





Helsinki Institute for Information Technology HIIT Annual Report 2014

Helsinki Institute for Information Technology HIIT Tietotekniikan tutkimuslaitos HIIT Forskningsinstitutet för Informationsteknologi HIIT

Helsinki Institute for Information Technology HIIT is a joint research institute of Aalto University and the University of Helsinki for basic and applied research on information technology. HIIT's mission is to conduct top-level research, seamlessly moving between fundamental methods and technologies to novel applications and their impact on people and society. HIIT is a strategic partnership of the two universities to take Helsinki IT to the world class and keep it there. HIIT's research is interdisciplinary, operating across departments and with industry. The current foci of research are computational modelling and data analysis, and ubiquitous ICT in the modern networked world.

HIIT groups belong to several national Centers of Excellence and co-operate with the information industry and with sciences applying information technology. HIIT works in a multidisciplinary way, with scientists from computer, natural, behavioural and social sciences, as well as from humanities and design. HIIT partners with several international and Finnish companies as well as with universities and research institutions in Europe, North America and Asia.

HIIT was founded in 1999 and now has a budget of over 10 million euros, realising in total about 134 person-years of work by almost 300 researchers. HIIT is located in two different sites in the Helsinki Metropolitan area: at Aalto University's Otaniemi campus and University of Helsinki's Kumpula campus, co-located with several departments at both places. HIIT's operating principle is that it adds to other departments' activities, instead of competing with them. The institute is led by Professor Samuel Kaski until July 2015 and by Professor Petri Myllymäki from August 2015 onwards. HIIT's research is funded by Aalto University, the University of Helsinki, Tekes, the Academy of Finland, the European Union, private companies, and foundations financing Finnish hi-tech research. The research by HIIT has been assessed and deemed excellent by its Scientific Advisory Board in 2008 and 2012; in the Aalto University Research Assessment Exercise in 2009; and in the Evaluation of Research and Doctoral Training at the University of Helsinki in 2011.



www.hiit.fi

Ella Bingham, Maria Lindqvist (eds.)

Helsinki Institute for Information Technology HIIT

Annual Report 2014

Ella Bingham and Maria Lindqvist (eds.)



Contact Information

Helsinki Institute for Information Technology HIIT Tietotekniikan tutkimuslaitos HIIT (in Finnish) Forskningsinstitutet för Informationsteknologi HIIT (in Swedish)

hiit-info@hiit.fi www.hiit.fi

Otaniemi Site

Postal address at Open Innovation House (OIH): Helsinki Institute for Information Technology HIIT PO Box 15600, FI-00076 Aalto, Finland

Street address: Aalto University, Open Innovation House (OIH), Otaniementie 19-21, Espoo Telephone: +358 9 47001

Postal address at Computer Science Building: Helsinki Institute for Information Technology HIIT PO Box 15400, FI-00076 Aalto, Finland

Street address: Computer Science Building, Konemiehentie 2, Espoo Telephone: +358 9 47001

Kumpula Site

Postal address: Helsinki Institute for Information Technology HIIT PO Box 68, FI-00014 University of Helsinki, Finland

Street address: University of Helsinki, Department of Computer Science, Exactum Gustaf Hällströmin katu 2b, Helsinki Telephone: +358 9 1911 Fax: +358 9 191 51120

Unigrafia, Espoo 2015 Copyright ® 2015 HIIT Cover image by Jukka Corander's group Lay-out: Anna Bergman, Maria Lindqvist

Table of Contents

| Review of Year 2014 | 5 |
|---|----|
| Augmented research, health and knowledge work, and other IT | 5 |
| Collaboration | 7 |
| Societal impact | 9 |
| Views for 2015 | 9 |
| Research | 11 |
| Algorithmic Data Analysis (ADA) | 11 |
| Computational Inference (CI) | 20 |
| Distributed and Mobile Cloud Systems (DMC) | 25 |
| Network Society (NS) | 31 |
| HIIT-Wide Focus Area | 35 |
| Administration | 36 |
| Personnel and funding | 36 |
| Board | 36 |
| Scientific Advisory Board (SAB) | 37 |

HIIT in brief

Helsinki Institute for Information Technology HIIT is a joint research institute of Aalto University and the University of Helsinki for basic and applied research on information technology. HIIT's mission is to conduct top-level research, seamlessly moving between fundamental methods and technologies to novel applications and their impact on people and society. HIIT is a strategic partnership of the two universities to take Helsinki IT to the world class and keep it there. HIIT's research is interdisciplinary, operating across departments and with industry. The current foci of research are computational modelling and data analysis, and ubiquitous ICT in the modern networked world.

HIIT groups belong to several national Centers of Excellence and co-operate with the information industry and with sciences applying information technology. HIIT works in a multidisciplinary way, with scientists from computer, natural, behavioural and social sciences, as well as from humanities and design. HIIT partners with several international and Finnish companies as well as with universities and research institutions in Europe, North America and Asia.

HIIT was founded in 1999 and now has a budget of over 10 million euros, realising in total about 134 person-years of work by almost 300 researchers. HIIT is located in two different sites in the Helsinki Metropolitan area: at Aalto University's Otaniemi campus and University of Helsinki's Kumpula campus, co-located with several departments at both places. HIIT's operating principle is that it adds to other departments' activities, instead of competing with them. The institute is led by Professor Samuel Kaski until July 2015 and by Professor Petri Myllymäki from August 2015 onwards. HIIT's research is funded by Aalto University, the University of Helsinki, Tekes, the Academy of Finland, the European Union, private companies, and foundations financing Finnish hi-tech research. The research by HIIT has been assessed and deemed excellent by its Scientific Advisory Board in 2008 and 2012; in the Aalto University Research Assessment Exercise in 2009; and in the Evaluation of Research and Doctoral Training at the University of Helsinki in 2011.

| Total numbers | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
|------------------------------|------|------|------|------|------|------|
| Funding M€ | 9,0 | 12,1 | 10,5 | 11,5 | 10,0 | 10,5 |
| External/competitive funding | 71 % | 83 % | 78 % | 83 % | 79 % | 82 % |
| Person-years | 185 | 175 | 135 | 144 | 154 | 134 |
| Refereed publications | 177 | 222 | 213 | 209 | 236 | 255 |

Review of Year 2014

Augmented research, health and knowledge work, and other IT

Two main goals of HIIT are to (1) conduct high level strategic research bridging the two universities, and (2) coordinate big common ICT issues across departments to form a common "Helsinki ICT". Year 2014 again witnessed success along both lines.

In 2014 HIIT contributed to three Centres of Excellence of the Academy of Finland: Computational Inference (COIN, with key contributions from HIIT's CI programme's groups), Inverse Problems, and Cancer Genetics.

Tekes's strategic research openings are visionary and challenging projects aiming to make serious breakthroughs that will lay the foundation for entirely new areas of business in the future. The first strategic openings were chosen in 2013 when Re:Know, Revolution of Knowledge Work, was one of the two. Re:Know is coordinated by HIIT and the partners are University of Helsinki, Aalto University and Institute for Occupational Health. In 2014, Tekes funded three new strategic openings, two of which HIIT's researchers participate in: Digital Health Revolution and Living Factories. The vision of the Digital Health Revolution initiative is that future healthcare will allow citizens to control and make use of their personal data. The goal of Living Factories is to realise the full potential of Synthetic Biology in Finland. Synthetic Biology is based on the design and engineering of newto-nature biological systems.

Professor Hiroshi Mamitsuka from the Institute for Chemical Research, Kyoto University, Japan, was selected as a FiDiPro Professor funded by Tekes, and he started at HIIT in the beginning of 2015.





Finnish Centre of Excellence in Inverse Problems Research

The Finnish Center of Excellence in Cancer Genetics Research

1 1 1 1

AA AL BAARANAA.













Nominations

Dr Matti Järvisalo got Academy of Finland's 5-year Academy Research Fellow funding for his research on "Decision Procedures for the Polynomial Hierarchy, Boolean Optimization, and Model Counting (PHantoM)". Dr Tuukka Ruotsalo got Academy of Finland's 3-year Postdoctoral Researcher funding for his research on "Multisource Intent-prediction for Interactive Information Retrieval" which is closely tied to HIIT Wide Focus Area. Dr Alexandru Ioan Tomescu also got Academy of Finland's 3-year Postdoctoral Researcher funding for his research where the aim is to provide novel and unifying models and efficient and accurate algorithmic solutions for the multi-assembly of genomic sequences. Dr Antti Ukkonen was awarded the Frontier Prize at the 13th International Symposium on Intelligent Data Analysis (IDA 2014) for his work on estimating the distribution of shortest path lengths in real-world social networks.

Augmented research

Augmented search, research, and knowledge work are the main themes of the HIIT-wide research initiative that is a strategic research focus of HIIT. The project investigates how Human-Computer Interaction and Machine Learning can be combined to a loop that will increase, by an order of magniture, the effectiveness of search and more generally knowledge work. Since 2011, this HIIT-wide focus area builds on HIIT's existing excellences and touches most of HIIT's research topics. The work focuses on two case studies: general augmented research, and computational cumulative biology. The focus area is now in full speed, accompanied with several big multi-group projects: Tekes big strategic opening "Re:Know", EU FP7 funded "MindSee" project and Academy of Finland's "Multivire" project. We have e.g. developed a technique called interactive intent modeling that allows humans to direct exploratory search. The technique has been implemented in a real-world search engine SciNet. The search engine anticipates user's search intents and visualizes them on a novel "Intent Radar" screen.





