Proceedings of the

Oxford Computer Science Conference 2018

DEPARTMENT OF
COMPUTER
SCIENCE

1st June 2018

Contents

Conference Chairs

Ulrik Lyngs, *General Chair*
Jaclyn Smith, *Programme Chair*
Temitope Ajileye, *Conference co-Chair*
Arianna Schuler Scott, *Conference co-Chair*

With *outstanding* help, support, and guidance from
Julie Sheppard, Sarah Retz, Lyn Hambridge,
Kelly Ryan, Leanne Carveth, and Martin Dehnel-Wild

Review Committee

David Tena Cucala, Alina Petrova, Vojtech Havlicek
Nitin Agrawal, Daniel Woods, Johan Wahlstrom, Amartya Sanyal

Conference Sponsors

Google DeepMind, Accenture, and Ocado

1 Programme

Lecture Theatre B

9:30 **Welcome Address**

Ulrik Lyngs, General Chair

9:35 **Session I: Security**

Formal analysis of 5G-AKA, and security vulnerability
Martin Dehnel-Wild and Cas Cremers

Evaluating smartwatch privacy games through a longitudinal study
Meredydd Williams, Jason Nurse and Sadie Creese

Leaky Wires: Information Leakage and Covert Communication Between FPGA Long Wires
Ilias Giechaskiel, Kasper Rasmussen and Ken Eguro

10:35 **Session II: Lightning Talks**

A Threat Model for EMI Attacks on Analog-to-Digital Converters
Ilias Giechaskiel, Youqian Zhang and Kasper Rasmussen.

A Simulation-based Framework for the Security Testing of Autonomous Cars
Eduardo Dos Santos

Inter-domain Deep Gaussian Processes
Tim Rudner and Dino Sejdinovic

Atrium 10:55 **Coffee break**

Lecture Theatre B

11:15 **Session III: Logic, Algorithms, and Information**

Consequence-based Reasoning for the Web Ontology Language
David Tena Cucala, Bernardo Cuenca Grau and Ian Horrocks

Low Rank Structure of Learned Representations
Amartya Sanyal, Varun Kanade and Philip H. S. Torr

An abstract model for higher-order incremental computation
Mario Alvarez-Picallo

Dual-STCs; creating robust stego objects for the noisy channel
Christy Kin-Cleaves and Andrew Ker

Atrium 12:35 **Poster session & lunch**

Lecture Theatre A	13:30	**Keynote panel: Coding the Future: Values in Computer Science** Michael Wooldridge, Mariarosaria Taddeo, Nigel Shadbolt, David Franke. Kenneth Cukier chairs.
Atrium	14:45	**Poster session & coffee**
Lecture Theatre B	15:20	**Session IV: Biology and Human Centered Computing** **Social Acceptability and Respectful Virtual Assistants** William Seymour **Investigation of time-dependent ECG biomarkers of heart rate for risk stratification in hypertrophic cardiomyopathy** Anna Bialas, Aurore Lyon and Alfonso Bueno-Orovio **Bayesian Verification of Chemical Reaction Networks** Gareth Molyneux, Viraj Brian Wijesuriya and Alessandro Abate
	16:20	**Session V: Lightning Talks** **Medical information extraction with deep neural networks** Maximilian Hofer and Alejo Nevado-Holgado **Communicating Recurrent Neural Networks for Resource Constrained Systems** Prince Makawa Abudu
	16:35	**Closing remarks**
Atrium	16:40	**Drinks reception & awards ceremony**
Kellogg College	18:30	**Conference dinner**

2 Talk Abstracts

Session I: Security

Formal analysis of 5G-AKA, and security vulnerability (Slides)
Martin Dehnel-Wild and Cas Cremers

The 5th Generation (5G) mobile networks and telecommunications standards are currently under development, and are nearly finalised. We analyse the security properties of the main 5G-AKA protocol within the February 2018 version of the draft standard. Our analysis reveals a security vulnerability in the proposed 5G-AKA protocol as specified within the standard. The discovered protocol vulnerability would allow a malicious actor (with no privileged network access) to impersonate another user to a Serving Network, for example in a roaming scenario. This could potentially allow malicious users to bill expensive phone calls or access charges to other legitimate users, after eavesdropping on their initial connection.
We found the vulnerability by performing formal symbolic analysis of the protocol standard using the Tamarin Prover. After describing the protocol and the vulnerability, we provide possible fixes, and verification of the fixes.

