MECHANICAL ENGINEERING I PHYSICS I PRESERVATION OF THE ARCHITECTURAL HERITAGE I SPATIAL PLANNING AND URBAN DEVELOPMENT I STRUCTURAL SEISMIC AND GEOTECHNICAL ENGINEERING I TECHNOLOGY AND DESIGN FOR ENVIRONMENT AND BUILDING I TERRITORIAL DESIGN AND GOVERNMENT I URBAN PLANNING. DESIGN AND POLICY I AEROSPACE ENGINEERING I ARCHITECTURAL AND URBAN DESIGN I ARCHITECTURAL COMPOSITION IARCHITECTURE. BUILT ENVIRONMENT AND CONSTRUCTION ENGINEERING I ARCHITECTURE, URBAN DESIGN, CONSERVATION OF HOUSING AND LANDSCAPE I BIOENGINEERING I DESIGN I ELECTRICAL ENGINEERING I ENERGY AND NUCLEAR SCIENCE AND TECHNOLOGY I ENVIRONMENTAL AND INFRASTRUCTURE ENGINEERING **IINDUSTRIAL CHEMISTRY AND CHEMICAL** ENGINEERING I INFORMATION TECHNOLOGY I INTERIOR ARCHITECTURE AND DESIGN I MANAGEMENT ENGINEERING I MATERIALS ENGINEERING I MATHEMATICAL MODELS AND METHODS IN ENGINEERING

PhD Yearbook | 2017



DOCTORAL PROGRAM IN INFORMATION TECHNOLOGY

Chair: Prof. Andrea Bonarini

Introduction

The PhD program in Information Technology (IT) goes back to the year 2001, when the two traditional programs in Automation- Computer Engineering and Electronics-Telecommunications were merged. As such, the unified course covers the research interests in four scientific areas, namely: Computer Science and Engineering, Electronics, Systems and Control, and Telecommunications. This broad variety of research activities is completely focused in the ICT area, and perfectly corresponds to the core mission of the Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB).

However, following the historical development of the Department, and the new trends of the modern society, some cross-related research fields are also developed, such as ecology, environmental modelling, operations research, and transportation systems. The PhD program in IT is the largest at the Politecnico in terms of number of students. There are more than 50 first year students and about 190 in total. The students are subject to an examination every year to evaluate the progress achieved in their research and course work.

Topics

The research carried out in the Department in the field of Information Technology is supported by 35 laboratories, and is organized in 4 main areas, listed here below.

Computer Science and Engineering (Vice-Coordinator: Prof. Cristiana Bolchini): Information systems, Database management, Information design for the web, Methods and applications for interactive multimedia, Embedded systems design and design methodologies, Dependable systems: performance, security and reliability, Autonomous robotics, Artificial intelligence, Computer vision and image analysis, Machine learning, Dependable Evolvable Pervasive Software Engineering, Compiler Technology, Natural Language Processing and Accessibility. *Electronics* (Vice-Coordinator: Prof. Angelo Geraci): Circuits and systems: theory and applications, Single-photon detectors and applications, Radiation detectors and low noise electronics, Electronic circuit design, Electron devices.

Systems and Control (Vice-Coordinator: Prof. Paolo Bolzern):

Control systems, Robotics and industrial automation, Optical measurements and laser instrumentation, Dynamics of complex system, Planning and management of environmental systems, Operations research and discrete optimization. *Telecommunications* (Vice-Coordinator: Andrea Virgilio Monti Guarnieri): Networking, Applied electromagnetics, Information transmission and radio communications, Optical communications, Wireless and space communications, Remote sensing, Signal processing for multimedia and telecommunications.

Industrial collaborations

Due to its intrinsic technological nature, the PhD curriculum is corroborated by many industrial collaborations. About 40% of the total number of scholarships are funded by companies or by international research projects involving industrial partners. In the school vision, the collaboration between university and companies is ideally based on the challenge of turning invention and scientific research into technological innovation. This shapes new technology frontiers and builds a fertile atmosphere for a mixture of world-class research at universities and in private and public companies. This also contributes to create a common terrain of friendly culture, to size the risk, and to believe in strong basic research. The external referee board is composed by members of public and private companies, working in management, production, and applied research. The board is in charge of monitoring the activities of the PhD program and giving suggestions for its development. The board meets once a year to point out the new emerging research areas worth to be investigated and to monitor the visibility of the course in the industrial and service world.

Educational aspects

The teaching organization and subject of the courses reflect the scientific interests of DEIB faculties. The curricula include a wide choice of courses (about 30 per year) of different nature offered by this PhD Program, and more than 30 courses for basic skill formation offered by the Polimi PhD School. Access to external courses and summer schools is also encouraged.

The challenge is to promote interdisciplinary research while offering advanced courses that spur innovative and cutting edge research.

Therefore, particular attention is devoted to help students to make their best choice according to an internal regulation scheme.

Internationalization

Every year, several courses are delivered by foreign professors. Moreover, the PhD program encourages joint curricula through agreements with foreign institutions. We have several Double Degree and Joint Degree agreements with countries in all continents. Every year we have more than 150 applications from foreign countries and about 20% of our PhD candidates come from there.

Conclusions

The core mission of this PhD Program is to offer an excellent PhD curriculum, through high quality courses, a truly interdisciplinary advanced education, cutting-edge research, international collaborations, and collaborations with companies.

PHD BOARD OF PROFESSORS					
Andrea Bonarini	Stefano Ceri	Pier Luca Lanzi			
Cesare Alippi	Matteo Cesana	Maurizio Magarini			
Francesco Amigoni	Giancarlo Ferrigno	Federico Malucelli			
Luciano Baresi	Carlo Ettore Fiorini	Andrea Virgilio Monti Guarnieri			
Luca Bascetta	Simone Garatti	Barbara Pernici			
Giuseppe Bertuccio	Gian Guido Gentili	Ivan Rech			
Cristiana Bolchini	Angelo Geraci	Alessandro Sottocornola Spinelli			
Paolo Bolzern	Carlo Ghezzi				
Andrea Francesco Castelletti	Paolo Giacomazzi				

PHD ADVISORY BOARD				
Claudio Bartolini	HP Labs (USA)			
Riccardo De Gaudenzi	European Space Agency			
Giuseppe Fogliazza	MCM S.p.A.			
Renato Marchi	Gruppo PAM			
Giorgio Parladori	SM Optics s.r.l.			
Fabrizio Renzi	IBM Italia S.p.A.			
Massimo Valla	TIM			
Stefano Verzura	Huawei Technologies Italia srl	Huawei Technologies Italia srl		

Prizes and awards

In 2016 the following awards have been obtained by PhD students:

Chorafas Foundation Award: Stefano Ambrogio, Matteo Pirotta Best Junior Carassa Award 2016: Andrea Annoni CPSWEEK Best PhD Forum Presentation Award: Naveed Anwar Bhatti Ericsson Innovation Award – 3rd place: Saleem Shahid iEMSs 2016 Student Best Presentation Award: Francesca Recanati IEEE CAS DAY 2016 Best Poster Award: Giacomo Gervasoni IEEE PRIME 2016 Golden Leaf Award: Dmytro Cherniak IEEE RNDM 2016 Best Paper Award: Ali Hmaity IEEE WACV Best paper Award 2016: Andrea Romanoni XILINX Open Hardware 2016 Award for PhD FPGA Category: Giuseppe Natale 23rd RAW Best Demo Award: Emanuele Del Sozzo IEEE RNDM 2016 Best Paper Award: Ali Hmaity ACM ASSETS Student Research Competition 2016: Mirko Gelsomini IEEE NSS Student Awards 2016 - 4th place: Giovanni Ludovico Montagnani INNS-BigData 2016 Best Regular Paper Award: Diego Carrera DeepVision Best Paper Award: Francesco Visin, Marco Ciccone

HIGH-PERFORMANCE INTEGRATED ELECTRONICS FOR HIGH-THROUGHPUT TIME CORRELATED SINGLE PHOTON COUNTING

Acconcia Giulia - Advisor: Prof. Ivan Rech

The analysis of optical signals by means of Single Photon Avalanche Diodes (SPADs) has been subject to a widespread and steadily increasing interest in recent years, driven by the need for ultimate sensitivity in various scientific and industrial applications such as Fluorescence Lifetime Imaging (FLIM) and Förster Resonance Energy Transfer (FRET) in life sciences, Laser Imaging Detection and Ranging (LIDAR) in remote objects sensing and Quantum Key Distribution (QKD) in cryptography and communication. In particular, the use of fluorescence lifetime spectroscopy as both analytical and research tool has increased markedly in recent years with remarkable applications in chemistry, biochemistry and biology. The Time Correlated Single Photon Counting (TCSPC) technique is a powerful tool by which fluorescence-lifetime measurements are carried out. In the complex environment of living cells, for instance, the study of the composite interactions between proteins and other cellular components is extremely challenging due to the reactions and uncertainties that mainly affect the intensity of the stimulated fluorescence emission. Nevertheless, a deep understanding of these processes is paramount for a

better understanding of the (patho)physiology of tissues and organisms and gives a base for gaining a better insight into key medical issues, such as the origin and growth mechanisms of tumors. In this case the exploitation of a TCSPC-based lifetime imaging technique makes it possible to overcome many issues, since it relies on phenomena that, unlike intensity, are mostly independent from probe concentration and all other parameters that cannot be easily controlled during cellular experiments.

Although TCSPC is a very effective and well-established technique. it can have a few drawbacks: the repetitive nature of this measurement, for example, makes the acquisition time intrinsically long. The need to speed up the acquisition time - in order to exploit this technique in all those applications where the duration of the measurement is crucial for example - has been driving the development of multichannel acquisition systems featuring a steadily increasing number of channels operating in parallel. However, the currently available systems either reported in the scientific literature or commercially available still suffer from a trade-off between the number

of channels and performance: state-of-art systems are typically limited to a few number of channels, while systems that feature hundreds of channels are much less performing than the aforementioned ones. The current trend towards densely integrated systems relies on the exploitation of standard CMOS technologies to integrate both the SPAD and the electronics on the same chip. This approach led to remarkable results in terms of number of channels operating in parallel. The exploitation of a standard process for the fabrication of the detector. though, is based on the re-use of structures (i.e. tubs, wells, etc.) that have been designed to optimize transistors performance and it prevents the designer from having the degrees of freedom necessary to pursue the best device performance. To date, best in class performance in terms of photon detection efficiency (especially at the wavelengths of most interest in medical and biological applications), in terms of temporal resolution, in terms of a combination of the formers and in terms of their timing figure of merit have been obtained resorting to technologies specifically dedicated and optimized for the fabrication of the SPAD, usually known as custom technologies.

Taking advantage of the remarkable features of custom technology SPAD devices in real acquisition systems, though, requires complex front end and processing electronics designed on purpose. Excellent results have been obtained with single channel modules and linear arrays featuring up to a few tens of detectors. Unfortunately, the solutions adopted in those works are not compatible with high-density systems. In order to overcome the limitations of the previously developed systems, the complete re-design of the read out and processing electronics in a multichannel perspective has been necessary.

The approach followed in this work is completely orthogonal to what other researchers currently pursue, both in terms of the technologies and of the architecture adopted. Highperformance, fully-integrated readout and processing electronics has been developed using two different technologies, each of them optimally suited for a single part of the system, targeting the exploitation of dense arrays of high-performance custom technology SPAD detectors combined with the electronics in a 3-D stacked architecture.

In order to extract the timing information from custom technology SPAD sensors with picosecond resolution, I designed a new current pick-up circuit in deep-submicron CMOS technology (180nm); this circuit is capable of sensing the fast leading edge of the photogenerated current at the anode terminal of the

SPAD with a timing jitter as low as few tens of picoseconds. The output of the pick-up circuit has to be fed to a time measurement circuit in order to compute the time of arrival of the photon that triggered the avalanche current growing. The use of a conversion chain for each detector, though, is not only expensive in terms of area and power dissipation but it is also unnecessary. In fact, in a TCSPC measurement the probability that one pixel detects a photon during each excitation cycle must be far less than one in order to avoid any distortion of the reconstructed light signal. This means that the number of conversion chains required in each excitation interval is actually much lower than the number of detectors in the system. Moreover, the amount of data potentially produced by a great number of independent time measurement circuits that operate in parallel could rapidly lead to a saturation of the transfer rate towards the PC or the capability to store them in real time on an internal hard drive. For all these reasons this work has been focused on the design of an entity able to join the advantages of having only a few acquisition chains - so that the resources in terms of area and power can be optimally used to achieve very high performance and numerous detectors that can provide a truly parallel spectral or spatial analysis of the sample. The operation of this intermediary circuit guarantees that the maximum sustainable throughput can always be achieved while also maintaining high performance

on each channel. The connection between the two groups must be dynamic in order to maximize the exploitation of the time measurement circuits and, mostly important, in order to avoid that the presence of some SPAD sensors operating at very high rates (e.g. hot pixels) could result into the masking of some area of the array. To this aim, a routing circuit has been designed, capable of dynamically binding a dense array of SPAD detectors to N time measurement circuits where the number N of outputs has been dimensioned in order to achieve a maximum sustainable throughput in the order of 10Gbit/s. The routed signals have to be fed to a time measurement circuit. In order to fully exploit the typical repetition rate of the excitation laser, I developed a TAC featuring an 80MHz conversion rate in 0.35µm Si-Ge technology. In this way the converter is ready to perform a conversion at any excitation pulse clock thus increasing the total efficiency of the system. Considering a 32x32 array of SPAD detectors requiring 10 bits to be univocally identified, with 80MHz conversion rate and two bytes to encode the timing information, this work targets the exploitation of five conversion chains, that would correspond to a maximum output rate of the system as high as 10.4 Gbit/s.

ADVANCE OPTICAL ROUTING TECHNIQUES IN THE SOFTWARE DEFINED ERA

Alvizu Gomez Rodolfo Enrique - Supervisor: Prof. Guido Maier

Paradoxically, with an everincreasing traffic demand, today transport-network operators experience a progressive erosion of their margins. The operational complexity, manual configuration and static nature of current technologies lead to: inefficient network utilization, overprovisioning of resources and very high Capital expenditures (CapEx) and Operational expenses (OpEx). The alarms of change are set, and Software Define Networking (SDN) is coming to the rescue with the promise of reducing CapEx and OpEx. Driven by economic needs and network innovation facilities. today transport SDN (T-SDN) is a reality. It gained big momentum in the last years, however in the networking industry, the transport network will be perhaps the last segment to embrace SDN, mainly due to the heterogeneous nature and complexity of the optical equipment composing it. This thesis starts with a deep dive into a fascinating technological adventure that provides an organic analysis of the T-SDN development and evolution. Our work is the first to consider contributions from the whole transport network ecosystem composed by: academic research, standardization bodies, industrial development, open source projects and alliances among them. After creating a

comprehensive picture of T-SDN, we provide an analysis of many open issues that are expected to need significant future work, and give our vision in this path towards a fully programmable and dynamic transport network.

Then, assuming the deployment of T-SDN technologies, we propose operational research formulations and heuristic algorithms for two advanced routing techniques: dynamic optical routing and multipath optical routing. In general, humans follow highly predictable daily movements. We commute from residential to working and/or educational areas in a daily basis, and we have a selection of commercial and recreational areas for the nights and weekends. We also use the mobile phone at regular hours, for example when commuting, during lunch break and at night. Such regular behavior creates predictable spatio-temporal fluctuations of traffic patterns, which in analogy to the periodical rise and fall of the sea levels, its known as tidal traffic phenomenon. We have exploited such predictability, and the centralized control plane and programmability of SDN to propose off-line optimization and on-line matheuristic to reduce the energy consumption in the optical layer of metro-core networks,

while providing 1+1 protection of the aggregated traffic. The matheuristic dynamically optimizes resource allocation by effectively adapting to predictable aggregated tidal traffic variations. The optimality of on-line decisions is achieved by generating optimal weighted graphs from the interaction with an off-line optimization phase. The off-line phase solves an ILP problem using regular tidal traffic patterns. Our results display energy savings of more than 20% by introducing load adaptive network operation, while the matheuristic closely follows the optimal results. We have also introduced a heuristic method to reduce service disruption due to routing changes, while preserving energy saving capability. The heuristic allows to reduce service disruption in the range of 38 to 80% with a small penalty on energy saving of less than 5%. Historically, Internet traffic has been routed over the shortest path: which was convenient for best-effort data traffic; but it is not always suitable for today's scenario. Multipath (MP) routing is an effective technique for applications imposing stringent requirements on bandwidth, delay and availability. MP routing overcomes few limitations of single path routing like increased throughput and network resource

utilization and resilience to possible link failures. However, different data paths over the network face different end-toend path delay which is called differential delay (D). In presence of D the destination of a MP connection receives a disordered version of the original packet sequence. Thus, D limits the performance gain of MP routing. We investigated two main approaches to mitigate the effects of D: 1) Differential delay-constrained MP routing problem (the 3D problem), and 2) Differential delay-compensated MP routing problem. The 3D problem is an optimization problem for MP routing involving maximization of the cardinality K of the disjoint-path set, under the D constraint as the primary objective, and then minimization of the average end-to-end transfer delay for the fixed (maximum) K under the same D constraint. The optimization approach is iterative, based on solving an inner mixedinteger programming subproblem to minimize the delay for a given value of K and D. We proposed a technique to equalize D that exploits unconventional routing cycles (e.g., loops). This solution is novel and creative, and increases the solution space exploiting the nature of optical communications in which delays are deterministic. Numerical results demonstrate the gain achieved by using controlled loops in comparison with a traditional loop-free approach. On the other hand, differential delay-compensated MP routing for optical transport networks normally performs differential delay compensation (DDC) at

destination, or distributed using high speed reconstruction buffers. Such techniques incur in extra costly and power hungry electro/ optical (E/O) conversions, that are otherwise avoided by routing all optical circuits. In this thesis we proposed, for the first time, the use of compact fiber delay lines (FDL)s to perform distributed all optical DDC. The FDLs are passive elements that overcome the problems of previous solutions. An integer linear programming formulation is presented for the MP routing with DD-minimization problem that combines electronic-DDC co-located with 3R regeneration points with the novel transparent-DDC based on FDLs. Numerical results show the advantages of combining transparent and electronic-DDC in realistic network scenarios. Finally, based on the background on T-SDN and the advanced routing algorithms, we present the implementation of two SDN use cases. The first use case is a transport network Orchestrator for multi-domain networks. The second use case is a path manager and the related automated testbed for evaluating Multipath TCP (MPTCP). We demonstrated a hierarchical control plane architecture for SDN-based Segment Routing (SR) in multi-domain networks. An orchestrator application, on top of multiple open source SDN

controllers, creates a hierarchical

control plane architecture using

northbound RESTFul APIs of

controllers. The orchestrator

has control, visibility and traffic

engineering capabilities to manage

multi-domain SR service creation.

Standard southbound interfaces with proper SR extensions are exploited to manage SR tunnels in the MPLS data plane. The Orchestrator has the following capabilities: cross-domain segment routing, application-aware path selection, and MP routing. MPTCP is an extension to TCP protocol that allows a single connection to be split across multiple paths. MPTCP overcomes few limitations of TCP thereby offering added benefits like increased throughput and network resource utilization and resilience to possible link failures. The performance of MPTCP is affected not only by the delay but also by the DD. MPTCP runs on the host, and it is not aware of the underlying network topology. Hosts cannot create optimal paths to destination without assistance from a network control element. We use SDN technology to provide this assistance to MPTCP and to setup best available paths for MPTCP connections. We design an SDN-based path manager for MPTCP that allowed us to deploy DD-aware algorithms. Moreover, we implemented an automated testbed for performance evaluation, to experimentally demonstrate the impact of delay and D on the overall performance of MPTCP.

THEATREBOT: STUDYING EMOTION PROJECTION AND EMOTION ENRICHMENT SYSTEM FOR AUTONOMOUS THEATRICAL ROBOT

Angel Fernandez Julian Mauricio - Supervisor: Prof. Andrea Bonarini

Robots should be able to express emotional states to interact with people as social agents. However there are cases where robots cannot have bio-inspired bodies, for instance because the task to be performed requires a specific shape, this is the case of home cleaners, package carriers, and many others platforms. In these cases, emotional states need to be represented by exploiting other features, such as movements of the body. The work presented in this thesis studies emotion expression in non-anthropomorphic platform. Therefore a set of case studies. aimed to identify specific values to convey emotion through changes in linear and angular velocities that might be applied on different non-anthropomorphic platforms were done. The work takes into account some of the most considered emotion expression theories and from emotion coding for people. The results show that people can recognize some emotional expressions better than others, and it is proposed some directions to express emotions exploiting only bio-neutral movements. In parallel to these studies, it was developed an emotional enrichment system, which let the parametrization, modulation and expression of emotions through



the change of actions' parameters and addition of required actions. This system was used in the last two cases studies, where the emotions elicited by the robot were totally generated by the system. Additional a simple test were done with other platforms to test the interoperability of the system among diverse platforms. Moreover, it was explored the possibility to use robots in theatre to study further social capabilities in robotics. To facilitate the use of robots in theatre, software architecture has been designed, which includes the emotional enrichment system.

CONTROL STRATEGIES IN LARGE PHOTONIC INTEGRATED CIRCUITS

Annoni Andrea - Supervisor: Prof. Andrea Ivano Melloni

Introduction

Nowadays, due to the everlasting increase of internet traffic and energy consumption for data transmission and data processing, integrated photonics technologies are envisioned as a viable alternative to power hungry electronics. To reach this objective, it is necessary to handle more and more complex functions in the optical domain; this objective can be realized only if photonic circuits reach a higher level of complexity, interconnecting many devices on the same chip to realize a so-called

``system-on-a-chip"; but the lack of adequate tools to reliably control the manufacturing and operation of photonic integrated circuits when scaled up to hundreds of optical functional elements, prevents to establish an equivalent of the Moore's law in photonics circuits. This thesis contributes to fill some of the existing gaps that are currently preventing to move from a single-device level to a new system-on-a-chip paradigm; the work shown here, lays in the framework of the FET European project BBOI (Breaking the Barrier of Optical Integration, www.bboi.eu) which objective is to remove the limitations of current Silicon-On-Insulator photonic platforms through the exploitation of radically new concepts and enabling technologies for the

realization and control of photonic devices.

Silicon photonic integrated circuits can integrate thousands of components in a mm2 but are also intrinsically imperfect due to the sensitivity to fabrication tolerances and temperature fluctuations: therefore, the response of fabricated PICs hardly matches the design performance. To solve this issue, it is not always possible to have an accurate and robust control by relying only on lookup tables, due to functional drifts and components ageing over time; hence, we need an automatic feedback control loop to set the new working point automatically and control the PIC to counteracts unwanted drifts. Moreover, it is possible that reconfiguration capabilities are needed to adapt the circuit response to new required functions (i.e. dynamic switching and routing, channels add/drop and so on). While the active control of single devices is a known issue and many control strategies are available; efficient and smart control of complex circuits, with many interconnected and interdependent degrees of freedom, is still an open issue that has to be tackled.

Results

To this aim, we first introduced a powerful tool for non-invasive

light monitor in semiconductor waveguides; the CLIPP obtain information on the optical power in a waveguide by monitoring the conductance changes associated with free carriers generated by the light itself without tapping any photon. We demonstrated that the CLIPP can be exploited for active feedback control on silicon integrated devices; active control by means of non-invasive CLIPPs has been achieved on silicon microring resonators, demonstrating advanced controls capabilities such as tuning, locking, and wavelength swapping. Moreover, channels can be monitored independently even when multiplexed on the same waveguide by means of low frequency pilot tone that does not impair the signal quality. We then exploited this non-invasive monitor to achieve control in complex silicon photonic circuits by integrating CLIPPs in strategic positions. We then demonstrated methods

and techniques to control arbitrarily large switch fabrics on an 8x8 Mach-Zehnder-Interferometers (MZIs) switch matrix. A strategic use of CLIPPs throughout the circuit allows to reduce this complex control problem, with many and interconnected degrees of freedom, to smaller and simpler problems with a reduced number of degrees of freedom (as in this case, up to just one). The problem of tuning the switching matrix to route an input channel, was reduced to the sequential tuning of each switching element involved in the routing; CLIPPs allow to track the path of the channels throughout the circuit and to use their signal as a feedback to tune and stabilize the switching elements. We also demonstrated that when many input signals are simultaneously injected in the switch matrix, each channel can be distinguished and monitored individually by exploiting on-chip generated labelling tones. By exploiting local monitoring and feedback loops, we counteracted the effects of thermal crosstalk between switching elements when reconfiguration events occurred in the circuits. Due to this thermal crosstalk, the working point of a MZI is affected by the status of



2. Layout of the Mode-Division-Multiplexing 4 channels unscrambler shown in Fig.1. This circuit implements all-optically functions that otherwise would require complex fast DSP techniques.

the other switching elements; when a section of the matrix is reconfigured, control algorithms are needed to keep stable the working point of each MZI switch of the circuit.

