryptanalysis on Symmetric Ciphers

Xianyun Wang (Tsinghua University – Beijing, CN)

License  Creative Commons BY 3.0 Unported license
© Xianyun Wang

This talk recalls the existing three main differential attacks: XOR differential attack, modular differential attack and conditional differential attack, and the bit cryptanalysis means the modular differential attack or the XOR differential attack by considering the bit conditions to ensure the differential path hold.

This talk introduces the details of the bit cryptanalysis in differential attack, linear attack and cube attack respectively. As a result, we get the best differential attacks and the linear hull attacks on the full 10 round-reduced SIMON versions, and the cube attack combining with bit cryptanalysis can results in the new key recovery attack on the reduced Keccak-MAC.

4 Panel discussions

4.1 Discussion on Secret Agency Crypto Standards

Orr Dunkelman (University of Haifa, IL)

License  Creative Commons BY 3.0 Unported license
© Orr Dunkelman

The discussion was about what should be the assurance level we need to require as community from cryptography developed by secret agencies.

Participants

- Elena Andreeva
KU Leuven, BE
- Frederik Armknecht
Universität Mannheim, DE
- Daniel J. Bernstein
Univ. of Illinois – Chicago, US
- Eli Biham
Technion – Haifa, IL
- Alex Biryukov
University of Luxembourg, LU
- Andrey Bogdanov
Technical University of Denmark
– Lyngby, DK
- Anne Canteaut
INRIA – Paris, FR
- Benoît Cogliati
University of Versailles, FR
- Joan Daemen
STMicroelectronics –
Diegem, BE
- Itai Dinur
Ben Gurion University – Beer
Sheva, IL
- Orr Dunkelman
University of Haifa, IL
- Henri Gilbert
ANSSI – Paris, FR
- Jian Guo
Nanyang TU – Singapore, SG
- Matthias Hamann
Universität Mannheim, DE
- Tetsu Iwata
Nagoya University, JP
- Jérémy Jean
ANSSI – Paris, FR
- Antoine Joux
UPMC – Paris, FR
- Dmitry Khovratovich
University of Luxembourg, LU
- Matthias Krause
Universität Mannheim, DE
- Nils Gregor Leander
Ruhr-Universität Bochum, DE
- Jooyoung Lee
Sejong University – Seoul, KR
- Gaëtan Leurent
INRIA – Paris, FR
- Stefan Lucks
Bauhaus-Universität Weimar, DE
- Willi Meier
FH Nordwestschweiz –
Windisch, CH
- Bart Mennink
KU Leuven, BE
- Kazuhiko Minematsu
NEC – Kawasaki, JP
- Nicky Mouha
KU Leuven, BE
- Chanathip Namprempre
Thammasat University –
Patumtani, TH
- Mridul Nandi
Indian Statistical Institute –
Kolkata, IN
- Ivica Nikolic
Nanyang TU – Singapore, SG
- Kaisa Nyberg
Aalto University, FI
- Jacques Patarin
University of Versailles, FR
- Léo Paul Perrin
University of Luxembourg, LU
- Bart Preneel
KU Leuven, BE
- Christian Rechberger
Technical University of Denmark
– Lyngby, DK
- Yu Sasaki
NTT Labs – Tokyo, JP
- Ernst Schulte-Geers
BSI – Bonn, DE
- Adi Shamir
Weizmann Inst. – Rehovot, IL
- John Steinberger
Tsinghua Univ. – Beijing, CN
- Marc Stevens
CWI – Amsterdam, NL
- Tyge Tiessen
Technical University of Denmark
– Lyngby, DK
- Meiqin Wang
Shandong Univ. – Jinan, CN
- Xianyun Wang
Tsinghua Univ. – Beijing, CN
- Kan Yasuda
NTT Labs – Tokyo, JP



Geometric and Graph-based Approaches to Collective Motion

Edited by

Giuseppe F. Italiano¹, Marc van Kreveld², Bettina Speckmann³,
and Guy Theraulaz⁴

1 University of Rome “Tor Vergata”, IT, giuseppe.italiano@uniroma2.it

2 Utrecht University, NL, m.j.vankreveld@uu.nl

3 TU Eindhoven, NL, b.speckmann@tue.nl

4 CNRS and Université Paul Sabatier – Toulouse, FR,
guy.theraulaz@univ-tlse3.fr

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16022 “Geometric and Graph-based Approaches to Collective Motion”.

The seminar brought together a group of enthusiastic researchers with a diverse background. To create a shared body of knowledge the seminar featured a number of survey talks that were planned early in the week. The survey talks were rather engaging: the audience learned for instance at what scale one should look at a painting of Van Gogh, how bats chase each other, what size of clumps mussels make and why, and how to interact with a computational geometer.

Seminar January 10–15, 2016 – <http://www.dagstuhl.de/16022>

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems, I.5 Pattern Recognition

Keywords and phrases Geometry, Graph, Interaction, Motion, Pattern, Trajectory

Digital Object Identifier 10.4230/DagRep.6.1.55

Edited in cooperation with Tim Ophelders

1 Executive Summary

Giuseppe F. Italiano

Marc van Kreveld

Bettina Speckmann

Guy Theraulaz

License © Creative Commons BY 3.0 Unported license
© Giuseppe F. Italiano, Marc van Kreveld, Bettina Speckmann, and Guy Theraulaz

A trajectory is a time-stamped sequence of locations which represents the movement of entities in space. Trajectories are often created by sampling GPS locations and attaching a time-stamp, but they can also originate from RFID tags, video, or radar analysis. Huge data sets exist for entities as diverse as birds, deer, traveling humans, sports players, vehicles, and hurricanes.

During recent years analysis tools for trajectory data have been developed within the areas of GIScience and algorithms. Analysis objectives include clustering, performing similarity analysis, segmenting a trajectory into characteristic sub-trajectories, finding patterns like flocking, and several others. Since these computations are mostly spatial, algorithmic



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

Geometric and Graph-based Approaches to Collective Motion, *Dagstuhl Reports*, Vol. 6, Issue 1, pp. 55–68
Editors: Giuseppe F. Italiano, Marc van Kreveld, Bettina Speckmann, and Guy Theraulaz



DAGSTUHL
REPORTS

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

solutions have been developed in the areas of computational geometry and GIScience. Although trajectories store only the location of a single point of reference on a moving entity, this is acceptable for the common large-scale analysis tasks. However, for the study of more complex phenomena like interaction and collective motion, it is often insufficient and the basic trajectory representation must be extended.

Simultaneously, in the area of ecology the study of motion of animals has also become a topic of increasing interest. Many animal species move in groups, with or without a specific leader. The motivation for motion can be foraging, escape from predators, changing climate, or it can be unknown. The mode of movement can be determined by social interactions, energy efficiency, possibility of discovery of resources, and of course the natural environment. The more fascinating aspects of ecology include interaction between entities and collective motion. These are harder to grasp in a formal manner, needed for modelling and automated analysis.

The seminar brought together a group of enthusiastic researchers with a diverse background. To create a shared body of knowledge the seminar featured a number of survey talks that were planned early in the week. The survey talks were rather engaging: the audience learned for instance at what scale one should look at a painting of Van Gogh, how bats chase each other, what size of clumps mussels make and why, and how to interact with a computational geometer.

Probably the main research result was a momentum started up by interaction and awareness of an exciting direction of research where a lot can still be accomplished.

More specific research accomplishments included a methodology for evaluating whether fish or other animals have their movement mostly influenced by closest neighbors, and how to reconstruct movement just based on counts at different time steps.

2 Table of Contents

Executive Summary

Giuseppe F. Italiano, Marc van Kreveld, Bettina Speckmann, and Guy Theraulaz . 55

Overview of Talks

Multidisciplinary challenges concerning self-organization in ecological systems. <i>Johan van de Koppel</i>	59
Multiscale inference in collective behaviour <i>Richard Philipp Mann</i>	59
Self-Organization in Complex Systems <i>Nicholas Ouellette</i>	60
Dynamic Graph Algorithms <i>Giuseppe F. Italiano</i>	60
Topological Data Analysis in Real Applications <i>Brittany Terese Fasy</i>	60
Analysing spatio-temporal patterns of delayed alignment interactions <i>Luca Giuggioli</i>	61
Computational Geometry for Collective Motion <i>Maarten Löffler</i>	61
Challenges for movement analytics: A GI science perspective <i>Matt Duckham</i>	61

Extra Event

DYNAMO: Dynamic Visualization of Movement and the Environment <i>Somayeh Dodge</i>	62
---	----

Working groups

Analysis on Check-in Data <i>Karl Bringmann, Oliver Burkhard, Brittany Terese Fasy, Giuseppe F. Italiano, Martin Nöllenburg, Frank Staals, Goce Trajcevski, and Carola Wenk</i>	63
Detecting Avoidance Interaction in Trajectory Data <i>Luca Giuggioli, Johan van de Koppel, Andrea Perna, Robert Weibel, and Carola Wenk</i>	63
Detecting Interactions Given Trajectory Information <i>Somayeh Dodge, Brittany Terese Fasy, Tim Ophelders, Nicholas Ouellette, and Kevin Verbeek</i>	64
Distinguishing Real and Artifactual Social Interactions <i>Martin Beye, Oliver Burkhard, Brittany Terese Fasy, Richard Philipp Mann, Bettina Speckmann, and Kevin Verbeek</i>	65
Identifying Influential Neighbors in Animal Flocking <i>Martin Beye, Anael Engel, Ramon Escobedo, Luca Giuggioli, Marc van Kreveld, Andrea Perna, Frank Staals, Guy Theraulaz, and Goce Trajcevski</i>	65
Mussel Bed Connectivity and its Influence on Survival <i>Johan van de Koppel, Maarten Löffler, and Tim Ophelders</i>	66

Schedule 67

Participants 68

3 Overview of Talks

3.1 Multidisciplinary challenges concerning self-organization in ecological systems.

Johan van de Koppel (Royal Netherlands Inst. for Sea Research – Yerseke, NL)

License  Creative Commons BY 3.0 Unported license
© Johan van de Koppel

I discussed the emergence of theory of spatial self-organization in ecology, focussing on the formation of regular spatial patterns. Past research on for instance arid systems or mudflats have highlighted Turing's activator-inhibitor principle to explain pattern formation. I have highlighted a different form of self-organization in mussel beds, where mussels aggregate to form regular patterns. This form of pattern formation follows a different mechanism that is similar to the physical process of phase-separation, as formulated by Cahn and Hilliard in 1958. Individual-based models of pattern formation in mussel beds indicate that patterns have a important effect on ecosystem functioning, increasing mussel bed resilience.

I finished with highlighting a number of outstanding challenges in the field of spatial self-organization of ecosystems:

- Understanding critical transitions: How to distinguish ecosystems with and without tipping points?
- Can we determine the process driving self-organization from the observed patterns?
- How do patterns affect ecosystem functioning?
- Can we find self-organization in “everyday”, human modified ecosystems?
- How do organisms adapt to/in self-organizing systems?
- How can we best translate individual behavior to population dynamics?

Technical issues

- Multidisciplinary – bridging the culture gap between physics, computer sciences, mathematics, and biology.
- How do we overcome scale differences in ecosystems?
- Communication – how to explain our results to the general public?

3.2 Multiscale inference in collective behaviour

Richard Philipp Mann (University of Leeds, GB)

License  Creative Commons BY 3.0 Unported license
© Richard Philipp Mann

How do groups of animals act cohesively and make collective decisions? How do complex patterns of collective motion emerge in groups of individually simple organisms? Simulation studies show that simple interaction rules between individuals and their local neighbours are sufficient to produce complex group behaviours, but the specific nature of these interactions is often unclear. Large scale group behaviours often fall into generic ‘universality’ classes such as spinning balls or polarised flocks, obscuring the precise interactions at the individual level.

In this talk I demonstrate a technique of theory-driven model comparison based on individual agent motions alongside group level observations. I phrase alternative hypothesised interactions as models which predict the behaviour of individuals, using a Bayesian model

comparison to select between competing theories, and combine this with model simulations to detect emergent effects. I show examples of this approach in the context of the collective motion of glass prawns and decision-making in damselfish, and discuss how to take this method forward.

3.3 Self-Organization in Complex Systems

Nicholas Ouellette (Stanford University, US)

License  Creative Commons BY 3.0 Unported license
© Nicholas Ouellette

Complex systems – that is, systems that consist of many simple but coupled degrees of freedom – generically and spontaneously form structure. Over the past several decades, this process of self-organization has been identified as driving the formation of patterns and structure at nearly every scale in nature. Here, I will give a brief overview of some of the key results that have come out of the study of self-organization from a physicist’s perspective. I will then connect these ideas to the study of collective behavior in animals, as well as outlining some caveats. Finally, I will pose some questions that deserve future study.

3.4 Dynamic Graph Algorithms

Giuseppe F. Italiano (University of Rome “Tor Vergata”, IT)

License  Creative Commons BY 3.0 Unported license
© Giuseppe F. Italiano

In my talk, I will survey dynamic graph algorithms. In particular, I will consider a fundamental problem in this area: the dynamic maintenance of shortest paths. Although research on this problem spans over almost 50 years, progress has been achieved only recently through the introduction of many novel algorithmic techniques. I will make a special effort to abstract some basic combinatorial properties that are at the base of some of those techniques. This will help presenting some of the most efficient algorithms in a unifying framework so that they can be better understood and deployed also by non-specialists.

3.5 Topological Data Analysis in Real Applications

Brittany Terese Fasy (Montana State University – Bozeman, US)

License  Creative Commons BY 3.0 Unported license
© Brittany Terese Fasy

Persistent homology is a method for probing topological properties of point clouds and functions. The mathematical concepts stem from Morse theory, but the use of topology in data analysis is fairly recent. In this talk, we draw an analogy between looking at homology at different parameter values and at a painting at different distances. These parameters (distances) give different insights as the values change. Persistence tells us which of these insights lasts through large intervals of the parameter. For example, the parameter can be time and the persistence may explain the dynamics of animals moving in a collective motion. After giving an intuition for persistent homology, we briefly explain how it can be used to describe, compare and analyze data.

3.6 Analysing spatio-temporal patterns of delayed alignment interactions

Luca Giuggioli (University of Bristol, GB)

License  Creative Commons BY 3.0 Unported license
© Luca Giuggioli

Animal coordinated movement interactions are commonly explained by assuming unspecified social forces of attraction, repulsion and alignment with parameters drawn from observed movement data. Very little is done to connect the sensory ecology with the movement and interaction of the individuals. On trajectories of two interacting trawling bats flying over a water pond we have shown how to extract delays with which individuals respond by copying each other's heading. Using that information it is possible to reconstruct the echolocation field strength and directionality of the bats [1].

References

- 1 L. Giuggioli, T.J. McKetterick & M. Holderied, Delayed response and biosonar perception explain movement coordination in trawling bats, *PLoS Comput. Biol.* 1(3):e1004089 (2015)

3.7 Computational Geometry for Collective Motion

Maarten Löffler (Utrecht University, NL)

License  Creative Commons BY 3.0 Unported license
© Maarten Löffler

The main challenge in applying techniques from computational geometry to solve real-world problems in collective motion analysis is deciding on the correct mathematical model. A mathematical model, or problem description, in this context, is a precise description of the input and output of a problem. In particular, given the problem description and an input, it must be clear (unambiguous) what the resulting output is.

On the one hand, a problem description must then be validated in the application. Both the input and output can be interpreted in collective motion – for instance biological or ecological – terms, and the ability to convert the input to the output has a clear value.

On the other hand, given a problem description, we can design efficient and provably correct algorithms to arrive at the desired output. For this, we leverage 50 years of research in geometric algorithms, combined with new solutions that are tailored to the unique challenges in the problem at hand.

3.8 Challenges for movement analytics: A GI science perspective

Matt Duckham (RMIT University – Melbourne, AU)

License  Creative Commons BY 3.0 Unported license
© Matt Duckham

This talk examined three of the key challenges and trends in movement analytics, from the perspective of the field of GI science. The first challenge concerns the structure of movement data. Whilst there has in the past been a strong focus on the Lagrangian trajectory data view of movement, Eulerian checkpoint or “cordon-structured” data is often

under researched. A wide range of familiar and voluminous data sources, including social media check-ins, electronic tolling, public transport smart cards, generate data structured as Eulerian checkpoints, rather than Lagrangian trajectories. The second challenge concerns the integration of information about the drivers of movement into movement analytics. In particular, the context and causes of movement are often of overriding importance in understanding movement. The third challenge relates to the interaction between moving objects in the production of collectives. These collectives are more than the sum of their parts, and so cannot be adequately captured purely by looking at the movement of individuals in the collective. Underlying these challenges is the maxim that there is “more to movement than geometry”, and a comprehensive approach to movement analytics must incorporate non-geometric movement structure, the context and causes of movement, and the role of collectives.

4 Extra Event

4.1 DYNAMO: Dynamic Visualization of Movement and the Environment

Somayeh Dodge (University of Colorado – Colorado Springs, US)

License © Creative Commons BY 3.0 Unported license
© Somayeh Dodge

Joint work of Somayeh Dodge, Glenn Xavier

URL <http://dynamovis.com>

Movement is highly influenced by its embedding spatiotemporal context, a geographic context that changes over time, such as the ambient environment, terrain, and landscape. In essence, movement occurs in both spatiotemporal space and a multidimensional attribute space (i.e. environmental and geographic context of movement). The syntheses of these two spaces need new tools suitable for dynamic visualization of the traversal through these dimensions. These tools can play a major role as a fundamental component of spatiotemporal computing systems for the comprehension and understanding of complex spatiotemporal processes and patterns of movement. This presentation provides an overview of a visualization tool, called “DYNAMO: Dynamic Visualization of Movement and the Environment” (<http://dynamovis.com>), developed for the exploratory analysis of movement in relation to the environment and geographic context. DYNAMO applies visual variables such as point and line width, color, and directional vector to visualize and animate movement tracks in their attribute space (e.g. movement parameters and context attributes).

5 Working groups

5.1 Analysis on Check-in Data

Karl Bringmann (MPI für Informatik – Saarbrücken, DE), Oliver Burkhard (Universität Zürich, CH), Brittany Terese Fasy (Montana State University – Bozeman, US), Giuseppe F. Italiano (University of Rome “Tor Vergata”, IT), Martin Nöllenburg (TU Wien, AT), Frank Staals (Aarhus University, DK), Goce Trajcevski (Northwestern University – Evanston, US), and Carola Wenk (Tulane University, US)

License © Creative Commons BY 3.0 Unported license
 © Karl Bringmann, Oliver Burkhard, Brittany Terese Fasy, Giuseppe F. Italiano, Martin Nöllenburg, Frank Staals, Goce Trajcevski, and Carola Wenk

Most current work in trajectory and movement analysis assumes that the input trajectories are precise enough to capture the exact movement of the entities, that is, it is assumed that the trajectories are either (piecewise linear) functions mapping time to a location, or a dense sample of (time,location) pairs from such a function. A large amount of such trajectory data is available. However, systems such as cell towers, wireless sensor networks, electronic travel cards (e.g. Oyster card, ov-chipkaart), and social networks such as Foursquare generate much sparser trajectory data. They capture the location of an entity only at very few, and often wide spread, times. This means that the usual assumption that the entity moves linearly in between two trajectory vertices does not make sense, and thus the traditional algorithms and analysis techniques are not applicable. We refer to such sparse trajectory data as *check-in data*, and identify several interesting questions and analysis tasks for check-in data.

We assume that the entities move in a network, which we model as a (planar) directed weighted graph, where the edge weights model travel time. Some of the edges are equipped with a *beacon* that registers when an entity traverses the edge. Beacons register the time at which an entity starts traversing the edge, and the identity of the entity. For every moving entity, we thus obtain a (*sparse*) *trajectory*, a sequence of (time,edge) pairs. Note that not every edge is equipped with a beacon, hence the trajectory of an entity will, in general, not be known completely. We are then interested in the following problem:

Given all sparse trajectories, an edge e of the graph, and a time interval I , compute how many entities start traversing edge e during I .

We identified and formalized several subproblems toward solving the above problem. Furthermore, we discussed variations of the model, e.g. incorporating uncertainty, delay, placing beacons at vertices rather than at edges etc. We aim to solve these problems in the near future.

5.2 Detecting Avoidance Interaction in Trajectory Data

Luca Giuggioli (University of Bristol, GB), Johan van de Koppel (Royal Netherlands Institute for Sea Research – Yerseke, NL), Andrea Perna (Paris Diderot University, FR), Robert Weibel (Universität Zürich, CH), and Carola Wenk (Tulane University, US)

License © Creative Commons BY 3.0 Unported license
 © Luca Giuggioli, Johan van de Koppel, Andrea Perna, Robert Weibel, and Carola Wenk

The purpose of the working group was to explore ideas and methods to detect avoidance behaviour by looking at trajectories of an habituated population of dwarf mongoose in the wild. The data consists of gps tracking at approximately 30 second sampling based on each

of the location on the researchers following simultaneously different groups of mongooses of 5–7 individuals. Biologists expect a scent-mediated interaction to occur as individuals detect fresh scent left by other groups and move away to avoid costly confrontations. As we played with the dataset at our disposal, we realized that the locations of individuals were potentially too far apart for scent to be detected and no-delayed crossing of the trajectories appeared. We hypothesized that individuals may respond instantaneously through visual detection of other groups from a distance. By looking at the vegetation in South Africa where data were collected we saw that trees and high bushes may constrain the distance at which such detection occurs. We thus concluded that sentinels on trees, usually used for detecting predators, may also be involved in identifying the presence of other mongoose groups.

5.3 Detecting Interactions Given Trajectory Information

Somayeh Dodge (University of Colorado – Colorado Springs, US), Brittany Terese Fasy (Montana State University – Bozeman, US), Tim Ophelders (TU Eindhoven, NL), Nicholas Ouellette (Stanford University, US), and Kevin Verbeek (TU Eindhoven, NL)

License  Creative Commons BY 3.0 Unported license
 © Somayeh Dodge, Brittany Terese Fasy, Tim Ophelders, Nicholas Ouellette, and Kevin Verbeek

Our group was tasked with developing strategies to answer a question that seems straightforward but that in practice is quite difficult: given trajectory information from one or more individuals, can we identify where and when interactions occurred?

This question is particularly difficult given that in general we do not know what the signature of an interaction is; and indeed, it will likely be different for different specific problems. We therefore modified the question to something we felt was more tractable. If we make the hypothesis that interactions result in some (possibly transient) change in behavior, then identifying these interaction events becomes a problem of segmenting the trajectory. This approach will miss interactions that do not result in a measurable change in behavior; such events, however, will likely never be detectable without additional information.

We had some discussion as to whether considering single trajectories is sufficient for this problem, or whether we must look at explicitly pairwise or higher-order quantities. Although we did not reach a general consensus on this point, we settled on considering only single trajectories initially, as in general the signature of interactions should be reflected in single trajectories as well as in pairs.

The problem is thus one of trajectory segmentation into “normal” and “unusual” parts. How to do this in a meaningful, objective way, however, is not obvious. We considered approaches based on simple global statistics, such as segmenting based on the mean speed or the velocity variance, but found these to be unsatisfying. Instead, we proposed a kind of “auto-segmentation”, which we expressed as an optimization problem. The ideal segmentation for our purposes would cut the trajectories into pieces that are as different as possible. That is, if we cut a trajectory into segments of type A and B, the information content in the A and B segments should be as different as possible. To accomplish this, we thought of taking an iterative approach. Suppose we begin with some initial, arbitrary segmentation. We can then calculate the statistics of the A and B segments separately, and define a cost function based on the difference between these statistics (the mutual information, for example, or the Kolmogorov distance between their PDFs). By then optimizing this cost function, one could come up with an objective segmentation, which could then hopefully be interpreted.

5.4 Distinguishing Real and Artifactual Social Interactions

Martin Beye (Universität Düsseldorf, DE), Oliver Burkhard (Universität Zürich, CH), Brittany Terese Fasy (Montana State University – Bozeman, US), Richard Philipp Mann (University of Leeds, GB), Bettina Speckmann (TU Eindhoven, NL), and Kevin Verbeek (TU Eindhoven, NL)

License © Creative Commons BY 3.0 Unported license
© Martin Beye, Oliver Burkhard, Brittany Terese Fasy, Richard Philipp Mann, Bettina Speckmann, and Kevin Verbeek

We started by considering whether one could use purely trajectory data from groups (e.g. pairs) of animals to decide whether a social interaction took place, as opposed to correlated responses to an environmental stimulus.

In general the problem appears insoluble, since an arbitrary unseen stimulus could exactly mimic another animal. Therefore in principle the two could not be distinguished.

However, if we have some idea what the external stimuli might look like we can either:

- Do experiments to determine response to these, and look for differences in later data. This is the approach taken by Gautrais et al. in PLoS Comput. Biol. 2012.
- Try to theorise what a response might look like, e.g. a fleeing response to a point source might be spherically symmetric around that point

We considered whether there would be differences in the noise structure between two individuals following each other versus two individuals following a common route, since the noise from one would become part of the signal the other follows. This might lead to an entangled noise signature that would be indicative of interactions.

Then we moved onto larger scale observations of collectives and whether we could assess whether an aggregation was due to self organisation rather than an environmental cue. This led us onto a detailed discussion of to what extent persistent homology methods could be used to classify different groups collective behaviour as being distinct from each other. Unfortunately it seems necessary to be quite sure what types of structure you want to pick out in the combined trajectories before designing a persistent homology metric to find them.

5.5 Identifying Influential Neighbors in Animal Flocking

Martin Beye (Universität Düsseldorf, DE), Anael Engel (The Hebrew University of Jerusalem, IL), Ramon Escobedo (Université Paul Sabatier – Toulouse, FR), Luca Giuggioli (University of Bristol, GB), Marc van Kreveld (Utrecht University, NL), Andrea Perna (Paris Diderot University, FR), Frank Staals (Aarhus University, DK), Guy Theraulaz (CNRS and Université Paul Sabatier – Toulouse, FR), and Goce Trajcevski (Northwestern University – Evanston, US)

License © Creative Commons BY 3.0 Unported license
© Martin Beye, Anael Engel, Ramon Escobedo, Luca Giuggioli, Marc van Kreveld, Andrea Perna, Frank Staals, Guy Theraulaz, and Goce Trajcevski

One important issue that is largely discussed in the community dealing with collective motion in animal groups concerns the number of neighbors each individual in a flock of birds or a school of fish is interacting with. Indeed, it has been shown that the properties that emerge at the level of a flock or a school largely depend on the number and position of neighbors each individual is paying attention to. Is it possible to analyze the trajectories of individuals moving in groups in such a way to get access to this information?

To identify the influential neighbors, we propose first to detect the correlation between the velocity changes of a focal individual and the corresponding velocity changes of individuals moving in its close vicinity. Obvious choices are to compute: (1) the normalized cross-velocity correlation (difference in headings) and (2) the non-normalized cross-velocity correlation. In addition one has to take into account the fact that the focal individual will also react with a certain time delay. One good method to extract the values of delays as a function of time from one or all of the correlation plots would be to use the pairwise method developed by Giuggioli et al. (2015). The problem with that method is that it is only pairwise, and in some instances of time the extracted delays have inconsistencies (the order is not maintained if one looks at all pairwise possibilities). Another method recently developed by Kevin Buchin (Konzack et al., 2015) that we also intend to use is in principle capable of extracting delays for N animals without these inconsistencies.

Then, we will analyze the time-average correlation value between a focal individual as a function of either the distance or the number of neighbors. This analysis should also distinguish individuals that are located either in the centre or in the periphery of the flock. The results can then be used to get histograms of the maximum correlation values as a function of the k -nearest neighbor or distance for each moment of time. If each individual pay attention to only a limited number of individuals within their perception field, one can expect that it can strongly affect the distribution of correlation values on these histograms. However one cannot know if a given individual focuses its attention on only one of its close neighbors at a time or if it responds to some “average information”. In order to test this hypothesis, one can perform the same analysis as the one described before, but with some average quantity associated with the (linear or non-linear) combinations of neighbors, e.g. the simplest one is the average.

In order to test our method, we will use controlled simulations of a model of collective movement with known rules (i.e. how many influential neighbors, if they influence by rank or by distance) and use our method to test if it can detect and distinguish between the different rules. In addition we will perform this analysis on trajectory data on groups of fish, moving in an annular arena and we will only focus on spontaneous U-turn events to minimize the effects of the constraining geometry.

References

- 1 Konzack, M., McKetterick, T., Wilcox, G. Buchin, M., Giuggioli, L., Gudmundsson, J., Westenberg, M.A., Buchin, K. *Analyzing Delays in Trajectories*. pp. 93–97 in Proc. of the IEEE Pacific Visualization Symp. 2015, pp. 14–17, Hangzhou, China, 2015
- 2 Giuggioli, L., McKetterick, T.J., Holderied, M. *Delayed Response and Biosonar Perception Explain Movement Coordination in Trawling Bats*. *LoS Comput. Biol.* 11(3):e1004089, 2015.

5.6 Mussel Bed Connectivity and its Influence on Survival

Johan van de Koppel (Royal Netherlands Inst. for Sea Research – Yerseke, NL), Maarten Löffler (Utrecht University, NL), and Tim Ophelders (TU Eindhoven, NL)

License  Creative Commons BY 3.0 Unported license
© Johan van de Koppel, Maarten Löffler, and Tim Ophelders

We want to understand how the spatial structure of a mussel bed influences the survival or persistence of groups of mussels. It seems that net-shaped structures may provide a more stable landscape, with less vulnerability to waves. Models that account for the effect of such grouping structure on mussel survival may provide a better understanding of 1) self-organization in mussels, and 2) stability of mussel beds as a key habitat to many species.

Waves put force on a limited section of the bed (say 25x25 cm). Small clumps of mussels get dislodged easily, while larger clumps that are connected to the larger bed, are not dislodged. A binary test is needed, that checks which of the mussels are sufficiently connected not to break free after a wave impact.

Given a geometric graph of mussels and a disk representing the wave impact zone, we define a score $F(M)$ for each set of mussels M . If $F(M)$ is above a certain threshold H , the set M gets dislodged. In order to define this $F(M)$, we are going to count three things:

1. C The number of connections between M and the rest of the mussels;
2. I The number of mussels of M that are inside the disk;
3. O The number of mussels of M that are outside the disk.

In addition, we define three weights:

1. W_C The strength of the glue between connected mussels;
2. W_I The force applied by a wave to the mussels inside the disk;
3. W_O The force required to move a mussel that is outside the disk.

We define $F(M) = W_C C(M) - W_I I(M) + W_O O(M)$, and take the minimum of $F(M)$ over all sets M to compare with our threshold H .

If the mussel connections are directed, we can model this as a minimum closure problem. Minimum closures can be computed in quadratic time using an approach based on MinCut. One way to get a directed graph is to direct all edges to the center of the wave impact zone. It is likely that undirected graphs can be handled by a similar approach.

6 Schedule

Monday

- 09.00–10.30: Survey lectures: Johan van de Koppel, Richard Mann
- 10.50–12.00: Quick introductions of participants, Dagstuhl explanations
- 14.00–15.30: Survey lectures: Nicholas Ouellette, Giuseppe Italiano
- 16.00–18.00: Open problems + break-out

Tuesday

- 09.00–10.15: Survey lectures: Brittany Therese Fasy, Luca Giuggioli
- 10.40–12.00: Continue break-out
- 13.00–15.30: Personal discussions with colleagues at Dagstuhl on collective motion
- 16.00–18.00: Continue break-out

Wednesday

- 09.00–10.15: Survey lectures: Maarten Löffler, Matt Duckham
- 10.40–12.00: Reporting back from break-out, discussion
- Afternoon excursion

Thursday

- 09.00–10.15: New open problems and groups
- 10.40–12.00: Second break-out set
- 14.00–15.30: Break-out
- 16.00–17.30: Break-out
- 17.30–18.00: Extra event

Friday

- 09.00–10.15: Reporting from second break-out
- 10.40–12.00: Future plans of research in collective motion

Participants

- Martin Beye
Universität Düsseldorf, DE
- Karl Bringmann
MPI für Informatik –
Saarbrücken, DE
- Kevin Buchin
TU Eindhoven, NL
- Maike Buchin
Ruhr-Universität Bochum, DE
- Oliver Burkhard
Universität Zürich, CH
- Somayeh Dodge
University of Colorado –
Colorado Springs, US
- Matt Duckham
RMIT Univ. – Melbourne, AU
- Anael Engel
The Hebrew University of
Jerusalem, IL
- Ramon Escobedo
Université Paul Sabatier –
Toulouse, FR
- Brittany Terese Fasy
Montana State University –
Bozeman, US
- Sándor Fekete
TU Braunschweig, DE
- Luca Giuggioli
University of Bristol, GB
- Giuseppe F. Italiano
University of Rome “Tor
Vergata”, IT
- Maarten Löffler
Utrecht University, NL
- Richard Philipp Mann
University of Leeds, GB
- Martin Nöllenburg
TU Wien, AT
- Tim Ophelders
TU Eindhoven, NL
- Nicholas Ouellette
Stanford University, US
- Andrea Perna
Paris Diderot University, FR
- Bettina Speckmann
TU Eindhoven, NL
- Frank Staals
Aarhus University, DK
- Guy Theraulaz
CNRS and Université Paul
Sabatier – Toulouse, FR
- Goce Trajcevski
Northwestern University –
Evanston, US
- Johan van de Koppel
Royal Netherlands Institute for
Sea Research – Yerseke, NL
- Marc van Kreveld
Utrecht University, NL
- Kevin Verbeek
TU Eindhoven, NL
- Robert Weibel
Universität Zürich, CH
- Carola Wenk
Tulane University, US



Report from Dagstuhl Seminar 16031

Well Quasi-Orders in Computer Science

Edited by

Jean Goubault-Larrecq¹, Monika Seisenberger², Victor Selivanov³,
and Andreas Weiermann⁴

1 ENS – Cachan, FR, goubault@lsv.ens-cachan.fr

2 Swansea University, GB, m.seisenberger@swansea.ac.uk

3 A. P. Ershov Institute – Novosibirsk, RU, vseliv@iis.nsk.su

4 Ghent University, BE, andreas.weiermann@ugent.be

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16031 “Well Quasi-Orders in Computer Science”, the first seminar devoted to the multiple and deep interactions between the theory of Well quasi-orders (known as the Wqo-Theory) and several fields of Computer Science (Verification and Termination of Infinite-State Systems, Automata and Formal Languages, Term Rewriting and Proof Theory, topological complexity of computational problems on continuous functions). Wqo-Theory is a highly developed part of Combinatorics with ever-growing number of applications in Mathematics and Computer Science, and Well quasi-orders are going to become an important unifying concept of Theoretical Computer Science. In this seminar, we brought together several communities from Computer Science and Mathematics in order to facilitate the knowledge transfer between Mathematicians and Computer Scientists as well as between established and younger researchers and thus to push forward the interaction between Wqo-Theory and Computer Science.

Seminar January 17–22, 2016 – <http://www.dagstuhl.de/16031>

1998 ACM Subject Classification B.2 Arithmetic and Logic Structures, F.3.1 Specifying and Verifying and Reasoning about Programs, F.4.1 Mathematical Logic

Keywords and phrases Better quasi-order, Well quasi-order, Hierarchy, Infinite State Machines, Logic, Noetherian space, Reducibility, Termination, Topological Complexity, Verification, Well-Structured Transition Systems

Digital Object Identifier 10.4230/DagRep.6.1.69

Edited in cooperation with Simon Halfon

1 Executive Summary

Jean Goubault-Larrecq

Monika Seisenberger

Victor Selivanov

Andreas Weiermann

License © Creative Commons BY 3.0 Unported license

© Jean Goubault-Larrecq, Monika Seisenberger, Victor Selivanov, and Andreas Weiermann

Computer Science, being a huge and complex conglomerate of theoretical disciplines, technological advances and social methodologies, strongly needs unifying concepts and techniques. In particular, relevant mathematical concepts and theories are required. The notion of well quasi-order (or almost-full relation, if transitivity is not required – a notion preferred by some authors) was discovered independently by several mathematicians in the 1950-s and



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

Well Quasi-Orders in Computer Science, *Dagstuhl Reports*, Vol. 6, Issue 1, pp. 69–98

Editors: Jean Goubault-Larrecq, Monika Seisenberger, Victor Selivanov, and Andreas Weiermann



Dagstuhl Reports

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

quickly evolved to a deep theory with many applications and remarkable results. Soon afterwards, well and better quasi-orders started to appear more and more frequently in different parts of theoretical computer science such as automata theory, term rewriting, verification of infinite-state systems, computations with infinite data, and others. Accordingly, an increasing number of researchers from different fields of computer science use notions and methods of Wqo-Theory. Therefore, it seemed to be the right time to have a broad discussion on how to speedup this process and to better understand the role of well quasi-orders in theoretical computer science.

Topics of the seminar

During this seminar we concentrated on the following four topics:

1. Logic and proofs
2. Automata and formal languages
3. Topological issues
4. Verification and termination problems

Logic and proofs

Well quasi-orders, originally introduced in algebra, soon played an important role in proof theory: Higman's Lemma and Kruskal's Theorem are examples of theorems that are not provable in Peano Arithmetic. Determining the proof-theoretic strength of these (types of) theorems, as well as classifying them in terms of Reverse Mathematics, constituted an important endeavor. The concept of a WQO naturally extends to the more complex concept of a better quasi-order (BQO) which deals with infinite structures. Again, the proof theoretic strength of theorems on BQOs has been/must be investigated, and the theorems themselves can be used for more sophisticated termination problems. One of the open challenges is the strength of Fraïssé's order type conjecture. Non-constructive proofs of this type of theorems (on WQOs) include proofs using the so-called minimal-bad-sequence argument. Investigating their strengths and also their computational content, via Friedman's A-translation or Gödel's Dialectica Interpretation, has led to interesting results. To optimize these techniques so that realistic programs can be extracted from these classical proofs, using bar recursion, update recursion, selection functions, etc., is ongoing work.

Automata and formal languages

Well quasi-orders have many-fold connections to automata theory and formal language theory. In particular, there are nice characterizations of regular and context-free languages in terms of well quasi-orders, some lower levels of the concatenation hierarchies admit characterizations in terms of the subword relation and its relatives. Such characterizations sometimes help in getting new results, say on decidability of some levels of the concatenation hierarchy (Glasser, Schmitz, Selivanov). The same applies to ω -languages, though in this case the relationships are less investigated.

On the topological level, it is known that Wadge reducibility (or reducibility by functions on ω -words computable by finite automata) are well quasi-orders on the class of ω -regular finite partitions of the Cantor space. Using some variants of the Kruskal theorem on quasi-orderings of labeled trees, Selivanov was able to completely characterize the corresponding

partial order, obtaining thus a complete extension of the Wagner hierarchy from sets of finite partitions.

The mentioned relationships between Wqo-theory and formal languages are currently not well systematized, and many natural questions remain open. Further insights in this topic is essential for the development of this field.

Topological issues

An important task in computing with infinite data is to distinguish between computable and non-computable functions and, in the latter case, to measure the degree of non-computability. Usually, functions are non-computable since they are not even continuous, hence a somewhat easier and more principal task is in fact to understand the degree of discontinuity of functions. This is achieved by defining appropriate hierarchies and reducibility relations.

In classical descriptive set theory, along with the well-known hierarchies, Wadge introduced and studied an important reducibility relation on subsets of the Baire space. As shown by van Engelen et al, von Stein, Weihrauch and Hertling, this reducibility of subsets of topological spaces can be generalized in various ways to a reducibility of functions on a topological space. In this way, the degrees of discontinuity of several important computational problems were classified. The transfer from sets to functions requires some notions and results of Wqo-theory in order to define and study hierarchies and reducibilities arising in this way.

Verification and termination problems

WQOs made their debut in computer science when Don Knuth suggested that Kruskal's Theorem might find an application in proving termination of programs. This was achieved a few years later by Nachum Dershowitz and the advent of recursive path orderings. Today, it is probably the area of software verification that provides the largest number of applications of WQOs in computer science. The decidability of coverability for well-structured transition systems (WSTS) crucially relies on the very properties of well quasi-orders. WSTS include Petri nets and their extensions, and more generally affine nets. They also include lossy channel systems, weak memory models, various process algebras, data nets, certain abstractions of timed Petri nets, and certain parametrized transition systems. The verification of new classes of transition systems prompts for new classes of WQOs. In addition to this, understanding the computational complexity of the resulting verification algorithms requires a finer analysis of minimal-bad-sequence arguments and their relation to hierarchies of recursive functions (Hardy, fast growing, etc.)

Report

One of the central purposes of the proposed seminar was to bring together researchers from Wqo-theory and those from the related areas of computer science who actively apply notions and techniques of Wqo-theory. We wanted thus to encourage more interaction between the different communities, leading finally to a significant development of the mentioned fields. Overall, the seminar was very stimulating. The initial concerns that the four topics of the seminar might remain separate was quickly brushed off, as verification talks relied on concepts from logic (e.g., maximal order types), topological issues resonated with verification (e.g., Noetherian spaces), and all participated actively in vivid sessions of setting up and discussing open questions.

To facilitate the interaction, each of the four topics of the seminar started with one or two introductory talks. In the topic **Logic and Proofs**, Diana Schmidt gave the first talk of the seminar “Ordinal notations, the maximal ordertypes of Kruskal’s Theorem and a tale of two cultures”, where she summarized work that started 40 years ago, but is still of interest to date. Andreas Weiermann complemented the introduction by addressing open problems concerning the so-called phase transitions (a topic which was then further elaborated by Lev Gordeev) in proof theory and the calculation of maximal ordertypes. A number of contributions focused on the formalization of Kruskal type theorems possibly including the extraction of computational content, leading to various practical exchange sessions in the evenings (see also next paragraph). The topic was continued on Tuesday by Alberto Marcone who gave a survey on WQOs and BQOs in Reverse Mathematics, followed various talks and discussions on open questions around better quasi-orderings. Another highlight of the second day were two contributions on the Graph Minor Theorem: Chun-Hung Liu reported on his recent proof of a conjecture by Robertson involving topological minors, and Michael Rathjen looked at the Graph Minor Theorem to determine its proof theoretic strength. Of the further contributions we want to specifically mention Julia Knight who made the connection between WQOs and Hahn-Fields.

The topic **Verification** was addressed by Philippe Schnoebelen in “Well quasi-orderings and program verification” on Monday afternoon, with a gentle introduction to the field of well-structured transition systems (WSTS). Although this was an introductory talk, he took the opportunity to introduce the class of priority channel systems as an example, whose decidability results rely on a form of well quasi-ordering with gap embedding. Such orderings require very high maximal order types, and led to the open question of giving a formula, or even lower and upper bounds, for those maximal order types. This has direct consequences on the complexity of verification. Alain Finkel proceeded to give another talk on the verification of WSTS, “Decidability results on infinitely branching WSTS”, which completed the introduction given by Philippe Schnoebelen, and quickly proceeded to explore the challenges of verifying WSTS that are infinitely branching. He explained that deciding termination, coverability, and boundedness could be done through computations that involve a so-called completion of the WSTS, generalizing the well-known Karp-Miller construction for Petri nets – some of these problems turn out to be decidable, some others not. Crucially, in the decidable cases, even if the WSTS is infinitely branching, its completion is always finitely branching. Maurice Pouzet made the observation that the key result used there, that every downwards closed set of a WQO is a union of finitely many ideals, is due to Kabil and himself in 1992, and that this was still true for the more general class of FAC (finite antichain) orders. FAC orders were the subject of the next presentation, by Mirna Džamonja, entitled “On the width of FAC orders”. She started from the fact that, while height, length and width of well quasi-orders are important notions, width generalizes to all FAC orders, and that we can compute the width of FAC orders defined from FAC constituents. One consequence is that for every ordinal α , there is a WQO of width exactly α . The main open question is to be able to compute the width of a finite product of FAC orders. The width of the product of two ordinals is known, but is given by a complex formula. Mirna Džamonja gave some new results on the question, in particular for some three-way products.

The final two Monday contributions were concerned with very different questions, namely mechanized proofs of Higman’s Lemma, one of the core results in well quasi-order theory. Christian Sternagel described a proof of Higman’s Lemma by so-called open induction, a concept akin to Scott induction in domain theory, as one of the participants noticed. The proof is a considerably simplified version of a proof of the same kind given by Alfons

Geser. Open induction was again the subject of Thomas Powell’s talk “Open induction and the Dialectica interpretation” on Friday morning. Helmut Schwichtenberg concluded the afternoon of Monday, just before a rump session dedicated to open questions. He gave a talk “on the computational content of Higman’s Lemma”, stressing that different constructive proofs of the same theorem yield programs – by the Curry-Howard correspondence – that behave differently in practice. This resonated with the last morning talk of the same day, “An axiom-free Coq proof of Kruskal’s Theorem”, by Dominique Larchey-Wendling, on a recent constructive proof of the much more complex Kruskal theorem, inspired from and generalizing a proof due to Wim Veldman.

Other verification-oriented talks were given by Mizuhito Ogawa on Thursday morning, “Notes on regularity and WQOs, and well-structured pushdown systems”, which gave new decidability results for coverability on extended forms of transition systems, through the use of so-called P-automata. In the evening, Roland Meyer explained how one can encode certain depth-bounded, breadth-bounded, and name-bounded processes of the pi-calculus into well-structured transition systems, and obtain decidability results through acceleration techniques, akin (again) to the Karp-Miller technique already mentioned above. Sylvain Schmitz provided us with a glimpse of the new complexity classes that had to be invented in recent years to characterize the complexity of the standard decidable problems for classical WSTS. These are classes of very high complexity: Ackermannian, hyper-Ackermannian, and others. This was a talk that unified the concerns of verification with the logical view, based on maximal ordertypes, and introduced by Andreas Weiermann and others. This provided a natural link with the last talk in the verification strand, “Trace universality for VASS”, given by Simon Halfon, who explained what the problem was, that it relied on the fact that all bad sequences of finite subsets of d -tuples of natural numbers are finite, that it is Ackermann-complete for $d = 1$, and he then proceeded to explain what was known in the cases $d \geq 2$, on which he is working.

The topic **Topological Issues** was introduced by Victor Selivanov in “Well quasi-orders and Descriptive Set Theory” on Tuesday afternoon, and by Jean Goubault-Larrecq in “A Gentle Introduction to Noetherian Spaces” on Wednesday morning. Selivanov gave a short survey of the relationships between Wqo-Theory and Descriptive Set Theory, as well as a discussion of several interesting open questions in this field. Even the definition and basics of BQOs are related to Descriptive Set Theory, as was demonstrated in the talks by Alberto Marcone and Yann Pequignot on Tuesday. Other connections were given by Raphael Carroy on Tuesday (discussing some WQOs on continuous functions), by Oleg Kudinov on Wednesday (discussing joint results with Selivanov on definability issues for some popular WQOs) and by Peter Hertling on Friday (giving an overview of results about Weihrauch reducibilities). Goubault-Larrecq introduced an important topological extension of the notion of WQO which motivates several interesting open questions. Some of these questions were addressed in the subsequent Wednesday talks by Matthew de Brecht and Arno Pauly. Both mentioned directions in the topic **Topological Issues** look very prospective and deserve additional attention.

The topic **Automata and Formal Languages** was surveyed by Mizuhito Ogawa on Thursday morning who summarized several interesting facts relating WQOs to popular classes of languages and ω -languages on words and trees (cf. also the topic **Verification**). Automata and Formal Languages were also addressed on Tuesday by Selivanov who related WQOs to his extension of the Wagner hierarchy from the case of sets to the case of k -partitions. Another related talk was given on Thursday by Willem Fouché who discussed some applications of Ramsey theory to unavoidable regularities in words. Although **Automata and Formal**

Languages turned out slightly underrepresented at this seminar, the relationship between Wqo-Theory and Formal Languages seems quite important and deserves further investigation which we hope to include in a future seminar.

Overall, our seminar attracted 44 participants (10 from Germany, 22 from other European countries, 12 from Canada, Japan, Russia, South Africa, and USA) who contributed 33 talks. In addition, we included several problem sessions where we summarized all problems mentioned in the seminar. As a result of these sessions we give a list of open problems at the end of this report. Looking at the feedback the seminar was very well received amongst the participants. Positively mentioned was that the seminar involved “people from different backgrounds” who “can still share interest”, or in other words “hearing people from different research areas discuss similar questions”, and that “one week is too short :-)”. Thoroughly enjoyed was also our two hour long walk in the snow on Wednesday afternoon. The great success of the seminar is not only due to the participants, but also to the staff in Saarbrücken and Dagstuhl, who did a splendid job in facilitating the seminar and making our stay a very pleasant one. Special thanks go to Susanne Bach-Bernhard for all the interaction related to the organization of the seminar and to Jutka Gasiorowski for her support in producing this report.

2 Table of Contents

Executive Summary

<i>Jean Goubault-Larrecq, Monika Seisenberger, Victor Selivanov, and Andreas Weiermann</i>	69
--	----

Overview of Talks

A question about bad arrays <i>Andreas R. Blass</i>	77
Ordering functions <i>Raphael Carroy</i>	77
Noetherian spaces and quantifier elimination <i>Matthew de Brecht</i>	78
Some uses of WQOs and BQOs in modal logic <i>Dick de Jongh</i>	78
Basics on (infinitely branching) WSTS <i>Alain Finkel</i>	79
On the width of FAC orders <i>Mirna Džamonja</i>	79
BQOs and where they come from <i>Thomas E Forster</i>	80
The algorithmic complexity of recognizing unavoidable regularities of words. <i>Willem L. Fouché</i>	80
On Harvey Friedman's Finite Phase Transitions <i>Lev Gordeev</i>	81
A gentle introduction to Noetherian spaces <i>Jean Goubault-Larrecq</i>	81
Trace Universality for VASS <i>Simon Halfon</i>	82
On Initial Segments of Topological Weihrauch Degrees <i>Peter Hertling</i>	83
Well quasi-orderings and Hahn fields <i>Julia Knight</i>	83
Definability in some well partial orders <i>Oleg Kudinov</i>	83
An axiom free Coq proof of Kruskal's tree theorem <i>Dominique Larchey-Wendling</i>	84
Robertson's conjecture on well quasi-ordering and topological minors <i>Chun-Hung Liu</i>	84
Wqo and Bqo Theory in Reverse Mathematics <i>Alberto Marcone</i>	85
Dimensions of Mobility <i>Roland Meyer</i>	85

Notes on regularity and well quasi-ordering <i>Mizuhito Ogawa</i>	86
Well-structured pushdown systems <i>Mizuhito Ogawa</i>	86
Noetherian Spaces in TTE <i>Arno Pauly</i>	87
From well to better: the space of ideals <i>Yann Pequignot and Raphael Carroy</i>	88
Problems on well quasi-orders and hereditary classes <i>Maurice Pouzet</i>	88
A constructive interpretation of open induction <i>Thomas Powell</i>	88
Bounds for the strength of the graph minor theorem <i>Michael Rathjen</i>	89
Ordinal notations, the maximal order types of Kruskal’s Tree Theorem, and a tale of two cultures <i>Diana Schmidt</i>	89
Complexity Classes Beyond Elementary <i>Sylvain Schmitz</i>	90
Well-quasi-orderings for program verification and computational complexity <i>Philippe Schnoebelen</i>	90
Linearization as Conservation <i>Peter M. Schuster, Davide Rinaldi, and Daniel Wessel</i>	91
Higman’s Lemma and its Computational Content <i>Helmut Schwichtenberg</i>	91
Well quasi-orders and descriptive set theory: some results and questions <i>Victor Selivanov</i>	91
A Mechanized Proof of Higman’s Lemma by Open Induction <i>Christian Sternagel</i>	92
Some Challenges Related to Wqo-Theory <i>Andreas Weiermann</i>	93
Discussion and Open Problems	93
Participants	98

3 Overview of Talks

3.1 A question about bad arrays

Andreas R. Blass (University of Michigan – Ann Arbor, US)

License  Creative Commons BY 3.0 Unported license
© Andreas R. Blass

If a quasi-ordering \leq of a set Q is not a better quasi-order, then this fact is witnessed by a continuous bad array, which means a continuous function F to Q (with the discrete topology) from the Baire space B of infinite subsets of the set \mathbb{N} of natural numbers (topologized as a subspace of the product $2^{\mathbb{N}}$ of discrete two-point spaces) such that no set X in B has $F(X) \leq F(X - \{\min(X)\})$. Identifying B with the set of paths through the tree T of finite increasing sequences of natural numbers, we have that each continuous map F from B to Q is given by a function into Q from a barrier in the tree T . The complexity of F can be measured by the ordinal height of the part of T lying between this barrier and the root. The smaller this height, the farther F is from being better quasi-ordered. In particular, the height is never 0, and it can take the value 1 if and only if Q is not even well quasi-ordered.

For any finite value h of the height, Ramsey's theorem allows one to arrange for F to have a very uniform structure, so that the order relation between $F(X)$ and $F(Y)$ depends only on the relative ordering, in \mathbb{N} , of the first h elements of X and the first h elements of Y . When $h = 2$, this uniformity makes the quasi-ordering \leq unique on the range of F ; the only such quasi-order is the Rado example.

I asked whether there is a similar uniqueness for larger finite values of h . Maurice Pouzet provided a negative answer. So I refined the question: How many such uniform examples are there, as a function of the height h ? This seems to be an open problem.

3.2 Ordering functions

Raphael Carroy (University of Torino, IT)

License  Creative Commons BY 3.0 Unported license
© Raphael Carroy

Main reference R. Carroy, "A quasi-order on continuous functions", Journal of Symbolic Logic, Vol. 78(2), pp. 633–648, 2013.)

URL <http://dx.doi.org/10.2178/jsl.7802150>

I examine a notion of reduction between functions: a function f is *continuously reducible* to a function g whenever there are two continuous functions σ and τ such that $f = \tau \circ g \circ \sigma$. This is the topological equivalent of the *strong Weihrauch reducibility*.

After briefly discussing the relevance of other quasi-orders existing in the literature, I begin to analyze continuous functions between zero-dimensional Polish spaces with respect to continuous reducibility. I prove that the identity is complete among continuous functions, and reduces to any Borel function with uncountable image.

I also prove that this well orders the family of continuous functions with compact domains. Concerning more general families of functions, I generalize the Cantor-Bendixson analysis of closed sets to continuous functions, showing that it stratifies those with countable image into countably many layers, and describing the general structure of these layers.

Finally, I apply this analysis to obtain that embeddability between closed subsets of the Baire space is a better quasi-order.

3.3 Noetherian spaces and quantifier elimination

Matthew de Brecht (NICT – Osaka, JP)

License  Creative Commons BY 3.0 Unported license
© Matthew de Brecht

Noetherian spaces are topological spaces which can be viewed as natural generalizations of well quasi-orders. They are defined as spaces whose open set lattice satisfies the ascending chain condition, or equivalently, as spaces in which every open set is compact. They are not Hausdorff in general.

We prove a simple quantifier elimination result for countably based sober Noetherian spaces. In particular, we show that if X and Y are countably based sober Noetherian spaces, and P is a boolean combination of open subsets of the product space $X \times Y$, then the projection of P onto Y is a boolean combination of open sets. This result is essentially a weak version of a well known theorem in algebraic geometry due to Chevalley concerning images of morphisms between schemes.

The result we present is purely topological, and can be viewed as an exercise to better understand Noetherian spaces. Our proof methods will use common tools from descriptive set theory, such as the Baire category theorem and the Hausdorff-Kuratowski theorem.

This work was supported by JSPS Core-to-Core Program, A. Advanced Research Networks and by JSPS KAKENHI Grant Number 15K15940.

3.4 Some uses of WQOs and BQOs in modal logic

Dick de Jongh (University of Amsterdam, NL)

License  Creative Commons BY 3.0 Unported license
© Dick de Jongh

In modal logic sometimes the class of models (or better: frames) can be seen as WQOs or BQOs. On that basis such logics can be shown to be finitely axiomatizable. I will sketch the case of the extensions of $S4.3$ and of $S5^2$. The latter is work of my student and now colleague Nick Bezhanishvili with Ian Hodkinson.

References

- 1 Nick Bezhanishvili, Ian Hodkinson. *All normal extension of $S5$ -squared are finitely axiomatizable*. *Studia Logica*, vol. 78, 443-457, 2004.