Evaluating smartwatch privacy games through a longitudinal study (Slides)
Meredydd Williams, Jason Nurse and Sadie Creese

Smartwatches are growing in popularity, offering useful apps and wearable connectivity. To provide these functions, they access messages, contacts, calendars and locations. Despite this ability, users rarely configure their privacy settings. Serious games have been found persuasive for behaviour change. Therefore, we developed the first smartwatch privacy game, carefully designed through Learning Science principles. To evaluate its influence, we conducted a two-month longitudinal study. 10 participants were given a smartwatch, with their behaviour logged for three weeks. The treatment half then installed the game, which they used for 9 days. This app assigned privacy challenges, and adjusted its tasks based on current behaviour. Our control group installed a non-privacy version, reducing bias from confounding factors. We then monitored posttest behaviour for three weeks, concluding with semi-structured interviews. Protective behaviour was successfully encouraged: GPS usage decreased by 40% while password usage rose by 43%. Our control group failed to differ, implying bias was not introduced. None of these users ever checked GPS or permissions, suggesting privacy would not be protected. In the interviews, the treatment group demonstrated risk awareness and smartwatch knowledge. In contrast, most control participants failed to perceive a threat. This suggests that interactive games can encourage privacy-protective behaviour.

Leaky Wires: Information Leakage and Covert Communication Between FPGA Long Wires

Ilias Giechaskiel, Kasper Rasmussen and Ken Eguro

Field-Programmable Gate Arrays (FPGAs) are integrated circuits that implement reconfigurable hardware. As the capacity of FPGAs grows, it is common for designers to incorporate implementations of algorithms and protocols from third-party sources. The monolithic nature of FPGAs, however, means that all on-chip circuits (including third-party designs) must share common on-chip infrastructure, such as routing resources. Our research shows that a "long" routing wire carrying a logical 1 reduces the propagation delay of adjacent but unconnected long wires in the FPGA interconnect, thereby leaking information about its state. We exploit this effect and propose a communication channel that can be used for both covert transmissions between circuits, and for exfiltration of secrets from the chip. We show that the effect is measurable for both static and dynamic signals, and that it can be detected using small on-board circuits. In our prototype, we are able to correctly infer the logical state of an adjacent long wire over 99% of the time, even without error correction, and for signals that are maintained for as little as 82 μs. Using a Manchester encoding scheme, our channel bandwidth is as high as 6kbps. We characterize the channel in detail and show that it is measurable even when multiple competing circuits are present and can be replicated on different generations and families of Xilinx devices (Virtex 5, Virtex 6, and Artix 7). Finally, we propose countermeasures that can be deployed by systems and tools designers to reduce the impact of this information leakage.

Session II: Lightning Talks

A Threat Model for EMI Attacks on Analog-to-Digital Converters

Ilias Giechaskiel, Youqian Zhang and Kasper Rasmussen

An Analog-to-Digital Converter (ADC) bridges the analog and the digital world by converting continuous analog properties into discrete digital signals. However, the ADC cannot authenticate its inputs, presenting a vulnerability which can be exploited through malicious Electromagnetic Interference (EMI) signals, which are picked up by the wires connected to the ADC. We present a threat model for ADCs under EMI attacks, and experimentally demonstrate how ADCs can act as demodulators of amplitude-modulated signals. Attackers can thus transmit high-frequency signals and cause malicious low-frequency signals to appear at the outputs of ADCs. For example, we demonstrate that “OK Google” commands can be successfully injected into Android smartphones through the headphone wires.

A Simulation-based Framework for the Security Testing of Autonomous Cars

Eduardo Dos Santos

Sensor attacks can lead to an inaccurate perception of the environment by autonomous cars. However, in the event of a particular sensor being subject to attack, the car may still be able to behave safely by means of other unaffected sensors, or via sensor redundancy. However, it is challenging to assess how true this requirement holds in real traffic situations because of the costs and the dangerous nature of testing `in the wild'. Simulation tools for autonomous driving research enable the simulation of the urban environment with high level of fidelity, thus providing a risk- and cost-free platform for data collection, training and testing of in-car perception systems. In this work, we extend an open-source simulator with models of camera attacks with a view to test the accuracy of built-in learning algorithms against these kinds of attacks. This can be assembled into a framework for the automated and systematic testing of autonomous car sensors. Our contribution can be used to guide model training and attack impact assessment, as well as support more robust machine learning models.