Collaboration

HIIT collaborates actively both within Helsinki area and internationally, and acts as a link between researchers, companies and public administration.

Collaboration in Helsinki IT

HIIT coordinates common activities of computer science research at Helsinki area in several ways: education, recruitments, and joint initiatives.

In doctoral education, HIIT coordinates **the Helsinki Doctoral Education Network in Information and Communications Technology (HICT)**. HICT is a collaborative doctoral education network hosted jointly by Aalto University and the University of Helsinki, that continues the successful tradition of collaboration in Helsinki area. Altogether 55 supervisors and over 200 students belong to the network that operates at three departments at Aalto University and one department at University of Helsinki.

HIIT also collaborates with Institute for **Molecular Medicine Finland FIMM** in recruiting PhD students to rotate between research groups at FIMM and HIIT. The collaboration is part of the Nordic EMBL Partnership for Molecular Medicine and the call is able to attract a large number of excellent doctoral candidates from all over the world.

Data Science is a big research focus for the whole HIIT, spanning over several application areas. It is a new discipline whose objective is to provide the underlying theory and the necessary tools to cope with the current data revolution. The education in this area has been boosted by several education programmes such as the Master's Programme in Machine learning and Data mining (Madacamia) at Aalto and Data Science study profile at University of Helsinki. For interdisciplinary studies, **Analytics and Data Science** is a new minor topic available for all MSc students at Aalto, jointly coordinated by professor Aristides Gionis from HIIT and professor Pekka Malo at Aalto BIZ. The goal of the minor is to educate students on how to become proficient in making sense of such big data, and how to apply data analysis skills on their own domain of expertise.

Recruitment is an important mean to boost collaboration. HIIT arranges joint Helsinki IT postdoc and senior researcher calls together with several departments in the field of computer science at Aalto University and University of Helsinki. These joint calls leverage on and further enhance the brand of Helsinki as a hub of computer science, and attract a wide range of high-level applicants.

Helsinki Distinguished Lecture Series on Future Information Technology, coordinated by HIIT, continues to attract world-class speakers and a wide audience. Speakers during 2014 were Corinna Cortes, Head of Google Research, New York; Professor Roderick Murray-Smith from University of Glasgow; and Professor Wojciech Szpankowski from Purdue University. The first speaker of 2015 was Mikko Hyppönen, Chief Research Officer of F-Secure.





HIIT Annual Report 2014

Aalto Digi Platform





Center for Science of Information NSF Science and Technology Center







Digi Platform is Aalto University's new collaborative initiative in the field of ICT and digitalization, to maximize Aalto's internal synergies in a nonexclusive manner, and to increase Aalto's external visibility. In the Digi platform, expertise across Aalto is brought together to maximize the potential of IT to boost other sciences, technologies and society. Currently in Aalto the field is very large but scattered, and the platform aims to coordinate and raise the profile of Aalto. HIIT coordinates the Digi Platform, and University of Helsinki participates in the platform via HIIT.

HIIT is active in the ICT SHOK **Digile** (Strategic Centre for Science, Technology and Innovation), shaping the collaboration between universities and companies in Finland. Professor Sasu Tarkoma is the academic coordinator of Digile's research programme Internet of Things, and Professor Petri Myllymäki is the academic coordinator of Data to Intelligence D2I programme. HIIT researchers participate also in the Health field SHOK SalWe.

International collaboration

All HIIT's activities are inherently international, mostly in a bottomup manner. Cooperation with key international research institutes and universities is active: Berkeley (ICSI and UC Berkeley), MIT, Centre for Computational Statistics and Machine Learning (CSML) at UCL, European Bioinformatics Institute EBI, Human Technology Lab (HTLab) at University of Padova, and Waseda University, to name a few.

HIIT coordinates Aalto University's and University of Helsinki's strategic partnership with NSF Science and Technology Center for Science of Information (Sol). The member universities of the Sol center are Purdue (leader), Bryn Mawr, Howard, MIT, Princeton, Stanford, Texas A&M, UC Berkeley, UCSD and Urbana-Champaign. Other strategic partners in Europe are ETH (Zürich, Switzerland) and the LINCS network (Paris, France).

EIT ICT Labs is an European initiative intended to turn Europe into a global leader in ICT innovation. It aims to fulfill this mission by establishing a new type of partnership between leading companies, research centres, and universities in Europe. One of the 5 nodes is Helsinki, and Professor Marko Turpeinen of HIIT is Helsinki Node Director. Several HIIT researchers have research projects with EIT ICT Labs and are thus contributors to the establishment and ramp up of the EIT ICT Labs activities in Helsinki.

HIIT's researcers are active in organizing high-profile international conferences and workshops: the most recent ones are 17th International Conference on Artificial Intelligence and Statistics (AISTATS 2014), IEEE International Conference on Data Mining (ICDM 2014), 13th International Symposium on Intelligent Data Analysis (IDA 2014), and Discovery Science (DS 2014).

HIIT Annual Report 2014

Societal impact

HIIT's researchers and alumni establish approximately one startup company per year, one of the newest ones being Etsimo whose mission is "discovery beyond search".

HIIT aims to boost the competitiveness of the Finnish information industry and information society in the long run by linking scientific research with the long-term high-risk research and development of the information industry, and by contributing to the development of the information society. HIIT both anticipates developments of the information society and influences them. New technologies developed at HIIT are taken into use at the ICT sector or other fields in the commercial and public sector.

HIIT researchers have been active in issues related to public policy having the profile of a vocal and visible participant in the public debate related to various important themes of information society, especially the discussion on emerging social forms of media use, immaterial rights, privacy and trust in the network society.



Views for 2015

A new Board of HIIT was nominated in early 2014, as described in more detail in Chapter 3 of this Annual Report. One of the first tasks of the Board was to open a call for a new Director in Autumn 2014. The selection was made in early 2015, and Professor Petri Myllymäki, a HIIT group leader and programme director, will start as a new Director in August 2015. Professor Samuel Kaski's 5-year term will end in July 2015.











ada

Research

Algorithmic Data Analysis (ADA) Programme

The mission of the Algorithmic Data Analysis research programme at HIIT is to develop useful algorithmic data analysis methods for other sciences and for industry. The work involves both basic research in computer science and applied work on problems arising from applications. The research focuses on the algorithmic and modeling problems of combinatorial pattern matching, data mining, and machine learning. The work is strongly interdisciplinary. Developing new concepts and algorithms is an iterative process consisting of interacting extensively with the application experts, formulating computational concepts, analyzing the properties of the concepts, designing algorithms and analyzing their performance, implementing and experimenting with the algorithms, and applying the results in practice. The main application areas of ADA are in biology, medicine, environmental studies, creativity, social networks, and neuroscience.



Combinatorial Pattern Matching, Professor Esko Ukkonen

The combinatorial pattern-matching group develops combinatorial algorithms and probabilistic modeling techniques for pattern search and synthesis problems in sequential and higher-dimensional data. The team is interested in the basic research of the theoretical aspects of the area as well as in various applications such as genome structure, gene regulation, and information retrieval.

Key publications:

- Emanuele Giaquinta, Kimmo Fredriksson, Szymon Grabowski, Alexandru I Tomescu and Esko Ukkonen. Motif matching using gapped patterns. Theoretical Computer Science. 548, p. 1-13 13 p., 2014
- Leena Salmela, Eric Rivals. LoRDEC: accurate and efficient long read error correction. Bioinformatics, vol 30, no. 24, pp. 3506-3514, 2014.
- Virpi Ahola, Rainer Lehtonen, Panu Somervuo, Leena Salmela, Patrik Koskinen, Pasi Rastas, Niko Valimaki, Lars Paulin, Jouni Kvist, Niklas Wahlberg, Jaakko Tanskanen, Emily A Hornett, Laura C Ferguson, Shiqi Luo, Zijuan Cao, Maaike A de Jong, Anne Duplouy, Olli-Pekka Smolander, Heiko Vogel, Rajiv C McCoy, Kui Qian, Swee Chong Wong, Qin Zhang, Freed Ahmad, Jani K Haukka, Aruj Joshi, Jarkko Salojarvi, Christopher W Wheat, Ewald Grosse-Wilde, Daniel Hughes, Riku Katainen, Esa Pitkanen, Johannes Ylinen, Robert M Waterhouse, Mikko Turunen, Anna Vaharautio, Sami P Ojanen, Alan H Schulman, Minna Taipale, Daniel Lawson, Esko Ukkonen, Veli Mäkinen, Marian R Goldsmith, Liisa Holm, Petri Auvinen, Mikko J Frilander, Ilkka Hanski. The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. Nature Communications, vol 5, 4737, 2014.
- Kimmo Fredriksson, Emanuele Giaquinta. On a compact encoding of the swap automaton. Information Processing Letters, vol 114, no. 7, pp. 392– 396, 2014.
- Domenico Cantone, Simone Faro, Emanuele Giaquinta. Text searching allowing for inversions and translocations of factors. Discrete Applied Mathematics, vol 163, pp. 247-257, 2014.