3.5 Basics on (infinitely branching) WSTS

Alain Finkel (ENS – Cachan, FR)

License © Creative Commons BY 3.0 Unported license
© Alain Finkel

Joint work of Michael Blondin, Jean Goubault-Larrecq, Pierre McKenzie, Alain Finkel

Main reference A. Finkel, J. Goubault-Larrecq, “Forward Analysis for WSTS, Part II: Complete WSTS”, Logical Methods in Computer Science, Vol. 8(3), pp. 1–35, 2012.

URL <http://arxiv.org/abs/1208.4549>

Main reference M. Blondin, A. Finkel, P. McKenzie, “Handling Infinitely Branching WSTS”, in Proc. of 41st Int’l Colloquium on Automata, Languages, and Programming (ICALP’14), LNCS, Vol. 8573, pp. 13–25, Springer, 2014.

URL http://dx.doi.org/10.1007/978-3-662-43951-7_2

WSTS (introduced in ICALP’87) is a model that allows verification of safety properties of infinite-state systems. We will recall the definition and the essential results of WSTS. Most decidability results concerning well-structured transition systems apply to the *finitely branching* variant. Yet some models (inserting automata, ω -Petri nets, ...) are naturally infinitely branching. Here we develop tools to handle infinitely branching WSTS by exploiting the crucial property that in the (ideal) completion of a well quasi-ordered set, downward closed sets are *finite* unions of ideals. Then, using these tools, we derive decidability results and we delineate the undecidability frontier in the case of the termination, the coverability and the boundedness.

3.6 On the width of FAC orders

Mirna Džamonja (University of East Anglia, Norwich, GB)

Joint work of Sylvain Schmitz, Philippe Schnoebelen, Mirna Džamonja

License © Creative Commons BY 3.0 Unported license
© Mirna Džamonja

We investigate the ordinal invariants height, length and width of well quasi-orders, with the particular emphasis on width, which is an invariant also interesting in the case of the larger class of orders with finite antichain condition (*FAC*). We show that the width in the class of *FAC* orders is completely determined by the width in the class of WQOs, in the sense that if we know how to calculate the width of any WQO then we have a procedure to calculate the width of any given *FAC* order. We give formulas for the behavior of the width function under various classically defined operations with partial orders and obtain as a consequence a theorem that shows that for any ordinal α there is a WQO poset whose width is α . We make some progress towards a Minimax Theorem for the width function, complementing what was known about the height and the length. In addition to the Minimax Theorem, one of the main remaining questions is to give a complete formula for the width of the Cartesian products of WQOs. Even the width of the product of two ordinals is only known through a complex recursive formula. Although we have not given a complete answer to this question we have advanced the state of knowledge by considering some more complex special cases and in particular by calculating the width of certain products containing three factors.

3.7 BQOs and where they come from

Thomas E. Forster (University of Cambridge, UK)

License  Creative Commons BY 3.0 Unported license
© Thomas E Forster

A quasi-order $\langle X, \leq_X \rangle$ can be lifted naturally to a quasi-order on $V_\infty(X)$, the cumulative hierarchy based on X as a set of atoms. \leq_X is BQO iff this lifted quasi-order is well founded. It turns out that this condition is equivalent to the condition that the lift to $H_{\aleph_1}(X)$ (the hereditary countable sets over X as a set of atoms) is well founded (but this uses DC). The proof is in [1]. Two questions: (i) can we dispense with DC? (ii) $H_{\aleph_1}(X)$ is a rather set-theoretic construction. Might we be able to use instead something like the free countable completion of $\langle X, \leq_X \rangle$?

References

- 1 “Better-quasi-orderings and Coinduction”, *Theoretical Computer Science*, **309**, Issues 1-3, 2 December 2003, pp. 111–123.

3.8 The algorithmic complexity of recognizing unavoidable regularities of words.

Willem L. Fouché (UNISA – Pretoria, ZA)

License  Creative Commons BY 3.0 Unported license
© Willem L. Fouché

Main reference W. L. Fouché, “Unavoidable regularities and factor permutations of words”, in Proc. of the Royal Society of Edinburgh: Section A Mathematics, Vol. 125(3):519–524, 1995.

URL <http://dx.doi.org/10.1017/S0308210500032650>

Many theorems in combinatorics can be interpreted as statements expressing unavoidable regularities in words. Examples are Van der Waerden’s theorem or Graham-Rothschild theorems in Ramsey theory. It is a highly non-trivial task to understand the complexity of the bounds in terms of time and space of where and how these phenomena become manifest. It sometimes took decades before it was established that some of these phenomena belong to primitive recursive arithmetic [Shelah]. On the other hand, many such results have been shown not to be provable in Peano arithmetic [Paris-Harrington].

These results are frequently finitisations of topological results, results which in themselves do have constructive content [Coquand].

In this talk we shall explore the algorithmic content of the following statement, first discovered by topological means together with invoking notions involving well quasi-orderings which was then shown to be presentable by an Ackermann type recursion, leaving open the problem whether we can do better than that.

The result in question is as follows:

For every n, r we can find some $N = N(n, r)$ such that any word W on r symbols of length N will contain, for any permutation p of $\{1, \dots, n\}$ a subword (factor) of the form $w_1 \dots w_n X w_{p(1)} \dots w_{p(n)}$, where w_1, \dots, w_n, X are all subwords of W .

We shall discuss the recursive complexity of $N(n, r)$.

3.9 On Harvey Friedman's Finite Phase Transitions

Lev Gordeev (Universität Tübingen, DE)

License  Creative Commons BY 3.0 Unported license
© Lev Gordeev

The *proof theoretic integer* of a given theory S (abbreviated $PTI(S)$) is the least integer n such that every arithmetical Σ_1 sentence that has a proof in S with at most 10,000 symbols, has witnesses less than n . A good source of examples is in the area surrounding Kruskal's theorem. Consider Friedman's finite form (unstructured).

► **Theorem** (H.F.). *For all non-negative k there exist $n > 0$ such that the following holds. For all finite rooted trees T_1, \dots, T_n , where $|T_i| < i + k - 1$, there exist $i < j$ such that T_i is homomorphically embeddable into T_j .*

► **Definition** (H.F.). *Let $F(k)$ be the least n such that Theorem holds.*

► **Question** (H.F.). *How fast grows function F and where are its jumps, i.e. phase transitions?*

► **Theorem** (L.G.).

1. $F(0) = 2$, $F(1) = 3$, $F(2) = 6$, $F(3) = 125$. However
2. $F(4) > H_{\epsilon_0}(10^{10^{100}})$ [H being the Hardy function].
3. $PTI(PA) < F(4) < PTI(PA + \text{Consis}(PA))$. Moreover
4. PA proves that $F(4)$ does exist, but the length of any proof thereof must exceed 10,000 symbols.

3.10 A gentle introduction to Noetherian spaces

Jean Goubault-Larrecq (ENS – Cachan, FR)

License  Creative Commons BY 3.0 Unported license
© Jean Goubault-Larrecq

To prepare for the next two talks, I'll explain what Noetherian spaces are, and how they generalize WQOs. The exposition will stress concepts that play similar roles on the WQO side and on the topological side: upward closed subsets become opens, monotone subsequences become self-convergent subnets, and so on.

For example, a WQO is a quasi-ordered set where every monotonic sequence of upward-closed subsets stabilizes; a Noetherian space is a topological space where every monotonic sequence of opens stabilizes. A WQO is a quasi-ordered set where every upward-closed subset is the upward closure of finitely many points; a Noetherian space is a topological space where every open is compact. A WQO is a quasi-ordered set where every sequence has a monotonic subsequence; a Noetherian space is a topological space where every net has a self-convergent subnet (a net is self-convergent if and only if it converges to every of its points. I will mention the Alexandroff topology below, and in such a topology, the self-convergent nets are exactly those that are eventually monotone; in particular, every monotone sequence gives you an example of a self-convergent net).

The precise connection is as follows: every topological space can be considered as a quasi-ordered set, with the so-called *specialization quasi-ordering*, defined by $x \leq y$ iff every open neighborhood of x contains y ; conversely, the *Alexandroff topology* of a quasi-ordered set is the collection of all its upward-closed subsets. Starting from a quasi-order \leq , building

its Alexandroff topology, then the specialization quasi-ordering of the latter, we retrieve \leq . Doing a similar round-trip from topological spaces to topological spaces does not give you back the original topology in general, unless it happened to be an Alexandroff topology: there are many topologies that have the same specialization quasi-ordering, and this is the source of some additional freedom that Noetherian spaces provide us, compared to WQOs. For example, the powerset of a Noetherian space, with the lower Vietoris topology, is again Noetherian. The analogous result for WQOs (that the powerset of a WQO under domination would be WQO) has been known to be false since Rado.

I will also mention that most of the constructions that are known to preserve WQO-ness also preserve Noetherianity: the space of all finite words on a Noetherian alphabet is Noetherian (an extension of Higman’s Lemma), the space of all finite trees with vertices labeled by a Noetherian alphabet is Noetherian (generalizing Kruskal’s Theorem), notably.

Finally, I’ll mention two open problems:

- Following on Diana Schmidt, Andreas Weiermann, and others, is there a notion equivalent to that of maximal order types for Noetherian spaces? I contend that the ordinal height of the lattice of closed subsets should be the right notion, but we now need a theory of those: what is it for products of spaces, for sums, for spaces of words, etc.?
- What would be the natural analogue of BQOs in that topological theory of Noetherian spaces? One possible answer would be Noetherian spaces themselves, since they are already closed under the powerset construction. I don’t think this is satisfactory, in particular in the view of Yann Péquignot’s talk (with Raphaël Carroy) that the ideal completion remainder of a WQO is BQO iff the WQO is already BQO itself.

References

- 1 Jean Goubault-Larrecq. *Non-Hausdorff topology and domain theory*. Cambridge University Press, New Mathematical Monographs 22, 2013. See in particular Section 9.7.

3.11 Trace Universality for VASS

Simon Halfon (ENS – Cachan, FR)

License  Creative Commons BY 3.0 Unported license
© Simon Halfon

In [1], Esparza et al. have shown that the problem of trace universality for Petri Nets is decidable. The algorithm relies on the finiteness of bad sequences in the WQO $\mathbb{P}_f(\mathbb{N}^d)$. The complexity of the problem is addressed in the case $d = 1$ in [2]. They have shown that the problem is Ackermann-complete, using the tools introduced in [3] for the upper bound. In this talk I will present what is known on the complexity of this problem.

References

- 1 Petr Jancar, Javier Esparza, Faron Moller. *Petri Nets and Regular Processes*. J. Comput. Syst. Sci. 59(3):476–503 (1999)
- 2 Piotr Hofman, Patrick Totzke. *Trace Inclusion for One-Counter Nets Revisited*. RP 2014:151–162
- 3 Diego Figueira, Santiago Figueira, Sylvain Schmitz, Philippe Schnoebelen. *Ackermannian and Primitive-Recursive Bounds with Dickson’s Lemma*. LICS 2011:269–278

3.12 On Initial Segments of Topological Weihrauch Degrees

Peter Hertling (*Universität der Bundeswehr – München, DE*)

License  Creative Commons BY 3.0 Unported license
© Peter Hertling

Topological Weihrauch reducibility gives a very fine way of measuring the topological complexity of computation problems. We present old and new results stating that initial segments of topological Weihrauch degrees of certain classes of computation problems can be characterized in a combinatorial way by reducibilities between forests.

3.13 Well quasi-orderings and Hahn fields

Julia Knight (*University of Notre Dame, US*)

License  Creative Commons BY 3.0 Unported license
© Julia Knight
Joint work of Julia Knight, Karen Lange

Mourgues and Ressayre [2] showed that every real closed field has an integer part, where this is a discrete ordered subring appropriate for the range of a floor function. The proof gives an explicit procedure for embedding the given real closed field in a Hahn field. We wanted to measure the complexity of this procedure. For this, we needed to bound the lengths of roots of polynomials over the Hahn field, in terms of the lengths of the coefficients. In [3], we gave a conjecture, which we proved in [4]. We used results of de Jongh-Parikh [1] on well quasi-orderings.

References

- 1 D. H. J. De Jongh and R. Parikh, “Well partial orderings and hierarchies”, *Proc. Kon. Ned. Akad. Sci., Series A*, vol. 80(1977), pp. 195–207.
- 2 M. H. Mourgues and J. P. Ressayre, “Every real closed field has an integer part”, *J. Symb. Logic*, vol. 58(1993), pp. 641–647.
- 3 J. F. Knight and K. Lange, “Complexity of structures associated with real closed fields”, *Proc. of London Math. Society*, vol. 107(2013), pp. 177–197.
- 4 J. F. Knight and K. Lange, “Lengths of developments in $K((G))$ ”, pre-print.

3.14 Definability in some well partial orders

Oleg Kudinov (*Sobolev Institute of Mathematics, Novosibirsk, RU*)

License  Creative Commons BY 3.0 Unported license
© Oleg Kudinov
Joint work of Oleg Kudinov, Victor Selivanov

Well quasi-orders appear in many fields of Mathematics and Computer Science, where usually they play a role of classification tool. To our knowledge, the structure of concrete important WQOs (especially, definability aspect) was not investigated in details so far. In this talk we present some results on definability in some concrete WQOs, in particular, in the subword order on words and in the homomorphic quasi-order on labeled forests. In the case when the rank of a WQO is ω we are able in several cases to characterize completely the first-order definability. For WQOs of higher rank we discuss partial results and some open problems.

3.15 An axiom free Coq proof of Kruskal's tree theorem

Dominique Larchey-Wendling (LORIA – Nancy, FR)

License © Creative Commons BY 3.0 Unported license
 © Dominique Larchey-Wendling
URL <http://www.loria.fr/~larchey/Kruskal>

We present a Coq (<http://coq.inria.fr>) implementation of a purely inductive proof of Kruskal's tree theorem:

If R is a well quasi-order on the type X then `homeo_embed(R)` is a WQO on the type of finite trees decorated by values in X .

Contrary to classical proofs, there are a few instances of intuitionistic proofs for the Kruskal tree theorem. Some of these proofs requires the further assumption that the ground relation R is decidable (e.g. Monika Seisenberger's proof [2] or Jean Goubault-Larrecq's proof [1]). Wim Veldman's proof [3] is the only published proof that does not require that assumption of decidability, but it requires *Brouwer's thesis*. Moreover, none of these proofs had been mechanized before.

We implement a typed variant of Wim Veldman's intuitionistic proof and we show that the use of the axiom called "Brouwer's thesis" is not necessary in that setting which makes our proof an axiom free one (w.r.t. the CIC on which Coq is based).

We use Thierry Coquand *et al.* [4] inductive definition of *Almost-Full* (AF) relations as an alternative to Wim Veldman's. We present the architecture of Wim Veldman's proof and its fundamental constituents: Ramsey's theorem, the Fan theorem, combinatorial principles and Evaluation maps. We show how to replace Wim Veldman *stump* based induction by lexicographic products of relations well founded upto a projection.

The source code for this project can be accessed at the following web page: <http://www.loria.fr/~larchey/Kruskal>.

References

- 1 Jean Goubault-Larrecq. *A Constructive Proof of the Topological Kruskal Theorem*. MFCS 2013, LNCS 8087, pp. 22–41.
- 2 Monika Seisenberger. *On the Constructive Content of Proofs*. PhD dissertation, München, 2003.
- 3 Wim Veldman. *An intuitionistic proof of Kruskal's theorem*. Archive for Mathematical Logic 43(2):215–264, 2004.
- 4 Dimitrios Vytiniotis, Thierry Coquand and David Wahlstedt. *Stop When You Are Almost-Full – Adventures in Constructive Termination*. ITP 2012, LNCS 7406, pp. 250–265.

3.16 Robertson's conjecture on well quasi-ordering and topological minors

Chun-Hung Liu (Princeton University, US)

License © Creative Commons BY 3.0 Unported license
 © Chun-Hung Liu
Joint work of Robin Thomas, Chun-Hung Liu

One of the most prominent results in graph theory is Robertson and Seymour's Graph Minor Theorem: finite graphs are well quasi-ordered by the minor relation [1]. They also proved

that finite graphs are well quasi-ordered by the weak immersion relation [2], confirming a conjecture of Nash-Williams.

Topological minor relation is a graph containment relation that is closely related to the minor and the immersion relations. Kruskal's Tree Theorem and Robertson and Seymour's Weak Immersion Theorem imply that finite trees and finite subcubic graphs, respectively, are well quasi-ordered by the topological minor relation. However, unlike the minor and the weak immersion relation, the topological minor relation does not well quasi-order finite graphs in general.

In the late 1980's, Robertson conjectured that the known obstruction is the only obstruction. More precisely, he conjectured that for every positive integer k , finite graphs that do not contain a topological minor isomorphic to the graph obtained from the path of length k by duplicating each edge are well quasi-ordered by the topological minor relation. Joining with Robin Thomas, we prove this conjecture. The proof will be sketched in the talk.

An application of this result is that every topological-minor-closed property on certain classes of graphs can be characterized by finitely many graphs. It leads to the existences of cubic time algorithms to test those properties. But more applications of this theorem remain requiring further investigations.

References

- 1 N. Robertson and P.D. Seymour. *Graph Minors. XX. Wagner's conjecture*. J. Combin. Theory, Ser. B, 92 (2004), pp. 325–357.
- 2 N. Robertson and P.D. Seymour. *Graph Minors. XXIII. Nash-Williams' immersion conjecture*. J. Combin. Theory, Ser. B, 100 (2010), pp. 181–205.

3.17 Wqo and Bqo Theory in Reverse Mathematics

Alberto Marcone (University of Udine, IT)

License  Creative Commons BY 3.0 Unported license
© Alberto Marcone

This is a survey talk about WQO and BQO theory in reverse mathematics.

3.18 Dimensions of Mobility

Roland Meyer (University of Kaiserslautern, DE)

License  Creative Commons BY 3.0 Unported license
© Roland Meyer

We study natural semantic fragments of the pi-calculus: depth-bounded processes (there is a bound on the longest communication path), breadth-bounded processes (there is a bound on the number of parallel processes sharing a name), and name-bounded processes (there is a bound on the number of shared names). We give a complete characterization of the decidability frontier for checking if a pi-calculus process in one subclass belongs to another. Our main construction is a general acceleration scheme for pi-calculus processes. Based on this acceleration, we define a Karp and Miller (KM) tree construction for the depth-bounded pi-calculus. The KM tree can be used to decide if a depth-bounded process is name-bounded, if a depth-bounded process is breadth-bounded by a constant k , and if a name-bounded

process is additionally breadth-bounded. Moreover, we give a procedure that decides whether an arbitrary process is bounded in depth by a given k .

We complement our positive results with undecidability results for the remaining cases. While depth- and name-boundedness are known to be Σ_1 -complete, we show that breadth-boundedness is Σ_2 -complete, and checking if a process has a breadth bound at most k is Π_1 -complete, even when the input process is promised to be breadth-bounded.

References

- 1 R. Hüchting, R. Majumdar, and R. Meyer. Bounds on mobility. In *Proc. of the 25th International Conference on Concurrency Theory, CONCUR*, volume 8704 of LNCS, pages 357–371. Springer, 2014.

3.19 Notes on regularity and well quasi-ordering

Mizuhito Ogawa (JAIST – Ishikawa, JP)

License  Creative Commons BY 3.0 Unported license
© Mizuhito Ogawa

In [2], Ehrenfeucht et al. showed that a set L of finite words (over finite alphabet) is regular if and only if L is \leq -closed under some monotone well quasi-order \leq over finite words. This note briefly surveys extensions to finite trees and ω -words [4]. They are obtained by similar proofs by modifying the standard congruence in Myhill-Nerode theorem to those in [3, 1]. The extensions are,

1. a tree language L is regular if and only if L is \leq -closed under some monotone well quasi-order \leq over finite trees.
2. an ω -language L is regular if and only if L is \preceq -closed under a *periodic* extension \preceq of some monotone WQO over finite words, and
3. an ω -language L is regular if and only if L is \preceq -closed under a WQO \preceq over ω -words that is a *continuous* extension of some monotone WQO over finite words.

References

- 1 A. Arnold. A syntactic congruence for rational ω -languages. *Theoretical Computer Science*, 39:333–335, 1985.
- 2 A. Ehrenfeucht, D. Hausser, and G. Rozenberg. On regularity of context-free languages. *Theoretical Computer Science*, 27:311–332, 1983.
- 3 D. Kozen. On the Myhill-Nerode Theorem for Trees. *Bull. EATCS*, 47:170–173, 1992.
- 4 M. Ogawa. Well-quasi-orders and Regular ω -languages. *Theoretical Computer Science*, 324(1):55–60, 2004.

3.20 Well-structured pushdown systems

Mizuhito Ogawa (JAIST – Ishikawa, JP)

License  Creative Commons BY 3.0 Unported license
© Mizuhito Ogawa

Joint work of Xiaojuan Cai, Mizuhito Ogawa

Well-structured transition systems (WSTS) have been widely investigated [2]. We introduce an extension of WSTS with the stack, called *well-structured pushdown systems* (WSPDS),

which is a pushdown system with well quasi-ordered states and stack alphabet [1, 3, 4]. The decidability of their properties, such as coverability, boundedness, and termination are discussed. The boundedness and the termination are decidable under the strong monotonicity and the monotonicity, respectively. The decidability of the coverability has been shown under certain conditions, e.g.,

- When the states are 1-dimensional vectors and the stack alphabet is finite [4].
- When the states are finite and the stack alphabet is well quasi-ordered [1].

The latter is based on P-automata techniques, and the convergence of a P-automaton with the minimization rules implies the decidability of the coverability.

References

- 1 Cai, X., Ogawa, M.: Well-Structured Pushdown Systems. CONCUR 2013, LNCS 8052 (2013) 121–136.
- 2 Finkel, A., Schnoebelen, P.: Well-structured transition systems everywhere! Theoretical Computer Science **256**(1–2) (2001) 63–92.
- 3 Leroux, J., Praveen, M., Sutre, G.: Hyper-Ackermannian bounds for pushdown vector addition systems. CSL-LICS'14 (2014) 63:1–63:10.
- 4 Leroux, J., Sutre, G., Totzke, P.: On the coverability problem for pushdown vector addition systems in one dimension. ICALP'15, part 2, LNCS 9135 (2015) 324–336.

3.21 Noetherian Spaces in TTE

Arno Pauly (University of Brussels, BE)

License © Creative Commons BY 3.0 Unported license
© Arno Pauly

Many topological notions have associated computable notions (eg [2]). Here, we investigate the computable counterpart of Noetherian space. Unfortunately, it turns out that no non-empty space is computably Noetherian in the straight-forward sense.

Based on the idea of relativizing topological notions w.r.t. some endofunctor introduced in [1], and then investigate the ∇ -computably Noetherian spaces. These turn out to be well-behaved, and constitute a prime candidate for the correct notion of being Noetherian within computable analysis.

References

- 1 Arno Pauly & Matthew de Brecht. *Towards Synthetic Descriptive Set Theory: An instantiation with represented spaces*. arXiv 1307.1850, 2013
- 2 Arno Pauly. *On the topological aspects of the theory of represented spaces*. Computability, 2015

3.22 From well to better: the space of ideals

Yann Pequignot (Universität Wien, AT) and Raphael Carroy (University of Torino, IT)

License © Creative Commons BY 3.0 Unported license
© Yann Pequignot and Raphael Carroy

Main reference R. Carroy and Y. Pequignot, “From well to better, the space of ideals”, in *Fundamenta Mathematicae*, Vol. 227, pp. 247–270, 2014.

URL <http://dx.doi.org/10.4064/fm227-3-2>

On the one hand, the ideals of a well quasi-order (WQO) naturally form a compact topological space into which the WQO embeds. On the other hand, Nash-Williams’ barriers are given a uniform structure by embedding them into the Cantor space. We prove that every map from a barrier into a WQO restricts on a barrier to a uniformly continuous map, and therefore extends to a continuous map from a countable closed subset of the Cantor space into the space of ideals of the WQO. We then prove that, by shrinking further, any such continuous map admits a canonical form with regard to the points whose image is not isolated. As a consequence, we obtain a simple proof of a result on better quasi-orders (BQO); namely, a WQO whose set of non-principal ideals is a BQO is actually a BQO.

3.23 Problems on well quasi-orders and hereditary classes

Maurice Pouzet (University Claude Bernard – Lyon, FR)

License © Creative Commons BY 3.0 Unported license
© Maurice Pouzet

I will present some problems on well quasi-ordering and their interactions with hereditary classes of relational structures. Among the dozen of problems presented, only two are recent and due respectively to Atmitas and Lozin (2015) and to Abraham, Bonnet and Kubis (2008). The others go back to the seventies and are about the ordinal length of hereditary classes of graphs; the relationship between WQO and BQO for hereditary classes; the effect of labeling members of a hereditary class by a WQO poset; the preservation of the WQO character by adding a linear order; the extension of Laver’s theorem to hereditary classes; the effect of the WQO character of a hereditary class of finite structure on the asymptotic growth of the enumerative function of that class; and some other problems.

3.24 A constructive interpretation of open induction

Thomas Powell (Universität Innsbruck, AT)

License © Creative Commons BY 3.0 Unported license
© Thomas Powell

Nash-Williams’ well known minimal-bad-sequence argument can be elegantly reformulated as an instance of open induction over the lexicographic ordering on infinite sequences. In this talk I focus on how Gödel’s Dialectica interpretation can be used to give a constructive interpretation to general induction principles, and in particular discuss the problem of giving a direct realizer to the Dialectica interpretation of open induction which can be used to extract natural programs from Nash-Williams proofs of Higman’s lemma and Kruskal’s theorem. I conclude by presenting several related open problems.

3.25 Bounds for the strength of the graph minor theorem

Michael Rathjen (University of Leeds, GB)

License © Creative Commons BY 3.0 Unported license
© Michael Rathjen

Joint work of M. Krombholz, M. Rathjen

The graph minor theorem, GM, is arguably the most important theorem of graph theory. The strength of GM exceeds that of the standard classification systems of Reverse Mathematics known as the “big five”. The plan is to survey the current knowledge about the strength of GM, presenting lower and upper bounds.

3.26 Ordinal notations, the maximal order types of Kruskal’s Tree Theorem, and a tale of two cultures

Diana Schmidt (Heilbronn, DE)

License © Creative Commons BY 3.0 Unported license
© Diana Schmidt

Main reference Diana Schmidt, “Well-Partial Orderings and their Maximal Order Types,” Habilitationsschrift, Mathematics Faculty, Heidelberg University, 1979.

1. Why ordinal notations are useful, and what they are: Ordinal notations are terms built by applying functions from the ordinals to the ordinals, starting with 0. They are used to represent large ordinals. Such ordinal notations correspond in a natural way to labeled trees such as those in Kruskal’s Tree Theorem.
2. What my 1979 Habilitationsschrift theorem means (intuitively) for ordinal notations: it computes the maximal order types of the ordinal notation systems which correspond to the tree well quasi orderings in Kruskal’s tree theorem.
3. How I came to prove that theorem, and who else contributed to the proof (Schütte, Gandy, de Jongh, Parikh).
4. A tale of two cultures: also in 1979, Nachum Dershowitz submitted his paper “Orderings for term rewriting systems”, which depends essentially on Kruskal’s tree theorem. There was no internet; he was a computer scientist, I a mathematician. It was not til 15 years later that Andreas Weiermann unearthed the Habilitationsschrift and bridged the culture gap.

References

- 1 Diana Schmidt. *Well-Partial Orderings and their Maximal Order Types*. Habilitationsschrift, Mathematics Faculty, Heidelberg University, 1979.

3.27 Complexity Classes Beyond Elementary

Sylvain Schmitz (ENS – Cachan, FR)

License © Creative Commons BY 3.0 Unported license
© Sylvain Schmitz

Main reference S. Schmitz, “Complexity hierarchies beyond Elementary”, *ACM Transactions on Computation Theory*, Vol. 8(1:3), 2016.

URL <http://dx.doi.org/10.1145/2858784>

Well quasi-orders provide termination or finiteness arguments for many algorithms, and miniaturized versions can furthermore be employed to prove complexity upper bounds for those algorithms. We have however an issue with these bounds: they go way beyond the familiar complexity classes used in complexity theory. I shall discuss a definition of complexity classes suitable for the task. In particular, unlike the subrecursive classes they are based on, those classes support the usual notions of reduction and completeness.

3.28 Well-quasi-orderings for program verification and computational complexity

Philippe Schnoebelen (ENS – Cachan, FR)

License © Creative Commons BY 3.0 Unported license
© Philippe Schnoebelen

Joint work of Sylvain Schmitz, Christoph Haase, Philippe Schnoebelen

Main reference S. Schmitz and Ph. Schnoebelen, “The power of well-structured systems”, in *Proc. of the 24th Int’l Conf. on Concurrency Theory (CONCUR’13)*, LNCS, Vol. 8052, pp. 5–24, Springer, 2013.

URL http://dx.doi.org/10.1007/978-3-642-40184-8_2

Well-structured systems (WSTS) are a generic family of computational models where transitions are monotonic w.r.t. an effective well quasi-ordering of the states. This allows generic decidability proofs and verification algorithms for the verification of behavioral properties (like safety, liveness, . . .) [1, 2] Recent work by the authors aim at extracting computational complexity bounds from decidability proofs that rely on well quasi-orderings [3, 4].

References

- 1 S. Schmitz and Ph. Schnoebelen. The power of well-structured systems. In volume 8052 of *Lecture Notes in Computer Science*, pages 5–24. Springer, 2013.
- 2 Ch. Haase, S. Schmitz, and Ph. Schnoebelen. The power of priority channel systems. *Logical Methods in Comp. Science*, 10(4:4), 2014.
- 3 S. Schmitz and Ph. Schnoebelen. Multiply-recursive upper bounds with Higman’s lemma. In volume 6756 of *Lecture Notes in Computer Science*, pages 441–452. Springer, 2011.
- 4 Sylvain Schmitz. Complexity hierarchies beyond Elementary. *ACM Trans. Comput. Theory* 8(1:3), 2016.

3.29 Linearization as Conservation

Peter M. Schuster (University of Verona, IT), Davide Rinaldi, and Daniel Wessel (University of Trento, IT)

License © Creative Commons BY 3.0 Unported license
© Peter M. Schuster, Davide Rinaldi, and Daniel Wessel

A variant of Szpilrajn's Order Extension Principle (OEP) says that every partial order can be extended to a linear order. While OEP as it stands is a form of the Axiom of Choice, Negri-von Plato-Coquand 2004 have proved a proof-theoretic, purely syntactical counterpart: for quasi-orders, linearity is conservative when it comes to prove Horn sequents. This has now turned out a special case of the universal Krull-Lindenbaum conservation theorem we have gained from a criterion for conservation given by Scott 1974.

3.30 Higman's Lemma and its Computational Content

Helmut Schwichtenberg (LMU, München)

Joint work of Monika Seisenberger (Swansea University), Franziskus Wiesnet (LMU München), Helmut Schwichtenberg
License © Creative Commons BY 3.0 Unported license
© Helmut Schwichtenberg

Higman's Lemma is a fascinating result in infinite combinatorics, with manyfold applications in logic and computer science, that has been proven using different methods several times. The aim of the talk is to look at Higman's Lemma from a computational point of view. We give a proof of Higman's Lemma that uses the same combinatorial idea as Nash-Williams' indirect proof using the so-called minimal-bad-sequence argument, but which is constructive. For the case of a two letter alphabet such a proof was given by Coquand. Using more flexible structures, we present a proof that works for an arbitrary well quasi-ordered alphabet. We report on a formalization of this proof in the proof assistant Minlog, and discuss machine extracted terms (in an extension of Gödel's system T) expressing its computational content.

3.31 Well quasi-orders and descriptive set theory: some results and questions

Victor Selivanov (A. P. Ershov Institute – Novosibirsk, RU)

License © Creative Commons BY 3.0 Unported license
© Victor Selivanov

The existing hierarchies of sets have very easy structure (their levels are almost well ordered under inclusion) and they are sufficient for expressing apparently all interesting topological properties of sets. In contrast, existing classifications of functions and equivalence relations seem to be insufficient to express many specific properties of these objects. The situation is relatively clear for functions with finite range and for equivalence relations with finitely many classes but is much less clear for more complex objects.

In this talk, we survey some earlier results and discuss some more recent results and open questions in the specified direction, considering classifications from descriptive set theory

and automata theory. We try to demonstrate that this topic is closely related to WQO- and BQO-theory. We give some relevant references for the interested reader.

References

- 1 R. Carroy. A quasi-order on continuous functions. In: *Journal of Symbolic Logic* 78.2 (2013), pp. 633–648.
- 2 F. van Engelen, A. Miller and J. Steel. Rigid Borel sets and better quasi-order theory. *Contemporary mathematics*, 65 (1987), pp. 199–222.
- 3 P. Hertling. Topologische Komplexitätsgrade von Funktionen mit endlichem Bild. *Informatik-Berichte* 152, 34 pages, Fernuniversität Hagen, December 1993.
- 4 P. Hertling. *Unstetigkeitsgrade von Funktionen in der effektiven Analysis*. PhD thesis, Fachbereich Informatik, FernUniversität Hagen, 1996.
- 5 Y. Pequignot. A Wadge hierarchy for second countable spaces. *Archive for Mathematical Logic*, 54, No 5 (2015), pp. 659–683.
- 6 V. L. Selivanov. Boolean hierarchy of partitions over reducible bases. *Algebra and Logic*, 43, N 1 (2004), pp. 44–61.
- 7 V. L. Selivanov. The quotient algebra of labeled forests modulo h-equivalence. *Algebra and Logic*, 46, N 2 (2007), pp. 120–133.
- 8 V. L. Selivanov. Hierarchies of Δ_2^0 -measurable k -partitions. *Math. Logic Quarterly*, 53 (2007), 446–461.
- 9 V. L. Selivanov. A fine hierarchy of ω -regular k -partitions. B. Löwe et.al. (Eds.): *CiE 2011*, LNCS 6735, pp. 260–269. Springer, Heidelberg (2011).
- 10 V. L. Selivanov. Fine hierarchies via Priestley duality. *Annals of Pure and Applied Logic*, 163 (2012), pp. 1075–1107.
- 11 V. L. Selivanov. Total Representations. *Logical Methods in Computer Science* 9(2) (2013), pp. 1–30.
- 12 V. L. Selivanov. Towards a descriptive theory of cb_0 -spaces. Arxiv (2014): 1406.3942v1.
- 13 K. Weihrauch. The degrees of discontinuity of some translators between representations of the real numbers. Technical Report TR-92-050, International Computer Science Institute, Berkeley, 1992.

3.32 A Mechanized Proof of Higman’s Lemma by Open Induction

Christian Sternagel (Universität Innsbruck, AT)

License  Creative Commons BY 3.0 Unported license
© Christian Sternagel

I present a recent Isabelle/HOL formalization of a short proof of Higman’s lemma using open induction. The proof is based on Geser’s technical report “A Proof of Higman’s Lemma by Open Induction (1996)” but considerably simplified and amending an intermediate lemma.

References

- 1 Alfons Geser. *A proof of Higman’s Lemma by open induction*. Technical Report MIP-9606, Universität Passau, April 1996.
- 2 Jean-Claude Raoult. *Proving open properties by induction*. *Information Processing Letters*, 29(1):19–23, 1988.
- 3 Mizuhito Ogawa and Christian Sternagel. *Open Induction*. *Archive of Formal Proofs*, November 2012.

3.33 Some Challenges Related to Wqo-Theory

Andreas Weiermann (University of Ghent, BE)

License  Creative Commons BY 3.0 Unported license
© Andreas Weiermann

An important result of De Jongh and Parikh states that every well partial order can be extended to a well order of maximal possible order type. We discuss the role of this order type in bounding complexities arising from applications of well quasi-orders to termination problems. We also discuss Friedman style miniaturizations of well quasi-orders and indicate possible phase transitions. One specific example concerns Kruskal's theorem for which the critical constant is 0.639577689994720133112899870565731384115276481914419... (these decimal numbers have been calculated with great accuracy by Moritz Firsching).

Remaining challenges will be to determine critical constants related to other WQO-principles (joint work with Lev Gordeev) and the calculation of maximal order types for more general tree classes (joint work with Michael Rathjen and Jeroen Van der Meeren).

References

- 1 A. Weiermann *An application of graphical enumeration to PA*. J. Symbolic Logic 68 (2003), no. 1, pp. 5–16.
- 2 L. Gordeev and A. Weiermann: *Phase transitions of iterated Higman-style well-partial-orderings*. Arch. Math. Logic 51 (2012), no. 1–2, pp. 127–161.

4 Discussion and Open Problems

The following list of open problems reflects the discussion at the workshop, and can be used for further reference.

1. (Philippe Schnoebelen) What extra parameter could make the maximal order type functional again? That is, writing $o(X)$ for the maximal order type of a WQO X , it holds that $o(X + Y)$, $o(X \times Y)$ and $o(X^*)$ are completely determined by $o(X)$ and $o(Y)$ (for instance), but $o(\mathbb{P}_f(X))$ does not depend just on $o(X)$. ($\mathbb{P}_f(X)$ is the finite powerset of X , quasi-ordered by domination: $A \leq^b B$ iff for every $a \in A$, there is $b \in B$ such that $a \leq b$.) We know that $1 + o(X) \leq o(\mathbb{P}_f(X)) \leq 2^{o(X)}$, and the bounds are reached for some X . Considering dependencies on width and height are still not enough.

Answers/suggestions:

- a. (Thomas Forster) The rank of the tree of bad quadratic arrays.
 - b. (Lev Beklemishev) Look at the simpler case of subsets of two elements.
 - c. (Julia Knight) Conjecture: $o([X]^2)$ might be $o([o(X)]^2)$ – for finite non-linear X , it is false.
2. (Andreas Weiermann) Let α be an ordinal and (X, \leq) be a well quasi-order. Let $S_\alpha(X)$ be the set of all sequences $f: \beta \rightarrow X$ with $\beta < \alpha$ such that f has a finite range. For $f, g \in S_\alpha(X)$ with $f: \beta \rightarrow X$ and $g: \gamma \rightarrow X$ let $f \leq g$ if there exists a strictly monotonic function $h: \beta \rightarrow \gamma$ such that $f(\delta) \leq g(h\delta)$ for all $\delta < \beta$. Nash-Williams proved that $S_\alpha(X)$ is a well quasi-order. Is it possible to give good bounds for the maximal order type for $S_\alpha(X)$ in terms of α and the maximal order type of X ? This problem has been considered in Diana Schmidt's Habilitationsschrift.

3. (Andreas Weiermann) What is the maximal order type of the set of finite trees, with labels $\{0, 1, \dots, n\}$, and a gap condition? The case with 2 labels is known, see Jeroen van der Meeren's PhD thesis.
4. (Alberto Marcone) Laver's theorem states that the collection of countable scattered linear orders is WQO (even BQO) under embedding. A theorem by Hausdorff states that those orders can all be obtained by a certain enrichment process, indexed by ordinals. The Hausdorff rank $rh_H(L)$ of such an order L is the least ordinal where we obtain it by this construction. We already know the maximal order type of the subcollection of countable scattered linear orders with finite Hausdorff rank. As an attempt to approach Laver's theorem from below, rather than from above, and its proof-theoretic strength, what is the maximal order type of the collection of scattered linear orders of Hausdorff rank $< \omega^2$? It is a long-standing open problem, maybe too dangerous to give to a PhD student (Andreas Weiermann).
5. (Jean Goubault-Larrecq) The proper analogue of the notion of maximal order type for Noetherian spaces seems to be the ordinal height of their poset of closed subsets. Can we develop their De Jongh-Parikh theory? E.g., what is it for products, for sums, for spaces of words, etc.?
6. (Maurice Pouzet) Let P be a WQO. Is it true that $rank_{CB}(Idl(P)) = rank_{CB}(Idl(o(P)))$, where $rank_{CB}$ denotes Cantor Bendixson rank, $Idl(P)$ is the set of ideals of P (up-directed, non-empty, downward closed), and $o(P)$, as an ordinal, is considered as a WQO itself? The point is that we have a formula for the right-hand side.
7. (Lev Beklemishev) Is there any relationship between Noetherian spaces and scattered spaces?
8. (Yann Pequignot) Informal conjecture by Nash-Williams: is every "naturally occurring" WQO actually a BQO? Have you encountered any counterexample in your research? See also next problem, and problem 14.
Variant (a): is every "naturally occurring" WQO an ω^2 -WQO? (And all other variants of the same type.)
9. (Thomas Forster) It is still unknown whether the minor relation on finite graphs is a BQO. Is it?
Variant (a): Are subcases of that relation, which were already known to be WQO before the Robertson-Seymour result, already BQO? Graphs of bounded tree-width are known to be BQO.
10. (Sylvain Schmitz) Can we develop a reverse mathematics programme for WSTS? Is the proof-theoretic ordinal of the statement "this property is decidable for that model of WSTS" always the maximal order type of the underlying WQO?
11. (Raphael Carroy) Are continuous functions a WQO under \leq_1 , where $f \leq_1 g$ iff there are continuous functions F, G such that $f = F \circ g \circ G$?
12. (Jean Goubault-Larrecq) While Noetherian spaces seem to be the proper topological analogue of the order-theoretic notion of WQO, what would be the analogue for BQOs?
13. (Thomas Forster) Is there another definition of BQO that would help us in any of the BQO-related questions, letting us have slicker proofs?
14. (Maurice Pouzet) If you have a hereditary class of finite graphs (w.r.t. embeddability) which is WQO, is that class BQO?
15. (Philippe Schnoebelen, Sylvain Schmitz)
(This is a "reference request" type of question, it was prompted by Dick de Jongh's talk Friday morning.)

A quasi-ordering (A, \leq) leads to a natural notion of embedding on $Mat[A]$, the set of (finite) rectangular matrices M, N, \dots with elements from A . One lets $M \leq_{Mat} N$ when there is a submatrix N' of N (i.e., a matrix derived from N by removing some lines and columns) s.t. M and N' have same dimensions and $M[i, j] \leq N'[i, j]$ for all i, j .

Asking for which qos (A, \leq) one has $(Mat[A], \leq_{Mat})$ WQO is an exercise or homework problem that we sometimes give to our students after teaching them Higman's Lemma. We won't spoil the fun by answering here. The question is: do you know of some work where this question is mentioned/answered? What would be the best reference?

16. (Julia Knight) An integer part for a real closed ordered field R is a discrete ordered subring I such that for all $r \in R$, there exists $i \in I$ with $i \leq r < i + 1$. Mourgues and Ressayre proved that every real closed ordered field has an integer part. If R is countable, with universe ω , then the procedure of Mourgues and Ressayre yields an integer part that is $\Delta_{\omega}^0(R)$. Is there one that is $\Delta_2^0(R)$? See the next question, suggested by Beklemishev.
17. (Lev Beklemishev) Analyze from the point of view of reverse mathematics the theorem of Mourgues and Ressayre saying that every real closed ordered field has an integer part.
18. (Julia Knight) A divisible ordered Abelian group G is *Archimedean* if for all $a, b \in G^{>0}$, there exist natural numbers m, n such that $ma > b$ and $nb > a$. Let G be an Archimedean divisible ordered Abelian group. Suppose $S \subseteq G^{\geq 0}$, and let $[S]$ be the semi-group generated by elements of S . Let α be a multiplicatively indecomposable ordinal. If S has order type at most α , then so does $[S]$. But what can we say if G is *not* Archimedean?
19. (Julia Knight) Let G be an Archimedean divisible ordered Abelian group and let K be a field that is algebraically closed, or real closed. Suppose $p(x)$ is a polynomial over $K((G))$ with $Supp(p) \subseteq G^{\geq 0}$, and let r be a root with $w(r) > 0$. Let α be multiplicatively indecomposable. If $Supp(p)$ has order type at most α , then r has length at most α . What can we say if G is *not* Archimedean?
20. (Lev Beklemishev) In (D. Gabelaia, A. Kurucz, F. Wolter, M. Zakharyashev. Non-primitive recursive decidability of products of modal logics with expanding domains. *Annals of Pure and Applied Logic* 142 (1), 245–268) a natural notion of *expanding product* of Kripke frames was considered. Let (W, R) be a Kripke frame, and let F be a function assigning to each $x \in W$ a Kripke frame $F(x) = (W_x, R_x)$. We assume that whenever $x, y \in W$ and xRy , the frame $F(x)$ is a subframe of $F(y)$ in the sense that $W_x \subseteq W_y$ and $R_x = R_y \cap (W_x^2)$. An *e-frame (expanding frame)* associated with (W, F) is the set

$$\bigsqcup_{x \in W} F(x) = \{(x, u) : x \in W, u \in W_x\}$$

equipped with two binary relations R_1, R_2 such that

$$(x, u)R_1(y, v) \iff (xRy \wedge u = v),$$

$$(x, u)R_2(y, v) \iff (x = y \wedge uR_x v).$$

Expanding frames are models of propositional bimodal logic. By using Kruskal's Theorem, the authors of the paper cited above show that the bimodal logic determined by the class of all e-products of finite irreflexive trees is decidable. However, they also prove an Ackermannian lower bound by reducing to it the decision problem for lossy channel systems. Similar results are obtained for several other classes of frames.

What are sharp upper and lower bounds on the complexity of the decision problem for the bimodal logic of the class of e-products of finite irreflexive trees? Similar questions are also open for several other natural classes of frames studied in the paper cited above, in particular for linear frames.

Variant (a): find an axiomatization of the bimodal logic determined by the class of all e-products of finite irreflexive trees.

See also problem 26.

21. (Victor Selivanov) Define and investigate new natural topologically relevant WQOs (reducibilities) on Borel measurable functions $f : X \rightarrow Y$ between topological spaces (the particular case when both X, Y coincide with the Baire space is already important). Previous results in this direction (due to Wadge, Carlson-Laver, van Engelen-Miller-Steel, Weihrauch, Hertling, Selivanov, Carroy and others, some references may be found in my conference presentation) show that this research programme might be of great interest for descriptive set theory but the reducibilities considered so far do not seem sufficient for a deep understanding of Borel measurable functions.
22. (Victor Selivanov) For a qo Q , let \mathcal{T}_Q (resp. $\tilde{\mathcal{T}}_Q$) be the set of finite (resp. of at most countable well founded) Q -labeled trees (T, c_T) equipped with the homomorphism qo \leq_h defined as follows: $(S, c_S) \leq_h (T, c_T)$ iff there is a monotone (not necessarily injective) function $\varphi : S \rightarrow T$ such that $\forall x \in S (c_S(x) \leq c_T(\varphi(x)))$. Several versions of these constructions that were introduced and studied in my publications (see e.g. LNCS 6735 (2011), p. 260-269, APAL 163 (2012), p. 1075-1107, Arxiv (2014): 1406.3942v1) turn out to be relevant to classifying some topological objects. From well known facts of WQO-theory it follows that \mathcal{T}_Q (resp. $\tilde{\mathcal{T}}_Q$) is WQO (resp. BQO) provided that Q is WQO (resp. BQO). The question is to understand the relationships between Q and \mathcal{T}_Q (resp. $\tilde{\mathcal{T}}_Q$), in particular to compute the ranks (heights), the maximal order types and other natural invariants (like the automorphism group of the corresponding quotient-orders) of $\mathcal{T}_Q, \tilde{\mathcal{T}}_Q$ for natural Q . The question is interesting and non-trivial also for iterations of these constructions (and their modifications), e.g. for $\mathcal{T}_{\mathcal{T}_k}$ and $\tilde{\mathcal{T}}_{\tilde{\mathcal{T}}_k}$ where k is the antichain with $k < \omega$ elements.
23. (Victor Selivanov) Continue the systematic investigation of (un)decidability and definability issues of natural WPOs on words, trees, forests, graphs and other structures relevant to WQO theory and Computer Science. Some interesting work in this direction is already done by Comon, Kuske, Selivanov, Kudinov, Schnoebelen, and many others.
One challenging concrete problem is: what is a precise estimate of the m -degree of first-order theory of the quotient-order of $\tilde{\mathcal{T}}_k$ (for $k \geq 3$) from the previous question?
Variant (a): can you characterize the first-order definable relations in the quotient-order of $\mathcal{T}_{\mathcal{T}_k}$ (also for $k \geq 3$)?
In solving such questions the tools developed in (Kudinov-Selivanov, LNCS 5635 (2009), p. 290-299) seem especially relevant since they probably generalize to many natural well founded partial orders.
24. (Victor Selivanov) Are the Weihrauch reducibilities \leq_1, \leq_2 WQO on the Borel-measurable functions from the Baire space \mathcal{N} to the discrete space with countably many points? Let us recall that $f \leq_1 g$ iff there are continuous functions F, G such that $f = F \circ g \circ G$, while $f \leq_2 g$ iff there are continuous functions F, G such that for every $x, f(x) = F(x, g(G(x)))$.
Variant (a): Is the continuous reducibility WQO on the Borel equivalence relations with countably many equivalence classes?
25. (Victor Selivanov, Oleg Kudinov) Let (F_P, \leq_h) denote the factorization (i.e., the quotient order) of the set of all finite forests with vertices labeled by elements from WQO (P, \leq) , and \leq_h is the homomorphic quasi-ordering of problem 22 (it is WQO again; please do not confuse it with the homeomorphic embedding as mentioned in Kruskal's Theorem). The detailed properties of such WQOs are not established so far even for finite P . So, the question is to characterize them in terms of finite P : 1) $\text{height}(F_P)$; 2) $o(F_P)$; 3) $Th(F_P)$. For the last point the conjecture is that this theory is decidable iff $\text{width}(P) < 3$.

26. (Sylvain Schmitz) Same questions as in problem 20 for one-variable FOLTL with counting over expanding domains on finite linear orders (C. Hampson and A. Kurucz. Undecidable propositional bimodal logics and one-variable first-order linear temporal logics with counting. *ACM Transactions on Computational Logic* 16(3:27), 2015).

Participants

- Lev D. Beklemishev
Steklov Institute – Moscow, RU
- Andreas R. Blass
University of Michigan –
Ann Arbor, US
- Raphael Carroy
University of Torino, IT
- Raphael Chane-Yack-Fa
University of Sherbrooke, CA
- Matthew de Brecht
NICT – Osaka, JP
- Dick De Jongh
University of Amsterdam, NL
- Mirna Dzamonja
University of East Anglia –
Norwich, GB
- Alain Finkel
ENS – Cachan, FR
- Thomas E. Forster
University of Cambridge, GB
- Willem L. Fouché
UNISA – Pretoria, ZA
- Christian Glasser
Universität Würzburg, DE
- Lev Gordeev
Universität Tübingen, DE
- Jean Goubault-Larrecq
ENS – Cachan, FR
- Simon Halfon
ENS – Cachan, FR
- Peter Hertling
Universität der Bundeswehr –
München, DE
- Takayuki Kihara
University of California –
Berkeley, US
- Julia Knight
University of Notre Dame, US
- Oleg V. Kudinov
Sobolev Institute of Mathematics
– Novosibirsk, RU
- Dominique Larchey-Wendling
LORIA – Nancy, FR
- Chun-Hung Liu
Princeton University, US
- Alberto Marcone
University of Udine, IT
- Roland Meyer
TU Kaiserslautern, DE
- Kenji Miyamoto
LMU München, DE
- Mizuhito Ogawa
JAIST – Ishikawa, JP
- Arno Pauly
University of Brussels, BE
- Yann Pequignot
Universität Wien, AT
- Maurice Pouzet
University Claude Bernard –
Lyon, FR
- Thomas Powell
Universität Innsbruck, AT
- Michael Rathjen
University of Leeds, GB
- Diana Schmidt
Heilbronn, DE
- Heinz Schmitz
Hochschule Trier, DE
- Sylvain Schmitz
ENS – Cachan, FR
- Philippe Schnoebelen
ENS – Cachan, FR
- Peter M. Schuster
University of Verona, IT
- Helmut Schwichtenberg
LMU München, DE
- Luc Segoufin
INRIA SENS-Cachan, FR
- Monika Seisenberger
Swansea University, GB
- Victor Selivanov
A. P. Ershov Institute –
Novosibirsk, RU
- Dieter Spreen
Universität Siegen, DE
- Christian Sternagel
Universität Innsbruck, AT
- Bill Wadge
University of Victoria, CA
- Andreas Weiermann
Ghent University, BE
- Klaus Weihrauch
FernUniversität in Hagen, DE
- Daniel Wessel
University of Trento, IT



Privacy and Security in Smart Energy Grids

Edited by

George Danezis¹, Stefan Katzenbeisser², Christiane Peters³, and
Bart Preneel⁴

1 University College London, GB, g.danezis@ucl.ac.uk

2 TU Darmstadt, DE, skatzenbeisser@acm.org

3 IBM Belgium, BE, christiane.pascale.peters@gmail.com

4 KU Leuven, BE, bart.preneel@esat.kuleuven.be

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16032 “Privacy and Security in Smart Energy Grids”. Smart electricity grids augment the electricity distribution network with modern communications and computerized control to improve efficiency, reliability, and security of electricity distribution, and more flexible production. This initiative has been greeted by consumers and utilities not only with enthusiasm but also concern. Consumers worry about their privacy. Utilities worry about the security of their assets. These outcries and reactions have triggered academics and industry to look into designing privacy friendly architectures for smart metering. The Dagstuhl Seminar 16032 brought together academic researchers as well as utility experts in order to start an open dialogue on smart grid privacy and security problems and potential solutions to support customers and utilities. A particular focus of the seminar were problems related to two timely use-cases for the smart grid, namely smart charging of electric vehicles and distribution automation.

Seminar January 17–20, 2016 – <http://www.dagstuhl.de/16032>

1998 ACM Subject Classification K.6.5 Security and Protection

Keywords and phrases Critical infrastructure protection, smart energy grids

Digital Object Identifier 10.4230/DagRep.6.1.99

1 Executive Summary

George Danezis

Stefan Katzenbeisser

Christiane Peters

Bart Preneel

License  Creative Commons BY 3.0 Unported license
© George Danezis, Stefan Katzenbeisser, Christiane Peters, and Bart Preneel

Smart electricity grids augment the electricity distribution network with modern communications and computerized control to improve efficiency, reliability, and security of electricity distribution, and more flexible production. This initiative has been greeted by consumers and utilities not only with enthusiasm but also concern. Consumers worry about their privacy. Utilities worry about the security of their assets.

Consumer organizations across the globe protested against smart meters and smart homes collecting all their data, warning that security breaches in the databases of the utilities would expose privacy-critical data to attackers, or open to secondary uses leading to increased insurance premiums, behavioral advertising or privacy invasion. These outcries and reactions



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

Privacy and Security in Smart Energy Grids, *Dagstuhl Reports*, Vol. 6, Issue 1, pp. 99–107

Editors: George Danezis, Stefan Katzenbeisser, Christiane Peters, and Bart Preneel



Dagstuhl Reports

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

have triggered academics and industry to look into designing privacy friendly architectures for smart metering.

The seminar 16032 in particular focused on the two use cases of smart charging of electric vehicles (EVs) and distribution automation. The seminar discussed these use cases with respect to the following challenges:

- security architectures,
- secure and privacy-friendly communication, and
- hardware and software security for constrained devices in the smart grid.

Smart Charging: Charging of electric vehicles is the next big challenge for privacy and security researchers: smart charging algorithms try to minimize loads on the grid by collecting various kinds of customer data, making it easy to reserve charging spots and book charge frequencies using smart-phone apps. The main motivation behind smart charging is to save copper for cables to match the load demands, given that an electric vehicle draws as much as a full household. Cables are designed to satisfy the demands at peak times. So profiling customers helps to foresee these demands and to calculate the cost of the needed grid infrastructure. Moreover, the cable designs use prediction algorithms to optimize loads, while assigning low priority to privacy issues, security architectures, and secure communication protocols.

Distribution Automation: Another problem lies in the task of automated electricity distribution. In a smart grid, safety critical events in transformer stations can be monitored and operated remotely. Adding communication also exposes assets to new vulnerabilities and attacks. Grid components are controlled by dedicated devices that pose a challenge in terms of their storage and computation capacities. Moreover, as with any critical infrastructure, security often conflicts with safety. As a consequence security often does not play any role in the design of communication protocols and devices, supported by the argument that most devices reside in physically protected substations. However, providing such physical security is expensive and hackers do not need physical access to the grid operator sites if they are connected to the utility's IT network.

The goal of this seminar was thus (i) to raise awareness of these critical problems affecting every European citizen now or at least in the foreseeable future, and (ii) to bring together academic researchers as well as utility experts in order to start an open dialogue on smart grid privacy and security problems and potential solutions to support customers and utilities.