Inter-domain Deep Gaussian Processes
Tim Rudner and Dino Sejdinovic

We propose a novel variational inference method for deep Gaussian processes (GPs), which combines doubly stochastic variational inference with variational Fourier features, an inter-domain approach that replaces inducing points-based inference with a framework that harnesses RKHS Fourier features. First experiments have shown that inter-domain deep Gaussian processes are able to achieve levels of predictive performance superior to shallow GPs and alternative deep GP models.

Session III: Logic, Algorithms, and Information

Consequence-based Reasoning for the Web Ontology Language
David Tena Cucala, Bernardo Cuenca Grau and Ian Horrocks

In this paper, we present a consequence-based calculus for concept subsumption and classification in the description logic ALCHOIQ, which supports all Boolean connectives, role hierarchies, inverse roles, number restrictions, and nominals. By using well-known transformations, our calculus extends to the description logic SROIQ, which covers all of OWL 2 DL except for datatypes. A key feature of our calculus is its pay-as-you-go behaviour— unlike existing reasoning algorithms for description logics, our calculus is worst-case optimal for all the well-known fragments of ALCHOIQ.

Low Rank Structure of Learned Representations (Slides**)**
Amartya Sanyal, Varun Kanade and Philip H. S. Torr

A key feature of neural networks, particularly deep convolutional neural networks, is their ability to “learn” useful representations from data. The very last layer of a neural network is then simply a linear model trained on these “learned” representations. Despite their numerous applications in other tasks such as classification, retrieval, clustering etc., a.k.a. transfer learning, not much work has been published that investigates the structure of these representations or whether structure can be imposed on them during the training process.

In this paper, we study the dimensionality of the learned representations by models that have proved highly successful for image classification. We focus on ResNet-18, ResNet-50 and VGG-19 and observe that when trained on CIFAR10 or CIFAR100 datasets, the learned representations exhibit a fairly low rank structure. We propose a modification to the training procedure, which further encourages low rank representations of activations at various stages in the neural network. Empirically, we show that this has implications for *compression* and robustness to *adversarial examples*.

An abstract model for higher-order incremental computation
Mario Alvarez-Picallo

Incremental computation is a technique to reduce the cost of applying the same function to an input that is modified over time, by instead applying it to an initial input and then using an incremental version of the function to update the output. This technique is well-known in the literature, but a recent work by Cai, Giarrusso, Rendel and Ostermann has opened a new avenue of research into the denotational aspects of this technique with their use of derivatives.

In this talk, I intend to give a more rigorous treatment of the semantics of incremental computation by introducing the notions of difference algebras and differentiable functions between these, which give rise to two Cartesian closed categories of incrementalizable programs. These can be further generalized to difference stacks and smooth functions, which again constitute a Cartesian closed category. My intention is to show that difference stacks can be used as a denotational model for higher-order languages equipped with an operator for automatically incrementalizing arbitrary expressions.

Dual-STCs; creating robust stego objects for the noisy channel (Slides**)**
Christy Kin-Cleaves and Andrew Ker

Steganography is the practice of encoding messages in inconspicuous media files, resulting in stego objects. Robust steganography creates stego objects with an element of robustness, able to convey the secret message over the noisy channel. In this paper we propose an extension of Syndrome Trellis Codes (STCs), to generate stego objects that are both distortion-minimizing and robust. These aims are in tension because low distortion requires choosing codewords that are close, but robustness requires codewords to be far apart. The noisy channel may present itself in many forms, including recompression. The ability to create robust stego objects would allow social media channels, such as Facebook etc., to be used as a method of distributing stego payloads. Since their introduction in the 2010 paper 'Minimizing Embedding Impact in steganography using Trellis-Coded Quantization', STCs have become the widely accepted state of the art for adaptive steganography. In our work, we show a method to extend the STC algorithm to create robust stego objects whilst still minimising distortion. We compare against, and show that we can improve, the state of the art with our proposed solution.