HIIT Annual Report 2014

ada

Data Mining, Professor Aristides Gionis

The data-mining group focuses on developing novel methods to extract knowledge from data, designing algorithms to summarize large volumes of data efficiently and effectively, and exploring new ways of using the extracted information. Specific areas of interest include: pattern discovery, clustering and outlier detection, graph mining, social-network analysis, analysis of information networks and social-network dynamics, and analysis of smart-city sensor data.

- Hongyu Su, Aristides Gionis, Juho Rousu. Structured Prediction of Network Response. International Conference on Machine Learning, pp. 442-450, 2014.
- Polina Rozenshtein, Nikolaj Tatti, Aristides Gionis. Discovering Dynamic Communities in Interaction Networks. European Conference on Machine Learning and Knowledge Discovery in Databases, ECML PKDD 2014, pp. 678-693, 2014.
- Polina Rozenshtein, Aris Anagnostopoulos, Aristides Gionis, Nikolaj Tatti. Event detection in activity networks. ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2014, pp. 1176-1185, 2014.
- Aristides Gionis, Theodoros Lappas, Konstantinos Pelechrinis, Evimaria Terzi. Customized tour recommendations in urban areas. ACM International Conference on Web Search and Data Mining, WSDM 2014, pp. 313-322, 2014.
- Jussi Korpela, Kai Puolamäki, Aristides Gionis. Confidence bands for time series data. Data Mining and Knowledge Discovery, Vol. 28, nro 5-6, pp. 1530-1553, 2014.
- Esther Galbrun, Aristides Gionis, Nikolaj Tatti. Overlapping community detection in labeled graphs. Data Mining and Knowledge Discovery, Vol. 28, nro 5-6, pp. 1586-1610, 2014.
- Francesco Bonchi, Aristides Gionis, Tamir Tassa. Identity obfuscation in graphs through the information theoretic lens. Information Sciences. Vol. 275, pp. 232-256, 2014.
- Nikolaj Tatti, Fabian Moerchen, Toon Calders. Finding Robust Item sets under Subsampling. ACM Trans. Database Syst. 39(3): 20, 2014.
- Hao Wu, Jilles Vreeken, Nikolaj Tatti, Naren Ramakrishnan. Uncovering the plot: detecting surprising coalitions of entities in multi-relational schemas. Data Min. Knowl. Discov. 28(5-6): 1398-1428, 2014.
- Orestis Kostakis. Classy: fast clustering streams of call-graphs. Data Min. Knowl. Discov. 28(5-6): 1554-1585, 2014.
- Ingmar Weber, Venkata Rama Kiran Garimella. Using Co-Following for Personalized Out-of-Context Twitter Friend Recommendation. ICWSM 2014.





BRANS UN ANY YOU are disguising yourself you precede the waves you kindled over deep with a coat of gold you knew Plato, Bon-Bon

[haba]

Discovery Group: Data Mining and Computational Creativity, Professor Hannu Toivonen

The Discovery group develops novel methods and tools for data mining and computational creativity. Our focus is on algorithmic methods for discovering links and patterns in data, and recently also on their use in creative systems. Computational creativity is interesting on its own right but also as an application area for data mining methods. We work on verbal creativity, i.e., computational poetry and humor, and also on automatic composition of music. Our emphasis is novel methods that minimize the need for manually coded or language-specific knowledge.

- Ravi Kumar, Hannu Toivonen, Jian Pei, Joshua Zhexue Huang, and Xindong Wu (Editors): Proceedings, 14th IEEE International Conference on Data Mining, IEEE Computer Society, Los Alamitos, CA. 2014.
- Fang Zhou, Hannu Toivonen, Ross D King. The Use of Weighted Graphs for Large-Scale Genome Analysis. PLoS One, vol 9, no. 3, 2014.
- Joonas Paalasmaa, Hannu Toivonen, Markku Partinen. Adaptive Heartbeat Modeling for Beat-to-Beat Heart Rate Measurement in Ballistocardiograms. IEEE Journal of Biomedical and Health Informatics, Early access, 2014.
- Anna Kantosalo, Jukka Toivanen, Ping Xiao, Hannu Toivonen. From Isolation to Involvement: Adapting Machine Creativity Software to Support Human-Computer Co-Creation. The Fifth International Conference on Computational Creativity, pp. 1-7, 2014.
- Jukka Toivanen, Oskar Gross, Hannu Toivonen. "The Officer Is Taller Than You, Who Race Yourself!" Using Document Specific Word Associations in Poetry Generation. The Fifth International Conference on Computational Creativity, 2014.
- Jiuyong Li, Jixue Liu, Hannu Toivonen, Kenji Satou, Youqiang Sun, Bingyu Sun. Discovering statistically non-redundant subgroups. Knowledge-Based Systems, vol 67, pp. 315-327, 2014.
- Oskar Gross, Antoine Doucet, Hannu Toivonen. Document summarization based on word associations. The 37th international ACM SIGIR conference on Research and development in information retrieval, pp. 1023-1026, 2014.

ada

Neuroinformatics, Professor Aapo Hyvärinen

Neuroinformatics is widely defined as the cross-fertilization of informationprocessing and mathematical sciences on the one hand, and neural and cognitive sciences on the other.

Our group works on different aspects of neuroinformatics related to machine learning. We model the visual system in the brain by analyzing the statistical structure of the natural input images. We apply machine learning models on neuroimaging data, in particular MEG. We also develop the relevant theory of statistical machine learning, typically unsupervised.

- Michael Gutmann, Valero Laparra, Aapo Hyvärinen, Jesús Malo. Spatio-Chromatic Adaptation via Higher-Order Canonical Correlation Analysis of Natural Images. PLoS One, vol 9, no. 2, 2014.
- Pavan Ramkumar, Lauri Parkkonen, Aapo Hyvärinen. Group-level spatial independent component analysis of Fourier envelopes of resting-state MEG data. NeuroImage, vol 86, pp. 480-491, 2014.
- Jouni Puuronen, Aapo Hyvärinen. A Bayesian inverse solution using independent component analysis. Neural Networks, vol 50, pp. 47-59, 2014.
- Tatsuya Tashiro, Shohei Shimizu, Aapo Hyvärinen, Takashi Washio. ParceLiNGAM: A Causal Ordering Method Robust Against Latent Confounders. Neural Computation, vol 26, no. 1, pp. 57-83, 2014.
- Ilmari Kurki, Jussi Saarinen, Aapo Hyvärinen. Investigating shape perception by classification images. Journal of vision, vol 14, no. 12, 24, 2014.
- Jun-ichiro Hirayama, Takeshi Ogawa, Aapo Hyvärinen. Simultaneous blind separation and clustering of coactivated EEG/MEG sources for analyzing spontaneous brain activity. 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society: (EMBC 2014), pp. 4932-4935, 2014.





Parsimonious Modeling, Dr. Jaakko Hollmén

The research group Parsimonious Modelling develops novel computational data analysis methods and applies these methods on two application fields: cancer genomics and environmental informatics. Parsimonious modeling aims at simple, compact, or sparse models as a result of learning from data in the presence of very little or no a priori information about the modeled problem. Simplicity of the models facilitates understanding of the problem domain by humans.

Both application fields present similar challenges to the data analysis problems: the high dimensionality of observed data and the presence of moderate or large noise levels are both factors that bear fundamental problems for any data analysis. Seeking new areas of application and interfacing the newest application domains with lots of novel types of generated data helps in finding new, unsolved settings of problems.

- Konsta Sirviö, Jaakko Hollmén. Multi-Step Ahead Forecasting of Road Condition Using Least Squares Support Vector Machines. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2014. pp. 661-666, 2014.
- Jesse Read, Jaakko Hollmén. A Deep Interpretation of Classifier Chains. International Symposium on Intelligent Data Analysis, IDA 2014, pp. 251-262, 2014.
- Prem Raj Adhikari, Anze Vavpetic, Jan Kralj, Nada Lavrac, Jaakko Hollmén. Explaining mixture models through semantic pattern mining and banded matrix visualization. International Conference on Discovery Science, DS2014, Springer Berlin Heidelberg, pp. 86-97, 2014.
- Indre Zliobaite, Jaakko Hollmén. Mobile Sensing Data for Urban Mobility Analysis: A Case Study in Preprocessing. CEUR workshop proceedings. pp. 309-314, 2014.
- Indre Zliobaite, Jaakko Hollmén, Heikki Junninen. Regression models tolerant to massively missing data: a case study in solar radiation nowcasting. Atmospheric Measurement Techniques Discussions, Vol. 7, nro 7, pp. 7137-7174, 2014.
- Indré Žliobaité, Jaakko Hollmén. Optimizing regression models for data streams with missing values. Machine Learning, In press.

ada

Kernel Machines, Pattern Analysis and Computational Biology, Professor Juho Rousu

The group develops machine learning methods, models and tools for computational sciences, in particular computational biology. The group focuses on kernel methods and regularized learning for multiple and structured targets, multiple views and ensembles. Applications of interest include network reconstruction (Pitkänen et al, 2014) and labeling, gene functional classification, metabolite identification (MIDAS, Academy of Finland grant, 2013-2017) and synthetic biology (LiF, Tekes strategic opening, 2014-2016). In our most recent approach in metabolite identification from tandem mass spectrometric data, an important bottleneck problem in metabolomics, we combine molecular fragmentation models with multiple kernel learning, thus achieving state-of-the-art predictive performance (Shen et al. 2014). In machine learning, in collaboration with University College London and Laval University, Canada, we have developed new methods for multilabel classification, based on ensemble learning on a collection of random output graphs imposed on the multilabels, and a kernel-based structured output learner as the base classifier (Marchand et al. 2014). Another example of our recent work is structured prediction for predicting subgraphs of a network responding to a stimulus (Su et al. 2014).