2 Table of Contents

Executive Summary

George Danezis, Stefan Katzenbeisser, Christiane Peters, and Bart Preneel 99

Overview of Talks

In praise of distributed trusted computing bases

George Danezis 102

EV Smart Charging Security Architectures

Benessa Defend 102

Lessons Learned from Implementing Privacy-Preserving Protocols for Smart Meters

Benessa Defend 102

The Interplay of Data Resolution and Privacy in Smart Metering

Dominik Engel 103

α -Signatures: Some observations on the integrity of measurements in the privacy-preserving smart grid

Florian Kerschbaum 103

Field experiences with securing RTUs

Carlos Montes Portela 104

Smart Charging of EVs

Carlos Montes Portela 104

Charging my EV at my Friend's House

Mustafa Mustafa 105

Security of EV charging

Erik Poll 105

What about the software?

Erik Poll 106

Participants 107

3 Overview of Talks

3.1 In praise of distributed trusted computing bases

George Danezis (University College London, GB)

License © Creative Commons BY 3.0 Unported license
© George Danezis

The Trusted Computing Base (TCB) is a foundational concept in computer security, defining the set of hardware, software and processes that need to be protected from the adversary to ensure the security properties of the system are not violated. However, the TCB as a monolithic entity is rather old fashioned. These days the integrity and confidentiality of the TCB is instead preserved through distributing the functionality across a number of components, each of which could fail. Such distributed TCBs have proved to be robust against extremely motivated and well-resourced adversaries, but engineering them remains a technical and cryptographic challenge.

3.2 EV Smart Charging Security Architectures

Benessa Defend (ENCS – The Hague, NL)

License © Creative Commons BY 3.0 Unported license
© Benessa Defend

Joint work of Defend, Benessa; Montes Portela, Carlos; Kursawe, Klaus

We present an overview of the electric vehicle (EV) charging architecture from generation to consumption. The architectural overview includes various key stakeholders and shows the information flows between multiple parties for charging and billing purposes. We zoom in on the components inside EV charging stations and explore a number of threat scenarios from theft of charging cables to large-scale attacks on charging stations that could lead to a blackout. The talk closes with security considerations from the point of view of electricity grid operators whose objective is to maintain a balanced electricity grid with minimal outages.

3.3 Lessons Learned from Implementing Privacy-Preserving Protocols for Smart Meters

Benessa Defend (ENCS – The Hague, NL)

License © Creative Commons BY 3.0 Unported license
© Benessa Defend

Main reference B. Defend, K. Kursawe, “Implementation of privacy-friendly aggregation for the smart grid”, in Proc. of the 1st ACM Workshop on Smart Energy Grid Security, pp. 65–74, ACM, 2013.

URL <http://dx.doi.org/10.1145/2516930.2516936>

This talk provides an update on the privacy-preserving smart meter aggregation activities that have continued since Dagstuhl Seminar 11511 that was held in December 2011. One of the privacy-preserving protocols developed by Kursawe et al. [1] was implemented on four real smart meters in collaboration with a smart meter manufacturer. Based on the success of this demonstration we teamed up with an electricity grid operator to conduct scalability and integration tests on 100 meters. We also cover several lessons-learned, including the importance of understanding use cases, framing privacy as a business enabler,

and the importance of good security metaphors. Currently the privacy-preserving aggregation protocol has been submitted to the DLMS/COSEM smart meter standard as an optional add-on; support from stakeholders is needed in order to finalize the adoption process.

References

- 1 Klaus Kursawe, George Danezis, Markulf Kohlweiss, “Privacy-Friendly Aggregation for the Smart-Grid,” in Privacy Enhancing Technologies Symposium’ (PETS), pp. 175–191, 2011.
- 2 Benessa Defend, Klaus Kursawe, “Implementation of privacy-friendly aggregation for the smart grid,” in SEGS@CCS 2013, pp. 65–74, 2013.

3.4 The Interplay of Data Resolution and Privacy in Smart Metering

Dominik Engel (FH Salzburg, AT)

License © Creative Commons BY 3.0 Unported license
© Dominik Engel

Joint work of Eibl, Günther; Engel, Dominik

Main reference D. Engel, G. Eibl, “Wavelet-Based Multiresolution Smart Meter Privacy”, IEEE Transactions on Smart Grid, 12 pages, 2015.

URL <http://dx.doi.org/10.1109/TSG.2015.2504395>

Through smart metering load profiles are measured per household. Personal data can be inferred from these load profiles, which has led to privacy concerns. Privacy is expected to increase with longer measurement intervals. In this talk, first the impact of data granularity on edge detection, a first step in appliance detection, is reviewed. Based on these insights, a method for generating multi-resolution representation of load profiles by using the wavelet transform is presented. By using a hierarchical keying scheme and different keys in the different keys on the various resolutions, users can decide which party can access their load profile at which resolution. Finally, open issues and further research directions are discussed.

3.5 α -Signatures: Some observations on the integrity of measurements in the privacy-preserving smart grid

Florian Kerschbaum (SAP SE – Karlsruhe, DE)

License © Creative Commons BY 3.0 Unported license
© Florian Kerschbaum

Main reference F. Kerschbaum, H. W. Lim, “Privacy-Preserving Observation in Public Spaces,” in Proc. of the 20th European Symp. on Research in Computer Security (ESORICS’15), LNCS, Vol. 9327, pp. 81–100, Springer, 2015.

URL http://dx.doi.org/10.1007/978-3-319-24177-7_5

The privacy architecture for the smart grid as put forward by Jawurek, Kerschbaum, Danezis and others hinges on some assumptions. Particularly integrity of measurements is ensured by secure hardware. In this talk I will investigate whether we can relax this assumption by using the concept of α -signatures – an extension of α -authentication introduced by Kerschbaum and Lim at ESORICS 2015.

3.6 Field experiences with securing RTUs

Carlos Montes Portela (Enexis B. V. – 's-Hertogenbosch, NL)

License © Creative Commons BY 3.0 Unported license
 © Carlos Montes Portela
URL <http://www.enecs.eu>

RTU stands for Remote Terminal Unit and is basically a PC for industrial purposes. Distribution System Operators (DSOs) use RTUs in their substations to monitor their electricity grids (eg. measurements of voltage and current levels) or to remotely switch parts of the grid off or on. Typically a master/slave architecture is used, where there is one master (control center of the DSO) and multiple slaves (the RTUs in the substations). The communication from and to the RTUs is done via a standardized protocol: IEC-60870-5-104 also known as the 104 protocol. Performance and reliability have had much more attention from the industry than security. As the number and magnitude of cyber-physical related attacks grows, it becomes more and more important for DSOs to have RTUs that are secure enough to cope with these risks appropriately.

DSO Enexis is rolling out station automation and wants to make sure that security is taken into account upfront. Therefore, ENCS (European Network for Cyber Security) was hired to setup security related requirements that can be used in European tenders. ENCS has delivered a good quality report that has already been used by Enexis during a tender. Other DSOs that are members of ENCS will do the same. As a result, RTU vendors will need to adhere to these requirements in order to win bids.

During this talk the RTU tendering process and the usage of the ENCS security requirements were discussed.

3.7 Smart Charging of EVs

Carlos Montes Portela (Enexis B. V. – 's-Hertogenbosch, NL)

License © Creative Commons BY 3.0 Unported license
 © Carlos Montes Portela
Joint work of Verheijen, Lennart; Klapwijk, Paul; Montes Portela, Carlos; Postma, André
Main reference C. Montes Portela, D. Geldtmeijer, M. van Eekelen, H. Slootweg, “A flexible and privacy friendly ICT architecture for Smart Charging of EV’s,” in Proc. of the 22nd Int’l Conf. on Electricity Distribution (CIRED’13), paper 0199, 2013.
URL http://www.cired.net/publications/cired2013/pdfs/CIRED2013_0199_final.pdf

Building infrastructures for charging electric vehicles (EVs) is a complex task, optimizing counteracting goals. The main aim is maximizing EV driver’s convenience, by using the available charging infrastructure and local grid capacity as efficiently as possible. By controlling the charging process, the DSO could optimize the grid usage and facilitate the integration of RES. Herewith additional investments necessary for (large scale) EV charging could be avoided or at least minimized. This is coined as ‘Smart Charging’ by Eurelectric [2]. During this talk the concept of smart charging has been explained.

Implementing Smart Charging in a liberalized context, calls for an interaction and corresponding information exchange between DSOs, Charge Spots, EVs, EV drivers, energy suppliers and possibly new market participants. Amongst the latter, one could count a Charge Service Provider (CSP) which deals with fulfilling the charge wish of the EV driver and a Charge Spot Operator (CSO), which deals with the operation of the Charge Spots. Without measures, one could derive the charge locations of EVs throughout time. If this

could be coupled to EV drivers, it would then become privacy sensitive data as it reveals the whereabouts of the latter. Based on the negative experiences with privacy during the roll-out of Smart Meters in the Netherlands, this could become a problem for the concept of Smart Charging. Furthermore, the interest of hackers and commercial parties for privacy sensitive data increases the likelihood of disclosure. Lastly, the fact that the proposed marked model for public EV charging is still evolving into its full maturity, calls for an ICT architecture that is flexible enough to deal with future changes. During this talk multiple variants of the evolving role model have been presented.

References

- 1 Montes Portela, Carlos, Geldtmeijer, Danny, van Eekelen, Marko & Slootweg, Han, “A flexible and privacy friendly ICT architecture for Smart Charging of EV’s,” in Proceedings Cired Conference 2013, paper 0199.
- 2 Geldtmeijer, Danny, Hommes, Klaas &, Postma, André, 2011, “Charging EVs in a liberalized electricity market,” in Proceedings Cired Conference 2011, paper 0889.

3.8 Charging my EV at my Friend’s House

Mustafa Mustafa (KU Leuven, BE)

License © Creative Commons BY 3.0 Unported license
© Mustafa Mustafa

Joint work of Mustafa Mustafa, Ning Zhang, Georgios Kalogridis, Zhong Fan

Main reference M. A. Mustafa, N. Zhang, G. Kalogridis, Z. Fan, “Roaming electric vehicle charging and billing: An anonymous multi-user protocol,” in Proc. of the 2014 IEEE Int’l Conf. on Smart Grid Communications (SmartGridComm’14), pp. 939–945, 2014.

URL <http://dx.doi.org/10.1109/SmartGridComm.2014.7007769>

In this talk, I will first briefly introduce the smart grid and how the current electricity markets work. Then, I will present on a high level a secure roaming electric vehicle (EV) charging protocol that helps preserve users’ privacy. This protocol protects the user’s identity privacy from other suppliers as well as the user’s privacy of location from its own supplier. Further, it allows the user’s contracted supplier to authenticate the EV and the user. Using two-factor authentication approach a multi-user EV charging is supported and different legitimate EV users (e.g. family members) can be held accountable for their charging sessions. Finally, I will give some high level ideas on how this protocol can be extended to integrate EV aggregators which could help EV users to participate in the balancing electricity market.

3.9 Security of EV charging

Erik Poll (Radboud University Nijmegen, NL)

License © Creative Commons BY 3.0 Unported license
© Erik Poll

Charging Electric Vehicles (EVs) involves many parties and many information flows between them. Besides privacy concerns, the big impact of EV charging on the grid raises important security concerns w.r.t. grid stability. However, currently little attention is paid to security (more specifically, to authentication and integrity) in protocols for information exchanges to support EV charging.

References

- 1 Fabian van den Broek, Erik Poll, and Bárbara Vieira, “Securing the information infrastructure for EV charging”, International Workshop on Communication Applications in Smart Grid (CASG 2015), LNICST Vol. 154, pp. 61–74, Springer, 2015.

3.10 What about the software?

Erik Poll (Radboud University Nijmegen, NL)

License © Creative Commons BY 3.0 Unported license
© Erik Poll

Joint work of Erik Poll, Joeri de Ruiter, Aleksy Schubert

Main reference E. Poll, J. de Ruiter, A. Schubert, “Protocol state machines and session languages: specification, implementation, and security flaws”, in Proc. of the 2015 IEEE Security and Privacy Workshops (SPW’15), pp. 125–133, IEEE, 2015.

URL <http://dx.doi.org/10.1109/SPW.2015.32>

Software – or rather, the presence of flaws in software – is a major root cause of security problems. Language-theoretic security is a collection of ideas to tackle an important class of security flaws in software, namely flaws in handling (possibly malicious) input. These ideas also seem highly relevant for the protocols using in the smart grids. We used these ideas in fuzzing GSM and in analysing the protocol state machines of TLS.

Participants

- Nikita Borisov
University of Illinois – Urbana
Champaign, US
- George Danezis
University College London, GB
- Benessa Defend
ENCs – The Hague, NL
- Dominik Engel
FH Salzburg, AT
- Zekeriya Erkin
TU Delft, NL
- Benedikt Gierlichs
KU Leuven, BE
- Stefan Katzenbeisser
TU Darmstadt, DE
- Florian Kerschbaum
SAP SE – Karlsruhe, DE
- Erwin Kooi
Alliander – Duiven, NL
- Klaus Kursawe
ENCs – The Hague, NL
- Éireann Leverett
University of Cambridge, GB
- Carlos Montes Portela
Enexis B.V. –
's-Hertogenbosch, NL
- Mustafa Mustafa
KU Leuven, BE
- Christiane Peters
IBM Belgium, BE
- Erik Poll
Radboud University Nijmegen,
NL
- Bart Preneel
KU Leuven, BE
- Ahmad-Reza Sadeghi
TU Darmstadt, DE
- Kazuo Sako
NEC – Kawasaki, JP
- Matthias Schunter
INTEL ICRI – Darmstadt, DE
- Neeraj Suri
TU Darmstadt, DE
- Makoto Takahashi
Tohoku University – Sendai, JP
- Pol Van Aubel
Radboud University
Nijmegen, NL
- Ingrid Verbauwhede
KU Leuven, BE
- Jos Weyers
TenneT – Arnhem, NL



Reproducibility of Data-Oriented Experiments in e-Science

Edited by

Juliana Freire¹, Norbert Fuhr², and Andreas Rauber³

1 New York University, US, juliana.freire@nyu.edu

2 Universität Duisburg-Essen, DE, norbert.fuhr@uni-due.de

3 TU Wien, AT, rauber@ifs.tuwien.ac.at

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16041 “Reproducibility of Data-Oriented Experiments in e-Science”. In many subfields of computer science, experiments play an important role. Besides theoretic properties of algorithms or methods, their effectiveness and performance often can only be validated via experimentation. In most of these cases, the experimental results depend on the input data, settings for input parameters, and potentially on characteristics of the computational environment where the experiments were designed and run. Unfortunately, most computational experiments are specified only informally in papers, where experimental results are briefly described in figure captions; the code that produced the results is seldom available.

This has serious implications. Scientific discoveries do not happen in isolation. Important advances are often the result of sequences of smaller, less significant steps. In the absence of results that are fully documented, reproducible, and generalizable, it becomes hard to re-use and extend these results. Besides hindering the ability of others to leverage our work, and consequently limiting the impact of our field, the absence of reproducibility experiments also puts our reputation at stake, since reliability and validity of empiric results are basic scientific principles.

This seminar brought together experts from various sub-fields of computer science to create a joint understanding of the problems of reproducibility of experiments, discussing existing solutions and impediments, and proposing ways to overcome current limitations.

Seminar January 24–29, 2016 – <http://www.dagstuhl.de/16041>

1998 ACM Subject Classification A.1 Introductory and Survey

Keywords and phrases Documentation, Reliability, Repeatability, Replicability, reproducibility, Software

Digital Object Identifier 10.4230/DagRep.6.1.108

Edited in cooperation with Daniel Garijo

1 Executive Summary

Norbert Fuhr

Juliana Freire

Andreas Rauber

License  Creative Commons BY 3.0 Unported license
© Norbert Fuhr, Juliana Freire, and Andreas Rauber

In many subfields of computer science, experiments play an important role. Besides theoretical properties of algorithms or methods, their effectiveness and performance often can only be validated via experimentation. In most of these cases, the experimental results depend



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

Reproducibility of Data-Oriented Experiments in e-Science, *Dagstuhl Reports*, Vol. 6, Issue 1, pp. 108–159

Editors: Juliana Freire, Norbert Fuhr, and Andreas Rauber



DAGSTUHL
REPORTS

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

on the input data, settings for input parameters, and potentially on characteristics of the computational environment where the experiments were designed and run. Unfortunately, most computational experiments are specified only informally in papers, where experimental results are briefly described in figure captions; the code that produced the results is seldom available.

This has serious implications. Scientific discoveries do not happen in isolation. Important advances are often the result of sequences of smaller, less significant steps. In the absence of results that are fully documented, reproducible, and generalizable, it becomes hard to re-use and extend these results. Besides hindering the ability of others to leverage our work, and consequently limiting the impact of our field, the absence of reproducibility experiments also puts our reputation at stake, since reliability and validity of empiric results are basic scientific principles.

Reproducible results are not just beneficial to others – in fact, they bring many direct benefits to the researchers themselves. Making an experiment reproducible forces the researcher to document execution pathways. This in turn enables the pathways to be analyzed (and audited). It also helps newcomers (e.g., new students and post-docs) to get acquainted with the problem and tools used. Furthermore, reproducibility facilitates portability, which simplifies the dissemination of the results. Last, but not least, preliminary evidence exists that reproducibility increases impact, visibility and research quality.

However, attaining reproducibility for computational experiments is challenging. It is hard both for authors to derive a compendium that encapsulates all the components (e.g., data, code, parameter settings, environment) needed to reproduce a result, and for reviewers to verify the results. There are also other barriers, from practical issues – including the use of proprietary data, software and specialized hardware, to social – for example, the lack of incentives for authors to spend the extra time making their experiments reproducible.

This seminar brought together experts from various sub-fields of Computer Science as well as experts from several scientific domains to create a joint understanding of the problems of reproducibility of experiments, discuss existing solutions and impediments, and propose ways to overcome current limitations.

Beyond a series of short presentations of tools, state of the art of reproducibility in various domains and “war stories” of things not working, participants specifically explored ways forward to overcome barriers to the adoption of reproducibility. A series of break-out sessions gradually built on top of each other, (1) identifying different types of repeatability and their merits; (2) the actors involved and the incentives and barriers they face; (3) guidelines for actors (specifically editors, authors and reviewers) on how to determine the level of reproducibility of papers and the merits of reproduction papers; and (4) the specific challenges faced by user-oriented experimentation in Information Retrieval.

This led to the definition of according typologies and guidelines as well as identification of specific open research problems. We defined a set of actions to reach out to stakeholders, notably publishers and funding agencies as well as identifying follow-up liaison with various reproducibility task forces in different communities including the ACM, FORCE11, STM, Science Europe.

The key message resulting from this workshop, copied from and elaborated in more detail in Section 6.5 is:

Transparency, openness, and reproducibility are vital features of science. Scientists embrace these features as disciplinary norms and values, and it follows that they should be integrated into daily research activities. These practices give confidence in the work; help research as a whole to be conducted at a higher standard and be undertaken more

efficiently; provide verifiability and falsifiability; and encourage a community of mutual cooperation. They also lead to a valuable form of paper, namely, reports on evaluation and reproduction of prior work. Outcomes that others can build upon and use for their own research, whether a theoretical construct or a reproducible experimental result, form a foundation on which science can progress. Papers that are structured and presented in a manner that facilitates and encourages such post-publication evaluations benefit from increased impact, recognition, and citation rates.

Experience in computing research has demonstrated that a range of straightforward mechanisms can be employed to encourage authors to produce reproducible work. These include: requiring an explicit commitment to an intended level of provision of reproducible materials as a routine part of each paper's structure; requiring a detailed methods section; separating the refereeing of the paper's scientific contribution and its technical process; and explicitly encouraging the creation and reuse of open resources (data, or code, or both).

2 Table of Contents

Executive Summary

<i>Norbert Fuhr, Juliana Freire, and Andreas Rauber</i>	108
---	-----

Overview of Tools

noWorkflow	
<i>Vanessa Braganholo</i>	113
ReproMatch	
<i>Fernando Chirigati and Juliana Freire</i>	113
ReproZip: Computational Reproducibility With Ease	
<i>Fernando Chirigati and Juliana Freire</i>	114
Janiform: Intra-Document Analytics for Reproducible Research	
<i>Jens Dittrich</i>	114
DIRECT and LOD-DIRECT	
<i>Nicola Ferro</i>	115
Research Objects, FAIRDOM and SEEK4Science	
<i>Carole Goble</i>	116
Moore/Sloan Data Science Environments Projects	
<i>Randall J. LeVeque</i>	116
YesWorkflow	
<i>Bertram Ludäscher</i>	117
Process Migration Framework	
<i>Rudolf Mayer</i>	118
ROHub	
<i>Raul Antonio Palma de Leon</i>	118
TIRA	
<i>Martin Potthast and Benno Stein</i>	119
CodaLab	
<i>Evelyne Viegas</i>	120

State of the Art in Different Areas of CS

State of the art trade-offs in IR Research	
<i>Shane Culpepper</i>	120
Managing and Curating IR Experimental Data	
<i>Nicola Ferro</i>	120
Reproducibility in Databases	
<i>Juliana Freire, Fernando Chirigati, Jens Dittrich, and Tanu Malik</i>	122
Reproducibility using semantics: An overview	
<i>Daniel Garijo</i>	123
Reproducibility in Earth Science: aspects and ongoing work	
<i>Raul Antonio Palma de Leon</i>	124

Reproducible Data Sets in Dynamic Settings: Recommendations of the RDA Working Group on Dynamic Data Citation <i>Andreas Rauber</i>	124
Reproducibility in Visualization <i>Paul Rosenthal</i>	126
Research Data Alliance: State of the Art <i>Rainer Stotzka</i>	126
War Stories	
Reimplementation study “Who wrote the Web?” <i>Martin Potthast</i>	127
Repeatability in Computer Systems Research <i>Christian Collberg</i>	127
Working groups	
PRIMAD – Information gained by different types of reproducibility <i>Andreas Rauber, Vanessa Braganholo, Jens Dittrich, Nicola Ferro, Juliana Freire, Norbert Fuhr, Daniel Garijo, Carole Goble, Kalervo Järvelin, Bertram Ludäscher, Benno Stein, and Rainer Stotzka</i>	128
Reproducibility Tools and Services <i>Tanu Malik, Vanessa Braganholo, Fernando Chirigati, Rudolf Mayer, Raul A. Palma de Leon</i>	132
Taxonomy of Actions Toward Reproducibility <i>Martin Potthast, Fernando Chirigati, David De Roure, Rudolf Mayer, and Benno Stein</i>	135
Actors in Reproducibility <i>Justin Zobel, Shane Culpepper, David De Roure, Arjen P. de Vries, Carole Goble, Randall J. LeVeque, Mihai Lupu, Alistair Moffat, Kevin Page, and Paul Rosenthal</i>	138
Guidelines for Authors, Editors, Reviewers, and Program Committee Chairs <i>Alistair Moffat, Shane Culpepper, Arjen P. de Vries, Carole Goble, Randall J. LeVeque, Mihai Lupu, Andreas Rauber, and Justin Zobel</i>	143
What Makes A Reproducibility Paper Publishable <i>Alistair Moffat, Shane Culpepper, Arjen P. de Vries, Carole Goble, Randall J. LeVeque, Mihai Lupu, Andreas Rauber, and Justin Zobel</i>	146
Incentives and barriers to reproducibility: investments and returns <i>Paul Rosenthal, Rudolf Mayer, Kevin Page, Rainer Stotzka, and Evelyne Viegas</i>	148
User Studies in IR <i>Nicola Ferro, Norbert Fuhr, Kalervo Järvelin, Noriko Kando, and Matthias Lippold</i>	152
Open problems	
Open Research Problems in Reproducibility <i>Carole Goble and Daniel Garijo</i>	157
Participants	159

3 Overview of Tools

3.1 noWorkflow

Vanessa Braganholo (Fluminense Federal University, BR)

License © Creative Commons BY 3.0 Unported license
© Vanessa Braganholo

Main reference L. G. P. Murta, V. Braganholo, F. S. Chirigati, D. Koop, J. Freire, “noWorkflow: Capturing and Analyzing Provenance of Scripts” in Proc. of the 5th Int’l Provenance and Annotation Workshop (IPAW’14), LNCS, Vol. 8628, pp. 71–83, Springer, 2014.

URL http://dx.doi.org/10.1007/978-3-319-16462-5_6

URL <https://github.com/gems-uff/noworkflow>

Capturing provenance in scientific experiments has been a major concern both for result comprehension and reproducibility. Although the scientific community often writes experiments using script languages, most of the existing provenance capture approaches require scientists to change the way they work, by wrapping their experiments in scientific workflow systems, installing version control systems, or modifying and instrumenting their scripts, which may be laborious and error prone. As a solution to these problems, noWorkflow is a non-intrusive tool that transparently captures provenance of scripts, keeping data about how they evolve over time, as well as about their execution. It systematically monitors script execution without requiring any modification to the source code. Provenance data can then be analyzed using graphical interfaces, SQL or Prolog queries. We also provide ways of comparing two different executions, highlighting their differences, and support Jupyter notebooks¹.

3.2 ReproMatch

Fernando Chirigati (NYU Tandon School of Engineering, US) and Juliana Freire (New York University, US)

License © Creative Commons BY 3.0 Unported license
© Fernando Chirigati and Juliana Freire

Joint work of Fernando Chirigati, Tommy Ellqvist, Juliana Freire

URL <http://repromatch.poly.edu/>

ReproMatch stands for Reproducibility Match and it was designed as a search engine to help you find the tool (or tools) that best matches your reproducibility needs. The tools in the ReproMatch catalog are classified according to different reproducibility tasks, which we organized in a taxonomy². Researchers can submit information about new tools, or corrections to existing information.

¹ <http://jupyter.org/>

² <http://repromatch.poly.edu/task-descriptions/>

3.3 ReproZip: Computational Reproducibility With Ease

Fernando Chirigati (NYU Tandon School of Engineering, US) and Juliana Freire (New York University, US)

License © Creative Commons BY 3.0 Unported license

© Fernando Chirigati and Juliana Freire

Joint work of Fernando Chirigati, Rémi Rampin, Juliana Freire, Dennis Shasha

Main reference F. Chirigati, R. Rampin, D. Shasha, J. Freire, “ReproZip: Computational Reproducibility With Ease”, in Proc. of the 2016 ACM SIGMOD Int’l Conf. on Management of Data (SIGMOD’16), Demo Session, pp. 2085–2088, ACM, 2016

URL <http://dx.doi.org/10.1145/2882903.2899401>

URL <https://vida-nyu.github.io/reprozip/>

ReproZip provides a lightweight solution that makes experiments reproducible without forethought. Researchers can create an experiment without thinking about reproducibility and use ReproZip to make it reproducible and portable to other machines. In a nutshell, ReproZip automatically and transparently captures the provenance of an existing experiment by tracing system calls, and uses this information to create a lightweight reproducible package that includes only the required files needed for its reproduction. It also adds important features and contributions, including: (1) portability – ReproZip provides unpackers that allow researchers to automatically create a VM or a Docker container encompassing the experiment, thus allowing it to be reproduced in different operating systems; it also generates a workflow specification for the experiment, which can be used to easily change parameters or modify the original dataflow; (2) extensibility – by implementing new unpackers, researchers can easily extend ReproZip to port experiments to other environments and systems while keeping compatibility with existing packaged experiments; (3) modifiability – ReproZip automatically identifies input files, parameters, and output files, allowing researchers to easily modify these for reuse purposes; and (4) usability – researchers have control over the collected trace and can customize the reproducible package; the tool also provides command-line interfaces that make it easier to setup, reproduce, and modify the original experiment.

ReproZip has been recommended for the SIGMOD Reproducibility Review³, and listed on the Artifact Evaluation Process guidelines⁴.

3.4 Janiform: Intra-Document Analytics for Reproducible Research

Jens Dittrich (Universität des Saarlandes, DE)

License © Creative Commons BY 3.0 Unported license

© Jens Dittrich

URL <https://github.com/uds-datalab/PDBF>

Peer-reviewed publication of research papers is a cornerstone of science. However, one of the many issues of our publication culture is that our publications only publish a summary of the final result of a long project. This means that we put well-polished graphs describing (some) of our experimental results into our publications. However, the algorithms, input datasets, benchmarks, raw result datasets, as well as scripts that were used to produce the graphs in the first place are rarely published and typically not available to other researchers. Often they are only available when personally asking the authors. In many cases, however,

³ <http://db-reproducibility.seas.harvard.edu/>

⁴ <http://www.artifact-eval.org/guidelines.html>

they are not available at all. This means from a long workflow that led to producing a graph for a research paper, we only publish the final result rather than the entire workflow. This is unfortunate and has been criticized in various scientific communities. In this demo we argue that one part of the problem is our dated view on what a “document” and hence “a publication” is, should, and can be. As a remedy, we introduce portable database files (PDbF). These files are jani-form, i.e. they are at the same time a standard static pdf as well as a highly dynamic (offline) HTML-document. PDbFs allow you to access the raw data behind a graph, perform OLAP-style analysis, and reproduce your own graphs from the raw data – all of this within a portable document. We demo a tool allowing you to create PDbFs smoothly from within LATEX. This tool allows you to preserve the workflow of raw measurement data to its final graphical output through all processing steps. Notice that this pdf already showcases our technology: rename this file to “.html” and see what happens (currently we support the desktop versions of Firefox, Chrome, and Safari). But please: do not try to rename this file to “.ova” and mount it in VirtualBox.

3.5 DIRECT and LOD-DIRECT

Nicola Ferro (University of Padova, IT)

License © Creative Commons BY 3.0 Unported license
© Nicola Ferro

Main reference M. Agosti, N. Ferro, “Towards an Evaluation Infrastructure for DL Performance Evaluation”, in G. Tsakonas, C. Papatheodorou (eds.), “Evaluation of Digital Libraries: An Insight to Useful Applications and Methods,” Chandos Publishing, Oxford, 2009.

Distributed Information Retrieval Evaluation Campaign Tool (DIRECT⁵) is a system which models IR experimental data and manages all the steps of an IR evaluation campaign, like creation of the topics, submission of system runs, creation of relevance judgements, computation of performance measures and so on. DIRECT not only supports IR evaluation campaigns but takes also care of archiving the IR experimental data in order to make the accessible and referenceable for future re-use. At the time of writing, DIRECT counts about 35 millions documents, 14 thousands topics, around 4 million relevance judgements, 5 thousands experiments and 20 millions measures. This data has been inserted and used by about 1,500 researchers from more than 70 countries world-wide. Overall, DIRECT counts around 650 visitors who accessed and downloaded the data. LOD-DIRECT⁶ is an evolution of DIRECT to model and make available a subset of its IR experimental data as Linked Open Data.

⁵ <http://direct.dei.unipd.it/>

⁶ <http://lod-direct.dei.unipd.it/>

3.6 Research Objects, FAIRDOM and SEEK4Science

Carole Goble (University of Manchester, GB)

License © Creative Commons BY 3.0 Unported license
© Carole Goble

Main reference S. Bechhofer, J. Ainsworth, J. Bhagat, I. Buchan, P. Couch, D. Cruickshank, M. Delderfield, I. Dunlop, M. Gamble, C. Goble, D. Michaelides, P. Missier, S. Owen, D. Newman, D. De Roure, S. Sufi, “Why Linked Data is Not Enough for Scientists”, in *Future Generation Computer Systems*, Vol. 29(2):599–611, 2013.

URL <http://dx.doi.org/10.1016/j.future.2011.08.004>

URL <http://www.researchobject.org/>

Making scientific experiments FAIR – findable, accessible, interoperable, reusable – is hard. To be reproducible means bundling, along with the narrative, the experimental methods, computational codes, data, algorithms, workflows, scripts – some of which might be hosted remotely, in many different repositories and with the potential to change. In this talk I presented a framework for Research Objects⁷ – a metadata framework for bundling, porting and linking resources and representing the context of experiments. Research Objects have a manifest and a container. The manifest uses off the shelf standards and ontologies to construct the manifest and describe the content held in a container. The description is tailored to the type of Research Object, for example a Systems Biology Experiment or a computational workflow. The description broadly covers provenance, dependencies, versioning and checklists (aka reporting guidelines). Containers are off the shelf packaging platforms like Zip, Docker, Bagit or bespoke platforms that are “RO native”.

In the talk I presented FAIRDOMHub⁸, a Systems Biology Commons for supporting the reporting and sharing of models, data and Standard Operating Procedures arising from projects. It is built on the RO-compliant SEEK4Science⁹ commons and cataloguing platform. The system gathers the metadata needed for reproducible modelling. Moreover it supports the packaging up of content to be exported and deposited into other repositories like Zenodo.

Finally I presented other implementations of the RO framework: the COMBINE Archive for Systems Biology models which uses zip, Workflow RO bundles using Bagit, which is part of the Common Workflow Language, the STELAR Asthma eLab which uses Docker and ATLAS LHC Experiments, which uses Docker and CDE.

3.7 Moore/Sloan Data Science Environments Projects

Randall J. LeVeque (University of Washington – Seattle, US)

License © Creative Commons BY 3.0 Unported license
© Randall J. LeVeque

URL <https://reproduciblescience.org>

The Moore and Sloan Foundations are funding a joint project between the University of Washington, NYU, and Berkeley on creating better environments for data science research within academics. There is a joint working group on Reproducibility and Open Science¹⁰ that is developing several tools of possible interest. This repository¹¹ contains short descriptions

⁷ <http://www.researchobject.org/>

⁸ <http://www.fairdomhub.org/>

⁹ <http://www.seek4science.org/>

¹⁰ <https://reproduciblescience.org/>

¹¹ <https://github.com/BIDS/repro-case-studies/tree/submissions/case-studies>

and diagrams of workflows as examples of how researchers from many different disciplines have approached collaboration, data management, and sharing of code and data. I also gave a status report on a project to develop a system of badges to acknowledge the steps people take to make their work open and reproducible¹², and as means to collect links of examples others can follow. The main goals are to provide incentives and education about what is possible.

3.8 YesWorkflow

Bertram Ludäscher (University of Illinois at Urbana-Champaign, US)

License © Creative Commons BY 3.0 Unported license
© Bertram Ludäscher

Main reference T. McPhillips, T. Song, T. Kolisnik, S. Aulenbach, K. Belhajjame, R. K. Bocinsky, Y. Cao, J. Cheney, F. Chirigati, S. Dey, J. Freire, C. Jones, J. Hanken, K. W. Kintigh, T. A. Kohler, D. Koop, J. A. Macklin, P. Missier, M. Schildhauer, C. Schwalm, Y. Wei, M. Bieda, B. Ludäscher, “YesWorkflow: A User-Oriented, Language-Independent Tool for Recovering Workflow Information from Scripts”, in *Int’l Journal of Digital Curation*, 10(1):298–313, 2015; pre-print available as arXiv:1502.02403v1 [cs.SE].

URL <http://dx.doi.org/10.2218/ijdc.v10i1.370>

URL <http://arxiv.org/abs/1502.02403v1>

URL <http://yesworkflow.org/>

Traditional workflow automation approaches are based on scripting languages and remain very popular, e.g., due to the availability of countless libraries, a gentle learning curve (e.g. for Python), and – last not least – the high productivity that users of scripting languages experience. New interactive environments such as iPython/Jupyter add further to the popularity of script-based approaches. The YesWorkflow toolkit aims to bring some of the advantages of scientific workflow systems to researchers who use scripting languages such as Python, R, or Matlab. YesWorkflow enables script writers to reveal the computational steps and flow of data within the scripts they write (i.e., prospective provenance) by annotating their code with special comments. YesWorkflow extracts and analyzes these comments, represents the scripts in terms of entities based on a typical scientific workflow model, and provides graphical renderings of this view of the scripts. YesWorkflow additionally enables researchers to reconstruct retrospective provenance of data products used by scripts, and to query both prospective and retrospective provenance, allowing users powerful insights into their script-based models and simulation runs.

¹²<http://uwescience.github.io/reproducible/badges.html>

3.9 Process Migration Framework

Rudolf Mayer (SBA Research – Wien, AT)

License © Creative Commons BY 3.0 Unported license
© Rudolf Mayer

Main reference A. Rauber, T. Miksa, R. Mayer, S. Proell, “Repeatability and Re-Usability in Scientific Processes: Process Context, Data Identification and Verification”, in Proc. of the 17th Int’l Conf. on Data Analytics and Management in Data Intensive Domains (DAMDID’15), CEUR Workshop Proceedings, Vol. 1536, pp. 246–256, 2015.

URL <http://ceur-ws.org/Vol-1536/paper33.pdf>

The process migration framework (PMF) aims to make a process such an e-Science experiment repeatable, by extracting the process from its original environment, and enabling to redeploy it in a dedicated virtual machine. To this end, PMF logs the resources a process is utilising during execution time. Based on these logs, required files (executables and data) are identified and copied to the new environment, where the process can be run again. In addition, the PMF creates a higher-level semantic description of the process. Based on the execution trace of the process, the PMF also creates a basic process model visualising the sequence of steps identified.

By further analysing and aggregating the identified resources to for example Linux software packages, the PMF creates a smaller and more human-readable description of the process dependencies.

Finally, specific emphasize is put on identifying and discovering resources external to the original system, such as calls to web services e.g. for data processing, or the connection to a database server.

This analysis informs the user on resources that are outside of his direct control, and on which manual emphasize on ensuring the long-term availability needs to be put on.

3.10 ROHub

Raul Antonio Palma de Leon (Poznan Supercomputing and Networking Center, PL)

License © Creative Commons BY 3.0 Unported license
© Raul Antonio Palma de Leon

Main reference R. Palma, O. Corcho, J. M. Gómez-Pérez, C. Mazurek, “ROHub – A Digital Library of Research Objects Supporting Scientists Towards Reproducible Science”, in Semantic Web Evaluation Challenge at ESWC’14, Communications in Computer and Information Science, Vol. 475, pp. 77–82, 2014.

URL http://dx.doi.org/10.1007/978-3-319-12024-9_9

URL <http://www.rohub.org>

ROHub is a digital library system supporting the storage, lifecycle management, sharing and preservation of research findings via Research Objects. It includes different features to help scientists throughout the research lifecycle: (i) to create and maintain ROs that are compliant with predefined quality requirements so that they can be interpreted and reproduced in the future; (ii) to collaborate along this process; (iii) to publish and search these objects and their associated metadata; (iv) to manage their evolution; and (v) to monitor and preserve them through time ensuring that they will remain accessible and reusable. ROHub has a modular structure, comprising a backend (rod1) that exposes a set of RESTful APIs and a SPARQL endpoint; and a frontend Web Portal providing a graphical interface for end users (scientists/researchers/etc.)

3.11 TIRA

Martin Potthast (Bauhaus-Universität Weimar, DE) and Benno Stein (Bauhaus-Universität Weimar, DE)

License © Creative Commons BY 3.0 Unported license

© Martin Potthast and Benno Stein

Main reference A. Hanbury, H. Müller, K. Balog, T. Brodt, G. V. Cormack, I. Eggel, T. Gollub, F. Hopfgartner, J. Kalpathy-Cramer, N. Kando, A. Krithara, J. Lin, S. Mercer, M. Potthast, “Evaluation-as-a-Service: Overview and Outlook”, arXiv:1512.07454v1 [cs.CY].

URL <http://arxiv.org/abs/1512.07454v1>

URL <http://www.tira.io>

The TIRA experimentation platform is a web service that supports organizers of shared tasks in computer science to accept the submission of executable software [1]. TIRA automates software submission to a point at which it imposes no significant overhead on organizers and participants alike. From the start, TIRA has been in active use: since 2012, TIRA is employed for the PAN shared task series on digital text forensics [2], and as of 2015, TIRA hosts the annual shared task of the CoNLL conference. TIRA’s technology stack relies primarily on a combination of low-level (LXC, Docker) and high-level (hypervisor) virtualization technology, server-side control software, and a web front end that allow for the remote management of shared tasks. TIRA distributes virtual machines across a number of TIRA hosts, which are remote-controlled by a master server. Every virtual machine is accessible from the outside by participants via SSH and remote desktop, and both Linux and Windows are supported as guest operating systems. This allows for a variety of development environments, so that participants in a shared task can directly work as they usually would. TIRA further hosts the datasets used in a shared task, split into training datasets and test datasets. The former are publicly visible to participants, including ground truth data, whereas the latter are accessible only to participant software in a secure execution environment that protects the test datasets from leaking to participants. Before executing the software on a test dataset, TIRA clones its virtual machine into the secure execution environment, where Internet access is disabled. After the software successfully executed on the test dataset, its output is copied, whereas the cloned virtual machine is deleted to prevent any potentially private files on its virtual hard disk from exiting the execution environment. In this way, participants in a shared task can run their software on the shared task’s test datasets, whereas its organizers need not worry about the data leaking. TIRA also enables the use of proprietary and sensitive data as evaluation data. Finally, TIRA hosts a special purpose virtual machine for each shared task, where the organizer deploys software for performance measurement. The output of participant software that was executed on a training dataset or a test dataset is fed directly into the performance measurement software at the click of a button. The results are displayed on a dedicated web page for the shared task on TIRA’s web front end.

References

- 1 Tim Gollub, Benno Stein, and Steven Burrows. Ousting Ivory Tower Research: Towards a Web Framework for Providing Experiments as a Service. In Bill Hersh, Jamie Callan, Yoelle Maarek, and Mark Sanderson, editors, 35th International ACM Conference on Research and Development in Information Retrieval (SIGIR 12), pages 1125–1126, August 2012. ACM.
- 2 Martin Potthast, Tim Gollub, Francisco Rangel, Paolo Rosso, Efstathios Stamatatos, and Benno Stein. Improving the Reproducibility of PAN’s Shared Tasks: Plagiarism Detection, Author Identification, and Author Profiling. In Evangelos Kanoulas et al, editors, Information Access Evaluation meets Multilinguality, Multimodality, and Visualization. 5th International Conference of the CLEF Initiative (CLEF 14), pages 268–299, Berlin Heidelberg New York, September 2014. Springer.

3.12 CodaLab

Evelyne Viegas (Microsoft Research – Redmond, US)

License  Creative Commons BY 3.0 Unported license
© Evelyne Viegas

Main reference <https://github.com/codalab>

CodaLab is an open source platform which goal is to accelerate the rate of research by enabling collaboration among researchers and scientists across disciplines and make science truly reproducible. CodaLab Worksheets¹³ focuses on accelerating data-driven research and making it more sound while enabling scientists to publish their research as executables papers with full provenance on data and code. CodaLab competitions¹⁴ is a powerful framework for running data-driven competitions that involve result and/or code submission. Users can either participate in an existing competition or host a new competition as an organiser. CodaLab enables coopetitions, a new collaborative framework where users with different expertise can work together in a new environment favouring cross-pollination of ideas.

4 State of the Art in Different Areas of CS

4.1 State of the art trade-offs in IR Research

Shane Culpepper (RMIT University – Melbourne, AU)

License  Creative Commons BY 3.0 Unported license
© Shane Culpepper

URL <https://github.com/lintool/IR-Reproducibility>

This talk briefly presented a state-of-the-art comparison of ad-hoc search engines for a common TREC task. By aggregating results from the IR Reproducibility Challenge in the 2015 ACM SIGIR Workshop on Reproducibility, Inexplicability, and Generalizability of Results (RIGOR), we contrast fully reproducible baseline runs and “best known” submissions from the TREC Adhoc Search Task between 2004–2006.

4.2 Managing and Curating IR Experimental Data

Nicola Ferro (University of Padova, IT)

License  Creative Commons BY 3.0 Unported license
© Nicola Ferro

Information Retrieval (IR) is a discipline deeply rooted in experimentation since its inception and, over the time, it has developed robust and shared methodologies for conducting experiments, relying on the so-called Cranfield Paradigm. In particular, the adoption of large-scale and shared experimental collections, typically used in international evaluation

¹³ <https://github.com/codalab/codalab-worksheets/wiki>

¹⁴ <https://competitions.codalab.org/>

campaigns like TREC¹⁵, CLEF¹⁶, and NTCIR¹⁷ and then available for further re-use by the community, provide the means for running comparable experiments. This experimental paradigm gives rise to three targets for reproducibility:

- **experimental collections:** they consist of documents, topics, which surrogate real user information needs, and relevance judgements, which determine which documents are relevant to which topics. Experimental collections are an integral part of the experimental design and they are often used for many different purposes after their creation. It is thus important to understand their limitations and their generalizability as well as to reproduce the process that led to their creation. This is not always trivial since, for example, topics may be sampled from real system logs or relevance judgements are made by humans and, more and more often, using crowdsourcing.
- **system runs:** they are the most common target for reproducibility since they are what is discussed in papers proposing new methods and algorithms.
- **meta-evaluation experiments:** IR has a strong tradition in assessing its own evaluation methodologies, such as robustness of the experimental collections, reliability of the adopted evaluation measures or appropriateness of the adopted statistical analysis methods. All these investigations strongly rely on existing experimental collections and gathered systems runs and their reproducibility should be a key concern, since they probe our own experimental methods.

All the above mentioned three targets for reproducibility heavily depend on experimental data. Unfortunately, even if IR has a long tradition in ensuring the due scientific rigor is guaranteed in producing such data, it has not a similar tradition in managing and taking care of such valuable data. There currently are several barriers to proper data curation for reproducibility. There is a lack of common formats for modelling and describing the experimental data as well as almost no metadata (descriptive, administrative, copyright, etc.) for annotating and enriching them. The semantics of the data themselves is often not explicit and it is demanded to the scripts typically used for processing them, which are often not well documented, rely on rigid assumptions on the data format or even on side effects in processing the data. Finally, IR lacks a commonly agreed mechanism for citing and linking data to the papers describing them.

All these issues may be addressed by adapting solutions developed in other fields with similar problems but the biggest issue is the community itself, which would need to evolve its experimental methodologies to take into account reproducibility and the actions needed to guarantee it. This calls for an orchestrated effort and a cultural change which are the most compelling challenges towards a proper management and curation of experimental data.

¹⁵ <http://trec.nist.gov/>

¹⁶ <http://www.clef-initiative.eu/>

¹⁷ <http://research.nii.ac.jp/ntcir/index-en.html>

4.3 Reproducibility in Databases

Juliana Freire (New York University, US), Fernando Chirigati (NYU Tandon School of Engineering, US), Jens Dittrich (Universität des Saarlandes, DE), and Tanu Malik (University of Chicago, US)

License © Creative Commons BY 3.0 Unported license
© Juliana Freire, Fernando Chirigati, Jens Dittrich, and Tanu Malik

While some authors in the community make their code available, reproducibility has not been widely adopted. In 2008, SIGMOD instituted a reproducibility review: authors of accepted papers are invited to submit their experiments for an independent review. Over the years, the rate of participation has varied, usually fewer than 35% of the accepted papers were evaluated for reproducibility. Authors have argued that making papers reproducible requires too much work. Reviewers have faced many challenges to install and run the experiments, due to incomplete instructions, missing dependencies, incompatible operating system [1]. Most of the reproducibility submissions include source code and data, together with instructions on how to build them. However, it is often the case that some of them fail the set-up phase because dependencies are not made explicit, which puts a high burden on reviewers; few submissions include a VM and a workflow. Authors complained that the process requires too much work for the benefit derived – and the process does require a substantial amount of work if there is no planning for reproducibility since the beginning of the project. The SIGMOD reproducibility review was revamped in 2015¹⁸ by Stratos Idreos, and ACM now allows papers to be marked as reproducible in ACM Digital Library. Elsevier’s Information Systems Journal now also has a Reproducibility Section, led by Dennis Shasha and Fernando Chirigati, where some accepted papers are invited to submit a reproducibility paper, explaining in detail how to run it and what the effort was to make it reproducible. The paper is submitted together with source code (GitHub), data (GitHub or Mendeley Data), and a Docker container/ReproZip package/VM (Mendeley Data) to ease the review process. Reviewers also become co-authors of the report, as they describe in the paper the efforts in reproducing the published experiment. This approach provides incentives for both authors and reviewers: by making their experiment reproducible, authors have a new paper, and by having to review and reproduce the experiment, reviewers are also included in the same publication [2, 3].

Some technical challenges must still be solved to help databases experiments to be reproduced. First, it is still unclear how to reproduce experiments that access networked resources, including Web services, remote databases, and HDFS. Tools such as LDV (Lightweight Database Virtualization) are a step towards this, but must be made more general to broaden the adoption, since currently users must use Postgres. A second challenge is how to enable reproducibility for distributed applications (e.g.: MPI, Hadoop, Spark). Many more variables and configurations are involved, and performance results are often important. Using a different compiler, compilation flags, or architecture may change these results. There are reproducibility challenges in data integration and data analysis, where in data from a large variety of sources are either aggregated into a database and/or modified/analyzed respectively along the process. Here reproducibility tools for database must interact with file-system tools to ensure cross-system reproducibility.

A noteworthy effort in the database community is the inauguration of the experiments&analyses track at VLDB 2008. This is a special conference track allowing researchers

¹⁸ <http://db-reproducibility.seas.harvard.edu/>

to submit experimental studies, rebuttals as well as negative results. Accepted papers on this track are not distinguished from standard “research” papers in the final conference program and the proceedings. Papers on E&A may run code used in other papers as blackboxes. However, in general, also often whiteboxing is done (in the sense of algorithm and implementation analysis as well as re-implementations) to make the experimental comparison and the comparability of algorithms and systems stronger. The E&A track has become quite popular in recent years. So far at least two best paper awards have been given to E&A track papers.

References

- 1 P. Bonnet, S. Manegold, M. Bjørling, W. Cao, J. Gonzalez, J. Granados, N. Hall, S. Idreos, M. Ivanova, R. Johnson, D. Koop, T. Kraska, R. Müller, D. Olteanu, P. Papotti, C. Reilly, D. Tsirogiannis, C. Yu, J. Freire, and D. Shasha. Repeatability and workability evaluation of sigmod 2011. *SIGMOD Record*, 40(2):45–48, 2011.
- 2 A. Wolke, M. Bichler, F. Chirigati, V. Steeves. Reproducible experiments on dynamic resource allocation in cloud data centers *Information Systems*, Available online 7 January 2016.
- 3 A. Wolke, B. Tsend-Ayush, C. Pfeiffer, M. Bichler. More than bin packing: Dynamic resource allocation strategies in cloud data centers *Information Systems*, Volume 52, August–September 2015, Pages 83–95

4.4 Reproducibility using semantics: An overview

Daniel Garijo (Technical University of Madrid, ES)

License  Creative Commons BY 3.0 Unported license

© Daniel Garijo

URL <http://www.slideshare.net/dgarijo/reproducibility-using-semantics-an-overview>

The Semantic Web has helped to create knowledge bases that link and facilitate accessibility to research data. However, how can it help scientists to make their experiments reproducible? This talks introduces an overview of the different initiatives led by the Ontology Engineering Group (UPM) to address reproducibility by using semantics. The initiatives are distributed among several disciplines, including the formalization of laboratory protocols to detect ambiguity and missing descriptions [1], documentation and publication of scientific workflows and their resources [2], capturing the infrastructure needed to reproduce a scientific experiment [3], achieving long term preservation of research objects and conditional access to resources based on their intellectual property rights¹⁹.

References

- 1 O. Giraldo , A. García, J. Figueredo and O. Corcho. Using Semantics and NLP in Experimental Protocols. 8th Semantic Web Applications and Tools for Life Sciences International Conference (SWAT4LS 2015). Cambridge, UK. 2015
- 2 D.Garijo and Y. Gil. A new approach for publishing workflows: abstractions, standards, and linked data. *Proceedings of the 6th workshop on Workflows in support of large-scale science (WORKS11)*, pp 47–5, Seattle, 2011.
- 3 I. Santana-Perez, R. Ferreira da Silva, M. Rynge, E. Deelman, M. S. Pérez-Hernández and O. Corcho. Reproducibility of execution environments in computational science using Semantics and Clouds. *Future Generation Computer Systems*, 2016.

¹⁹<http://licensius.com/>

4.5 Reproducibility in Earth Science: aspects and ongoing work

Raul Antonio Palma de Leon (Poznan Supercomputing and Networking Center, PL)

License © Creative Commons BY 3.0 Unported license
© Raul Antonio Palma de Leon

URL <http://www.slideshare.net/rapw3k/aspects-of-reproducibility-in-earth-science>

The “Earth Science Research and Information Lifecycle” can be regarded as a continuous, iterative and ongoing process used by scientists for conducting, validating and disseminating scientific knowledge. It can undergo an unlimited number of iterations that lead to the development of new and innovative ideas, concepts, techniques and technologies, which ultimately benefit both science and society. The life cycle can be briefly summarized into four main phases that involve multiple categories of stakeholders: (i) scientists access information and (usually) share results; (ii) shared results and information are analysed and interpretative models are generated and discussed with other colleagues; (iii) discussions lead to novel ideas and concepts which might need validation through further experimentation or data acquisition; (iv) new results are validated and shared so that other scientists can access them and start the process again.

This presentation introduces the ongoing work of the EU project EVEREST that aims at establishing a Virtual Research Environment (VRE) e-infrastructure for Earth Science. The VRE is being validated in four communities: sea monitoring, natural hazards, land monitoring and supersites, and is applying the Research Objects concepts and technologies as the mean for sharing information and establish more effective collaboration in the VRE. Regarding the reproducibility in their domain, they have a slightly different vision as other disciplines like experimental science that often aims at testing a hypothesis. For instance Supersite community can be described as an historical science that is mostly based on past observations. For such community, the main goals involve measuring geophysical parameters in the natural environment, derive information on the effects of the phenomena, model this information to generate space/time representations and provide these representations to risk management and other relevant stakeholders, and only complementary scientists may use this information to develop theories or confirm hypothesis. Hence, in such communities, reproducibility is mainly concerned about the execution of common or community-agreed workflows for data analysis and modelling, and for testing algorithms and data products. Nevertheless, there are still several limitations for achieving reproducibility in these communities: they are not yet using formalised (computational) workflows, the data necessary is not always available or known, workflows usually require considerable human intervention, etc. These are some of the challenges currently being addressed in EVEREST.

4.6 Reproducible Data Sets in Dynamic Settings: Recommendations of the RDA Working Group on Dynamic Data Citation

Andreas Rauber (TU Wien, AT)

License © Creative Commons BY 3.0 Unported license
© Andreas Rauber

URL https://rd-alliance.org/system/files/documents/RDA-DC-Recommendations_151020.pdf

One key requirement to enable repeatability in data-driven processes is the ability to specify precisely the data that was used as input, and to be able to re-create this identical input data for any re-execution of an analysis. This proves challenging due to two reasons:

1. granularity of the data: Databases contain massive amounts of data, of which specific subsets are selected and used for analysis. To enable reproducibility or comparability of results we need to be able to specify precisely, which subset was extracted from a larger dataset. Providing a verbal description of the extraction process (i.e. specifying the rows and columns selected and filter criteria used, or describing an arbitrary geographic region in natural language is error-prone and requires significant effort to recreate exactly the same data set again. Another solution encountered frequently, i.e. storing a backup dump of each subset used, does not scale and leads to massive data management overheads.
2. Dynamic data: In many settings, the data available for analysis changes, by new data being added continuously, erroneous data being deleted or corrected. Again, we need to ensure that we can obtain earlier “versions” of data as used in a study to enable repeatable and comparable results. Solutions such as artificially defining e.g. annual batch releases of data delay the availability of current data and again lead to massive overhead in storing duplicate batches of unchanged data.

The Working Group on Dynamic Data Citation of the Research Data Alliance (RDA WGDC) has elaborated a set of recommendations [1] to solve these challenges. In a nutshell, the solution is based on (1) time-stamped and versioned data to ensure that earlier versions of data remain available, and (2) storing the queries used to select arbitrary subsets of data with a timestamp. A persistent identifier (PID, e.g. a DOI) is added to such a query together with additional metadata such as hash keys for fixity information, to ensure the time-stamped query can be re-executed against the time-stamped database to retrieve an identical subset. This approach allows retrieving the data both as it existed at a given point in time as well as the current view on it, by re-executing the same query with the stored or current timestamp, thus benefiting from all corrections made since the query was originally issued. This allows tracing changes of data sets over the time and comparing the effects on the result set. The query stored as a basis for identifying the data set provides valuable provenance information on the way the specific data set was constructed, thus being semantically more explicit than a mere data export. The query store also offers a valuable, central basis for analyzing data usage. Metadata such as checksums support the verification of correctness and authenticity of data sets retrieved.

The recommendations are applicable across different types of data representation and data characteristics (big or small data; static or highly dynamic; identifying single values or the entire data set). Pilot implementations have been used to evaluate this approach in different settings including data stored in relational databases (RDBMS, e.g. MySQL), XML databases (e.g. X-Base), and comma separated value files (CSV). More details are available from the homepage of the working Group²⁰.