At last we demonstrated is an all-optical mode division multiplexing unscrambler for mode division multiplexing optical communications, the silicon photonic circuit has been wire bonded to CMOS circuitry for control purposes (**fig.1**); the



1. Silicon photonic chip (middle) wire bonded to two CMOS electronic chips with dedicated control circuitry.

device is capable of realizing a function that otherwise needs to be carried out by powerhungry MIMO DSP techniques. The circuit relies on MZIs and phases shifter to revert the mixing effect of a multi-mode channel by applying an inverse transformation with respect to the one from the channel: noninvasive detectors are needed to avoid differential losses that otherwise will impair the channel reconstruction. Moreover, the circuit architecture (fig.2) does not allow to be tuned only by relying on signals from external monitors since the outputs are not related univocally to the working point of the actuators in the circuit. We tackled this issue by integrating non-invasive CLIPPs in circuit, thus allowing to look at each unscrambler stage independently; we simplified the control problem by decomposing it in the sequential tuning of the circuit stages, by means of local feedback loop similar to the one exploited to automatically reconfigure the 8x8 switch fabric. **NFORMATION TECHNOLOGY**

COMPILER AUTOTUNING USING MACHINE LEARNING TECHNIQUES

Ashouri Amir Hossein - Supervisor: Prof. Cristina Silvano

Although there were several longstanding problems with compiler optimizations, they have not been adequately addressed because optimizations were yielding performance improvements. These problems included knowing what optimizations to apply, and in which configuration (e.g., the tile size in loop tiling) and, in which order to apply them for the best improvement. The former yields the so-called problem of selecting the best compiler optimizations and the latter is the phase-ordering problem of compiler optimizations. Compiler optimization problem polarizes over two major sub-problems based on (i) whether we take into account the enabling/disabling the optimizations only (optimization selection problem) or (ii) changing the ordering of those optimizations (phase-ordering problem). Here we briefly discuss the different optimization space of the two. Challenge 1. Design Space Exploration of Compiler Parameters Embedded systems design traditionally exploits the knowledge of the target domain, e.g. telecommunication, multimedia, home automation etc., to customize the HW/ SW coefficients found onto the deployed computing devices. Although the functionalities of these devices are differed, the

computational structure and design are tightly connected with the platform in which they rely on. Platform-based design has been proposed as a promising alternative for designing complex systems by redefining the problem of designing into that of finely tuning specific parameters of the platform template. Although a significant amount of research has been conducted on exploring and optimizing VLIW architectural parameters and introducing specific compiler optimization for VLIW processors, there are limited references regarding the analysis of the impacts of conventional compiler transformations onto VLIW architectures and moreover how these transformations are correlating with the underlying architectural configuration. Challenge 2. The Selection Problem Several compiler optimizations form an optimization sequence. When we disregard the ordering of these optimizations and focus on whether or not to apply the optimization, we define the scope of selecting the best compiler optimizations. Previous researchers have shown that the interdependencies and interaction between enabling/disabling optimizations in a sequence can dramatically alter performance of a running code even by ignoring

the order of phases. This process of selecting the right optimizations for each code segment is typically done manually and the sequence of optimizations is constructed with little insight into the interaction between the preceding compiler optimizations in the sequence. The task of constructing heuristics to select the right sequence of compiler optimizations is infeasible given the evergrowing number of compiler optimizations being integrated into compiler frameworks. As an example, GCC has more than 200 compiler passes, referred to as compiler options, and LLVMclang and LLVM-opt both have more than 100 transformation passes each. Additionally, these optimizations are applied at very different phases of the compilation, including analysis passes and loop-nest passes. Most optimization flags are turned off by default and compiler developers rely on software developers to know, which optimizations will be beneficial for their code. Compiler developers provide standard optimization levels, e.g. -01, -02, -Os, etc. to introduce a fixed-sequence of compiler optimizations that, on average, bring good performance on a set of benchmarks compiler developers tested the optimization

levels with. However, using predefined optimizations usually is not good enough to bring the best achievable application-specific performance. One of the key approaches that are used recently in literature in order to find the best optimizations to apply given an application is inducing prediction models using different classes of machine learning. Approaches, which leverage machine learning to find the best optimizations to apply, will be the center focus of this chapter. Challenge 3. The Phase-ordering Problem The phase-ordering problem has been an openproblem in the field of compiler research for many decades. The inability of researchers to solve the phase-ordering problem has led to advances in the simpler problem of selecting the right set of optimizations, but even this problem has yet to be solved. Compiler designers must consider the order in which optimization phases are performed; a pair of phases may be interdependent in the sense that each phase could benefit from information produced by the other. When having both the selection and the ordering are of importance, the phase-ordering problem is formed. It is once of the longstanding problems of compilation field and has its peer problems in numerous other

sub-fields of compiler design such as register allocation, codegeneration and compaction. Dissertation Contribution Addressing Challenge. 1 In order to address the exploration of design space parameters in embedded domain, we propose a methodology that provides the designer with an integrated framework to automatically (i) generate optimized applicationspecific VLIW architectural configurations and (ii) analyze compiler level transformations, enabling application-specific compiler tuning over customized VLIW system architectures. We based the aforementioned analysis on a Design of Experiments (DoEs) procedure that captures in a statistical manner the higher order effects among different sets of activated compiler transformations. Applying the proposed methodology onto real-case embedded application scenarios, we show that (i) only a limited set of compiler transformations exposes high confidence level (over 95%) in affecting the performance and (ii) using them we could be able to achieve gains between (16-23)% in comparison to the default optimization levels. Chapter 3 discusses more in details the aforementioned contribution. Addressing Challenge. 2 In

order to address the problem of selecting the right set of compiler optimizations, we proposes COBAYN: COmpiler autotuning framework using BAYesian Networks, an approach for a compiler autotuning methodology using machine learning to speed up application performance and to reduce the cost of the compiler optimization phases. The proposed framework is based on the application characterization done dynamically by using independent micro-architecture features and Bayesian networks. The chapter also presents an evaluation based on using static analysis and hybrid feature collection approaches. In addition, the chapter compares Bayesian networks with respect to several state-of-the-art machinelearning models. Chapter 4 discusses more in details the aforementioned contribution. Addressing Challenge. 3 Mitigating the phase-ordering problem, we propose two different approaches: (i) an intermediate speedup predictor and (ii) a complete sequence predictor approaches. We elaborate on the pros and cons of each approach and provide extensive experimental comparison against standard optimization levels and state- ofthe-art techniques. Chapters 5 and 6 discusses more in details the aforementioned contribution.

PhD Yearbook | 2017

MODELING, DESIGN, CONTROL AND VALIDATION OF A FULL HYBRID ELECTRIC BICYCLE

Berretta Daniele - Supervisor: Prof. Matteo Corno

Co-Supervisor: Prof. S.M. Savaresi

This Thesis deals with the project and design of a full hybrid electric bicycle. The main characteristics of this novel vehicle are:

- Small additional weight with respect to a traditional bike
- A main power source (the cyclist) and a secondary power source (the battery pack)
- Charge sustaining mode (bike does not need to be recharged from electric grid differently from all the other electric bicycle present in the market)
- Globally decrease of main power source effort (i.e. the human effort)

Two main sections compose the research work: the first part describes the design and validation of the hybrid electric bike algorithm, the second part introduces some features that allow the assistance algorithm to work properly and to improve its safety. The first part describes the system setup, underlining substantial equivalency in terms of weight and appearance with respect to a traditional bike. The modelling and validation of a complete electro-mechanical model of the bike follow. Then, a control-oriented analysis of the cyclist metabolic efficiency is carried out. For this purpose, three features are employed: an ad-hoc defined equivalent

cycling efficiency based on oxygen consumption, a dynamic model for the State of Fatigue (SoF) and Heart Rate (HR) measurements. The analysis of the equivalent cycling efficiency and SoF dynamics guide the design of the charge-sustaining assistance algorithm. The idea is to recover energy from the cyclist when he is most efficient and then return it during low efficiency pedalling. The proposed system is extensively tested on several users, showing that it can achieve a 25 % improvement in equivalent cycling efficiency, and a reduction in the peak HR and SoF during urban cycling. In other words, the first part of this work further develops the charge sustaining idea for EPACs along the following directions:

- A new closed-loop chargesustaining algorithm is presented that also accounts for road slope and regenerative braking. The closed-loop nature makes it robust to uncertainties and viable for real world usage. The proposed algorithm does not need biomedical feedback (measured nor estimated).
- An exhaustive control-oriented analysis of cyclist effort is discussed using three different features: an ad-hoc defined equivalent cycling efficiency,

a State of Fatigue model and

- Heart Rate measurements.
- A rigorous simulation and model based tuning of the control algorithm is presented.
- An exhaustive experimental validation of the approach is presented using tests performed on a closed track and in a city with several subjects. The results are analysed from both a quantitative and qualitative point of view.
- The second part of the Thesis presents two additional features of the HEB algorithm: an innovative Kalman filter estimator for road slope and an algorithm to improve the safety of the HEB algorithm when the cyclist walks with the bicycle. The assistance algorithm needs to know the road slope to work properly and to improve the cyclist efficiency while he is cycling. The Kalman filter estimator fulfills this necessity. On the other side, as the European normative states for safety reasons, the HEB algorithm should enable the assistance only when the cyclist is pedalling. Unfortunately, due to the mechanical configuration, the system can activate improperly the assistance. In fact, the all-inone architecture has no torque sensor to prevent this issue. Thus, an algorithm that detects if the

cyclist is on-seat or he is walking with the bicycle is designed. The cyclist detection is only based on bicycle speed, cadence and inertial measurements. The proposed system is based on the hierarchical analysis of three different bicycle dynamics: longitudinal, lateral and vertical dynamics. The right management of the intrinsic signal information along each dynamic aims to resolve this problem and it increases the safety of the whole system. In fact, cyclist on-seat detection algorithm prevents some critical situations that could happen when the cyclist walks with the bicycle.

LINKING SIGNAL AND SEMANTIC REPRESENTATIONS OF MUSICAL CONTENT FOR MUSIC INFORMATION RETRIEVAL

Buccoli Michele - Supervisor: Prof. Augusto Sarti

Co-Supervisor: Dr. Massimiliano Zanoni

In recent years, the evolution of technology and connectivity has led to a major revolution in the music industry, which has led to novel content consumption scenarios. People, in fact, are suddenly given the possibility to easily access millions of songs through multiple browsing/diffusion/streaming platforms and are therefore in need of new navigational/browsing tools, music recommendation assistants. virtual clerks, etc. Music Information Retrieval (MIR) is a multi-disciplinary research field that addresses the issues raised by the design of search/navigation/annotation strategies, in order to develop those tools, like content classification and rich annotation algorithms, that are required to support such scenarios. The design of tools and applications for MIR-related scenarios requires us to model and account for two different levels of abstraction in the musical content description: the semantic level and the signalbased level. The former concerns how we subjectively perceive and interpret the musical properties, which terms we choose to describe them and how we use such terms to discuss about music. The latter involves nailing down the objective properties of the musical signals, ranging from those that can be directly computed from the signal to those that can be inferred from

it, such as the rhythmic, tonal, etc. Such two description levels. however, are very separated from each other, which is why MIR research today is focusing exactly on bridging this critical gap. In order to develop MIR applications, the modern approaches must take into account all abstraction levels in the description of musical content. This means that, in addition to considering the signal domain and the semantic domain, such approaches focus on the linking function between them. In this work we follow a schema that involves the formalization of the signal and the semantic domains, as well as the design of the related linking function.

The thesis begins with the discussion of the possible formalization of the signal domain based on the extraction of a feature representation of the musical content. This formalization requires a deep knowledge of the musical properties and the features that are able to capture them. As part of this formalization process, we show an example of application of feature-based analysis which is suitable for the scenario of Networked Music Performance. Through this scenario we show how to estimate some of the musical properties that appear to have an impact on the overall quality of the networked performance experience. We also discuss how deep learning techniques are suitable for automatically extracting (or learning) an effective feature representation of the musical content. We first describe how to design the linking function by means of rulebased techniques, which follow a manually-designed algorithm. More often, however, the link between the two domains is not clear and it is hard to design it by means of a



1. Visualization of the schema for generalization

procedural solution. In such cases, we show how to use machine learning techniques to automatically learn and predict the relation between the two domains. The semantic domain can be instead formalized following two main approaches: the categorical approach, that defines which descriptors are feasible to represent a given song; and the dimensional approach, that also specifies how much the aforementioned descriptors represent the given song. The set of semantic descriptors can be extended by including a semantic similarity among them, either manuallydefined or automatically-inferred; the resulting set of semantic descriptors and similarities is referred to as semantic model. In this regard, we conduct a specific research activity to enrich a manually-defined generic dataset of dimensional descriptors with musicspecific information automatically inferred from users' annotations. We also consider the structure of songs as a special case of formalization of

the semantic domain. After the main components are defined and formalized, we apply the schema to a number of MIRrelated application scenarios of gradually increasing complexity.

The first application scenario involves the analysis and extraction of the song's structure. We rely on deep learning techniques to automatically extract a featurebased representation of the signal domain. By doing so, we address the uncertainty issue about which properties are suitable for describing the musical content for the task of musical structure analysis. Since the semantic domain is well formalized, we are able to apply rule-based techniques from the MIR literature to retrieve the structure. In the second application scenario, we address the detection of bootlegs, i.e., unauthorized recordings of live performances. We formalize the semantic domain by following the categorical approach, which consists of describing a given song as either a bootleg, or an official live performance, or a studio recording. We employ learned features to formalize the signal domain and machine learning techniques to tackle the complexity of the definition of the linking function.

The third application scenario concerns the automatic annotation of the recording of violins with their timbral qualities. The timbral properties of violins are described by rather imprecise semantics, which we formalize through of a set of six dimensional descriptors. Albeit the number of employed descriptors is rather limited, the adoption of a dimensional approach helps us increase the expressiveness of this semantic model. As done in the previous application scenarios, we employ learned features to address the uncertainty in the formalization of the signal domain and machine learning techniques to automatically design the linking function and ultimately assess the timbral qualities of the instrument. In the last scenario we address the definition of a semantic model by considering the ambiguities that commonly occur in natural language. We investigate a formalization of the semantic domain that is able to address the ambiguity issues raised by polysemy, which occurs when descriptors take on different meanings when used within different semantic contexts. We embed the defined semantic model (which is based on overlapping semantic contexts and on context-dependent semantic similarities) in a prototype of music search engine. This prototype represents a novel application that allows users to retrieve musical content by using natural language queries.

Application scenarios				
Application	Signal Domain	Linking Function	Semantic Domain	
Music Structure Analysis	Poorly known: learned features	Rule-Based Techniques	Well known	
Bootleg Detection	Poorly known: learned features	Machine Learning Techniques	Defined but unkown	
Timbral description of violins	Poorly known: learned features	Machine Learning Techniques	Fuzzy: semantic model	
Semantic model for description and retrieval	Musical content	Information Retrieval system	Ambiguous: semantic model	

TIME-GATED SINGLE-PHOTON COUNTING INSTRUMENTATION AND APPLICATIONS

Buttafava Mauro - Supervisor: Prof. Alberto Tosi

The measurement of fast and very faint light signals, down to the singlephoton level, has proven to be an effective tool for many research and industrial applications, like: fluorescent decays measurement in physics, chemistry, biology and material science, luminescence microscopy, diffuse optics spectroscopy, laser ranging and telemetry, just to mention few of them.

Among different kinds of photodetectors, silicon SinglePhoton Avalanche Diodes (SPADs) are becoming increasingly widespread thanks to their good detection efficiency (around 45% at 500 nm and higher than 10% at 830 nm), low noise (hundreds of counts per second at ambient temperature), good timing resolution (few tens of picoseconds) and ease of operation (low bias voltages, no mandatory cooling). Moreover, SPADs can be guickly and effectively gated from OFF to ON condition, by modulating their bias voltage from below to above the breakdown value. A timegated SPAD can be very useful in all of those situations

where it is needed to perform a time filtering of the incoming photons. For example, it is an effective way to reject undesired strong light pulses preceding (or

following) the signal of interest, turning the device ON only during a welldefined time period, when the signal of interest has to be acquired.

The Ph.D. research focused on design and characterization of several single-photon counting instruments, starting from a fastgated SPAD module able to turn ON and OFF a silicon SPAD with transition times down to 110 ps and a maximum repetition rate of 80 MHz (figure 1). Very uniform response of the detector during its ON-time was achieved, with an overall time resolution of less than 30 ps, thanks to a differential front-end circuit. The instrument can operate SPADs of various active area diameters (from 20 μm to 200 μm) built in different technologies (both custom and standard CMOS) and is also capable of free-running operation mode.



1. (a) The complete fast-gated SPAD module, with control unit (left) and detection head (right), with the packaged SPAD on the front. Dimensions are 60 x 60 x 120 mm3 for the detection head and 235 x 250 x 90 mm3 for the control unit. (b) Internal PCB assembly of the detection head, with rigid-flex structure and the TO-8 SPAD package



2. Photographs of the developed TCSPC single-photon detection module (a) and its time-measurement unit (b). The system is housed

in a compact 2-inches long, 1-inch diameter aluminum tube. The round-shaped detection board (c) hosts the SPAD. The time-measurement unit mounts a time-to-digital converter chip, an FPGA and a USB 2.0 communication controller

Thanks to its unique features, this module has been exploited in collaboration with international research groups. Experiments at Istituto Italiano di Tecnologia (in Genoa) successfully proved the advantages of using a time-gated detection in super-resolution Stimulated Emission Depletion (STED) microscopy and with the University of Wisconsin (in Madison) the module was exploited for reconstructing images of hidden objects (outside the field of view), relying on multiple light bounces through the scene.

Following the directions of miniaturization and low-power consumption, a set of extremely compact Time-Correlated Single-Photon Counting (TCSPC)

3. Picture of the developed twowavelength NIRS are on the rear panel.

The instrument is enclosed in an aluminum box of size 200 x 160 x 50 mm3.

> (figure 3). Employing selected laser diodes and highspeed drivers, it is able to deliver optical pulses with 250 ps duration at repetition rates up to 40 MHz, with average output power of 2 mW. It is based on a Silicon Photomultiplier (SiPM) singlephoton detector, and can collect real-time TCSPC waveforms with 10 ps resolution, providing a measurement stability better than ±1% over several hours of operation. Specifically tailored for portability and ease of operation, this system aims to pave the way towards the wide diffusion of timedomain NIRS techniques and instrumentation.



instrument. The front panel hosts three optical fiber ports (one for collection, two for injection), while power supply, USB and trigger IN/OUT connectors

and portable high sensitivity instruments (for monitoring and diagnostic applications). Based on a customdesigned timemeasurement unit, these instruments are able to operate single SPADs in both free-running and fastgating modes, or arrays of multiple SPADs with an overall timing precision down to 50 ps, linearity better than 1.2% of LSB (RMS) and conversion rates up to 5 Mconv/s. Finally, exploiting the same timemeasurement unit, a

systems have been developed

increasing interest in handheld

(figure 2), to answer for the

twowavelength (670 and 830 nm) optical timeresolved NearInfrared Spectroscopy (NIRS) system has been designed, tested and used

NFORMATION TECHNOLOGY

ADVANCED SINGLE-PHOTON DETECTORS IN SEMICONDUCTORS AND SUPERCONDUCTORS

Calandri Niccolo' - Supervisor: Prof. Alberto Tosi

Nowadays many applications of various scientific fields (such as physics, biomedicine, electronics, etc.) need to detect verv faint and fast light signals, down to the singlephoton level. Different kinds of single-photon detectors have been developed in the past, but solidstate detectors have the advantage of high reliability and robustness, thus enabling the development of very compact and portable modulus, while superconductive cryogenic detectors show the best performances achieved for a singlephoton detector but within bulky crvostats.

The present Ph.D. dissertation is focused on the design, development, characterization and modeling of novel near-infrared single-photon detectors and the associated readout architectures. An extensive characterization of different single-photon is proposed, new structures are presented proving outstanding performance. New physical models are envisioned and validated, guaranteeing new interpretations for phenomena often unclear. Thank to successful international collaboration with other research groups, new detectors and read-out circuits are developed, enabling measurements previously considered unfeasible. In the following, a brief description of the main topics discussed in this is reported. The Ph.D. work can be divided in three macro topic:



1. Scanning Electron Microscope image of the InGaAs/InP SPAD cross-section, with layers, Zn diffusion and other information sketched on it. SEM image not in scale and post processed in order to underline the details

InGaAs/InP SPAD: the first part describes the InGaAs/InP single-photon avalanche diode (SPAD) designed, developed and characterized during the Ph.D. activity. The structure is presented (see Figure 1) and the dependence of the main parameters (i.e. photon detection efficiency, noise and time resolution) on the operating conditions (e.g. temperature and excess bias voltage) is analyzed, showing better performance compared to other state-of-the art near-infrared SPADs: . The charge persistence effect is investigated and a new physical model is proposed, proving a new theoretical description for this noise source that arises when the detector is off. Moreover, a comprehensive study of the build-up process is carried out, showing the dependence

of SPAD timing jitter on photon absorption position. Finally It is demonstrated that the optical crosstalk in SPAD arrays can be reduced, without degrading the performance of the device, by means of metal-filled trenches fabricated with focused ion beam (FIB) technique. InAlAs/InP SPAD: the first part presents design and characterization of a InGaAs/ InAIAs SPAD. New MESA structures are designed in collaboration with Sheffield University, where these SPADs ware developed and fabricated. The main parameters of InGaAs/InAlAs detectors are studied (see Figure 2) and the main issues are outlined. New innovative double MESA structures are designed in order to avoid field distortion due to MESA walls and charge trapping.



The SPAD will guarantee better performance than other state-ofthe-art InAlAs-based devices. NbN SNSPD: the last part describes the design and the experimental characterization of SNSPDs and SNSPD cryogenic readout circuits in collaboration with CQC2T (Sydney, Australia) and MIT (Boston, USA). A novel readout technique has been demonstrated for obtaining a single-photon 2. Temporal response at different excess bias voltages, with a 1550 nm pulsed laser focused in the center of the active area. The inset shows the FWHM of the temporal response

detector at telecom wavelength with high count rate and low timing jitter. The time jitter components were analyzed demonstrating a different jitter contribution from geometrical effects. Moreover employing the differential readout we demonstrated new innovative structures able to sense the photon absorption position. We fabricated the first monolithically integrated SNSPD imager (SNSPI) with 20 µm pixels (**see Figure 3**).



3. Superconducting nanowire single-photon imager (SNSPI). a, Architecture concept of the SNSPI. The nanowire transmission line (TL) and the impedance transforms are fabricated in coplanar-waveguide structures, where the ground plane is not shown in the sketch. The SNSPI is biased at a constant current $l_{\rm B}$ and read out by room-temperature amplifiers (amp.). When a photon is absorbed, two pulses propagate towards the nanowire's two ends with delays τ_1 and τ_2 , from which the photon landing location and the photon arrival time can be determined simultaneously. b, A scanning-electron-microscopy image of the top nine rows (15 rows in total) of the SNSPI that meandered from a single 19.7 mm long nanowire over an active imaging area of 286 µm × 193 µm. The dimensions are $q = 13.0 \ \mu m$, $h = 9.7 \ \mu m$, and $p = 5.4 \ \mu m$. The scale bars

in the insert figures are 2 µm (left) and 300 nm (right). c, The single-photon image of a metal mesh taken by the SNSPI with a 780 nm incident light. The image contains 427,905 photon detections. The colour of the map shows the normalized photon counts at each location. The metal mesh has an opening size of 43 µm and a wire diameter of 30 µm, which is indicated from the circular patterns formed by multiple rows.

343

INFORMATION TECHNOLOGY

INTERNET BANKING FRAUD ANALYSIS AND DETECTION

Carminati Michele - Supervisor: Prof. Stefano Zanero

The significant growth of online banking frauds, fueled by the underground economy of malware, raised the need for effective defense systems. Consequently, in last the years, banks have upgraded their security measures to protect online transactions from frauds. However, fraudsters have adapted to these countermeasures developing more and more sophisticated banking Trojans, malicious software that aims at stealing banking credentials and hijacking transactions. This thesis work details our research on Internet banking fraud and financial malware analysis, trying to correlate frauds with malware campaigns to provide a sound detection of the two key aspects of modern online frauds, anomalous transactions and malware activity.

Internet Banking Fraud Analysis and Detection

Internet banking frauds are difficult to analyze and detect because malicious behavior is dynamic and dispersed in large datasets. Despite the importance of the problem, the development of new online banking fraud detection methods is made difficult by the very limited available information due to privacy and security issues related to the banking field. Most state-of-the-art approaches build black-box models that are not very useful in manual investigation, making the process slower, and are usually trained with synthetic datasets, which often implies low detection performance when applied to real data. In addition, none of them provides an in-depth security analysis and manages the "under-training" problem (i.e., users whose number of previous transactions is insufficient to build a consistent profile).

Financial Malware Analysis and Detection

The purpose of banking Trojans is to steal credentials and any other kind of private information. More precisely, they load code in memory and hook the networkrelated Windows APIs used by the browser. With this technique, they can intercept any sensitive data that is processed by the browser even in cases the connection is encrypted. Moreover, this banking Trojan contains a WebInject module that allows the cybercriminals to write scriptable procedures that modify a web page right before it is rendered. Given its flexibility, WebInjectbased malware has become a popular information-stealing mechanism. Different works have been done regarding the analysis

and detection of banking Trojans but most of them are dependent on a specific malware family and require a considerable effort to be constantly adapted to new emerging techniques.