Session IV: Biology and Human Centered Computing

Social Acceptability and Respectful Virtual Assistants
William Seymour

Underneath the friendly facade, do you feel like there is something sinister going on with Siri? This talk introduces a new way of conceptualising many of the interaction problems we see with modern virtual assistants: respect. Beginning by identifying a lack of respect in interactions as an overarching link between instances of device behaviour seen as creepy, problematic, or just generally unwanted, it then briefly explains the semantics of respectful devices, including the four most applicable types of respect that have been identified by philosophers. Finally, current research being undertaken by the human centred computing theme at Oxford is presented, which is beginning to explore how people react to devices that exhibit different thresholds of respectful behaviour.

Investigation of time-dependent ECG biomarkers of heart rate for risk stratification in hypertrophic cardiomyopathy (Slides)
Anna Bialas, Aurore Lyon and Alfonso Bueno-Orovio

Hypertrophic Cardiomyopathy (HCM) is the most common genetic heart disease, affecting one in five hundred people. Although many HCM patients remain asymptomatic, others are characterised by a high risk of sudden cardiac death due to ventricular arrhythmias. Importantly, the methodology to identify such subjects at risk is not yet well established. In this work, we investigate time-dependent biomarkers of heart rate and heart rate variability in HCM, integrated into machine learning techniques for risk stratification. Multi-resolution heart rate features were derived from the 24-hour Holter electrocardiogram (ECG) recordings of 49 HCM patients and 29 healthy subjects, providing extensive information on heart rate. Different dimensionality techniques were analysed for feature selection in the complex multi-featured medical records and the classification of HCM patients from controls via support vector machines.

Dimensionality reduction by Laplacian Eigenmaps successfully clustered the HCM cohort into four groups, organised from normal to severe abnormalities in ventricular activation. The last group displays further subgroups, which identify patients with specific abnormalities promoting ventricular arrhythmias. These results, previously not captured by former research on morphological ECG features, greatly increase the potential of machine learning techniques for risk stratification in HCM.

Bayesian Verification of Chemical Reaction Networks (Slides)
Gareth Molyneux, Viraj Brian Wijesuriya and Alessandro Abate

We present a data-driven verification approach that determines whether or not a biological system satisfies a given property, expressed as a formula in a modal logic. Our approach

consists of three phases, integrating formal verification over models with learning from data. First, we consider a parametric set of possible models based on the known stoichiometry and classify them against the property of interest. Secondly, we utilise Bayesian inference techniques to update a probability distribution over a model class with data gathered from the underlying system. In the third and final stage, we combine the results of both steps to compute the probability that the underlying system satisfies the property. We apply the new approach on two case studies and compare it to standard statistical model checking. We argue that our approach is data-efficient compared to standard statistical model checking techniques.

Session V: Lightning Talks

Medical information extraction with deep neural networks (Slides)
Maximilian Hofer and Alejo Nevado-Holgado

Electronic Health Records (EHRs) are the databases used by hospital and general practitioners to daily log all the information they record from patients (i.e. disorders, taken medications, symptoms, medical tests, etc.). Most of the information held in EHRs is in the form of natural language text (written by the physician during each session with each patient), making it inaccessible for research. Unlocking all this information would bring a very significant advancement to biomedical research, multiplying the quantity and variety of scientifically usable data.

In Artificial Intelligence, Information Extraction (IE) is the task of systematically extracting information from natural language text in a form that can later be processed by a computer. For instance, if a physician describes the symptoms and full treatment of a patient, an IE task could be identifying from the text alone all the drugs prescribed to the patient and their dosages. This challenge remains an unsolved problem in complex cases (e.g. badly structured language; few labelled samples), which is more akin the text typically found in EHRs. Namely, physicians tend to use badly formatted shorthand and non-widespread acronyms (e.g. 'transport pt to OT tid via W/C' for 'transport patient to occupational therapy three times a day via wheelchair'), while labelled records are scarce (ranging in the hundreds for a given task).