- Mario Marchand, Hongyu Su, Emilie Morvant, Juho Rousu, John Shawe-Taylor. Multilabel Structured Output Learning with Random Spanning Trees of Max-Margin Markov Networks. Annual Conference on Neural Information Processing Systems, pp. 873-881, 2014.
- Esa Pitkänen, Paula Jouhten, Jian Hou, Muhammad Syed Fahad, Peter Blomberg, Jana Kludas, Merja Oja, Liisa Holm, Merja Penttilä, Juho Rousu. Comparative Genome-Scale Reconstruction of Gapless Metabolic Networks for Present and Ancestral Species. PLOS Computational Biology, Vol. 10, nro 2, 2014.
- Shen Huibin, Kai Dührkop, Sebastian Böcker, Juho Rousu. Metabolite identification through multiple kernel learning on fragmentation trees. Bioinformatics 30, no. 12, pp.i157-i164, 2014.
- Hongyu Su, Aristides Gionis, Juho Rousu. Structured Prediction of Network Response. International Conference on Machine Learning, pp. 442-450, 2014.



Genome-Scale Algorithms, Professor Veli Mäkinen

We develop algorithms and data structures for the analysis of genomescale data. Such data is abundant due to modern molecular biology measurement techniques like high-throughput sequencing. We are especially interested in applications of compressed data structures, that make it possible to analyse the often highly redundant data within the space of their information content. We also study other scalability aspects like distributed computation/storage around genome-scale data.

- Veli Mäkinen. Editorial: Special Issue on Algorithms for Sequence Analysis and Storage. Algorithms, vol 7, no. 1, pp. 186-187, 2014.
- Veli Mäkinen, Daniel Valenzuela. Recombination-aware alignment of diploid individuals. BMC Genomics, vol 15, no. Suppl 6, pp. S15, 2014.
- Jouni Sirén, Niko Välimäki, Veli Mäkinen. Indexing Graphs for Path Queries with Applications in Genome Research. IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol 11, no. 2, pp. 375-388, 2014.
- Simon Gog, Kalle Karhu, Juha Kärkkäinen, Veli Mäkinen, Niko Välimäki. Multi-pattern matching with bidirectional indexes. Journal of Discrete Algorithms, vol 24, pp. 26-39, 2014.
- Romeo Rizzi, Alexandru I Tomescu, Veli Mäkinen. On the complexity of Minimum Path Cover with Subpath Constraints for multi-assembly. BMC Bioinformatics, vol 15, no. S9, pp. S5, 2014.
- Virpi Ahola, Rainer Lehtonen, Panu Somervuo, Leena Salmela, Patrik Koskinen, Pasi Rastas, Niko Välimaki, Lars Paulin, Jouni Kvist, Niklas Wahlberg, Jaakko Tanskanen, Emily A Hornett, Laura C Ferguson, Shiqi Luo, Zijuan Cao, Maaike A. de Jong, Anne Duplouy, Olli-Pekka Smolander, Heiko Vogel, Rajiv C McCoy, Kui Qian, Wong Swee Chong, Qin Zhang, Freed Ahmad, Jani K Haukka, Aruj Joshi, Jarkko Salojärvi, Christopher W Wheat, Ewald Grosse-Wilde, Daniel Hughes, Riku Katainen, Esa Pitkänen, Johannes Ylinen, Robert M Waterhouse, Mikko Turunen, Anna Vähärautio, Sami P Ojanen, Alan H Schulman, Minna Taipale, Daniel Lawson, Esko Ukkonen, Veli Mäkinen, Marian R Goldsmith, Liisa Holm, Petri Auvinen, Mikko J Frilander, Ilkka Hanski. The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. Nature Communications, vol 5, 4737, 2014.

ada

New Paradigms in Computing, Professor Petteri Kaski

The group performs basic research at the intersection of core computer science (algorithm design and analysis) and discrete mathematics, with an emphasis towards novel techniques and less studied models of computation. We invest substantial effort to high-risk, high-yield research problems of relatively broad theoretical interest, selected on both problem and method driven basis. However, we also aim at rapid publication of more specific, smaller observations. We particularly seek and value solid results with mathematical elegance and simplicity.

The current research themes of the group are threefold. (1) Exploring the interplay between algebraic, combinatorial, and geometric techniques in the design of exact deterministic algorithms. For example, many combinatorial problems can be cast in algebraic form, whereby a nontrivial algebraic algorithm yields a more efficient solution compared with direct combinatorial tools. (2) Restricted models of computation and tradeoffs in resources and/or objectives. For example, one fundamental limitation in modern large-scale distributed systems is the infeasibility of central control. In practice, the system must be operated by a distributed algorithm in which each computational node operates based on the information available in its local neighbourhood only. Assuming this setting, is it possible to achieve globally optimal or near-optimal operation? What is the tradeoff between the available information and the degree of approximation for the optimum? (3) While we are a theory group, we occasionally engage in practical algorithm implementation. Examples include attacks on combinatorial classification problems and applications in computational geometry.



- Andreas Björklund, Petteri Kaski, Łukasz Kowalik: Counting thin subgraphs via packings faster than meet-in-the-middle time. Proc. 25th ACM-SIAM Symposium on Discrete Algorithms (SODA 2014), pp. 594– 603, Society for Industrial and Applied Mathematics, 2014.
- Mika Göös, Juho Hirvonen, Jukka Suomela: Linear-in-∆ lower bounds in the LOCAL model. Proc. 33rd ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC 2014), pp. 86–95, ACM Press, 2014.
- Kustaa Kangas, Mikko Koivisto, Teppo Niinimäki. Learning chordal Markov networks by dynamic programming. Advances in Neural Information Processing Systems 27, NIPS 2014, pp. 2357–2365, 2014.
- Brandon Malone, Kustaa Kangas, Matti Järvisalo, Mikko Koivisto, Petri Myllymäki. Predicting the hardness of learning Bayesian networks. AAAI Conference on Artificial Intelligence. AAAI 2014, Association for the Advancement of Artificial Intelligence, pp. 2460–2466, 2014.



Computational Inference (CI) Programme

The four groups of the CI programme are all members of the Finnish Centre of Excellence in Computational Inference Research (COIN), and the objectives of the programme are closely intertwined with those of COIN.

The main objective of CI is to develop methods for transforming the data produced by the current data revolution into useful information. The key methodology for achieving this goal is statistical and computational inference based on the data. The emphasis is on large data collections and computationally demanding modelling and inference algorithms. Our mission is to push the boundary towards both more complex problems, requiring more structured data models, and towards extremely rapid inference. We address a set of carefully chosen interdisciplinary "grand challenge" -level problems with high societal impact where solving the data intensive problems requires novel methodologies that can only result from combining the expertise of separate subfields.

Our mission brings up four partially overlapping methodological focus areas: Learning of massive data-driven models; Learning from multiple sources; Statistical inference in highly structured stochastic models; and Extreme inference engine.

In applied research our work is also motivated by the big data and ubiquitous computing vision, where adaptivity, context-awareness and personalisation are key enablers. We see that our four methodological research areas support strongly each other, and they all address from a different perspective the key technological problems we face in our future "big data" information society. In our applied research we link our strong basic research work in machine learning and constraint reasoning to wellmotivated applied research activities involving prototype applications and real-world deployments.



Statistical Machine Learning and Bioinformatics, Professor Samuel Kaski

We develop new methods for machine learning, computational inference, and probabilistic modeling. We focus on models for learning from multiple data sources, including multi-view learning and multi-task learning, and methods combining mechanistic models and probabilistic inference. Our primary application areas are computational systems biology and medicine, bioinformatics, user interaction, data visualization, as well as neuroinformatics. Our main lines of work include Bayesian methodologies for data integration and modeling dependencies between multiple data sets with co-occurring samples, model-driven methods for retrieving and visualizing data sets, and Gaussian process models of genomic time series as well as quantitative analysis of sequencing data.