References

- 1 Andreas Rauber, Ari Asmi, Dieter van Uytvanck and Stefan Pröll. Data Citation of Evolving Data: Recommendations of the Working Group on Data Citation (WGDC). Research Data Alliance, October 20 2015.

²⁰ <https://rd-alliance.org/node/141>

4.7 Reproducibility in Visualization

Paul Rosenthal (TU Chemnitz, DE)

License  Creative Commons BY 3.0 Unported license
© Paul Rosenthal

During the last decades, visualization has made its way towards a serious science. However, finalizing this procedure would also require that publications and results based upon a verifiable basis, i. e. that they are reproducible. Introducing a culture of reproducibility within this community would also increase the acceptance of visualization methods in other communities and applications, make contributions more well-grounded, and speed up the development of the community by making comparisons and advancements much easier.

In this talk, the recent efforts around the EuroRV³, the EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization²¹, which goes into its fourth year in 2016, were presented. The important role and the benefits of a strong culture of reproducibility have been discussed over the last years in the EuroRV³ workshops in detail. And there were only a few cases identified where very special requirements prevent the achieving of basic reproducibility at all. However, it was also observed that, in current publications, providing reproducibility is often limited to the fraction that is needed to be accepted to a venue. This seems to be due to the fact that reproducibility is not rewarded at all.

Consequently, efforts arise to encourage good reproducibility in future publications by introducing a badge of honor or a similar sign of appreciation for all papers fulfilling a set of criteria. The main goal is to point out and honor such work and strengthen the community in its effort to establish a common culture of reproducibility. The efforts in this direction are still in the phase of planning and negotiations.

4.8 Research Data Alliance: State of the Art

Rainer Stotzka (KIT – Karlsruher Institut für Technologie, DE)

License  Creative Commons BY 3.0 Unported license
© Rainer Stotzka
URL <https://rd-alliance.org>

The Research Data Alliance (RDA) is an international organization focused on the development of infrastructure and community activities that reduce barriers to data sharing and exchange, and the acceleration of data driven innovation worldwide. With more than 3,200 members globally representing more than 100 countries, RDA includes data science professionals from multiple disciplines, including but not limited to academia, library sciences, earth science, astronomy and meteorology. RDA is building the social and technical bridges that enable open sharing of data to achieve research reproducibility and transparency.

²¹ <http://www.eurorvvv.org/>

5 War Stories

5.1 Reimplementation study “Who wrote the Web?”

Martin Potthast (Bauhaus-Universität Weimar, DE)

License © Creative Commons BY 3.0 Unported license
© Martin Potthast

Main reference M. Potthast, S. Braun, T. Buz, F. Duffhauss, F. Friedrich, J. M. Güllow, J. Köhler, W. Löttsch, F. Müller, M. E. Müller, R. Paßmann, B. Reinke, L. Rettenmeier, T. Rometsch, T. Sommer, M. Träger, S. Wilhelm, B. Stein, E. Stamatatos, M. Hagen, “Who Wrote the Web? Revisiting Influential Author Identification Research Applicable to Information Retrieval”, in Proc. of the 38th European Conf. on IR Research – Advances in Information Retrieval (ECIR’16), LNCS, Vol. 9626, pp. 393–407, Springer, 2016; pre-print available from author’s webpage.

URL http://dx.doi.org/10.1007/978-3-319-30671-1_29

URL http://www.uni-weimar.de/medien/webis/publications/papers/stein_2016d.pdf

We revisited author identification research by conducting a new kind of large-scale reproducibility study: we selected 15 of the most influential papers for author identification and recruited a group of students to reimplement them from scratch. Since no open source implementations have been released for the selected papers to date, our public release will have a significant impact on researchers entering the field. This way, we lay the groundwork for integrating author identification with information retrieval to eventually scale the former to the web. Furthermore, we assess the reproducibility of all reimplemented papers in detail, and conduct the first comparative evaluation of all approaches on three well-known corpora.

5.2 Repeatability in Computer Systems Research

Christian Collberg (University of Arizona – Tucson, US)

License © Creative Commons BY 3.0 Unported license
© Christian Collberg

Main reference C. Collberg and T. Proebsting, “Repeatability in Computer Systems Research”, Communications of the ACM, Vol. 59(3):62–69, 2016.

URL <http://dx.doi.org/10.1145/2812803>

URL <http://repeatability.cs.arizona.edu/>

We give anecdotal as well as empirical evidence that Computer Systems researchers are generally unwilling or unable to share research artifacts (code and data) and that, when they do, their code often does not build. As a result, the minimum requirements for the reproducibility of applied Computer Systems research (that code used in experiments is available and that it builds) are generally not met.

We give an account of our own failed attempt at reproducing the results in a published research paper, a description of an empirical study into the reproducibility of the research published in 601 papers that appeared in the last two years in ACM publications, and a recommendation for how researchers, public funding agencies, and academic publishers can improve the reproducibility of Computer Science research.

6 Working groups

6.1 PRIMAD – Information gained by different types of reproducibility

Andreas Rauber (TU Wien, AT), Vanessa Braganholo (Fluminense Federal University, BR), Jens Dittrich (Universität des Saarlandes, DE), Nicola Ferro (University of Padova, IT), Juliana Freire (New York University, US), Norbert Fuhr (Universität Duisburg-Essen, DE), Daniel Garijo (Technical University of Madrid, ES), Carole Goble (University of Manchester, GB), Kalervo Järvelin (University of Tampere, FI), Bertram Ludäscher (University of Illinois at Urbana-Champaign, US), Benno Stein (Bauhaus-Universität Weimar, DE), and Rainer Stotzka (KIT – Karlsruher Institut für Technologie, DE)

License © Creative Commons BY 3.0 Unported license

© Andreas Rauber, Vanessa Braganholo, Jens Dittrich, Nicola Ferro, Juliana Freire, Norbert Fuhr, Daniel Garijo, Carole Goble, Kalervo Järvelin, Bertram Ludäscher, Benno Stein, and Rainer Stotzka

6.1.1 What is Reproducibility

What is “reproducibility” anyways? And how is it different from “repeatability”, “replicability”, or any of the other r-words? There are already a number of attempts at defining and sorting out these different notions. De Roure [1] lists 21 different r-words grouped into 6 categories, stating that reproducibility means reusing a research object with a change to some circumstances, inputs, resources or components in order to see if the same results are achieved independent of those changes. Often these notions are context-sensitive (e.g., validation vs verification have rather precise and very different meanings in different communities).

As an alternative approach to sort out terminological confusions, we attempted to look at a different perspective. When trying to reproduce a study, what are the things that are kept the same (e.g., the overall method or algorithm) and what is changed (e.g., the input data or implementation language, etc.)? More importantly, while changing these things, what information is gained by successfully reproducing (or failing to reproduce) a study?

6.1.2 The PRIMAD Model

As a starting point, we defined a preliminary list of “variables” that could potentially be changed:

- (R) or (O) Research Objectives / Goals
- (M) Methods / Algorithms
- (I) Implementation / Code / Source-Code
- (P) Platform / Execution Environment / Context
- (A) Actors / Persons
- (D) Data (input data and parameter values)

This spells: OMIPAD. Rearranging the letters that we use to represent the several aspects that can be changed, it can be remembered as PRIMAD: (P)latform, (R)esearch Goal, (I)mplementation, (M)ethod, (A)ctor, (D)ata (both input and parameter data), which allows us to ask: What variables have you “primed” in your reproducibility study?

As a concrete example of the meaning of these variables, let’s assume our (R)esearch objective is to sort a data set. We could use Quick Sort as the sorting (M)ethod (algorithm), which could be (I)mplemented as a script in Python and run over a Python 2.7 compiler on an iMac running MacOS 10 (and this would be the execution (P)latform). We could run this

over a specific (D)ataset (data.csv) using 0 as the pivot parameter. The (A)ctor, in this case, is the researcher that is executing the sorting. Summarizing:

- Research goal: sorting the input
- Method: quick sort
- Implementation: script in Python
- Platform: Python 2.7, MacOS, iMac, etc
- Input Data: the data that is to be sorted
- Parameter: the position of the pivot
- Actor: user that is executing the experiment

As a more concrete example, we can take Tandy Warnow’s statistically binning paper and the controversy around it²². In this case, the controversy was that her initial approach (we will call it method M, proposed by team T1) was claimed (by team T2) to be non-reproducible. More specifically, team T2 implemented method M and could not reproduce the original results obtained by team T1. So, in this example, we have the following scenario:

- Research Objective: Improve state-of-the-art in phylogenetic tree construction
- Method: Statistical binning (supposedly $M = M'$, but one side is arguing that $M \neq M'$)
- Implementation: two available, by team T1 and by the “opposing” team T2
- Platform: various (we suppose)
- (input) Data: different datasets – some arguments were made about the suitability here as well, since apparently team T2 did not respect some premises of how the input data should be organized.

To describe this reproducibility study in terms of these variables, only the research objective R and the method M are fixed; everything else is varied (team T1 actually argues that the implementation I2 isn’t of the method M, but of another method M’). To represent what changed, we use primed variables.

In this case, T1 argues: P’R’I’M’A’D’, while T2 argues P’R’I’M’A’D’ (variables with apostrophe were changed, and non-apostrophe variables were kept the same). Thus, both teams actually disagree on whether $M = M'$ or not!

6.1.3 Gains from different types of reproducibility

Reproducibility in its various forms, however, is never a goal in itself. We do it in order to gain something. By changing some (or several) of these variables, we gain different kind of knowledge. For example, if one keeps R, M, and I fixed, but varies the platform $P \rightarrow P'$, then the reproducibility study tests the portability, stability, or platform-independence of the experiment.

Figure 1 shows an attempt to categorize and label the various types of reproducibility and to summarize the gain they bring to a computational experiment. The precise terminology to use is still subject to further debate and no final agreement could be reached, specifically with respect to the labels and the mapping to the terminology found in the literature to describe different types of reproducibility. This may be partially due to the fact that many of the terms used describe repeatability settings refer to combinations of the above, e.g. to differentiate between obtaining a certain level of repeatability within the same lab or by an external lab. But even independent of this combinatorial issues the exact terminology proves to be difficult to agree upon already within a computing setting, not to mention beyond this domain.

²² https://youtu.be/-0jd0x7Kg90?list=PLO8UWE9gZT1AgHZPaxQbpUNY0T26zeL_f

Label	Data		Platform / Stack	Implementation	Method	Research Objective	Actor	Gain
	Parameters	Raw Data						
Repeat	-	-	-	-	-	-	-	Determinism
Param. Sweep	x	-	-	-	-	-	-	Robustness / Sensitivity
Generalize	(x)	x	-	-	-	-	-	Applicability across different settings
Port	-	-	x	-	-	-	-	Portability across platforms, flexibility
Re-code	-	-	(x)	x	-	-	-	Correctness of implementation, flexibility, adoption, efficiency
Validate	(x)	(x)	(x)	(x)	x	-	-	Correctness of hypothesis, validation via different approach
Re-use	-	-	-	-	-	x	-	Apply code in different settings, Re-purpose
Independent x (orthogonal)							x	Sufficiency of information, independent verification

■ **Figure 1** PRIMAD Model: Categorizing the various types of reproducibility by varying the (P)latform, (R)esearch Objective, (I)mplementation, (M)ethod, (A)ctor and (D)ata, analyzing the gain they bring to computational experiments. x denotes the variable primed i.e. changed, (x) a variable that may need to be changed as a consequence, whereas – denotes no change.

We now elaborate on the various aspects that can be changed, and how we could “label” reproducibility studies that use such combinations of changes.

1. ϵ equals to not changing anything, simply repeating an earlier experiment within the same computational environment, using the same code and data, allows to verify that the computed results are deterministically consistent. **Suggested label: [repeat]**
2. **Data → Parameters:** changing the parameter settings (e.g. parameter sweep, 10-fold cross-validation, etc.) allows to determine the: robustness/sensitivity of an experiment wrt. the specific parameters. Suggested label: **[rerun: robustness check, parameter sweep]**
3. **Data → Raw (Input) data:** changing the raw data processed by an experiment allows to verify how far the statements made hold across a larger part of the input space. Depending on the degree of similarity/difference in the input data, statements on the generality can be made. It also allows to evaluate whether the data originally used is representative/comparable for a given domain. **Suggested label: [rerun: check generality]**
4. **Platform:** changing the execution platform (i.e. the context, execution environment, including the software and hardware stack, i.e. a Java virtual Machine, running on a specific version of some operating system, within some hypervisor, running on specific HW) allows to test the platform independence/portability of an experiment. It may gain wider adoption or higher stability by being runnable on a wider range of platforms. **Suggested label: [port]**
5. **Implementation:** changing the implementation allows to verify the correctness of the previous implementation. It may also gain you higher efficiency, provide broader set of execution platforms, leading to higher adoption in different communities. Note that changing the implementation may incur a change of the execution platform. **Suggested label: [re-code]**

6. **Method:** changing the method allows to validate the correctness of a hypothesis using a different methodological approach. This provides a method-independent verification, or may provide a more efficient method to support the claims made. Note that a change in the method by definition will incur a change in the implementation, and possibly also of the execution platform. **Suggested label:** [validate]
7. **Research Objective:** changing the research objective (hypothesis) basically constitutes a re-purposing / re-use of an earlier experiment, allowing science to progress faster, opening new avenues for research. It requires trustworthy results/components to offer a solid basis. **Suggested label:** [repurpose]
8. **Actor:** changing the actor is orthogonal to all changes discussed above. It allows both independent verification of the characteristics, and also determines whether the information provided is sufficient to achieve such independent verification. **Suggested label:** [experimenter-independent <activity>]

Consistency: success or failure of a reproducibility study has to be evaluated wrt. the consistency of the outcomes. The criterion to apply thus is not whether the outcomes of priming any of the above variables leads to identical results, but whether results are consistent with the previous ones. Depending on the setting, this may require identity of results, but may also be lessened to consistency within certain error bounds or allow differences that are not statistically significant.

Transparency: Another dimension to be considered is transparency. It denotes the ability to look into all necessary components to be able to understand the path from the hypothesis to the results. While many of the changes above can be performed on a black-box level (repeating a run using binary code, performing the repeatability evaluation on a virtual machine provided by the original authors) it does not allow to make qualified inspections on the internal functioning on the respective levels. Thus, the degree of transparency should be used as a measure for the degree of inspection possible.

6.1.4 Variations on PRIMAD

After analyzing the various aspects that can be changed, we realized that using just one letter to represent both input data and parameters may not be enough. We are also aware of the fact that the differences between these attributes may not always be very clear-cut, as e.g. the fuzzy distinction between parameter and data to be supplied to an algorithm, or the boundary between an implementation and the execution platform becoming less clear-cut via the use of static or dynamically linked libraries. Yet, we find that the current set of variables helps in distinguishing core concepts and challenges to repeatable experiments relying on computation. Thus, we tried to identify possible letters we could use to represent each of the aspects we discussed:

- (O,R,G) Research Objectives / Goals
- (M,A) Methods / Algorithms
- (I,C,S) Implementation / Code / Source-Code
- (E,C) Platform / Execution Environment / Context
- (D,I,R) Input Data (“raw” data)
- (P) Parameter values
- (A) Actors / Persons

In the future, we may define a new acronym using these letters to better represent all the possible variations. Some possibilities are APDEIMO, PDEIMO, AOMIEDP, OMIEDPA,

OMIEPAD. We may also need a deeper analysis of the various attributes and their changes, seeing in how far these can be mapped to, first of all, the different definitions of types of reproducibility being used in different communities. Furthermore, with most scientific work today spanning several disciplines and crossing methodological boundaries we need to investigate, in how far the concept of fixing and changing various attributes can be applied in more general settings. However, while the precise labels being used may change, we have the feeling that having a precise definition and understanding of the attributes that are fixed or changed is essential to define the various types of reproducibility studies and, specifically, to understand the benefit we gain from them. Reproducibility is not a means to its own end. While showing deterministic results by simply repeating a computation without changing anything may already be an exciting fact in some settings we very likely will want to go beyond such basic settings of reproducibility studies, gaining deeper insights into scientific work and establishing trust in results, methods and tools for the benefit of science.

References

- 1 De Roure, D., (2014). The future of scholarly communications. *Insights*. 27(3), pp. 233–238.

6.2 Reproducibility Tools and Services

Tanu Malik (University of Chicago, US), Vanessa Braganholo (Fluminense Federal University, BR), Fernando Chirigati (NYU Tandon School of Engineering, US), Rudolf Mayer (SBA Research – Wien, AT), and Raul A. Palma de Leon (Poznan Supercomputing and Networking Center, PL)

License © Creative Commons BY 3.0 Unported license

© Tanu Malik, Vanessa Braganholo, Fernando Chirigati, Rudolf Mayer, Raul A. Palma de Leon

Sharing code and data increase reproducibility, but such sharing may not reflect the overall method, which is typically published in research papers. The current format of research papers (text-based) does not link code and data at finer granularity, the page-limit restricts detailed description of analyses and/or reporting of negative results, and authors have little motivation to describe in detail on a companion website. The consequence is built-up of scientific bias, which can be hard to break, given long cycles of publishing and funding.

Consequently, there is a critical need for reproducibility tools that, along with the changing culture of reproducibility, can also help researchers achieve the desired state of reproducibility in an efficient manner. However, before developing and/or applying a tool-suite to solve a reproducibility problem, several issues at hand must be understood. These range from:

1. **Precise identification of gaps in the research lifecycle.** A precise identification of gap in the research life-cycles is needed to understand which tool is applicable for solving the problem. Three gaps are often identified in the research lifecycle. The first one is related to the lack of motivation from researchers to apply reproducibility on their research. Better methods to incentivize reproducibility are needed, e.g.: having regulations and funding agencies to “force” the practice of reproducible research. A second possible gap is due to the poor linking between computational assets and text-based research outputs: there is rarely a connection between computational artifacts (research material, data, samples, software, models, methods, etc.) and the published results (paper and review process). This gap is very much discipline specific: some disciplines have developed standards on how to handle these artefacts and document the

procedures (e.g.: systems biology), while others have done less so (e.g.: computational sciences).

Last, a third possible gap is the lack of sufficient tools to help researchers do reproducible research according to their requirements, which, again, are domain-specific.

While the development of tools and services helps fill the last two gaps, the lack of motivation remains a barrier that must be addressed to broaden the adoption of reproducibility. It does not matter if we can build the most useful and easy-to-use reproducibility tools: unless there are proper incentives, these tools will be pointless.

2. **Domain-related issues.** In several domains, the term “reproducibility” is vaguely defined, and when defined, can substantially differ among communities. The main reason for such disparity is due to the fact that different domains have different requirements regarding reproducible research. For instance, while numerical analysis does not often handle large amounts of data, a tool or service constructed for the databases area must take into account how to share terabytes of data for reproducibility purposes. Therefore, a distinction must be made between common and discipline-specific reproducibility requirements. Also, different domains use different technologies (e.g.: programming languages, protocols, and types of data), which often influences the development of domain-specific tools.
3. **Are there tools that can solve the problem at hand?** It is unclear whether we need to build new tools to solve the existing problems, and the reason is twofold: first, we do not know what the problems are, as the requirements from different domains must be better understood; second, it is hard to know all the tools available and all the features they provide. Assuming there are sufficient tools, it may be a matter of just improving existing tools, either by creating new features or tackling new requirements. In addition, since different tools address different issues, a question that comes up is whether they are integrated enough with the existing environment/infrastructure to achieve the desired reproducibility.

6.2.1 Tool Landscape

The primary challenge for the user is to how to navigate the tool landscape with minimal effort but improved reproducibility. Thus, given L reproducibility levels, and N dimensions of assessing the reproducibility, the objective in tool landscape is to guide the user to move from tool A to tool B such that there is minimal effort but a gain in the reproducibility level along one or more dimension.

Reproducibility is a continuous process that is achieved over time, but in several cases it can be discretized to various levels that provide a different state of reproducibility of the experiment. For instance, if we do not change any of the PRIMAD attributes, then at the bare minimum we demonstrate determinism and consistent behavior. If the user moves from simple scripting – an environment in which the user is already satisfied – to a “workflow” environment (e.g.: YesWorkflow, noWorkflow, VisTrails, Taverna, Wings), then an effort is required to transition depending on the domains and experiment, while gaining both a demonstration of tool independence / correctness of the implementations, as well as higher portability or easier adoption / re-use in different settings (e.g.: transitioning from Python scripting to noWorkflow is straightforward). Packaging tools such as CDE, PTU, Docker, and RepoZip may significantly improve reproducibility with small effort.

6.2.2 Addressing the gap

First, we need to understand **what is needed**, rather than **what is possible**: not everything that can be done and developed is needed (and wanted) when it comes down to reuse and reproducibility. Therefore, it is of crucial importance to collect the different common and domain-specific requirements, and then recognize what we are missing. Related to the requirements, another important issue is knowing the target group that is interested in reproducibility, which helps determining the needs and requirements of appropriate tools and services.

Regarding the gap associated to the poor linking between computational resources and dissemination reports, papers, etc., the concept and approach of research objects may be one of the means to support scientists and the research community in filling this gap. Research objects (ROs) are aggregating objects that bundle together resources that are essential to a computational scientific study or investigation (data used/produced, methods applied, results, publications, people, etc.), along with semantic annotations on the bundle or the resources needed for the understanding and interpretation of the scientific outcomes, including provenance and evolution information, descriptions of the computational methods, dependency information and settings about the experiment execution. There is also a plethora of tools that can create executable papers, such as Janiform, Galaxy, and VisTrails. In addition, there is a set of literate programming tools that help linking documents to code and data (e.g.: Jupyter notebooks).

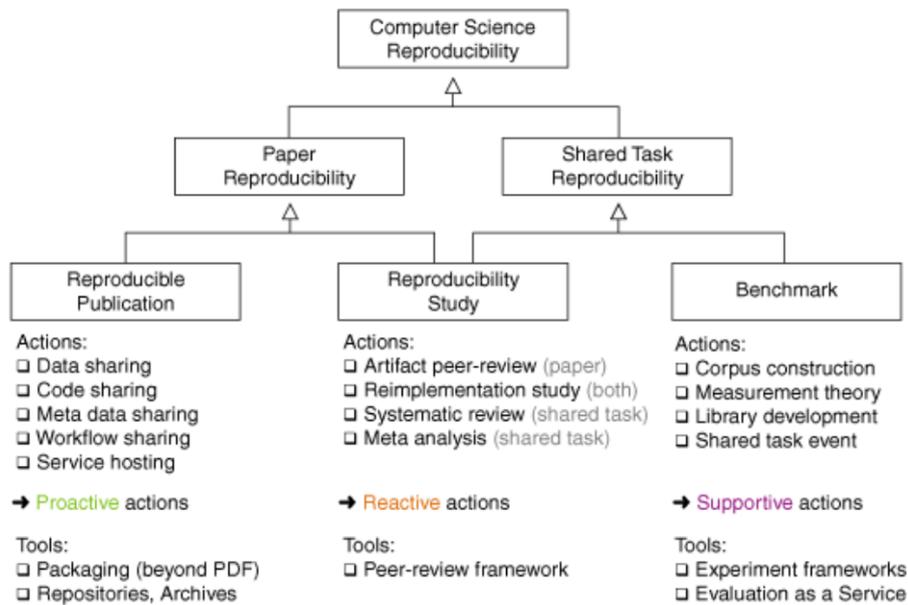
In terms of gaps in the existing suite of tools and services, there might be enough of them available. However, these may still not serve the intended purpose and may need to be improved according to the collected requirements. Such an approach may be preferred rather than developing and implementing yet new ones. In addition, integrating such tools may be useful for not having to reinvent the wheel, or at least enabling their interoperability through common or established models and formats.

One of the main issues is understanding **which tools are available, what each of them support, and which types of problem they solve**. There is a plethora of available tools for reproducible research, and these can be categorized in different ways: (1) features provided (provenance capture, representation, portability, archiving, longevity, access to remote services, etc.); (2) target domain areas; (3) reproducibility modes (planning for reproducibility, such as scientific workflow systems and configuration management tools, vs. reproducibility as an afterthought, such as packaging tools); and many others. Depending on these different tags, researchers may need different tools. A search engine for reproducibility tools, for instance, would be useful. ReproMatch²³ is a step towards this.

There are some well established and widely **accepted infrastructures**, such as DropBox, GitHub, and Zenodo. It is questionable, however, how effective these are for the intended reproducibility and whether the granularity of the stored artifacts is sufficient. For instance, these examples are entitled to provide pure storage and versioning, and not curation; also, the lineage between code and data, and the reasoning behind the versioning are not captured. If this is a requirement, certainly these tools do not address the needs.

It may be helpful to reflect on a **basic research environment**, which allows to automatically track and record individual steps and milestones during the research and developing process. This may include an electronic notebook providing automatic documentation support. Such infrastructure would provide stable and standardized code, including a version

²³ <http://repromatch.poly.edu/>



■ **Figure 2** Taxonomy of actions towards improving reproducibility in computer science.

control and a notification service in case anything changes in any related software package, library, or operating system. It is unclear, however, if a single environment can be developed for different domains.

6.3 Taxonomy of Actions Toward Reproducibility

Martin Potthast (Bauhaus-Universität Weimar, DE), Fernando Chirigati (NYU Tandon School of Engineering, US), David De Roure (University of Oxford, GB), Rudolf Mayer (SBA Research – Wien, AT), and Benno Stein (Bauhaus-Universität Weimar, DE)

License © Creative Commons BY 3.0 Unported license
 © Martin Potthast, Fernando Chirigati, David De Roure, Rudolf Mayer, and Benno Stein

There are a number of well-known actions that researchers may take today to improve the reproducibility of computer science, whereas many of them are at best partially supported by tools, or not at all. Figure 2 organizes these actions within a taxonomy.

The taxonomy comprises two levels, where the first divides reproducibility into two disjunct categories: “Paper reproducibility” comprises only actions that serve to improve the reproducibility of individual papers, whereas “shared task reproducibility” comprises actions aimed at improving the reproducibility of groups of independent papers that address a common problem of interest (i.e., a shared task). The distinction is important, since the actions that can be taken in each case significantly differ.

On the second level of the taxonomy, the artifacts that may result from taking action toward improving reproducibility are given. There are basically three categories, namely a “reproducible publication,” a “reproducibility study,” and a “benchmark.” Reproducibility studies can be done for both, individual papers as well as shared tasks, whereas some actions apply only to one, the other, or both.

Below these categories, specific actions are listed that ultimately yield artifacts belonging to their respective category. All of the actions can be distinguished by who takes them, and their relation to reproducibility: actions that lead toward reproducible publications must be taken proactively by authors before publication. Reproducibility studies, however, are always in reaction to publication of new results. They can be done at publication time, e.g., by peer-reviewers, but are more often conducted long after publication by researchers working on the same problem. For shared tasks, supportive actions are often taken in order to create standardized benchmarks, and to ensure comparability across papers.

These actions and the required information can be nicely mapped to the PRIMAD model of distinguishing different types of reproducibility by varying (priming) specific aspects of a study while keeping others unchanged. This requires the unchanged aspects (data, code, execution environment) to be shared and deliver different gains in knowledge on the original study.

Proactive actions: to share the data and to share the code underlying a paper are perhaps the most important actions that authors can take to create a reproducible publication. Moreover, hosting a prototype service and providing metadata and workflow information may prove to be key assets to understanding the runtime behavior of a given contribution.

All of the aforementioned artifacts can be shared today with some extra effort, since there are places on the web where data, code, and other artifacts belonging to a given paper can be hosted. However, spreading artifacts across different platforms is hardly straightforward to follow up, much less maintainable. Rather, scientists are used to obtaining research at one place, namely the PDF hosted at publisher site. Since no common standards for sharing scientific artifacts beyond the PDF have emerged to date, the landscape is disorganized, with a number of individual solutions as well as a number of community-specific tools. Two particular kinds tools can be identified in this respect, namely packaging technology that envelopes all artifacts resulting from a given piece of research, and repositories and archives that allow for retrieval, long-term storage, and maintenance of published artifacts.

Reactive actions: ideally, submitted publications would be immediately checked for reproducibility, e.g., within an artifact peer-review, but this not, yet, commonplace. Otherwise, reproducibility studies are the most effective way to ensure independent reproducibility. Specifically, reimplementing a given paper including all of its experiments, or reimplementing individual approaches proposed in a group of papers on a shared task without reproducing all experiments of all papers. In addition, for shared tasks, systematic reviews and meta analyses may shed light onto the state of reproducibility in a given shared task. In this connection, we emphasize the distinction between systematic reviews and literature reviews (i.e., surveys): systematic reviews abstract over a subject matter (e.g., by unifying terminology, by organizing existing contributions with regard to previously unconsidered criteria, or by recasting the problem of interest in a new way), whereas literature reviews merely collect the existing contributions with little to no abstraction. Hardly any tool support beyond the existing academic search engines has been invented so far, since systematic reviews and meta analyses rely on abstract thinking over the original papers that are studied. Also the peer-review of artifacts is currently hardly supported within conference management systems.

Supportive actions: benchmarks for software-driven and data-driven computer science are perhaps one of the few cases where computer science already excels: once a given shared task is studied more frequently, researches often build specific evaluation corpora, they study the theory of measuring performance of proposed solutions, and they develop software libraries to collect state-of-the-art algorithms for the shared task. If the community surrounding a shared

task agrees on a benchmark, papers published henceforth are comparable without further need for coordination among researchers. In some cases, shared task events are organized, where researchers compete to build the best solution for the shared task's underlying problem, and where submitted solutions are evaluated within a standardized evaluation setup that often becomes a new benchmark.

Regarding tools for supportive actions, experiment frameworks allow for the structured execution of experiment series for particular tasks. Such frameworks are often tailored to particular research domains and shared tasks. Moreover, since not all datasets can be shared publicly for reasons of privacy and copyright, among others, this prevents some important benchmarks from becoming widely available. To mitigate these limitations, it has been proposed to move evaluation to the cloud under the recently proposed evaluation as a service paradigm [1]: under this paradigm, software that solves a given shared task is deployed within a cloud infrastructure, and the software's processing rights for the sensitive datasets are managed and controlled, so that data cannot leak.

Automation: all of the aforementioned tools together point into an interesting new direction for experimental, data-driven and software-driven science, namely automation. Much of the process of optimizing scientific software to a problem includes parameter optimization, which often boils down to an (informed) search in hyperparameter space. The expanding capabilities of cloud computing can be exploited to tune scientific software deployed under the evaluation as a service paradigm, maximizing expected performance compared to manual or semi-automatic optimization. Moreover, considering individual papers, a more standardized way of annotating the scientific process and its outcome in the form of papers and other artifacts, will allow for their inclusion in the linked data cloud and, eventually, inference on top of that.

Social interaction: the tool support to improve reproducibility will not be based solely on standardized interfaces. Rather, the web services that will eventually emerge will likely include social networks. Unlike papers, artifacts may not always be perfectly documented, which question to be answered and solutions to be discussed, in order to complete a documentation or fix issues with previously published artifacts. This is especially true when the original authors of the artifacts are not available, anymore, long after publication.

References

- 1 Allan Hanbury, Henning Müller, Krisztian Balog, Torben Brodt, Gordon V. Cormack, Ivan Eggel, Tim Gollub, Frank Hopfgartner, Jayashree Kalpathy-Cramer, Noriko Kando, Anastasia Krithara, Jimmy Lin, Simon Mercer, and Martin Potthast. Evaluation-as-a-Service: Overview and Outlook. ArXiv e-prints, December 2015.

6.4 Actors in Reproducibility

Justin Zobel (The University of Melbourne, AU), Shane Culpepper (RMIT University – Melbourne, AU), David De Roure (University of Oxford, GB), Arjen P. de Vries (Radboud University Nijmegen, NL), Carole Goble (University of Manchester, GB), Randall J. LeVeque (University of Washington – Seattle, US), Mihai Lupu (TU Wien, AT), Alistair Moffat (The University of Melbourne, AU), Kevin Page (University of Oxford, GB), and Paul Rosenthal (TU Chemnitz, DE)

License © Creative Commons BY 3.0 Unported license
 © Justin Zobel, Shane Culpepper, David De Roure, Arjen P. de Vries, Carole Goble, Randall J. LeVeque, Mihai Lupu, Alistair Moffat, Kevin Page, and Paul Rosenthal

Reproducibility is a component of a greater activity (e.g. reviewing, reusing) undertaken by actors (e.g. reviewer, author) who have their own behaviours (inherent or induced by external drivers). Interventions to motivate reproducibility behaviours, through positive incentives or the removal of obstacles, requires us to first classify actors and then layout behavioural standard

6.4.1 Actors

- **Creators:** authors, academic leaders/lab directors, research software engineers, thesis supervisors
- **Consumers:** readers, authors, students, policy makers, educators, adopters, technical communities, IT services, industry, user, research software engineers, PhD students
- **Moderators:** editors
- **Examiners:** reviewers, thesis examiners, research evaluation committees,
- **Enablers:** funders, publishers, institutions, academic leaders/lab directors, data providers, thesis supervisors, digital archives, professional societies, industry, research software engineers
- **Auditors:** funders, policy makers, institutions, professional societies

6.4.2 Questions

- What are the properties of reproducibility for each actor?
- What are the interventions they can invoke?
- What are the current behaviours, and how might they shift?
- What aspects of behaviour are important to whom?
- What timeframes apply?
- What are the obstacles to good behaviour?
- What are the incentives to encourage change in behaviour?
- What are the interventions to action change in behaviour?

6.4.3 Authors

This section summarizes the main obstacles and expectations for an author.

6.4.3.1 Obstacles (real or perceived) to good behaviour for authors

Obstacles may be external drivers over which the authors have limited control, or internal where the authors can be responsible for their own behaviour. Table 1 describes the obstacles in detail.

■ **Table 1** Obstacles for authors.

Recognition	Lack of explicit recognition of the need for reproducibility within a lab Lack of credit for achieving reproducibility
Cultural pressure	Lab culture Publication (volume) pressure Time pressure
Ambition/Personal Pressure	Paranoia – fear of losing competitive advantage Embarrassment, limitations as a developer Fear of having mistakes exposed (security through obscurity)
Awareness	Ignorance of the benefits of reproducibility, lack of mentoring and guidance Misjudgement of the difficulty of achieving reproducibility Lack of planning for reproducibility – it cannot be an afterthought Perception of achievability
Intention	Code/data was meant to be disposable (ephemeral)
Resources	Lack of access to appropriate resources Inertia, apathy, lack of incentives
Institutional restrictions	Legal and licensing issues, Corporate privacy requirements
Innate restrictions	Code or data cannot be encapsulated

Three tiers of standard – sufficient, better, exemplary – set out a rubric of expected behaviour. Interventions and incentives have the capacity to move up the reproducibility ramp.

6.4.3.2 Standards: Sufficient

These elements, if present in a paper and appropriate to that paper, represent a minimum expectation of authors – with regard to both ethical requirements and the demands of reproducibility.

- Methods section – to a level that allows imitation of the work
- Appropriate comparison to appropriate benchmark
- Data accurately described
- Can re-run the experiment
- Verify on demand (provide evidence that the work was done as described)
- Ethical considerations noted, clearances listed
- Conflicts noted, contributions and responsibilities noted
- Use of other authors' reproducibility materials should respect the original work and reflect an attempt to get best-possible results from those materials

6.4.3.3 Standards: Better

Addition of elements such as these represent a substantial increment beyond sufficient, while not yet being best practice.

- Black/white box
- Code is made available, in the form used for the experiments
- Accessible or providable data

■ **Table 2** Obstacles to good behaviour for reviewers.

Recognition	Lack of explicit recognition of the need for reproducibility within the discipline Lack of credit for examining reproducibility
Cultural pressure	Time pressure Volume pressure
Ambition/Personal Pressure	Embarrassment, technical limitations Lack of understanding of why reproduction failed – is it really the fault of the reviewer or authors?
Awareness	Ignorance of the benefits of reproducibility, lack of mentoring and guidance Misjudgment of the difficulty of examining reproducibility Perception of achievability
Intention	None
Resources	Lack of access to appropriate resources – technical, personnel Inertia, apathy, lack of incentives
Institutional restrictions	None
Innate restrictions	None

6.4.3.4 Standards: Exemplary

Addition of these elements, in or accompanying a paper, represent best practice for authors.

- Open-source software
- Engineered for re-use
- Accessible data
- Published in trustworthy, enduring repository
- Data recipes, to allow construction of similar data
- Data properly annotated and curated
- Executable version of the paper; one-click installation and execution

6.4.4 Reviewers

Noting the potential for reviewers to be explicitly assigned to provide either technical review or scientific review:

6.4.4.1 Obstacles (real or perceived) to good behaviour for reviewers

Table 2 describes the obstacles in detail.

6.4.4.2 Standards: Sufficient

- Assesses reproducibility
- Fair assessment, respect of strengths and weaknesses
- Clarity on what was assessed and what the limits of the review are
- Conflicts noted

6.4.4.3 Standards: Better

- Checks that reproducibility is in fact possible

■ **Table 3** Obstacles to good behaviour for editors.

Recognition	Lack of explicit recognition of the need for reproducibility within the discipline Lack of credit for examining reproducibility
Cultural pressure	Time pressure Volume pressure
Ambition/Personal Pressure	None
Awareness	Ignorance of the benefits of reproducibility, lack of mentoring and guidance Misjudgment of the difficulty of examining reproducibility Perception of achievability
Intention	None
Resources	Inability to find technically accomplished reviewers
Institutional restrictions	None
Innate restrictions	None

6.4.4.4 Standards: Exemplary

- Reproducible, within limits of materials and resources
- Timely reviews

6.4.5 Editors

6.4.5.1 Obstacles (real or perceived) to good behaviour for editors

Table 3 describes the obstacles in detail.

6.4.5.2 Standards: Sufficient

- Find reviewers who can assess the science
- Have reviewing policies that require examination of reproducibility/methodology
- Have instructions for authors on expectations with regard to reproducibility/methodology
- ‘Reproducibility compacts’ (or contracts) for authors, in which they must state availability of code and so on [1]

6.4.5.3 Standards: Better

- Find reviewers who can assess the technical contribution
- Separation of assessment of papers on science grounds from reproducibility/methodology grounds
- Have processes for working with authors to improve reproducibility

6.4.5.4 Standards: Exemplary

- Advocacy to the publisher of requirements for reproducibility
- Advocacy of standards
- Leadership regarding all aspects of reproducibility
- Participation in relevant advocacy bodies

■ **Table 4** Obstacles to good behaviour for institutions.

Recognition	Lack of explicit recognition of the need for reproducibility Lack of credit for achieving reproducibility
Cultural pressure	Publication (volume) pressure Fear of having mistakes exposed (security through obscurity)
Ambition/Personal Pressure	Lack of enduring commitment – long-term budgeting Lack of communication plans Resistance to openness Paranoia – fear of losing competitive advantage Fear of having mistakes exposed (security through obscurity)
Awareness	Ignorance of the benefits of reproducibility, lack of mentoring and guidance Misjudgment of the difficulty of examining reproducibility Perception of achievability Legal and licensing issues
Intention	None
Resources	Resources, services, infrastructure, repositories Lack of standards and tools Lack of access to appropriate resources Lack of understanding of the resources required Inertia, apathy, lack of incentives
Institutional restrictions	Confused lines of responsibility, mixed ownership of the problem Human resources structures: mentoring, training, staffing Mismatch between academic and organizational goals Conflicting or missing or ill-informed policies Legal and licensing issues Corporate privacy requirements
Innate restrictions	None

6.4.6 Institutions (also as transmitted via academic leaders)

6.4.6.1 Obstacles (real or perceived) to good behaviour for institutions

Table 4 describes the obstacles in detail.

6.4.6.2 Standards: Sufficient

- Clear policies on reproducibility, ethic

6.4.6.3 Standards: Better

- Compliance framework
- Resourcing of reproduction – technical, financial
- Constructive environment with recognition of demands of reproduction

6.4.6.4 Standards: Exemplary

- Trusted, enduring repository
- Reproduction as a primary research goal

References

- 1 C. Collberg, T. Proebsting and A. M. Warren. Repeatability and Benefaction in Computer Systems Research. University of Arizona TR 14-04.

6.5 Guidelines for Authors, Editors, Reviewers, and Program Committee Chairs

Alistair Moffat (The University of Melbourne, AU), Shane Culpepper (RMIT University – Melbourne, AU), Arjen P. de Vries (Radboud University Nijmegen, NL), Carole Goble (University of Manchester, GB), Randall J. LeVeque (University of Washington – Seattle, US), Mihai Lupu (TU Wien, AT), Andreas Rauber (TU Wien, AT), and Justin Zobel (The University of Melbourne, AU)

License © Creative Commons BY 3.0 Unported license
© Alistair Moffat, Shane Culpepper, Arjen P. de Vries, Carole Goble, Randall J. LeVeque, Mihai Lupu, Andreas Rauber, and Justin Zobel

A framework for explaining the need for reproducibility is to describe it in three separate elements, the *what*, the *why*, and the *how*. The *what* consists of articulating the authors' goals in the context of the instructions provided to reviewers and editors (or PC Chairs). The *why* consists of communicating an understanding of desirability of reproducibility, and of helping to convey the distinctions in the key terminology (reproducibility, repeatability). The *how* entails guidelines as to the means by which papers are assessed as a consequence of introduction of expectations regarding reproducibility.

The SIGMOD reproducibility guidelines²⁴ are a description of *what*: it is stated that it is desirable that papers' materials have shareability, coverage, and flexibility; noting that in some cases mechanical descriptions (recipes) rather than code may be necessary to couple shared resources to the external world. Another extension is the need for authors to re-use open materials in an appropriately scientific way – reusability materials are a scientific resource, not conventional open-source software.

An explanation of *why* appears in Section (see Section 6.7) of this report, and can be summarized as providing:

- Confidence in the work
- Acceleration of the science and of the state-of-the-art
- Verifiability
- Falsifiability
- Participation in the community, contribution to the community

A key requirement to achieve these outcomes is in the instructions that are provided to referees and authors. The European Conference on Information Retrieval 2016 has provided guidance in this regard²⁵, to which we add a final sentence (cf. also the PRIMAD model and the gains of different types of reproducibility in Section 6.1):

*Reproducibility is key for establishing research to be reliable, referenceable and extensible for the future. Experimental papers are therefore most useful when their results can be tested and generalized by peers. This track specifically invites submission of papers reproducing a single or a group of papers, from a third-party where you have **not** been directly involved (e.g. **not** been an author or a collaborator). Emphasize your motivation for selecting the paper/papers, the process of how results have been attempted to be reproduced (successful or not), the communication that was necessary to gather all information, the potential difficulties encountered and the result of the process. A successful reproduction of the work is not a requirement, but it is important to provide*

²⁴ <http://db-reproducibility.seas.harvard.edu/>

²⁵ http://ecir2016.dei.unipd.it/call_for_papers.html

*a clear and rigid evaluation of the process to allow lessons to be learned for the future.”
It is not sufficient for a reproduction to be a simple re-execution of the existing code on the original data.*

A key *how* is the encouragement and definition of papers that have as their primary objective the reproduction and extension of previous work, and the provision of a refereeing process that recognizes the merits of such papers, and evaluates them accordingly. Encouraging behaviors that facilitate independent reproduction is another key goal. SIGMOD again provides an illustration of how this is done, with an independent evaluation process, different from the regular scientific acceptance review process. Authors should be encouraged to plan for reproducibility right from the commencement of each investigation, with a clear plan in place for how they will develop methods and artifacts that can be communicated to and reused by others.

Collberg et al.’s [1] evaluation describes the concept of a compact (or contract) that authors are expected to add to their paper at submission time and retain in the final version, in which they make claims about the likely reproducibility of their work. We regard the routine adoption of such a statement in published work as a low-cost but effective instrument for reproducibility.

Other key interventions that might benefit one or more of the various actors involved in the processes of scientific research, funding, and publishing include:

- A template for journals of instructions for authors and reviewers, in a form that allows adaptation; include explanation of how to respond when attempts to reproduce struggle or fail; include explanations in a digestible form that promotes reproducibility
- A template for thesis examination
- Managerial appraisal questions; compliance requirements at all levels
- A consolidated and maintained resource of information about reproducibility in computing.
- Recruitment of advocates and champions
- Information packs targeted at particular categories of actor
- Recognition systems – effective, visible – promotion, tenure, within institution, within journal/conference

Drawing on similar statements (including Science [2]; the SIGMOD²⁶ and ICTIR²⁷ Calls for Papers; and other sources), **we suggest the following message be sent to a wide pool of journal editors and conference chairs**, to be then used by them in communications with other senior members of the various communities:

Transparency, openness, and reproducibility are vital features of science. Scientists embrace these features as disciplinary norms and values, and it follows that they should be integrated into daily research activities. These practices give confidence in the work; help research as a whole to be conducted at a higher standard and be undertaken more efficiently; provide verifiability and falsifiability; and encourage a community of mutual cooperation. They also lead to a valuable form of paper, namely, reports on evaluation and reproduction of prior work. Outcomes that others can build upon and use for their own research, whether a theoretical construct or a reproducible experimental result, form a foundation on which science can progress. Papers that are structured and presented in a manner that facilitates and encourages

²⁶ <http://db-reproducibility.seas.harvard.edu/>

²⁷ <http://ictir2015.org/cfp>

such post-publication evaluations benefit from increased impact, recognition, and citation rates.

Experience in computing research has demonstrated that a range of straightforward mechanisms can be employed to encourage authors to produce reproducible work. These include: requiring an explicit commitment to an intended level of provision of reproducible materials as a routine part of each paper's structure; requiring a detailed methods section; separating the refereeing of the paper's scientific contribution and its technical process; and explicitly encouraging the creation and reuse of open resources (data, or code, or both).

- This document provides links and resources to the following:
- template instructions for authors
- examples of authorial statements of commitment
- template guidelines for reviewers
- lists of resources (such as trustworthy repositories and tools)
- lists of examples of publication venues that have implemented such measures
- list of exemplary papers

The list of objects in the bulleted points would need to be assembled and made available.

6.5.1 Template Instructions for Authors

Following are two simple examples that capture the “pay it forward” benefit to the community of having papers that are explicitly designed with reproducibility in mind.

Version 1

The [insert name of journal/conference] encourages authors to provide their work in a way that enables reproduction of their outcomes. Just as you have benefited as an author from the work you cite in your paper, and the tools and resources that others have provided, your efforts will also assist the community, including your future collaborators, if you provide access to and understanding of the tools and resources that you have used and created while carrying out your project. We therefore encourage authors to include in their papers detailed explanations of how their work might be reproduced by others in the field, and to accompany their papers with links to data and source code.

Version 2

The [insert name of journal/conference] encourages authors to provide their work in a way that enables reproduction of their outcomes. Just as you have benefited as an author from the work you cite in your paper, and the tools and resources that others have provided, your efforts will also assist the community, including your future collaborators, if you provide access to and understanding of the tools and resources that you have used and created while carrying out your project. We therefore request that authors include in their papers detailed explanations of how their work might be reproduced by others in the field, and to accompany their papers with links to data and source code if it is possible to do so. Authors can request separate reviewing of the reproducibility of their work, a category of publication that we explicitly acknowledge.

Example of supplementary statement

In order to support these expectations authors are encouraged to include a detailed methods section in their paper that describes the techniques, tools, data resources, and code resources that enables readers to easily reproduce the work. Such a methods section is of greatest benefit to the reader when it is linked to materials stored in a trusted open repository, and these materials include illustrative or complete data, and code that can easily be re-used and understood.

References

- 1 Christian Collberg, Todd A. Proebsting: Repeatability in Computer Systems Research. Communications of the ACM, Vol. 59 No. 3, Pages 62-69.
- 2 B. A. Nosek, G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, T. Yarkoni: Promoting an open research culture. Science 26 Jun 2015, Vol. 348, Issue 6242, pp. 1422-1425

6.6 What Makes A Reproducibility Paper Publishable

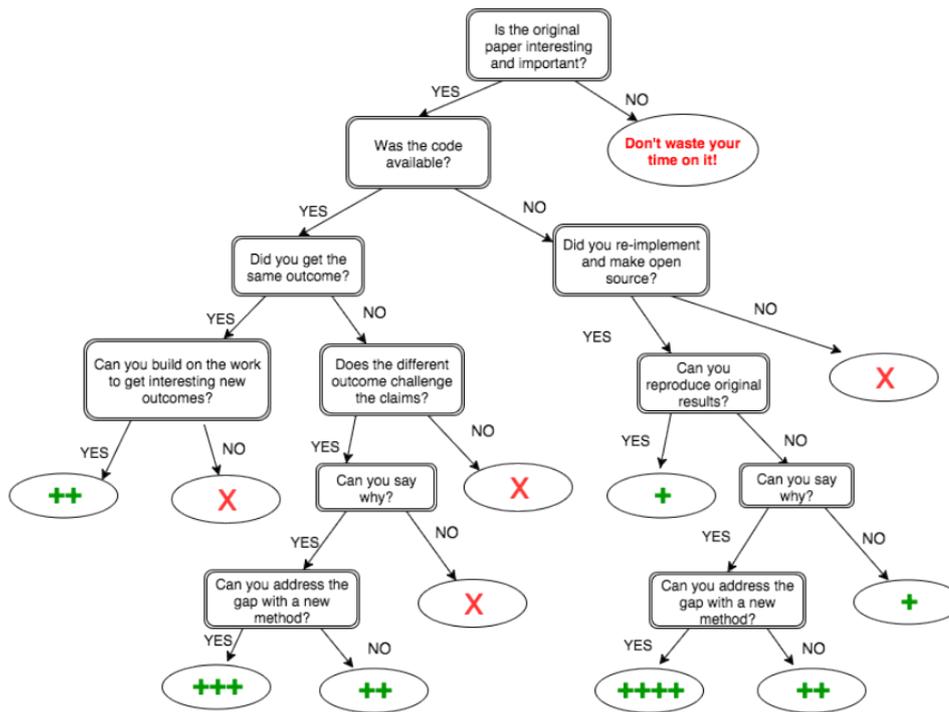
Alistair Moffat (The University of Melbourne, AU), Shane Culpepper (RMIT University – Melbourne, AU), Arjen P. de Vries (Radboud University Nijmegen, NL), Carole Goble (University of Manchester, GB), Randall J. LeVeque (University of Washington – Seattle, US), Mihai Lupu (TU Wien, AT), Andreas Rauber (TU Wien, AT), and Justin Zobel (The University of Melbourne, AU)

License © Creative Commons BY 3.0 Unported license
 © Alistair Moffat, Shane Culpepper, Arjen P. de Vries, Carole Goble, Randall J. LeVeque, Mihai Lupu, Andreas Rauber, and Justin Zobel

The nature of science is to improve our understanding, and to build on the work that precedes our own. This means that there will be occasions when of necessity we repeat, review, re-implement, or re-execute experimental work that has been published by others. The goals of such a review may include: extending, establishing limitations/scope/applicability of, confirming, validating, and in some cases invalidating, the previous work. Particular cases might include ones in which analysis is provided that yields a mathematical underpinning to empirical results.

Figure 3 provides guidance as to what might be regarded as being interesting findings in such work, and hence publishable. In some ways, what is being described in the diagram is the essence of the scientific method. It should not be regarded as being a failure, or even unusual, for previous work to be shown to be of limited applicability, or inaccurate when viewed through a more precise lens. Authors who improve on the work of others should always acknowledge their debt to the authors of the prior work and be graceful in their criticism, even when pointing out flaws and errors. The key attitude must be one of “standing on the shoulders” rather than “kicking in the shins”.

To warrant publication, a paper must do more than merely take existing code and existing data and repeat a previous experiment to obtain the same result. Work becomes interesting when it adds value or insights or defines applicability by documenting extensions, or in a



■ **Figure 3** How publishable is a paper attempting to independently replicate/reproduce an earlier paper? The decision tree is intended to give guidance to someone considering writing a paper that reports on an independent study of a previous paper (or an editor/reviewer considering such a paper). In some cases “code” could be replaced by “code and data”. Data considerations may be different, but similar principles apply.

limiting sense, crystallizing exceptions. With this in mind, the leaves in the diagram that are marked as “X” represent outcomes that are less informative, or unsurprising, outcomes. Conversely, the leaves marked with “+” represent outcomes where ongoing research activities will benefit from publication of this new work.

The number of “+” marks in each leaf is intended to provide some level of guidance as to the value of such a paper, but should not be regarded as being prescriptive. The single node marked as “++++” represents a clear and unambiguous contribution, and in some ways is the ideal. But note that the pathway to that node commenced with a careful review of some previous work – a baseline or starting point – and must still be regarded as having been initiated by a reproduction of that prior activity. As already noted, that is the nature of science: to demonstrate in deeds that previous thought and understandings of the subject can be enhanced or improved. From the point of view of the authors of the prior paper, having that work selected for an in-depth re-evaluation should be regarded as being a sign of respect, and of recognition of contribution. Einstein improved upon Newton, and yet Newton was by no means dumb.

Because of its simplicity, the diagram fails in some cases. One might choose to independently reimplement a method even if code exists; in this case, the results are strengthened if the new code is validated relative to the old experiments before any new experiments are undertaken. This would add to the new work, rather than weaken it.

In summary, papers should always be evaluated on their merits, rather than formulaic requirements. Editors and reviewers need to be aware of the benefits of work that reproduces,

extends, or otherwise refines the work of others, and be encouraging and supportive of authors who pursue these goals. The work that gets encapsulated in such publications should not be regarded as being second-class, or be dismissed as being insufficiently novel.

6.7 Incentives and barriers to reproducibility: investments and returns

Paul Rosenthal (TU Chemnitz, DE), Rudolf Mayer (SBA Research – Wien, AT), Kevin Page (University of Oxford, GB), Rainer Stotzka (KIT – Karlsruher Institut für Technologie, DE), and Evelyne Viegas (Microsoft Research – Redmond, US)

License © Creative Commons BY 3.0 Unported license
© Paul Rosenthal, Rudolf Mayer, Kevin Page, Rainer Stotzka, and Evelyne Viegas

There have been many studies on what motivates people to change their behavior in their personal or professional lives [1]. Simply put, motivation is driven by need: the greater the need the greater the motivation. Other studies have focused on incentives as a means to change behaviours [2] while emphasizing the necessity to engineer the incentive to align with the need and goal to achieve and avoid the incentive to backfire [3]. For instance, as part of reaching better reproducibility, one should not build an incentive towards just making (any) code available, but rather the incentive should be built to make good quality code.

6.7.1 Investments of value

Table 5 lists the different investments of value from the different actors to achieve good reproducibility. The definition of actors is inspired by Section 6.4, with minor differences:

- **Creators** are persons responsible for the creation of possibly reproducible scientific artifacts, i.e. students, authors, academic leaders/lab directors, research software engineers, thesis supervisors, industrial researchers
- **Enablers** are persons and institutions enabling research to be conducted and published, e.g. funders, publishers, editors, institutions, academic leaders/lab directors, data providers, thesis supervisors, digital archives, professional societies, industry, research software engineers
- **Consumers** are persons and institutions consuming and utilizing scientific artifacts, e.g. readers, authors, students, policy makers, educators, adopters, technical communities, IT services, industrial researchers, users, research software engineers, PhD students
- **Examiners** are persons examining the quality of scientific artifacts, e.g. reviewers, thesis examiners, research evaluation committees, funders, policy makers, institutions, professional societies

Table 5 indicates a set of examples for each type of investment and which actors have to invest them.

6.7.2 Returns of Value

Table 6 indicates the returns of value from good reproducibility broken down with respect to the different actors.

6.7.3 Incentives

Having documented the investments, returns of value and the needs for reproducibility per actor (see Section 6.4), in the following we look at incentives required to transition from

■ **Table 5** Investments of value for creators (creat.), enablers (enabl.), consumers (consum.) and examiners (exam.)