Communicating Recurrent Neural Networks for Resource Constrained Systems (Slides)
Prince Makawa Abudu

Applications that require heterogeneous sensor deployments such as wildlife tracking,construction site surveillance, bridge activity monitoring and military surveillance systems continue to face practical challenges related to energy efficiency, computational power and reliability, tedious design implementations, effective communication, optimal

sampling and accurate event classification, prediction and detection. Owing to resource constraints in environments where sensors are operating, there is need for effective ways of selecting a sensing strategy that maximises detection accuracy for events of interest using available resources and data-driven approaches. Inspired by those limitations, we ask two fundamental questions: whether state-of-the-art Recurrent Neural Networks can observe different series of data and communicate in their hidden states to collectively solve an objective in a distributed fashion. We realise our answer by conducting a series comprehensive and systematic analyses of communicating recurrent neural network architectures on varying time-steps, objective functions and number of nodes. Our experimental setup models tasks synonymous with those in Wireless Sensor Networks. Our contributions show that Recurrent Neural Networks can communicate in their hidden states and we achieve promising results from our quantitative analyses of our architectures' performances.

3 Poster Abstracts

Learning the Intuitive Physics of Non-Rigid Object Deformations
Stefano Rosa, Zhihua Wang, Andrew Markham and Niki Trigoni

The ability to interact and understand the environment is a fundamental prerequisite for a wide range of applications from robotics to augmented reality. In particular, predicting how deformable objects will react to applied forces in real time is a significant challenge. This is further confounded by the fact that shape information about encountered objects in the real world is often impaired by occlusions, noise and missing regions e.g. a robot manipulating an object will only be able to observe a partial view of the entire solid. In this work we present a framework, 3D-PhysNet, which is able to predict how a three-dimensional solid will deform under an applied force using intuitive physics modelling.

In particular, we propose a new method to encode the physical properties of the material and the applied force, enabling generalisation over materials. The key is to combine deep variational autoencoders with adversarial training, conditioned on the applied force and the material properties. We further propose a cascaded architecture that takes a single 2.5D depth view of the object and predicts its deformation. Training data is provided by a physics simulator. The network is fast enough to be used in real-time applications from partial views. Experimental results show the viability and the generalisation properties of the proposed architecture.

A Cognitive Design Space for Supporting Self-Regulation of ICT Use
Ulrik Lyngs

A growing number of surveys and interview studies find that a majority of users of smartphones and laptops often struggle with effective self-control over their device use. In response, HCI research - as well as a rapidly growing commercial market for 'anti-distraction tools' - have begun to develop apps, browser plugins, and other tools to help users understand and regulate their ICT use. Such tools tend to either remove or block distractions (e.g. Freedom, Newsfeed Eradicator), track and visualise device use (e.g. RescueTime, Moment), or reward intended/punish unintended use (e.g. Forest, Obtract).

However, human-computer interaction research on how to design ICTs to support self-regulation of use is in its early days. Existing developments of anE-distracEon tools tend to rely mostly on user studies or a limited number of theoretical frameworks including Social Cognitive Theory and Goal-Setting Theory. We suggest that bridging current research efforts to a wider body of self-regulation research within cognitive neuroscience and behavioural economics will provide a useful framework for understanding and predicting relevant design features in terms of the major cognitive mechanisms involved in self-regulation.

Paranoid Android: Identifying Impostors in the App Marketplace
Alexandru Okros and Reuben Binns

Smartphone app marketplaces often feature `impostor apps', which deceive users into downloading them by mimicking better-known apps. This paper focuses on the impostors of two popular applications, Facebook Messenger and Candy Crush, from the Google Play Store. Based on these case studies, we introduce a typology of impostor apps which aims to capture the differences in intentions, benefits and risks associated with them. We then investigate the privacy implications for a user who installs such apps. Finally, we explore a possible solution to the problem of detecting impostor apps using machine learning techniques. The results suggest that this is feasible using only features that are exposed publicly via the Play Store, and we highlight the most promising features to be used for this purpose.