Research Objectives

Our research is inspired and rooted around the idea of constructing user's profiles from historical data to detect suspicious deviations. Instead of focusing on pure detection approaches, we believe that more research efforts are needed toward systems that support investigations. We aim at providing the analyst of a modular framework, able of big data behavioral analysis and where analyst's feedback is put together with machine learning techniques to build a dynamic and auto-adaptive fraud analysis and detection system. This research is made possible thanks to the collaboration with a leading national banking group, which gave us the great opportunity to work on a real dataset. In parallel, our research analyzes banking Trojans that base their attack technique on DOM modifications. We focus on the development of an automatic family-independent banking Trojans analysis platform, able to reconstruct the client-side behavior of financial Trojans that perform WebInjects. By doing

this, we aim to correlate frauds to malware campaigns to provide a sound detection of the two key aspects of modern online frauds: anomalous transactions and malware activity.

Main Contributions

-We provide an **in-depth** analysis of the online banking dataset in our possession and we propose a **framework for** the generation of synthetic transactions and frauds, through a statistical modeling of each feature distribution. Malicious transactions are built in collaboration with domain experts based on fraud scenarios that replicate real-world attacks. -We propose BankSealer, an unsupervised decision support framework for fraud analysis and detection that is based on a combination of different models built on historical user data. First, it quantifies the anomaly of each transaction with respect to the customer local profile. Second, it finds global clusters of customers with similar spending habits. Third, it uses a temporal threshold system that measures the anomaly of the current spending pattern of each customer with respect to his or her past spending behavior. With this threefold profiling approach, BankSealer is adaptive to non-stationary sources, it mitigates the under-training problem and the evolution of user's spending habit over time. At runtime, it supports analysts by ranking new transactions that deviate from the learned profiles, with an output that has a clear statistical meaning. An

in-depth performance evaluation on a real dataset against fraud scenarios that replicate real-world attacks showed that BankSealer's approach correctly ranks complex frauds with up to 98% detection rate. In parallel, we evaluate the influence on detection performance of the granularity at which the spending habits are modeled. We compare usercentric modeling, which builds a model for each user, with systemcentric modeling, which builds a model for the whole system. Finally, we assess the robustness of BankSealer against mimicry attacks, which allow an attacker, equipped with the system knowledge (i.e., is aware of how spending habits are modeled), to cloak its frauds to avoid detection. -We analyze the distribution of transactions in the time **domain** to detect "legitimatelooking" frauds committed by repeatedly stealing small amounts of money over time. We propose a **temporal framework** that aims at detecting frauds by exploiting a precise modeling of recurrent vs. non-recurrent spending patterns. To do so, our solution apply signalprocessing techniques to extract temporal patterns "hidden" in the time series, automatically learns end user's temporal profiles, by means of spending patterns, and guantifies the deviation of the current user's behavior from the average one learned. Our solution outperforms BankSealer temporal profile by detecting 40% more frauds. -We propose a supervised

-we propose a **supervised** learning framework, based on a **multi-objective genetic** algorithm, that can exploit the knowledge of the analysts through their feedbacks to **automatically tune** BankSealer's **parameters** and **improve** its **detection** performances (i.e., push real frauds higher in the ranking). The evaluation showed that this module can increase the detection performance of BankSealer up to a factor of 35%.

-We propose a **dynamic** supervised learning framework, based on a Random Forest classifier, able to learn and detect frauds pattern. We evaluate this system against stateof-the-art literature works keeping into consideration the detection performances and a cost function that depends on misclassifications. Our solution shows overall better performances detecting up to 87% of frauds with a false positive rate of 0.3%, always keeping a low-cost function value.

-We propose a framework able to analyze advanced information stealers and generate behavioral signatures of the WebInject components. Its approach, based on memory forensic techniques and "webpage differential analysis" that consists in analyzing visible DOM modifications that information stealers cause inside the web-pages and extracting these modifications through a comparison of the web-pages with the original ones. We evaluate our solution against real-world, online websites and a dataset of distinct variants of financial Trojans. We show that our approach correctly recognizes known variants of WebInject-based malware and successfully extracts the WebInject targets.

345

MODEL PREDICTIVE CONTROL IN MANUFACTURING PLANTS

346

Cataldo Andrea - Supervisor: Prof. Riccardo Scattolini

Over the years manufacturing industries have become more and more demanding due to the complexity of production processes and to tighter rules and regulations. In the global market, enhancing the efficiency and productivity of manufacturing systems is mandatory to maintain high levels of competitiveness. Moreover the energy efficiency of manufacturing production systems is becoming a topic of paramount interest for many reasons, such as the need to minimize the energy consumption of industrial plants, to resize the factory energy supply infrastructures and to limit the CO2 emissions. Among the main issues in these fields, the development of advanced control strategies, such as Model Predictive Control (MPC), has an important role for the solution of many significant problems such as lotsizing, scheduling, packing, inventory, resource allocation, energy efficiency. MPC is a control method nowadays widely used in the process industry in view of its capability of dealing with complex systems. This is due to the possibility with MPC to enable reformulation of control problems in to optimisation ones, which gives the opportunity to explicitly add constraints on the control inputs and the controlled variables. Nevertheless. model

predictive controllers are mostly used with continuously varying systems, while they are less frequently applied to discreteevent systems, typical of the manufacturing field. The scarce use of MPC for discrete-event systems can be explained by the fact that they are characterised by integer or Boolean decision variables, so that the use of MPC could lead to large combinatorial optimisation problems to be solved on-line, which is seen as a computational bottleneck. In this thesis, in order to prove the applicability of MPC to control problems typical of manufacturing systems, different industrial applications have been taken into account and the adequacy of MPC in terms of "easy to design and use" and performances has been proven.

Firstly, in Chapter 2 the efficient routing of the pallets in networks made by machines and transportation lines is studied in order to avoid bottlenecks, starvation, congestion, and to maximize throughputs. However, the design of optimal routing strategies is difficult due to the combinatorial nature of the problem and the implemented control laws are often based on heuristic logic rules tuned by means of simulation studies. In this scenario, MPC has been

applied to a de-manufacturing transport line in which a multipallet, dynamic multi-target problem has to be solved. The dynamic system of the transport line has been formulated as a Mixed Logical Dynamical (MLD) system. Then, a performance index has been defined and the optimal control sequence has been recursively computed and applied according to a receding horizon approach. The MPC algorithm developed has been used to control the transport line placed at the Institute of Industrial Technology and Automation -National Research Council (ITIA-CNR). Secondly, in Chapter 3 the problem of optimizing on-line the production scheduling and the buffer management of a multipleline production plant composed by L machines Mi, i = 1,...,L, which can operate at different speeds corresponding to different energy demands, has been considered. The path from a common source node, where the part to be processed is assumed to be always available, to each machine may differ in the number of buffer nodes and the energy required to move the part along these transportation lines must be suitably considered in the computation of the overall energy consumption. Therefore, the control problem

consists of computing, at each sampling instant, the sequence of commands to be applied to the transportation lines and the processing speed of the machines in order to optimize the throughput of the system and to limit the overall energy consumption. In Chapter 4 a laboratory stacker crane, a specific type of Automated Storage / Retrieval System (AS/RS), has been considered. The AS/RS system has been modeled in terms of MLD system and the control problem has been reformulated as an integer linear programming problem. Finally, in order to be able to compute the energy consumption for a manufacturing plant, in Chapter 5 a specific new energy consumption computation method has been defined and validated, named aCtuatorS Methodology (CSM), based on the Discrete Event System (DES) approach for the computation of the energy consumption of discrete systems, i.e. systems where the energy consumption is mainly due to the on/off switching of the actuators governed by the control logic. In order to evaluate the MPC

performances in the considered applications, many experiments have been performed. Concerning the de-manufacturing transport line, the experimental results show the very satisfactory behavior of the proposed algorithm when applied both to the discrete event simulation model and to the real system. In the problem related to the multiple-line production plant, the simulation results show that the algorithm is highly flexible and its performance can be easily adapted to obtain different behaviours by means of the tuning of simple and easy-tounderstand parameters of the cost function. Moreover, the proposed method allows to cope with dynamic changes of the minimum production and maximum absorbed power and to choose the constraints to be violated in case of infeasibility, features that are very difficult to be achieved with standard scheduling techniques based on the solution of MILP problems or on heuristics. Finally, the laboratory stacker crane example witness the potentialities of this control method also for this class of problems. According to the considered manufacturing applications, future research activity could be respectively aimed at: customizing the algorithm described in Chapter 2 to production lines with operating machines whose working function settings can be dynamically changed in order to further optimize the

production line efficiency;

non deterministic behaviours of the machines described in Chapter 3; improving the modelling phase by reducing the complexity of the AS/RS model considered in Chapter 4 and decreasing the required computation time. The focus hereby should be on the reduction of the number of integer variables, since these determine the complexity of an integer linear programming problem. The current model can easily be extended, e.g. to a form with multiple final storage nodes. Next to these improvements and extensions it would be interesting to compare the MPC method to other control methods such as time instant optimisation MPC and heuristics.

including constraints on the early

production of parts or considering

ANALYSIS, CONTROL AND FORECAST OF SCHISTOSOMIASIS SPATIOTEMPORAL DYNAMICS VIA NETWORK MODELLING

Ciddio Manuela - Supervisor: Prof. Marino Gatto

Waterborne diseases belong to a group of infections caused by pathogenic microorganisms that most commonly are transmitted in contaminated freshwater. Typically, infection occurs during bathing, washing, drinking, or ingesting contaminated food. For these reasons, they are especially endemic among low-income populations in developing regions and represent a serious public health problem where people lack access to clean water and adequate health care resources. In this work, novel models are developed for the transmission of schistosomiasis, a parasitic, waterrelated disease that is prevalent in tropical and subtropical areas of the world. Current measures for schistosomiasis control are principally focused on preventive

chemotherapy (Fig. 1), which however does not confer permanent immunity to humans. Schistosomiasis is caused by a snail-transmitted trematode. which may infect the urinary tract or the intestines. The specific aim of this work is to include in new modelling frameworks some important features of the transmission cycle that are still underestimated, yet crucial, such as the impact of social interactions and physical interconnections between populations. Therefore, the developed models incorporate both the biological complexity associated with the parasite's life cycle (including secondary hosts, Fig. 2) and the mechanisms that influence the spatiotemporal dynamics of the disease.



1. Number of individuals requiring preventive chemotherapy (PC) for schistosomiasis in 2015.



2. Schistosoma life cycle.

First, the mechanisms that drive the temporal variability of disease severity and prevalence are explored introducing nonlinearities in demographic and epidemiological dynamics. Then, the impacts of different sources of local and spatial heterogeneity are investigated, together with their implications on effectiveness of possible intervention strategies. These models are used to analyze transmission patterns in different settings, ranging from purely theoretical ones to real case studies. In particular, the analysis is performed by integrating proxies of human mobility (inferred from a very large database of mobile phone traces) with a geospatial analysis which includes georeferenced data on demography, water supply/sanitation, and urogenital



3. Fluxes of individuals as obtained from Call Detail Records (on the left) and simulated schistosomiasis prevalence across Senegal (on the right)

schistosomiasis prevalence. The developed modelling frameworks are then applied to Senegal (**Fig. 3**), where the urogenital form of the infection is widespread. In the country, schistosomiasis represents a major health problem, being the third disease in terms of years lived with disability. A spatially explicit model is first applied to medium-to-large spatial scales, at which human mobility is retained as the main mechanism for the spatial spread of the disease. Over finer spatial scales, instead, connectivity via hydrological transport and snail dispersal increases the risk of disease propagation. A multidimensional network model accounting for both social and environmental connectivity is thus applied to a

set of connected villages in the area of the Lower Basin of the Senegal River.

Results are presented and discussed in the perspective of using epidemiological models as tools for disease control. The spatial structure is used to quantitatively evaluate the impacts of different sanitary and humanitarian efforts, such as preventing exposure and contamination of environmental freshwater, spreading awareness about disease transmission, and biological control of snail hosts. In this respect, a quantitative decision-support tool can help guide resource allocation in the fight against the disease, identifying the focal hotspots of disease transmission to reduce its impact on society.

As most significant contribution to the state-of-the-art in the field of schistosomiasis transmission modelling, this work presents a comprehensive approach to the epidemiological problem, underlying the importance of integrating different perspectives, from environmental to social and behavioural. On the one hand, the intermediate snail host population dynamics play an important role in generating variations of disease transmission patterns over time, with potential implications for long-term disease dynamics. On the other, socioeconomic conditions and water availability are fundamental in the definition of the infection risk of different communities, in particular with respect to human habits and quality of snails habitats. In this framework, anonymized mobile phone data represent a useful source of information as a proxy of human movements which, together with hydrologically mediated processes, are crucial in the spread, persistence and infection intensity of infectious diseases.

349

NFORMATION TECHNOLOGY

MODELLING RESIDENTIAL WATER CONSUMERS' BEHAVIOR - FROM SMART METERED DATA TO DEMAND MANAGEMENT

Cominola Andrea - Supervisor: Prof. Andrea Castelletti

Water demand in the residential sector is forecasted to grow in the next decades, under human population growth, urbanization, and climate change. Water supply expansion through the construction of new storage and distribution infrastructures might be an option to escape water stress in some situations. Yet, geographical or financial limitations, as well as social, environmental and economic impacts of infrastructural interventions, largely restrict such options in most countries. Worldwide experiences have been proving that demand management strategies based on technological, financial, legislative, maintenance, and educational mechanisms can complement supply-side management to meet future demands, potentially leading to significant reductions in residential water consumption, as well as reducing short- and long-term utilities' costs. However, the design of effective demand-side management strategies relies on our understanding of consumers' behaviors. Indeed, we can think of water consumers as agents who decide when and how much water to use during the day, as well as for which specific enduse, depending on their needs. Consequently, the achievement of demand management objectives

depends on the aggregate effect of actions undertaken by several independent and diverse agents. For this reason, models that quantitatively describe how water demand is influenced and varies in relation to exogenous uncontrolled drivers, water consumers' and household characteristics, and demand management actions are key to explore water users' response to alternative water demand management strategies, ultimately supporting strategic planning and policy design. On this regard, the advent of smart meters in the late 1990s made available new water consumption data at very high spatial (household) and temporal (from several minutes up to few seconds) resolution, enabling the development and application of data analytics tools and mathematical models to accurately characterize sub-daily water consumption behaviors of individual users, as well as enduse consumption profiles. These constitute essential information for designing and evaluating consumer-tailored WDMSs, which, in turn, can more effectively modify the users' attitude favoring water saving behaviors. While data from traditional meters, usually collected with time sampling frequencies of a few months or a year, were mainly used only for

billing purposes, high-resolution metering is making available an amount of data that was not available just a few years ago. and is going to increase in the next future with the large scale deployment of smart metering systems. Yet, apart from the technical challenges posed by the storage and management of big datasets, a paramount challenge present in all the fields where big data and big data analysis are involved is how to turn these data into useful information. The main goal of this thesis is advancing data analysis and mathematical models to extract information on water consumers' behavior out of smart metered data, ultimately informing and proposing recommendations to customized demand management. More specifically, in this thesis we contribute novel methodologies for profiling, analyzing, and modeling residential water consumers based on high temporal and spatial resolution data of residential water consumption, and also coupled with several qualitative and quantitative data describing consumers' psychographic features. The innovative aspects of this research are two-fold. First, innovation is present in the methodologies we propose, partially developed by adapting,

further extending and integrating data mining techniques from several data mining and modelling research fields, i.e., signal processing, information extraction, data reduction, and customer behaviour studies. Secondly, the set of modelling tools we develop in this research cover all the phases of residential water demand modeling and management, from big smart metered data processing, to enduses classification, user modelling and customer segmentation and recommendations to water (and electricity) demand management. Such integration can constitute a suite of data analysis and modelling tools supporting the decisional processes needed for the design of customized demand-side management strategies. While the main focus is on water demand modelling and management, in this thesis we either assess the inter-portability of the methodologies between the water and energy fields of applications, or present integrated water-energy applications. The first outcome of this research is the first published comprehensive review of more than 130 studies, published between 1990 and 2015, on high and low resolution residential water demand modelling and management. Apart from being the first comprehensive review on this topic, it proposes also a methodological framework for the classification and comparison of studies within that field, and a "roadmap" listing the main research challenges that application of smart metering technologies will face in the near

future.

Secondly, we developed two novel Non-Intrusive Load Monitoring algorithms demonstrated to achieve high electric power and water disaggregation performance. The first algorithm addresses the disaggregation problem as a least-square error minimization problem, with a convex penalty term aiming at enforcing the disaggregate signals to be piecewise constant over the time. The second algorithm, called HSID, is based on the combination of FHMMs, which provides an initial approximation of the enduse trajectories, and Iterative Subsequence Dynamic Time Warping (ISDTW), which processes the end-use trajectories in order to match the typical power/water consumption pattern of each appliance. The application of such algorithms on both real-world and synthetically generated water and energy data provided insights on the robustness of the two methods to signal noise, data resolution, as well as their portability between the two application fields of water and energy. Thirdly, we developed two novel descriptive and predictive modelling tools to infer water consumers' habits and routines, as well as identify the most relevant determinants of their water consuming or saving behaviors, at the household level. More specifically, we first develop a descriptive modeling procedure, based on a combination of clustering and principal component analysis, which allows performing water users' segmentation on the basis of their routines, automatically

identified from smart metered consumption data. Secondly, we propose a predictive modeling approach, consisting of a two-step procedure that extracts the most relevant determinants of users' consumption and identifies and learns a predictive model of water consumers' profile. Finally, we implemented a threephase data-mining procedure composed of data dimensionality reduction, customer segmentation, and factor mapping to capture heterogeneous water-electricity consumption profiles, highlighting differences between daily time-ofuse of water and electricity, and allowing for the characterization of users based on psychographic and behavioral factors. Applications of the developed methods onto synthetic data, as well as onto real-world case studies in Switzerland, Australia, and USA, demonstrated that our tools constitute important progress for an effective and efficient exploitation of smart metering data to develop models of water-energy users' behavior at the household scale, and advance the customization of demandside management strategies. Moreover, we believe that the technological outcomes and policy implications of our research can support utilities, planners, meter producers, and data scientists when planning new smart metering deployments.

ON UNIFIED STREAM REASONING

Dell'Aglio Daniele - Supervisor: Prof. Emanuele Della Valle

The real-time integration of huge volumes of dynamic data from heterogeneous sources is getting more and more attention, as the number of data stream sources is keeping growing and changing at very high pace. Cities and the Internet of Things are perfect illustrations of such need. For instance, in the urban setting, semantic interpretation of road sensors and social networks can supply (directly and indirectly) continuous and up-to-date information about the traffic causes and their impacts, the progress of city-scale events or the trending activities around a user. While Data Stream and Event Processing deal with data streams and reactiveness, reasoning is a potential solution for the data heterogeneity: ontologies are key to access the data streams from the different sources and to make explicit hidden information. Stream Reasoning aims at bringing together those areas, with techniques to perform continuous reasoning tasks over data streams. Even if the Stream Reasoning goal can be easily stated, bridging those two worlds together is not straightforward: there are multiple ways on which these approaches can be combined. The combination can bring problems at both theoretical level – to create comprehensive data and processing models - and practical level - to guarantee the reactiveness, satisfying the realtime computation requirements. Since 2008, Stream Reasoning has moved forward. Different research groups proposed (i) data models and vocabularies to capture data streams through RDF and ontologies; (ii) continuous query models, languages and prototypes, inspired by SPARQL and collected under the RDF Stream Processing (RSP) label; (iii) extensions of the reasoning tasks over streams, as consistency check and closure; (iv) applications built on top of the aforementioned results. The growth of Stream Reasoning brought several positive results, but also some drawbacks that can bring new challenges to be addressed in the future research. Most of the solutions have been designed and proposed with regards to different use cases and input data. Rigorous comparisons are missing, and as a result, the connections among current approaches are missing as well. This fact drastically limits the possibility to determine when one approach is preferable to another one. Moreover, it is an obstacle to interoperability, making hard, if not impossible, to let these different solutions interoperate in

complex architectures. Even when solutions seem to be comparable, i.e. they operate on the same input, the query languages are similar and consequently the user may expect the same outputs, we observe that actually it does not happen. In other words, two RSP engines with similar guery models can produce different results given the same continuous query that processes the same data stream. And, surprisingly, the same system can produce different results while processing the same input in different runs. The main goal of this research is to study how to unify current Stream Reasoning results. A comprehensive framework to describe continuous query answering and reasoning over streams of RDF data is missing. In addition to comparison and interoperability, mentioned above, such framework would be a basis to study RDF Stream Processing related problems and a candidate to build a standard RSP query language. To investigate this problem, hard to tackle in its entirety, we asked ourselves two research questions, related to different (but still connected) parts of the problem. The first question states: RQ.1: How can the behaviour of existing RDF Stream Processing

systems be captured and compared when reasoning processes are not involved? At the beginning of the research, we work under the assumption of absence of reasoning tasks. SPARQL, the language for querying RDF repositories, defines this setting as query answering under the simple entailment regime. The goal is to verify if it is possible to capture in a common formalism the semantics of existing RDF Stream Processing engines: if we do not succeed in the absence of inference processes, it follows that we cannot do it when we relax this constraint. The answer to this question requires putting together elements from Data Stream Processing, Event Processing, SPARQL semantics and the current state of the art on RDF Stream Processing. If the answer is affirmative, it is possible to design a reference model that captures the guery models of the different systems. At this point, we can move a step forward and consider entailment regimes different from the simple one. We do it through the following question: RQ.2: What is the correct behaviour of a continuous query engine while processing a semantic stream under entailment regimes? To address the question, we investigate how SPARQL entailment regime can be used in this context of continuous query answering over semantics streams. The main outcomes of this research can be summarised in three results: RSEP-QL, CSRBench and TripleWave.

RSEP-QL. RSEP-QL is a reference model to continuously query RDF streams. It extends SPAROL by introducing operators to process and produce streaming data. RSEP-QL is built in a modular way. The first block, RSP-QL, is a model that defines the continuous evaluation semantics; the sliding windows – a typical data Stream Processing operator to manage subsequences of the stream and limit the amount of data to be gueried; a new dataset that extends SPARQL dataset to include also streams; a set of streaming operators – used to build the output stream. The model extends the SPARQL semantics and is inspired by two Stream Processing models, CQL and SECRET. The second block, RSEP-QL, extends RSP-QL by adding Event Processing operators. That means, it defines (i) a new window operator, the landmark window, used to have a large view over the underlying stream; and (ii) event pattern operators to identify complex events on the stream. The evaluation semantics of the new operators is built on the results of RSP-QL. Finally, the third block defines the entailment regimes in RSEP-QL. It

entailment regimes in RSEP-QL. It extends the SPARQL entailment regimes to take into account the presence of streams in the dataset and of event pattern operators. We adopt the ontology stream notion to capture the portion of the stream to be considered and to describe the inference processes w.r.t. a conceptual model. At the same time, RSEP-QL entailment regimes

maintain the compatibility with

description logics without time extensions, e.g. EL++ and DL-Lite. This fact is important with regards to computational complexity. **CSRBench.** To supply evidence on the fact that RSEP-QL captures the semantics of existing systems, we use it in the context of correctness assessment. We show that RSP-QL captures the semantics of a set of existent RDF Stream Processing solutions, and we propose CSRBench, a benchmark that extends SRBench with correctness tests. We also designed and implemented an open source framework to automatically run the correctness tests. TripleWave. To fill an existent gap on RDF stream availability on the Web, the last contribution is TripleWave, a reusable and generic tool to spread and exchange RDF streams on the Web. The features of TripleWave have been derived from requirements of real use-cases, and consider a diverse set of scenarios, independent of any specific RSP implementation. TripleWave is fed with existing Web streams (e.g. Twitter and Wikipedia streams). It can also be invoked through both pull- and push-based mechanisms, thus enabling RSP engines to register and receive data from TripleWave automatically.

353

INFORMATION TECHNOLOGY

MEMS GYROSCOPES BASED ON NEMS PIEZORESISTIVE ELEMENTS: A COMPREHENSIVE ANALYSIS OF DESIGN, RELIABILITY AND EXPERIMENTS

Dellea Stefano - Supervisor: Prof. Giacomo Langfelder

The Thesis discusses the design, the characterization and the development of innovative MEMS gyroscopes based on piezoresistive NEMS sensing elements. At the same time, the Thesis provides a characterization and a knowledge improvement on several aspects of the used, innovative, piezoresistive technology.

The Thesis introduces in Chapter 1 the piezoresistive technology that provides nano sensing elements integrated in a MEMS device. Then, the basic principles of the rate gyroscope and the piezoresistive effect are presented and fused into the fundamental equations of the piezoresistive rate gyroscope.

Then, Chapter 2 reports the mechanical design of the consumer and medical gyroscopes. A motivated list of design choices leads to the definition of the structures. The devices are coupled with electronic boards designed for the characterization of the gyroscopes.

In particular, the Thesis presents the design, the layout and a detailed characterization of triaxial gyroscopes for consumer and for medical applications. These two different application fields share the same requirements: low power consumption and miniaturization. In parallel with the design of the gyroscopes, dedicated test structures are developed in the framework of the Thesis and thus described in this work. These devices enable particular characterization of mechanical and technological issues of the NEMS gauges, that are the peculiar elements of the technology.