Investments	Creat.	Enabl.	Consum.	Exam.
Artifact preparation (clean code, negotiate rights of code and data, annotate, document code and data, support incentives to promote reproducibility)	x			
Research documentation (carefully and reproducibly document all research steps, enable access to documentation)	x			
Education (training good reproducibility structures and methods, examine reproducibility knowledge)	x			x
Infrastructure (establish systems for reproducing computations, to publish reproducible research, and to review reproducibility)	x	x		
Citation (careful creating/citing literature, software, and data)	x		x	
project resources (enable researchers to make investments into reproducibility with respect to time and work force)		x		x
publication guidelines (prepare guidelines for authors and reviewers of publications with focus on reproducibility)		x		x
time for reviewers (give reviewers time to assess the reproducibility documentation of publications)		x		x
principles (create software citation, data management, and reproducibility plan principles)		x		x
requirements validation (establish methods to validate proposal with respect to established principles, adapt panel behaviours to follow principles)		x		x
credit mechanisms (establish community and institutional mechanisms to credit reproducible research)	x	x		

current behaviours to the desired ones to reach reproducibility. We do so in the context of four categories (natural, moral, financial and coercive), and their relationship to the actors. We adapted the categories of incentives from McClelland [3] and Dalkir [4] to address reproducibility as follows:

- **Natural Incentives:** an actor applies her/his curiosity towards PRIMAD, searches for the pursuit of true science, or wants to participate to accelerating research and innovation for the benefit of the social good in the world.
- **Moral incentives:** the choice made by the actor of embracing PRIMAD (see Section 6.1) to make her/his work reproducibility is widely regarded as the right thing to do, or as admirable and the actor can expect a sense of self-esteem, while the failure to adopt PRIMAD is condemned as the wrong thing to do, or as condemnable, and the actor can expect a sense of guilt.
- **Financial incentives:** the actor can expect some form of material reward (e.g. prize, grant, and more generally money) – in exchange for making her/his work reproducible.
- **Coercive incentives:** the actor failing to embrace PRIMAD will see her/his reputation shaken, portfolio of opportunities (e.g. grants, government budget) diminished.

■ **Table 6** Possible returns of value for creators (creat.), enablers (enabl.), consumers (consum.) and examiners (exam.)

Returns of Value	Creat.	Enabl.	Consum.	Exam.
Publicity (more citations and promotion for papers, code, and data, awareness of own and other communities, visibility for possible industry partners)	x	x		
Insight (better estimation of costs for reproducibility, easy incorporation into future proposal and plans)	x			
Impact in industry (commercialisation of research, recognition of results in industry, impact of research in industry)	x	x		
entry into industry (knowledge entry into industry eased, acceleration of transfer, wider economic value and relevance of research, easier reuse for industry)		x	x	
personal satisfaction (providing good reproducibility can give good conscience and satisfaction due to the good cause)	x	x		
incorporation in teaching (reduced preparation time and costs for education by using reproducible research artifacts)			x	
research reuse (easier, quicker, and more reliable research, building on reproducible results, code, data, and methods)	x	x	x	
innovation (more innovation through saving time to reproduce)			x	
ease of reproducibility (well-established mechanisms of reproducibility and accountability through introduction of common culture)	x	x	x	
funding effectivity (funding agencies get reproducible and reusable research, ineffective duplicated investigation and implementation is avoided, greater funds are conserved for novel research)		x	x	
interdisciplinarity (research between agencies, institutions, and labs becomes easier through a common ground of reproducibility)	x	x	x	
comparability (easier comparison with state of the art methods)			x	x

The principles underlined in the 4 reproducibility incentives categories are proposed to help design incentives that meet the reproducibility needs of each community and we expect that they will vary across communities, cultures and actors.

We propose below some examples of incentives per actor:

- The researcher who embraces PRIMAD creates a financial incentive (e.g. app research store) to get more investment from the funders, from industry into her/his research
- The researcher/community who embraces PRIMAD creates a coercive incentive (e.g. “no PRIMAD” stamp) for funders who ignore PRIMAD cost in research
- The community creates a natural incentive (e.g. best reproducibility award) for the researcher to make her/his research reproducible.
- The community creates a moral incentive (e.g. hall of fame) for the researcher to make her/his research reproducible
- The funders create a natural incentive (e.g. interdisciplinary badge) for the researcher to make her/his research reproducible where research is reused across scientific areas
- The funders create a coercive incentive (e.g. grant application section on reproducibility) for the researcher to make her/his research reproducible
- The funders create a financial incentive (e.g. grant, in kind resources) for the researcher to make her/his research reproducible

In structuring these incentives we also note the potential for deferred returns of value to act as a barrier for adoption and implementation of reproducibility. Where an actor must make an investment of value (Table 5), frequently as an individual, a significant period of time before reaping an equivalent or greater return of value (Table 6), often through membership of a community, the interim “debt” may become a disincentive to make that investment; i.e. beyond principled or altruistic motivations it may be difficult to justify that investment above the many other demands for priority faced by researchers and their organisations. As such, despite the long-term sustainability of reproducibility as an economic system through a beneficial cycle of investment and returns, it may be desirable – perhaps necessary – for enabling organisations to provide an initial pump priming investment of value to provide a “bridging loan” to creators until the system is self-sustained.

References

- 1 The World Bank Group. Theories of Behavior Change, Communication for Governance and Accountability Program.
- 2 Rothman AJ. Initiatives to Motivate Change: A Review of Theory and Practice and Their Implications for Older Adults. In: National Research Council (US) Committee on Aging Frontiers in Social Psychology, Personality, and Adult Developmental Psychology; Carstensen LL, Hartel CR, editors. When I’m 64. Washington (DC): National Academies Press (US); 2006.
- 3 McClelland, David C. (1987). Human Motivation. CUP Archive.
- 4 Dalkir, Kimiz (2013). Knowledge management in theory and practice. Routledge. McClelland, David C. (1987). Human Motivation. CUP Archive.

6.8 User Studies in IR

Nicola Ferro (University of Padova, IT), Norbert Fuhr (Universität Duisburg-Essen, DE), Kalervo Järvelin (University of Tampere, FI), Noriko Kando (National Institute of Informatics – Tokyo, JP), and Matthias Lippold (Universität Duisburg-Essen, DE)

License  Creative Commons BY 3.0 Unported license
© Nicola Ferro, Norbert Fuhr, Kalervo Järvelin, Noriko Kando, and Matthias Lippold

The goal of information retrieval (IR) is to best serve a user information need by presenting him/her with a list of documents (information objects) potentially relevant to this need. This calls for specific evaluation methodologies which take into account the user, since determining the quality of a produced ranking, i.e. the effectiveness of a system, is directly depending on the user notion of what is satisfactory for his/her information need.

This setting is quite different from what we have, for example, in databases, where queries are exact and the correctness of results is not an issue, putting the emphasis on efficiency rather than effectiveness.

Therefore, it becomes central to understand what reproducibility is and how it can be achieved when users are in the loop.

6.8.1 Methodological Background

6.8.1.1 Experiments in psychology

The knowledge acquired in psychology is based on empirical results of experiments. An experiment is a research method in which one or more independent variables (IV) are manipulated to determine the effect(s) on a dependent variable. Other relevant factors need to be controlled in this setting. For instance, in the case of a user experiment in information retrieval, the independent variable could be a different search algorithm and the dependent variable could be the time to finish the search.

Psychological experiments needs to fulfil three criteria: **validity**, **objectivity** and **reliability**.

6.8.1.2 Validity

They need to be valid, which it is when the measures what it claims to measure is really measured. A problem could be that some participants might not be paying attention during the experiment, because of a lack of motivation. In some cases a manipulations check, which tests the attention of the user can be useful.

6.8.1.3 Objectivity

Objectivity is also important. An experiment has to be objective in two ways, the result of the experiment should not be influenced by the experimenter and that the interpretation of the data should not depend on the examiner.

6.8.1.4 Reliability

An experiment has to be reliable. When you repeat your experiment or another person repeats your experiment should come to a similar result. To ensure reliability, scientists have to specify their experimental design, they have to describe the conditions, under which the experiment is conducted and share information about the participants. The material and the

raw data of the experiments needs to be stored and shared on demand by the corresponding author.

6.8.1.5 Reproducibility crisis in psychology

In a recent study (Open Science Collaboration, 2015) the results from 100 experiments from four top journals could just be partially replicated. That started a big discussion about the reasons.

6.8.1.6 Reasons for failed replication

Theoretical reason can be in the theories selection itself. If you have an **ill-defined theory**, which does not specify the outcome of the experiment and you use the result of the experiment as evidence for you theory, then the result did not matter and most likely can not be reproduced. For the IR experiments it might be necessary to define for which population the tools are produced and if the result can be generalized for all possible users. Older people might use the search engines in a different way than students do, which usually are the participants of the experiments.

Another theoretical threat are post theories and **post hypotheses or predictions**. If the hypotheses and the theoretical background are selected after the result of the experiment is known, you can not claim that you knew before. When this is happening the probabilities and the p-values are wrong.

Concerning the methodology, this is also a problem in psychology. Researchers rely almost exclusively on the p-value and **do not consider the effect sizes**, which are more important. The question in IR should not primarily be, is there a difference, but how big is the difference and would the user actually notice this difference. Furthermore, a lot of experiments are conducted with **low statistical power**, so the effect in this kind of experiment might not be the real effect and a replication can not find this result.

6.8.2 Context of User-oriented IR Evaluation

In IR, we have different kinds of user studies:

- laboratory experiments, where users are observed in the lab
- in situ observation of users at their workplace
- living labs, where the researcher analyses the system logs and possibly also manipulates the system employed by the users for their daily work.

Besides these types of experiments, there are studies that focus mainly on data collection methods, for which the discussion below only partially applies:

- exploratory user studies,
- focus groups, where researchers interview users
- longitudinal studies of users.

For discussing the reproducibility issues for the specific case of user studies, we follow the PRIMAD model (see Section 6.1) described above:

- Research objective is the research question to be addressed. In most cases, this part should also include the hypotheses to be tested with the experiment described in the remainder of the research paper.
- Model relates here to the experimental settings, which are used for testing the hypotheses specified before, So, besides the type of study, also the relevant aspects of the settings that refer to the research objectives are part of the model

- Implementation and Platform correspond here to the environment in which the study was carried out. Besides the system used for the study, also the group of users participating in the study as well as the exact conditions under which they participated belong to this aspect.
- Actor is the experimenter. In cases where the experimenter has direct contact with the users, the actor might have influence on the results of the study. Thus the actor should be kept constant throughout the study
- Data has a twofold meaning in user studies. First, there is the data that comprises the so-called testbed, like the document collection, the tasks carried out by the users, etc.. Second, there is the observation data collected throughout the study (thus, the user is regarded here as a data generator))

For enabling reproducibility, a researcher should share this context with other users to the maximum extent possible. Research objective and method are usually described in the research paper. In the past, the main research objective was the effectiveness of the methods investigated. Nowadays, also other aspects are considered, which are either more closely related to the actual user task, or to more subjective factors such as user satisfaction or engagement (which, in turn, can be measured via different variables). The more factors are considered, the more it becomes important to state the research hypotheses before actually carrying out the study, in order to achieve statistically valid results.

The environment usually can only be shared partially (mainly the system), while most other aspects (e.g. the users, the hardware, etc.) should be described at a reasonable level of detail in order to ease reproducibility. The same holds for the actor.

For the data, sharing testbeds is widely accepted nowadays, since the state of the art does not allow yet to characterize testbeds to such an extent that an independent researcher would be able to create a comparable testbed that could be expected to give the same results. The observation data, on the other hand, is essential for verifying the claims of a research paper. To a limited extent, it also can be used for simulation studies, depending on the degree of interactivity involved in the study (in classical IR experiments, the only data of this kind are relevance judgments).

6.8.3 Barriers/Obstacles

The research on Information Retrieval (IR) using computer started in 1950s and is said that IR is the first area in computer science using the human judgement as a success criteria of the technology [1]. This makes IR interesting and complex, and therefore the IR community has a strong tradition on evaluation to cope with how users incorporated in the experiment and testing, and make the reasonable comparison across the systems and the algorithms in the same system. Moreover, the commercial online search services started in 1960s and then the issue of working with real users in an interactive environment came up.

Since the Cranfield project in early 1960 [2], researchers constructed and shared testbeds called “test collections” which consist of the three types of data: document collections, the set of search requests, and the static set of human relevance judgment for each search request on the document collection. Such re-usable static testbeds were shared by the community as an infrastructure for the comparative evaluation and as one of the major elements for reproducibility of the experiments.

However, the technology and the society evolved tremendously: interactive online search for various purposes by ordinary persons became pervasive in everyday life, the various data collections including various web services and the social media are enhanced, search

on multiple devices for multi-tasking become more common. The traditional evaluation paradigm (based on the batch-style one-time judgments) can not cover all the problems in the IR research and we are facing various new challenges and obstacles to make the research reproducible:

For studying users in interactive IR, there are various barriers and obstacles for reproducibility:

- Privacy/limitations of anonymizing
- Confidentiality
- Volatility of data (live streams, when the same situation never happens again, etc)
- Validity of the data: the data is so multidimensional that it is difficult to ensure the external validity of the experiment. This complexity is present also in IR test collection, and even more if we consider dynamic test collections.

Online web services are generally based on algorithms using user behaviour data in some way. This data is intrinsically rich in privacy and often includes confidential information. With interactive research IR systems, the situation is similar. Although various research efforts have targeted anonymization, there are still limitations, and these make it difficult to release the user behaviour data for external research groups, which, in turn, hampers reproducibility.

Large-scale users logs are generated in commercial search services and substantial studies on modeling and predicting users behaviour have been conducted based on these data, but again, the underlying data is not accessible for other researchers. Not only user modeling studies, but also various operational search mechanisms exploit user behavior data in the search and ranking algorithms, thus making it difficult to reproduce these methods.

To tackle the problems of the document data with privacy information and/or copyright problems, various evaluation-as-a-service approaches have been proposed and some of them were implemented successfully. However, these are still not sufficient for all the data produced by the users in-situ and lab environments.

For volatility, IR experiments can be conducted on live streaming data or commercial search services, in which the data and algorithms are continuously changing and the same data will never be obtained again. Also, user experiments can not be “re-run” with the same users as the users learn from the previous experience.

In IR, interactivity and user behaviour or search experience through whole search sessions (or sometimes even a task involving multiple sessions) become more important, in order to consider real-world contexts. Various algorithms and softwares to support such interactions have been studied and proposed. The data obtained from the users in such task-based or whole session-based studies are highly complex, comprising e.g. the nature of the tasks conducted as well as the characteristics of each user. More research is needed for developing a framework that is able to describe such complex, multi-dimensional data as well as for devising methods for proper scientific analysis of the data collected.

6.8.4 Actions to Improve reproducibility

Actions to improve reproducibility of user-oriented experiments include checklists for authors (and reviewers, editors, chairs, etc.), sample exemplary papers, method inventories, extended methodological sections in papers, and critical discussions on the components/tools/other data used. These are considered briefly below:

Checklists should be provided on the methodology applied in the study. Kelly [3] is a useful source for constructing a checklist for user-oriented IR studies. Examples of items to check are:

- Research questions and experimental design (latin square, intra/inter subject, etc.)
- Participant characteristics and the population they are claimed to represent
- Methods of data collection, including the experimental protocol, environmental conditions, and variables used in the study (how to describe, how to measure, operational definition, observables)
- Experimenter
- Retrieval systems and their interfaces
- Methods for data analysis, including assumptions of statistical analyses (and adjustments if assumptions are not met)
- Degree of control on the system by the experimenter

Exemplary papers representing various types of user experiments could be offered in some community-based repository and annotated for their strong features, see also next section.

Inventories of typical variables various types of user experiments and standard ways to operationalize and measure them in different (sample) study settings could be provided by the community), as further discussed below.

Methodological sections could be emphasized in document templates, author guidelines and review guidelines. More space might be allocated to these sections and or authors encouraged to provide methodological appendixes or technical reports.

Finally, authors could be encouraged to *critically discuss* how suitable the set of tools and collections is to answer the research questions, what claims can the tools/collection support, describing the generalizability of the findings on the basis of the tools/collection that have been used.

6.8.5 Community Support to Reproducibility

In order to embody the vision described above and foster reproducibility in user-oriented studies, the involvement of the research community is crucial and it should consist of two complementary actions:

1. Support to the creation of shared resources;
2. Taking up and implementation of shared practices.

When it comes to *shared resources*, we can foresee several examples of them:

- **Inventories:** in order to streamline the reproducibility process, there is a need for catalogues accounting for the most appropriate experimental designs, the kind of independent and dependent variables you typically encounter in these settings, how to describe and measure such variables, the proper data analysis methodologies and statistical validation methods to apply to these variables in the different experimental designs, and so on;
- **Do's and don'ts:** in order to facilitate the understanding and adoption of the above facilitators of reproducibility, real and hands-on examples of appropriate and inappropriate ways to carry out user-oriented experiments are needed to clearly explain why a seemingly appropriate experimental setup is or is not working as expected. This could be partnered with a selection of exemplary and well-known papers, which should be annotated and enriched with links and explanations related to the above inventories, in order to clarify the researcher how and when to apply a given approach by means of concrete and remarkable case studies;
- **Repositories:** the adoption of shared repositories to gather collections of documents, interaction data, tasks and topics, and more is a key step to extend the reach of reproducibility in user-oriented experimentation;

- **Data formats:** the development of commonly understood and well-documented data formats, which can be extended to specific needs, as well as the introduction of proper metadata (descriptive, administrative, copyright, etc.) to model, describe, and annotate the data and the experimental outcomes is a crucial factor in lowering the barriers to reproducibility in user-oriented experimentation.

The methodological instruments, the checklists, the critical discussions, the different kinds of shared resources previously described are all key “ingredients” for successfully reproducing user-oriented experiments but the actual catalyst is the systematic and wide adoption by the community of *shared practices*, effectively exploiting all of these “ingredients”, as also discussed in Sections 6.4 and 6.7.

References

- 1 W. B. Croft. Information retrieval and computer science: an evolving relationship. Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2–3. July 28 – August 1, 2003, Toronto, Canada.
- 2 C. Cleverdon. The Cranfield tests on index language devices. *Aslib Proceedings* 16, 6:173–194. 1967.
- 3 D. Kelly. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval*, 3(1–2), pp. 1–224. 2009.

7 Open problems

7.1 Open Research Problems in Reproducibility

Carole Goble (University of Manchester, GB) and Daniel Garijo (Technical University of Madrid, ES)

License © Creative Commons BY 3.0 Unported license
© Carole Goble and Daniel Garijo

When referring to reproducibility, we can distinguish two main types of research agendas, each with their scope and social implications. There is a **macro research agenda**, which consists of the topics of interest of the main funding agencies, and a **micro research agenda**, which would consist of the particular topics for new PhD students. While the macro agenda is influenced by the political tendencies of the moment, the micro agenda is influenced by the particular interests of researchers. Reproducibility initiatives may work fine for specific domains, but they may collapse when applying them at a macro level. Since most of the people in the group did not belong to funding agencies, the discussion focused on the micro agenda.

Regarding the social implications of reproducibility, an agenda should be issued in terms of productivity. Reproducibility can be seen as an investment for productivity, and part of its agenda should study and make explicit the correlation between these two features. Another challenge is addressing how the quality and quantity of the research work is affected by reproducibility. Currently, when given the opportunity, a researcher will choose to publish two publications rather than a highly reproducible one. **It is important to be able to show the long term value of high quality reproducible work.**

Another important aspect to take into consideration is the analysis of infrastructure, which includes the improvement of record keeping. The best way of holding trusted resources is to convince institutions to get involved. **Labs, companies and people are temporary,**

while institutions tend to last even centuries. In order to achieve this, it is critical to address the intellectual property rights of the resources to be archived. Having company-friendly licenses may help in their adoption.

7.1.1 Open Research Challenges

We summarize the main challenges related to reproducibility in the list below

1. What are the interventions needed to change of behavior of the researchers? Making a paper reproducible is often related to the ways people are used to work within a given community. Knowing which are the necessary changes to change the behaviour of scientists towards adopting reproducibility may help to make the transition in a more effective way.
2. Do reproducibility and replicability translate in long term impact for your work? By showing empirical proof of the impact of reproducible versus non reproducible work on a community, more authors may be convinced on adopting a reproducible approach.
3. How do we set the research environment for enabling reproducibility? If making a paper reproducible takes a lot of time, people will not do it. Instead, working towards the creation of environments for enabling reproducibility seems like a more sensible approach.
4. How can we obtain long term digital archiving? Having a long lasting record of the resources used for a paper is a crucial requirement for reproducibility. Existing institutions (libraries, church) have archived successfully knowledge for centuries, and we should learn their methods and apply it to software as well.
5. How can we track the components that are part of the materials that have been used in a project? A researcher may forget to include data considered trivial in an experiment, but that same data may crucial for another researcher that aims to reproduce it years later. Is it possible to auto-document your research?
6. Is it possible to define roles of contributors for getting the credit for a work? Capturing the finer grain of contributions is crucial to provide appropriate credit to all the contributors of a research work. Some initiatives have started proposing sharing taxonomies²⁸ and distribute credit [1], which are a first step towards addressing this challenge.
7. Can we measure the cost of reproducibility/repeatability/documentation? What are the difficulties for newcomers? Understanding how to lower the barrier for adopting reproducibility and its costs is another of the key aspects to take into consideration when convincing a community to make their work reproducible.

References

- 1 D. S. Katz, and A. M. Smith. (2014). Implementing transitive credit with JSON-LD. arXiv preprint arXiv:1407.5117.

²⁸ <http://casrai.org/CRediT>

Participants

- Vanessa Braganholo
Fluminense Federal
University, BR
- Fernando Chirigati
NYU Tandon School of
Engineering, US
- Christian Collberg
University of Arizona –
Tucson, US
- Shane Culpepper
RMIT University –
Melbourne, AU
- David De Roure
University of Oxford, GB
- Arjen P. de Vries
Radboud University
Nijmegen, NL
- Jens Dittrich
Universität des Saarlandes, DE
- Nicola Ferro
University of Padova, IT
- Juliana Freire
New York University, US
- Norbert Fuhr
Universität Duisburg-Essen, DE
- Daniel Garijo
Technical University
of Madrid, ES
- Carole Goble
University of Manchester, GB
- Kalervo Järvelin
University of Tampere, FI
- Noriko Kando
National Institute of Informatics –
Tokyo, JP
- Randall J. LeVeque
University of Washington –
Seattle, US
- Matthias Lippold
Universität Duisburg-Essen, DE
- Bertram Ludäscher
University of Illinois at
Urbana-Champaign, US
- Mihai Lupu
TU Wien, AT
- Tanu Malik
University of Chicago, US
- Rudolf Mayer
SBA Research – Wien, AT
- Alistair Moffat
The University of Melbourne, AU
- Kevin Page
University of Oxford, GB
- Raul Antonio Palma de Leon
Poznan Supercomputing and
Networking Center, PL
- Martin Potthast
Bauhaus-Universität Weimar, DE
- Andreas Rauber
TU Wien, AT
- Paul Rosenthal
TU Chemnitz, DE
- Claudio T. Silva
New York University, US
- Stian Soiland-Reyes
University of Manchester, GB
- Benno Stein
Bauhaus-Universität Weimar, DE
- Rainer Stotzka
KIT – Karlsruher Institut für
Technologie, DE
- Evelyne Viegas
Microsoft Research –
Redmond, US
- Stefan Winkler-Nees
DFG – Bonn, DE
- Torsten Zesch
Universität Duisburg-Essen, DE
- Justin Zobel
The University of Melbourne, AU



Eyewear Computing – Augmenting the Human with Head-Mounted Wearable Assistants

Edited by

Andreas Bulling¹, Ozan Cakmakci², Kai Kunze³, and James M. Rehg⁴

1 Max-Planck-Institut für Informatik – Saarbrücken, DE, bulling@mpi-inf.mpg.de

2 Google Inc. – Mountain View, US, ozancakmakci@google.com

3 Keio University – Yokohama, JP, kai@kmd.keio.ac.jp

4 Georgia Institute of Technology – Atlanta, US, rehg@gatech.edu

Abstract

The seminar was composed of workshops and tutorials on head-mounted eye tracking, egocentric vision, optics, and head-mounted displays. The seminar welcomed 30 academic and industry researchers from Europe, the US, and Asia with a diverse background, including wearable and ubiquitous computing, computer vision, developmental psychology, optics, and human-computer interaction. In contrast to several previous Dagstuhl seminars, we used an ignite talk format to reduce the time of talks to one half-day and to leave the rest of the week for hands-on sessions, group work, general discussions, and socialising. The key results of this seminar are 1) the identification of key research challenges and summaries of breakout groups on multimodal eyewear computing, egocentric vision, security and privacy issues, skill augmentation and task guidance, eyewear computing for gaming, as well as prototyping of VR applications, 2) a list of datasets and research tools for eyewear computing, 3) three small-scale datasets recorded during the seminar, 4) an article in ACM Interactions entitled “Eyewear Computers for Human-Computer Interaction”, as well as 5) two follow-up workshops on “Egocentric Perception, Interaction, and Computing” at the European Conference on Computer Vision (ECCV) as well as “Eyewear Computing” at the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp).

Seminar January 24–29, 2016 – <http://www.dagstuhl.de/16042>

1998 ACM Subject Classification H.5 Information Interfaces and Presentation (e.g., HCI), I.4 Image Processing and Computer Vision, K.4 Computer and Society

Keywords and phrases Augmented Human, Cognition-Aware Computing, Wearable Computing, Egocentric Vision, Head-Mounted Eye Tracking, Optics, Displays, Human-Computer Interaction, Security and Privacy

Digital Object Identifier 10.4230/DagRep.6.1.160



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

Eyewear Computing – Augmenting the Human with Head-mounted Wearable Assistants, *Dagstuhl Reports*, Vol. 6, Issue 1, pp. 160–206

Editors: Andreas Bulling, Ozan Cakmakci, Kai Kunze, and James M. Rehg



DAGSTUHL
REPORTS Dagstuhl Reports

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

1 Executive Summary

Andreas Bulling

Ozan Cakmakci

Kai Kunze

James M. Rehg

License © Creative Commons BY 3.0 Unported license

© Andreas Bulling, Ozan Cakmakci, Kai Kunze, and James M. Rehg

Main reference A. Bulling, K. Kunze, “Eyewear Computers for Human-Computer Interaction”, ACM Interactions, 23(3):70–73, 2016.

Computing devices worn on the human body have a long history in academic and industrial research, most importantly in wearable computing, mobile eye tracking, and mobile mixed and augmented reality. In contrast to traditional systems, body-worn devices are always with the user and therefore have the potential to perceive the world and reason about it from the user’s point of view. At the same time, given that on-body computing is subject to ever-changing usage conditions, on-body computing also poses unique research challenges.

This is particularly true for devices worn on the head. As humans receive most of their sensory input via the head, it is a particularly interesting body location for simultaneous sensing and interaction as well as cognitive assistance. Early egocentric vision devices were rather bulky, expensive, and their battery lifetime severely limited their use to short durations of time. Building on existing work in wearable computing, recent commercial egocentric vision devices and mobile eye trackers, such as Google Glass, PUPIL, and J!NS meme, pave the way for a new generation of “smart eyewear” that are light-weight, low-power, convenient to use, and increasingly look like ordinary glasses. This last characteristic is particularly important as it makes these devices attractive for the general public, thereby holding the potential to provide a research and product platform of unprecedented scale, quality, and flexibility.

While hearing aids and mobile headsets became widely accepted as head-worn devices, users in public spaces often consider novel head-attached sensors and devices as uncomfortable, irritating, or stigmatising. Yet with the advances in the following technologies, we believe eyewear computing will be a very prominent research field in the future:

- Increase in storage/battery capacity and computational power allows users to run eyewear computers continuously for more than a day (charging over night) gathering data to enable new types of life-logging applications.
- Miniaturization and integration of sensing, processing, and interaction functionality can enable a wide array of applications focusing on micro-interactions and intelligent assistance.
- Recent advances in real-life tracking of cognitive activities (e.g. reading, detection of fatigue, concentration) are additional enabling technologies for new application fields towards a quantified self for the mind. Smart eyewear and recognizing cognitive states go hand in hand, as naturally most research work in this field requires sensors.
- Cognitive scientists and psychologists have now a better understanding of user behavior and what induces behavior change. Therefore, smart eyewear could help users in achieving behaviour change towards their long term goals.

Eyewear computing has the potential to fundamentally transform the way machines perceive and understand the world around us and to assist humans in measurably and significantly improved ways. The seminar brought together researchers from a wide range of computing disciplines, such as mobile and ubiquitous computing, head-mounted eye tracking,

optics, computer vision, human vision and perception, privacy and security, usability, as well as systems research. Attendees discussed how smart eyewear can change existing research and how it may open up new research opportunities. For example, future research in this area could fundamentally change our understanding of how people interact with the world around them, how to augment these interactions, and may have a transformational impact on all spheres of life – the workplace, family life, education, and psychological well-being.

2 Table of Contents

Executive Summary

Andreas Bulling, Ozan Cakmakci, Kai Kunze, and James M. Rehg 161

Ignite Talks

Pervasive Sensing, Analysis, and Use of Visual Attention
Andreas Bulling 165

Egocentric Vision for social and cultural experiences
Rita Cucchiara 165

User Interfaces for Eyewear Computing
Steven K. Feiner 166

Action and Attention in First-person Vision
Kristen Grauman 167

Technology for Learning in Virtual and Augmented Reality
Scott W. Greenwald 168

Detecting Mental Processes and States from Visual Behaviour
Sabrina Hoppe 169

In pursuit of the intuitive interaction in our daily life
Masahiko Inami 169

Cognitive Activity Recognition in Real Life Scenario
Shoya Ishimaru 170

Reading-Life Log
Koichi Kise 171

Redesigning Vision
Kiyoshi Kiyokawa 172

Augmenting the Embodied Mind
Kai Kunze 173

Egocentric discovery of task-relevant interactions for guidance
Walterio Mayol-Cuevas 173

Haptic gaze interaction for wearable and hand-held mobile devices
Päivi Majaranta 174

From EyeWear Comptuing to Head Centered Sensing and Interaction
Paul Lukowicz 175

Eyewear Computing: Do we talk about security and privacy yet?
René Mayrhofer 175

A multimodal wearable system for logging personal behaviors and interests
Masashi Nakatani 176

Pupil – accessible open source tools for mobile eye tracking and egocentric vision research
Will Patera and Moritz Kassner 176

Smart Eyewear for Cognitive Interaction Technology
Thies Pfeiffer 177

Activity Recognition and Eyewear Computing
Philipp M. Scholl 178

Eyes, heads and hands: The natural statistics of infant visual experience <i>Linda B. Smith</i>	178
Head-Worn Displays (HWDs) for Everyday Use <i>Thad Starner</i>	179
Unsupervised Discovery of Everyday Activities from Human Visual Behaviour and 3D Gaze Estimation <i>Julian Steil</i>	179
Calibration-Free Gaze Estimation and Gaze-Assisted Computer Vision <i>Yusuke Sugano</i>	180
Wearable barcode scanning <i>Gábor Sörös</i>	180
Workshops and Tutorials	
Workshop: PUPIL <i>Moritz Kassner, Will Patera</i>	181
Workshop: Google Cardboard <i>Scott W. Greenwald</i>	184
Workshop: J!NS Meme <i>Shoya Ishimaru, Yuji Uema, Koichi Kise</i>	186
Tutorial: Head-Worn Displays <i>Kiyoshi Kiyokawa, Thad Starner, and Ozan Cakmakci</i>	187
Tutorial: Egocentric Vision <i>James M. Rehg, Kristen Grauman</i>	188
Challenges	189
Breakout Sessions	190
Group 1: Multimodal EyeWear Computing <i>Masashi Nakatani</i>	190
Group 2: Egocentric Vision <i>Rita Cucchiara, Kristen Grauman, James M. Rehg, Walterio W. Mayol-Cuevas</i> . .	193
Group 3: Security and Privacy <i>René Mayrhofer</i>	197
Group 4: Eyewear Computing for Skill Augmentation and Task Guidance <i>Thies Pfeiffer, Steven K. Feiner, Walterio W. Mayol-Cuevas</i>	199
Group 5: EyeWear Computing for Gaming <i>Thad Starner</i>	203
Group 6: Prototyping of AR Applications using VR Technology <i>Scott W. Greenwald</i>	204
Community Support	204
Datasets	204
Tools	205
Participants	206

3 Ignite Talks

3.1 Pervasive Sensing, Analysis, and Use of Visual Attention

Andreas Bulling (*Max-Planck-Institut für Informatik – Saarbrücken, DE*)

License © Creative Commons BY 3.0 Unported license
© Andreas Bulling

Main reference A. Bulling, “Pervasive Attentive User Interfaces”, *IEEE Computer*, 49(1):94–98, 2016.
URL <http://dx.doi.org/10.1109/MC.2016.32>

In this talk I motivated the need for new computational methods to sense, analyse and – most importantly – manage user attention continuously in everyday settings. This is important to enable future human-machine systems to cope with the ever-increasing number of displays and interruptions they cause. I provided a short overview of selected recent works in our group towards this vision, specifically head-mounted and appearance-based remote gaze estimation [2], short and long-term visual behaviour modelling and activity recognition, cognition-aware computing [3], attention modelling in graphical user interfaces [1], as well as collaborative human-machine vision systems for visual search target prediction and visual object detection.

References

- 1 Pingmei Xu, Yusuke Sugano, Andreas Bulling. *Spatio-Temporal Modeling and Prediction of Visual Attention in Graphical User Interfaces*. Proc. of the 34th ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), 2016. <http://dx.doi.org/10.1145/2858036.2858479>
- 2 Xucong Zhang, Yusuke Sugano, Mario Fritz, Andreas Bulling. *Appearance-Based Gaze Estimation in the Wild*. Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4511-4520, 2015. <http://dx.doi.org/10.1109/CVPR.2015.7299081>
- 3 Andreas Bulling, Thorsten Zander. *Cognition-Aware Computing*. *IEEE Pervasive Computing* 13 (3), pp. 80-83, 2014. <http://dx.doi.org/10.1109/mprv.2014.42>

3.2 Egocentric Vision for social and cultural experiences

Rita Cucchiara (*University of Modena, IT*)

License © Creative Commons BY 3.0 Unported license
© Rita Cucchiara

Joint work of R.Cucchiara, A. Alletto, G. Serra
Main reference A. Alletto, G. Serra, S. Calderara, R. Cucchiara, “Understanding Social Relationships in Egocentric Vision”, *Pattern Recognition*, 48(12):4082–4096, 2015.
URL <http://dx.doi.org/10.1016/j.patcog.2015.06.006>

Egocentric Vision concerns computer vision models and techniques for understanding what a person sees, from the first person’s point of view, with eyewear devices and centered on the human perceptual needs. Computer vision problems become more challenging when applied in such an unconstrained scenario, with an unknown camera motion due to the body and head movements of the camera wearer. In this research, we address how egocentric vision can be exploited to augment cultural and social experiences. It can be done in many ways. A first application is in attention driven life-logging: whenever vision solutions will be able to recognize the interest of the persons, what they are looking at and how much they do, personalized video summarization will be available to keep the memory of the experience,

to share them and to use it also for educational purpose. The second is more general: it concerns recognition of targets of interests in indoor museums or in outdoor unconstrained setting, as for instance in a cultural visit around a historical/artistic area. Here egocentric vision is required to understand the social relationships among people and friends, recognize what persons would like to see and be engaged with, and localize the person's position to suggest useful information in real time. It can be done in a simplified manner by using image search for similarity or more precisely by providing 2D and 3D registration. Here, there are many open problems such as, for instance, video registration in real-time, matching images taken at different weather or time conditions. New generation of eyewear devices, possibly provided with eye-tracking and with high connectivity to allow a fast image processing, will provide a big leap- forward in this field, and will open new research areas in egocentric vision for interaction with environment.

References

- 1 P. Varini, G. Serra, R. Cucchiara. *Egocentric Video Summarization of Cultural Tour based on User Preferences*. Proc. of the 23rd ACM International Conference on Multimedia, 2015.
- 2 R. Cucchiara, A. Del Bimbo. *Visions for Augmented Cultural Heritage Experience*. IEEE Multimedia 2014.
- 3 A. Alletto, G. Serra, R. Cucchiara, V. Mighali, G. Del Fiore, L. Patrono, L. Mainetti, *An Indoor Location-aware System for an IoTbased Smart Museum*. Internet of Things Journal, 2016.

3.3 User Interfaces for Eyewear Computing

Steven K. Feiner (Columbia University, US)

License © Creative Commons BY 3.0 Unported license
© Steven K. Feiner

Joint work of Feiner, Steven K.; Elvezio, Carmine; Oda, Ohan; Sukan, Mengu; Tversky, Barbara
Main reference O. Oda, C. Elvezio, M. Sukan, S.K. Feiner, B. Tversky, "Virtual Replicas for Remote Assistance in Virtual and Augmented Reality", in Proc. of the 28th Annual ACM Symp. on User Interface Software & Technology (UIST'15), pp. 405–415, ACM, 2015.
URL <http://dx.doi.org/10.1145/2807442.2807497>

Our research investigates how we can create effective everyday user interfaces that employ eyewear alone and in synergistic combination with other modalities. We are especially interested in developing multimodal interaction techniques that span multiple users, displays, and input devices, adapting as we move in and out of their presence. Domains that we have addressed include outdoor wayfinding, entertainment, and job performance. One approach that we employ is *environment management* – supervising UI design across space, time, and devices. (A specific example is *view management*, in which we automate the spatial layout of information by imposing and maintaining visual constraints on the locations of objects in the environment and their projections in our view.) A second approach is the creation of *hybrid user interfaces*, in which heterogeneous interface technologies are combined to complement each other. For example, we have used a video-see-through head-worn display to overlay 3D building models on their footprints presented on a multitouch horizontal tabletop display.

A research theme underlying much of our research is collaboration: developing user interfaces that support multiple users wearing eyewear working together, both co-located and remote. We are especially interested in remote task assistance, in which a remote subject-matter expert helps a less knowledgeable local user perform a skilled task. Two issues here are helping the local user quickly find an appropriate location at which to perform the

task, and getting them to perform the task correctly. In one project, we have developed and evaluated ParaFrustum, a 3D interaction and visualization technique that presents a range of appropriate positions and orientations from which to perform the task [3]. In a second project, we have developed and evaluated 3D interaction techniques that allow a remote expert to create and manipulate virtual replicas of physical objects in the local environment to refer to parts of those physical objects and to indicate actions on them [2].

Finally, much of our research uses stereoscopic eyewear with a relatively wide field of view to present augmented reality in which virtual media are geometrically registered and integrated with our perception of the physical world. We are also exploring the use of eyewear with a small monoscopic field of view, such as Google Glass. In this work, we are developing and evaluating approaches that do not rely on geometric registration [1].

References

- 1 Carmine Elvezio, Mengu Sukan, Steven Feiner, and Barbara Tversky. *Interactive Visualizations for Monoscopic Eyewear to Assist in Manually Orienting Objects in 3D*. Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2015), 180–181. <http://dx.doi.org/10.1109/ISMAR.2015.54>
- 2 Ohan Oda, Carmine Elvezio, Mengu Sukan, Steven Feiner, and Barbara Tversky. *Virtual replicas for remote assistance in virtual and augmented reality*. Proc. ACM Symposium on User Interface Software and Technology (UIST 2015), pp. 405–415. <http://dx.doi.org/10.1145/2807442.2807497>
- 3 Mengu Sukan, Carmine Elvezio, Ohan Oda, Steven Feiner, and Barbara Tversky. *Para-Frustum: Visualization techniques for guiding a user to a constrained set of viewing positions and orientations*. Proc. ACM Symposium on User Interface Software and Technology (UIST 2014), pp. 331–340. <http://dx.doi.org/10.1145/2642918.2647417>

3.4 Action and Attention in First-person Vision

Kristen Grauman (University of Texas at Austin, USA)

License © Creative Commons BY 3.0 Unported license
© Kristen Grauman

Joint work of K. Grauman, D. Jayaraman, Yong Jae Lee, Bo Xiong, Lu Zheng
URL <http://www.cs.utexas.edu/~grauman/>

A traditional third-person camera passively watches the world, typically from a stationary position. In contrast, a first-person (wearable) camera is inherently linked to the ongoing experiences of its wearer. It encounters the visual world in the context of the wearer’s physical activity, behavior, and goals. This distinction has many intriguing implications for computer vision research, in topics ranging from fundamental visual recognition problems to high-level multimedia applications.

I will present our recent work in this space, driven by the notion that the camera wearer is an active participant in the visual observations received. First, I will show how to exploit egomotion when learning image representations [1]. Cognitive science tells us that proper development of visual perception requires internalizing the link between “how I move” and “what I see” – yet today’s best recognition methods are deprived of this link, learning solely from bags of images downloaded from the Web. We introduce a deep feature learning approach that embeds information not only from the video stream the observer sees, but also the motor actions he simultaneously makes. We demonstrate the impact for recognition, including a scenario where features learned from ego-video on an autonomous car

substantially improve large-scale scene recognition. Next, I will present our work exploring video summarization from the first person perspective [3, 2]. Leveraging cues about ego-attention and interactions to infer a storyline, we automatically detect the highlights in long videos. We show how hours of wearable camera data can be distilled to a succinct visual storyboard that is understandable in just moments, and examine the possibility of person- and scene-independent cues for heightened attention. Overall, whether considering action or attention, the first-person setting offers exciting new opportunities for large-scale visual learning.

References

- 1 D. Jayaraman and K. Grauman. *Learning Image Representations Tied to Ego-Motion*. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, Dec 2015.
- 2 Y. J. Lee and K. Grauman. *Predicting Important Objects for Egocentric Video Summarization*. International Journal on Computer Vision, Volume 114, Issue 1, pp. 38–55, August 2015.
- 3 B. Xiong and K. Grauman. *Detecting Snap Points in Egocentric Video with a Web Photo Prior*. In Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, Sept 2014.

3.5 Technology for Learning in Virtual and Augmented Reality

Scott W. Greenwald (MIT – Cambridge, US)

License © Creative Commons BY 3.0 Unported license
© Scott W. Greenwald

Main reference S. W. Greenwald, M. Khan, C. D. Vazquez, P. Maes, “TagAlong: Informal Learning from a Remote Companion with Mobile Perspective Sharing”, in Proc. of the 12th IADIS Int’l Conf. on Cognition and Exploratory Learning in Digital Age (CELDA’15), 2015.

URL <http://dspace.mit.edu/handle/1721.1/100242>

Face-to-face communication is highly effective for teaching and learning. When the student is remote (e.g. in a situated context) or immersed in virtual reality, effective teaching can be much harder. I propose that in these settings, it can be done as well or better than face-to-face using a system that provides the teacher with the first-person perspective, along with real-time sensor data on the cognitive and attentional state of the learner. This may include signals such as EEG, eye gaze, and facial expressions. The paradigm of eyewear computing – to provide contextual feedback using the egocentric perspective as an input and output medium – is critical in the realization of such a system. In my current work, I investigate key communication affordances for effective teaching and learning in such settings.

3.6 Detecting Mental Processes and States from Visual Behaviour

Sabrina Hoppe (*Max-Planck-Institut für Informatik – Saarbrücken, DE*)

License © Creative Commons BY 3.0 Unported license
© Sabrina Hoppe

Joint work of Stephanie Morey, Tobias Loetscher, Andreas Bulling

Main reference S. Hoppe, T. Loetscher, S. Morey, A. Bulling, “Recognition of Curiosity Using Eye Movement Analysis”, in *Adjunct Proc. of the ACM Int’l Joint Conf. on Pervasive and Ubiquitous Computing (UbiComp’15)*, pp. 185–188, ACM, 2015.

URL <http://dx.doi.org/10.1145/2800835.2800910>

It is well known that visual behaviour in everyday life changes with activities and certain environmental factors like lightening conditions, but how about mental processes? In this talk, I introduced some of our initial work in this direction: we have shown that curiosity, as an exemplar personality trait, is reflected in visual behaviour and can be inferred from gaze data [1]. However, personality is just one out of many potential mental states to be investigated in this context – further processes of interest include confusion [3] and mental health [2]. If and how this information can be leveraged from gaze data remains an interesting open question in the field of eye wear computing.

References

- 1 Sabrina Hoppe, Tobias Loetscher, Stephanie Morey, Andreas Bulling. *Recognition of Curiosity Using Eye Movement Analysis*, Adj. Proc. of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), pp. 185–188, 2015.
- 2 Tobias Loetscher, Celia Chen, Sophie Wignall, Andreas Bulling, Sabrina Hoppe, Owen Churches, Nicole Thomas. *A study on the natural history of scanning behaviour in patients with visual field defects after stroke*, BMC Neurology, 15(64), 2015.
- 3 Nikolina Koleva, Sabrina Hoppe, Mohammed Mehdi Moniri, Maria Staudte, Andreas Bulling. *On the interplay between spontaneous spoken instructions and human visual behaviour in an indoor guidance task*, Proc. of the 37th Annual Meeting of the Cognitive Science Society, 2015.

3.7 In pursuit of the intuitive interaction in our daily life

Masahiko Inami (*University of Tokyo, JP*)

License © Creative Commons BY 3.0 Unported license
© Masahiko Inami

Joint work of Fan, Kevin; Huber, Jochen; Nanayakkara, Suranga; Kise, Koichi; Kunze, Kai; Ishimaru, Shoya; Tanaka, Katsuma; Uema, Yuji

Main reference K. Fan, J. Huber, S. Nanayakkara, M. Inami, “SpiderVision: Extending the Human Field of View for Augmented Awareness”, in *Proc. of the 5th Augmented Human Int’l Conf. (AH’14)*, Article No. 49, ACM, 2014.

URL <http://dx.doi.org/10.1145/2582051.2582100>

We have established the Living Lab Tokyo at National Museum of Emerging Science and Innovation (Miraikan). The main focus is to create a living environment for users, by embedding mini sensors to sense the various interactions between users and the environment for various purposes such as entertainment, safety, enhancing communication between family members and much more. Our group worked hand-in-hand with users to learn and explore more about the users’ needs at the Living Lab Tokyo. We have been achieved some interactive systems such as Senskin [1] and JINS MEME [2]. Recently, we are trying to expand our research target from a living room to a play ground. Then we have start a new research laboratory on Superhuman sports, which is a new challenge to reinvent sports that anyone can

enjoy, anywhere and anytime. Based on this concept we have developed a new head mounted device, SpiderVision [3] that extends the human field of view to augment a user's awareness of things happening behind one's back. SpiderVision leverages a front and back camera to enable users to focus on the front view while employing intelligent interface techniques to cue the user about activity in the back view. The extended back view is only blended in when the scene captured by the back camera is analyzed to be dynamically changing. In this project, we explore factors that affect the blended extension, such as view abstraction and blending area. We succeeded using the system on a sky field. I would like to discuss possible application to enhance our physical activities with a smart eyewear.

References

- 1 Masa Ogata, Yuta Sugiura, Yasutoshi Makino, Masahiko Inami, and Michita Imai. 2013. *SenSkin: adapting skin as a soft interface*. In Proceedings of the 26th annual ACM symposium on User interface software and technology (UIST'13). ACM, New York, NY, USA, 539-544. <http://dx.doi.org/10.1145/2501988.2502039>
- 2 Shoya Ishimaru, Kai Kunze, Katsuma Tanaka, Yuji Uema, Koichi Kise and Masahiko Inami. *Smarter Eyewear – Using Commercial EOG Glasses for Activity Recognition*. Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication (UbiComp2014). September 2014.
- 3 Kevin Fan, Jochen Huber, Suranga Nanayakkara and Masahiko Inami. *SpiderVision: Extending the Human Field of View for Augmented Awareness*. In Proceedings of the 5th Augmented Human International Conference (AH'14), ACM, Article 49, 2014.

3.8 Cognitive Activity Recognition in Real Life Scenario

Shoya Ishimaru (Osaka Prefecture University, JP)

License © Creative Commons BY 3.0 Unported license

© Shoya Ishimaru

Main reference S. Ishimaru, K. Kunze, K. Tanaka, Y. Uema, K. Kise, M. Inami, “Smart Eyewear for Interaction and Activity Recognition”, in Proc. of the 33rd Annual ACM Conf. Extended Abstracts on Human Factors in Computing Systems (CHI EA'15), pp. 307–310, ACM, 2015.

URL <http://dx.doi.org/10.1145/2702613.2725449>

As people can be motivated to keep physical fitness by looking back their step counts, tracking cognitive activities (e.g. the number of words they read in a day) can help them to improve their cognitive lifestyles. While most of the physical activities can be recognized with body-mounted motion sensors, recognizing cognitive activities is still a challenging task because body movements during the activities are limited and we need additional sensors. The sensors also should be for everyday use to track daily life. In this talk, I introduced three projects tracking our cognitive activities with affordable technologies. The first project is eye tracking on mobile tablets. Most of the mobile tablets like iPad have a front camera for video chat. We have analyzed the facial image from the camera and detected where the user is looking at [1]. The second project is activity recognition based on eye blinks and head motions. We have detected eye blinks by using the sensor built in Google Glass and combined head motion and eye blink for the classification [2]. The last project is signal analysis on commercial EOG glasses. We have detected eye blinks and horizontal eye movement by using three electrodes on J!NS MEME and used them for cognitive activity recognition [3]. In addition to detecting reading activity, the number of words a user read can be estimated from small eye movements appeared on the EOG signal. After presenting the three projects, I proposed an important topic we should tackle at Dagstuhl. Because of the rising of deep

learning, the critical issue for classification tasks might be switched from “What is the best feature” to “How can we create a large dataset with labeling”. Unlike image analysis, it’s difficult to correct human’s sensor data with ground truth. So I would like to discuss with participants how to organize large scale recording in real life through the seminar.

References

- 1 Kai Kunze, Shoya Ishimaru, Yuzuko Utsumi and Koichi Kise. *My reading life: towards utilizing eyetracking on unmodified tablets and phones*. Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication (UbiComp2013). September 2013.
- 2 Shoya Ishimaru, Jens Weppner, Kai Kunze, Koichi Kise, Andreas Dengel, Paul Lukowicz and Andreas Bulling. *In the Blink of an Eye – Combining Head Motion and Eye Blink Frequency for Activity Recognition with Google Glass*. Proceedings of the 5th Augmented Human International Conference (AH2014). March 2014.
- 3 Shoya Ishimaru, Kai Kunze, Katsuma Tanaka, Yuji Uema, Koichi Kise and Masahiko Inami. *Smart Eyewear for Interaction and Activity Recognition*. Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI2015). April 2015.

3.9 Reading-Life Log

Koichi Kise (Osaka Prefecture University, JP)

License © Creative Commons BY 3.0 Unported license
© Koichi Kise

Main reference K. Yoshimura, K. Kise, K. Kunze, “The Eye as the Window of the Language Ability: Estimation of English Skills by Analyzing Eye Movement While Reading Documents”, in Proc. of 13th Int’l Conf. on Document Analysis and Recognition (ICDAR’15), pp. 251–255, IEEE, 2015.

URL <http://dx.doi.org/10.1109/ICDAR.2015.7333762>

Reading to the mind is what exercise is to the body. It is a well-known sentence that indicates the importance of reading. Actually we are spending so much amount of time for everyday reading. In other words, it is rare to spend a whole day without reading anything. Unfortunately, however, this activity has not been recorded and thus we are not able to use it. Our project called “Reading-Life Log” is to record our reading activities at various levels by using a variety of sensors. As levels, we have proposed the amount of reading, the period of reading, the type of documents, the log of read words, and the level of understanding and interests. The first one, the amount of reading is measured by a method called “wordometer” [1] which estimates the number of read words based on eye movement or gaze. The period of reading is estimated by analyzing eye movements or gaze. Document types are recognized by using first person vision, or eye gaze. Read words are listed with the help of document image retrieval and eye gaze. As the estimation of the level of understanding, we have developed a method to estimate an English proficiency by analyzing eye gaze [2].

As a possible research direction, I also introduced our notion of “experiential supplement”, which is to record people’s experiences to give them to persons for their better experiences. This is based on our notion that ordinary people can be best helped by other ordinary people who have similar background knowledge and experiences.

As an important research topic for us I have pointed out that we need a way of fractional distillation of the data obtained by eye wears, because the data contains many factors such as a person, an object the person looks, his/her interests, difficulties, etc.

References

- 1 Kai Kunze, Hitoshi Kawaichi, Koichi Kise and Kazuyo Yoshimura, *The Wordometer – Estimating the Number of Words Read Using Document Image Retrieval and Mobile Eye Tracking*, Proc. 12th International Conference on Document Analysis and Recognition (ICDAR 2013), pp. 25–29 (2013-8).
- 2 Kazuyo Yoshimura, Koichi Kise and Kai Kunze, *The Eye as the Window of the Language Ability: Estimation of English Skills by Analyzing Eye Movement While Reading Documents*, Proc. 13th International Conference on Document Analysis and Recognition (ICDAR 2015), pp. 251–255 (2015-8).

3.10 Redesigning Vision

Kiyoshi Kiyokawa (Osaka University, JP)

License © Creative Commons BY 3.0 Unported license

© Kiyoshi Kiyokawa

Joint work of Kiyoshi Kiyokawa, Alexandor Plopski, Jason Orlosky, Yuta Itoh, Christian Nitschke, Takumi Toyama, Daniel Sonntag, Ernst Kruijff, Kenny Moser, J. Edward Swan II, Dieter Schmalstieg, Gudrun Klinker, and Haruo Takemura

Main reference A. Plopski, Y. Itoh, C. Nitschke, K. Kiyokawa, G. Klinker, and H. Takemura, “Corneal-Imaging Calibration for Optical See-Through Head-Mounted Displays”, in IEEE Transaction on Visualization and Computer Graphics (TVCG), 21(4):481–490, 2015.

URL <http://doi.ieeecomputersociety.org/10.1109/TVCG.2015.2391857>

Ideally, head worn displays (HWDs) are expected to produce any visual experience we imagine, however, compared to this ultimate goal, current HWDs are still far from perfect. One of fundamental problems of optical see-through HWDs is that the user’s exact view is not accessible as image overlay happens on the user’s retina. Our new concept, corneal feedback augmented reality (AR) is a promising approach to realize a closed feedback loop for better registration [1], color correction, contrast adjustment, accurate eye tracking and scene understanding, by continuously analyzing the reflections in the eye.

In the case of video see-through HWDs, our vision can be more flexibly redesigned. With a modular video see-through HWD we developed [2], our natural vision can be switched to a super eyesight on demand such as a super wide view and a super zoom view, which is controlled by natural eye gesture such as squinting. By combining advanced 3D reconstruction system, our vision can even be free from physical constraints so that we can change our viewpoint or the size of the world on demand.

References

- 1 Alexandor Plopski, Yuta Itoh, Christian Nitschke, Kiyoshi Kiyokawa, Gudrun Klinker, and Haruo Takemura, *Corneal-Imaging Calibration for Optical See-Through Head-Mounted Displays*, IEEE Transaction on Visualization and Computer Graphics (TVCG), Special Issue on IEEE Virtual Reality (VR) 2015, Vol. 21, No. 4, pp. 481–490, 2015.
- 2 Jason Orlosky, Takumi Toyama, Kiyoshi Kiyokawa, and Daniel Sonntag, *ModuLAR: Eye-controlled Vision Augmentations for Head Mounted Displays*, IEEE Transactions on Visualization and Computer Graphics (TVCG), Special Issue on International Symposium on Mixed and Augmented Reality (ISMAR) 2015, Vol. 21, No. 11, pp. 1259–1268, 2015.

3.11 Augmenting the Embodied Mind

Kai Kunze (Keio University – Yokohama, JP)

License © Creative Commons BY 3.0 Unported license
© Kai Kunze

People use mobile computing technology to track their health and fitness progress, from simple step counting to monitoring food intake to measuring how long and well they sleep. Can also quantify cognitive tasks in real-world requirements and in a second step can we use them to change behavior? There are patterns in the physiological signals and behavior of humans (facial expressions, nose temperature, eye movements, blinks etc.) that can reveal information about mental conditions and cognitive functions. We explore how to use these patterns to recognize mental states and in a second step searches for interactions to change human mental states by stimulating the users to change these patterns. We believe all sensing and actuation will be embedded in smart eye wear. We continue our research recognizing reading comprehension detecting cognitive load using eye movement analysis, first with slightly altered stationary setups (baseline experiments) then with extending to a more mobile, unconstrained setup, where we also use our pervasive reading detection/quantification methods (recording how much a person reads with unobtrusive smart glasses and looking for patterns in to detect what are healthy reading habits to improve comprehension) and our work to record seamless the user’s facial expression using an unobtrusive glasses design (affective wear). In initial trials with affective wear, we found indications that facial expressions change in relation to cognitive functions.

References

- 1 Amft, O., Wahl, F., Ishimaru, S., and Kunze, K. *Making regular eyeglasses smart*. IEEE Pervasive Computing 14(3):32–43, 2015.