- James C Costello, Laura M Heiser, Elisabeth Georgii, Mehmet Gönen, Michael P Menden, Nicholas J Wang, Mukesh Bansal, Muhammad Ammad-ud-din, Petteri Hintsanen, Suleiman A Khan, John-Patrick Mpindi, Olli Kallioniemi, Antti Honkela, Tero Aittokallio, Krister Wennerberg, NCI DREAM Community, James J Collins, Dan Gallahan, Dinah Singer, Julio Saez-Rodriguez, Samuel Kaski, Joe W Gray, Gustavo Stolovitzky. A community effort to assess and improve drug sensitivity prediction algorithms. Nature Biotechnology, Vol. 32, pp. 1202-1212, 2014.
- Manuel J A Eugster, Tuukka Ruotsalo, Michiel M Spapé, Ilkka Kosunen, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski.
 Predicting term-relevance from brain signals. In Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 425–434, New York, NY, 2014
- Mehmet Gönen, Samuel Kaski. Kernelized Bayesian Matrix Factorization. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 36, nro 10, pp. 2047-2060, 2014.
- Samuel Kaski, Jukka Corander. Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics, JMLR W&CP, Vol. 33, 2014.
- Pekka Marttinen, Matti Pirinen, Antti-Pekka Sarin, Jussi Gillberg, Johannes Kettunen, Ida Surakka, Antti J Kangas, Pasi Soininen, Paul O'Reilly, Marika Kaakinen, Mika Kähönen, Terho Lehtimäki, Mika Ala-Korpela, Olli T Raitakari, Veikko Salomaa, Marjo-Riitta Järvelin, Samuli Ripatti, Samuel Kaski. Assessing multivariate gene-metabolome associations with rare variants using Bayesian reduced rank regression. Bioinformatics, Vol. 30, nro 14, pp. 2026-2034, 2014.
- Sohan Seth, Niko Välimäki, Samuel Kaski, Antti Honkela. Exploration and retrieval of whole-metagenome sequencing samples. Bioinformatics, Vol. 30, nro 17, pp. 2471-2479, 2014.





We do research on theoretical and applied machine learning, biometry, bioinformatics and forensic statistics. Some specific areas of interest are clustering (see e.g. Blomstedt et al. 2014, IEEE TPAMI) and statistical inference in highly structured stochastic models (see e.g. Numminen et al. Royal Proceedings B). Our main application area is computational biology, in particular models for evolution and transmission of bacteria and viruses, with several highlight publications published in Nature Genetics and Proceedings of the National Academy of Sciences. These findings have been enabled by our better scalable Bayesian inference methods for the analysis of bacterial whole-genome data that are several orders of magnitude faster than recent other methods based on standard Bayesian computation. For further details and complete publication list see www. helsinki.fi/bsg

- Paul Blomstedt et al. A Bayesian predictive model for clustering data of mixed discrete and continuous type. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 489-498, 2014.
- Luca Martino et al. An adaptive population importance sampler. IEEE international conference on acoustics, speech and signal processing, pp. 8038- 8042, 2014.
- Johan Pensar et al. Labeled Directed Acyclic Graphs: a generalization of context-specific independence in directed graphical models. Data Mining and Knowledge Discovery, pp.503-533, 2014.
- Nicola Casali et al. Evolution and transmission of drug-resistant tuberculosis in a Russian population. Nature Genetics, vol 46, no. 3, pp. 279-286, 2014.
- Astrid von Mentzer et al. Identification of enterotoxigenic Escherichia coli (ETEC) clades with long-term global distribution. Nature Genetics, vol 46, no. 12, pp. 1321-1326, 2014.
- Yaara Oren et al. Transfer of noncoding DNA drives regulatory rewiring in bacteria. The National Academy of Sciences of the United States of America, vol 111, no. 45, pp. 16112-16117, 2014.
- Claire Chewapreecha et al. Dense genomic sampling identifies highways of pneumococcal recombination. Nature Genetics, vol 46, no. 3, pp. 305-309, 2014.
- Elina Numminen et al. Two-phase importance sampling for inference about transmission trees. The Royal Society B. Biological Sciences, vol 281, no. 1794, 2014.
- Gemma C. Langridge et al. Patterns of Genome Evolution That Have Accompanied Host Adaptation in Salmonella. The National Academy of Sciences, vol 112, no. 3, pp. 863-868, 2014.
- Samuel K Sheppard et al. Cryptic ecology among host generalist Campylobacter jejuni in domestic animals. Molecular Ecology, vol 23, no. 10, pp. 2442-2451, 2014.

Complex Systems Computation (CoSCo), Professor Petri Myllymäki

The Complex Systems Computation Research Group investigates computational modeling issues in complex systems, and the related implementation aspects, focusing on prediction and model selection tasks. The work has both a strong basic research component, being at the intersection of computer science, information theory and mathematical statistics, and an applied component where the results have been used for solving problems in various areas from social and historical sciences to criminology, ecology, medicine and industrial engineering.

The research areas addressed include probabilistic modeling and data analysis, information theoretical approaches to inference, constraint reasoning and ubiquitous computing. Central themes in the applied research include information retrieval, user modeling, context-awareness and data visualization.



- Jeremias Berg, Matti Järvisalo, Brandon Malone. Learning Optimal Bounded Treewidth Bayesian Networks via Maximum Satisfiability. The 17th International Conference on Artificial Intelligence and Statistics (AISTATS 2014), pp. 86-95, 2014.
- Ralf Eggeling, Teemu Roos, Petri Myllymäki, Ivo Grosse. Robust learning of inhomogeneous PMMs. The 17th International Conference on Artificial Intelligence and Statistics (AISTATS-2014), pp. 229-237.
- Brandon Malone, Kustaa Kangas, Matti Järvisalo, Mikko Koivisto, Petri Myllymäki. Predicting the Hardness of Learning Bayesian Networks. The 28th AAAI Conference on Artificial Intelligence (AAAI 2014), pp. 2460-2466.
- Kerstin Bunte, Matti Järvisalo, Jeremias Berg, Petri Myllymäki, Jaakko Peltonen, Samuel Kaski. Optimal Neighborhood Preserving Visualization by Maximum Satisfiability. The 28th AAAI Conference on Artificial Intelligence (AAAI 2014), pp. 1694-1700.
- Wolfgang Dvořák, Matti Järvisalo, Johannes Peter Wallner, and Stefan Woltran. Complexity-Sensitive Decision Procedures for Abstract Argumentation. Artificial Intelligence 206:53-78, 2014.
- Antti Hyttinen, Frederik Eberhardt, Matti Järvisalo. Constraint-based Causal Discovery: Conflict Resolution with Answer Set Programming. The 30th Conference on Uncertainty in Artificial Intelligence (UAI 2014), pp. 340-349, 2014.
- Kumaripaba Athukorala, Antti Oulasvirta, Dorota Głowacka, Jilles Vreeken, and Giulio Jacucci. Narrow or Broad?: Estimating Subjective Specificity in Exploratory Search. In Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM '14), 819-828.
- Michael Sherman, Gradeigh Clark, Yulong Yang, Shridatt Sugrim, Arttu Modig, Janne Lindqvist, Antti Oulasvirta, Teemu Roos. User-Generated Free-Form Gestures for Authentication: Security and Memorability. MobiSys '14 Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services, pp. 176-189, 2014.







Computational Logic, Professor Ilkka Niemelä

The group develops automated reasoning techniques for solving challenging computational problems in engineering and science. The current focus is on developing efficient computational methods for large constraint satisfaction problems solved, e.g., as Boolean satisfiability problems (SAT) or using rule-based constraints of answer-set programming (ASP). The group has a strong track record in research on verification and testing of automation systems and software, as well as applying formal methods in the analysis of distributed systems.

Many knowledge representation tasks involve trees or similar acyclic structures as abstract datatypes. In 2014 our research concentrated on compact encodings of acyclicity properties in terms of constraints. In addition to tree structures, the encodings developed enable the description of directed acyclic graphs and chordal graphs such as the one illustrated by the side. The applications of acyclicity constraints are manifold and it is worth mentioning graphical models in machine learning. Their structure can be compactly described so that constraint solvers with optimization capabilities can be harnessed for machine learning tasks. To implement efficient reasoning in the presence of acyclicity constraints, we developed a new extension of SAT solvers with effective propagation mechanisms for acyclicity. This extension provides an alternative way to implement ASP and we developed a new translation from ASP to SAT for this purpose. In addition, we devised new techniques for the normalization of weight rules to better support extended rule types in ASP translations. Our research on constraint-based planning and scheduling focused on automated methods for analyzing temporal models. The main outcome was a novel, very general algorithm for extracting invariants from temporal models. The invariants represent dependencies between variables in the model.

Key publications:

- Martin Gebser, Tomi Janhunen, Jussi Rintanen. ASP Encodings of Acyclicity Properties. International Conference on Knowledge Representation and Reasoning, KR 2014, AAAI Press, pp. 634-637, 2014.
- Martin Gebser, Tomi Janhunen, Jussi Rintanen. SAT modulo Graphs: Acyclicity. European Conference on Logics in Artificial Intelligence, JELIA 2014, Springer-Verlag, pp. 137-151, 2014.
- Martin Gebser, Tomi Janhunen, Jussi Rintanen. Answer Set Programming as SAT modulo Acyclicity. European Conference on Artificial Intelligence, ECAI 2014, IOS Press, pp. 351-356, 2014.
- Jori Bomanson, Martin Gebser, Tomi Janhunen. Improving the Normalization of Weight Rules in Answer Set Programs. European Conference on Logics in Artificial Intelligence, JELIA 2014, Springer, pp. 166-180, 2014.
- Jussi Rintanen. Constraint-Based Algorithm for Computing Temporal Invariants. European Conference on Logics in Artificial Intelligence, Springer-Verlag, pp. 665-673, 2014.
- Guohua Liu, Tomi Janhunen, Ilkka Niemelä. Introducing Real Variables and Integer Objective Functions to Answer Set Programming. International Conference on Applications of Declarative Programming and Knowledge Management, INAP 2013, Springer, pp. 118-135, 2014.