The characterization tests performed on the gyroscopes are described in Chapter 3. The obtained results satisfy the defined performance given by the targets specifications for both applications.

Then, an exhaustive experimental campaign for the study of the device reliability is reported, as well as the investigations among some fundamental properties of the technology. Then, the performance in terms of linearity, reliability, critical tensile and compressive stress and fatigue are discussed. Chapter 4 presents the design of an innovative fully piezoresistive sensed gyroscope. The

advantages related to this new device are critically discussed. Chapter 5 shows the perspectives of the piezoresistive gyroscopes and the possibility to become, from a promising device, a new standard. Finally, the conclusions

are drawn. The end of the Thesis is the presentation of a gyroscope which includes a further step of innovation: the nano gauge transduction is here used not only for the sense mode, but also for the drive motion detection, showing angle random walk in the order of few hundreds dps/√Hz with voltage applied at the gyroscope lower than 1.5 V. Such a level of voltage is considerably lower, at least one order of magnitude, than for a capacitive gyroscope with the same noise performance.

A MIDDLEWARE FOR PEER-TO-PEER INTERACTION OF SMART THINGS

Derakhshan Naser - Supervisor: Prof. Luciano Baresi

In these years, we are assisting to the rapid development of small, low-power, and low-cost wireless computation/communication devices, which have served as enablers for the so-called "Internet of Things" (IoT). Within the Internet of Things, devices adopt wireless communication to cooperate to provide services and added value to users. probably in large-scale dynamic environments. Devices that can connect to the Internet can exploit cloud-based services to enable interaction in an IoT scenario. However, an Internet connection might not always be available, or it might be too expensive to use. It is clear that within the Internet of Things we cannot impose the requirement of having an always-on functional Internet connection. Instead, we need to shift our focus to a new wave of interaction called "proximitybased interactions". In this kind of interactions, applications operate in an infrastructure-less scenario, possibly interacting in a peer-to-peer (P2P) manner. Among P2P communication protocols for mobile devices, Wi-Fi Direct has recently gained attention. Although many works have already tried to exploit Wi-Fi Direct in social interactions among proximal smart devices, there is no work to the best of

our knowledge that attempts to exploit Wi-Fi Direct in large-scale dynamic application domains. This thesis bridges this gap by proposing a middleware infrastructure, called "MAGNET", a novel middleware infrastructure that exploits Wi-Fi Direct to provide a reliable and stable communication means for large numbers of mobile devices. This self-organizing middleware abstracts the multihop communication process by autonomously maintaining connectivity among devices. MAGNET also provides a discovery mechanism that exploits the MAC address of the different devices or the services they offer. The layered view of Figure 1 illustrates the functionality embedded in MAGNET and exemplifies the

components installed on each Android device. The architecture consists of three main abstraction layers. Each layer comprises two main components: a user-configurable part and an autonomous part. Most of the functionality provided by MAGNET is carried out autonomously to keep a Wi-Fi Direct network connected without any user intervention. Nevertheless, some user preferences and configurations must be considered.

The first layer is a platformdependent layer that abstracts the device. This layer consists of two main components. Component Radio Controller utilizes the Wi-Fi and Wi-Fi Direct APIs provided by Android to perform discovery and communication. Component







2. Overview of WiDiSi

Dynamic Intention Calculator is in charge of computing the value of the aforementioned intention, and providing it to the radio controller when needed. The second layer is in charge of ensuring that the device is part of a Wi-Fi Direct group in any circumstance. For instance, if the device moves close to a group, moves away from it, or must face saturation, this layer works to keep the node connected. The second layer is in charge of ensuring that the device is part of a Wi-Fi Direct group in any circumstance. For instance, if the device moves close to a group, moves away from it, or must face saturation, this layer works to keep the node connected. Component "Group Bootstrapping Manager" is responsible for the actual creation of groups. It is also in charge of managing connection requests, failures, new arrivals, saturations, and empty groups (owners only).

The third layer is in charge of inter-connecting Wi-Fi Direct groups and providing multi-hop

connectivity between devices. The main goal of this layer is to ensure that a group is connected to as many groups as possible in any circumstance. Since Android's Wi-Fi Direct does not allow a single device to be both the owner of a group and a client of another one, we cannot connect groups through Wi-Fi Direct, but we must exploit other ``legacy" protocols like standard Wi-Fi or Bluetooth. Since a Wi-Fi Direct group can also host legacy Wi-Fi clients, which see the owner as a standard access point, MAGNET exploits Wi-Fi connectivity, which results in better integration compared to using Bluetooth. To evaluate the effectiveness of

MAGNET, a Wi-Fi Direct simulator has also been developed, called "WiDiSi". WiDiSi is a researchoriented prototype simulator for Wi-Fi Direct networks. WiDiSi is provided as an extension of PeerSim, a widely-used, opensource simulation framework for large-scale peer to peer networks. WiDiSi's main goal is to allow Android Wi-Fi direct applications PhD Yearbook | 2017

runs a set of **Applications**. These applications represent the behaviors that the node would have on a real Android device. Each node also includes the **Wi-Fi Direct Interface**. This interface is what implements the Wi-Fi P2P protocol inside the simulator; more specifically it pretends to be the Android implementation of the specification. On top of the container we have various components that are used to setup the simulation, manage it, and analyze its behavior through logging and advanced visualization.

The proposed solution has been tested on real eighteen Android devices and on the simulation environment. Realistic scenarios are utilized in the thesis to showcase and evaluate the key features of the proposed solution. The evaluation results illustrate the effectiveness of such a solution in providing a stable communication between many mobile devices using Wi-Fi Direct.

Di Rienzo Antonella - Supervisor: Prof. Franca Garzotto

Co-Advisor: Ing. Cristina Frà

Smart Spaces (defined as environments having sensing and adaptation capabilities, a multiplicity of connected devices and various forms of interaction) open up an enormous potential for new added value services in many domains.

Designing applications for Smart Spaces requires a multi-faceted research that explores people's needs with regard to technological feasibility, as well as limitations and opportunities of novel interaction paradigms.

This thesis discusses how novel developments in ubiquitous computing, findings from empirical studies, and a usercentered design approach help to shape new added-value services in Smart Spaces. Our main focus is on novel interaction paradigms across different devices and interfaces (e.g., full-body interaction), exploring how their integration leads to new solutions that benefit the intended user, and how UX can be personalized to the specific needs and preferences of each single person or target group.

To this end, various design concepts and alternatives have been defined and analyzed in different domains under the general framework in , and empirical studies have led to explore how users interact with interconnected devices in Smart Spaces. Attention has been paid on the usage of alternative control and interaction mechanisms. effectiveness of the approach, we have created some domainspecific technological set-ups, exploiting prototypes built in collaboration with national and



1. General Framework

A number of technological solutions have been developed to empower control and interaction that (i) encompass design patterns of interaction that address all aspects of the services as perceived by users; (ii) take into account context changing and awareness; (iii) support interface adaptation, exploiting the knowledge about the current user, the interaction flow, and his/ her tasks; (iv) master "information overload" by offering the proper content on the right channel at the proper time. To empirically measure the

international partners, and we have collected data on users' behaviors in ecological settings - public and semi-public spaces: a Department Entrance Hall (the Department of Electronics, Information and Bioengineering at Politecnico di Milano); two fashion stores in Milan (see Figure 2 and Figure 3); a smart office building (Open House Innovation in Helsinki); a mobile telephony store (TIM Shop in Milan). Our experiences gained through field studies highlight issues that need to be addressed by any public social space that is



2. Large Interactive Display in a fashion store



3. Smart Lighting in retail shop windows

intended to sense and respond to the people and activities taking place nearby in a proactive manner. These issues include the identification/classification of people (or objects) in the vicinity of the ICT tools, the acquisition and maintenance of profiles about people, selecting appropriate content and designing useful interaction models. Reconsidering the core contributions of the presented research, on the one hand in regards to the research community of HCI and on the other hand contributions targeted to beneficiaries beyond the academic environment, they target four different levels and can be classified into the following categories:

- Theoretical contributions.
- Technological contributions.
- (Interaction) design specific contributions.
- Empirical contributions.

Based on the analysis of related literature and prior art, we developed an application domain classification model. On the one hand, this model is a valuable contribution to the HCI research community. In an academic context, such a model can be used to identify comparable solutions to a novel one which is helpful, for instance, when searching for baseline approaches within the context of a comparative study. On the other hand, this model can be useful to practitioners in the field of application and interaction design, who are working on solutions suitable for smart interactive social spaces. The second, and one of the core contributions of this thesis, is a set of multiple technological insights and advances for HCI systems, which includes software aspects introducing novel ways to apply existing technologies in the context of different domains. The third aspect which describes a contribution of this thesis are insights regarding interaction design for smart social spaces. This includes best-practices as well as pros and cons of interaction design patterns. Finally, this thesis contributes empirical insights and study results that help to develop a deeper understanding of smart social spaces interaction techniques. These empirical insights are based on multiple user studies (including

techniques. These empirical insights are based on multiple user studies (including quantitative and qualitative study designs), for which each technique was implemented as fully functional prototype which allowed examining the solution on real field environments. 359

NFORMATION TECHNOLOGY

ENABLING POWER-AWARENESS FOR MULTI-TENANT SYSTEMS

Ferroni Matteo - Supervisor: Prof. Marco Domenico Santambrogio

Power consumption has become a major concern for almost every digital system: from the smallest embedded devices to the biggest data centers, energy and power budgets are always constraining the performance of the system. On the one hand, this is the case of battery-powered devices like smartphones and tablets: their batteries have to be small and lightweight, as the device has to be portable, but this of course constrains their potentials. On the other hand, power consumption remains an open issue also in those contexts that do not involve batteries: for instance. data centers providers aim to reduce it as much as possible to decrease operating costs and to improve system reliability. Power consumption consolidation through applications colocation and migration then becomes critical for the Cloud Computing paradigm.

360

Even though these use cases are quite different from multiple perspectives (e.g., computational purposes, domain fields, target users and so on), they share an important characteristic: their consumption is strongly affected by their current "working regime", from idle to heavy-load conditions. Moreover, they both need to deal with hardware heterogeneity, software multitenancy and input variability over time. It is then difficult to make accurate predictions on the power consumed by the whole system over time, as it is subject to constantly changing operating conditions: under this perspective, any attempt to optimize these system then becomes far from trivial.

The aforementioned scenarios have been extensively discussed throughout the thesis work, addressing the following two questions: (1) "how much power is a system going to consume, given certain working conditions?" and (2) "is it possible to control a system to consume less power, still satisfying its functional requirements?". An important aspect we need to face is the fact that heterogeneity, multi-tenancy and time variability make a comprehensive profiling of these systems unfeasible: in fact, it is not reasonable to explore all the possible system's "working regimes" before the final deployment of the system, given the huge number of hardware features available, the possible combinations of concurrently running applications, together with their input fluctuations. However, we can tackle these uncertainties starting from the assumption that the system will probably behave in the future as it did in the past: this suggests that the system can learn

from experience, trying to improve and optimize its performance in its real working conditions. This preliminary analysis of the context leads to the following outcomes: (1) the best strategy is to observe the behavior of the system at runtime, during its real working conditions; (2) we can then learn from the experience which are the variables that mostly affect the system's power consumption; (3) this knowledge can then be used to decide what to do and *actuate* the best strategy with respect to performance goals and power constraints. These steps must be performed in automation throughout the whole lifetime of the system. In literature, this approach is known with the name of Observe-Decide-Act (ODA) control loop: here, I introduce an additional Learning phase to decouple the knowledge learnt (i.e., self-awareness) and the decision phase (i.e., selfexpression), thus leading to an architecture style can then be described as an OLDA control loop. My thesis work exploits this approach to enable powerawareness for a wide range of autonomous agents: I designed a general power modeling methodology that can be tailored on the actual system's features, extracting a specific power model able to describe and predict the

future behavior of the observed entity. This methodology is meant to be provided in an "asa-service" fashion: at first, the target system is instrumented to collect power metrics and workload statistics in its real usage context (Observe phase); then, the collected measurements are sent to a remote server, where data is processed using well known techniques (e.g., Principal Components Analysis, Markov Decision Chains, ARX models, etc., *Learning* phase); finally, an accurate power model is built as a function of the metrics monitored on the instrumented system. This information will support a power-aware orchestrator in the estimation of the best trade-off between global performance and power consumption (Decide and Act phases), still providing the required Quality-of-Service (QoS) to the guests applications. I moved my first steps in the field of power-aware systems with my M.Sc. thesis: it consisted in the MPower project, a mobile application able to predict how long the battery of a smartphone is going to last, given the current hardware status of the device. Durgin my Ph.D., I generalized the same concepts towards a comprehensive and general methodology: the observed system does not need to be a smartphone but it could be a generic system, i.e., an "agent", that wants to become power-aware or, in a wider sense, "resource-aware". The generalized data-driven methodology for resource consumption modeling has then been implemented into MARC, a Cloud-service platform designed

to Model and Analyze Resource Consumption trends (MARC), supporting the aforementioned "as-a-service" paradigm. Regression tests show how it is able to reproduce the same precision of the results obtained with the MPower application, thus showing how the same methodology has been generalized and abstracted consistently. This generalization allowed to bring power-awareness into a completely different context: power consumption models for virtual machines in a multi-tenant virtualized system. As I needed to observe the behavior of the system at runtime during its real working conditions, I designed and implemented XeMPower, a lightweight monitoring solution for the Xen hypervisor: it precisely accounts hardware events to guest workloads, enabling attribution of CPU power consumption to individual tenants. Results show that *XeMPower* introduces negligible overhead in power consumption, aiming to be a reference design to monitor power-aware virtualized environments. The power traces produced by XeMPower represent a perfect input for *MARC* to build data-driven power consumption models for multi-tenant virtualized infrastructures. Results show a modeling relative error of around 2% on average, and under 4% in almost all the cases and on different workload classes, outperforming previous research in the field. Moreover, I discussed model portability across similar architectures, showing how they can be also used to evaluate tenants colocation in a multitenant infrastructure. In order to thoroughly explore the topic, I finally included workloads' performances in the loop. At first, I discuss XeMPUPiL. a performance-aware power capping orchestrator for the Xen hypervisor: it implements a hybrid hardware-software power capping solution that aims at maximizing the performance of a workload under a power cap. *XeMPUPiL* has been validated with just one guest application running at a time: this condition may not be very common in a real production environment, where multiple guests may be hosted on the same node, each one with different performance requirements. In order to tackle these issues, a smart resource manager must be put in place to deal with resource allocation, power constraints and performance requirements. These issues are addressed in the last chapter, that discusses preliminary results and opportunities of *containerization*, i.e., a different approach to multi-tenancy: the proposed solution is called *DockerCap* and represents the first step towards power-awareness and QoS-awareness for Docker container orchestration, laying the basis for further research work.

Florio Luca - Supervisor: Prof. Elisabetta Di Nitto

Self-adaptation is the capability of a system to adapt in an autonomous way to every change in the scenario where it operates. This capability is fundamental in distributed systems, since they are composed of thousands of elements that work in a very dynamic and unpredictable environment. Distributed systems plays a central role in the current software development landscape: the advent of technologies like cloud computing changed the development methodologies, moving from applications built as a single and monolithic piece of software to systems composed of a great number of decoupled elements distributed on a large scale. Thus, distributed selfadaptive systems are an actual need and require to be studied in order to understand the most effective way to design and manage them. The literature provides different approaches to deal with such systems. However, while several algorithms for self-adaptation have been proposed, stable platforms and comprehensive software engineering approaches for the application of such algorithms to a concrete context are still to come. This thesis addresses these challenges, providing methods and tools to design and manage distributed systems able to

autonomously adapt to changes and operate in a concrete and dvnamic context. We started from the engineering of an existing decentralized self-adaptive system called the SelfLet Framework, that has been improved and deployed in a cloud computing environment in order to be evaluated not through simulation. but with a concrete case study. Using the SelfLet Framework the user is able to define the application as a set of services connected between them. The core element is the SelfLet, an autonomous computational unit which can act as a selfadaptive service container. Different services can be provided by a SelfLet, which integrates a feedback loop that actuates self-adaptive actions defined in a policy-based fashion. The selfadaptive actions comprehend the scaling of services, the load balancing of requests, and the change of the implementation of services. SelfLets can be deployed on a cloud platform and are able to self-organize in neighbors, i.e., intercommunicating sets of peers, maintaining application intent in a steady-state. The SelfLet Framework is completely decentralized: each SelfLet decides which self-adaptive action to actuate solely on the basis of the information coming from its

neighborhood. In this way there is no centralized unit which warrants self-adaptation, but rather the self-adaptive capability becomes emergent from the local decisions made by each SelfLet on the basis of the partial knowledge of the status of the system. We evaluated the SelfLet framework using experimental research on a real-life case-study, that is, online video provisioning. We used our framework to create a service-based application able to provide video-generation on demand and to actuate the most common key operations featured for these videos (apply subtitles, translate them, change video format, etc.). This scenario has been inspired by modern video-on-demand services such as Netflix, that relies on a cloud infrastructure based on the microservices architecture pattern, where the application is composed of several services deployed in virtual containers (e.g., Docker) distributed in thousands of nodes. For experimental purposes we deployed our case-study system on the Amazon Web Service cloud platform using 50 Virtual Machines, and sending web requests for the video provisioning service. We used a workload that changes over time and that has been extracted from the data obtained monitoring the traffic

of a real website. The analysis of the results shows the capability of the system to provide the requested service without violating any Service Level Agreement by self-adapting the application to workload changes. Despite the results of the evaluation highlight the selfadaptive capabilities of the SelfLet Framework, we identified some issues for its concrete adoption by practitioners: the use of a framework requires to learn how to use it, and imposes constraints on the technologies adopted for the development of the application. This led to the study and design of a tool that inherits the advantages of the SelfLets (i.e., decentralization, emergent behavior, encapsulation of the business logic), but that is as decoupled as possible from the technology stack of the application that should be managed. Such tool enables the transparent adaptation of the application, i.e, it actuates the self-adaptive process externally from the application, introducing self-adaptive capabilities without requiring its modification. In order to validate our approach, we designed and implemented Gru, a tool that exploits the concept of virtual container to introduce transparent adaptation into microservices-based applications. Microservices represent a good application domain for our tool: they have been adopted by the major companies in the IT industry due to their advantages, like high scalability and fault-tolerance, but are very complex and difficult to manage, because microservices-

based applications are composed of thousands of small independent elements communicating between them. Self-adaptation can help to handle such complexity, making the application able to manage itself to a certain level, but developers have to think how to integrate this self-adaptive layer inside the application during its design and implementation. Gru has been studied to address this issue, because it can be easily integrated with the application, autonomously managing the microservices running in Docker containers. Gru is composed of autonomous agents that are deployed in each node of the cluster running the application. Agents communicate directly with the Docker containers to gather information about the status of the services, and to actuate self-adaptive actions on the containers (e.g., the start and stop of containers). In this way the application does not require modifications to interact with Gru, which can be applied seamlessly to every application running in Docker containers. Each agent is independent and communicates with a subset of the total number of peers, exchanging information on the status of the system. Using this approach, Gru is totally decentralized and able to actuate actions on the basis of a partial knowledge of the system: this makes it well suited for the management of a large-scale distributed system, since the possible communication overhead is reduced to the minimum. By using Gru, developers can focus on the functional part of the application, without worrying

about the self-adaptive layer. We evaluated the prototype of Gru with the same video-provisioner application used for the evaluation of the SelfLets, deploying the services composing the application in Docker containers. We tested the system in a cloud environment composed of 30 Virtual Machines provided by PoliCloud, the cloud infrastructure of Politecnico di Milano, sending requests to the video provisioning service. We used the same workload adopted for the evaluation of the SelfLets. Despite the limitations of our prototype, the results of the evaluation show that Gru can adapt the application to the changes in the workload, and validate our approach for the transparent adaptation of a distributed system. In the future, we plan to improve the self-adaptive capabilities that Gru can introduce in the application, and evaluate it through its application to an industrial system and in a different application domain (e.g. Internet of Things).

EMBODIMENT IN ROBOTICS: THE SENSORIMOTOR LOOP TOWARD LEARNING AND COGNITION

Franchi Alessio Mauro - Supervisor: Prof. Giuseppina Gini

Robots nowadays are not widespread among human beings but are still confined into plants and research laboratories; they are in some sense "intelligent", but they are not autonomous in its real meaning. Besides lacking of smart autonomous behaviors. robots are also not able to learn and extend their knowledge as humans are: these deficits result in a very slow rate of adaptation in response to changes both in the environment and also to their own physical structure, such as faults. Summing up in a simple sentence, robots still miss what is called cognitive development, which for humans refers to the evolution of perception, memory, language, thinking and problem solving attitude; these are probably those factors influencing the most our growth and shaping our behaviors. Classical approaches to the problem of robot autonomy struggle to find good solutions; they have been so far oriented at solving a single problem at a time, either planning, or vision, or movement learning. Instead, a quick look at natural systems and their internal mechanisms reveals that all the systems together concur for the autonomy of the living being. Modeling autonomous systems is a very hard multidisciplinary problem

indeed; even more, we actually do not even know how the simplest animals are able to develop the behaviors they show and the human brain is still almost considered as a black box. A new view in robotics and Artificial Intelligence (AI) is now emerging; it is called embodiment and states that all the cognitive processes are strongly tied to the physical structure of the body and to its interaction with the environment. This theory is in contrast with the functionalism and dualism and classical AI; its central concept is the importance for learning of having a body interacting with the environment. This innovative perspective is getting everyday more common among Al experts, and affects different aspects of the cognitive processes, ranging for example from language and symbolic reasoning, to motor learning; a well known example of these concepts in robotics is the iCub platform, which aims at highlighting the enactive approach to cognition, also with the implications for phylogenetic configuration, i.e., the information hard coded in the DNA and passed down to generations, and the necessity for ontogenetic development, which on the contrary is related to a specific individual. This thesis explores the idea of "sensorimotor loop",

which is central for the embodied cognition theory, and refers to the continuous collaboration of sensing and acting as the basis for cognition. This dissertation deals in particular with the problem of modeling an embedded cognitive system with learning capability, focusing on the sensorimotor loop, which is known to play a kev role in our own continuous growth from newborns to adults. This loop is so considered a prerequisite for cognition and the consequent autonomy of robots; it affects cognition through motor and language development, and through the functions of the working memory. Moreover cognition requires the exploration of sensing data and of the agent body, activities driven by learning and new goal generation. Although many cognitive systems have been developed in AI, little effort has been spent to actually integrate those systems with models of the motor system to create a system suitable for controlling a robot. Different algorithms of large use in AI and robotics are selected and explored as the basic tools to implement the important functions of clustering, classification, imitation, memory, moving, and learning. Their biological mechanisms are functionally explored, including

some details about neural coding. that is the way our brain stores concepts and information. The selected methods are used to build high level models of sub-areas of the human brain involved in cognitive processes. In this dissertation the first model developed shows the capability of new goal generation through unsupervised learning. It exploits the functional organization of the brain areas of thalamus, amygdala and cortex. Providing it with a basic set of primitive instincts, it is able to exploit sensing data to autonomously generate new motivations. The second model develops a working memory, that allows the learning of sequences of action and delayed rewards of actions, and goes in addition to the generation of new goals. The third model extends the concept of motor primitives to learn trajectories



1. The transposition of the amygdala-thalauscortex model into the IDRA architecture

starting from the auto-generated goal of reaching a target. This kind of learning is similar to the motor babbling observed in infants. The model can, as well, learn movements from a human teacher using imitation learning. Moreover, the model of motion primitives is applied with minor modification to language learning, showing that the capability to generate meaningful words from basic simple sounds is similar to the capability to generate complex trajectories from basic simple movements. While developing those bio-inspired models, their integration into an architecture is considered. The whole system architecture, called IDRA, acts as a substratum of the whole functions.

All the proposed model are supported by a full implementation in Matlab, along with a variety of experiments tests are defined for humanoid autonomous robots. However our goal is not to outperform others' work but to check whether our main hypothesis and their implementation are able to induce in robots a child-like basic learning mechanism in terms of cognition, movements imitation and language learning. Our experimental results positively compare with biological evidences. In conclusion, we show a bio-inspired integration of different classical research areas as planning, sensing, and motor control, all tied together by basic learning mechanisms. This is a very first step in the broad research area of embodied cognitive systems, and several open problems still requires attention, but it shows a promising strategy to achieve cognition in robots

in simulation and on a real

show that the models are a

plausible implementations of

the natural system they try to

our results with the state of

the art, since no standardized

imitate. It is not easy to compare

humanoid robot. The experiments

Gervasoni Giacomo - Supervisor: Prof. Giorgio Ferrari

Thesis synthesis Lock-in amplifiers (LIAs) are extensively used for synchronous (phase-sensitive) AC signals detection and measurement in a wide range of scientific fields, from Atomic Force Microscopy (AFM), Raman spectroscopy, to sensors and actuators (e.g. MEMS). Such wide application is justified firstly by the ubiquitous presence of 1/f noise and secondly to the physics itself of the experiments, which in certain cases imposes an AC measurement.