3.12 Egocentric discovery of task-relevant interactions for guidance

Walterio Mayol-Cuevas (University of Bristol, UK)

License © Creative Commons BY 3.0 Unported license
© Walterio Mayol-Cuevas

Joint work of Walterio Mayol-Cuevas; Dima Damen; Teesid Leelasawassuk

Main reference D. Damen, T. Leelasawassuk, O. Haines, A. Calway, W. Mayol-Cuevas, “You-Do, I-Learn: Discovering Task Relevant Objects and their Modes of Interaction from Multi-User Egocentric Video”, in Proc. of the British Machine Vision Conf. (BMVC’14), BMVA Press, 2014.

URL <http://dx.doi.org/10.5244/C.28.30>

How to decide what is relevant from the world? What objects and situations are important to record and which ones to ignore? These are key questions for efficient egocentric perception. One first step toward egocentric relevance determination is the building of real-time models of attention which can help in building systems that are better at online data logging, systems that are assistive by knowing what to show, as well as understanding more about attention is part of the process needed to inform the type of sensors and feedback hardware needed. At Bristol, we have been developing methods that gate visual input into small snippets that contain information that is task relevant. These include objects that have been interacted with, the ways in which these objects have been used as well as video segments representative of the interactions that can later be offered to people for guidance. The ultimate goal of this work is to allow people to feel they can do much more than what they can currently do – to

have such a superhuman feeling will transform how wearables are perceived and make them more useful beyond simple recording devices. Our work is underpinned by various research strands we are exploring including attention estimation from IMU signals [1], the ability to fuse information from SLAM and gaze patterns with visual appearance for interaction relevance determination [2] and lightweight object recognition for fast and onboard operation [3]. Overall, the understanding of what is important and useful to provide guidance will ultimately lead to better informed algorithmic and hardware choices in eyewear computing.

References

- 1 T Leelasawassuk, D Damen, W Mayol-Cuevas. *Estimating Visual Attention from a Head Mounted IMU*. International Symposium on Wearable Computers (ISWC). 2015.
- 2 D Damen, T Leelasawassuk, O Haines, A Calway, W Mayol-Cuevas. *You-Do, I-Learn: Discovering Task Relevant Objects and their Modes of Interaction from Multi-User Egocentric Video*. British Machine Vision Conference (BMVC), Nottingham, UK. 2014.
- 3 Dima Damen, Pished Bunnun, Andrew Calway, Walterio Mayol-Cuevas, *Real-time Learning and Detection of 3D Texture-less Objects: A Scalable Approach*. British Machine Vision Conference. September 2012.

3.13 Haptic gaze interaction for wearable and hand-held mobile devices

Päivi Majaranta (University of Tampere, FI)

License  Creative Commons BY 3.0 Unported license

© Päivi Majaranta

Joint work of Majaranta, Päivi; Kangas, Jari; Isokoski, Poika; Špakov, Oleg; Rantala, Jussi; Akkil, Deepak; Raisamo, Roope

Gaze-based interfaces provide means for communication and control. It is known from previous research that gaze interaction is highly beneficial for people with disabilities (see e.g. work done within the COGAIN network reported in [3]). Gaze interaction could also potentially revolutionize pervasive and mobile interaction. Haptics provide a private channel for feedback, as it is only felt by the person wearing or holding the device, e.g. on eye glass frames [1] or hand-held devices. We found such haptic feedback useful for example in controlling a mobile device with gaze gestures. Haptic feedback can significantly reduce error rates and improve performance and user satisfaction [2]. In addition to gaze interaction with mobile devices, we see a lot of potential in using gaze as a channel to interact with our surroundings in the era of Internet-of-Things. One challenging question that we wish to tackle is how to enable more direct interaction with the objects instead of sending the commands via a screen.

References

- 1 Kangas, J., Akkil, D., Rantala, J., Isokoski, P., Majaranta, P., and Raisamo, R. (2014a). *Using Gaze Gestures with Haptic Feedback on Glasses*. Proc. 8th Nordic Conference on Human-Computer Interaction. ACM, 1047-1050. <http://dx.doi.org/10.1145/2639189.2670272>
- 2 Kangas, J., Akkil, D., Rantala, J., Isokoski, P., Majaranta, P. and Raisamo, R. (2014b). *Gaze Gestures and Haptic Feedback in Mobile Devices*. Proc. SIGCHI Conference on Human Factors in Computing Systems. ACM, 435-438. <http://dx.doi.org/10.1145/2556288.2557040>
- 3 Majaranta, P., Aoki, H., Donegan, M., Hansen, D.W., Hansen, J.P., Hyrskykari, A., and Riih a, K-J. (Eds.) *Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies*, IGI Global, 2012. <http://dx.doi.org/10.4018/978-1-61350-098-9>

3.14 From EyeWear Computing to Head Centered Sensing and Interaction

Paul Lukowicz (DFKI Kaiserslautern, DE)

License  Creative Commons BY 3.0 Unported license
 © Paul Lukowicz
 URL <http://ei.dfki.de/en/home/>

Eyewear Computing in the for currently implemented by e.g. Google Glass predominately aims at making switching between the digital and the physical world faster and easier. This is the proposition on which consumers are likely to be initially adopting this technology. However, in the long run, the more profound impact is the fact that it creates the possibility of putting advanced sensing and interaction modalities at the users' head. This in turn creates access to a broad range of information sources not accessible at any other body location which is likely to revolutionize activity and context recognition.

3.15 Eyewear Computing: Do we talk about security and privacy yet?

René Mayrhofer (Universität Linz, AT)

License  Creative Commons BY 3.0 Unported license
 © René Mayrhofer

Like other wearable devices, eyewear is subject to typical threats to security and privacy. In my talk, I outlined four categories of threats that seem especially relevant to eyewear computing. Three of these are shared with other mobile device categories such as smart phones: *device-to-user authentication* is required to make sure that a device has not been physically replaced or tampered with (and which has not been sufficiently addressed for most device shapes [1]); *emphuser-to-device authentication* to prevent malicious use of devices by other users (which is mostly solved for smart phones citehintze-locked-device-usage, but still largely open for eyewear devices); and *emphdevice-to-device authentication* to ensure that wireless links are established correctly (strongly depending on the scenario, some solutions exist that may be directly applicable to eyewear computing [3]). The fourth threat of *privacy* is also common to all wearable devices, but due to typical inclusion of cameras and microphones, is potentially harder to address for eyewear devices. We expect the application n of cross-device authentication methods to have significant impact especially for security in eyewear computing.

References

- 1 Rainhard D. Findling, Rene Mayrhofer. *Towards Device-to-User Authentication: Protecting Against Phishing Hardware by Ensuring Mobile Device Authenticity using Vibration Patterns*. 14th International Conference on Mobile and Ubiquitous Multimedia (MUM'15), 2015
- 2 Daniel Hintze, Rainhard D. Findling, Muhammad Muaaz, Sebastian Scholz, Rene Mayrhofer. *Diversity in Locked and Unlocked Mobile Device Usage*. Adjunct Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2014). 379–384, 2014
- 3 Rene Mayrhofer. *Ubiquitous Computing Security: Authenticating Spontaneous Interactions*. Vienna University, 2008

3.16 A multimodal wearable system for logging personal behaviors and interests

Masashi Nakatani (University of Tokyo, JP)

License © Creative Commons BY 3.0 Unported license
© Masashi Nakatani

Joint work of Liang, Feng; Miyazaki, Hazuki; Minamizawa, Kouta; Nakatani, Masashi; Tachi, Susumu

URL <http://www.merkel.jp/research>

We propose a wearable/mobile system that can capture our daily life experiences with multimodality (vision, sound, and haptics). This device is consisted of wearable devices that captures ego-centric view through a camera, microphones for audio information and nine-axis inertial motion sensor as haptics information. Collected data is analyzed based on individual interests to the environment, then can be used for predicting users' interest based on statistics data. By combining with the state-of-art smart eyewear technology, we aim at segmenting captured audio-visual scene based on measured dataset (mainly haptics information) as well as personal interest as ground truth. This system would be helpful for providing custom-made service based on personal interests, such as guided tour of the city, hike planning, and trail running etc. This system may also provide benefit for elderly people who may need life support in their daily lives.

3.17 Pupil – accessible open source tools for mobile eye tracking and egocentric vision research

Will Patera (Pupil Labs – Berlin, DE) and Moritz Kassner (Pupil Labs – Berlin, DE)

License © Creative Commons BY 3.0 Unported license
© Will Patera and Moritz Kassner

Main reference M. Kassner, W. Patera, A. Bulling, “Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction”, in Proc. of the ACM Int'l Joint Conf. on Pervasive and Ubiquitous Computing (UbiComp'14), pp. 1151–1160, ACM, 2014.

URL <http://dx.doi.org/10.1145/2638728.2641695>

URL <https://pupil-labs.com/>

Pupil is an accessible, affordable, and extensible open source platform for eye tracking and egocentric vision research. It comprises a lightweight modular eye tracking and egocentric vision headset as well as an open source software framework for mobile eye tracking. Accuracy of 0.6 degrees and precision 0.08 degrees can be obtained. Slippage compensation is implemented using a temporal 3D model of the eye. Pupil is used by a diverse group of researchers around the world. The primary mission of Pupil is to create an open community around eye tracking methods and tools.

References

- 1 Moritz Kassner, William Patera, Andreas Bulling, *Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction*. Proc. of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), pp. 1151–1160, 2014.

3.18 Smart Eyewear for Cognitive Interaction Technology

Thies Pfeiffer (Universität Bielefeld, DE)

License © Creative Commons BY 3.0 Unported license
© Thies Pfeiffer

Main reference T. Pfeiffer, P. Renner, N. Pfeiffer-Leßmann, “EyeSee3D 2.0: Model-based Real-time Analysis of Mobile Eye-Tracking in Static and Dynamic Three-Dimensional Scenes”, in Proc. of the 9th Biennial ACM Symp. on Eye Tracking Research & Applications, pp. 189–196, ACM, 2016.

URL <http://dx.doi.org/10.1145/2857491.2857532>

We are currently following a line of research in which we focus on assistance systems for humans that are adaptive to the competences and the current cognitive state of the user. In the project ADAMAAS (Adaptive and Mobile Action Assistance in Daily Living Activities), for example, we are addressing the groups of elderly and people with mental handicaps. One main research question is, what kind of assistance to provide in which situation (including competence and cognitive state) to support the person in maintaining an autonomous life. Other scenarios we are targeting with different projects are sports training, assistance in decision tasks and chess playing.

We believe that smart eyewear is a key technology here, as in the targeted activities people have to make use of their hands. Coupled with sensor technologies, such as sensors for egocentric vision or eye tracking, we aim to assess the current cognitive state. Paired with display technologies for augmented reality (visual or acoustic) smart eyewear can provide contextual feedback that is adaptive to the current progress and level of expertise of the wearer.

One method we are applying to follow this goal is virtual reality simulation for a rapid prototyping of different design alternatives for the hardware and the user interface. For this we make use of either a fully immersive CAVE or HMDs, such as Oculus Rift or Samsung Gear VR. In one of the breakout sessions here at Dagstuhl, we were able to discuss merits and challenges of this approach in more details and from different perspectives.

References

- 1 Thies Pfeiffer, Patrick Renner, Nadine Pfeiffer-Leßmann (2016, to appear). *EyeSee3D 2.0: Model-based Real-time Analysis of Mobile Eye-Tracking in Static and Dynamic Three-Dimensional Scenes*. In ETRA’16: 2016 Symposium on Eye Tracking Research and Applications Proceedings. <http://dx.doi.org/10.1145/2857491.2857532>
- 2 Patrick Renner, Thies Pfeiffer (2015). *Online Visual Attention Monitoring for Mobile Assistive Systems*. In SAGA 2015: 2nd International Workshop on Solutions for Automatic Gaze Data Analysis (pp. 14-15). eCollections Bielefeld. <http://biecoll.ub.uni-bielefeld.de/volltexte/2015/5382/>
- 3 Jella Pfeiffer, Thies Pfeiffer, Martin Meißner (2015). *Towards attentive in-store recommender Systems*. In Annals of Information Systems: Vol. 18. Reshaping Society through Analytics, Collaboration, and Decision Support (pp. 161-173). Springer International Publishing.

3.19 Activity Recognition and Eyewear Computing

Philipp M. Scholl (Universität Freiburg, DE)

License © Creative Commons BY 3.0 Unported license
© Philipp M. Scholl

Main reference P. Scholl, K. Van Laerhoven, “Wearable digitization of life science experiments”, in Proc. of the 2014 ACM Int’l Joint Conf. on Pervasive and Ubiquitous Computing: Adjunct Publication, pp. 1381–1388, ACM, 2014.

URL <http://dx.doi.org/10.1145/2638728.2641719>

Combining motion-based Activity Recognition and Eyewear Computing could allow for new ways of documenting manual work. In a wetlab environment, protective garment is necessary to avoid contamination of the experiment subjects, as well as protecting the experimenter from harmful agents. However, the taken procedure needs to minutiously documented. In the sense of Vannevar Bush’s vision[1] of a scientist wearing a head-mounted camera records his experiment. These recordings still need to be indexed to be useful. The idea is to detect well-defined activities from wrist motion, and similar sensors, which can serve as an additional index to these recordings, serving as external memories for scientists. These recordings can either be reviewed post-experiment or accessed implicitly while the experiment is on-going.

References

- 1 Bush, Vannevar. *As we may think*. The atlantic monthly, pp. 101–108, 1945.

3.20 Eyes, heads and hands: The natural statistics of infant visual experience

Linda B. Smith (Indiana University – Bloomington, US)

License © Creative Commons BY 3.0 Unported license
© Linda B. Smith

Main reference S. Jayaraman, C. Fausey, L. B. Smith, “The Faces in Infant-Perspective Scenes Change over the First Year of Life”, PLoS ONE, 10(5):e0123780, 2015.

URL <http://dx.doi.org/10.1371/journal.pone.0123780>
URL <http://www.iub.edu/~cogdev>

The visual objects labeled by common nouns – truck, dog, cup – are so readily recognized and their names so readily generalized by young children that theorists have suggested that these “basic-level” categories “carve nature at its joints.” This idea is at odds with contemporary understanding of visual object recognition, both in human visual science and in computational vision. In these literatures, object recognition is seen as a hard and unsolved. If object categories are “givens” for young perceivers, theorists of human and machine vision do not yet know how they are given. Understanding the properties of infant and toddler egocentric vision – and the coupling of those properties to the changing motor abilities provides potentially transformative new insights into typical and atypical human developmental process, and, perhaps, to machine vision. The core idea is that that the visual regularities that t young perceivers encounter is constrained by the limits of time and place and by the young child’s behavior, including the behavior of eyes, heads and hands. Changes in infant motor abilities gate and order regularities and in a sense, guide the infant through a series of sequential tasks and through a search space for an optimal solution to the problem of recognizing objects under varied and nonoptimal conditions Although the natural statistics of infant experience may not quite “carve nature at its joints,” but we propose they make those joints easier to find. In pursuit of this idea, we have collected and are analyzing a large corpus of infant-perspective scenes, about 500 total hours of video, 54 million images.

3.21 Head-Worn Displays (HWDs) for Everyday Use

Thad Starner (*Georgia Institute of Technology – Atlanta, US*)

License © Creative Commons BY 3.0 Unported license
© Thad Starner

Consumer purchase and use of eyewear is driven more by fashion than just about any other feature. Creating head-worn displays (HWDs) for everyday use must put fashion first. Due to lack of familiarity with the use of the devices, HWD consumers often desire features that are impractical or unnecessary. Transparent, see-through displays require more power, are more expensive, are difficult to see in bright environments, and result in inferior performance on many practical virtual and physical world tasks. Yet consumers will less readily try opaque, see-around displays because they are not aware of the visual illusion that makes these displays appear see-through. Many desire wide field-of-view (FOV) displays. However, the human visual system only sees in high resolution in a small 2 degree area called the fovea. Small FOV consumer displays (i.e., smart phones) have been highly successful for reading books, email, messaging, mobile purchasing, etc. A typical smart phone has a 8.8x15.6 degree FOV. Small FOV displays are much easier to manufacture with fashionable eyewear and will likely be the first successful (> 2m) consumer HWD product. Head weight should be kept under 75g for everyday use. A traveling exhibit of previous HWDs can be found at <http://wcc.gatech.edu/exhibition>

References

- 1 Thad Starner. *The Enigmatic Display*. IEEE Pervasive Computing 2(1):133-135, 2003.

3.22 Unsupervised Discovery of Everyday Activities from Human Visual Behaviour and 3D Gaze Estimation

Julian Steil (*Max-Planck-Institut für Informatik – Saarbrücken, DE*)

License © Creative Commons BY 3.0 Unported license
© Julian Steil

Joint work of Andreas Bulling, Yusuke Sugano, Mohsen Mansouryar

Main reference J. Steil and A. Bulling, “Discovery of Everyday Human Activities From Long-Term Visual Behaviour Using Topic Models”, in Proc. of the 2015 ACM Int’l Joint Conf. on Pervasive and Ubiquitous Computing (UbiComp), pp. 75–85, 2015.

URL <http://dx.doi.org/10.1145/2750858.2807520>

Practically everything that we do in our lives involves our eyes, and the way we move our eyes is closely linked to our goals, tasks, and intentions. Thus, the human visual behaviour has significant potential for activity recognition and computational behaviour analysis. In my talk I briefly discussed the ability of an unsupervised method to discover everyday human activities only using long-term eye movement video data [1]. Moreover, I presented a novel 3D gaze estimation method for monocular head-mounted eye trackers which directly maps 2D pupil positions to 3D gaze directions [2]. My current research is driven by the idea to shed some light on attention allocation in the real world.

References

- 1 Julian Steil and Andreas Bulling. *Discovery of Everyday Human Activities From Long-Term Visual Behaviour Using Topic Models*. Proc. of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), pp. 75-85, 2015. <http://dx.doi.org/10.1145/2750858.2807520>

- 2 Mohsen Mansouryar and Julian Steil and Yusuke Sugano and Andreas Bulling. *3D Gaze Estimation from 2D Pupil Positions on Monocular Head-Mounted Eye Trackers*. Proc. of the 9th ACM International Symposium on Eye Tracking Research & Applications (ETRA), pp. 197-200, 2016. <http://dx.doi.org/10.1145/2857491.2857530>

3.23 Calibration-Free Gaze Estimation and Gaze-Assisted Computer Vision

Yusuke Sugano (Max-Planck-Institut für Informatik – Saarbrücken, DE)

License © Creative Commons BY 3.0 Unported license

© Yusuke Sugano

Joint work of Bulling, Andreas; Matsushita, Yasuyuki; Sato, Yoichi

Main reference Y. Sugano and A. Bulling, “Self-Calibrating Head-Mounted Eye Trackers Using Egocentric Visual Saliency”, in Proc. of the 28th ACM Symp. on User Interface Software and Technology (UIST’15), pp. 363–372, 2015.

URL <http://dx.doi.org/10.1145/2807442.2807445>

Human gaze can provide a valuable resource for eyewear computing systems to understand both internal states of the users (e.g., their activities) and external environments (e.g., scene categories). In this ignite talk I presented a brief overview of my previous researches from calibration-free gaze estimation to gaze-assisted computer vision techniques. One approach for calibration-free gaze estimation is to use visual saliency maps estimated from the scene video as probabilistic training data [1]. Another approach is to take a purely machine learning-based approach to train an image-based gaze estimator using a large amount of eye images with ground-truth gaze direction labels [2]. If gaze estimation can be naturally integrated into daily-life scenarios with these techniques, collaboration between eyewear computers and human attention will have larger potential for future investigation. For example, gaze can guide computers to find semantically important objects in the images [3], and is expected to provide important information for user- and task-specific image understanding.

References

- 1 Yusuke Sugano and Andreas Bulling, *Self-Calibrating Head-Mounted Eye Trackers Using Egocentric Visual Saliency*, in Proc. of the 28th ACM Symposium on User Interface Software and Technology (UIST), pp. 363-372, 2015.
- 2 Yusuke Sugano, Yasuyuki Matsushita and Yoichi Sato, *Learning-by-Synthesis for Appearance-based 3D Gaze Estimation*, Proc. 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014), June 2014.
- 3 Yusuke Sugano, Yasuyuki Matsushita and Yoichi Sato, *Graph-based Joint Clustering of Fixations and Visual Entities*, ACM Transactions on Applied Perception (TAP), 10(2):1-16, June 2013.

3.24 Wearable barcode scanning

Gábor Sörös (ETH Zürich, CH)

License © Creative Commons BY 3.0 Unported license

© Gábor Sörös

URL <http://people.inf.ethz.ch/soeroesg/>

Ubiquitous visual tags like barcodes and QR codes are the most prevalent links between physical objects and digital information. Technological advancements in wearable computing

and mobile computer vision may radically expand the adoption of visual tags because smart glasses and other wearable devices enable instant scanning on the go. I develop robust and fast methods to overcome limitations and add advanced features that can make wearable barcode scanning an attractive alternative of traditional laser scanners. I present an overview of my research on tag localization, motion blur compensation, and gesture control on resource-constrained wearable computers.

References

- 1 Gábor Sörös, Stephan Semmler, Luc Humair, Otmar Hilliges, *Fast Blur Removal for Wearable QR Code Scanners*, Proc. of the 19th International Symposium on Wearable Computers (ISWC 2015)
- 2 Jie Song, Gábor Sörös, Fabrizio Pece, Sean Fanello, Shahram Izadi, Cem Keskin, Otmar Hilliges, *In-air Gestures Around Unmodified Mobile Devices*, Proc. of the 27th ACM User Interface Software and Technology Symposium (UIST 2014)
- 3 Gábor Sörös, Christian Floerkemeier, *Blur-Resistant Joint 1D and 2D Barcode Localization for Smartphones*, Proc. of the 12th International Conference on Mobile and Ubiquitous Multimedia (MUM 2013)

4 Workshops and Tutorials

4.1 Workshop: PUPIL

Moritz Kassner (*Pupil Labs UG*)

Will Patera (*Pupil Labs UG*)

License  Creative Commons BY 3.0 Unported license
© Moritz Kassner, Will Patera

Introduction

Will Patera and Moritz Kassner (co-founders of Pupil Labs) conducted a workshop using Pupil – eye tracking and egocentric vision research tool with the members of the Dagstuhl seminar. The workshop was organized in four general parts: an overview of to the Pupil platform, setup and demonstration demo, data collection in groups using Pupil, and concluding presentation of results from each group and feedback on the tool.

Hardware

Pupil is a head-mounted video based eye tracking and egocentric vision research tool. The headset material is laser sintered polyamide 12. Pupil attempts to reduce slippage by deforming the geometry of the headset prior to fabrication to make a comfortable and snug fit, and keeping the headset weight low. A binocular configuration with two 120hz eye cameras and one 120hz scene camera weighs 47 grams. The headset is modular and can be set up for egocentric vision, monocular eye tracking, or eye only cameras. The headset connects to a computer via high speed USB 2.0. Cameras used in the Pupil headset are UVC compliant.



■ **Figure 1** In the eye tracking workshop organised by Moritz Kassner and Will Partera (Pupil Labs UG, Berlin), seminar participants obtained hands-on experience with PUPIL, an open source platform for head-mounted eye tracking and egocentric vision research.

Software

The software is open source¹ Striving for concise, readable implementation and a plugin architecture. There are two main software applications/bundles; Pupil Capture for real-time and Pupil Player for offline visualization and analysis.

Working Principles

Pupil is a video based tracker with one camera looking at the scene and another camera looking at your pupil. Dark pupil tracking technique. Eye is illuminated in the infrared. Pupil detection is based on contour ellipse fitting and described in detail here in [1]. Performance was evaluated using the Swirski dataset described in [2]. For gaze mapping we need a transformation to go from pupil space to gaze space. One method is regression based gaze mapping. Mapping parameters are obtained from a 9 point marker calibration. Accuracy is 0.6 deg, precision 0.08 deg are nominal. Pupil dilation and movement of the headset degrades accuracy.

Recent Work

Current trackers assume the headset is "screwed to the head". You need to compensate for the movement of the headset. Additionally, a model-less search for the ellipse yields poor results when the pupil is partially obstructed by reflections or eyelashes or eyelids. With the use of a 3D model of the eye and a pinhole model of the camera based on Swirski's work in [3] we can model the eyeball as a sphere and the pupil as a disk on that sphere. The sphere used is based on an average human eyeball diameter is 24mm. The state of the model is

¹ <https://github.com/pupil-labs/pupil>

the position of the sphere in eye camera space and two rotation vectors that describe the location of the pupil on the sphere.

Using temporal constraints and competing eye models we can detect and compensate for slippage events when 2D pupil evidence is strong. In case of weak 2D evidence we can use constraint from existing models to robustly fit pupils with considerably less evidence than before.

With a 3D location of the eye and 3D vectors of gaze we don't have to rely of polynomials for gaze mapping. Instead we use a geometric gaze mapping approach. We model the world camera as a pinhole camera with distortion, and project pupil line of sight vectors onto the world image. For this we need to know the rotation translation of world and eye camera. This rigid transformation is obtained in a 3 point calibration routine. At the time of writing we simply assume that the targets are equally far away and minimize the distance of the obtained point pairs. This will be extended to infer distances during calibration.

Workshop Groups

At the workshop, participants split into eight groups to explore various topics related to head-mounted eye tracking and collect three small-scale datasets.

- Group 1 – Päivi Majaranta, Julian Steil, Philipp M. Scholl, Sabrina Hoppe – “Mutual Gaze” – used Pupil to study two people playing table tennis and synchronized videos.
- Group 2 – René Mayrhofer, Scott Greenwald – used Pupil to study the iris and see if eye cameras on Pupil could be used for iris detection.
- Group 3 – Thies Pfeiffer, Masashi Nakatani – used fiducial markers on the hands and computer screen and proposed a method to study typing skills on a “new keyboard” (e.g. German language keyboard vs American English language keyboard)
- Group 4 – Thad Starner, Paul Lukowicz, Rita – “Qualitative tests” – tested calibration and stability of the system
- Group 5 – Yusuke Sugano, Walterio Mayol-Cuevas, Gábor Sörös – “Indoor outdoor gaze” – recorded datasets in varying environments (indoor, outdoor, transition between) and varying activities: cycling, walking, vacuuming.
- Group 6 – James M. Rehg, Linda B. Smith, Ozan Cakmakci, Kristen Grauman – “Social Gaze Explorers” recorded gaze data during a 3 speaker interaction and experimented with simple approaches to detect when a wearer switches their attention from one speaker to another.
- Group 7 – Shoya Ishimaru, Koichi Kise, Yuji Uema – JINS Meme + Pupil – This group demonstrated a proof of concept combining JINS MEME EOG eye tracking device with Pupil. Pupil could be used to collect ground truth data for EOG systems like JINS MEME.
- Group 8 – Steven K. Feiner, Kiyoshi Kiyokawa, Masahiko Inami – used pupil while performing everyday tasks like making coffee and having conversations in a group.

References

- 1 Moritz Kassner, Will Patera, Andreas Bulling. *Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction*. Adj. Proc. of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), pp. 1151-1160, 2014. <http://dx.doi.org/10.1145/2638728.2641695>
- 2 Lech Świrski, Andreas Bulling, Neil Dodgson. *Robust, real-time pupil tracking in highly off-axis images*. Proc. of the 7th International Symposium on Eye Tracking Research and Applications (ETRA), pp. 173-176, 2012. <http://dx.doi.org/10.1145/2168556.2168585>

- 3 Lech Świrski, Neil A. Dodgson. *A fully-automatic, temporal approach to single camera, glint-free 3D eye model fitting*. Proc. of the 3rd International Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction (PETMEI), 2013.

4.2 Workshop: Google Cardboard

Scott W. Greenwald (MIT)

License  Creative Commons BY 3.0 Unported license
© Scott W. Greenwald

Scott Greenwald led a workshop on the Google Cardboard. Google Cardboard is a physical enclosure, consisting primarily of cardboard and two plastic lenses, and matching set of software constructs that turns almost any smartphone into a VR headset. Thanks to Ozan (Google Inc), the workshop was able to provide a Google Cardboard device for every participant to use and take home. Scott prepared a website that was used to document and distribute three live examples and a demo. The live examples could be opened directly on the website with the participants own mobile device. The examples and demo are described in the sections that follow.

Example 1: WebVR Boilerplate

When loading example using a web browser on a mobile device, users see an immersive animated video clip. They can pan around using touch gestures or moving their device in space. By pressing a Cardboard icon, they can switch into stereoscopic mode for viewing using Google Cardboard.

The example illustrates the seamless process of distributing cross-platform VR content



■ **Figure 2** Another workshop organised by Scott Greenwald (MIT) enabled participants to experience Google Cardboard using 30 devices kindly donated by Google Inc, US.

viewable on Cardboard. Mobile web browser technology allows this functionality to be built into a webpage and viewed on any sufficiently-equipped mobile device. In addition to being easier to distribute, this also accelerates the development workflow. The next important observation is that the same content was view with and without the VR headset. This is the concept of Responsive WebVR , an important step for making the VR-enabled web accessible on all devices.

The 3D environment the user sees in the example consists of a single sphere. The virtual scene camera through which the viewer is looking is located inside the sphere, and can be rotated using either touch gestures or physically moving the mobile device. The core mobile web technology that makes this possible is the WebGL part of the HTML5 specification. WebGL allows web pages to leverage mobile graphics hardware previously inaccessible inside the browser. Although it is possible to use the browser's built in WebGL language constructs directly, Scott advises using the Three.js framework which raises the level of abstraction and thereby simplifies the process of using the browser's WebGL capabilities.

Example 2: See-through video AR

This Cardboard-only video see-through app allows the user to use a button interaction to place playful markers in field of view that track the environment. It demonstrates a computer vision algorithm running on a live video stream in the mobile web browser. This is made possible by an HTML5 Media Capture API called `getUserMedia`. It allows the browser to access webcam and microphone hardware on the host device. For optical flow tracking, the example uses `jsfeat`, a JavaScript-based computer vision library.

Example 3: Social with WebRTC (Demo)

I demonstrate a two-device configuration where a cardboard user sees a video see-through application, and is able to see visual markers sent by a remote companion on a tablet and desktop system. The important aspect of this system is that it uses an HTML5 real-time communication technology called WebRTC. The example implementation shows the robustness of both `getUserMedia`, also shown in the previous example, and WebRTC for real-time sending of marker coordinates.

Example 4: Neuron Viewer

This example shows a high-resolution 3D model using the Responsive WebVR boilerplate (see Example 1). In contrast to the previous example building responsive WebVR, this example shows a high-resolution 3D geometry. This showcases the capabilities of mobile graphics hardware to not only show photographic content, but also polygon-based geometric content.

Conclusion

Most participants were able to successfully run most of the examples. Some participants' devices, however were insufficient or incompatible with the HTML5 technologies required. This demonstrates that, at this moment in time, the cross-platform vision of the mobile web is not yet fully realized in practice. Every device from the latest generation, as well as some devices several years old, were able to view the content. From this one can conclude that it is only a matter of time before this technology will represent a truly universal method for rapidly prototyping and distributing web-based VR and AR experiences for mobile devices.

4.3 Workshop: J!NS Meme

Shoya Ishimaru (Osaka Prefecture University)

Yuji Uema (JIN Co. Lt.)

Koichi Kise (Osaka Prefecture University)

License © Creative Commons BY 3.0 Unported license
© Shoya Ishimaru, Yuji Uema, Koichi Kise

Introduction of J!NS MEME

J!NS MEME is developed and released November 2015 by Japanese eyewear company, JIN CO.,LTD. Two significant differences between J!NS MEME and other traditional smart eyewear are long battery life and physical appearance. They last for around one day and look very close to normal eye wear.

They can stream sensor data to a second device (e.g. smart phone or computer) using Bluetooth LE. Sensor data includes vertical and horizontal EOG channels and accelerometer/gyroscope data. The runtime of the device is 12 hours enabling long term recording and, more important, long term real-time streaming of eye and head movement.

Applications of J!NS MEME to Reading-Life Log

I introduced our research of reading-life log by using J!NS MEME. The device J!NS MEME is not an eye tracker so that it is not possible for us to get eye gaze information; what we can get is the information about eye movement and eye blinks. Based on this information, we have implemented a wordometer, which counts the number of read words, and reading detector, which detects the period of reading in our daily activities.

These are examples for inspiring participants to think how to use J!NS MEME for their new applications.

Software for J!NS MEME

We introduced software development with J!NS MEME. According to primary asking, we had prepared three types of devices and their sensor logging softwares on several OS platforms (iOS/Windows/Mac). All materials and tutorial are available on <http://shoya.io/dagstuhl/>

Experiences of using J!NS MEME

We had 4 teams that tackled the following research topics with J!NS MEME.

- Head and Eye Gestures with MEME: The team tried several eye gestures and head gestures for user interaction and could present some initial ideas for unobtrusive gestures.
- Combination with Pupil: The team worked on integrating a Pupil eye-tracker with MEME and could show a working prototype.
- Haptic Feedback on EyeWear: The group presented a quick haptic feedback system attached to MEME for unobtrusive feedback.
- Eye movement influence on cognitive states: The group presented ideas on how to use the simple sensors of MEME to detect cognitive states (e.g. fatigue, disorientation, Alzheimer's and drug influence).

4.4 Tutorial: Head-Worn Displays

Kiyoshi Kiyokawa (Osaka University), Thad Starner (Georgia Institute of Technology), Ozan Cakmakci (Google)

License © Creative Commons BY 3.0 Unported license
© Kiyoshi Kiyokawa, Thad Starner, and Ozan Cakmakci

The head-worn display tutorial was organized into three sections: an ideal perspective, a user perspective, and an optical design perspective.

Ideally, head worn displays (HWDs) are expected to produce any visual experience we imagine, however, compared to this ultimate goal, current optical see-through (OST) HWDs are still far from perfect. In the first part of the HWD tutorial, we introduced introduced a number of issues that need to be tackled to make a 'perfect' OST-HWD. They include size, weight, the field of view, resolution, accommodation, occlusion, color purity, and latency. These issues are discussed by taking a number of research prototypes as example solutions. Then another fundamental problem of OST-HWDs is introduced that is, the user's exact view is not accessible as image overlay happens on the user's retina. A new concept of corneal feedback augmented reality (AR) is then introduced as a promising approach to realize a closed feedback loop for better registration, color correction, contrast adjustment, accurate eye tracking and scene understanding, by continuously analysing the reflections in the eye. In the end of the talk, by taking visualization of motion prediction as an example, it is emphasized that sensing is the key to success not only for better visual experience but also for more advanced applications.

Consumer purchase and use of eyewear is driven more by fashion than just about any



■ **Figure 3** The first tutorial provided an introduction to head-mounted displays (Kiyoshi Kiyokawa, Osaka University), insights into the development of Google Glass (Thad Starner, Georgia Institute of Technology), as well as a 101 on the design of optical systems (Ozan Cakmakci, Google).

other feature. Creating head-worn displays (HWDs) for everyday use must put fashion first. Due to lack of familiarity with the use of the devices, HWD consumers often desire features that are impractical or unnecessary. Many desire wide field-of-view (FOV) displays. However, the human visual system only sees in high resolution in a small 2 degree area called the fovea. Small FOV consumer displays (i.e., smart phones) have been highly successful for reading books, email, messaging, mobile purchasing, etc. A typical smart phone has a 8.8x15.6 degree FOV. Small FOV displays are much easier to manufacture with fashionable eyewear and will likely be the first successful (> 2m) consumer HWD product. Head weight should be kept under 75g for everyday use.

From the point of view of optical design, most researchers in the field of eyewear computing rely on existing commercially available head-worn displays to do their research and development. In the last part of the tutorial, we illustrated the optical design process by taking a deep sea diving telemetry display as a case study. The optical design process of a visual instrument involves understanding the design parameters, for example, image quality, field of view, eyebox, wavelength band, and eyerelief. We started the optical design monochromatically with a singlet lens to illustrate the limiting monochromatic aberrations. Next, we evaluated the optical performance with a photopic spectrum, and discussed approaches color correction. The use of transverse ray aberration plots was emphasized as a debugging tool. The main point in the last part of the tutorial was to remind researchers that it is possible to design custom optics at reasonable cost in contrast to relying solely on commercially available displays.

The participants got a chance to experience Google Glass (mirror based magnifier), Optinvent ORA (flat lightguide with collimation optics), Epson BT-200 (flat lightguide with freeform outcoupler), Triplett Visualizeyezer (concept of a beamsplitter), and Rift 1 (single lens based magnifier) during the workshop as part of the tutorial. We used the various hardware to illustrate eyebox, field of view, chromatic aberrations, distortion, and brightness.

4.5 Tutorial: Egocentric Vision

James M. Rehg (Georgia Institute of Technology)

Kristen Grauman (University of Texas)

License © Creative Commons BY 3.0 Unported license
© James M. Rehg, Kristen Grauman

The video produced by a head-worn camera possesses a key property – the motion of the camera through the scene is fundamentally guided by the intentions and goals of the camera-wearer. As a consequence, first person video implicitly contains potent cues for scene understanding that can be leveraged for video analysis. This tutorial consisted of two parts. The first part discussed specific egocentric cues, such as head motion and the locations of the hands, that can be extracted from egocentric video and utilized for predicting the attention and activities of the camera-wearer. The second part discussed representation learning and the inference of attention to drive the summarization of egocentric video.

Egocentric video provides a unique opportunity to capture and analyze the visual experiences of an individual as they go about their daily routines. The first part of the tutorial demonstrated that egocentric cues such as head motion and hand location can be used to predict the attention of the first person as they perform hand-eye coordination tasks such as activities of daily living. Using a dataset of cooking activities (GTEA+), we demonstrated

that accurate predictions of the the user's gaze could be obtained through temporal fusion of egocentric cues. These same cues can also be used as features in performing activity recognition. We showed that egocentric features can be used to complement and improve classical features such as Improved Dense Trajectories (IDT) for an activity recognition task. We presented a method for stabilizing egocentric video to remove the effect of head motion on IDT, leading to increased accuracy in using that feature. We also demonstrated the added benefit of using head motion and hand positions as features, beyond the baseline provided by the standard IDT approach. These findings demonstrate the unique properties of egocentric video and the existence of specific egocentric features which provide an improvement beyond classical video representations.

The fact that the camera wearer is an active participant in the visual observations received has important implications for both (1) representation learning from egocentric video and (2) understanding cues for attention. The second part of the tutorial overviewed work on both of these fronts. In terms of representation learning, we discussed how egomotion that accompanies unlabeled video can be viewed as an implicit and free source of weak supervision. The goal is to learn the connection between how the agent moves and how its visual observations change. Whereas modern visual recognition systems are focused on learning category models from "disembodied" Web photos, often for object class labeling, the next generation of visual recognition algorithms may benefit from an embodied learning paradigm. We presented one such attempt, based on deep learning for representation learning that regularizes the classification task with the requirement that learned features be equivariant. In terms of attention cues, we surveyed ideas for exploiting the information about a camera wearer's gaze, pose, and attention to (a) compute compact summaries from long egocentric videos, (b) estimate when a video frame passively captured on the camera looks like an intentionally framed shot, and (c) detecting moments of heightened engagement where the camera wearer is examining his/her environment to gather more information about something. In all cases, learning based methods are being developed to capture the special informative structure from egocentric video. Key future directions for these lines of work include dealing with multi-modal sensing, transforming the learned view predictive representations to tackle active vision problems, exploring novel forms of visualization of a summary, and identifying scalable methods for quantitative evaluation of video summaries.

5 Challenges

To take advantage of the opportunities for sensing and interaction offered by near eye (or more general head mounted) systems significant computational resources are needed, especially for processing o video signals, advanced information fusion and long term learning. While in some cases processing can be outsourced to remote devices (or the cloud), in others local processing on the device is desirable. Such processing must take into account stringent power, thermal and space constraints specific to the near eye location. Thus, appropriate digital and mixed signals architectures specifically combining special purpose circuits with appropriate, reconfigurable components and low power general purpose processing capabilities need to be developed.

In most worn computing systems, four major inter-related technical challenges must be addressed: power & heat, networking (on & off-body), interface, and privacy. For eyewear computing, interface can be further divided into several sub challenges: head weight, fashion, and attention. Advances in electronics will continue to drive improvements in power &

heat, networking, and head weight. However, there is much work yet to be done in creating low-attention interfaces that are appropriate for use in everyday environments. Eyewear computing interfaces should be designed to be secondary to a user's primary task and distract minimally from it. Similarly, depending on the social situation of use, the appearance of the electronics (fashion) and the use of the electronics (interaction) should not distract bystanders, colleagues, or conversational partners from their primary tasks.

This seminar has suggested possible directions toward improving interactions with eyewear computing. Can context-awareness, leveraging the access to the user's senses afforded by eyewear computing, help deliver content to the user unobtrusively and at the right moment? Can eye, head, and or body movement be used to provide both explicit and implicit input to a eyewear computing agent that can assist the user in day-to-day tasks? These questions hint at another, more ambitious agenda: can we use eyewear computing's unique first person perspective to create an intelligent assistant that learns to live and interact in the human world? Given the recent interest and improvements in techniques that leverage "big data" to gain new insights into human-level problems like speech recognition and parsing images, initiatives leveraging eyewear computing might be well poised to gain levels of understanding of what it means to be human that systems restricted to data from the web can not attain.

6 Breakout Sessions

6.1 Group 1: Multimodal EyeWear Computing

Masashi Nakatani (University of Tokyo, JP)

License  Creative Commons BY 3.0 Unported license
© Masashi Nakatani

Joint work of Gábor Sörös, Moritz Kassner, Ozan Cakmakci, Päivi Majaranta, Sabrina Hoppe, Will Patera, Yuji Uema

Recent advancement of eyewear device and computing allows us to use the equipment in multiple contexts. Sensory substitution and sensuarization is one example, in which the system can make captured infomation into more perceptually detectable [1]. The use of eyewear computing should not be limited to visual information. Recent advancement of multimodal study in perceptual psychology and engineering implentation allow to integrate eyewear into multimodal interface. One representative applications of multimodal integrations is sensory substitution, in which people who has sensory impairments (mostoly vision and auditory sensation).

In addition to multimodal integration as a device, potential collaborative study with cognitive psychology is now getting to be expected. As old syaing states, "Look into my eyes and hear what I'm not saying, for my eyes speak louder than my voice ever will." Based on this intuition, a bridge between eye-wear and cognition attracts attentions from old ages.

Taking into these aspects into account, this section describes potential collaborative research fields in eyewear computing, and discuss possible research directions in the future.

Auditory

Gaze pointing has been proposed to use for producing sound and music composition [3] [4].

In virtual reality application, 3D spatial audio information is important for producing the feeling of immersion and presence of presented graphics. Based on this insight, 3D audio information has been already integrated in the application of Google Cardboard and Oculus Rift, and Apple iOS is plausible platform for including gaze based audio feedback.

Haptic

Touch modality has been considered to be beneficial for hearing impaired people to give feedback through the skin. Haptic feedback is provided based on visual [2] and audio information[5] well as audio feedback[6].

The advantage of using haptic feedback is that this information is relatively private comparing with visual and audio feedback.

Haptic Gaze Interaction is one of promising research direction of eyewear computing with haptic feedback[7]. A series of this project has been initiating basic research on combining of eye-pointing and haptic interaction. There are various open opportunities for natural and efficient human-computer interaction.

Olfactory

Olfactory sensation has a good connection with memory. One person may retrieve his/her memory by smelling certain odor, and neuroscientists are pursuing the neural basis of this phenomenon.

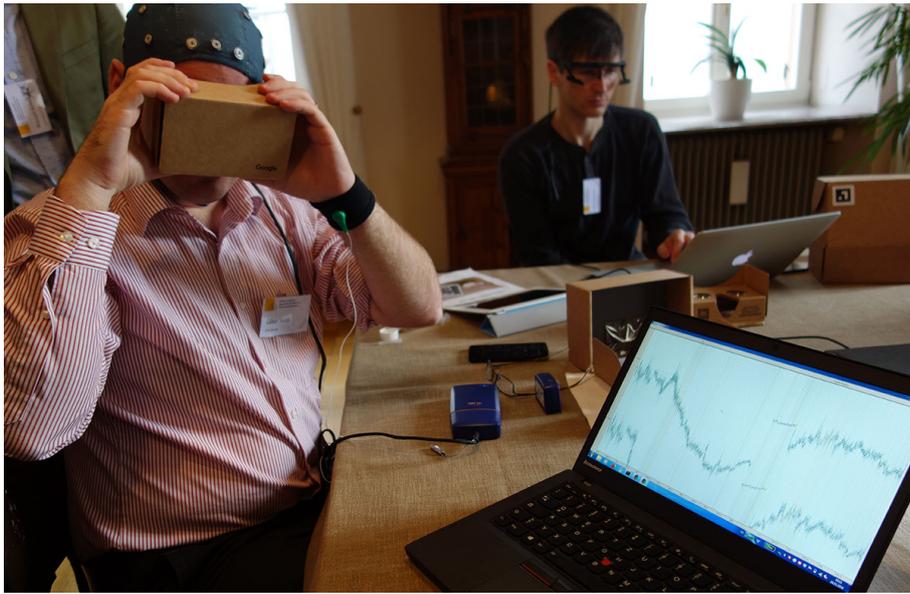
Consideration the combination with eyewear, it is also interesting hypothesis to test whether vision-odor conversion may increase the ability of memorizing certain tasks in the office or in the class. Zoladz and Raudenbush reported that odorant may enhance cognitive processing [8]. This suggests a potential that visual-scene-odor conversion may also beneficial for office workers.

Brain Activity

Electroencephalography (EEG in short) is now popular method in cognitive psychology to access brain activity during daily behaviors. Comparing with other brain imaging techniques (fMRI, MEG, NIRS), participants can relatively freely move during the task. In addition to that wearable EEG is commercially available (cite Emotive) with reasonable price. Exploiting EEG as an even more direct pathway to the brain to create adaptive or assistive user interfaces.

Examining brain activity in virtual reality environment may attract attentions from industrial and academic interests. Not only using conventional VR environment using multiple screens and projectors, but also recent proposal of Google cardboard helps to test the relationship between eye-movements and visual feedback in immersive environment.

One potential disadvantage is that to obtained information from EEG and traditional eye wear sensors may be redundant, so that experimenters need to consider carefully before conducting actual experiment. Eyewear researchers will benefit from collaborative study with cognitive scientists.



■ **Figure 4** Hands-on trial of an EEG measurement while experiencing VR through the Google Cardboard. Seminar participants combined their expertise on site and conducted preliminary experiments.

Facial Expressions

Electromyography enables using facial expressions such as smiling or frowning to select things that are currently under visual focus. Wearing such a head-mounted "face interface" enables interaction based on voluntary gaze direction and muscle activation. For more information, see e.g. [9].

Memory

Eyewear computing has also impacts on the therapy of trauma. Psychotherapy study using Eye Movement Desensitization (EMD) reported that this technique is beneficial for decreasing traumatic experience (reference). EMD is the procedure in which patients access to their traumatic memories in the context of a safe environment, the hypothesis is that information processing is enhanced, with new associations forged between the traumatic memory and more adaptive memories or information. These new associations result in complete information processing, new learning, elimination of emotional distress, and the development of cognitive insights about the memories. Although eyes move for any memory, EMD is still effective for some patients. If eyewear computing research can collaborate with medical and cognitive scientists studying memory and its psychotherapy, both side would get benefits in the advancement of device and human emotions.

Conclusion

There are multiple potentials of eyewear computing by combining with other modality or for practical use in medical science. Close relationship between eyewear computing community

and other research fields would be critical for the prosperity of both side of the research. We ended up the discussion by encouraging researchers to have multiple interests in basic and applied study of eyewear computing.

References

- 1 Péter Galambos, András Róka, Gábor Sörös, Péter Korondi. *Visual feedback techniques for telemanipulation and system status sensualization*, Proc. of IEEE 8th International Symposium on Applied Machine Intelligence and Informatics (SAMI), pp. 145–151, 2010.
- 2 Matthias Berning, Florian Braun, Till Riedel, and Michael Beigl. *ProximityHat: A Head-worn System for Subtle Sensory Augmentation with Tactile Stimulation*. ISWC'15 Proceedings of the 2015 ACM International Symposium on Wearable Computers, pp. 31–38, 2015.
- 3 Hornof, A., Sato, L. (2004, June). EyeMusic: making music with the eyes. In Proceedings of the 2004 conference on New interfaces for musical expression (pp. 185–188). National University of Singapore.
- 4 Hornof, A. J., and Vessey, K. E. (2011, September). The Sound of One Eye Clapping Tapping an Accurate Rhythm With Eye Movements. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (55,1pp. 1225–1229). SAGE Publications.
- 5 Novich S, Eagleman D. *Using space and time to encode vibrotactile information: toward an estimate of the skin's achievable throughput*. Experimental Brain Research. 233(10), pp. 2777-2788, 2015.
- 6 Peter B. L. Meijer. *An Experimental System for Auditory Image Representations*. IEEE Transactions on Biomedical Engineering, vol. 39, no. 2, pp. 112–121, 1992.
- 7 Jari Kangas, Deepak Akkil, Jussi Rantala, Poika Isokoski, Päivi Majaranta, and Roope Raisamo. *Gaze gestures and haptic feedback in mobile devices*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'14). ACM, New York, NY, USA, pp. 435–438, 2014.
- 8 Phillip R. Zoladz, Bryan Raudenbush, *Cognitive Enhancement Through Stimulation of the Chemical Senses*, North American Journal of Psychology, Vol. 7 Issue 1, pp. 125–140, 2005.
- 9 Tuisku, O., Surakka, V., Vanhala, T., Rantanen, V., and Lekkala, J. (2012). Wireless Face Interface: Using voluntary gaze direction and facial muscle activations for human–computer interaction. *Interacting with Computers*, 24(1), 1–9.

6.2 Group 2: Egocentric Vision

Rita Cucchiara (University of Modena)

Kristen Grauman (University of Texas)

James M. Rehg (Georgia Institute of Technology)

Walterio W. Mayol-Cuevas (University of Bristol)

License © Creative Commons BY 3.0 Unported license

© Rita Cucchiara, Kristen Grauman, James M. Rehg, Walterio W. Mayol-Cuevas

Joint work of Andreas Bulling, Rita Cucchiara, Kristen Grauman, Kiyoshi Kiyokawa, James M. Rehg, Linda B. Smith, Julian Steil, Yusuke Sugano, Walterio W. Mayol-Cuevas, and Masahiko Inami

The ego-centric view is one created by the wearer's own actions and momentary goals. The visual properties of these ego-centric scenes and videos have their own properties (hands, center bias for attended objects, information reduction). They are also inherently connected to the wearer's movements – eye gaze, head movements, hand movements, whole body movements. Egocentric vision is an emerging field of computer vision, specially devoted to defining models, algorithms and techniques dealing with egocentric video, generally captured

by eyewear devices. It exploits geometry, pattern recognition and machine learning paradigms to understand ego-centric scenes, recognizing objects, actions and the interaction between the wearer, the persons and the surrounding environment. It addresses problems of personalized video summarization [19], relevance determination for detecting what is important and how things are used for user guidance [11], and the prediction of attention [20] and activities.

In this breakout session, we discussed (a) the links between movement and visual learning in animals, humans, and machines, (b) its relation to supervised and unsupervised learning, (c) the challenges of egocentric vision in the wild considerable body movements and lighting changes and (d) and relation other frames of reference for capturing the visual information.

Movement and Visual Learning

Visual experience in the context of planned and goal-directed movements are known to enhance learning and change patterns of brain connectivity in animals and humans [1, 2]. There are possible synergies between computational and theoretical approaches cross biological and machine learning. Particularly relevant may be leveraging the structure in multimodal information in goal-directed action. In biological learning, computational models have been commonly based on prediction [3] and re-entrance (see also [5] for recent review).

Different actions and different tasks may prevent different solutions for computer vision systems as they yield different patterns of egocentric vision and different patterns of eye, head, and body movements. The role of hands, object manipulations and different patterns of movements in goal directed tasks of different kinds also merit consideration. When the larger multimodal properties of human are considered in natural real-world tasks, the question arises as to whether foveation and saccades are of central importance or whether these might be considered attributes of human vision specific to specific tasks such as reading but which do not deeply inform human visual behavior in more active contexts nor perhaps computer vision systems ([6, 7]).

The current status of computational vision doesn't speak to how children learn to recognize objects, which is robust at a very early age (see, [8]). Current computational vision with its emphasis on static images may not be very useful for autonomous vehicles or for building learning robots that improve through their own experience. Egocentric vision, by leveraging the properties of goal-directed purposeful action, provides potentially transformative domain for the classic questions of machine vision including object segmentation and object recognition.

Issues in Weakly/Unsupervised Learning

A current trend in computer vision is exploiting machine learning, and in particular Deep Learning, to support the tasks of object and activity recognition. This trend holds for egocentric vision as well. Nevertheless, although recent improvements in recognition performance may seem impressive, these gains have been made in closed world datasets where the space of object labels is fixed. In contrast, when a person is moving in the real world, they continuously encounter novel objects and unique situations, and classifiers trained with closed world datasets cannot be easily adapted. This raises the question of egocentric learning from weak supervision. Large amounts of video can be easily acquired, how can we utilize this data if labels are available only sparsely? For example, can we acquire models of objects

based on observations of how they are used in performing actions? Semi-supervised learning can be used for recognizing action patterns or objects that have very few annotated examples: an example is hand gesture recognition, which is frequently characterized by personal, and often unique, gestures. In this case a robust hand segmentation process can be coupled with a semi-supervised process of recognizing actions from very few examples. [19].

Egocentric Vision and Neuroscience

The neuroscience aims at understanding how information sensed by eyes become vision, how thoughts become memory, how visual behavior comes from biology. This is very important to understand the human vision and how can this be exploited in computer vision too. According with Kandel (Kandel 2012) the cortical view suggests that understanding what is important from an observer is a mix of recognition and tracking. That can be viewed in the manner artists approached the problem of objects' shape, movement and 3D representation in paintings, and is strongly connected with their exploitation of the cortical vision in the "way of what" and the "way of where". Therefore combining semi-supervised recognition, object tracking, action inferring from a first person view is still the big issue.

Egocentric Vision and Other Sensors

Just as humans perceive via multiple modalities, so egocentric perception could benefit from leveraging additional sensing modalities. It is an open issue how to use other sensors and augmentation techniques to provide additional cues to inform image understanding. From power standpoint, vision is a very expensive sensor, and it could be interesting to exploit less expensive sensors to cue vision and reduce the overall power budget. One challenge we face is that while on-body signals from a variety of wearable sensors are likely to be inter-related, the nature of the relationships can change with the task and over time. For example, signals are related as a result of basic physiological processes [21] as well as via the task the subject is performing. In the case of object manipulation tasks, two sources of difficulty are the challenge of reliably estimating motion and the lack of effective object-level representations.

Relationship Between First Person and Third Person Vision

What is the relationship between first person vision and other types of camera positions and movements? We could define three kinds of egocentric vision: Eyeball imaging, Head-mounted imaging, Omnidirectional imaging. This first, still far to be commonly used, will extract the precise information of where the people are fixating the gaze. What are the natural statistics of the visual world via egocentric vision? For example, since cameras move smoothly through the scene there is continuity of space and time which produces a power law distribution of types and tokens in video sequences. So temporal continuity potentially provides a very strong prior that should help to solve the problem. We may need to revisit all of the standard problems in vision from the standpoint of egocentric vision. Segmentation, perception of form, recognition, etc. But there are other applications in eye wear computing such as human augmentation and of accessibility applications such as visually-impaired.

Attention Models

Egocentric vision may require a rethink about our attention models so that what is sensed, recorded or processed is relevant to a task or tasks. Such an *active sensing* approach finds similarities with what can be inferred from the way people make high-level decisions on where to focus visual and processing resources. Eye gaze patterns being a clear example of a gated visual interaction with the world. Current attention models used in the egocentric vision literature are based on eye fixations [10, 11], image features [12] or head motion [13]. However there is evidence in the human vision literature, of tasks where fixations may not be used [14, 6] and in egocentric systems where a broader peripheral sensing may be sufficient for activity detection or navigation tasks [16, 15]. There is thus a need to consider what attention models our eyewear systems will benefit from with the possibility that these are task-driven.