HIIT Annual Report 2014

dmc

Distributed and Mobile Cloud Systems (DMC)

The focus of the programme is to bridge the gap between mobile devices and the cloud based server backend systems into a single seamless distributed and mobile computing platform. The main motivation behind this is that mobile devices are by their very nature very resource constrained in available battery power, CPU, memory, network, as well as storage capacity compared to the server hardware available in the cloud backend systems. Thus mobile devices need to be tightly integrated to the cloud backend systems in order to do computational tasks that are too heavy for them. However, this basic setup is not yet sufficient for highly interactive applications. The wide area network (WAN) communication latencies between the mobile device and the possibly quite physically remote cloud backend can often be too large for interactive mobile applications, e.g., for interactive augmented reality applications such as Google Glass, as well as computationally intensive mobile intelligent information access applications. Therefore, sometimes an additional layer of computing called cloudlets is called for, that is a collection of local cloud servers that acts as local computing service for mobile devices in its own network neighborhood area. This idea is also known as cyber foraging in the literature.



Distributed Networking and Security, Professor Antti Ylä-Jääski

Distributed networking and security research group is working on two research areas, namely (a) energy aware computing and communications and (b) distributed systems and services.

Energy-aware computing and communications is a very timely and important topic. Energy consumption is a concern with today's mobile devices. While the capabilities of the devices have improved rapidly over the last ten years transforming also the way these devices are being utilized, battery technology has not been able to keep up with this evolution. As a consequence, there is an increasing gap between the battery capacity and the amount of energy required for typical usage. We address these challenges by building models of energy consumption through experiments and measurements based on which we develop new more energy-efficient protocols and services.

Distributed systems and services address architectures, platforms, and protocols for flexible, scalable, and easily usable services. Cloud computing uses virtual resources in the Internet for computing and storage, and is able to elastically scale to match changing resource needs. Merging cloud technologies with mobile domain has potential to offer new technology innovations and business opportunities for operators, vendors, and developers, as well as novel services for end users. Mobile cloud computing provides versatile research topics varying from the virtualisation layer up to the service layer.

Key publications:

- Zhonghong Ou, Hao Zhuang, Andrey Lukyanenko, Jukka Nurminen, Pan Hui, Vladimir Mazalov, Antti Ylä-Jääski. Is the Same Instance Type Created Equal? Exploiting Heterogeneity of Public Clouds. IEEE Transactions on Cloud Computing. Vol. 1, No.2, 201-214, 2014.
- Yu Xiao, Yong Cui, Petri Savolainen, Matti Siekkinen, An Wang, Liu Yang, Antti Ylä-Jääski, Sasu Tarkoma. Modeling Energy Consumption of Data Transmission over Wi-Fi. IEEE Transactions on Mobile Computing. Vol. 13, No. 8, pp. 1760-1773, 2014.
- Teemu Kämäräinen, Matti Siekkinen, Yu Xiao, Antti Ylä-Jääski. Towards Pervasive and Mobile Gaming with Distributed Cloud Infrastructure. The 13th Annual Workshop on Network and Systems Support for Games (NetGames 2014), 2014.
- Hieu Nguyen Trung, Mario Di Francesco, Antti Ylä-Jääski. A Virtual Machine Placement Algorithm for Balanced Resource Utilization in Cloud Data Centers. In the IEEE 7th International Conference on Cloud Computing (CLOUD 2014). pp. 474-481, 2014.

HIIT Annual Report 2014

dmc

Distributed Computing, Associate Professor Keijo Heljanko

The group has two main research topics: distributed computing and development methods and tools for distributed systems. On the distributed computing side the main focus is on Big Data processing and cloud computing, together with the underlying distributed algorithms and technologies, allowing the creation of novel applications in key areas of science and engineering. On the development methods and tools for distributed systems side the group focuses on automated tools for analysis of distributed systems, including dynamic symbolic execution based testing techniques, parallel and distributed verification methods, and techniques for the development and quality assurance of safety critical systems.



- Xiang Gan, Jori Dubrovin, Keijo Heljanko. A symbolic model checking approach to verifying satellite onboard software. Science of Computer Programming, 82: 44-55, 2014.
- Kari Kähkönen, Keijo Heljanko. Lightweight State Capturing for Automated Testing of Multithreaded Programs. Tests and Proofs (TAP), pp 187-203, 2014.
- Andre Schumacher, Luca Pireddu, Matti Niemenmaa, Aleksi Kallio, Eija Korpelainen, Gianluigi Zanetti, Keijo Heljanko. SeqPig: Simple and scalable scripting for large sequencing data sets in Hadoop. Bioinformatics, Vol. 30, nro 1, pp. 119-120, 2014.



Mobile Computing, Professor Sasu Tarkoma

The group investigates different aspects of wireless and mobile communications. The group has a strong focus on Internet of Things, urban analytics and energy awareness. Recent research of the group pertains to energy modeling and optimization, wearable devices, and urban mobility.

Our vision consists of cognitive devices, applications, and networks that are aware of user intents and the properties of the computing environment. These smarter systems adapt to changing conditions and deliver enriched user experience. A key challenge is how to distribute service functionality and logic in the distributed environment. The aim of our work is to enable efficient, secure, always-on, and reliable connectivity irrespective of the access network and terminal device. Moreover, service access and usage must be personalized and adapted to the current operating context.

Our central research items include mobile data analytics, energy efficiency, Internet of Things, and more recently secure wearable systems.

- Sasu Tarkoma, Matti Siekkinen, Eemil Lagerspetz, Yu Xiao. Smartphone energy consumption: modeling and optimization. Cambridge University Press, Cambridge.
- Mohammad Asharful Hoque, Matti Siekkinen, Jukka K Nurminen, Sasu Tarkoma, Mika Aalto. Saving Energy in Mobile Devices for On-Demand Multimedia Streaming: A Cross-Layer Approach. ACM Transactions on Multimedia Computing Communications and Applications, Vol. 10, No. 3, 25, 2014.
- Aaron Y Ding, Jouni Korhonen, Teemu Savolainen, Markku Kojo, Jörg Ott, Sasu Tarkoma, Jon Crowcroft. Bridging the Gap between Internet Standardization and Networking Research. Computer Communications Review, Vol. 44, No. 1, pp. 56-62, 2014.
- Aaron Y Ding, Jon Crowcroft, Sasu Tarkoma, Hannu Flinck. Software defined networking for security enhancement in wireless mobile networks. Computer Networks, Vol. 66, pp. 94-101, 2014.
- Kumaripaba Athukorala, Eemil Lagerspetz, Maria von Kugelgen, Antti Jylhä, Adam J Oliner, Sasu Tarkoma, Giulio Jacucci. How Carat Affects User Behavior: Implications for Mobile Battery Awareness Applications. CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1029-1038, 2014.
- Samuli Hemminki, Petteri Nurmi, Sasu Tarkoma. Gravity and Linear Acceleration Estimation on Mobile Devices, ACM Mobiquitous 2014.
- Hien Truong, Eemil Lagerspetz, Petteri Nurmi, Adam J Oliner, Sasu Tarkoma, N Asokan, Sourav Bhattacharya. The Company You Keep: Mobile Malware Infection Rates and Inexpensive Risk Indicators. WWW '14: Proceedings of the 23rd international conference on World Wide Web, pp. 39-50, 2014.

dmc

Collaborative Networking, Professor Jussi Kangasharju

The Collaborative Networking (CoNe) group's research focuses on largescale distributed systems and network applications which are based on nodes cooperating voluntarily. Examples of such systems are informationcentric networks and mobile opportunistic networks. The work focuses on investigating architectures and mechanisms for designing, prototyping, and testing of future networks. The goal of the work is to understand how future networks should be designed and built. The group has also worked on data center energy efficiency and green networking.

- Jinyoung Han, Taejoong Chung, Seungbae Kim, Hyun-chul Kim, Jussi Kangasharju, Ted Taekyoung Kwon, Yanghee Choi. Strategic bundling for content availability and fast distribution in BitTorrent. Computer Communications, vol 43, pp. 64-73, 2014.
- Liang Wang, Suzan Bayhan, Jussi Kangasharju. Effects of Cooperation Policy and Network Topology on Performance of In-Network Caching. IEEE Communications Letters, vol 18, no. 4, pp. 680-683, 2014.
- Suzan Bayhan, Jussi Kangasharju. Extending Cognitive Radios with New Perspectives. The Sixth International Conference on Ubiquitous and Future Networks (ICUFN), pp. 363-368, 2014.
- Stefanie Roos, Liang Wang, Thorsten Strufe, Jussi Kangasharju. Enhancing Compact Routing in CCN with Prefix Embedding and Topology-Aware Hashing. MobiArch '14: Proceedings of the 9th ACM workshop on Mobility in the evolving internet architecture, pp. 49-54, 2014.
- Esa Hyytiä, Suzan Bayhan, Jörg Ott, Jussi Kangasharju. Searching a Needle in (Linear) Opportunistic Networks. The 17th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems, pp. 187-196, 2014.
- Salim Eryigit, Gurkan Gur, Suzan Bayhan, Tuna Tugcu. Energy Efficiency Is a Subtle Concept: Fundamental Trade-offs for Cognitive Radio Networks. IEEE Communications Magazine, vol 52, no. 7, pp. 30-36, 2014.