I used such technique for the development of a first custom 4-channel FPGA-based lock-in platform operating up to medium frequencies (100 kHz), aimed at the characterization and readout of advanced magnetic sensors fabricated at the PoliFab facility of Politecnico di Milano. The platform, implementing a specifically designed transimpedance amplifier, shows better noise performance (by a factor between 6 and 10 depending on measurement conditions) compared to the most common state-of-the-art commercial instrument, the HF2LI by Zurich Instruments, used in combination with the HF2TA transimpendance amplifier from the same company. The second part of the thesis

treats the resolution limit of lock-in amplifiers, leading to the development of an extremely high resolution instrument. Ideally, when using LIAs, it would be possible to measure a signal variation given by the instrument input equivalent noise and the chosen filtering bandwidth, as for the minimum detectable signal. Instead, a detailed characterization of the digital lock-in amplifiers (one example in Figure 1) has shown an unforeseen 1/f noise at the lock-in demodulated output, proportional to the total signal amplitude. Consequently, the minimum detectable signal variation turns out to be proportionally dependent to the total signal itself and orders of magnitude greater than the theoretically expected

1. Noise spectral density of the demodulated LIA output obtained directly connecting output and input of a HF2LI lock-in amplifier from Zurich Instruments. value. The signal-proportionality and 1/f nature of this noise, pose a fundamental limit to the LIAs achievable resolution, defined as the ratio between the minimum detectable signal variation and the total signal amplitude, typically expressed in ppm. This limit has been found to be dependent from the instrument maximum operating frequency, from few ppm for lock-in operating up to few hundreds of kHz, to few tens of ppm for instruments operating up to few MHz or tens of MHz. I find out that the measured signal-proportional 1/f noise is due to slow gain fluctuations that the modulated signal experiences from the generation stage to the acquisition one. In particular, the DAC and ADC used for signal generation and acquisition, as well





2. Architecture of a switched ratiometric LIA. The equivalent gain in a period 1/fSW experienced by the two reconstructed signals is the same and is equal to the mean of the two ADC gains. The effect of the gain fluctuations is canceled out by performing the ratio between the amplitudes of the DUT and STIM signals.

as the analog stages, introduce these gain fluctuations, causing random amplitude modulation of the signal. This random amplitude modulation is successively downconverted through demodulation, causing random fluctuations of the LIA output.

To compensate these gain fluctuations, I conceived a switched ratiometric technique based on two ADCs alternately acquiring the signal coming from the DUT and the stimulus (STIM) signal (Figure 2). The idea is that both DUT and STIM signals should experience the same gain fluctuations (due to DAC and ADC), which can be successively removed by means of a division on the outputs of the synchronous demodulation, allowing a high-resolution measurement of the DUT response.

Finally, I realized an FPGA-based lock-in amplifier working up to 10 MHz and implementing the novel technique. Results (Figure 3) demonstrate a resolution improvement of more than an order of magnitude compared to standard implementations working up to similar frequencies (from tens of ppm down to sub-ppm values), not requiring additional external elements or accurate case by case calibrations, two typical constraints of the alternative differential technique. The technique has been patented and a company is investigating its possible commercial exploitations.



3. Demonstration of the technique effectiveness in making the DUT and STIM signals experience the same gain fluctuations allowing their cancellation by means of a simple ratio. 367

NFORMATION TECHNOLOGY

TOWARDS SUSTAINABLE SOLUTIONS FOR APPLICATIONS IN CLOUD COMPUTING AND BIG DATA

Ho Thi Thao Nguyen - Advisor: Prof. Barbara Pernici

Sustainability is a development paradigm where sustainable growth is placed at the center, forcing any involved activities to prioritize positive and long-term growth for the future of mankind. This pursuit of sustainability involves economic, social and environmental aspects, in which one of the goals is to minimize environmental impact. While the introduction of recent technologies in ICT sector such as Cloud Computing and Big Data opens new opportunities for economic growth and scientific discoveries, they also raise serious concerns about environmental impact due to the rapid growth of energy consumption and consequently CO₂ emissions. As reported recently by Global e-Sustainability Initiative, ICT sector is estimated to release 1.25Gt CO₂₀ in 2030, contributing 1.97% of global CO₂ emissions. One of the main reasons of energy inefficiency in Cloud Computing is resources overprovisioning, leading to most of resources are under-utilized. Thus recent research in energy efficiency has been focused on reducing underutilized resources, and state-oftheart techniques have been targeted at hardware and infrastructure levels of data

centers. On the other hand, while Big Data is bringing evolutionary

breakthroughs in scientific and business disciplines, the exponential growth of data is raising challenges in all aspects related to storage, accessibility, control, and energy consumption. In this era, the size of data sets to be analyzed is extraordinarily big, measured by Terabytes and even Zettabytes for a single data set. Processing, storing, maintaining and transmitting such big volume of data require a big cost. As energy cost is an important component in data center's strategy and management, and the energy consumption trend continues increasing, the challenge still remains to search for more ways to reduce the energy consumption, and consequently the environmental impact of ICT sector. While the research in energy efficiency in Cloud Computing is continuously growing in the last years, only a limited attention has been paid to the consuming energy behaviors of applications and the way computing resources are used by different application types and the consequent impact on the energy efficiency of the systems.

It is still not clearly understood, for example, how the energy consumption changes if we execute a given application in different ways; and whether there

are chances to further reduce energy consumption by just better configuring the application. From user's perspective, it still remains open questions on how much control they have to actively improve energy efficiency of their applications. Can they, at certain extent, actively exploit the dynamic of running environment to reduce their energy cost? With respect to Big Data challenges, several interesting questions have been raised: Are all of the generated data equally bringing the knowledge that users are seeking for? How much data should be generated? How much of generated data should be stored, maintained and analyzed? The answers to these questions will decide how much it will cost to go through all the data life cycles, from generation, capture, storage, and transmission to analysis. To this end, the research goal of

this thesis is aimed to look for new opportunities to further improve energy efficiency at the level of applications. Particularly, this research proposes to tackle the energy inefficiency problem in the context of Cloud Computing and Big Data from two different angles: the possibility to reduce the energy consumption of applications and the potential to reduce the amount of data to be processed. Adopting both modelingbased and experimental-based techniques, the first approach proposes analytical models to estimate the energy consumption of cloud-based applications and quantify the energy consumed per job. The models are built based on cloud infrastructure's and application's parameters and are validated experimentally in a real cloud infrastructure. Once validated, the models are used to estimate power and energy consumption of applications under different configurations. The estimated energy consumption is the basis to quantify the energy per job, which in turn is used as the main driver to analyze energy consuming behaviors of applications and to assess possibilities to improve their energy efficiency. Our methodology is applied to three different types of workloads: batch, transactional and I/O intensive workloads. Through the selected case studies,

we demonstrate that application's energy efficiency can be further improved at the application level by better configuring their execution or better distributing the workloads. For example, we show that for batch workloads, shared resources with different access policies have significant impact on the energy consumption. Particularly when there is the bottleneck in the system, adding more resources is not always useful for improving performance and might result in lower energy efficiency. For transactional workloads, we show that users can actively reduce the energy consumption of their applications

by considering the power profiles of their VMs and deciding the optimal load distribution policy; whereas for I/O intensive workloads, the energy inefficiency can be reduced through avoiding interference among applications. As a final result, the proposed method provides a means to both application's users and cloud infrastructure's providers to actively control and optimize the energy usage of cloud-based applications. In the second proposal, the thesis proposes a novel approach for

proposes a novel approach for achieving sustainability in the context of Big Data through considering the potential to reduce data volume. Our main motivation is that data are not equally valuable, thus data volume reduction can be achieved through identifying and removing irrelevant/unimportant data with respect to user's goals. In our proposal, we first characterize the Big Data workloads' characteristics, and select data dimensions that can represent these properties. Based on the selected dimensions, we propose a novel metric called *data value* to represent the valuable level of data. The *data value* indicates how much important or valuable the considered data are. A high value implies that the data are important for users' goals and vice versa. Based on this value, a reduction of data volume can be achieved by reducing or discarding data that are not useful for users. Finally we propose a generic energy-efficient framework (Figure 1) utilizing this value, and demonstrate its applicability under different case studies. We also illustrate the inference of *data value* through a specific technique called data correlation. We also design and implement an efficient and adaptive algorithm to search for temporal correlations in the context of big data sets, and apply it to a case study to evaluate its effectiveness and performance.



1. Data-value-driven energy-efficient adaptation framework

369

NFORMATION TECHNOLOGY

TOWARDS SUSTAINABLE SOLUTIONS FOR APPLICATIONS IN CLOUD COMPUTING AND BIG DATA

Jalili Vahid - Supervisor: Prof. Matteo Matteucci

Advances in next generation sequencing (NGS), also known as high-throughput sequencing, ubiquitize DNA sequencing as a flexible tool for genome exploration. NGS has opened the possibility of a comprehensive characterization of the genomic and epigenomic landscapes, giving answers to fundamental questions for biological and clinical research, e.g., how DNA-protein interactions and chromatin structure affect gene activity, how cancer develops, how much complex diseases such as diabetes or cancer depend on personal (epi)genomic traits. This is opening the road to personalized and precision medicine. A distinguished aspect of NGS-based experiments is the large amount of data they produce. The generated data are broadly applicable and facilitate various functional analysis, including studies about DNAprotein interaction or histone modification (using Chromatin immunoprecipitation followed by massively parallel DNA sequencing (ChIP-seq)), transcriptional regulation (using RNA-seq), long range chromatin interactions explained by \textit{de novo} spatial structure of genome (using Hi-C6). Recent studies combine these studies into larger assays for in-depth interpretations of sequencing data. Yet such

interpretations, and \textit{making sense} of data, demand complex computation and large scale data retrieval systems. The present dissertation has focused on \textit{sense-making}, e.g., discovering how heterogeneous DNA regions concur to determine particular biological processes or phenotypes. Towards such discovery, characteristic operations to be performed on region data regard identifying co-occurrences of regions, from different biological tests and/or of distinct semantic types, possibly within a certain distance from each others and/or from DNA regions with known structural or functional properties.

The manuscript explains Di4 (1D interval incremental inverted index) and its predecessor Di3 (1D interval inverted index). Di4 and Di3 are single-dimension (1D) multi-resolution indexing frameworks, designed to be comprehensive, generic, extensible, and scalable back-end data structures for information retrieval on NGS interval-based data. Di4 and Di3 are defined at data access layer, agnostic to data, business logic, and presentation layers; this design makes them adaptable to any underlying persistence technology based on key-value pairs, spanning from classical B+tree to LevelDB and

Apache HBase, and it makes them suitable for different business logic and presentation layer scenarios. Benchmarking Di4 and Di3 on real and simulated datasets and a comparison with common tools in bioinformatics realm, demonstrate the effectiveness of Di4 and Di3 as a back-end for general purpose genomic region manipulation. The applicability of Di4 and Di3 to different business logic and presentation scenarios, and extensibility to applicationspecific functions, is assessed in comparative evaluation of ChIP-seg samples. The ChIP-seg technology identifies protein-DNA interactions using enriched regions on DNA when the significance measure (p-value) is below a stringency threshold. Replicated samples are expected to have a degree of repeated evidence, which can locally lower the minimum significance required to accept an enriched region. The present dissertation discusses a method for the joint analysis of ChIP-seg replicates, which confirms overlapping enriched regions if their comparative evaluation complies a set of userdefined thresholds. The method is implemented using Di3 to demonstrate the extensibility of the frameworks. Additionally, Di3 is used as back-end data structure to implement common ChIP-

seq data assessment methods, such as functional analysis, correlation assessment, nearest feature distance distribution, chromosome-wide statistics, and genome browser.

Keywords:

Genomic computing; indexing framework, domain-specific data indexing; region-based operations and calculus; data integration; comparative analysis. 372

PhD Yearbook | 2017

Kaitoua Abdulrahman - Supervisor: Prof. Stefano Ceri

The recent emergence of Next Generation Sequencing technologies, in genomics field, produced vast amounts of genomic data. NGS resulted in dropped the cost of sequencing ("reading" in general terms) genomic material very fast. There exist many methods to extract signals from the genomic data, that associate a region of the genome with some interesting information - such as a mutation or a peak of expression. Thus, a new problem is emerging: making sense of these signals, heterogeneous in nature, through new kind of languages that can extract relevant information from various heterogeneous sources, integrate them in a new data management system, and compute interesting results. Biologists say that a huge amount of information is undiscovered within the repositories that have been built in the last decade therefore, the focus of genomic data management for the next decade is querying and analysing heterogeneous genomic data. This thesis is about scalable data management and processing for genomic data. We developed a new system which consists of a new query language called GenoMetric Query Language (GMQL), a new data model for heterogeneous data (called

Genomic Data Model - GDM), and a new processing engine which embeds scalable genomic algorithms implemented on several data flow engines. The name (GMQL) derives from the ability of the language of dealing with region-based operations which take into account the regions topological location to the reference genome. GDM mediates all existing heterogeneous data formats. In combination, GDM and GMQL introduce a paradigm shift, by providing a high-level, declarative query language which supports data-driven computations.

GMQL is a collective effort which has involved a group of students and professors from Politecnico di Milano. The work reported in this PhD thesis is focused on the design, implementation and validation of the scalable genomic data management and processing system; The architecture of various prototypes of GMQL, GDM, and the scalable genomic processing algorithms.

GMQL is a domain-specific language. In this thesis, we developed several scalable algorithms for genomic processing for serving the needs of GMQL queries. By the use of data flow engines as our target implementations, we capitalize upon existing frameworks which are available today and will be developed in the future by the data management community. In order to increase the parallelism of the genomic algorithms in data flow engines, we proposed new data binning methodologies that are suitable for genomic data and for the nature of the data flow engines. We also used our genomic algorithms for comparing dataflow engines and we developed versions of the scalable algorithms that take advantage of the nature of data flow engines - for example in SciDB, based on multidimensional arrays, we make use of the fast access to slices of the array.

Along with the design of the system and the scalable genomic algorithms, the implementation has gone through phases, and specifically we delivered a first implementation of GMQL, called GMQL V1, based on Hadoop 1 and the target systems Pig; and a second implementation, called GMQL V2, based on Hadoop 2 and the target systems Spark, Flink, and SciDB. The thesis describes the rationale of the two implementations and the process that led from the first to the second prototype. GMQL engine has a well-designed

system architecture with modular organization; system modules can be easily tested, maintained or replaced. GMQL is translated to an intermediate, targetindependent representation, based on a Directed Acyclic Graph (DAG), which describes workflows of basic operation nodes. Each node is a primitive operation that implements a specific functionality of GMQL, and operations are mapped to specific target systems; in this way, we can support several implementations to several target systems.

The GMQL engine architecture includes a repository abstraction which is technology-independent, hence several options (local file system, Hadoop File system, scientific database management) are made available by simply including a different repository interface.

Several aspects of the architecture are designed for fast execution on big data sets. This thesis also includes a thorough performance analysis, by comparing GMQL engine V1 to V2, Flink to Spark implementations, and Spark to SciDB implementations. From such studies, we learnt about optimal parameters settings for the scalable genomic algorithms on diverse data sizes and platforms. We also include some preliminary results of a study which we are conducting in order to fieldtest GMQL applicability. We implemented a pipeline that

uses GMQL for studying gene expression in normal and cancer cells in the context of DNA 3D structure; the study is based on big data sets for about twenty available tissues, thousands of samples for either normal or tumor cases are considered in the study.

In summary, this thesis is a step forward in the development of a systemic approach to scalable genomic data management and processing. Whereas other approaches focused on extracting genomic features from data, our approach is on combining these heterogeneous features so as to solve complex biological problems. We believe that the importance of a systemic approach to genomic data management will grow in the near future, with the availability of huge repositories of genomic data sets.

Laghi Giacomo - Supervisor: Prof. Antonio Longoni

The aim of the project was to develop a 3-axis Lorentz-forcebased MEMS magnetometer. for heading and navigation applications, capable of overcoming several limitations of other types of commercial magnetometers (e.g. AMR, Hall and Fluxgate devices) and compatible with a co-integration with other MEMS inertial sensors (accelerometers and gyroscopes). The development of single-chip 9-axis sensors (acceleration, rotation speed, magnetic field) is indeed one of the most interesting innovations in the MEMS research

374

area. The target applications for this sensor (designed for the consumer market) are the heading determination and the pedestrian navigation. The design requirements were related to the sensor size (to be minimized to reduce the cost), to the low power consumption (for portable equipment) and good enough performances (an accuracy in the determination of the terrestrial magnetic field of about 50 for heading purposes to about 0.10 for navigation). A first prototype has been developed based on three single-axis sensors co-



1. SEM image of the complete 3-axis module, composed by one Out-Of-Plane (OOP) field sensitive device, two In-Plane (IP) field sensitive device and a Tang resonator to select the drive frequency. White arrows indicate the current path integrated with a suitable MEMS resonator.

The most relevant technique exploited to enhance the performance of these sensors are: (i) design of multi-loop structures: in this work three 1-axis 10-loop structures are presented. Making the driving current recirculate within the three devices - in series - a 30-fold improvement is gained in terms of power dissipation required to obtain the desired resolution; (ii) this can be achieved, together with the stability with temperature and the solution of the bandwidth-noise density trade-off typical of Lorentz-Force MEMS magnetometers, thanks to: (i) off-resonance operation: injection of a driving current at a frequency slightly lower than the resonance one; (ii) MEMS resonator fabricated within the same package of the sensor. As a side but not negligible effect, this enables the possibility of re-using electronic stages well refined for MEMS gyroscopes, which have been deeply investigated in the last decades. The prototype has been

characterized through a custom discrete components electronics and a custom (hardware and software) measurement setup. The possibility to achieve with this kind of structure performances



2. (a) Complete schematic of the designed PCB for the drive signal generation and readout of capacitive 3-axis Lorentz-Force based MEMS magnetic field sensor, and (b) characterization setup

comparable with or beyond the state-of-the-art of commercial products has been demonstrated. The industrial orientation of the work is also underlined by two additional factors: (i) the extensive characterization performed on the sensor: aspects as the stability with temperature and vibration rejection have been taken into consideration. Usually in the research field these second order effects are not characterized, but they are essentials for the correct operation of a MEMS device as a commercial product; (ii) they are fabricated in a standard industrial technology (ThELMA by STMicroelectronics), respecting all its design rules and constraints. This potentially makes the sensor ready to be fabricated and put into the consumer market with strong guaranteed reliability and repeatability.

A further step in the project has then been in the direction of strong reduction of the area occupation and the residual offset drift. The area occupation issue is addressed by the re-design of the structure, switching from

three 1-axis structures to a single 3-axis device. This adds complexity to the geometry, both because of the need of having three well decoupled oscillating modes and because of the dimension targets. The development has been carried out with the aim to fit the magnetometer into a 2x2 mm2 package together with a 3-axis accelerometer, in order to obtain a full 6-axis electronic Compass. The development of this device is presently in the ending of its layout phase (for masks definition). A related patent has been submitted. A 3-axis dedicate ASIC will be designed. Meanwhile, a slightly different fabrication process is under development, allowing the metal layer in which the driving current flows to be isolated with respect to the structural polysilicon layer. A new magnetometer will be designed based on this new process. The offset and its associated drifts will be reduced of orders of magnitude, allowing both a sensitivity increase and a further dimension reduction.

An experimental study on the

damping coefficient reduction for MEMS structures operating within certain pressure conditions, which would enhance the minimum detectable field, have been carried out and it is currently under investigation. From preliminary characterizations on test structures with different sensing geometries and different complexities a general predictive model for the damping coefficient has been produced. Allowing to predict the quality factor of a generic structure in the design phase, it would make possible the identification of sensing geometries for damping coefficient reduction.

PHYLOGENETIC ANALYSIS **OF VIDEO SEQUENCES**

376

Lameri Silvia - Supervisor: Prof. Stefano Tubaro

In the last few years, the availability of inexpensive, portable, and highly-usable digital multimedia devices (such as cameras, mobilephones, digital recorders, etc.) has increased the possibility of generating digital audiovisual data without any time, location, and network-related constraints. In addition, the versatility of the digital support, concurrently with the advent of social networks. allows copying, editing and distributing multimedia data with no effort. Indeed, once one image or video is shared by a user, it can easily go viral, being republished by many other users. Moreover, some users can create a new edited version of this content (e.g., by adding overlaying text or splicing some other content) that, in turn, goes viral. One example is the diffusion of the so called Internet memes, i.e., pieces of media which spreads, often as mimicry, from person to person across different channels on the Web. This scenario inevitably presents

an alarming situation: besides the many legitimate uses of digital documents, there are also various cases of criminal activities, including copyright infringement, sharing of illegal or abusive contents (e.g., child pornography, persons in a bullying situation, fake and defamatory images of celebrities or politicians), forgeries

in scientific publications and digital material used as evidence in court, among others. The need of methods and tools that enable to assess the origin and authenticity of multimedia data is therefore more than an urgent necessity. Recently, many forensic techniques have been proposed to detect malicious modifications of images and videos. These tools typically analyse the single object per se. However, great part of user-generated content distributed online is often created by modifying already existing material. In this scenario, multiple correlated copies of the same object are available to a forensic analyst, who can synergistically take into account the information coming from all of them to perform even deeper forensic analyses. In the literature, multimedia objects obtained by applying different content preserving transformations (e.g., colour correction, filtering, resize, crop, logo insertion, text overlay addition, etc.) to an original content are defined as nearduplicates (NDs). The study of near-duplicate objects and their causal relationships is the main focus of a recently developed field, Multimedia Phylogeny. More specifically, multimedia phylogeny aims to jointly analyse pools of near-duplicate objects

instead of single instances of the same object separately, in order to infer how these near-duplicates are related to each other and study their provenance. This kind of analysis has immediate applications in the forensic field. As an example, it paves the way to the development of more effective copyright infringement analysis tools. Moreover, it enables forensic analysts to identify the main responsible for the diffusion of illegal material. Although a significant research effort has been devoted to

phylogenetic analysis of still images, video phylogeny is at its early stages. Therefore, the goal of this thesis is to develop a series of video analysis techniques tailored to video phylogeny. Specifically, rather than focusing on a comprehensive study on a very narrow and specific aspect of video phylogenetic analysis, this work aims at providing a basic set of solutions that enables to develop video phylogeny applications from head to tail. To this purpose, we first focus

on solving two primary yet paramount problems for video phylogeny analysis: detection of near-duplicate videos and their temporal synchronization. We initially propose a method that addresses these problems in a simplified framework in which only



near-duplicate videos or unrelated ones are considered. Then we extend the proposed algorithms in order to jointly detect and temporally align near-duplicate videos together with sequences depicting the same scene from different viewpoints (semantically similar videos).

After providing these basic tools for video detection and synchronization, we show how to exploit them as building blocks for more complex applications in the video phylogeny scenario. First we propose a method that, starting from the analysis of a pool of near-duplicate videos, reconstructs an estimate of the original source video used to generate them. This approach allows us to study how content is reused and to recover original content when it is no longer available. An interesting case of study can be found in the news tracking service field. Indeed, being able to reconstruct the spreading pattern of an image or video can help companies to

understand demographics and effectiveness of an ad campaign of a product.

The second developed application retrieves the causal relationship between video pairs within a pool of near-duplicate videos, and represents them by means of a phylogeny tree. This is done by taking inspiration from biology studies and assimilating the process of videos changing into different versions over time to the mutation of living organisms. The presented application allows us to assess the authenticity of a video content and identify its origin. This possibility plays an important role in the detection of copyright and patent violations or in investigations regarding the diffusion of illegal material (such as child pornography, cyber bullying or terroristic propaganda). For instance, let us consider the previously described scenario, in which images and videos containing illegal or abusive content are

PhD Yearbook | 2017 377

first posted on the Internet and then reacquired, modified and redistributed by different users. In this case, application of phylogeny algorithms allows us to find the original content (or the least

Examples of near-duplicates

modified one), which can also give hints about the creator of these contents. The developed algorithms contribute to the widening panorama of multimedia phylogeny, representing a first step towards the analysis of video

sequences. In order to validate the proposed approaches, a set of experiments on real video sequences has been carried out. Moreover, also some use-cases on video sequences downloaded from YouTube have been analysed to prove the effectiveness of our approaches in a real-world scenario.