Deep learning based QRS Multilead Delineator in Electrocardiogram signals
Julià Camps, Blanca Rodriguez and Ana Minchole

The surface electrocardiogram (ECG) is the most widely adopted test to diagnose cardiac diseases. Extracting critical biomarkers from these signals, such as the QRS width, requires delineating the fundamental waves in them. However, even though ECG signals significantly change depending on the recording methodology and the patient's cardiac condition, the available QRS delineators are hard to adapt to non-considered cases. This study presents a deep learning-based multilead ECG delineation method which can successfully delineate QRS complexes. Our approach reached root-mean-square errors (RMSE) of 12.1+-0.5 and 18.5+-1.1 ms for QRS onset and offset, respectively, when evaluated on the QT database. These results are similar to the RMSE calculated from differences between the two cardiologists that annotated this database, namely, 14.7 ms for the QRS onset and 17.2 ms for the offset. Therefore, the proposed method is comparable to the state-of-the-art.

Preserving Privacy in Smart Homes - A Socio-Cultural Perspective
Martin Johannes Kraemer

Smart homes are expected to reduce the time we spend on routine activities. Devices which enable this collect and process data. Therefore, smart homes also have the potential to exaggerate bewilderment and resistance, feelings people express when their privacy is infringed. Because privacy is being influenced by socio-cultural factors and shaped by technology, this work argues for a thorough understanding of the home’s socio-cultural context. We aim to provide a grounded-in-data, contextual smart home privacy model. The model will be applicable to product and policy design, and also inform future work in privacy research for ubiquitous computing settings.

4 Keynote Panel

Coding the Future: Values in Computer Science

From the controversy around Facebook's handling of personal data to calls by the ACM for peer review to consider negative side effects of computing, discussions about ethics in computing are surging. Yet, in many computer science departments considerations of societal impact and values in design play a marginal role at best. How will the discipline of computer science adapt to a world where social problems are increasingly solved – and created – by algorithms, and how do we best prepare our students for this future?

Michael Wooldridge, Mariarosaria Taddeo, Nigel Shadbolt, and David Franke. Chaired by Kenneth Cukier.

Michael Wooldridge
Michael Wooldridge is Head of Department and Professor of Computer Science at the Department of Computer Science, University of Oxford. He is a lead researcher on the project 'Towards a code of ethics for AI research', one of 37 projects backed by a $10m donation from Elon Musk to ensure that AI 'remains beneficial'. He recently won the 2018 IFAAMAS Influential Paper Award, for influential and long-lasting contributions to the field of autonomous agents and multiagent systems.

Mariarosaria Taddeo
Mariarosaria Taddeo is a Research Fellow at the Oxford Internet Institute and Deputy Director of the Digital Ethics Lab, as well as a Fellow on the World Economic Forum's Council on the Future of Cybersecurity. In 2010, she was awarded the Simon Award for Outstanding Research in Computing and Philosophy.

Nigel Shadbolt
Principal of Jesus College and Professorial Research Fellow in Computer Science, Sir Nigel Shadbolt is head of the Human Centred Computing theme at the Department of Computer Science, University of Oxford. He is chairman of the Open Data Institute, which he co-founded with Tim Berners-Lee, and former Information Advisor to the UK government.

David Franke
UK Government Affairs Manager at Microsoft, David Franke leads a number of areas within Microsoft's government affairs programmes in the UK, including Accessibility, AI, and Cyber Security. He is former Private Secretary to successive Communication Ministers, as well as to the Intellectual Property Minister, during the 2010-15 coalition.

Kenneth Cukier (chair)
Senior editor of digital products at *The Economist*, Kenneth Cukier is co-author of *Big Data: A Revolution that Will Transform How We Work, Live and Think*, and *Learning with Big Data: The Future of Education* with Viktor Mayer-Schönberger, Professor of Internet Governance and Regulation at the Oxford Internet Institute. *Big Data* was shortlisted for FT Business Book of the Year 2013.

5 Awards

Prizes were awarded to the following outstanding submissions and presentations:

Best Talk

1st (£200): Martin Dehnel-Wild
2nd (£75): Gareth Molyneux

Best Poster

1st (£150): Meredydd Williams
2nd (£75): Prince Abudu

Best Abstract

1st (£100): Anna Bialas
2nd (£50): David Tena Cucala

These awards were made possible by kind co-sponsorship from Accenture (main sponsor of all prizes) and Ocado (co-sponsor of best talk 1st prize).

The main conference and evening dinner would not have been possible without kind sponsorship from Google DeepMind.