Networking Research, Professor Andrei Gurtov

We study a wide range of data communication problems, starting from the medium access control, traffic analysis, multicast and IP mobility up to hierarchical peer-to-peer architectures and distributed algorithms. Our particular focus is security aspects and game-theoretic approach to model selfish user behavior in the modern Internet. We are actively involved in Internet Engineering and Research Task Force events contributing to creation of Internet standards. We carry out projects on Internet-of-Things security and medical ICT, the locator/identifier split, web security and trust among Internet users, and smart spaces. The group collaborates tightly with top-notch scientists from the International Computer Science Institute (ICSI) in Berkeley. The latest directions include fair bandwidth allocation in datacenter networks and resilient routing. In 2014, Dmitriy Kuptsov defended his PhD thesis on "Improving dependability of networks with penalty and revocation mechanisms". According to ResearchGate portal, the group leader was the most downloaded author of HIIT.

- Dmitriy Kuptsov, Boris Nechaev, Andrey Lukyanenko, Andrei Gurtov. How Penalty Leads to Improvement: a Measurement Study of Wireless Backoff. Computer Networks, Volume 75, Part A, pages 37-57, 2014.
- Pawani Porambage, Corinna Schmitt, Pardeep Kumar, Andrei Gurtov, Mika Ylianttila. PAuthKey: A Pervasive Authentication Protocol and Key Establishment Scheme for Wireless Sensor Networks in Distributed IoT Applications. International Journal of Distributed Sensor Networks, Nro 357430, 2014.
- Dmitry Korzun, Andrei Gurtov. Hierarchical architectures in structured peer-to-peer overlay networks. Peer-to-Peer Networking and Applications, Vol. 7, nro 4, 359-395, 2014.
- Ilya Nikolaevskiy, Dmitry Korzun, Andrei Gurtov. Security for Medical Sensor Networks in Mobile Health Systems. The Third IEEE Workshop on the Internet of Things: Smart Objects and Services 2014 (IEEE IoT-SoS 2014), 2014.
- Pardeep Kumar, Mika Ylianttila, Andrei Gurtov, Sang-Gon Lee, Hoon-Jae Lee. An Efficient and Adaptive Mutual Authentication Framework for Heterogeneous Wireless Sensor Network-Based Applications. MDPI Sensors, Vol. 14, nro 2, 2732-2755, 2014.
- Zoltán Faigl, Jani Pellikka, László Bokor, Andrei Gurtov. Performance evaluation of current and emerging authentication schemes for future 3GPP network architectures. Computer Networks, Vol. 60, nro February, 60–74, 2014.

Network Society (NS) Programme

The mission of the Network Society research programme is to empower ubiquitous users with transparent and resourceful ICT with the following overarching objectives:

- Multimodal interaction and adaptive information: resources are provided to users to interact with information ubiquitously. This includes studies of modalities such as manual ergonomics and skills, tactile interaction and cross modal rendering. Adaptation and situated computing is pursued to contribute to issues such as data deluge and persuasion.
- Emergent social media platforms and practices. The study of emergent forms of ICT including gaming, basic studies of practices and behavior in social media with special attention to knowledge work, privacy and self expression.
- Experience of Mediated Interaction. Emotional and cognitive processes during mediated social interaction and the neuroscience of social ICT.

While there are three research groups their research objectives contribute to a common agenda.

These three challenges are integrated to solve in an adjacent or combined way problems or anticipate future ICT. As examples new ubiquitous or multimodal interaction can contribute to novel forms of social media such as public and pervasive displays. The neuro-psychology of mediated interaction can contribute to understanding social media practices.

NS collaborates with other programmes, in particular in the HIIT wide focus area focusing on information exploration and knowledge work.



Digital Content Communities, Professor Marko Turpeinen

The goal of our research is to enable and encourage people to belong to communities of content creators, to develop service design methodology and business model development for social media, to study how people interact with each other using digital content, and to understand what new business opportunities and social structures emerge around communitygenerated content and technology. One of our strong focus areas is the rapidly expanding social forms of gaming and how to apply the findings from online games to other non-gaming domains. This includes virtual economies and virtual consumerism, and applying economics and business studies in virtual worlds.

Another viewpoint is legal. In relation to future technologies, it is not enough to be able to apply only traditional jurisprudencial methods, but we must also collect and analyze empirical data. Therefore, in addition to conventional legal studies that focus on the analysis of statutes, court cases, etc, we also seek to understand practices and business reasons, as well as availing of user studies to find real human, business, and societal needs. Likewise, futures research methods, e.g. user scenarios, may give us important information on forthcoming issues. Digital Content Communities is a joint research group of both HIIT and EIT ICT Labs Helsinki node.

- Esa Pitkänen, Paula Jouhten, Jian Hou, Muhammad Syed Fahad, Peter Blomberg, Jana Kludas, Merja Oja, Liisa Holm, Merja Penttilä, Juho Rousu. Comparative Genome-Scale Reconstruction of Gapless Metabolic Networks for Present and Ancestral Species. PLOS Computational Biology, Vol. 10, nro 2, 2014.
- Francesco Bonchi, Aristides Gionis, Francesco Gullo, Antti Ukkonen. Distance oracles in edge-labeled graphs. 17th International Conference on Extending Database Technology (EDBT), 2014.
- Matthijs van Leeuwen, Antti Ukkonen. Fast Estimation of the Pattern Frequency Spectrum. The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD), pp. 114-129, 2014.
- Antti Ukkonen. Indirect Estimation of Shortest Path Distributions with Small-World Experiments. The Thirteenth International Symposium on Intelligent Data Analysis (IDA 2014), pp. 333-344, 2014.
- Kai Kuikkaniemi, Vilma Lehtinen, Matti Nelimarkka, Max Vilkki, Jouni Ojala, Giulio Jacucci. Designing for presenters at public walk-up-and-use displays. Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction - TEI '14.
- Teemu Leinonen, Eva Durall, Kai Kuikkaniemi, Teemu Mikkonen, Antti Syvänen, Matti Nelimarkka, Tarmo Toikkanen. Design for Learning: Enhancing Participation in Learning through Design Thinking. World Conference on Educational Multimedia, Hypermedia and Telecommunications, Tampere 2014.
- Mikael Johnson, Sampsa Hyysalo, Samuli Mäkinen, Pia Helminen, Kaisa Savolainen, Louna Hakkarainen. From recipes to meals... and dietary regimes: method mixes as key emerging topic in human-centred design. The NordiCHI14 conference, ACM, pp. 343-352, 2014.



Social Interaction and Emotion (SIE), Professor Niklas Ravaja

The mission of SIE is to increase our understanding of ICT-mediated social interaction. Our studies focus on: (a) emotional and cognitive processes during mediated social interaction, including mediated touch, interaction with virtual humans, and knowledge work, (b) the neuroscience of social ICT, (c) socially-mediated behavioral phenomena in SNSs, (d) the use of gamification to support the adoption of healthier lifestyles, and (e) user experience of ICT.

We use a wide methodological approach, from surveys and qualitative methods to neurophysiological recordings (e.g., EEG), to study users in the laboratory as well as in real-life contexts. By providing both fundamental and applied insight into social interaction, we help to create ICT and services that take the human social and emotional processes into account.

- Niklas Ravaja, Jari Kätsyri. Suboptimal facial expression primes in textual media messages: Evidence for the affective congruency effect. Computers in Human Behavior, 40, 64–77, 2014.
- Michiel Spapé, Eve Hoggan, Giulio Jacucci, Niklas Ravaja. The meaning of the virtual Midas touch: An ERP study in economic decision making. Psychophysiology, 2014.
- Niklas Ravaja, Pekka Aula, Alessio Falco, Salla-Maaria Laaksonen, Mikko Salminen, Antti Ainamo. Online news and corporate reputation: A neurophysiological investigation. Journal of Media Psychology, 2014.
- Michiel Spape, Imtiaj Ahmed, Giulio Jacucci, Niklas Ravaja. The self in conflict: Actors and agency in the mediated sequantial Simon Task. Frontiers in Psychology, 2014.
- Matias Kivikangas, Jari Kätsyri, Simo Järvelä, Niklas Ravaja. Gender differences in emotional responses to cooperative and competitive game play. PLoS One, 9(7): e100318, 2014.