PHYLOGENETIC ANALYSIS OF VIDEO SEQUENCES

Li Yu - Supervisor: Prof. Andrea Castelletti

Most of the hydrological models developed in the last decades, and ordinarily used for decision support in water system, focus on the natural component in the water cycle whilst human activities are predominantly regarded as external forces that marginally impact the hydrological system. This is, however, contradict to the presence of human signature in many river basins worldwide, and whose increasing impacts are rivaling with the natural forces themselves in conditioning the natural processes and transforming the hydrosphere. The feedback between the human system and the natural ones also imply such process a complex co-evolutionary one, with the projected nonstationary climate change further exacerbating our ability to predict the future conditions of these systems. To this end, there is a



1. The conceptualized diagram of CHNS

growing interest in shifting from traditional human-excluded modeling practices to a more integrated approach, where the human related entities have to be included in the modeling framework to better describe such Coupled Human-Natural Systems (CHNS, see Figure 1). In this work we present several contributions to CHNS studies. starting from an innovative modeling framework of constructing a fully coupled CHNS, followed by a series of demonstrations to assess the potential of proposed framework. In particular, we focus on a coupled human-water system as subset of, namely Lake Como (Italy) water system, where the natural subsystems are lake and agricultural irrigation areas, while the human subsystems are composed of lake operator operating the lake, and farmer

stakeholders engaged in agricultural practices, respectively. By applying an agent-based model (ABM, see Figure 2) we construct a fully coupled model for Lake Como, denoted as DistriLake framework, in which the hydrological sphere is described by traditional processed based models, while the humans' behaviors are represented by implicit/normative, optimization based behavioral models. Our results are disseminated into three contributions. In the first study, we use proposed decision-analytic modeling framework to study water supplydemand co-adaptation in Lake Como CHNS under changing climatic boundary conditions. Specifically, we dynamically couple water supply and demand with a twofold goal: i) to improve our understanding of present functioning and future evolution of agricultural systems by better describing the feedbacks between the natural and the decision-making processes driving the dynamics of these complex systems; ii) to assess the space for improving current water management practices by removing policy inertia, thus compressing the time of the transition toward more efficient solutions. The application to the Lake Como system confirmed



2. The conceptual representation of ABM framework in modeling CHNS

the validity of implicit behavioral models in capturing the overall rational behaviors of agents, and the co-adaptive method provides a useful framework for contrasting different policy options, assess the impacts and costs of policy inertia and explore the potential co-benefit of a more dynamic management of the water supplydemand interactions. In the second one, we advance traditional robustness decision frameworks by replacing the centralized social planner with a bottom-up, agent-based approach where stakeholders are modeled as individuals, and represented as potentially self-interested agents. Our framework enables a more explicit exploration of the potential inequities and asymmetries in the distribution of the system-wide benefit. Results show that temperature change

appears to be the major stressor for causing the non-robust situation at both system level and agent level, although the crops sale prices is more important than the inflow when evaluated in terms of system-level expected profitability. Analysis conducted at the farmer-agent scale highlights alternatively that inflow change is more critical than socio-economic factors in leading system failures. Finally, we show that the robustness of the considered adaptation options varies spatially, strongly influenced by stakeholders' context, the metrics used to define robust performance, and the assumed preferences for lake operations in balancing urban flooding and agricultural productivity. At last, we apply the DistriLake framework to assess the actual value of long-term weather

forecast products in improving the agricultural cropping practices. We show that how the agent-based behavioral models provide a distinct perspective by assessing the operational value of the innovations from the simulation of the actual decision-making process with the forecast information, in which the deterministic forecast of climatic variables is translated into the decision relevant outputs, i.e., the crop production and final profit. Such analysis contributes to a step further assessment of weather forecast products, in addition to traditional measures of forecast quality, by accounting for endusers' context, e.g., risk attitudes and/or behavioral patterns, which could alter the actual value of forecast information. To summarize, our specific aim is to propose and test an innovative decision-analytic framework by coupling implicit behavioral models with traditional natural process models, which should provide a new method and tool to study CHNS, and to advance the extant water system analysis techniques towards a more comprehensive assessment.

379

ANALYSIS AND DEVELOPMENT OF INTELLIGENT SYSTEMS FOR SAFETY IMPROVEMENT IN VEHICLES

Lucchetti Alberto - Supervisor: Prof. Matteo Corno

From the appearance of the first motorized vehicle, automotive researchers focused their attention to vehicles' safety, performance. and also to the quality of the driving experience. Due to the emergence of electronics, the vehicles' dynamics control became the major issue for researchers. Nowadays, most of the innovations in vehicles are directly associated with electronic innovation, related not only to pure built-in electronic but also to other devices such as smartphones or navigators. Vehicles in general are now composed by several subsystems. Every subsystem is focused on specific vehicles' dynamics, with the aim of keeping the vehicle in a safe state, whenever possible. First generation control systems act in parallel with the driver command in a feedforward way. Examples of this generation are: Anti-lock Braking System (ABS), Electronic Stability Control (ESC) and Traction Control (TC). This kind of systems has an input requested by the driver and tries to reach it without compromising the vehicle stability. Second generation control systems completely alter the characteristics of the vehicles' responsiveness to the driver action. Examples of this are: Automatic Transmission (AT), Steer-By-Wire (SBW) and Active Steering (AS).

These systems work independently

from the user's demand and. because of this, they have a great potential to improve vehicles' safety and performance. Lastly, third generation control systems melt the action of the previous systems in order to obtain a global controlling system. The interaction between subcomponents can strongly increase the efficiency of each component. These systems can be considered as traditional systems. Nowadays, traditional systems are standardized in produced vehicles, and automotive industries are continuously improving their robustness. In the last decade, a massive effort in the automotive research field has been focused on pioneering new solutions and technologies in order to increase road safety. More specifically, the so-called Advanced Driver Assistance Systems (ADAS's) have been deeply analyzed and developed. These new systems are focused on supporting the drivers during various driving tasks. As previously mentioned, traditional vehicles' dynamics control is focused only on the vehicle itself, while the ADAS's help the driver in daily traveling with the purpose of increasing road safety and driving comfort. The scope of the new ADAS's ranges from systems which support the driver to systems that directly

stabilize the vehicle. Automotive researchers change their focus from vehicle behavior to external elements behavior. The focus is no longer the interaction of the vehicle with its own state, but the interaction with mainly three other actors:

- The infrastructure. In this context, the infrastructure plays a coordination role by gathering global information on traffic and road conditions. With this information, it is possible to inform vehicles about their own state. In this way it is possible, for instance, to redirect vehicles towards another route in case of crashes or traffic jam. The communication is bidirectional: the vehicle can automatically inform the infrastructure about a great quantity of information.
- Other vehicles. The so called Vehicle-To-Vehicle interaction aims at organizing the interaction among vehicles, and possibly developing collaborations among them. There are two ways to gather information from other vehicles. It is possible to detect them with various sensors, such as a radar or a camera, but also by directly establishing a sort of communication among them. The former is the nowadays adopted solution,

due to the growing accuracy and reliability of the sensors, while the latter needs a new module in vehicle, such as GSM, to create a connection to a net.
The driver. In traditional control systems, this interaction is conceived as unidirectional. The

of reducing the number of deaths

under twenty thousand by the year

2010 resulted in the development

of a large variety of advanced

driving assistance systems. In

2007, there were 1.24 million

deaths on the world's roads and

a similar number of deaths were

registered in 2010 even if there

was a corresponding 15% global

increase of registered motorized

vehicles. This fact suggests that

interventions to improve global

road safety mitigate the number

ADAS's require new development

• One of the main problems of

techniques and a new definition of

of deaths.

security paradigms:

driver influences the vehicle's behavior acting as reference input generator, while in ADAS's the aim of this interaction is to capture signals from the driver, such as drowsiness, drunkenness, or driving attitude. By detecting the driver's state, it is possible to react in terms of performance and safety, and to instruct the drivers to act in a safer way. The definition of this interaction becomes bidirectional. The ultimate focus of automotive researchers is road safety, and the strength of the ADAS's relies on the communication between road actors. As a matter of fact the decision of the European Union to improve road safety with the aim

systems. These systems must take decisions similar to those of a human during driving; thus they have to make direct decisions, as is the case of collision avoidance systems, or to warn the driver and implicitly ask for emergency maneuvers, as in the case of lane departure warning. In both cases, the system cannot ignore the existence of the driver and has to act like a human. In this way, the action is accepted by the user. For this reason, in ADAS's development, it is more suitable to speak about scenarios and not tests, since there is a huge number of situations that need to be tested and, a unique definition of the correct behavior is not possible. Traditional systems are focused on a single task independent from the state of the environment. In this context, to define the correct scenario/situation of ADAS exploring, for instance, vehicle to vehicle interaction, is not a trivial operation. In this dissertation, the methods proposed in order to extract such scenarios from real driving experience have the scope of being as general as possible. As a consequence, the detection and the extrapolation of particular scenarios can be easily modified to adapt the procedure to different aspects. In traditional control problems, as previously mentioned, the target is generally clear and

developing ADAS's is that there

are several possible definitions

of what is more correct and

safer with respect to previous

consolidated. Security is in practice the stabilization of certain dynamics in order to reach the best performance. Safe performances are clear and they are, in a certain sense, implicit in controlling the vehicle's dynamics. With ADAS's, safety has to be intended in a more general way. Thanks to nowadays high connectivity, the safety improvements coincide with sharing information in order to react to unexpected events or to predict future movements of other vehicles. Thus the definition of safety is a sort of integrated safety of all the actors of the driving experience, and not only of the host vehicle. Vehicle to infrastructure and vehicle to driver interactions are seen as a powerful tool to automatically share the driver's state.

Luperto Matteo - Supervisor: Prof. Francesco Amigoni

Understanding the environments in which they operate is an important capability for autonomous mobile robots. These robots can perform many different tasks to help humans during their activities or to replace them in hazardous environments and in routine operations. In several of these tasks, robots are required to operate without any human supervision, especially in environments that human beings cannot access, as in rescue settings.

382

When we consider indoor tasks, robots have to interact with environments that are specifically designed for human activities and for interaction between humans, *buildings*. Buildings are strongly structured environments that are organized in regular patterns. For instance, rooms typically have a geometrical structure that is characterized by walls perpendicular to the floor and to the ceiling, and by a layout that can be approximated by a box-like model

In order to increase their ability to autonomously operate in indoor environments, robots must have a good understanding of buildings, similarly to the one human beings exploit during their everyday activities. This is even more important when the robot have not any previous knowledge on the

environment in which it operates, but incrementally acquires uncertain and noisy data from the environment while performing its task.

If we consider how people and robots interact and move within indoor environments, it can be said that people naturally understand and "read" buildings as humanmade environments (and act in them accordingly), and that this is hardly the case for autonomous mobile robots.

The standard method for representation of environments in robotics are metric maps, which describe the space occupation and do not explicitly contain any knowledge about the building structure.

Semantic maps have been proposed as an abstraction of metric maps aimed to represent the meaning of parts of the perceived environment by describing heterogeneous concepts that can be useful for robots, such as the nature of objects within the environment and the kind of activity performed in a room.

In this work, we focus on semantic mapping on indoor environments. Specifically, we consider a semantic map as a representation that identifies rooms by partitioning the metric map, recognizes how rooms are connected, and assigns to each room a semantic label

indicating its function, such as `corridor', `classroom', `office', or `bathroom'.

The main contribution of this work starts from the consideration that the structure of buildings, which is often neglected when building semantic maps, could be exploited to increase the autonomous abilities of robots when operating in indoor environments by improving standard semantic mapping techniques overcoming some of their actual limitations, in order to make a step towards filling the gap between how humans and robots interact with the environment they inhabit. Semantic mapping is usually performed starting from a metric map, which is built on data acquired incrementally with perception and which implicitly assumes that that has not been seen by the robot does not exists. As a consequence, a semantic map represents and provides to the robot only information about the subset of the environment that has been already visited by the robot. This form of interaction with the environment is radically different from that of humans and constitutes a limitation of semantic mapping.

Another limitation regards the fact that semantic mapping methods are usually trained and tested on a limited set of environments. A



third limitation regards the fact that semantic labels are assigned to parts of the environment often considering only information acquired directly there or, at most, on a *local* neighbour of those parts of environment. No information about the global structure of the environment is usually considered. In order to overcome such limitations, we consider an entire building as a single object, by identifying relations between different (and potentially unconnected) parts of the building using both the metric map of the environment (for example by identifying that rooms in different parts of the building share one or more walls), and the topology of the environment, namely how rooms are connected with each other, and for example observing that parts of the building with a similar function have a similar structure.

Moreover, we consider each building in relation with other buildings with the same function. The function of a building is represented by the main activity that each building is designed for and is captured by the concept of

building type. Examples of building types are schools, offices, hospitals, and others. We analyse data sets of buildings belonging to the same building type in order to identify the features that are shared between all the buildings of the same type.

We then model the *structure* of a building by identifying some of its *features*, which are measurable quantities, features that are intended to describe *global* characteristics of the environment and that cannot be determined by applying standard *local* approaches. The decision of considering the global structure as the principal target of our analysis leads us to focus on the 2D structure of a building, collected by laser range scanners. At first we extract automatically the structure and the connectivity of rooms of a building using a method that reconstruct the floor plan from a metric map obtained by a robot or from a blueprint. Such method can be also used to complete the layout of partially explored rooms. An example can be seen in the first figure. We evaluate this extracted

SAMPLED structure against those of similar buildings obtained from data sets, finding relations between parts of the target building and those of the data sets. Using Statistical Relational Learning techniques, besides semantic mapping, we enable reasoning tasks such as

> categorizing an entire building according to its function. The topology of the building is then represented as a graph, which is used for identifying the building structure. Using Markov Chain Monte Carlo and Graph Kernels we develop a generative model of the graph structure of the buildings, which is used for sampling new instance of graphs representing building topologies and for predicting the structure of unvisited parts of a building. An example of a topological structure of a floor of a real school compared to that of school sampled with our method is provided in the second figure.

ORIGINAL

TIME-OF-FLIGHT CMOS SINGLE-PHOTON CAMERAS FOR HIGH FRAME RATE 2D IMAGING AND 3D RANGING

Lussana Rudi - Supervisor: Prof. Franco Zapppa

During the last decade, the ability to acquire high frame-rate (higher than standard video-rate). both two- and three-dimensional videos of scenes with very faint illumination has become more and more important in many fields, such as ambient surveillance, road safety, identification of people and objects, gaming, biomedical imaging, and studies on physics of materials. In particular, there is a growing interest in devices capable of acquiring high framerate 3D distance-resolved videos with measurement ranges up to some hundreds of meters, to be employed as sensors for completely autonomous vehicles. My PhD research aimed at developing high-end cameras able to acquire such videos, both in 2D and 3D mode, based on direct Time-of-Flight (ToF) technique and deploy them in real scientific applications through international collaborations. The cameras are based on SPAD (Single-Photon Avalanche Diode) detectors which, thanks to their extreme sensitivity, make it possible to achieve both relatively high (some 100.000 frames/s) framerates and long distance ranges. Compared to other existing solutions, SPAD imagers are the only devices capable of achieving both single-photon sensitivity and high frame-rate at the same

time with no need for bulky and complex cooling systems. Other single-photon sensitive sensors, like photo-multiplier tubes and micro-channel plates, already exist on the market, but they cannot be integrated in an array together with complex CMOS electronics, thus preventing their exploitation as imaging devices. In detail, during my PhD I developed compact camera modules based on previously developed SPAD+TDC array chips with either 32x32, 16x32 or 60x1 pixels, for acquiring time-resolved optical signals by means of the Time-Correlated Single-Photon Counting technique. Thanks to the exploitation of on-board FPGA data-processing and a USB 3.0 data link, the acquired data is quickly transferred towards a PC (up to 1.5 Gbps data transfer), enabling the acquisition of real-time 2D and 3D videos with a frame-rate as high as 200 frames/s. I also conceived and implemented a novel algorithm to perform real-time data compression by automatically neglecting infrequent ToF values (associated to background light and dark counts), and keeping stored, per each pixel, only the few of the 1,024 histogram bins where the echoed signal is actually detected. In this way, the memory requirements are

reduced by a factor 10x compared to standard processing, thus enabling the implementation of such an algorithm inside the on-board Spartan 6 FPGA. This also gave me the possibility to control the camera with a Raspberry microcomputer, therefore miniaturizing the system, increasing its portability and minimizing cost, weight, and power consumption. Moreover, by exploiting Time-to-Digital Converter (TDC) refolding I conceived a technique for increasing the photon detection probability without reducing the maximum range duration, by repeatedly illuminating the scene several times in each frame. I conducted a theoretical analysis on the achievable 3D ranging precision, also accounting for background contribution, as opposed to traditional equations. By considering unrealistically ideal conditions without background light, the latter can underestimate the measurement uncertainty even by an order of magnitude. I validated my equation by means of both Monte Carlo simulations and experimental acquisitions. Based on this analysis, I compared the "direct" ToF 3D ranging technique with the "indirect" (i-ToF) ToF approach, in the same operating conditions and with same technology and detector, the



1. Assembled camera module, made by four boards interfaced to the 32x32 SPAD+TDC array chip

former showing better precision (by about an order of magnitude). I also conducted an analysis on the accuracy of the single-photon i-ToF technique, highlighting an unexpected dependence on hold-off duration (causing measurement errors even in the order of tens of cm), and I determined the proper hold-off value in order to minimize its effects.

I performed a preliminary study to exploit the developed ToF camera for a 3D-ranging system in two different application scenarios: for a lander on Europa (Jupiter's moon), in collaboration with NASA, and for an autonomous driving system, in collaboration with an automotive company. Miniaturization and power consumption reduction are essential for both the applications, which therefore incredibly benefit from the already introduced novel FPGA ToF reconstruction algorithm. Moreover, in order to increase the effective fill-factor (starting from the 3.14% fillfactor of the SPAD+TDC array), I conceived an optical setup making use of patterned illumination, thus maximizing signal detection

rate without increasing background contribution. To this aim, I devised, developed, and tested indoor a setup based on a Diffractive Optical Element. which resulted in a 20x increased signal detection efficiency that allowed me to conduct 3-D ranging measurements at 10 m distance with just 500 µW average optical power at standard video frame-rate. By reducing the required illumination power and by rejecting background light, this setup enables the possibility to perform LiDAR acquisitions at longer distance range, with higher frame-rate and better eye-safety. For the Europa lander system, I also analytically studied a solution to minimize the impact of background detection rate by using SPADs connected in parallel, as arrays of Silicon PhotoMultipliers. Besides 3D ranging, the singlephoton sensitivity offered by SPAD detectors also makes it possible to detect each photon's ToF information, thus enabling the reconstruction of very faint and TCSPC (Time-Correlated Single-

reconstruction of very faint and fast optical signals by means of the TCSPC (Time-Correlated Single-Photon Counting) technique, applied on each independent pixel. Such capability also gave me the possibility to further exploit these SPAD cameras in different scientific fields, including Raman Spectroscopy and Fluorescence Lifetime Imaging Microscopy, which all get important benefits from single-photon sensitivity (e.g., in all biological applications a lowpower illumination is compulsory, in order not to damage the sample under study). Finally,

other important applications are



2. Combined 2D+3D image of a 10x10x20 cm3 statue acquired by the SPAD+TDC camera in 300 ms dwell time, with 50 scanning steps for increasing the lateral resolution up to 50 kpixels. Depth information, rendered in false color, reaches sub-millimeter precision

growing up in quantum mechanics and quantum computing, thanks to the possibility to study the behavior of each individual detected photon. In collaboration with several Universities and Research Centers, I exploited the cameras developed during my PhD in many of these applications, ranging from quantum mechanics to plethysmography and to computational imaging. **NFORMATION TECHNOLOGY**

MODELLING, DESIGN AND CHARACTERISATION OF RECONFIGURABLE TERAHERTZ ANTENNAS BASED ON GRAPHENE

Handouk Ahmed Hamada Radwan - Supervisor: Prof. Michele D'Amico

The crowded wireless communication bands in the gigahertz (GHz) frequency range and the ever -increasing demand for more bandwidth has motivated the exploitation of the terahertz (THz) band (0.1-10 THz). Antenna as a front component of THz systems is required to have a wide band, good radiation performances and sometimes switchable ability. To obtain the switchable ability of the antenna, the concept of a reconfigurable antenna was proposed. Reconfiguring an antenna is achieved through deliberately changing its frequency, polarization, or radiation characteristics. Design of reconfigurable antennas provides an excellent opportunity to combine stateof-the-art technologies with the antenna theory in an attempt to obtain additional degrees of freedom in system performance.

The main objective of this research is the development of design methodologies and their validation through the synthesis of reconfigurable antennas operating in the THz frequency based on the utilization of graphene, depend on the recent demonstration of the electric field effect in graphene, which have sparked enormous interest as a novel material in electronics. Therefore, it is significant to predict the performances of using graphene for applications in THz region. We started working on modeling of the Graphene's properties in simulation tools then continued on designing and simulation several reconfigurable THz antennas based on graphene with the ability to provide reconfigurability of the frequency band and radiation pattern through its chemical properties. In addition we worked on array of split ring resonators (SRR)

also made of graphene; this array actually behaves like a metamaterial, both bandwidth and radiation properties are optimized to obtain good results. Then reconfigurable beam antenna is proposed for THz application, which is based on a switchable high-impedance surface (HIS) using a single-layer of graphene. We deigned a beam reconfigurable THz antenna, which is based on a switchable Graphene High-Impedance Surface (G-HIS) that acts as a reflector for a primary radiator made of gold. We achieved a way to control the radiation properties over wide range of beam angles. Finally, we measured the reflection properties of the antenna based G-HIS at two cases (biased and unbiased graphene) and it demonstrated the effect of the change of applied voltage in terms of the input impedance and the reflection properties.

Marfia Fabio - Advisor: Prof. Marco Colombetti

Policies are a widespread approach for protecting users' privacy and security, and for allowing or enforcing users to abide by different norms and laws. More and more complex and distributed scenarios are requesting access control solutions, in the last years, where a centralized framework is required to manage a large set of functionalities related to policies. Those functionalities include Policy Editing, Policy Storing, Policy Harmonization, Policy Decision (or Policy Evaluation), and Policy Explanation. Such functionalities are described in detail subsequently.

388

- Policy Editing: it is the functionality of providing the user with an interface for creating, modifying and deleting formal policies, that are the expression of norms and laws in reality.
- Policy Storing: the functionality of storing permanently the policies, handling their consistency and coherence, managing backups and access concurrence to information.
- Policy Harmonization: managing conflicts between policies, as, e.g., policies including other policies, redundancy, or contradictory policies.

 Policy Decision: the functionality of providing an authorization decision (permit or deny) when a user asks for performing an action on an atomic resource, by reading the available set of formal policies.

Policy Explanation: providing the user with a readable explanation, together with the Policy Decision response, about the reasons for which the system returned such specific decision.

Different standards are presented in literature, modelling the general architectures and language protocols involved in such a type of framework. While different solutions where developed during the last decade, the current *de facto* standard in access control policy languages is represented by XACML.

The XACML standard defines protocols for transmitting credentials, requesting resources, defining and storing access policies; together with the definition of a general security layer, made up of different and specialised software components. Such a layer deals with the aforementioned tasks of allowing policy administrators to edit and store policies, handling conflicts between contradictory decrees, providing a ultimate response for access requests, together with an explanation of such a response eventually.

I present an implementation of an XACML-compliant framework in the thesis, based on the OWL 2 technological implementation for annotating semantics in the form of Description Logics (DL), and technology for reasoning on such semantics, as the Java reasoners Pellet and Hermit. Such implementation is able to perform the functionalities of Policy Editing, Policy Storing, Policy Decision, and Policy Explanation.

The expression and application of deontological propositions is well known in literature. However, as far as I know, this is the first time they are applied with the specific aim of providing a solution for an XACML security layer, even if activities for formalizing XACML policies with DL were done in the past, for Policy Harmonization purposes.

The approach allows the expression of complex and expressive policies as it is requested, I believe, from nowadays pluralistic scenarios. Furthermore, a knowledge base represents an interesting representation for policies, alternative to XACML, for advanced automatic components wishing to portray the regulation state for an end-user, or components that are intended to modify their behaviour by reasoning on provided directives. All of that, while allowing policy administrators to edit and store policies without revising any hard-coded software. Relying many core functionalities on DL reasoning activities requiring significant computational resources, Policy Decision performances are worse than any known XACML engine's. Consequently, the present approach has to be considered in cases where current XACML technologies are unable to capture the expressiveness of the policies involved, or if providing other advanced functionalities, as, e.g., Policy Explanation, is a necessary requirement.

The subsequent architectural components are defined in the XACML standard, and, therefore, in our framework:

- Policy Enforcement Point (PEP): Point which intercepts user's access request to a resource, makes a decision request to the PDP to obtain the access decision (i.e. access to the resource is approved or rejected), and acts on the received decision;
- Policy Decision Point (PDP): Point which evaluates access requests against authorization policies before issuing access decisions;
- **Policy Administration Point**

(PAP): Point which manages access authorization policies; Policy Information Point (PIP):

- The system entity that acts as a source of attribute values; Context Handler: the Context
- Handler deals with the coordination of the communications between PDP, PEP and PIP; in particular, it acts in order to return the output of the PDP to the PEP as a response for an access request, consisting eventually in a retrieved resource.

the PEP is the component that acts as interface between the external agent that sends an action request and the framework. For example, such component may deal with the task of checking for credentials sent by the user with the use of encryption algorithms, or with the task of monitoring or enforcing other agents in the system in their acts. The technology behind the PEP is strongly domain-dependent and it is not matter of the current work. However, it is present in the framework being a necessary part of each framework in each domain.

The **PDP** is the component that receives information about the user and the context from the PIP (in the form of the Domain TBox ontology and a subset of the Domain ABox ontology), and retrieves information about the active policies from the PAP, in the form of the Policy TBox ontology. It relies on the use of an OWL reasoner and it uses inference for generating the needed Policy Decision response for an action request (permit/denied). It also generates policy explanations, if requested.

The **PAP** is the component that deals with the tasks of Policy Editing and Policy Storing. In our solution, the policies are both stored in XACML format (for compatibility reasons of editor that are thought for such format), and an axiomatic representation (Policy TBox). A Policy Administrator edits the XACML policies that are converted into the axiomatic representation automatically.

The **PIP** is the component that stores the information about the domain context (Domain TBox) and the states of facts about the domain (Domain ABox). It returns information as a subset of such ontologies, if requested by the Context Handler.