References

- 1 Held, R., Hein, A. (1963). *Movement-produced stimulation in the development of visually guided behavior*. Journal of comparative and physiological psychology, 56(5), 872.
- 2 James, Karin Harman. *Sensori-motor experience leads to changes in visual processing in the developing brain*. Developmental science 13, no. 2 (2010): 279-288.
- 3 Lake, B. M., Salakhutdinov, R., and Tenenbaum, J. B. (2015). *Human-level concept learning through probabilistic program induction*. Science, 350(6266), 1332-1338.
- 4 Sporns, O., Gally, J. A., Reeke, G. N., Edelman, G. M. (1989). *Reentrant signaling among simulated neuronal groups leads to coherency in their oscillatory activity*. Proceedings of the National Academy of Sciences, 86(18), 7265-7269.
- 5 Byrge, L., Sporns, O., Smith, L. B. (2014). *Developmental process emerges from extended brain-body-behavior networks*. Trends in cognitive sciences, 18(8), 395-403.
- 6 Foulsham, T., Walker, E., Kingstone, A. (2011). *The where, what and when of gaze allocation in the lab and the natural environment*. Vision research, 51(17), 1920-1931.
- 7 Kingstone, Alan, Daniel Smilek, Jelena Ristic, Chris Kelland Friesen, and John D. Eastwood. *Attention, researchers! It is time to take a look at the real world*. Current Directions in Psychological Science 12, no. 5 (2003): 176-180.
- 8 Bergelson, E., Swingle, D. (2012). *At 6–9 months, human infants know the meanings of many common nouns*. Proceedings of the National Academy of Sciences, 109(9), 3253-3258.
- 9 Eric Kandel *The Age of Insight*, 2012
- 10 Alireza Fathi, Yin Li, James M. Rehg *Learning to Recognize Daily Actions using Gaze*. ECCV, 2012.
- 11 D Damen, T Leelasawassuk, O Haines, A Calway, W Mayol-Cuevas. *You-Do, I-Learn: Discovering Task Relevant Objects and their Modes of Interaction from Multi-User Egocentric Video*. British Machine Vision Conference (BMVC), Nottingham, UK. 2014.
- 12 B. Xiong and K. Grauman. *Detecting Snap Points in Egocentric Video with a Web Photo Prior*. In Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, Sept 2014.
- 13 T Leelasawassuk, D Damen, W Mayol-Cuevas. *Estimating Visual Attention from a Head Mounted IMU*. International Symposium on Wearable Computers (ISWC). 2015.
- 14 Franchak, J. M., Adolph, K. E. (2010). *Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults*. Vision research, 50(24), 2766-2774.
- 15 M Milford, *Visual Route Recognition with a Handful of Bits*. Robotics: Science and Systems, 2012.
- 16 B Clarkson, K Mase, A Pentland, *Recognising User's context from wearable sensors: baseline system*, MIT VisMod Tech Report No 519, March, 2000.

- 17 Baraldi, L., Paci, F., Serra, G., Benini, L. Cucchiara, R. *Gesture Recognition in Ego-Centric Videos using Dense Trajectories and Hand Segmentation*, Proc. of 10th IEEE Embedded Vision Workshop at CVPR14, 2014
- 18 Lee, K.J., Grauman, K., *Predicting important objects for egocentric video summarization*, International Journal of Computer Vision, 1-18n 2015.
- 19 Cucchiara, R., Varini, P., Serra, G. *Personalized Egocentric Video Summarization for Cultural Experience* Proc. of the ACM Int. Conf. on Multimedia Retrieval, 2015
- 20 Li, Y., Fathi, A., and Rehg, J.M. *Learning to Predict Gaze in Egocentric Video*, In Proc. IEEE Intl. Conf. on Computer Vision (ICCV), Sydney, Australia, Dec 2013.
- 21 Hernandez, J., Li, Y., Rehg, J.M., and Picard, R.W. *BioGlass: Physiological Parameter Estimation Using a Head-mounted Wearable Device*, In Proc. Intl. Conf. on Wireless Mobile Communication and Healthcare (MobiHealth), 2014

6.3 Group 3: Security and Privacy

René Mayrhofer (Johannes Kepler University Linz)

License  Creative Commons BY 3.0 Unported license
© René Mayrhofer

As a relatively new scientific area, Eyewear Computing has not yet attracted much attention to security and privacy issues. Nonetheless, even current products and prototypes show some of the challenges that will need to be addressed before wide adoption is possible (or should be aimed for):

Device-to-user authentication counters the threat of devices being physically replaced by similar-looking but modified devices [1]. If users cannot be sure that the device they are about to put on is really their own, then they may divulge information to third parties or be presented with false information. Device-to-user authentication is highly related to device security – the former addresses the threat physical tampering, the latter of software tampering.

User-to-device authentication prevents the opposite threat of other (potentially malicious) people using a device and abusing the credentials associated with it or the data directly stored locally on the device. This can be (mostly) considered a solved problem for smart phones (with fingerprint readers as one reasonable compromise between security and usability, PIN/passwords, unlock patterns, and other methods already being used in practice [2]), but is still an open problem for eyewear devices.

Device-to-device authentication is required as soon as devices communicate wireless with other devices, as wireless radio links are not open to human senses. Proving to human users which other devices their own are communicating with (sometimes referred to as the human-in-the-loop property [3]) is typically achieved with *pairing* of devices, e.g. for Bluetooth connections. Pairing is an appropriate approach to long-term device links (such as eyewear devices to smart phones), but scales poorly for short-term interactions (such as using printers, projectors, etc. in the infrastructure).

Device security itself is an open challenge for many device categories (with smart phones being the most well-known at this time), and any eyewear devices will face the same issues due to complexity of the software stack they run [4]. However, assuming a similar set of sensors, communication interfaces, and platform, the potential implications of eyewear devices being subverted remotely are much more severe: using subliminal signals

on displays that are always visible, the risk for users being susceptible to manipulation is significantly higher.

Privacy of the wearer is an issue for all wearable devices connecting to and transmitting data into cloud services. The solution space is currently being explored for smart phones and will probably apply to eyewear display when considering that some additional sensors may be continuously recording.

Privacy of others is more difficult to guarantee, given the typical integration of cameras and microphones. Depending on the relevant legal system, it may not even be permissible to use standard eyewear devices with currently missing privacy safeguards in some settings. New technical approaches such as privacy-preserving filters close to the respective sensor may be required.

Although there are many unsolved issues, there is also the potential for eyewear devices to make some security issues easier to address. By relying on the fact that these devices are typically very close to the body during the day, continuous cross-device authentication may be used to more easily and more quickly unlock other devices of the same user [5]. By utilizing the camera, microphone, or other built-in sensors, eyewear devices could become proxies for spontaneous authentication to infrastructure services. By giving feedback on other devices and services to their users, they could improve awareness of potential security and privacy issues and provide improved transparency. Summarizing, security and privacy challenges are still largely unexplored in the specific context of eyewear computing, and novel solutions may be significantly different from current approaches to smart phone security.

References

- 1 Rainhard D. Findling, Rene Mayrhofer. *Towards Device-to-User Authentication: Protecting Against Phishing Hardware by Ensuring Mobile Device Authenticity using Vibration Patterns*. 14th International Conference on Mobile and Ubiquitous Multimedia (MUM'15), 2015
- 2 Daniel Hintze, Rainhard D. Findling, Muhammad Muaaz, Sebastian Scholz, Rene Mayrhofer. *Diversity in Locked and Unlocked Mobile Device Usage*. Adjunct Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2014). 379-384, 2014
- 3 Rene Mayrhofer. *Ubiquitous Computing Security: Authenticating Spontaneous Interactions*. Vienna University, 2008
- 4 Rene Mayrhofer. *An Architecture for Secure Mobile Devices*. Security and Communication Networks, 2014
- 5 Daniel Hintze, Rainhard D. Findling, Muhammad Muaaz, Eckhard Koch, Rene Mayrhofer. *Cormorant: towards continuous risk-aware multi-modal cross-device authentication*. Adjunct Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing and Symposium on Wearable Computers. 169-172, 2015

6.4 Group 4: Eyewear Computing for Skill Augmentation and Task Guidance

Thies Pfeiffer (Bielefeld University)

Steven K. Feiner (Columbia University)

Walterio W. Mayol-Cuevas (University of Bristol)

License © Creative Commons BY 3.0 Unported license

© Thies Pfeiffer, Steven K. Feiner, Walterio W. Mayol-Cuevas

Joint work of Steven K. Feiner, Scott Greenwald, Shoya Ishimaru, Koichi Kise, Kai Kunze, Walterio W. Mayol-Cuevas, René Mayrhofer, Masashi Nakatani, Thies Pfeiffer, Philipp M. Scholl, Gábor Sörös

Definitions/Distinctions

Assistance provided by eyewear computing can be classified as either skill extension or skill training. In *skill extension*, smart glasses provide assistance that goes beyond human capabilities, essentially offering super-human abilities. Examples include see-through views, such as now featured for some cars, and automatic language translation. On the other hand, in *skill training* or *skill support*, users are helped to be better at their normal skills (e.g., in terms of accuracy, speed, or quality).

A second important distinction is whether assistance is provided before, during, or after task performance. Of these possibilities, providing assistance during task performance places real-time constraints on the process. Assistance provided in advance can prepare and train the user, while assistance provided after the task, can help the user evaluate how well they did.

Assistance can also be categorized by the degree to which it is responsive to the user's actions. For example, eyewear might passively overlay relevant information on the user's environment, such as the unchanging 3×3 grid that many camera user interfaces can display to encourage the user to use the “rule of thirds” in composing an image. In contrast, eyewear could actively guide the user, in this case by steering the user to achieve an image composition based on this heuristic [14].

Application Areas for Eyewear Computing Related to Skills

- **Learning / Teaching.** Many countries have identified digital learning and online learning as an important component of the current and future educational system. Eyewear computing could support adaptive, personalized learning experiences in contexts beyond the desktop (i.e., beyond the classroom or the office). Learning material could be restructured and paced according to the personal needs, learning progress, and competences of the learner, even in group learning situations. One important motivational aspect could be gamification. Eyewear could not only educate the learner, but also provide valuable feedback to the teacher. There is a chance that personalized learning materials could also support better social integration by balancing expertise effects in learning situations by presenting different levels of assistance to individual learners. For example, if the teacher raises a question, advanced learners could be presented with four answers that are difficult to decide upon, whereas others could have easier choices. While the number of the correct answer could be the same for each student, the decision task could thus be adapted to their individual competences. The key point is that the learners would not be directly aware of that and thus social implications could be reduced. This might motivate weaker pupils to participate better during class.
- **Music / Physical Skill Training.** Training physical skills is a different challenge. How, for example, could eyewear improve the skills of a pianist? The answer is not straightforward,

but we discussed several possible approaches: visualizing correct hand postures, visualizing target keys (how useful this would be is questionable, as early hardware versions have been employed since the mid 20th century – e.g., Thomas Organ Color-Glo, https://en.wikipedia.org/wiki/Thomas_Organ_Company), or having an augmented score that adapts to the current level of expertise of the pianist (e.g., automatically updating so that problematic passages are repeated more often or are varied according to a training program).

- **Privacy.** There is a common conception that smart glasses can compromise privacy; however, there are some cases in which they might instead make it possible to maintain privacy. For example, in a future scenario in which smart machines adapt to a user's level of expertise, their reactions toward a specific user could reveal the user's competences to all onlookers (e.g., has difficulty using a coffee maker, confuses left and right, or forgets account information). Similarly, personalized advertising that is viewable by others could be embarrassing. Smart glasses could instead keep this information private. This could be especially attractive to professionals who would like to look up information on the fly while interacting with others. For example, a physician might not want to demonstrate their lack of knowledge about a recent study when talking with a patient who has just heard a news story about that study. Smart glasses could make it possible for the physician to find relevant information during the conversation; with the right user interface that could be done surreptitiously.
- **Communication Training / Therapy.** Communication training has been successfully used for children with autism using VR [11]. VR training applications for public speaking have long been developed and evaluated [10, 2] and are now available for consumer devices, such as Google Cardboard and Samsung Gear VR [13].
- **Task Motivation.** Gamification and other concepts are not directly affecting skill training, but help to maintain or create a necessary level of motivation. One important challenge is to design a system that helps the user to maintain a high level of motivation in the long run. In summary, that depends on the feedback and the intrinsic motivation of the users. This would add a strong A.I. component to eyewear computing.
- **Digital Memory.** A remembrance aid could use face and name recognition, and remind the user of their previous interactions with someone whom they may have forgotten.
- **Replacing/Augmenting/Assisting Senses.** Possibilities include improved hearing with noise cancellation and diminished reality [8] to suppress parts of the environment that the user wishes to avoid (e.g., deleting advertisements). But, note that expurgating things that the user wishes to avoid may keep important problems from being addressed.
- **Behavior Change.** Some aspects of interacting with a machine might increase the openness of persons in comparison with how they talk with other people [4].
- **Surveillance.** Eyewear can make surveillance possible without the need to consciously attend to capturing video. For example in a disaster scenario, such as an earthquake or nuclear plant accident, users need to concentrate on their own safety and that of others.
- **User Behavior Studies.** In particular in working contexts. For example in ISS (even though most of behaviors are already captured in current system)

Key “Selling Points” of Eyewear Computing for Augmentation/Guidance

This is a non-exhaustive list of selling points for eyewear computing that were collected from comments during the discussion.

- **Embedded display.** AR allows us to present information directly embedded in the context of real-world action, eliminating the need for the user to look back and forth between the real world action and a separate information source [5]. This can be accomplished by no other technology.
- **Subtle cueing.** Sometimes we do not need a fully featured AR, but subtle cueing, such as stimulating a rhythm while doing physical exercises or displaying subtle, possibly imperceptible, visual prompts to direct a user's attention [12, 6].
- **Shared attention.** Eyewear shares the wearer's attention and is thus very close by to the current interaction context.
- **Combined sensing and action/display.** Eyewear provides sensing and contextual presentation in one device.

Challenges

- Knowledge about multimodal communication/feedback channels between participants, in particular in a learning environment with teacher–pupil interactions, is essential for creating supportive eyewear applications for education.
- Context awareness could be a key feature of AR and smart glasses, yet to identify the state (emotional, cognitive, physiological) and the competences of the individual user (wearer) or interaction partner is a huge challenge. This needs to be tackled to be able to provide the right feedback to the user.
- Haptics could be important (e.g., for skill training), but it is an open question how to provide haptic feedback in a generalized way.
- Regarding learning, there appears to be more work on teacher–pupil interaction than on the effects of the peer group. Would this be helpful to consider in the future? We see many ways to add that to AR and VR simulations (This was followed by a discussion of Oculus Social Alpha, the peer-couchsurfing app).
- Authoring of applications for smart glasses, in particular for training, is a big issue that comes with its own problems.
- Measuring user interactions.
- Social acceptance: Will people want their use of mental/skill augmentation to be private? That is, will it be embarrassing to be seen using these augmentations? Compare to hearing aids, where only the most expensive ones are currently unobtrusive. Is it going to be scary that some people have skill augmentation? That is, will others feel threatened by losing out on a perceived (or very real) advantage? The core of this issue is the feeling of access to the technology. If someone feels they can have access to the augmentation and chooses not to use it, it is not a social issue anymore. For example, a photographer using a film camera has chosen not to use the functionality provided by a digital camera, and is unlikely to feel threatened or minimized by a photographer using a digital camera, as s/he could choose to use one if desired. (Indeed, the film photographer might feel superior to the digital photographer.) What if a company were to make and distribute for free to anyone interested a state-of-the-art wearable, in return for access to its data? Would this be any more privacy-compromising than providing state-of-the-art search facilities?
- We need a more sophisticated and nuanced understanding of the task so that there can be a more informed decision-making process about the type of guidance (3D, 2D, audio, tactile, etc.) that is most effective for the user and situation. This decision-making process is currently arbitrary at all levels: which hardware to use, the data-processing strategy, and the format of the augmentation. A better understanding of what is most suitable can lead to less expensive and more widely acceptable eyewear.

Questions

- What is the most effective tool for skill augmentation and task guidance?
- Do the same principles that can be thought of as relevant for training and supporting physical skills also hold for mental/cognitive skills?

Low-Hanging Fruit Applications

Which applications would seem to clearly benefit from being included with eyewear computing for augmenting skills based on existing applications/frameworks/tools? Some could be prototyped as undergrad \leq 4-week projects, if given access to existing tools, but could be much harder to do well.

- Checklists for to-do tasks
- Shopping lists
- Live translation (especially with a low-cognitive load format: <http://spritzinc.com/>, teleprompter)
- Context-sensitive calendar view
- Remembrance agent. Vannevar Bush [3] proposed a stereoscopic head-worn camera: “As the scientist of the future moves about the laboratory or the field, every time he looks at something worthy of the record, he trips the shutter and in it goes, without even an audible click.” This could be augmented with a display and search capabilities. Rhodes [9] describes an early text-based implementation.
- Augmenting Conversations Using Dual-Purpose Speech [7, 1]

References

- 1 Ackerman JM, Nocera CC, Bargh JA. *Incidental haptic sensations influence social judgments and decisions*, *Science*, 328(5986):1712–1715, 2010.
- 2 Page Anderson, Elana Zimand, Larry F. Hodges, and Barbara O. Rothbaum. Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and Anxiety*, 22(3), 156–158, 2005.
- 3 V Bush, *As We may Think*. The Atlantic Magazine. July 1945. <http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>
- 4 Jonathan Gratch, Gale M. Lucas, Aisha Aisha King, and Louis-Philippe Morency. It’s only a computer: the impact of human-agent interaction in clinical interviews. *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS’14)*. 85–92, 2014.
- 5 Steven Henderson and Steven Feiner, Augmented reality in the psychomotor phase of a procedural task, *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Basel, Switzerland, 191–200. 2011. <http://dx.doi.org/10.1109/ISMAR.2011.6092386>
- 6 Weiquan Lu, Henry B.-L. Duh, Steven Feiner, and Qi Zhao. Attributes of subtle cues for facilitating visual search in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 20(3), 404–412. March 2014. <http://dx.doi.org/10.1109/TVCG.2013.241>
- 7 Kent Lyons, Christopher Skeels, Thad Starner, Cornelis M. Snoeck, Benjamin A. Wong, and Daniel Ashbrook. *Augmenting conversations using dual-purpose speech*. Proceedings of the 17th annual ACM symposium on User Interface Software and Technology (UIST), pp. 237–246, 2004.
- 8 S. Mann and J. Fung. EyeTap devices for augmented, deliberately diminished, or otherwise altered visual perception of rigid planar patches of real-world scenes, *Presence*, 11(2):158–175, April 2002. <http://dx.doi.org/10.1162/1054746021470603>

- 9 Bradley J. Rhodes. The wearable remembrance agent: A system for augmented memory. *Personal Technologies*, 1(4), 218–224. December 1997.
- 10 M. Slater, D.P. Pertaub and A. Steed, Public speaking in virtual reality: facing an audience of avatars. *IEEE Computer Graphics and Applications*, 19(2), 6–9, March/April 1999. <http://dx.doi.org/10.1109/38.749116>
- 11 Penny J. Standen and David J. Brown. Virtual reality in the rehabilitation of people with intellectual disabilities: Review. *CyberPsychology & Behavior*, 8(3), 272–282, June 2005. <http://dx.doi.org/10.1089/cpb.2005.8.272>
- 12 Eduardo E. Veas, Erick Mendez, Steven K. Feiner, and Dieter Schmalstieg. Directing attention and influencing memory with visual saliency modulation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*. 1471–1480. 2011. <http://dx.doi.org/10.1145/1978942.1979158>
- 13 VirtualSpeech application. <http://virtualspeech.co.uk/2016>
- 14 Yan Xu, Joshua Ratcliff, James Scovell, Gheric Speiginer, and Ronald Azuma. Real-time Guidance Camera Interface to Enhance Photo Aesthetic Quality. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI'15)*. ACM, New York, NY, USA, pp. 1183–1186. <http://dx.doi.org/10.1145/2702123.2702418>

6.5 Group 5: EyeWear Computing for Gaming

Thad Starner (Georgia Institute of Technology – Atlanta, US)

License © Creative Commons BY 3.0 Unported license
© Thad Starner

Joint work of Sabrina Hoppe, Masahiko Inami, Moritz Kassner, Päivi Majaranta, Will Patera, Thad Starner, Julian Steil, Yusuke Sugano

Gaming seems an ideal platform for demonstrating and experimenting with EyeWear. Short, snack-style games can be used as probes for new interaction techniques, while longer form games might be used to create naturalistic and controlled scenarios for basic gaze research. Access to eye, face, and head motion enables new gaming mechanisms. For example, an advanced form of the popular game “Fruit Ninja” might require players to use relative eye motion to select a target from a group of distractors. Horror games may use knowledge about the user’s area of focus to guarantee that a new monster is rendered in the user’s peripheral field of view, increasing the level of surprise and startlement. Similarly, timing changes in the interface to correspond with eye blinks and saccadic blindness might enable a higher or lower level of stress in the game depending on the maker’s intention. Staring and squinting could be coupled with mechanisms of selection or zooming, and games might require the user make certain facial expressions to encourage certain moods in the game (e.g., snarling at an opponent or smiling at a dog that the player is trying to befriend). Using EyeWear to support gamification of everyday activities, like what Fitbit does for walking, might encourage healthy or desired behaviors. Gamification approaches can be also useful to collect large-scale data required for other related research areas. Everyday-use EyeWear games could also be used for studies, skill creation and rehearsal, persuasive interfaces, physical therapy, physical conditioning, meditation, or just relaxation. In short, EyeWear games seems a promising and enjoyable approach for rapid iteration of eye and head interfaces that might be adopted later to other applications.

6.6 Group 6: Prototyping of AR Applications using VR Technology

Scott W. Greenwald (MIT Media Lab – Cambridge, MA, USA)

License  Creative Commons BY 3.0 Unported license
© Scott W. Greenwald

Joint work of Rita Cucchiara, Ozan Cakmakci, Thies Pfeiffer

We discussed a methodology for iteratively designing and building eyewear applications using a sequence of mixed-reality prototype systems that incrementally approach the final user experience. The methodology allows researchers to derive and benefit from user-centered design insights without needing a complete prototype. Although it is already common to use paper-prototyping and “wizard of oz” techniques to pilot candidate designs, we are observing that in many cases there should be more intermediate steps between the paper prototype and the system based on the final hardware configuration. That is, many augmented reality user interactions and system affordances can be simulated with varying fidelities that are useful but cheaper or more rapid to implement. One example would be simulating an optical see-through eyewear system using a video see-through system. We also observe that small prototyping steps can be made within each iteration of the system. For example, in a system where the world is simulated using a “cave” surround projection system, one can move from simulating the use of a mobile device using the projection system to using a physical mobile device before moving completely out of the cave into the physical world. Furthermore, the intermediate design steps put bounds on hardware design requirements (such as field of view on a head-worn display). This approach aims at minimizing unnecessary hardware builds, which accelerates the hardware design cycle while reducing surprises at the end. A concrete outcome from this working group is to expand these ideas into a position paper on this subject.

7 Community Support

Another reoccurring topic discussed at the seminar was community support, i.e. how can we share tools, datasets and practices used in the different research communities present at the seminar. The following list of datasets and tools is the result of these discussions.

7.1 Datasets

- **Bristol egocentric object interactions.** 6 activities, 5 people, mobile eye tracker and egocentric camera. [2]
- **CAMSynthesEyes.** The dataset contains 11,382 synthesized close-up images of eyes. [3]
- **Databrary.** Databrary is a video data library for developmental science. [4]
- **MPI long-term visual behaviour.** Over 80 hours of visual behaviour data in everyday settings. [5]
- **Georgia tech egocentric datasets.** List of datasets on egocentric vision. [6]
- **Unimore Egocentric Vision.** Egocentric vision data sets focusing on social relationships. [8]
- **Swirski Dataset.** Pupil detection data set. [1]
- **Labelled Pupils in the Wild (LPW).** Pupil detection data set. [7]

References

- 1 <https://www.cl.cam.ac.uk/research/rainbow/projects/pupiltracking/datasets/>
- 2 Bristol Egocentric Object Interactions. <https://www.cs.bris.ac.uk/~damen/BEOID/>
- 3 AM SynthesEyes. <https://www.cl.cam.ac.uk/research/rainbow/projects/syntheseeyes/>
- 4 Databrary <https://nyu.databrary.org>
- 5 Discovery of Everyday Human Activities From Long-term Visual Behaviour. <https://www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/human-activity-recognition/discovery-of-everyday-human-activities-from-long-term-visual-behaviour-using-topic-models/>
- 6 Georgia Tech Egocentric Vision <http://cbi.gatech.edu/egocentric/datasets.htm>
- 7 MPII Labelled pupils in the wild (LPW) <http://mpii.de/LPW>
- 8 UNIMORE Egocentric vision for social relationship <http://imabelab.ing.unimore.it/imabelab/researchactivity.asp?idAttivita=23>

7.2 Tools

In the following, we enumerate some of the tools used by the communities present at the seminar.

- **WearScript.** Can be used to capture sensor data from Android, including camera frames (useful for Glass) <http://www.wearscript.com/>
- **Pupil head-mounted eye tracking.** <https://pupil-labs.com/pupil>
- **J!NS MEME Logger.** Records data from developer version of J!NS MEME on iOS. <https://github.com/shoya140/MEMELogger-iOS-developers>
- **Google Glass Logger.** Records all sensor data from glass. <https://github.com/shoya140/GlassLogger>
- Software to render ground truth annotated eye images for pupil detection and tracking / gaze estimation. <https://www.cl.cam.ac.uk/research/rainbow/projects/eyerender/>
- Caffe code for unsupervised feature learning with unlabeled video accompanied by egomotion sensor data. <http://www.cs.utexas.edu/~dineshj/projects/4-egoEquiv/>
- **TraQuMe.** Tool for checking the quality of tracking – this tool supports several trackers via COGAIN EtuDriver. <http://www.uta.fi/sis/tauchi/virg/traqume.html>
- **Snap point detection code.** http://vision.cs.utexas.edu/projects/ego_snappoints/
- Logging multimodal information for cognitive psychology – collecting applied force to the fingerpad. <http://www.tecghan.co.jp/english/p2.htm>
- **Binaural microphone for collecting soundscapes.** Roland CS-10EM In-Ear Monitors. <http://www.amazon.co.jp/dp/B003QGPCTE>
- **iMotions Biometric Research Platform** Record physiological signals synchronized with stimuli and videos of subjects. <http://imotions.com/>
- **Rapid Gesture Recognition Toolkit.** A Unix command line utility for doing quick evaluations of machine learning algorithms for gesture recognition. <http://gt2k.cc.gatech.edu/>

Participants

- Andreas Bulling
Max-Planck-Institut für
Informatik – Saarbrücken, DE
- Ozan Cakmakci
Google Inc. –
Mountain View, US
- Rita Cucchiara
University of Modena, IT
- Steven K. Feiner
Columbia University, US
- Kristen Grauman
University of Texas – Austin, US
- Scott Greenwald
MIT – Cambridge, US
- Sabrina Hoppe
Max-Planck-Institut für
Informatik – Saarbrücken, DE
- Masahiko Inami
Keio University – Yokohama, JP
- Shoya Ishimaru
Osaka Prefecture University, JP
- Moritz Kassner
Pupil Labs – Berlin, DE
- Koichi Kise
Osaka Prefecture University, JP
- Kiyoshi Kiyokawa
Osaka University – Osaka, JP
- Kai Kunze
Keio University – Yokohama, JP
- Yin Li
Georgia Institute of Technology –
Atlanta, US
- Paul Lukowicz
DFKI – Kaiserslautern, DE
- Päivi Majaranta
University of Tampere, FI
- Walterio W. Mayol-Cuevas
University of Bristol, GB
- René Mayrhofer
Universität Linz, AT
- Masashi Nakatani
University of Tokyo, JP
- Will Patera
Pupil Labs – Berlin, DE
- Thies Pfeiffer
Universität Bielefeld, DE
- James M. Rehg
Georgia Institute of Technology –
Atlanta, US
- Philipp M. Scholl
Universität Freiburg, DE
- Linda B. Smith
Indiana University –
Bloomington, US
- Gábor Sörös
ETH Zurich, CH
- Thad Starner
Georgia Institute of Technology –
Atlanta, US
- Julian Steil
Max-Planck-Institut für
Informatik – Saarbrücken, DE
- Yusuke Sugano
Max-Planck-Institut für
Informatik – Saarbrücken, DE
- Yuji Uema
J!NS – Tokyo, JP



Modern Cryptography and Security: An Inter-Community Dialogue

Edited by

Kristin Lauter¹, Radu Sion², and Nigel P. Smart³

1 Microsoft Research – Redmond, US, klauter@microsoft.com

2 National Security Institute – Stony Brook, US, radu@digitalpiglet.org

3 University of Bristol, GB, nigel@cs.bris.ac.uk

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16051 “Modern Cryptography and Security: An Inter-Community Dialogue”.

Seminar January 31 to February 5, 2016 – <http://www.dagstuhl.de/16051>

1998 ACM Subject Classification E.3 Data Encryption, K.6.5 Security and Protection

Keywords and phrases anti-surveillance/anti-censorship systems, homomorphic encryption, post-quantum cryptography, secure hardware design, secure multi-party computation, secure outsourcing, side/covert channels, virtualization security

Digital Object Identifier 10.4230/DagRep.6.1.207

1 Executive Summary

Nigel P. Smart

License © Creative Commons BY 3.0 Unported license
© Nigel P. Smart

The seminar aimed to bring together communities with different backgrounds and form a bridge between them.

The outcomes ranged from a series of bridging exercises where participants summarized the current thoughts in existing areas; these included areas such as

- Hardware Attacks: Where we summarized the known attacks in this space.
- Computing on Encrypted Data: Various aspects of this were discussed, including Secure Guard Extensions (SGX), Searchable Symmetric Encryption (SSE), Multi Party Computation (MPC), and Fully Homomorphic Encryption (FHE).

We then went on to discuss more technical aspects, rather than just summarizing work,

- Cyberphysical Systems and IoT: Where the research challenges of performing work in this new area were discussed. A reliance on practical experimental was noted in the current research landscape.
- Mass Surveillance, Trapdoors, Secure Randomness: The recent “backdooring” of the DUAL-EC random number generator formed the background of this discussion. The seminar examined different aspects of this area, both in preventing, creating and detecting backdoors.
- Anonymous Payment Systems: This was a rather broad discussion which examined a number of issues around payments in general, and how cryptography could solve address these issues.



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

Modern Cryptography and Security: An Inter-Community Dialogue, *Dagstuhl Reports*, Vol. 6, Issue 1, pp. 207–223

Editors: Kristin Lauter, Radu Sion, and Nigel P. Smart



DAGSTUHL
REPORTS

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

We also discussed aspects related to the process of research in this field. In particular focusing on the problem of the lack of expository writing. Here we identified a number of disincentives in the research culture which prevents the creation of more discursive writing and expository articles. A number of solutions both existing, and proposed, were discussed to solve this issue. In another small breakout we discussed the lack of incentives to work on the underlying hard problems upon which our security infrastructure rests.

In summary the seminar found more problems with our current research trends, than solutions.

2 Table of Contents

Executive Summary	
<i>Nigel P. Smart</i>	207
Format	210
Working groups	210
Hardware Attacks: Threat Models for Secure Hardware	
<i>Ferdinand Brasser, Raad Bahmani, Dieter Gollmann, Florian Kerschbaum, Yongdae Kim, Kristin Lauter, and Radu Sion</i>	210
Cyberphysical Systems and IoT Security	
<i>Dieter Gollmann, Alex Biryukov, Marc C. Dacier, George Danezis, Yevgeniy Dodis, Christian Grothoff, Stefan Katzenbeisser, Yongdae Kim, Moni Naor, Claudio Orlandi, Andreas Peter, and Martina Angela Sasse</i>	212
Mass Surveillance, Trapdoors, Secure Randomness	
<i>Nadia Heninger, Alexandra Boldyreva, Nikita Borisov, Marc C. Dacier, Yevgeniy Dodis, Stefan Katzenbeisser, Kenneth G. Paterson, and Andreas Peter</i>	213
Incentivizing Expository Writing	
<i>Aaron Michael Johnson, Allison Bishop, Alexandra Boldyreva, Nikita Borisov, George Danezis, Krista Grothoff, Nadia Heninger, Sarah Meiklejohn, and Radu Sion</i>	215
Computing on Encrypted Data, Secure Databases, Encrypted Cloud	
<i>Florian Kerschbaum, Melissa Chase, Jung Hee Cheon, Maria Dubovitskaya, Kristin Lauter, Giuseppe Persiano, and Benny Pinkas</i>	217
MPC: killer applications and threat models for applications	
<i>Giuseppe Persiano, Christian Grothoff, Aaron Michael Johnson, Yehuda Lindell, Claudio Orlandi, and Benny Pinkas</i>	219
Anonymous Payment Systems	
<i>Nigel P. Smart, Alex Biryukov, Allison Bishop, Bogdan Carbunar, Melissa Chase, George Danezis, Maria Dubovitskaya, Christian Grothoff, and Martina Angela Sasse</i>	220
Cryptographic Hardness Assumptions	
<i>Nigel P. Smart, Yehuda Lindell, and Kenneth G. Paterson</i>	221
Participants	223

3 Format

Unlike many Dagstuhl seminars this seminar was run on a very different format, with most time devoted to small group discussion, and one-on-one meetings. We thus reduced the total amount of “plenary” time to the minimum. The goal was to foster a dialogue between people working in two distinct but related fields, each with their own methodologies of working and presenting results.

The programme started with a series of one-on-one discussions, followed by a reporting back phase. This phase was to break the ice between participants and get participants to understand the area of research of someone from a very different background. The second phase consisted of the group selecting some common themes from the initial phase and then engaging in breakout discussions, followed by a series of plenary reports back. It is these reports back which we summarize in the abstracts contained in this document. As such the abstracts represent the combined brain storming of all the seminar participants.

This more interactive format was found to be highly successful by the participants, although very tiring, as it required concentration at all points during the week with little down time to “zone out” during someone else’s talk. All the plenary sessions were highly interactive with questions and answers coming from the floor, with the main speaker purely leading the discussion.

As can be seen from the abstracts herein, we eventually discussed a wide variety of topics from the mechanics of how our science is performed, through to detailed discussions on specific technical topics. There is no doubt that a number of new collaborations and contacts ensued from the programme, and we hope this more intense style format can be adopted by other seminars at Schloss Dagstuhl in future.

4 Working groups

4.1 Hardware Attacks: Threat Models for Secure Hardware

Ferdinand Brasser (TU Darmstadt, DE), Raad Bahmani (TU Darmstadt, DE), Dieter Gollmann (TU Hamburg-Harburg, DE), Florian Kerschbaum (SAP SE – Karlsruhe, DE), Yongdae Kim (KAIST – Daejeon, KR), Kristin Lauter (Microsoft Research – Redmond, US), and Radu Sion (National Security Institute – Stony Brook, US)

License © Creative Commons BY 3.0 Unported license
© Ferdinand Brasser, Raad Bahmani, Dieter Gollmann, Florian Kerschbaum, Yongdae Kim, Kristin Lauter, and Radu Sion

The recent developments in the area of secure hardware, in particular the introduction of Intel’s Software Guard Extensions (SGX), has yield the question under which condition secure hardware can be useful. To answer this question a threat model for secure hardware is required. This document provides an (incomplete) discussion of different classes and implementation of secure hardware with regard to a set of attack vectors. Attack vectors for secure hardware can be divided into two main groups, software attacks and hardware attacks.

4.1.1 Software Attacks

Software attacks can be carried out without physical proximity to the target system. Fault injection attacks aim to bring a secure hardware system into an invalid state to extract secret

information, certain Smartcards are known to be vulnerable; secrets from the Smartcard can be extracted through sequences of interactions with the Smartcard's interface.

Side-channels exist due to the use of shared resources. In the case of SGX, the caches of the CPU are used by the isolated environment (called enclave) and untrusted software on the same system, hence, SGX is vulnerable to cache side-channel attack. However, cache side-channels are dependent on the software executed in the enclave and can be countered by using side-channel resilient algorithm.

Memory access pattern might also leak information about the internal state of a SGX enclave. A malicious OS could observe all memory access of an enclave at page granularity. This attack can be countered by side-channel resilient algorithm, too.

TrustZone can be implemented in a way that cache side-channels are not possible. Cache flushes on transition between normal world and secure world render those attacks ineffective. Page fault side-channel are non-existing in TrustZone due to the fact that the secure world is in charge of handling page faults itself.

Dedicated secure hardware systems, like HSMs and Smartcards, do not share resources with untrusted software and are therefore not vulnerable to software-exploitable side-channels.

4.1.2 Hardware Attacks

Hardware attacks can be further divided into invasive and non-invasive attacks. Physical side-channel, like power consumption, heat, radio emission, etc., are non-invasive and can leak information about secret information processed inside secure hardware.

Protection methods against those attacks exist, however, they are specific to individual attacks. Hence, to achieve comprehensive protection secure hardware needs to implement mechanism against each possible side-channel. Although some HSMs and Smartcards are known to provide certain protections mechanism it is not possible to make general statements about entire classes of devices.

SGX and TrustZone do not provide explicit protection methods against hardware side-channels and therefore must be assumed to be vulnerable.

Destructive physical attacks, like etching of layers of hardware to extract keys stored in hardware, can again be countered by explicit methods. Processors with SGX or TrustZone are produced with state-of-the-art production methods leading to very dense designs impeding those attacks or making them extremely expensive.

Hardware trojan are another threat to secure hardware against which protection methods exist. However, the implementation of hardware trojans requires significant resources on the attacker's side (e.g., to manipulate the production process of the hardware). Given that easier attack vectors exist for most secure hardware systems it is more likely that an attacker would exploit those.

4.1.3 Conclusion

A general threat model for secure hardware cannot be constructed due to the diversity of secure hardware solutions. Even within a class of hardware systems (e.g., Smartcards) the in-homogeneity forbids general statements.

When using secure hardware available solutions must be evaluated against individual requirements.

4.2 Cyberphysical Systems and IoT Security

Dieter Gollmann (TU Hamburg-Harburg, DE), Alex Biryukov (University of Luxembourg, LU), Marc C. Dacier (QCRI – Doha, QA), George Danezis (University College London, GB), Yevgeniy Dodis (New York University, US), Christian Grothoff (INRIA – Rennes, FR), Stefan Katzenbeisser (TU Darmstadt, DE), Yongdae Kim (KAIST – Daejeon, KR), Moni Naor (Weizmann Institute – Rehovot, IL), Claudio Orlandi (Aarhus University, DK), Andreas Peter (University of Twente, NL), and Martina Angela Sasse (University College London, GB)

License © Creative Commons BY 3.0 Unported license

© Dieter Gollmann, Alex Biryukov, Marc C. Dacier, George Danezis, Yevgeniy Dodis, Christian Grothoff, Stefan Katzenbeisser, Yongdae Kim, Moni Naor, Claudio Orlandi, Andreas Peter, and Martina Angela Sasse

A CPS is a “physical and engineered systems whose operations are monitored, coordinated, controlled and integrated by a computing and communication core” [NSF]. The focus is on control and physical impact. IoT is about the networking of “things” (is a person a thing?), not necessarily via the internet. The focus is on networking.

CPS and IoT security more than a new playground for old techniques? The field is industry and business driven; performance drives demand for networking; previously air-gapped systems are opened to the Internet, but systems are there for 10+ years, old hardware, no patches for 10+ years. Incidents include Stuxnet, using legal commands to centrifuges to gradually decrease their performance, switching off a heart pacemaker by sending the heartbeat signal of a healthy person, and destroying a fail-safe pump by turning it on and off at a frequency too high to be picked up by the safety mechanism. Conclusion: CPS security is more than adapting and deploying familiar security mechanisms.

Distinguishing features? The interplay of safety and security: “fail-safe” systems depend on physical assumptions an attack may break; most work on safety builds on a world model that uses probabilities; this is not warranted in security. Inputs from and attacks on the physical layer need to be integrated in the security design of a system. Input validation is not sufficient, as shown by the pacemaker example. The attacker needs to understand how to manipulate inputs so that the system changes state in a way desired by the attacker. Challenge: One needs to understand the physical processes. Finding failure conditions is hard, process models of engineers cover how systems behave in normal operation but not necessarily in extreme situations.

Which methodology to apply to make progress in CPS security? Current research is hugely experimental, indicative of an early phase of a new field? Simulators, but where to get models and data from? Real data on real systems may be company secrets and not be made available to researchers. Does this matter? Testbeds, allow to observe real physical impact, but are limited to subsystems. Fake products widen the gap between model and reality. How to distinguish fakes from genuine items? How to guarantee that hardware or software was not modified/backdoored? Maybe one can build on the physical structure of objects; there is also the issue of supply-chain management. Final note: Is CPS security a matter of research or a matter of education? Security people need to learn about chemical plants, power grids, and other critical infrastructures. Operators need to learn thinking “security”.

4.3 Mass Surveillance, Trapdoors, Secure Randomness

Nadia Heninger (University of Pennsylvania – Philadelphia, US), Alexandra Boldyreva (Georgia Institute of Technology – Atlanta, US), Nikita Borisov (University of Illinois – Urbana Champaign, US), Marc C. Dacier (QCRI – Doha, QA), Yevgeniy Dodis (New York University, US), Stefan Katzenbeisser (TU Darmstadt, DE), Kenneth G. Paterson (Royal Holloway University of London, GB), and Andreas Peter (University of Twente, NL)

License © Creative Commons BY 3.0 Unported license
© Nadia Heninger, Alexandra Boldyreva, Nikita Borisov, Marc C. Dacier, Yevgeniy Dodis, Stefan Katzenbeisser, Kenneth G. Paterson, and Andreas Peter

4.3.1 Problem

We began by discussing what the scope of mass surveillance is. Surveillance is directed at a target population, and “mass” means the surveillance is directed at the vast majority of the targeted population. This can be done by both governments and companies.

We know that mass surveillance is happening on medium and large scales. More “local” levels of mass surveillance include surveillance of cell phone towers, IMSI catchers, local ISPs, or smaller countries performing surveillance of connections between their countries and the rest of the world.

We discussed the model by which countries would man-in-the-middle connections transiting borders. For HTTPS, there are several examples of countries obtaining fraudulent HTTPS certificates for companies, and using false DNS records or similar to redirect vulnerable traffic to middleboxes who can then impersonate the end site to the users within the country.

Global-scale issues include backdoors being built into communications and cryptographic infrastructure. We discussed the specific case of the Dual-EC DRBG, where the construction of the random number generator allows an entity who generates the input parameters adversarially to recover the state and future outputs from a single output. The standards are known to be influenced by the NSA and GCHQ, and contain a recommended set of parameters that was generated by the NSA, instead of specifying that users generate their own. Additionally, there are widespread rumors from participants on standards committees of agency interference to weaken cryptographic standards. We learned in December 2015 that Juniper used this random number generator in NetScreen products, with parameters they generated themselves and feeding the output into another PRNG that should hide the direct output. However, an unknown party replaced the parameters with different parameters, and the implementation contained a subtle bug that caused the direct output of the RNG cascade to be raw Dual-EC output.

4.3.2 Solutions and ideas for man-in-the-middling

For local levels of surveillance, there are existing technical solutions that can detect and sometimes prevent some kinds of surveillance.

For example, for the problem of countries man-in-the-middling connections that transit fixed exit points, HTTPS with certificate pinning, certificate transparency, and two-factor authentication works in practice. Solutions like certificate pinning don’t scale to all types of hosts. If there were a worldwide public-key infrastructure for clients, then TLS could do a two-way authenticated key exchange, but this is not in wide use for clients. Additional client authentication tends to take place after a one-way authenticated TLS session is established, and the client authenticates inside of the channel via a password or temporary code provided via a two-factor authentication device.

We talked about ways to make these schemes more cryptographically integrated to prevent the man-in-the-middle from successfully authenticating as a client even if they are successful in impersonating the site to the client, for example performing a challenge-response protocol combining a host's public key with a secret obtained out-of-band via two-factor authentication. This does not seem to be possible to implement securely in a browser user interface, because a naive implementation would allow the MITM to display a UI element to the host to enter a secret, and there is no way to validate the page/Javascript source implementing this. HTTPS certificate information is also not accessible via the page DOM. It may be possible to implement such a scheme using a browser extension and a TLS cipher suite with PSK authentication.

We also discussed the approach of using plausible deniability to increase the workload of an entity performing mass surveillance. Users could produce a large amount of spurious or cover traffic to try to confuse automated tests. However, it's unclear whether this would actually deter those implementing these schemes, or merely cause more false positives.

4.3.3 Ideas for crypto backdoors

We discussed ways of detecting cryptographic backdoors. We know that theoretical "kleptography" systems can be perfect: one can design a cryptographically undetectable backdoor. But implementors are imperfect. In practice, OpenSSL failed to implement the Dual EC DRBG correctly. (Yet it still passed FIPS certification, raising some questions.) Mistakes in implementations might reveal the presence of backdoors.

Additionally, the Dual EC backdoor is not cryptographically hidden: we can see that the parameters can be backdoored, even if we can't prove that a particular set of parameters was maliciously generated. For this case, there is a known bias that might allow discovery of this traffic in a black-box way given a large quantity of traffic, but for other implementations it's unclear what can be done.

Companies might want to prove to clients that implementations do not contain backdoors. They can publish open-source code, but customers don't have assurance that the code corresponds to the binaries they download. The Tor Project has been doing work on deterministic builds to allow external parties to verify this.

There has been some theoretical work on immunizing schemes against backdoors, or designing cryptographic schemes that cannot be backdoored. This is an area of current and future research.

4.4 Incentivizing Expository Writing

Aaron Michael Johnson (NRL – Washington, US), Allison Bishop (Columbia University – New York, US), Alexandra Boldyreva (Georgia Institute of Technology – Atlanta, US), Nikita Borisov (University of Illinois – Urbana Champaign, US), George Danezis (University College London, GB), Krista Grothoff (GNUNet e. V. – Rennes, FR), Nadia Heninger (University of Pennsylvania – Philadelphia, US), Sarah Meiklejohn (University College London, GB), and Radu Sion (National Security Institute – Stony Brook, US)

License © Creative Commons BY 3.0 Unported license

© Aaron Michael Johnson, Allison Bishop, Alexandra Boldyreva, Nikita Borisov, George Danezis, Krista Grothoff, Nadia Heninger, Sarah Meiklejohn, and Radu Sion

4.4.1 Problem

Expository writing helps consolidate research knowledge and communicate it to groups outside of the community of active researchers. However, there is a general lack of incentives for high-quality research exposition of security and cryptography research. The problem is particularly acute for cryptography because it has a higher technical barrier to entry. As a result, research fields are unnecessarily difficult for new researchers to enter, and their results challenging for outside communities to make use of.

Several kinds of expository writing is being undervalued by the security and cryptography research communities. These include (i) systematizations of recent results for other researchers, (ii) writing for practitioners (e.g. people who implement systems, companies looking to commercialize technology), and (iii) popularization of research results for a general audience. Research surveys and lecture notes are at least generally recognized by researchers as having some value, but writing for practitioners or a general audience is infrequently rewarded, and these latter two types of writing have substantial value. For example, implementing cryptographic protocols based on research papers is very difficult for non-researchers. Also, the general audience is vastly larger, and currently it is primarily informed by journalists.

The lack of incentives for expository writing arises primarily within hiring and tenure committees. Tenure committees heavily weight top-tier research publications, which makes expository writing not a good use of time for pre-tenure faculty. Hiring committees also count citations weighted by venue reputation. This issue is similar to the risk of doing interdisciplinary work in that hiring and tenure candidates are frequently evaluated mostly based on their relative publication success within a given research community.

We also note that research communities themselves frequently benefit from the obfuscation of their own results. Appearing simple or easy to understand can lower the perceived value of a paper, particularly among theoreticians. Also, producing quality writing is difficult, but it is not a main consideration for accepting conference submissions, and thus producing high-quality writing is not always a maximally rewarding use of time for researchers. This seems to be a worse problem in some communities than others (e.g. security seems to value simplicity and clarity while PL and theory seems to value complexity more)

We do observe that funding bodies attempt to incentivize impacts beyond citations within a narrow field of research. For example, the NSF in the US explicitly requests “broader impact” statements, DARPA in the US often runs programs with the goal of transitioning technology to industry or government, and REF in the UK values fewer papers with larger impact.

4.4.2 Some existing solutions

There are some types of exposition that are currently working moderately well within the security and cryptography research communities, the successes and failures of which we can learn from. For example,

1. There are journals for survey papers (e.g. ACM Computing Surveys), and the surveys can receive significant numbers of citations. However, such journals are not valued highly by hiring and tenure committees.
2. Professors often produce lecture notes or books for their courses.
3. Certain security and cryptography conferences solicit “systematization of knowledge” (SoK) submissions, including IEEE Security & Privacy and the Proceedings of Privacy Enhancing Technologies.
4. There do exist widely-recognized publications that popularize security research, including USENIX ;login:, Communications of the ACM, and IEEE S&P magazine.
5. Some individual or group research blogs reach wide audiences. Blogs aren’t highly valued by research committees, and may be labors of love, but they may also have important second-order benefits such as attracting collaborators, students, and funding.
6. Some researchers serve as contacts for the news media. This can eventually lead to increased funding, but this is highly variable, and talking to reporters can be time consuming.
7. Massive Online Open Courses (MOOCs) have been produced on relatively new topics, and they can contain new texts in addition to videos.
8. Books (e.g. research monographs) are quite valuable when produced, although they can be too slow for fast-moving research fields, and their limited value to the researcher’s career means they frequently end up being written by tenured faculty or by people outside the core research community.

4.4.3 Future solutions

We propose several potential solutions to incentivize expository writing:

1. The most direct way to promote expository writing would be for top security and cryptography conferences to request systematization of knowledge (SoK) papers in their calls for papers. As mentioned, some such conferences already do this, and it has resulted in many valuable expository papers on important current topics, including Bitcoin, secure messaging, and website fingerprinting. The short format often required in conferences isn’t ideal for exposition, however, and so journals should adopt this strategy as well. However, extra length should not be taken as an invitation for simple laundry lists of previous research, as topic surveys can easily become.
2. Journals or conferences can invite specific researchers to contribute high-quality exposition on a given topic. Conferences could combine this invitation with a keynote or tutorial invitation. It would likely be recognized as a valuable contribution because of the reputation of the journal or conference.
3. Specific “exposition retreats” or “SoK workshops” can be organized with a primary goal of producing a written exposition of a given topic. Contributors would produce different sections or chapters of a cohesive paper or book. The contributors could be invited or could propose beforehand and be selected competitively. This can be combined with summer/winter schools that are already common in cryptography by asking presenters to contribute written versions of their lectures to be combined into a set of lecture notes.

4. Graduate students might be encouraged or expected perform this function as part of their degree. Some universities and professors essentially already require this (e.g. as a qualifying exam or as part of a master's thesis). However, graduate students may lack the perspective of more experienced researchers to produce especially broad or deep exposition, and this may not suffice in fields with a high barrier to entry, such as heavily theoretical areas.

4.5 Computing on Encrypted Data, Secure Databases, Encrypted Cloud

Florian Kerschbaum (SAP SE – Karlsruhe, DE), Melissa Chase (Microsoft Corporation – Redmond, US), Jung Hee Cheon (Seoul National University, KR), Maria Dubovitskaya (IBM Research Zürich, CH), Kristin Lauter (Microsoft Research – Redmond, US), Giuseppe Persiano (University of Salerno, IT), and Benny Pinkas (Bar-Ilan University – Ramat Gan, IL)

License  Creative Commons BY 3.0 Unported license
© Florian Kerschbaum, Melissa Chase, Jung Hee Cheon, Maria Dubovitskaya, Kristin Lauter, Giuseppe Persiano, and Benny Pinkas

4.5.1 Objective and Methodology

The objective of the session was to discuss different technologies for encrypted computation. The discussion should result in a comparison of advantages and disadvantages, best-fitting use cases, and future direction of research. A single use case of outsourced, private, potentially verifiable computation of arbitrary functions in the cloud was chosen, i.e. no restriction to type of application, e.g. DRM or search. For this use case we compared the technologies of SGX, FHE, MPC and SSE/OPE. Each technology was first discussed in general terms attempting to reach a common understanding of its (security) functionality. Then we compared properties, assumptions and attacks – mostly from a security perspective, but also from economic or functionality aspects. The summary and conclusion of this discussion is listed below.

4.5.2 SGX

SGX provides a unique private, public key pair per processor. This key can be used to send encrypted data, sign the loaded code and messages. The public key must be managed in a PKI by Intel. Messages can be sent to the enclave openable only under the condition the code has been attested.

SGX provides integrity of the loaded code by mechanism comparable to remote attestation. The private key is protected by hardware. The memory of the enclave is protected by encryption, i.e. there is limited interference with other process on the cloud. The public key can also be used to tie a program to a specific processor. Management of keys for server farms can still be challenged. For communication with the client a session key needs to be established. Secure channels need to be implemented. Data stored outside of the enclave – disc or memory – needs to be encrypted.

Intel is trusted to securely generate the private key and not maintain a copy. Intel is trusted to securely manage the root CA key. All code inside the enclave is trusted. Intel and the cloud provider are trusted not to collude.

There seem to be side-channels that can leak either the private key of the processor or data, including the session key. These side channels could be timing, energy consumption or

memory access patterns. Access patterns could be read via cache-timing attacks or physically from the bus. These could be combined with known attacks, on e.g. AES, for the session key.

The code inside the enclave could be vulnerable or contain backdoors. This includes code for establishing a secure channel (SSL, etc.). The code must be secure against replay attacks – potentially also from permanent storage, like disk.

There could be simulation or man-in-the-middle attacks, if the PKI fails. Generally, it is not clear how to transfer the key of the processor (or server farm) to the client. Group signatures could help.

There could be hardware attacks, e.g. sniffing the bus. While attacks by sysadmins are harder, they do have such access.

SGX seems only agreeably secure for computationally intensive tasks with little data on the client's self-written code. From an economic point SGX seems well suited. The cloud provider makes an initial investment, but can charge a higher fee.

4.5.3 FHE

There are different type of homomorphic encryption schemes. Partially (1 operation) homomorphic schemes, like Paillier, are known for a long time. Efficient fully, somewhat (2 operation) homomorphic schemes are based on the RLWE or LWE property. They support low-depth circuits and are reasonably fast. For deep circuits the error amplification requires either large parameters or bootstrapping which is complicated. Their challenge are the large key and ciphertext size.

Verification of the function is not included per se. The privacy of the computation is based only on a security assumption (RLWE/LWE). Schemes provide at least IND-CPA style security. Some information about the circuit leaks.

The only assumptions are cryptographic, such as RLWE or LWE.

Attacks could arise from side-information, such as the result or other consequences, used as a partial decryption oracle. This may break the IND-CPA model.

4.5.4 MPC

MPC allows any function to be computed on encrypted data by a set of servers. The servers may be split across organizational or legislative borders or be within a single domain (but different sysadmins). Sysadmins are a frequent target of attacks, hence MPC may help. A split may also be necessitated by legal obligations. The economic motivation for splitting the computation is challenging. A secure computation service may help under certain circumstances. MPC has fault-tolerance (availability) built-in.

MPC provides fully encrypted computation, the data is never in the clear. However, the function is usually known to all servers. A universal circuit can help avoid this leakage. MPC is general, i.e. for any circuit.

The servers share interest in carrying out the joint computation. However, the servers are assumed not to collude. There is no cryptographic assumption for secret shares. Broadcast or secure channels may be required.

There are many security models, such as semi-honest, covert, malicious. The semi-honest model only guarantees confidentiality, if integrity is preserved. This is similar to SGX where the code must be attested or confidentiality of the data cannot be guaranteed.

Obviously there are collusion attacks, but also secure channels are established by cryptographic means and can be attacked.

4.5.5 SSE/OPE

Searchable and order-preserving encryption are limited to search only. They are symmetric key crypto systems and the key is only held by the client. However, they are highly optimized and very, very, fast.

They are extremely fast, but targeted for predefined, limited functionality. SSE requires a specific implementation (search procedure) whereas OPE can be retrofitted into existing applications.

Custom security models, such IND-CKA1, IND-CKA2, IND-OCPA, IND-FAOCPA, are devised for new schemes.

Security against malicious attackers or IND-CCA2 security is likely not achievable, although desirable. These attacks assume the worst-case and hence make minimal assumptions providing longer lasting security and the strongest security guarantee. Three different type of attacks can distinguished: Attacks based on static leakage have already been demonstrated. Attacks from dynamic information, such as queries and access patterns, are likely. Attacks based on updates are not even yet fully included in the models and algorithms. All attacks can be based on different assumptions about the adversary's knowledge or choice of plaintexts and ciphertexts.