Ubiquitous Interaction (UiX), Professor Giulio Jacucci

Ubiquitous Interaction studies the design, development and evaluation of interactivity with users in ubiquitous and mobile computing systems. The research is carried out coupling in-depth user studies with design in the area of novel interfaces. The goal is to contribute to technology development efforts by deepening understanding of human and design perspectives. Currently, research is conducted in different areas: multitouch displays, adaptivite and affective interfaces, multimodal interaction, exploratory search and social computing. Application areas range from sustainability and environmental awareness, home computing, walk-up-and-use display, information exploration, art and culture, scientific and knowledge work.

- Ruotsalo, Tuukka, Jacucci, Giulio, Myllymäki, Petri, Kaski, Samuel. Interactive intent modeling: information discovery beyond search. Communications of the ACM 58.1 (2014): 86-92.
- Michiel Spapé, Eve Hoggan, Giulio Jacucci, Niklas Ravaja. The meaning of the virtual Midas touch: An ERP study in economic decision making. Psychophysiology, 2014.
- Jung-Joo Lee, Siân Lindley, Salu Ylirisku, Tim Regan, Markus Nurminen, and Giulio Jacucci. Domestic appropriations of tokens to the web. The 2014 conference on Designing interactive systems, pp. 53-62. ACM, 2014, Best paper nominee.
- Giulio Jacucci, Luciano Gamberini, Jonathan Freeman, Anna Spagnolli (Eds.), Symbiotic Interaction, Third International Workshop, Symbiotic 2014.
- Kumaripaba Athukorala, Eemil Lagerspetz, Maria von Kugelgen, Antti Jylhä, Adam J Oliner, Sasu Tarkoma, Giulio Jacucci. How Carat Affects User Behavior: Implications for Mobile Battery Awareness Applications. CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1029-1038, 2014.
- Kumaripaba Athukorala, Antti Oulasvirta, Dorota Glowacka, Jilles Vreeken, Giulio Jacucci. Narrow or Broad? Estimating Subjective Specificity in Exploratory Search. CIKM'14 - ACM International Conference on Information and Knowledge Management, 2014.
- Manuel Eugster, Tuukka Ruotsalo, Michiel Spape, Ilkka Kosunen, Oswald Barral Mery de Bellegarde, Niklas Ravaja, Giulio Jacucci, Samuel Kaski. Predicting term-relevance from brain signals. SIGIR '14 Proceedings of the 37th international ACM SIGIR conference on Research and development in information retrieval, pp. 425-434, 2014.
- Gabor Aranyi, Sid Kouider, Alan Lindsay, Hielke Prins, Imtiaj Ahmed, Giulio Jacucci, Paolo Negri, Luciano Gamberini, David Pizzi, Marc Cavazza. Subliminal Cueing of Selection Behavior in a Virtual Environment. Presence, vol 23, no. 1, pp. 33-50, 2014.
- Silvia Gabrielli, Paula Forbes, Antti Jylhä, Simon Wells, Miika Sirén, Samuli Hemminki, Petteri Nurmi, Rosa Maimone, Judith Masthoff, Giulio Jacucci. Design challenges in motivating change for sustainable urban mobility. Computers in Human Behavior, vol 41, pp. 416-423, 2014.



focus

HIIT-Wide Focus Area: Augmented Science, Dr. Antti Honkela, Dr. Tuukka Ruotsalo (Coordinators)

The advances in science depend on effectively building upon the results that others have achieved, which are based on previously collected data. This brings to the forefront a challenge for computational science to better utilize the massive explosion in digital scientific data, whether it is scientific literature, or raw measurement data from previous studies. HIIT has begun focusing resources on a HIIT-wide focus area, titled Augmented Science, which builds on the existing excellences and touches most of HIIT research topics. Augmented science develops information technology methods and pilot applications for making the data-driven fields such as modern biology cumulative, and revolutionizing the way we search and access scientific resources and literature. Much of the most interesting biomedical data comes from individual people. We are developing methods that ensure the privacy of these sensitive data while allowing their use. Improving the general problem solving method of science, in collaboration with the other fields, is the best way for HIIT to contribute to solving the grand challenges of the humanity.

We aim to support scientific information access by enabling better coordination of communicating ideas and scientific results within the scientific community. Results of scientific efforts are traditionally published as articles and communicated personally as presentations or other related materials. The rapid communication of knowledge between researchers is a key success factor to enable better science. The volume of scientific output is estimated to be millions of publications worldwide per year; the growth rate of PubMed alone is more than 1 article per minute. The problem of communication that the scientific community is facing is shifting from publishing and sharing the information to finding and filtering the suitable materials to support every day work of researchers. We aim to help scientists to better find and manage the content that they use in their everyday work.

- Tuukka Ruotsalo, Giulio Jacucci, Petri Myllymäki, and Samuel Kaski. Interactive intent modeling: information discovery beyond search. Commun. ACM 58, 1 pp 86-92, 2014.
- Kumaripaba Athukorala, Antti Oulasvirta, Dorota Głowacka, Jilles Vreeken, and Giulio Jacucci. Narrow or Broad?: Estimating Subjective Specificity in Exploratory Search. The 23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM '14), pp 819-828, 2014.
- Manuel J.A. Eugster, Tuukka Ruotsalo, Michiel M. Spapé, Ilkka Kosunen, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski.
 Predicting term-relevance from brain signals. The 37th international ACM SIGIR conference on Research & development in information retrieval (SIGIR '14). pp 425-434, 2014.
- Tuukka Ruotsalo, Jaakko Peltonen, Manuel J.A. Eugster, Dorota Glowacka, Aki Reijonen, Giulio Jacucci, Petri Myllymäki, and Samuel Kaski. Intentradar: search user interface that anticipates user's search intents. CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14), pp 455-458, 2014.



adm

Administration





HELSINGIN YLIOPISTO HELSINGFORS UNIVERSITET UNIVERSITY OF HELSINKI

Personnel and funding

HIIT is a joint research institute of Aalto University and the University of Helsinki. At Aalto University, HIIT researchers currently work in the Department of Computer Science or HIIT's own "department" at Open Innovation House; at University of Helsinki, our operations are at the Department of Computer Science, Department of Mathematics and Statistics, and Department of Social Sciences. The personnel of HIIT are employed by the two parent universities. Many of HIIT's personnel have double or even triple affiliations. Most common is an affiliation with other Departments of one or both of the parent universities, but there are also some who share their time between HIIT and some other organisation. Thus the total number of HIIT affiliated personnel (about 275, of which 39 per cent are of foreign origin) is much higher than the number of personyears, 134.

The total funding of HIIT in 2014 was 10,5 MEur, of which 82 per cent was competitive. The main funding sources were Academy of Finland, Tekes, EU and the mother universities. More details will be given in the Facts and Figures document found via http://www.hiit.fi/abouthiit

Board

The highest decision-making body of HIIT is the Board. It decides on HIIT's overall research strategy and research programmes. The statutory tasks of the Board are to approve the annual budget and activity plans, and follow up and comment on the work of HIIT through regular activity updates given by the Director of HIIT. In 2014 the Board convened four times.

The HIIT Board consists of nine members who are appointed for four years at a time. Each university appoints four board members, two of which are university staff and two are not employed by the university, and personal deputies to each of them. The staff of HIIT selects one board member and his or her deputy from among their colleagues. A new Board for the term 1 April 2014 - 31 March 2018 was nominated in early 2014, and the Board now consists of the following members (personal deputies in parentheses):

- Chairman: Dean, Professor Risto Nieminen, Aalto (Vice Dean, Professor Jouko Lampinen, Aalto)
- Vice chairman: Dean, Professor Jouko Väänänen, UH (Vice Dean, Professor Esko Ukkonen, UH)
- Professor Pekka Orponen, Aalto (Professor Lauri Savioja, Aalto)
- Professor Jukka Paakki, UH (Professor Jyrki Kivinen, UH)







adm

- Vice President Hannu Kauppinen, Nokia Oyj (Director Jyri Huopaniemi, Nokia Oyj)
- Docent Kari-Pekka Estola (Innovation Director Ville Peltola, IBM)
- Director Mervi Karikorpi, The Federation of Finnish Technology Industries (Director Petri Vasara, Pöyry Management Consulting Oy)
- Principal Scientist Harri Valpola, Zenrobotics (CEO Kimmo Kiviluoto, Enreach Solutions)
- Personnel representative: Docent Antti Honkela, HIIT (Dr. Antti Jylhä, HIIT)

The Director of HIIT Samuel Kaski was responsible for preparing and submitting propositions to the Board. Board Secretary was Research Coordinator, Docent Ella Bingham.

Scientific Advisory Board (SAB)

The SAB consists of internationally prominent scholars who are invited by the HIIT Board. The SAB convened in 2012 and its recommendations were actively taken into account during 2013 and 2014. SAB Members in 2012 are listed in the following.

- Professor Randy Katz, University of California at Berkeley, Chairman of SAB
- · Professor Alberto Apostolico, Georgia Tech
- · Professor Christos Faloutsos, Carnegie Mellon University
- · Professor Jodi Forlizzi, Carnegie Mellon University
- · Professor Bengt Jonsson, Uppsala University
- · Professor Martin Kersten, University of Amsterdam and CWI
- · Professor Kari-Jouko Räihä, University of Tampere
- · Professor Mart Saarma, University of Helsinki
- Professor John Shawe-Taylor, University College London











HIIT Annual Report 2014