The **Context Handler** is the central component that deals with the management of the request/response procedure workflow. The name comes from the definition of the standard user's request in the XACML specification, that is called Context. 389

NFORMATION TECHNOLOGY

Pagano Roberto - Supervisor: Prof. Paolo Cremonesi

Recommender systems always relied on personalization while context has been used only to improve personalization performance. Algorithms that decouple the context exploitation from personalization have not been explored. My work conceptualizes Context-Driven Recommender Systems, a new family of recommendation algorithms that moves from personalization to contextualization, exploring its applicability in five application domains. My research aimed at finding the best way to leverage context to overcome the limits of state-of-the-art recommendation algorithms, given an application domain and the challenges it poses to recommender systems. The "Contextual Turn" defines the context as the combination of what is going on around the user in the moment of the interaction with the system (situation) and on what she is trying to accomplish (intent) by taking the position that people have more in common with other people in the same situation, or with the same goals, than they do with past versions of themselves.

390

Several factors make contextdriven recommendation feasible: a huge amount of contextual data is generated by smartphones and IoT devices and then stored on the

cloud, and big data technologies make the computational power to process it finally available. The availability of such amount of data regarding the context paves the way for a revolution of the recommender systems: the shift from the static nature of the personalization, where user preferences are carved into the stones of their profiles and do not allow for the discovery of new items and interests, to the contextualization, where the user taste is drifting according to the goal to be accomplished. In the context-driven paradigm each user is not identified by its id, but by her contextual states, thus dissociating users from their historical behavior. In this way, users have more room to develop and are not tied to their past actions. Context-Driven algorithms are needed when an application domain exhibits at least one of the following challenges: taste development, seasonality, ephemeral catalogue, bounded capacity, life-cycle, and task and intent. Many application domains, such as Linear TV, Music, E-tourism, Mobile App and Job recommendation, call for recommendations that are actually driven by the context, thus relaxing the constraint that personalization involves recommendation for specific individuals.

The ephemeral catalogue property breaks the assumption on the steadiness of the item catalogue in the linear TV and Job domains. New TV shows or jobs are added to the catalogue very frequently as well as TV programs will not be aired anymore or jobs that are published for a long time lose their effectiveness. The same holds for seasonality in TV and E-tourism domains: many TV shows or hotels or flights are very much dependent on the season. For example, very few people go to ski facilities when there is no snow to ski on, or go to the sea in the winter season. For TV, many programs are aired in a particular season of every year. Human tastes naturally evolve and recommenders should be able to cope with it. The same person before or after being graduated, can look for different kinds of jobs (e.g. a part time job during studies and a well payed full time job that meets its aspirations when graduated). The same holds for Music or TV: the same user may change consumption patterns over time and situation. The availability of items is another issue that was not present in the very first application domain in which recommender systems were applied, i.e. digital movies. There are different domains where items cannot be consumed an unlimited number of times: this is the case of



e-tourism, where hotels or flights have a maximum capacity that limits their availability, or jobs, where each job request can be met only by a limited number of individuals.

The life-cycle property is relative to the item consumption curve in time: many items are very popular for a limited time, decreasing their appeal after this period. This property holds in the music domain, where the very popular songs start to fade out after being on the top of the charts for some time. The same happens for TV, where the first episode of a new show or series is very popular. This is also true for the app domain: many apps, especially games, are installed and uninstalled very frequently to give room to newer and fresher apps. Finally, the intent or task makes preferences to change or to be inconsistent with themselves according to situation, i.e.

preferences do not depend on user and item. There are domains where the task or the intent of the user is far more important than her past preferences. In the music domain, for example, the user might want to work out, so she might enjoy music that gives strength. Or the user might want to work, so she might enjoy relaxing and concentrating music. The task is even more important in the app usage prediction: if the user's intent is to look for a restaurant to have dinner, she may open a totally different set of apps with respect to the set of apps used for reading news. There are many advantages for context-driven recommendation. First of all, it solves the cold start problem in the algorithmic fabric, since recommendations target context instead of past preferences. This means that also non-logged or nonregistered users can get their

recommendations. Secondly, it can actually "pop" the filter bubble, the effect of receiving a narrower and narrower set of recommendation because of the bias of past user preferences. Then the explainability of recommendations can benefit from the contextdriven paradigm, because the dominant contextual variable that led to the recommendation can be used for explanation. Lastly, it has advantages for the privacy of users, since personal information does not have to be stored, and for the compliancy with the new EU regulations on data retention (14th April 2016).

Palluzzi Fernando - Advisor: Prof. Gaetano Ivan Dellino, PhD (IEO, Istituto

Europeo di Oncologia)

Genomic Interaction Modeling

Genomes are **complex systems**. and not just static and monolithic entities. They represent the ensemble of the hereditary information of an organism or, in evolutionary terms, of a species. Sequencing sciences offered through the last two decades the possibility to dive into this complexity. This scientific revolution led to a series of paradigm changes, and the collateral need of efficient technical and methodological innovations. However, this process was not gradual, exposing genome research to an explosive **data deluge**, thus causing a series of **bottlenecks**: technological, computational, and **methodological**, due to the existence of efficient data reduction and analysis algorithm, capable of capturing the **informative** portion of genomic complexity. In the last years, there have been huge efforts towards the generation of unified standards and databases, ontologies, integrated platforms, integrative methodologies, to provide a strong theoretical background, capable of capturing true biological variation, often masked behind marginal synergic processes. The goal of this thesis was to shed light on this last problem, trying to expose a

unifying theoretical foundation needed to approach the study of genome complexity. We showed that this unifying theory actually exists, and it is represented by the **Kullback-Leibler Divergence** (KLD):

$I(f,F) = \int f(X) \ln \frac{f(X)}{F(X)} dX$

KLD is further supported by the sampling theory, that allowed us to **produce** and **rank** different representations of reality. We gave a generic definition of our modeling paradigm as the **Genomic** Interaction Model (GIM), that is deeply founded over data (i.e. it is data-driven), and expressed through the generic **Genomic** Vector Space (GVS) data structure, composed by a **sampling space** *S*, a variable space X (that is defined by data itself, represented by sequencing tag density profiles), an **architecture** over data *H(X)*, and a monotonic variable *T* defining its evolution. In other terms, GVS = { *S*, *H*(*X*) }_{*r*}. The architecture *H* can be represented in different ways, from the more traditional **Generalized Linear Models**

(GLM), to the more complex interaction networks, to the less used Structural Equation Models (SEM). Responding to the need for parallel computational solutions, we developed the GenoMetric Query Language (GMQL) as an Apache/Pig-based language for genomic cloud computing, ensuring clear and compact user coding for: (i) searching for relevant biological data from several sources: (ii) formalize biological concepts into rules for genomic data and region association; (iii) integrate heterogeneous kinds of genomic data; (iv) integrate publicly available and custom data. However, similarly to other available methods for genomic data modelling and algebra implementation (such as BEDTools, BEDOPS, GQL, and GROK), a declarative-like approach may underperform in modelling complex interactions. In general, feature definition may involve multiple combinations of variables, including **distance**, a (continuous) dynamic range of binding, transcriptional, and epigenetic signals, with interaction terms connecting all or some of these variables. Here we come to one fundamental definition of this work: In general, a **genomic feature** is represented by a specific chromatin state, defined through *context-specific* variables that can be profiled using sequencing technologies, and which relationships can be modeled and tested through an interaction model. In this way, we can represent every probable chromatin state (as a



KNOWLEDGE

-

functional element) and possible alteration (i.e. **physiological and pathological functional change**).

There are at least five main conceptual improvements in this definition:

- (i) Genomic features are not fixed entities.(ii) Genomic features are
- knowledge-independent. (iii) Unknown genomic features are included in the modeling framework.
- (iv) Time-variant interactions are part of the feature definition.
- (v) Genomic coordinates identify an instance of a chromatin state.

Research applications in Molecular Medicine

Our approach demonstrated its flexibility and efficacy in two different human complex disorders. Specifically, our interaction-based statistical modeling strategy allowed us to investigate the basal mechanisms of DNA damage occurrence in **breast cancer** (BC), and to discover new, subtle and synergic biomarkers in the etiology of frontotemporal dementia (FTD). In the former, we applied an ensemble logistic-RFC model (based on Random Forest Classifiers), that evidenced the

ing. Pre-Pro

Distribution (HTTP, FTP, DAS, ...) **Big-Data Centers**

(NGS, GWAS, Microarray, ...)

Drect Acquisition

(In-house hardware facility)
Profiling

(Viewing, Descriptive statistics, Clustering, Machine Learning)

Global Association Testing

Signature

Funcional Element

sociated Genomic

relation, Stratification, Trend Enrichment Analysis)

accumulation and causal effect of gene length and RNA Polymerase Il in promoters of damaged genes, significantly associated with the occurrence of translocations and gene fusions in breast cancer patients. Our methodology enabled us also to quantitatively measure thresholds for single and interacting predictors, excluding other potential causes of DNA damage (including gene expression and the presence of common fragile sites). The overall predictive accuracy of our algorithms was ~83% (i.e. we correctly predicted more than 8 damaged promoters every 10 trials, based only on functional signals at promoters, including gene length, RNA Polymerase II, histone H3K4me1, and gene expression). We also demonstrated that there are specific genomic regions, i.e. intronexon junctions, may be prone to pathogenic translocation events in BC, in which the probability of a DNA double-strand break (DSB) to cause a gene fusion is directly

proportional to the distance between the DSB event and the iunction. In the latter project, we identified novel FTD biomarkers, and characterized the perturbation of their interactions, to generate a novel oxidative stress-based theory for the etiology of this complex neurological and behavioral disorder. In this case, we instantiated the genomic interaction model into a SEM**based network**, coupled with a weighted shortest-path Steiner's tree-based heuristic algorithm. Firstly, the molecular basis of this theory includes Calcium/ cAMP homeostasis and energetic metabolism impairments as the primary causes of loss of neuroprotection and neural cell damage, respectively. Secondly, the altered postsynaptic membrane potentiation due to the activation of stress-induced Serine/Threonine kinases leads to neurodegeneration. Our study investigated the molecular underpinnings of these processes, evidencing key genes and gene interactions that may account for a significant fraction of unexplained

FTD etiology.

NFORMATION TECHNOLOGY

Validation

(Cross-Validation, Accuracy, ROC, Enrichment Analysis)

Interaction Model Building Linear/Structural Modelling, Factor Analysis, Non-linear Modelling)

Data Reduction (FDR, PCA, Kernel-based, Treebased, Network-based)

Genome Vector Space

Pinoli Pietro - Supervisor: Prof. Francesco Amigoni

In the last decade, new technologies for fast DNA sequencing, known under the name of next-generation sequencing (NGS), had just become available; they enable reading the whole genome much faster, at higher resolution, and at lower cost, thereby giving us the data to answer fundamental biological questions and opening the ground to personalized genetic medicine. Huge investments were, and they still are, targeted to sequencing the DNA of large populations, with repositories of wellcurated sequence data being collected and made public for investigation. Answers to many biomedical problems are hidden in these data, e.g. how cancer arises, how driving mutations occur, how much cancer is dependent on environment. While genetic sequencing are mature, a quantum leap is needed for building the computing infrastructure at the receiving end of DNA sequencing machines. In particular, genomic data management is struggling on the initial problem of storing the data which are fast produced by biologists in their laboratories. A powerful data infrastructure is required for going beyond pure storage, and enabling viewing, querying, analyzing, mining, and

searching over a world-wide available collection of genetic data.

This Thesis describes the GenData System. The first innovation proposed within GenData is the organization of genomic data through a uniform data model, the Genomic Data Model (GDM), capable of expressing the various features that are embedded in the produced biomolecular data, or in their correlated phenotypic data that can be extracted from clinical databases, or in the information inferred by applying data analysis methods to them. The most relevant feature of GDM is the subdivision of data into region data and metadata. Region data consists of structured data that captures the processed outcome of any NGS experiment. Metadata is a set of free attribute value pairs meant for representing biological and technical details of the experiments as well as clinical and phenotypic information about the sample donor. The vision of the project is completed by methods for querying, searching, and analyzing genomic data. While several prototypes were developed, this Thesis focuses on the GenoMetric Query Language (GMQL), a novel declarative high level query language, aimed to express complex queries on GDM instances for answering

biological questions. It combines relational algebra and domain specific operators for genomics. GMOL has been implemented upon cloud based platforms; this choice was dictated by the big-data nature of genomic data. The Thesis describes many aspect of such implementation, mainly regarding the translation strategy for GMQL, the generation of an execution plan from a query, its logical optimization and the efficient algorithms designed for computationally heavy domain specific operations. In its last release, GMQL is initially compiled down to an intermediate representation, which consists of a graph where nodes represents atomic transformations and edges dependencies among the operations. The intermediate representation of a query is then interpreted and executed by an engine that actually elaborates genomic data stored within the GenData repository. Having an intermediate representation provides a framework for developing many optimization strategies, based both on the re-writing of the graph and on efficient implementation of single blocks. Most notable results comprise the so called "meta-first" optimization and the "binning algorithms". The "meta-first" is a runtime optimization that

leverages the fact that metadata are usually much smaller than region data; by tracking the execution of metadata it is possible to avoid unnecessary and heavy computation on region data. The "binning algorithms" are a class of distributed and highly parallel algorithms for the implementation of genomic operations; they are based on the splitting of genome into small portions, named bins. Although binning algorithms have been designed with genomic operations in mind, they are suitable for processing any interval based information, such as spatial or temporal data. Two real world case studies, where the GenData system was used for answering complex biological questions, conclude the manuscript. The first of them consists of a study of the 3D organization of the DNA across several human tissues and an investigation of how a variation of that structure may be linked with the rise of cancer. The latter explores how certain mutations, occurring within enhancers (particular DNA regions involved in the regulation of the gene expression), are associated with genetic diseases or phenotypes.

AUTOMATED MALWARE BEHAVIORAL ANALYSIS

Polino Mario - Advisor: Prof. Stefano Zanero

In 1987, Cohen introduced the concept of computer viruses. suggesting theoretically the possibility for a program to perform malicious actions, act out of the control of the users, and remain undetected. Since when it was just a theoretical concern, malicious software has evolved enormously. At first malware samples were written to prove author's skills and talent, like the Morris worm that in 1988 caused a massive denial of service on computer connected to the Internet of that time. Of course, there was no financial motivation behind the development of Morris' experiment: "The goal of this program was to demonstrate the inadequacies of current security measures on computer networks by exploiting the security defects that Morris had discovered". Instead, today's malware programs are a prominent threat for users of personal computers and mobile devices alike. Malicious programs

devices alike. Malicious programs are written to steal data and gain profit, implementing behaviors that range from stealing online banking credentials to locking or encrypting files belonging to unaware victims to ask for a ransom payment.

Writing malware is not anymore an art of a small elite, nowadays malware can be bought online and configured with practical user

interfaces. Such an easy availability induces an enormous increasing of new malware samples. In fact, KasperskyLab reports 121,262,075 unique malware samples have been detected by their AV during 2015. Moreover, Symantec reports that the daily release of new malware samples is approaching 1 million units. Many researchers around the world every day have to analyze hundreds of thousand new malicious samples. Even if the identification of different samples of the same malware is possible using simple comparison methods, to understand a never seen before malware binary we need the process of reverse engineering. The increasing interest around reverse engineering of untrusted or malicious binaries demands for automated tools that aid the analysts. Past works such as Howard or TIE propose intelligent solutions to tackle some of the hard aspects of "reversing". We believe that research efforts in this direction are needed to turn reverse engineering from an art into a more structured engineering discipline with appropriate methodologies, tools, and computer-supported processes. Avoiding tedious and timeconsuming human labor in reverse engineering is key in

order to achieve a faster and more complete understanding when analyzing a new binary. One of the main challenges in reverse engineering is achieving automation in order to overcome the shortage of skilled analysts. To tackle all aspects of reversing, an automated tool needs to be able to overcome also malware authors' attempts to avoid analysis. The various approaches to perform automated malware analysis are usually divided in two categories: Static and Dynamic. We usually refer to the first for all the techniques that do not necessitate to execute the binary. The latter category includes all approaches that execute the analyzed binary. A variety of static- and dynamicanalysis tools exist and are very useful. A further step was achieved with the so-called "hybrid analysis" (i.e., the use of both type static and dynamic approaches) that balances their symmetric strengths and weaknesses. We believe that hybrid approaches can be pushed forward and leveraged to obtain better reverse engineering tools. The core aim of hybrid analysis techniques is to help bridging the semantic gap between static and dynamic analysis, using as a pivot the concept of behavior, expressed in different ways and abstraction levels (e.g., groups of API or system calls, instruction sequences). Thus,

the automatic identification of such behaviors is an interesting and challenging research problem with immediate and profound practical impact. Such an output could, for example, be used by a plugin for reverse-engineering tools to automatically highlight and annotate certain portions of the CFG to prioritize the analysis based on the information extracted from a large, collaborative back-end database, freeing up valuable analyst time to focus on novel, interesting behaviors. Such behaviors could then be fed back into the behavior database To make the analysts' task harder, malware authors developed various techniques to counteract and evade both static and dynamic analysis. Malware that employ such defensive techniques is defined evasive malware. To defeat static analysis, malware authors started to implement tools that now are known as packers. Because static analysis rely on understanding binary code, packers encode the whole binary so it becomes not intelligible to the analyst. In order to be executed, that binary needs to be decoded, so such packers insert a stub code that is executed at the beginning to decode the real malware code. Packers can be much more complex than the one aforementioned. Because our goal is an automated tool, we need to develop a technique that is able to automatically recover the original code of a malware sample. Such tools are known as unpackers. The problem of unpacking is well studied in literature several works have been proposed both

for enhancing the end user's protection and supporting the malware analysts in their work. Different approaches exist in order to build a generic unpacker like debuggers, kernel modules, hypervisor modules, and dynamic binary instrumentation (DBI). Dynamic analysis allows to detect all the interactions between a malicious program and the underlying operating system making us to understand behaviors of an analyzed program. Usually dynamic analysis is implemented using instrumentation of a virtualization environment. In this way malware, can be easily contained in the sandbox without causing any harm to sensible systems. However, malware implement a lot of techniques in order to change or hide their behavior if they detect an attempt of dynamic analysis. These countermeasures are called in literature anti-debugging and anti-sandboxing techniques. Often those techniques rely on imperfections of virtualization environments. Our research through malware analysis led us to develop a tool named *lackdaw* that is able to extrapolate from several malware

samples common behavioral

most evasion techniques

pattern. Jackdaw is assisted by two

modules that are able to overcome

implemented by malware authors.

The output of our system is a

list of high-level dynamic traces

enriched with static information

it can be used as input to other

hybrid analysis systems that

need behavior definitions to

work, thus relieving the analyst

from the burden of producing behavior specifications manually. It could also be used to build binary clustering or classification techniques. Finally, it could be used by a plugin for reverseengineering tools in order to automatically annotate portions of the CFG that implement behaviors. In summary, our contributions are the following:

- We develop a system to assist malware analyst during reverse engineering of new samples.
- We present an unsupervised approach to ease and systematize the task of reverse engineering by automatically extracting high-level behavior descriptions from a large dataset of binaries.
- We remove a time consuming manual step in hybrid staticdynamic analysis processes, namely the definition of high level behaviors.
- We propose an automatic algorithm to associate semantic tags to behaviors, allowing (inexperienced) analysts to understand their actions.
- We build a generic unpacker exploiting the functionality of a DBI framework.
- We develop a dynamic protection framework that can be used to defend PIN, one of the most used and supported DBI, against antiinstrumentation attacks.
- We implemented a generic as possible countermeasure to defeat every class identified.

396

397

_

INFORMATION TECHNOLOGY

MODEL PREDICTIVE CONTROL BASED METHODS FOR MICROGRID ENERGY MANAGEMENT

Raimondi Cominesi Stefano - Supervisor: Prof. Riccardo Scattolini

The past twenty years have witnessed a continuous growth in the diffusion of distributed energy resources, driven by the improved efficiency of the most recent cogeneration units, as well as by the increment of renewable energy production, connected to a global effort to reduce greenhouse gases emissions. In this scenario, with the electrical generation steadily transitioning from few large power systems to a multitude of small-size scattered units, the need for a decentralized control infrastructure has arisen. The most promising solution to the issue, at the present day, is the one suggested by the microgrid paradigm.

The term microgrid is used to identify a portion of the lowvoltage distribution network where generation capacities are present, together with electrical loads and, eventually, energy storage devices. Located downstream with respect to an electrical distribution substation, a microgrid is perceived as an individual unit from the utility grid, to whom it is connected by means of a single interface point: the Point of Common Coupling. Opposed to the classical grid configuration (often referred to as *macrogrid*), where a limited number of power stations are entitled to provide the energy for all the loads, supplied

by means of a capillary network, the introduction of microgrids is meant to define a new paradigm for the electrical network management. By regulating the energy exchange through the point of common coupling, in fact, these subsystems of the distribution network become active resources for the main grid energy management, potentially capable to take part in the energy market, as well as to provide ancillary services to the network itself. This actual decentralization of the main grid control, would furthermore result in an increase of the flexibility of the overall system since the microgrids acquire the ability, if needed, to interrupt the connection with the rest of the electrical network and self-sustain themselves in what is referred to as *islanded operation*. The benefits connected to the microgrid paradigm implementation are multiple and easily conceivable: the possibility to have production units in proximity of the consumption, for once, allows to considerably reduce energy transmission losses and, potentially, to curtail the very cost of the overall transmission infrastructure. The ability to detach itself from the rest of the utility network at need, ensures the microgrid to be independent and resilient to any

external malfunction, with clear benefits not only for hospitals and military bases, but also for those manufacturing industries and activities that are gravely endangered by blackouts and brownouts. Furthermore, the chance to tailor the production on the consumption itself, and the possibility to introduce cogeneration, may also result in a considerable increase of the overall system efficiency, with clear economical benefits, and consistently with the global target to pursue a significant carbon footprint reduction. To enable these advantages, though, to successfully implement, and fully exploit the potential of the microgrid concept, the development of a dedicated smart metering infrastructure able to provide the means for communication between all the devices is required, as well as an efficient control architecture, capable to properly manage the resources of the system and efficiently regulate the energy exchange with the main grid. The major threats to an efficient microgrid energy management are represented by the intrinsically non-linear dynamics of the system (e.g., switch-on/switchoff dynamics of the controllable units), and the uncertainty due to the presence of non-dispatchable

generation and non controllable loads. Fluctuations of noncontrollable components are even more critical for the islanded operation of the microgrid. In this case, in fact, due to the lacking connection to the main grid, the controller has necessarily to integrate additional functionalities to preserve the stability of the system.

Consistently with the multiple operational constraints characterizing different microgrids, as well as to exploit the eventually available forecasts, the implementation of these control schemes requires complex, modelbased algorithms. The purpose of this work is to present novel efficient energy management control algorithms for microgrids, both in grid-connected and islanded operation, to effectively cope with the uncertainty associated to renewable energy source generation and electrical load fluctuation. In the first part of the Thesis, microgrids connected to the main grid are considered, and some

relevant management and control problems are studied, while in the final Chapter the main issues related to microgrids operating in islanded mode are addressed. Specifically, a novel two-layer control scheme is firstly proposed: the higher layer of the structure is in charge of coordinating the microgrid resources, i.e. dispatchable and non-dispatchable generators and storage devices, to optimally manage their use, as well as the energy exchange with the main grid. At this stage, simplified models of the microgrid components are considered,

logical constraints related to their use are included, and an economic criterion is adopted. The resulting control problem is solved with Model Predictive Control (MPC) working at a slow time scale and with a long prediction horizon. Typically, the adopted sampling time is fifteen minutes, and a twenty-four hours time horizon is considered. At the lower level, a stochastic MPC (sMPC) algorithm is proposed to compensate for the uncertainties due both to renewable generators based on wind or solar radiation, and to unpredictable load fluctuations. At this stage, the main goal is to guarantee, to the maximum possible extent, that the energy exchanged with the national grid is equal to the value negotiated beforehand to avoid penalties. In this analysis, a quite simple microgrid is considered, including a dispatchable microturbine, a non-dispatchable photovoltaic system, an energy storage device, and a critical load. The analysis is then partially extended to the case of microgrids equipped with thermal units. It is in fact well known that the microgrid's overall efficiency greatly improves when electrical and thermal production are combined and integrated. A microgrid comprising a trigenerative unit, a heat pump,

a boiler and a thermal storage

simulation analysis.

system is hence considered in a

The two-layer control structure

described, however, can potentially

lead to computational expensive

optimization problems when the

number of devices of the microgrid

becomes large. To tackle this issue,

an innovative hierarchical scheme is subsequently described, where it is proposed to group devices with similar behavior in specific clusters, so as to obtain cumulative simplified models, and effectively reduce the computational effort of the centralized structure managing the microgrid.

In the final chapter, instead, a hierarchical control structure for the islanded operation of a microgrid is proposed. A MPCbased secondary control is designed to regulate the network variables of the microgrid, as well as to economically redistribute the production among the devices, while a decentralized primary regulator – based on the droop control paradigm – is entrusted with the stability of the system. The potential of the presented algorithms have been illustrated by means of numerical simulations.