4.6 MPC: killer applications and threat models for applications

Giuseppe Persiano (University of Salerno, IT), Christian Grothoff (INRIA – Rennes, FR), Aaron Michael Johnson (NRL – Washington, US), Yehuda Lindell (Bar-Ilan University – Ramat Gan, IL), Claudio Orlandi (Aarhus University, DK), and Benny Pinkas (Bar-Ilan University – Ramat Gan, IL)

License © Creative Commons BY 3.0 Unported license
© Giuseppe Persiano, Christian Grothoff, Aaron Michael Johnson, Yehuda Lindell, Claudio Orlandi, and Benny Pinkas

4.6.1 Killer Applications

We have identified the following as applications domain where the need for MPC is needed.

- Statistics over distributed systems.
Several organizations have a large user base and would like to collect statistics of the users. The type of statistics that one is interested in has a big potential impact on the efficiency of the protocol. Arithmetic statistics (like average, standard deviation) are easier to compute than other more robust statistics (like median).
- Sharing sensitive data.
One way to protect keys is to share them in a secure way and to distribute the shares to parties (possibly running different OSes). In this way if one of the parties is compromised then the key is still safe. Whenever the key is needed to access encrypted data (or to perform entity/data authentication, . . .) the parties holding the shares will perform the action required (decryption, authentication, . . .) by means of MPC.
- Privacy preserving.
MPC can also be used to protect privacy of users that are really concerned about their private data and would resort to MPC whenever their personal data was needed.
- Auction.

The domain of electronic marketplaces seems to be another area that might benefit from the use of MPC. The first example is action that could be very efficiently implemented in a secure way and would provide an added level of privacy to the users that desire so.

4.6.2 Model threats for applications

Research effort in MPC has focused primarily in obtaining results that would guarantee the best possible security along with several other desirable properties like correctness, fairness, termination, . . . This approach has been very successful and has led to very general results. We observed though that in several applications not all properties are needed and one could then obtain more efficient/practical construction. The primary example that came up is about termination. If one is running MPC within an organization for the purpose of protecting keys, it would be actually desirable to learn that one of the parties is compromised by observing some execution being aborted.

Dishonest majority arises naturally in the context of secure two-party computation where the problem with honest majority is trivial.

An interesting model that seems to have been adopted by most (if not all) current industrial implementations of MPC relies on a restricted number (as little as two) servers that perform MPC over the shares of data provided by the users. Having a restricted number of players in an MPC has the obvious advantage of increasing efficiency and, in addition, it is much easier to assess the trustworthiness of few parties. This model seems particularly fit for the problem of computing statistics over large distributed system that have already identified a restricted number of nodes for other administrative tasks.

4.7 Anonymous Payment Systems

Nigel P. Smart (University of Bristol, GB), Alex Biryukov (University of Luxembourg, LU), Allison Bishop (Columbia University – New York, US), Bogdan Carbunar (Florida International University – Miami, US), Melissa Chase (Microsoft Corporation – Redmond, US), George Danezis (University College London, GB), Maria Dubovitskaya (IBM Research Zürich, CH), Christian Grothoff (INRIA – Rennes, FR), and Martina Angela Sasse (University College London, GB)

License © Creative Commons BY 3.0 Unported license
© Nigel P. Smart, Alex Biryukov, Allison Bishop, Bogdan Carbunar, Melissa Chase, George Danezis, Maria Dubovitskaya, Christian Grothoff, and Martina Angela Sasse

The group discussed a number of topics related to payments in general. These ranged from payment systems which are bitcoin like, through to systems based on store loyalty points, credit card systems, smart metering and bartering. An issue with all systems was to define the nature of anonymity, and to whom anonymity is maintained. For example early systems proposed for online banking transactions like SET tried to maintain a cryptographic separation of data based on a “need to know principle”. This never took off, and has since been replaced by “best practice” requirements as in the PCI standards. This means that merchants are exposed to identifying information about customers, even when they do not need to be (for example in the purchase of digital goods).

We discussed issues of exchange between various reserves of value; for example altcoin exchanges are already in existence. Members of the group discussed the benefit of usage of

exchange between non-currency based reserves of value. For example the trading of frequent flyer miles with, say, store card points.

An example discussed in detail was the anonymity requirements in smart metering, which is essentially a payment system for electricity. The different stakeholders in the system were discussed and how their requirements for visibility of the transactions conflict with each other parties utility. An interesting aspect of the smart metering case study is that anonymity is not required for the identities, but is required for the amounts. This is the opposite of the case in bitcoin. Many of the aspects of the smart metering example also apply to digital goods such as Spotify.

4.8 Cryptographic Hardness Assumptions

Nigel P. Smart (University of Bristol, GB), Yehuda Lindell (Bar-Ilan University – Ramat Gan, IL), and Kenneth G. Paterson (Royal Holloway University of London, GB)

License © Creative Commons BY 3.0 Unported license
© Nigel P. Smart, Yehuda Lindell, and Kenneth G. Paterson

There are two problems we face at the moment

- People are not working on breaking of hard problems.
- People are not trying to build cryptosystems under minimal assumptions.

As an example of the first problem, few people are seriously working on factoring, discrete logarithms (bar characteristic two fields), or bilinear assumptions. This is a problem of incentives; for example a mathematician who might be interested in working on discrete logarithms would take a long time to come up with any result (if any result is possible), and would end up publishing outside their field. Thus their publication record would be damaged by engaging in working on hard problems. It seems a difficult to counteract this problem of incentives, as it goes to the heart of what is needed to become a successful academic.

The second problem is typified by what we see in the area of iO currently. Researchers seem to be incentivized in coming up with applications of iO and are less incentivized in understanding the underlying problems. This is interesting when compared to the similar “bandwagon” created by FHE: with that bandwagon, researchers worked both on simplifying the underlying constructions and improving the hardness assumptions (e.g. the creation of the LWE and NTRU based schemes compared to the original Gentry scheme), as well as looking at potential applications in areas such as verifiable computation.

There are plenty of problems in the space which are relevant in the real world but are not being addressed. For example, there are plenty of constructions (efficient ZKPoKs) for which we have no analogue in the post-quantum world. Are we likely to end up with a plethora of assumptions in the world of PQC as soon as there is an urgent need for PQC methods beyond encryption and signatures? This need is likely to become a pressing need with the next few years as companies start to look at deploying post-quantum systems.

There seems to be an incentive in research into creating new applications and functionalities within cryptography, as opposed to looking at old problems and finding solutions under better assumptions, or looking at making old applications better (where the metric is “better” is either security, implementation techniques etc). If we look at the top conferences, papers which create new applications (e.g. iO, FHE, etc) seem to get more traction than papers which implement things better. This is despite the CFP for CRYPTO stating implementation or industrially relevant research being welcome for a number of years.

One solution suggested by Angela, would be to have a conference in which half worked out exploratory ideas could be batted around and discussed. Each paper is presented and then discussed by the audience. With post-proceedings which contain the feedback from the audience. This model has apparently worked well in the security community. The question is when did this start becoming a problem? When did “irrational exuberance”¹ take over our field? Perhaps the plethora of pairing based assumptions in the early 2000’s led to our current state of affairs. The authors feel that researchers need to get back to cryptographic basics which have high impact:

- Encourage work looking at schemes based on more standard assumptions.
- Encourage work which tries to improve the efficiency of schemes and their practicality.

¹ A phrase borrowed from Phil Rogaway.

Participants

- Raad Bahmani
TU Darmstadt, DE
- Daniel J. Bernstein
University of Illinois –
Chicago, US
- Konstantin Beznosov
University of British Columbia –
Vancouver, CA
- Alex Biryukov
University of Luxembourg, LU
- Allison Bishop
Columbia University –
New York, US
- Alexandra Boldyreva
Georgia Institute of Technology –
Atlanta, US
- Nikita Borisov
University of Illinois –
Urbana Champaign, US
- Ferdinand Brasser
TU Darmstadt, DE
- Christian Cachin
IBM Research Zürich, CH
- Bogdan Carbunar
Florida International University –
Miami, US
- Melissa Chase
Microsoft Corporation –
Redmond, US
- Jung Hee Cheon
Seoul National University, KR
- Marc C. Dacier
QCRI – Doha, QA
- George Danezis
University College London, GB
- Yevgeniy Dodis
New York University, US
- Maria Dubovitskaya
IBM Research Zürich, CH
- Dieter Gollmann
TU Hamburg-Harburg, DE
- Christian Grothoff
INRIA – Rennes, FR
- Krista Grothoff
GNUNet e.V. – Rennes, FR
- Nadia Heninger
University of Pennsylvania –
Philadelphia, US
- Aaron Michael Johnson
NRL – Washington, US
- Stefan Katzenbeisser
TU Darmstadt, DE
- Florian Kerschbaum
SAP SE – Karlsruhe, DE
- Yongdae Kim
KAIST – Daejeon, KR
- Tanja Lange
TU Eindhoven, NL
- Kristin Lauter
Microsoft Research –
Redmond, US
- Yehuda Lindell
Bar-Ilan University –
Ramat Gan, IL
- Sarah Meiklejohn
University College London, GB
- Refik Molva
EURECOM –
Sophia Antipolis, FR
- Moni Naor
Weizmann Institute –
Rehovot, IL
- Claudio Orlandi
Aarhus University, DK
- Kenneth G. Paterson
Royal Holloway University of
London, GB
- Adrian Perrig
ETH Zürich, CH
- Giuseppe Persiano
University of Salerno, IT
- Andreas Peter
University of Twente, NL
- Benny Pinkas
Bar-Ilan University –
Ramat Gan, IL
- Martina Angela Sasse
University College London, GB
- Vitaly Shmatikov
Cornell Tech NYC, US
- Radu Sion
National Security Institute –
Stony Brook, US
- Nigel P. Smart
University of Bristol, GB
- Gene Tsudik
University of California –
Irvine, US
- Avishai Wool
Tel Aviv University, IL



Dark Silicon: From Embedded to HPC Systems

Edited by

Hans Michael Gerndt¹, Michael Glaß², Sri Parameswaran³, and Barry L. Rountree⁴

- 1 TU München, DE, gerndt@in.tum.de
- 2 FAU Erlangen-Nürnberg, DE, michael.glass@fau.de
- 3 UNSW – Sydney, AU, sridevan@cse.unsw.edu.au
- 4 LLNL – Livermore, US, routree@llnl.gov

Abstract

Semiconductor industry is hitting the utilization wall and puts focus on parallel and heterogeneous many-core architectures. While continuous technological scaling enables the high integration of 100s–1000s of cores and, thus, enormous processing capabilities, the resulting power consumption per area (the power density) increases in an unsustainable way. With this density, the problem of Dark Silicon will become prevalent in future technology nodes: It will be infeasible to operate all on-chip components at full performance at the same time due to the thermal constraints (peak temperature, spatial and temporal thermal gradients etc.). However, this is not only an emerging threat for SoC and MPSoC designers, HPC faces a similar problem as well: The power supplied by the energy companies as well as the cooling capacity does not allow to run the entire machine at highest performance anymore. The goal of Dagstuhl Seminar 16052 “Dark Silicon: From Embedded to HPC Systems” was to increase the awareness of the research communities of those similarities and to work and explore common solutions based on more flexible thermal/power/resource management techniques both for runtime, design time as well as hybrid solutions.

Seminar January 31 to February 3, 2016 – <http://www.dagstuhl.de/16052>

1998 ACM Subject Classification C.3 [Special-Purpose and Application-based Systems] Real-time and embedded systems, C.1.4 [Processor Architectures] Parallel Architectures

Keywords and phrases dark silicon, embedded, hpc, parallel computing, performance analysis and tuning, power density, power modelling, programming tools, resource management

Digital Object Identifier 10.4230/DagRep.6.1.224

Edited in cooperation with Michael Glaß

1 Executive Summary

Hans Michael Gerndt

Michael Glaß

Sri Parameswaran

Barry L. Rountree

License  Creative Commons BY 3.0 Unported license
© Hans Michael Gerndt, Michael Glaß, Sri Parameswaran, and Barry L. Rountree

Topic

Dark Silicon

Semiconductor industry is hitting the utilization wall and puts focus on parallel and heterogeneous many-core architectures. While continuous technological scaling enables the high



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

Dark Silicon: From Embedded to HPC Systems, *Dagstuhl Reports*, Vol. 6, Issue 1, pp. 224–244

Editors: Hans Michael Gerndt, Michael Glaß, Sri Parameswaran, and Barry L. Rountree



Dagstuhl Reports

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

integration of 100s-1000s of cores and, thus, enormous processing capabilities, the resulting power consumption per area (the power density) increases in an unsustainable way. With this density, the problem of Dark Silicon will become prevalent in future technology nodes: It will be infeasible to operate all on-chip components at full performance at the same time due to the thermal constraints (peak temperature, spatial and temporal thermal gradients etc.).

Recent research work on power management for Dark Silicon aims at efficiently utilizing the TDP (Thermal Design Power) budget to maximize the performance or to allocate full power budget for boosting single-application performance by running a single core at the maximum voltage or multiple cores at nominal level for a very short time period. Control-based frameworks are proposed to find the optimal trade-off between power and performance of many-core systems under a given power budget. The controllers are coordinated to throttle down the power when the system exceeds the TDP and to assign the task to the most suitable core to get the optimal performance. The work on near-threshold computing (NTC) enables operating multiple cores at a voltage close to the threshold voltage. Though this approach favors applications with thread-level parallelism at low power, it severely suffers from errors or inefficiency due to process variations and voltage fluctuations. On the other hand, the computational sprinting approach leverages Dark Silicon to power-on many extra cores for a very short time period (100s of millisecond) to facilitate sub-second bursts of parallel computations through multi-threading but thereby wasting a significant amount of energy due to leakage current. When doing so, it consumes power that significantly exceeds the sustainable TDP budget. Therefore, these cores are subsequently power-gated after the computational sprint. Alternate methods are Intel's Turbo Boost and AMD's Turbo CORE technologies that leverage the temperature headroom to favor high-ILP applications by increasing the voltage/frequency of a core while power-gating other cores. These techniques violate the TDP constraint for a short period (typically in terms of 10s of seconds) until the critical temperature is reached and then switches to a nominal operation. However, in case of dependent workloads, boosting of one core may throttle the other due to thermal coupling (i.e. heat exchange between different cores sharing the same die). Therefore, these boosting techniques lack efficiency in case dependent tasks of an application mapped to two different cores or, in general, for multiple concurrently executing applications with distinctive/dependent workloads.

State-of-the-art boosting techniques assume a chip with only 10-20 cores (typically 16) and accordingly a full chip temperature violation for short time. However, in a large-scale system (with 100s-1000s cores), temperature hot spots may occur on certain chip portions far before the full chip's average temperature exceeds the critical temperature. Therefore, a chip may either get damaged before reaching the full chip critical temperature or TDP needs to be pessimistically designed. Advanced power management techniques are required to overcome these challenges in large-scale environments.

HPC – Dark Power

The energy consumption of HPC systems is steadily growing. The costs for energy in the five year lifetime of large scale supercomputers already almost equal the cost of the machine. It is a necessity to carefully tune systems, infrastructure and applications to reduce the overall energy consumption. In addition, the computing centers running very big systems face the problem of limited power provided by the energy providers and of the requirement for an almost constant power draw from the grid. The big machines, especially future exascale systems, are able to use more power if they are run at highest performance of all components than can be provided by the energy company. Thus, a carefully optimized power distribution

is necessary to make most efficient use of the provided power. The second aspect is the requirement of an almost constant power draw: Sudden changes from 20 MW to 10 MW for example, will be dangerous for the components of the power grid. In addition, the contracts with the energy companies force the centers use the same power all the time by charging more, if it drops below or exceeds certain limits. These challenges also require a careful and flexible power and resource management for HPC systems.

For a certain class of high-end supercomputer, there is a standard pattern of power consumption: During burn-in (and perhaps while getting a result to go onto the top-500 list) the machine will run dozens or hundreds of instances of Linpack. This code is quite simple and often hand-optimized, resulting in an unusually well-balanced execution that manages to keep vector units, cache lines and DRAM busy simultaneously. The percent of allocated power often reaches 95 % or greater, with one instance in recent memory exceeding 100 % and blowing circuit breakers. After these initial runs, however, the mission-critical simulation codes begin to execute and they rarely exceed 60 % of allocated power. The remaining 40 % of electrical capacity is dark: just as unused and just as inaccessible as dark silicon. While we would like to increase the power consumption (and thus performance) of these simulation codes, a more realistic solution in the exascale timeframe is hardware overprovisioning. This solution requires buying more compute resources than can be executed at maximum power draw simultaneously. For example, if most codes are expected to use 50 % of allocated power, the optimal cluster would have twice as many nodes.

Making this a feasible design requires management of power as a first-class resource at the level of the scheduler, the run-time system, and on individual nodes. Hardware power capping must be present. Given this, we can theoretically move power within and across jobs, using all allocated power to maximize throughput. The purpose of this seminar is to find this optimal level.

Hybrid (Design-time & Run-time) Resource Management

Today's complex applications need to exploit the available parallelism and heterogeneity of – non-darkened – cores to meet their functional and non-functional requirements and to gain performance improvements. From a resource management's point of view, modern many-core systems come with significant challenges: (a) Highly dynamic usage scenarios as already observable in today's "smart devices" result in a varying number of applications with different characteristics that are running concurrently at different points in time on the system. (b) Due to the constraints imposed by the power density, the frequency at which cores can be operated as well as their availability as a whole, are subject to change. Thus, resource management techniques are required that enable a resource assignment to applications that satisfies their requirements but at the same time can consider the challenging dynamics of modern many-cores as a result of Dark Silicon.

Traditional techniques to provide a binding or pinning of applications to processor that are optimal and predictable with respect to performance, timing, energy consumption, etc. are typically applied at design time and result in a kind of static system design. Such a static design may, on the one hand, be too optimistic by assuming that all assigned resources are always available or it may require for a kind of over-allocation of cores to compensate for worst-case scenarios, e.g., a frequent unavailability of cores due to Dark Silicon. Hence, the dynamic effects imposed in Dark Silicon require for novel modeling techniques already at design time.

Approaches that focus on pure run-time resource management are typically designed with flexibility in mind and should inherently be able to dynamically react to changing applications

as well as to the described effects of Dark Silicon. But, future run-time resource management should not only react to a possible violation of a maximum power-density constraint, but also be able to proactively avoid such situations. The latter is an important aspect of the system's dependability as well. At the same time, such dynamic resource management is also required to regard the applications' requirements. Here, a careful consideration on whether pure run-time management strategies enable the amount of predictability of execution qualities required by some applications becomes necessary.

A recent research direction focuses on hybrid (design-time and run-time) approaches that explore this field of tension between a high predictability of design-time approaches and the dynamic adaptivity of run-time resource management. In such approaches, design-time analysis and optimization of the individual applications is carried out to capture information like core allocation, task binding, or message routing and predict resulting quality numbers like timeliness, energy consumption, or throughput. This information is then passed to the run-time resource management that then dynamically selects between the pre-optimized application embeddings. Such strategies may not only be able to achieve application requirements even in such highly dynamic scenarios, but could even balance the requirements of the individual applications with the system's requirements – in particular the maximum power density. On the other hand, coarse-grained resource management as required for core allocation etc. may be considered to happen on a longer time scale. The effects of Dark Silicon are instead on a smaller time scale with temperature almost immediately following changing workloads, thus, requiring for an intervention of the resource-management infrastructure. Therefore, novel concepts are required that enable a fine-grained resource management in the presence of Dark Silicon – both in the context of abstraction layer and time scale – without sacrificing the required efficiency but also predictable realization of application requirements via coarse-grained resource management.

Goals

Traditionally, resource management techniques play an important role in both domains – targeting very different systems. But, as outlined before, resource management may be the key to tackle the problem of dark silicon that both communities face. The aim of this seminar is to give an overview of the state of the art in the area of both embedded and HPC. It will make both groups aware of similarities and differences. Here, the competences, experiences, and existing solutions of both communities shall stimulate discussions and co-operations that hopefully manifest in innovative research directions for many-core resource management in the dark silicon era.

Overview of Contributions

This seminar presentations on the state-of-the-art in power and energy management in HPC and on techniques mitigating the Dark Silicon problem in embedded systems. In a joint session commonalities and differences as well as collaboration potential in the area of Dark Silicon were explored. This subsection gives an overview of the topics covered by the individual speakers in the seminar. Please refer to the included abstracts to learn more about individual presentations.

The HPC-related presentations were started with an overview presentation by Barry Rountree from the Lawrence Livermore National Laboratory. He introduced the field of HPC

and of exascale systems. The new challenge is that these systems will be power limited and the hardware is overprovisioned. Techniques increasing the efficient usage of the available power need to be developed. Exascale systems will be heterogeneous, even systems with homogeneous cores become heterogeneous due to production variability which takes effect under power limits. Careful distribution of power among jobs and within jobs as well as application and system configurations for jobs will be important techniques for these power limited and overprovisioned systems.

Axel Auweter added to this introduction deep insights into the electricity market in Germany, its complex price structure, and the challenges for German compute centers to act successfully on that market.

An introduction from the embedded field to Dark Silicon was given by Sri Parameswaran from the University of New South Wales. The continuous decrease in feature size without an appropriate decrease in the threshold voltage leads to increased power density. Between 50% and 90% of dark silicon is expected in future chips. Mitigation techniques are energy reduction techniques as well as spatial and temporal dimming of cores. Considerable energy reduction can be achieved from heterogeneity on various levels, e.g., heterogeneous cores and the DarkNoC approach.

Dark Silicon due to Power Density

Several techniques were presented to mitigate the effect of power density. Santiago Pagani presented *spatial and temporal dimming of cores* to make best use of the thermal distribution on the chip. He and Andrey Semin talked also about *boosting* the core frequency to exceed the power limit for a short time period to speedup computation. Sergio Bampi presented *near threshold computing* as a potential solution based on further lowering the threshold voltage. Michael Niemier explored the potential of *new transistor technology* to mitigate the Dark Silicon effect.

Dark Silicon due to Limited Power

Mitigation techniques in this field are quite similar in mobile computing and HPC, although the overall objective is a bit different. While in mobile computing the minimal power required to meet the QoS requirements of applications is the goal, in HPC it is to go as fast as possible with the available power, may be considering energy efficiency and system throughput as well.

The following approaches relevant for mobile computing and HPC were presented: *Heterogeneity* in various hardware aspects can be used to reduce the energy consumption of computations. Siddarth Garg and Tulika Mitra covered *performance heterogeneity* in scheduling tasks for big/little core combinations. Tulika Mitra and Andrea Bartolini talked about using *function heterogeneity*, e.g. accelerators, in mobile computing and HPC to increase energy efficiency. The *Heterogeneous Tile Architecture* was introduced in the presentations of Sri Parameswaran and Santiago Pagani as a general architecture enabling exploitation of heterogeneity to mitigate the Dark Silicon effect.

Another approach is to determine the most efficient *application and system configuration*. *Static tuning* of parameters, such as the power budget of an application, were presented by Michael Knobloch and Tapasya Patki. *Dynamic tuning* techniques were covered in the presentations of Michael Gerndt, Martin Schulz, and Per Gunnar Kjeldsberg. Jonathan Eastep introduced the GEO run-time infrastructure for distributed machine-learning based power and performance management.

Kirk Cameron highlighted the unexpected effects of changing the core frequency due to non-linear dependencies. Jürgen Teich talked about *Invasive Computing* providing dynamic resource management not only for improving certain non-functional application aspects but also for increasing the predictability of those aspects.

Wolfgang Nagel and Sri Parameswaran presented *energy efficient network architectures*. They covered heterogeneous on-chip network architectures and wireless communication within compute clusters.

Approximate computing was presented by Sergio Bampi. It allows trading off accuracy and energy. Pietro Cicotti covered in his presentation *data movement optimization* within a CPU to save energy.

Application and system monitoring is a pre-requisite for many of the above techniques. Michael Knobloch, Wolfgang Nagel, and Kathleen Shoga presented application and system monitoring techniques based on software as well as hardware instrumentation. Many compute centers are installing infrastructures to gather sensor values from the whole facility to enable future analysis. In addition to performance and energy measurements for application, higher level information about the application characteristics is useful in taking tuning decisions. Tapasay Patki presented *application workflows* as a mean to gather such information.

Besides these generally applicable techniques, some presentations covered also techniques that are specific to HPC installations with their batch processing approach and large compute systems.

Andrea Bartolini highlighted in his presentation the holistic multiscale aspect of power-limited HPC. The application, the compute system, and the *cooling infrastructure* have to be seen as a complex integrated system. *Power-aware scheduling*, presented by Tapasay Patki and Andrea Bartolini, can significantly improve the throughput of power-limit HPC systems and *moldable jobs* can improve the effect of power-aware scheduling significantly. Isaias Compres presented *Invasive MPI*, an extension of MPI for programming moldable application.

Conclusion

At the end of the seminar a list of takeaway messages was collected based on working-group discussions followed by an extensive discussion of all participants:

1. Dark silicon is a thermal problem in embedded and a power problem in HPC. HPC can cool down while in the embedded world you can't. Therefore HPC can power up everything if they have enough power. But the costs for providing enough power for rare use cases have to be rectified.
2. Better tools are required on both sides to understand and optimize applications.
3. Better support for optimizations is required through the whole stack from high level languages down to the hardware.
4. In both communities run-time systems will get more important. Applications will have to be written in a way that run-time systems can work effectively.
5. Task migration is of interest to both groups in combination with appropriate run-time management techniques.
6. Embedded also looks at specialized hardware designs while HPC has to use COTS. In HPC, the machine architecture might be tailored towards the application areas. Centers are specialized for certain customers.
7. Heterogeneity on architecture level is important to both groups for energy reduction.

8. Better analyzable programming models are required, providing composable performance models.
9. HPC will have to live with variability. The whole tuning step has to change since reproducibility will no longer be given.
10. Hardware-software co-design will get more important for both groups.
11. Both areas will see accelerator-rich architectures. Some silicon has to be switched off anyway, thus these can be accelerators that might not be useful for the current applications.

2 Table of Contents

Executive Summary

Hans Michael Gerndt, Michael Glaß, Sri Parameswaran, and Barry L. Rountree . 224

Overview of Talks

Beyond Power Capping – Coping with the Complexity of the German Electricity Market <i>Axel Auweter</i>	233
Dark Power: Applying Lesson from Dark Silicon to Power-Constrained High Performance Computing <i>Barry Rountree</i>	233
Multiscale Energy-Thermal Management for Green Supercomputers <i>Andrea Bartolini</i>	233
The Law of Unintended Consequences for Dark Silicon <i>Kirk W. Cameron</i>	234
Data Movement in Dark Silicon Systems <i>Pietro Cicotti</i>	235
Elastic Execution Models and Energy Aware Job Scheduling in HPC <i>Isaias Alberto Compres Urena</i>	235
An Introduction to GEO: A New Open Source Extensible Power Management Framework from Intel <i>Jonathan Eastep</i>	235
Scheduling for Dark Silicon Servers <i>Siddharth Garg</i>	236
Energy Efficiency Tuning: From Autotune to READEX <i>Hans Michael Gerndt</i>	236
Scenario based design of dynamic embedded applications <i>Per Gunnar Kjeldsberg</i>	237
Energy-efficient HPC – A Tools Perspective <i>Michael Knobloch</i>	237
The impact of new transistor technologies on core scaling trends (and dark silicon) <i>Michael Niemier</i>	238
Improving Energy-Efficiency through Heterogeneity in Mobile Platforms <i>Tulika Mitra</i>	239
Workflow Analysis – A map between Applications and System Resource Needs <i>David Montoya</i>	239
Performance, Energy, Structure, and Materials: What we have to learn, and how we will address the Challenges! <i>Wolfgang E. Nagel</i>	240
Mitigating the Power Density and Temperature Problems in the Nano-Era <i>Santiago Pagani</i>	240

System-Wide Power Management in High-Performance Computing <i>Tapasya Patki</i>	241
System Software for Power Limited HPC Systems: Challenges and Solutions <i>Martin Schulz</i>	241
Processor power and performance variability and impact on applications performance <i>Andrey Semin</i>	242
Livermore Computing Monitoring Infrastructure <i>Kathleen Sumiko Shoga</i>	242
Introduction to Dark Silicon – Problems and Techniques <i>Sri Parameswaran</i>	242
Adaptive Restriction and Isolation for Increasing *-Predictability <i>Jürgen Teich</i>	243
Participants	244

3 Overview of Talks

3.1 Beyond Power Capping – Coping with the Complexity of the German Electricity Market

Axel Auweter (LRZ – München, DE)

License  Creative Commons BY 3.0 Unported license
© Axel Auweter

Germany is one of the countries with the highest costs for electricity in the world. On top, the regulations and pricing scheme is overly complex. Yet, the availability of power saving and capping techniques, intelligent resource management and power consumption prediction models in HPC opens up the possibility for leveraging this complexity in smart ways. This presentation explains the pricing of electricity in Germany and how current and future developments from the EEHPC domain might help optimize the TCO for German HPC centers.

3.2 Dark Power: Applying Lesson from Dark Silicon to Power-Constrained High Performance Computing

Barry Rountree (LLNL – Livermore, US)

License  Creative Commons BY 3.0 Unported license
© Barry Rountree

The field of high performance computing is experiencing a sea change: where previous machines were limited by the number of compute nodes that could be purchased, future exascale machines will be primarily limited by the amount of power than can be brought into the center. The US Department of Energy has a target for the first exaflop machine to consume no more than 20 megawatts. Compared to early petaflop machines, this effectively means a 1000 x increase in performance for an 3x increase in power.

This change calls into question how we think about performance. If power is going to be the limiting factor, then we should be maximizing its utilization. Current HPC codes make use of only 60 % of allocated power. Scaling these codes up naively to exascale and beyond implies that 40 % of the electrical infrastructure would remain idle for most of the lifetime of the machine. In short, we need a new model for machine design and evaluation that uses all available power to maximize job performance and system throughput.

3.3 Multiscale Energy-Thermal Management for Green Supercomputers

Andrea Bartolini (University of Bologna, IT & ETH Zürich, CH)

License  Creative Commons BY 3.0 Unported license
© Andrea Bartolini

In the last decade large high performance computing systems as well as processing elements have become power and energy limited. At system scale energy provisioning and cooling power and facility design limit the available power budget for each machine installation while

at component scale the end of Dennard’s scaling makes the power consumption the limiting factor for the performance of the computing devices. Today’s processors performances are thermally and power limited, while today’s supercomputers performance are power, cooling and cost limited. In this talk I will present a set of tools, methodology and research results on the evaluation of the impact of temperature on the energy-efficiency of the supercomputer and internal components and opportunities for advanced and holistic management of thermally constrained large scale computing systems.

3.4 The Law of Unintended Consequences for Dark Silicon

Kirk W. Cameron (Virginia Polytechnic Institute – Blacksburg, US)

License © Creative Commons BY 3.0 Unported license
© Kirk W. Cameron

Joint work of Hung-Ching Chang, Bo Li, Godmar Back, Ali Raza Butt, Kirk W. Cameron
Main reference H.-C. Chang, B. Li, G. Back, A. R. Butt, K. W. Cameron, “LUC: Limiting the Unintended Consequences of Power Scaling on Parallel Transaction-Oriented Workloads”, in Proc. of the 2015 IEEE Int’l Parallel and Distributed Processing Symp. (IPDPS’15), pp. 324–333, IEEE, 2015.
URL <http://dx.doi.org/10.1109/IPDPS.2015.99>

In 1936, Harvard University sociologist Robert Morton wrote a paper entitled “The unanticipated consequences of purposive social action”, where he described how government policies often result in both positive and negative unintended consequences. The lesson from Morton’s work was that unexpected consequences in complex social systems, at the time relegated to theology or chance, should be evaluated scientifically.

Since the performance effects of dark silicon management are largely unknown, these potentially valuable features introduce risk and uncertainty in complex, large-scale, high-performance systems. While dark silicon promises to address power and energy limitations for emergent systems, Morton teaches us that relegating performance behavior to chance is just as likely to result in negative consequences. For example, there is mounting evidence that when processors are fixed at fully-powered, highest frequency (i.e., disabling dynamic frequency scaling), performance can worsen. Thus, challenges and opportunities abound for dark silicon to be adopted by the HPC community.

In this presentation, I will demonstrate that “faster is NOT always better” when managing power and performance of large scale systems. In essence, slowing down CPU frequency (or powering down system components) can speed up performance as much as 50 % for some I/O intensive applications. I show that identifying the root cause of such slowdowns is wrought with challenges. I will describe how modeling and runtime systems can limit these anomalies but that dark silicon will undoubtedly lead to more unintended consequences for high-performance systems.

References

- 1 *LUC: Limiting the Unintended Consequences of Power Scaling on Parallel Transaction-Oriented Workloads*. IPDPS 2015: 324-333, Hyderabad, India, 2015.

3.5 Data Movement in Dark Silicon Systems

Pietro Cicotti (San Diego Supercomputer Center, US)

License © Creative Commons BY 3.0 Unported license
© Pietro Cicotti

Joint work of Laura Carrington, Pietro Cicotti

As power limitations induce the presence of dark silicon, a new dimension appears in the configuration and optimization space of applications and systems. In order to optimize performance and efficiency, execution must be combined with an understanding of available resources (darkened and not) and potential gain/cost tradeoffs in using them.

A fundamental aspect of this optimization problem is associated with the need to move data. Leveraging dark silicon implies that data must be moved to powered resources, and then fetched back. For example, in computational sprinting and invasive computing, data must move to the claimed resources and then be flushed back when the resources are released. In addition, with the ability to finely select and configure the resources claimed, it is important to correctly estimate and select an optimal configuration of resources.

In this context, carefully tuning systems and applications requires understanding data access patterns and the effect of different memory hierarchies. In this presentation, I will discuss our work in modeling memory and creating tools to analyze and eventually manage data movement dynamically.

3.6 Elastic Execution Models and Energy Aware Job Scheduling in HPC

Isaias Alberto Compres Urena (TU München, DE)

License © Creative Commons BY 3.0 Unported license
© Isaias Alberto Compres Urena

Power density has increased in recent computing Integrated Circuits (IC), while computer hardware designs have maintained largely the same heat dissipation properties. This situation has led to Dark Silicon scenarios, where large parts of an IC must remain powered off to keep it in safe operating levels. Parallels can be drawn in HPC systems, where power limits require nodes or partitions to be turned off. Elastic execution models allow distributed memory applications to adapt to changes in resources. Job schedulers can manipulate such applications individually to achieve global energy requirements such as power level stabilization.

3.7 An Introduction to GEO: A New Open Source Extensible Power Management Framework from Intel

Jonathan Eastep (Intel – Hillsboro, US)

License © Creative Commons BY 3.0 Unported license
© Jonathan Eastep

URL <http://geopm.github.io/geopm>

In this talk, I will provide an intro to GEO (Global Energy Optimization). GEO is an open source, scalable, extensible runtime system and power management framework for HPC

systems from Intel. It provides out-of-the-box power management technology to mitigate application load imbalance by redistributing power to the application’s critical path, and it provides extensibility to new power management strategies through a plug-in architecture. A goal of the project is to provide a convenient platform that HPC power researchers can build their research on and accelerate innovation in HPC power management. The runtime as well as plug-ins are licensed with a permissive BSD license to encourage community and industry adoption and collaboration. See geopm.github.io/geopm for more project information and a link to the source code.

3.8 Scheduling for Dark Silicon Servers

Siddharth Garg (New York University, US)

License  Creative Commons BY 3.0 Unported license
© Siddharth Garg

Joint work of Siddharth Garg, Umit Ogras

Heterogeneous processors with multiple core types, for example, the so-called “big-little” processors, are becoming increasingly common-place. This talk will focus on energy and thermally-aware scheduling for heterogeneous servers; in particular, we will discuss a family of so-called “threshold” policies that preferentially assign jobs to power efficient cores and utilize larger cores only when the number of outstanding jobs exceeds a threshold. We will also discuss policies for parallelizable jobs.

3.9 Energy Efficiency Tuning: From Autotune to READEX

Hans Michael Gerndt (TU München, DE)

License  Creative Commons BY 3.0 Unported license
© Hans Michael Gerndt

The European AutoTune project developed the Periscope Tuning Framework (<http://periscope.in.tum.de>) for pre-production tuning of HPC applications [1]. Tuning plugins capture expert knowledge for a certain tuning aspect and search for optimal tuning parameter settings. Periscope provides a rich framework to tuning plugins, e.g., standard search algorithms, performance analysis services, static program information, and automatic experiment executions. Plugins use expert knowledge to structure the search process and to reduce the search space. Tuning plugins can use performance analysis information to, for example, determine the size distribution of messages and using this information to restrict the range of values for the eager threshold of the MPI library. Standard search algorithms are used to finally generate scenarios that are experimentally evaluated. The scenario is specified by the plugin and the real execution and measurement of the objective function is done automatically by Periscope.

The focus of automatic tuning in the AutoTune project was on design time tuning. The best setting for the tuning parameters is determined before production runs of the application. This best setting is then static for the execution. In the new Horizon 2020 project READEX (<http://www.readex.eu>) this work is extended for runtime tuning [2]. It follows the approach of scenario-based optimization from embedded systems. At design time the application is analyzed and a tuning model is constructed that captures the best configurations for various

runtime situations. This tuning model is passed to the runtime and dynamic switching happens between configuration if new runtime situations are encountered. The Periscope Tuning Framework is used for the design time analysis and Score-P is extended with the READEX Runtime Library for dynamic configuration switching.

References

- 1 Michael Gerndt, Eduardo César and Siegfried Benkner (Eds.). *Automatic Tuning of HPC Applications – The Periscope Tuning Framework*. Shaker Verlag, ISBN 978-3-8440-3517-9, 2015
- 2 Y. Oleynik, M. Gerndt, J. Schuchart, P. G. Kjeldsberg, W. E. Nagel. *Run-Time Exploitation of Application Dynamism for Energy-Efficient Exascale Computing (READEX)*. IEEE 18th International Conference on Computational Science and Engineering (CSE), pp. 347–350, 2015

3.10 Scenario based design of dynamic embedded applications

Per Gunnar Kjeldsberg (NTNU – Trondheim, NO)

License © Creative Commons BY 3.0 Unported license
© Per Gunnar Kjeldsberg

Joint work of Special Interest Group on Scenario Driven Design for Embedded Systems

URL <http://www.es.ele.tue.nl/scenarios/>

System scenario methodologies propose the use of different scenarios, e.g., different heterogeneous platform configurations, in order to exploit variations in computational and memory needs during the lifetime of an application. The system scenario methodology consists of a design-time and a run-time stage. The application is analyzed at design-time and different execution paths and variations in processing and memory demands are identified. Situations with similar N-dimensional Pareto cost requirements are grouped into a limited number of scenarios. At run-time the current situation is detected, and the platform is reconfigured accordingly, e.g., through remapping of tasks on processors, voltage and frequency scaling, changing power modes of memories, turning on and off processing cores and accelerators, etc. Compared with use-case scenarios, system scenarios exploit detailed knowledge of the application, giving rise to much larger performance and energy gains.

This talk will present the system scenario design methodology including results from implementation examples in the embedded systems domain.

3.11 Energy-efficient HPC – A Tools Perspective

Michael Knobloch (Jülich Supercomputing Centre, DE)

License © Creative Commons BY 3.0 Unported license
© Michael Knobloch

Main reference R. Schöne, J. Treibig, M. F. Dolz, C. Guillen, C. B. Navarrete, M. Knobloch, B. Rountree, “Tools and methods for measuring and tuning the energy efficiency of HPC systems”, *Scientific Programming*, 22(4):273–283, 2014.

URL <http://dx.doi.org/10.3233/SPR-140393>

Energy consumption of applications and power draw of large-scale installations has become a major topic in HPC on the road to Exascale. A detailed analysis of power and energy consumption is necessary in order to understand system and application characteristics and

control them for maximum efficiency. However, traditional performance analysis tools face multiple challenges obtaining power and energy relevant data. In this talk I present the work done by JSC and its partners in multiple energy-efficiency related projects and discuss the requirements on hardware in order to improve power and energy consumption analysis.

3.12 The impact of new transistor technologies on core scaling trends (and dark silicon)

Michael Niemier

License  Creative Commons BY 3.0 Unported license

© Michael Niemier

Joint work of Michael Niemier, Robert Perricone, X. Sharon Hu, Joseph Nahas

Continued transistor scaling no longer yields exponential performance gains due in part to the growth of dark silicon (DS). Both industrial and government sponsors are actively pursuing the development of new transistor technologies that may re-enable voltage scaling, offer I-V characteristics that lead to simpler and/or more efficient circuits when compared to CMOS functional equivalents, etc.

This talk will discuss architectural-level modeling efforts that build upon the framework developed in [1, 2], which originally considered DS in the context of core scaling efforts via the PARSEC benchmark suite [3]. The Notre Dame group has worked to unify architectural-level benchmarking [1, 2] with device-level benchmarking [4, 5] (that considers devices being studied under the umbrellas of various SRC and DARPA initiatives) to provide insight as to how voltage scaling could impact the viability of core scaling – and hence the spread of DS.

Interestingly, for high thermal design power TDPs (125 W), projections suggest that low voltage devices achieve a speedup of just 2X on average in the best case when compared to 15 nm high performance (HP) CMOS. Moreover, per [5] the 15 nm CMOS datapoint is representative of 2018 technology – which will undoubtedly come to market well-before emerging low voltage devices (which may presently exist in simulation only). Not surprisingly, for lower TDPs (5W), emerging low voltage devices fare much better when compared to HP CMOS – and speedups of approximately 10X appear possible. However, speedups of approximately 2.5X are projected over 2018 low power (LP) CMOS. In addition to these results, other benchmarking data and possible paths forward will be highlighted.

References

- 1 H. Esmaeilzadeh, E. Blem, R. St. Amant, K. Sankaralingam, and D. Burger, “Dark silicon and the end of multicore scaling,” in *Computer Architecture (ISCA)*, 2011 38th Annual International Symposium on, June 2011, pp. 365–376.
- 2 H. Esmaeilzadeh, E. Blem, R. S. Amant, K. Sankaralingam, and D. Burger, “Power challenges may end the multicore era,” *Commun. ACM*, vol. 56, no. 2, pp. 93102, Feb. 2013.
- 3 C. Bienia, S. Kumar, J. P. Singh, and K. Li, “The parsec benchmark suite: Characterization and architectural implications,” in *Proceedings of the 17th International Conference on Parallel Architectures and Compilation Techniques*, ser. PACT’08. New York, NY, USA: ACM, 2008, pp. 72–81.
- 4 D. Nikonov and I. Young, “Overview of beyond-cmos devices and a uniform methodology for their benchmarking,” *Proceedings of the IEEE*, vol. 101, no. 12, pp. 2498–2533, 2013.
- 5 D. Nikonov and I. Young, “Benchmarking of beyond-cmos exploratory devices for logic integrated circuits,” *Exploratory Solid-State Computational Devices and Circuits*, *IEEE Journal on*, vol. 1, pp. 3–11, Dec 2015.

3.13 Improving Energy-Efficiency through Heterogeneity in Mobile Platforms

Tulika Mitra (National University of Singapore, SG)

License © Creative Commons BY 3.0 Unported license
© Tulika Mitra

Joint work of Tulika Mitra, Thannirmalai Somu Muthukaruppan, Anuj Pathania, Alok Prakash, Mihai Pricopi, Vanchinathan Venkataramani, Sanjay Vishin

Main reference T. S. Muthukaruppan, A. Pathania, T. Mitra, “Price theory based power management for heterogeneous multi-cores”, in Proc. of the 19th ACM Int’l Conf. on Architectural Support for Programming Languages and Operating Systems (ASPLOS’14), pp. 161–176, ACM, 2014.

URL <http://dx.doi.org/10.1145/2541940.2541974>

The impact of dark silicon is more pronounced in the mobile platforms due to the absence of active cooling in these systems. In order to cope with the effect of dark silicon, mobile system-on-chip designs embrace heterogeneous multi-core architectures where cores with different functional characteristics (CPU, GPU, DSP, non-programmable accelerators) and/or power-performance characteristics (simple versus complex micro-architecture) co-exist on the same die. Given an application, only the cores that best fit the application can be switched on leading to faster and energy efficient computing. We present application-aware, software-level runtime management strategies to leverage the potential of heterogeneous multi-core architectures.

3.14 Workflow Analysis – A map between Applications and System Resource Needs

David Montoya (Los Alamos National Lab., US)

License © Creative Commons BY 3.0 Unported license
© David Montoya

Workflow has always been used to describe jobs and applications progressing and interacting with systems. As systems become more tightly integrated with varied architectures and with varied feedback loops added to better balance resource utilization, do we have a map that includes the application? When you envision the overall HPC environment being made up of applications and system components that are made up of workflows interacting with each other, it becomes apparent that we don’t have the tools to assess this interaction.

In this presentation I will describe an effort at LANL where we have started by developing a workflow taxonomy with layers describing the application stack, how we have used it for initial assessment for future machine needs, and the potential to further define lower layers to integrate with mapping of machine layers of workflow. As this evolves it brings in application and system performance collection, deriving workflow performance, and system monitoring as key initial capabilities.

3.15 Performance, Energy, Structure, and Materials: What we have to learn, and how we will address the Challenges!

Wolfgang E. Nagel (TU Dresden, DE)

License © Creative Commons BY 3.0 Unported license
© Wolfgang E. Nagel

Parallelism on chips and technology improvements nowadays have led to dark silicon, power budgets, and temperature and energy variations, which heavily depend on hardware features and usage profiles of the applications. Running thousands of these sockets in parallel lead to reasonable challenges not only in the field of load balancing, but also in the energy usage. The talk describes research work of the Dresden group in the field of energy measurement on the microsecond level, embedded in the collaborative research center HAEC (highly adaptive energy-efficient computing). In HAEC, research technologies are developed to enable computing systems with high energy-efficiency without compromising on high performance. As part of that, a novel concept (HAEC Box) of how computers can be built by utilizing innovative ideas of optical and wireless chip-to-chip communication is explored. This CRC is embedded in the excellence cluster cfAED (Center for Advanced Electronics Dresden) where also material research is done to shape the time after CMOS. The talk will describe the general approach and the implications on programming challenges in future systems.

3.16 Mitigating the Power Density and Temperature Problems in the Nano-Era

Santiago Pagani (KIT – Karlsruher Institut für Technologie, DE)

License © Creative Commons BY 3.0 Unported license
© Santiago Pagani

Joint work of Santiago Pagani, Heba Khdr, Waqaas Munawar, Dennis Gnad, Muhammad Shafique, Siddharth Garg, Minming Li, Jian-Jia Chen, Jörg Henkel

Main reference S. Pagani, H. Khdr, W. Munawar, J.-J. Chen, M. Shafique, M. Li, J. Henkel, “TSP: Thermal Safe Power – Efficient power budgeting for many-core systems in dark silicon”, in Proc. of the 2014 Int’l Conf. on Hardware/Software Codesign and System Synthesis (CODES’14), Art. 10, ACM, 2014.

URL <http://dx.doi.org/10.1145/2656075.2656103>

In this talk we introduce the Dark Silicon problem and discuss our research efforts for mitigating the associated power density and temperature issues. Specifically, we talk about mapping/patterning and how making smart mapping decisions can reduce the peak temperature on the chip. We introduce the Thermal Safe Power (TSP) concept for efficient power budgeting, in which the power budget depends on the number of active cores. We present some of our experiments comparing single/constant frequency solutions against boosting techniques. Finally, we discuss MatEx, an efficient analytical transient and peak temperature computation tool. In conclusion, power and performance efficiency should be jointly optimized at multiple hardware and software layers of the system stack. If all this is considered, there is a good chance that the Dark Silicon problem can be avoided.

3.17 System-Wide Power Management in High-Performance Computing

Tapasya Patki (LLNL – Livermore, US)

License © Creative Commons BY 3.0 Unported license
© Tapasya Patki

Main reference T. Patki, D. K. Lowenthal, A. Sasidharan, M. Maiterth, B. Rountree, M. Schulz, B. R. de Supinski, “Practical Resource Management in Power-Constrained, High Performance Computing”, in Proc. of the 24th Int’l Symposium on High-Performance Parallel and Distributed Computing (HPDC’15), pp. 121–132, ACM, 2015.

URL <http://dx.doi.org/10.1145/2749246.2749262>

One of the key challenges on the path to exascale supercomputing is power management. Supercomputing centers today are designed to be worst-case power provisioned, leading to two main problems: limited application performance and under-utilization of procured power. This talk will introduce hardware overprovisioning: a power-efficient design approach for future supercomputing centers that addresses the aforementioned problems. Power-aware resource management policies targeted toward overprovisioned HPC systems will also be discussed.

3.18 System Software for Power Limited HPC Systems: Challenges and Solutions

Martin Schulz (LLNL – Livermore, US)

License © Creative Commons BY 3.0 Unported license
© Martin Schulz

Power and energy consumption are critical design factors for any next generation large-scale HPC system. The costs for energy are shifting the budgets from investment to operating costs, and more and more often the size of systems will be determined by its power needs. As a consequence, it is likely that we will end up with power limited systems that can no longer power all their components at peak power. In these systems, system software must manage power caps at all layers of the system to ensure only the available power is used in the system and that this available power is used efficiently. In this talk, I will discuss the need and opportunities of power-limited systems and the challenges they pose and present system software techniques at different levels of the software stack that can help users successfully and efficiently exploit power-limited systems. In particular, I will present an approach to mitigate processor manufacturing variability at large scale [1], a runtime system – Conductor [2] – to steer power within an MPI application in order to maximize the utilization of an available power budget, and an operating system component – PowSched [3] – to exploit unused power resources in power constrained HPC systems.

References

- 1 Y. Inadomi, T. Patki, K. Inoue, M. Aoyagi, B. Rountree, M. Schulz, D. Lowenthal, Y. Wada, K. Fukazawa, M. Ueda, M. Kondo, and I. Miyoshi. Analyzing and mitigating the impact of manufacturing variability in power-constrained supercomputing. In *Proce. of the Int’l Conf. for High Performance Computing, Networking, Storage and Analysis, SC’15*, pp. 78:1–78:12, USA, 2015. ACM.
- 2 A. Marathe, P. E. Bailey, D. K. Lowenthal, B. Rountree, M. Schulz, and B. R. Supinski. *High Performance Computing: 30th Int’l Confe., ISC High Performance 2015, Frank-*

furt, Germany, July 12-16, 2015, Proceedings, chapter A Run-Time System for Power-Constrained HPC Applications, pp. 394–408. Springer Int’l Publishing, Cham, 2015.

- 3 D. A. Ellsworth, A. D. Malony, B. Rountree, and M. Schulz. Dynamic power sharing for higher job throughput. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC’15*, pp. 80:1–80:11, USA, 2015. ACM.

3.19 Processor power and performance variability and impact on applications performance

Andrey Semin (Intel GmbH – Feldkirchen, DE)

License  Creative Commons BY 3.0 Unported license
© Andrey Semin

Modern Intel microprocessors contain embedded controller called PCU (power control unit) that is in full control of processor execution state and mode of operation. Many power management and performance-related features are implemented with the use of this controller logic. At the same time some of PCU control operations presents challenges for the performance and parallel scaling of the HPC applications, specifically in the area of performance reproducibility. One of the specific challenges is that PCU makes CPU frequency dependent on consumed power, while we note that power is dependent on frequency, voltage, as well as temperature of the circuit. The observed variability in performance and power is noted by many HPC system users, and these observations are summarized in this presentation. In the conclusion we propose an “uncertainty principle” that governs power and frequency (or the cycle time) variability dependencies.

3.20 Livermore Computing Monitoring Infrastructure

Kathleen Sumiko Shoga (LLNL – Livermore, US)

License  Creative Commons BY 3.0 Unported license
© Kathleen Sumiko Shoga

Power, energy, and thermal constraints require us to make smarter use of our resources to get the best performance. There are, however, many factors that go into the performance of applications run in a large computing center. At Livermore Computing, we are deploying monitoring across the center in a multi-level fashion from the facilities level down to the hardware performance counter level. Gathering and analyzing this data will enable us to make better choices when it comes to tradeoffs for resources and future system designs.

3.21 Introduction to Dark Silicon – Problems and Techniques

Sri Parameswaran

License  Creative Commons BY 3.0 Unported license
© Sri Parameswaran

Joint work of Muhammad Shafique, Haseeb Bokhari, Joerg Henkel, Sri Parameswaran

In this talk we discuss aspects of dark silicon, starting with the definition of dark silicon, its origins and the reason why modern chips are becoming larger, yet are unable to be powered

on completely. This problem is exacerbated in embedded systems, where cooling is limited and thus only a fraction of the chip can be turned on at any one time. The second part of the talk explains some of the methods used to overcome the issues arising from dark silicon, and explains some of the opportunities afforded to designers and users of modern chips. Finally, we delve deeper in to one of the methods used to mitigate the problems of dark silicon and talk about creating a NoC which utilizes the area afforded by dark silicon to improve reliability and energy efficiency.

3.22 Adaptive Restriction and Isolation for Increasing *-Predictability

Jürgen Teich (Universität Erlangen-Nürnberg, DE)

License  Creative Commons BY 3.0 Unported license
© Jürgen Teich

Resource sharing and interferences of multiple threads of one, but even worse between multiple application programs running concurrently on a Multi-Processor System-on-a-Chip (MPSoC) today make it very hard to provide any timing or throughput-critical applications with time bounds. Additional interferences result from the interaction of OS functions such as thread multiplexing and scheduling as well as complex resource (e.g., cache) reservation protocols used heavily today. Finally, dynamic power and temperature management on a chip might also throttle down processor speed at arbitrary times leading to additional variations and jitter in execution time. This may be intolerable for many safety-critical applications such as medical imaging or automotive driver assistance systems.

Static solutions to provide the required isolation by allocating distinct resources to safety-critical applications may not be feasible for reasons of cost and due to the lack of efficiency and inflexibility.

In this Dagstuhl presentation, we first review definitions of predictability. We distinguish two techniques for improving predictability called restriction and isolation and present new definitions for predictability. Subsequently, new techniques for adaptive isolation of resources including processor, I/O, memory as well as communication resources on demand on an MPSoC are introduced based on the paradigm of Invasive Computing. In Invasive Computing, a programmer may specify bounds on the execution quality of a program or even segment of a program followed by an invade command that returns a constellation of exclusive resources called a claim that is subsequently used in a by-default non-shared way until being released again by the invader. Through this principle, it becomes possible to isolate applications automatically and in an on-demand manner. In invasive computing, isolation is supported on all levels of hardware and software including an invasive OS. Together with restriction (of input uncertainty), the level of on-demand predictability of program execution qualities may be fundamentally increased.

For a broad class of streaming applications, and a particular demonstration based on a complex object detection application algorithm chain taken from robot vision, we show how jitter-minimized implementations become possible, even for statically unknown arrivals of other concurrent applications.

Participants

- Axel Auweter
LRZ – München, DE
- Sergio Bampi
Federal University of Rio Grande
do Sul, BR
- Andrea Bartolini
University of Bologna, IT &
ETH Zürich, CH
- Kirk W. Cameron
Virginia Polytechnic Institute –
Blacksburg, US
- Pietro Cicotti
San Diego Supercomputer
Center, US
- Isaias Alberto Compres Urena
TU München, DE
- Jonathan Eastep
Intel – Hillsboro, US
- Siddharth Garg
New York University, US
- Hans Michael Gerndt
TU München, DE
- Michael Glaß
Universität
Erlangen-Nürnberg, DE
- Per Gunnar Kjeldsberg
NTNU – Trondheim, NO
- Michael Knobloch
Jülich Supercomputing
Centre, DE
- Tulika Mitra
National University of
Singapore, SG
- David Montoya
Los Alamos National Lab., US
- Wolfgang E. Nagel
TU Dresden, DE
- Michael Niemier
University of Notre Dame, US
- Santiago Pagani
KIT – Karlsruher Institut für
Technologie, DE
- Sri Parameswaran
UNSW – Sydney, AU
- Tapasya Patki
LLNL – Livermore, US
- Barry L. Rountree
LLNL – Livermore, US
- Martin Schulz
LLNL – Livermore, US
- Andrey Semin
Intel GmbH – Feldkirchen, DE
- Kathleen Shoga
LLNL – Livermore, US
- Jürgen Teich
Universität
Erlangen-Nürnberg, DE