A SYNTACTIC-SEMANTIC APPROACH FOR INCREMENTAL PROGRAM VERIFICATION OF MATCHING LOGIC PROPERTIES

Rizzi Alessandro Maria - Advisor: Prof. Carlo Ghezzi

Software is not built monolithically. Its construction requires a sequence of small steps. This vision is supported by modern software development methodologies, such as Agile approaches. In fact, software is continuously subjected to changes; either in the specification and requirements, due to changes in the environment or in user needs: either in the implementation, in order to fix a software fault or to implement a missing functionality. In both cases, change is a distinctive feature of software systems, affecting almost every phase of its life-cycle, which should be considered in every development process. Formal software verification techniques are becoming more and more effective in helping the developer of software systems; on one side due to improvements of the formal techniques, which allows more powerful abstraction models, e.g., separation logic; on the other side thanks to the fast growing pace of hardware capabilities, which allows bigger and bigger amount of computational power and memory. However, formal verification technology is still not widely adopted for ensuring software correctness. One of the main reasons is the high demand of computation

resource which affects every formal verification technique due to the high computation complexity of the problems involved. For example, let us consider *model checking*, which is a major formal technique based on the exhaustive exploration of the reachable software states. Its applicability is hindered by the so called state explosion problem, i.e., a combinatorial explosion of the number of states which have to be visited by the model checker. It is possible to exploits the incremental nature of software systems to develop more efficient verification techniques. Many of the existing formal verification approaches do not exploit the fact of having a sequence of different software versions which change relatively little one to another; instead they reapply the verification process from scratch to each version. I address this problem by developing formal verification techniques capable of incrementality: techniques which are able to reuse the intermediate results obtained with the verification of an early version of a software system when a new version of the software system is considered. My goal is to provide formal verification as a feasible and practical way

of assessing quality in software

products. Nowadays, the most applied verification technique is *testing* which has gained popularity with the introduction of Agile methodologies. It is applied to evolving software system in the form of *continuous testing*: software quality is assessed by automatically applying a set of tests to each version of a product. My ultimate goal is to provide an incremental formal verification approach which can complement testing, as a practical and reliable verification technique still providing strong formal guarantees of correctness. I target incremental verification techniques based on the syntactical structure of the software artifact being verified. The formal language which defines the software artifact can discover the changes occurred at a syntactical level: Software artifacts are expressed by operatorprecedence grammars, which, thanks to properties such as locality, provide ways to efficiently reparse only the parts affected by the change. This incrementality obtained at the syntactic level is then translated to semantics by means of the attribute grammar formalism. In particular, by using synthesized-only attribute grammars, the semantic effect of a change is limited to a portion of the abstract syntax tree defined

by the program. This idea has been developed in the SiDECAR framework, a system for defining a incremental verification technique on a given programming language. In my work, this general technique is applied to matching logic properties verification. Matching logic is a Hoare-like verification system rooted on a formal definition of programming language semantics. I have analyzed the case of KernelC language. KernelC is a subset of the C language having interesting features as heap, loops, function calls and recursion. In my work, I have developed a matching logic verification framework expressed as an evaluation of semantic attributes. Upon this approach I have built my incremental verification technique. This approach has been implemented using the framework SiDECAR, for the incremental matching logic verification of KernelC programs. The advantages of this approach, with respect to existing incremental verification techniques, are 1) an extremely fine-grained change discovery: with this approach changed can be confined even within a single statement, while common state of the art techniques target the entire procedure; 2) no preliminary division of the software: this approach can

detect the part affected by the changes in software without prior subdivision: 3) completely automatic verification: no human intervention is required, after the specification are provided, the tool can automatically discover the changed parts which are automatically checked; 4) support for any kind of change: the approach can process any change which can possibly be made without being specifically configured for it. I experimentally evaluated my incremental verification approach over a large suite of programs, containing many versions of each program; along with the stateof-the-art, non-incremental tool MATCHC for the verification of KernelC programs annotated with matching logic properties. These benchmark includes samples provided with MATCHC, examples from state of the art, software verification benchmarks, and wellknown data structure algorithms. The obtained results show that, even if program sequentiality limits incremental speedup, incrementality has always benefits and, in modular programs, this gain is pretty important.

INCREMENTAL LARGE-SCALE VISUAL 3D MESH RECONSTRUCTION

Romanoni Andrea - Supervisor: Prof. Matteo Matteucci

In the last decade, the growing interest about autonomous driving brought many computer vision and robotics researchers to focus on autonomous vehicles mapping and understanding their surrounding environment. In this context, a map is needed to plan a path to reach a specific destination, or to estimate the current vehicle position by comparing its percepts against a reference model. In robotics, a suitable mapping/reconstruction algorithm needs to be scalable, incremental and to provide a dense map. Scalability is needed especially in large environments; incrementality allows map update as new data are acquired; density enables a consistent and coherent navigability.

Researchers in computer vision have focused on accurate and dense results, disregarding any incremental processing, and only few works show large-scale capabilities. Conversely, in robotics the focus has been mainly on incremental algorithms, but the output maps are usually point clouds and only a very limited amount of works estimate dense and continuous surfaces, however limited to small scale scenes. In this thesis we build a bridge between these two communities through the first automatic and incremental pipeline able to build

a dense, scalable and manifold mesh, from a set of images. The key aspect of our proposal is to represent the scene as a triangular manifold mesh; the mesh representation allows us to reconstruct large-scale scenes though continuous meshes; the manifold property allows for coherent photometric mesh refinement. The building blocks of our pipeline are essentially two: incremental reconstruction from sparse points and incremental mesh refinement. The former step estimates a manifold mesh from the output of a Structure from Motion algorithm or any algorithm providing camera poses, sparse point clouds, and camera-to-point viewing rays (e.g., RGB-D and laser based systems). The latter step refines the resulting mesh according to the appearance provided by the images. Our incremental reconstruction relies on a volumetric representation of the space based on the Delaunay triangulation; each tetrahedron of the triangulation is voted by the camera-to-point viewing rays crossing it and the manifold mesh is extracted as the boundary between those tetrahedra receiving high votes and those traversed by few or no rays. We investigated which kind of 3D points are suitable to build a

Delaunay triangulation for 3D reconstruction, in particular in case of urban video sequences. Since 3D points on real-world edges usually project on points belonging to images edges, named 2D edge-points, we inverted this process by estimating the 3D position of 2D edge-points by tracking and triangulating them. When we build the triangulation upon reconstructed 3D edgepoints, we bias the edges of tetrahedra connecting them to lay on real-world edges. This improves the representativeness of the triangulation and the accuracy of the reconstruction. We further improve our incremental algorithm with a novel voting scheme that generalizes state of the art results and improves the accuracy of the reconstruction; finally, we propose a simple, but effective, method to efficiently handle moving point inside the triangulation.

In low frequency sequences or in an unordered set of images where feature tracking fails, we noticed that a manifold mesh estimated from sparse data can sometimes be too far from the true 3D model of the scene. This would cause the subsequent mesh refinement process to get stuck in local minima or to converge slowly to the solution. Therefore, we improved the accuracy of the manifold mesh in order to speed up the refinement convergence and improve its the accuracy. To obtain this, we sweep the initial mesh in the space and we add to the point clouds the new 3D points for which the photometric matching score induced in a pair of camera is very high, which likely means that these points belong to the real scene.

Incremental mesh reconstruction. and mesh sweep refinement, output incrementally a lowresolution manifold mesh that feeds the refinement process of a novel incremental photometric refinement pipeline. Indeed, we proposed a novel approach to build an accurate map of the environment according to which, as a new partial mesh is reconstructed obtained, we evolve it photometrically and we merge it with an already refined mesh thanks to a novel merging algorithm which keeps the manifold property valid. Aside from our pipeline, and motivated by the availability of lidar data often collected with images in autonomous driving applications, we tested and extended our mapping pipeline to work with hybrid lidar and image data. We proposed a complete mapping framework that leverages on the high accuracy of the 3D lidar data and on the



1. Result on Middlebury dataset

appearance and density of images. By detecting and removing the moving points from lidar point clouds, we reconstruct the map of the scene with our reconstruction algorithm and we are able to refine it by neglecting the moving objects from the photometric optimization. Finally we recover a full textured map where moving points have been filtered out.



2. Result on EPFL dataset

We tested our proposals against different publicly available dataset, which provide varied scenarios in order to stress the flexibility of our approach: the Middlebury dataset for small scenes (**Figure 1**); the EPFL dataset for larger scenes with few high resolution images (**Figure 2**), the KITTI dataset for large scenes captured by long image sequences (**Figure 3**).



3. Result on KITTI dataset

403

NFORMATION TECHNOLOGY

CONTROL OF AN UNMANNED AERIAL VEHICLE EQUIPPED WITH A ROBOTIC ARM

Rossi Roberto - Supervisor: Prof. Paolo Rocco

Robots ability to autonomously move in the terrain, underwater and aerial domains opened a large number of new applications in industry, service robotics and search and rescue activities. In particular, the impressive growth of aerial robotics during the last decade paved the way to the birth and subsequent development of the new research field of aerial manipulation. The increase of payload capacity and dexterity of aerial vehicles, in conjunction with advances in the design of light-weight arms, enabled the conception of flying robots equipped with articulated manipulators. Although first demonstrations of aerial manipulation activities are promising, this research field is still characterized by low maturity, and several open research challenges in the areas of perception, control and planning require further investigation. The work aims at increasing the degree of autonomy and improving the control performance of an aerial robot composed of a quadrotor equipped with a robotic manipulator underneath. One of the most critical research issues aimed at providing autonomy to aerial manipulators is the stability control of the entire system. The stability control of unmanned aerial vehicles with a

robotic arm is per se a challenging problem. The forces and torques exchanged at the physical junction between the two subsystems, the flying base and the arm, impose a dynamic coupling that, in most of the cases, cannot be neglected in the model and control strategy. The compensation of these mutual actions is difficult for the typical actuation structure of multi-rotors aerial vehicle. In fact, vehicles as helicopters, guadrotors and similar configuration aircrafts, are under-actuated, thus they present a number of actuators lower than the number of degrees of freedom of the body. As a consequence, the compensation of disturbances is performed with a variation of the vehicle attitude, in order to orient the propellers axes in the suitable direction to provide the required control actions. When a robotic arm is attached to the vehicle base, the 3D rotation of the aircraft is affecting the absolute position of the robot itself, and inertial torques and forces are exchanged between the two systems. In the thesis, a control approach addressed to this specific problem is presented. In particular, the

main advantage of such platforms, the dexterity and the redundancy of the system, is exploited in order to compensate the motion of roll and pitch angles, needed

for the control of the translational degrees of freedom of the quadrotor, with the redundant variables of the system. A suitable change of variables, entailing the mentioned compensation, is applied, and an inverse dynamics control is proposed. The stability of the control system is demonstrated by means of Lyapunov analysis. Notice that the effectiveness of model-based controls that consider the coupling on the dynamical level, as the present one, can be mitigated by inaccurate sensors information. model errors or communication delays. Nevertheless, the validity of the proposed approach is demonstrated with an extensive simulation campaign, and an improvement of the tracking performance is obtained with respect to a state of the art algorithm. The application of the proposed algorithm can guarantee higher impact on aerial manipulation performance for systems with accurate sensors information, compared to the size of the system itself. Notwithstanding the described control challenges, the configuration composed of a multi-rotor aerial vehicle and an articulated robotic arm is characterized by high dexterity, as mentioned above. Such systems commonly present kinematic



redundancy, thus they are characterized by a number of degrees of freedom greater than the main task to be performed. Thus, the additional degrees of freedom can be used to perform secondary tasks, or to maximize some performance criteria as a function of the robot joint configuration. To this extent, a trajectory generation algorithm, capable of solving the kinematic redundancy in order to satisfy some given tasks, takes particular relevance in the field of aerial manipulation. Moreover, the mentioned stability issues and the complexity of the physical structure of the system impose hard constraints to the set of joint configurations which can be taken by the system. Therefore, an algorithm for generating a suitable joint trajectory is not only recommended in order to exploit the redundancy of the system to perform multiple tasks, but it is also a critical requirement to avoid dangerous system conditions.

in the experiments, consisting on a 3DR-X8 coaxial multirotor platform with a custom-built 4 degrees-of-freedom serial

In the thesis, two trajectory generation algorithms based on constrained optimization are presented. To the author knowledge, they represents the first demonstrations of optimal trajectory generation performed on line and on board of an aerial manipulator. Both methods allow to perform a number of prescribed tasks, while satisfying equality and inequality constraints. The first approach is based on a second order local approximation of the system, and the generation of an optimal trajectory is converted into a quadratic programming problem, whose regressor variables are joint accelerations. The second approach applies a nonlinear model predictive control strategy, where the control action is represented by joint velocities. The differences and advantages between the two methods is compared and discussed. With respect the perception research field, the suitable sensing

obtain an estimate of the state of a small size aerial manipulators are analyzed and described. In this part, the attention is drawn also on technological issues. A specific theoretical contribution consists in a method for the calibration and performance evaluation of a camera, mounted on the quadrotor, employing an industrial robot. It represents a method particularly useful for academic laboratory and industry, where motion capture systems or similar technology are not available. Finally, a contribution to the control of interaction of aerial manipulators with the external environment is provided. In particular, a method to limit the damages in a potential impact of the aerial manipulator with an external object is proposed. The approach has been conceived in the context of safe human-robot interactions, in order to bound the potential injuries to human workers caused by accidental collision with collaborative robots. The application of the method to the aerial manipulation field is useful when particularly fragile equipments or surfaces are manipulated, whose integrity can be undermined by the aerial robots itself. An index to define the energy dissipated in potential inelastic impacts for a generic manipulator is proposed, which describes the severity of damages occurring as a consequence of a potential impact. Moreover, a constraint-based hierarchical control strategy is applied to perform prescribed tasks while bounding the proposed index under a certain threshold.

system and algorithms needed to

NEAR-INFRARED SINGLE-PHOTON DETECTORS IN EMERGING TECHNOLOGIES

Sanzaro Mirko - Supervisor: Prof. Alberto Tosi

A growing number of applications rely on single-photon detectors with high detection efficiency in the near-infrared (NIR) range. Advanced applications can be found in many fields, such as, as measurement of fluorescence decays (in physics, chemistry and biology), non-invasive testing of VLSI circuits, eye-safe timeof-flight laser ranging (LIDAR) and three dimensional imaging, quantum cryptography and quantum computing. In order to bring these applications out of the laboratories, singlephoton detectors have to be compact, low power and rugged. Solutions based on photomultiplier tubes (PMTs) and superconductive detectors, such as Superconductive Nanowire Single-Photon Detector (SNSPDs), are either bulky and fragile or require cryogenic cooling (< 4 K). On the other hand, Single-Photon Avalanche Diodes (SPADs), being solid-state devices, are the preferred solution for most applications. The aim of this doctoral work was to design novel NIR SPAD structures and to contribute to the advance of the state-of-the-art in the field. NIR radiation (0.75 µm – 1.7 μm) is quite broad for a single semiconductor detector, and is usually broken down in two subranges of interest. The sub-range

0.95 µm – 1.65 µm, which contains optical fiber communications wavelengths, is important due to the increasing interest in quantum information applications, such as quantum computing and quantum cryptography, and is served by InGaAs/InP SPADs. Even though InGaAs/InP SPADs have been available for the last ten years. noise and afterpulsing are still strong limitations that impair their wide-spread exploitation. Hence, a primary objective of my Ph.D. research was the development of low noise InGaAs/InP SPAD detectors.

With respect to previous InGaAs/ InP SPAD generations developed at Politecnico di Milano, critical heterostructure design and fabrication parameters have been identified and optimized. This has resulted in significant performance improvement. The new 25 µm active-area diameter InGaAs/InP SPADs, when cooled at 225 K and biased at 5 V excess bias, exhibit a dark count rate (DCR) of just few kilo-counts per second, photon detection efficiency (PDE) higher than 25% at 1550 nm (with good uniformity across the active area), timing jitter of approximately 100 ps (FWHM) and moderately low afterpulsing. These devices are currently mounted in commercially available singlephoton counting modules. In addition to discrete InGaAs/ InP SPADs, I conceived a novel mixed-quenching approach based on an integrated quenching resistor, which promptly limits the avalanche current buildup, and a fast external active circuit that completes the quenching and enforces a welldefined hold-off. This solution alleviates the trade-off between afterpulsing reduction and recovery time, which is typically present in similar detectors, such as negative-feedback avalanche diodes (NFADs), that rely only on integrated passive quenching. Furthermore, a novel implementation of the integrated quenching resistor that reduces fabrication costs by exploiting already available fabrication process steps has been proposed and fabricated. Finally, I designed a compact singlephoton counting module based on this detector. Experimental characterization confirmed the excellent performance of the system. Electroluminescence measurements demonstrated fainter and shorter (1 ns) avalanche current waveforms compared to other state-of-the-art quenching solutions, an indication of significant reduction of the avalanche charge. Accordingly, the measured afterpulsing



1. Scanning electron microscope image (SEM) of the InGaAs/InP SPAD cross section, with layers structure and Zn diffusion profile highlighted



probability is also strongly reduced, permitting gated mode operation with hold-off times in the µs range without the need for sub-ns gate-on windows. Finally, the current-mode readout frontend circuit guarantees very low timing jitter (70 ps FWHM). These characteristics open the way to future widespread use of these modules.

In the 0.75 µm – 1 µm NIR sub-range, Silicon SPADs can be employed. Silicon SPADs fabricated with custom epitaxial technologies provide best-in-class performance. However, custom technologies are expensive and, owing to the inherent limitations to system integration, best

2. Photograph of a 4x4 BCD SPAD array

suited for small (up to about a hundred pixels) SPAD arrays. As a consequence, the increasing demand of large SPAD arrays in many applications, ranging from fluorescence lifetime imaging (FLIM) to LIDAR and medical imaging, has driven the research toward the fabrication of SPADs in conventional CMOS processes, with the aim of producing complex systems-on-chip at low cost and with high reproducibility. The second research topic of this Ph.D. work was to design, in collaboration with STMicroelectronics, the first SPADs fabricated in a BCD (Bipolar, CMOS, DMOS) technology. With the aid of process TCAD simulations, calibrated on the STMicroelectronics technology process, I designed dedicated process steps to be added to a basic BCD process flow chart for integrating very dense arrays

of SPADs with high detection performance. The experimental characterization shows that, at 5 V excess bias, the room temperature DCR per unit area (< 0.2 cps/µm2) is comparable to those of the best-in-class CMOS SPADs. At the same operating conditions, the PDE (> 40% at 450 nm and > 5% at 800 nm) matches that of the verv best CMOS SPADs presented in literature. The timing response, featuring a very narrow Gaussian peak (~ 30 ps FWHM) followed by a fast exponential tail with less than 100 ps time constant, is definitely among the best presented in literature for either CMOS or custom SPADs. The good overall detection performance arises from the high cleanness and reliability of the STMicroelectronics foundry, and the precise tailoring of the SPAD electric field. In addition, I designed a novel device geometry that overcomes the intrinsic PDE limitation of typical CMOS SPADs, i.e. the avalanche is mainly initiated by holes, which have a lower avalanche triggering probability compared to electrons in Silicon. The achieved PDE (60 % at 500 nm and 15% at 800 nm) demonstrates the effectiveness of

the proposed approach.

OVERCOMING PERFORMANCE, COST AND ENERGY ISSUES IN FIXED AND MOBILE CONVERGENT NETWORKS

Savi Marco - Advisor: Prof. Giacomo Verticale

The thesis work aims at investigating novel technologies, i.e., Fixed and Mobile Convergence (FMC), Distributed Content Delivery and Network Function Virtualization (NFV) aimed at solving the performance, cost and energy issues arising in metro and access networks due to the Internet traffic explosion.

The work is inserted in the framework of the COMBO (Converged of fixed and Mobile BrOadband access/aggregation network) Project, funded by the European Community Seventh Framework Programme FP7/2013-2015, under grant agreement no. 317762. Specifically, our work is related to Working Package 3 (WP3). WP3 deals with the definition of new FMC metro/access networks both in terms of data plane and control plane. FMC metro/access networks can be jointly designed and optimized both from a functional convergence perspective (i.e., by unifying network functionalities among heterogeneous network types) and a structural convergence perspective (i.e., by sharing equipment and infrastructures among heterogeneous network types).

In a first part of the work we focused on the performance evaluation of video/content replication in different metro/ access network architectures. We

considered three different metro/ access network architectures for Fixed and Mobile Convergence. One architecture is based on a metro ring (Metro Ring Architecture), while two architectures are based on a tree topology. The first of such two tree-based architectures. called Active Architecture, includes active nodes (i.e., switches) in all the stages of the metro/access network and can be improved by adding mesh links in strategic positions of the tree topology. The second architecture, called Passive Architecture, is a Long-Reach Passive Optical Network (LR-PON), where nodes in the intermediate stages of the metro/access network are passive devices (i.e., splitters). In these metro/access network architectures a Video-on-Demand (VoD) content distribution system can be deployed. One or more content servers, called Metro Servers (MSs), can be placed in the network to enable core traffic offloading and to reduce latency for users VoD requests, achieving a better Quality of Experience (QoE). We evaluated both using simulative and analytical methodologies how content providers could benefit from the deployment of Metro Servers in terms of blocking probability of the VoD requests. Results show that the placement of MSs in different positions of such network architectures allows

network operators to significantly decrease the blocking probability of VoD requests. Moreover, the performance of the Active Architecture can be consistently improved by adding mesh links in strategic positions. The mesh link addition in the Active Architecture allows to gain better performance than the Passive Architecture and the Metro Ring Architecture. On the contrary, the Passive Architecture is the most beneficial from a cost perspective, but it leads to the worst performance. We also explored the role of network capacity and content replicability constraints for the MSs: in our condition of popularity, less than 10% of the contents should be replicated in the MSs in order to reach the minimum blocking probability. Moreover, savings can be achieved also in the network capacity of the MSs: network operators can dimension the minimum capacity to obtain the best performance without wasting network resources. In a second part of our work we focused on the definition of energyefficient strategies to reduce the overall power consumption when such Metro Servers are deployed in the metro/access network. First of all, we mainly focused our attention on a dynamic scenario, in which the VoD traffic load varies over the day. In such evaluation,

we decided to consider the aforementioned Active Architecture as reference network architecture. by defining a power consumption model for MSs and switches. We then designed and evaluated the performance of an algorithm, the Best Energy-Aware Pattern Algorithm (BEPA), which is able to set up a strategy for switching on/off network interfaces and MSs depending on traffic load conditions, by exploiting the tradeoff between the transport energy consumption (due to the switches) and caching energy consumption (due to the MSs). Results show that the BEPA algorithm always outperforms in terms of energy consumption other simpler strategies, i.e., keeping the MSs always on or always off. We then dealt with a static scenario aimed at planning the deployment of Metro Servers or, more in general, of content caches in the network. By using the power consumption models shown in and by making realistic assumptions about the content popularity for fixed and mobile users, we defined an optimization problem based on ILP (Integer Linear Programming) for an energy-efficient placement of caches in core, metro and fixed/mobile access nodes of the network. Such optimization problem can be used by telecom operators to intelligently plan

the content distribution in the network and turning on only selected caches and needed network interfaces. Results show that that the placement strategy provided by our optimization problem always outperforms other simple strategies, such as turning on or turning off all the caches. Moreover, they show that an evolution towards a FMC metro/ access network, where fixed and mobile users can share caches, can reduce the energy consumed for VoD content delivery. In the last part of our work, we investigated how the Network Functions Virtualization (NFV) paradigm can be exploited to ease service deployment in FMC metro/ access networks. First of all, we investigated the impact of some processingresource sharing costs that occur when multiple Virtual Network Functions (VNFs) are hosted in the same NFV generic-purpose hardware. We identified two different types of costs: the context switching costs and the upscaling costs. Context switching costs occur when multiple VNFs are hosted by the same hardware and share the same CPU cores, while upscaling costs occur when a VNF must process such a high quantity of traffic to require a multi-core implementation and, thus, traffic must be balanced

PhD Yearbook | 2017

409

the size and the number of VNFs hosted by the hardware. We first designed an ILP model to deal with Service Chain (SC, i.e., chain of multiple VNFs to provide a specific service) placement to minimize the NFV generic-purpose hardware active in the network. We then developed a heuristic algorithm to deal with scalability issues of the aforementioned ILP model, whose performance has been validated. In general, results show that context switching costs have a greater impact on VNF consolidation than upscaling costs. Besides, such costs strongly affect NFV consolidation when SCs with a very strict latency requirement are embedded in the network.

among different CPU cores. Such

costs lead to a trade-off between

Finally, we deepened our investigation on the impact of latency in the centralization/ distribution of VNFs in the metro/ access networks. Results show that the adoption of a FMC metro/ access network architecture allows a better consolidation of VNFs than a non-convergent architecture. Moreover, they also show that SCs with very strict latency requirements require a massive distribution of VNFs in the metro/ access network, while SCs with a loose latency requirement can consolidate all the chained VNFs in datacenters far from the users.

A NEWTON-EULER APPROACH TO MODELLING AND CONTROL OF FLEXIBLE MANIPULATORS

Scaglioni Bruno - Supervisor: Prof. Gianni Ferretti

Modelling and control of flexible lightweight manipulators has been matter of research since the late 80's. Nevertheless, the interest on this topic is still intense. Manipulators where structural flexibility deserve to be considered in the design stage can be found in a range of applications including traditional industrial robots, nuclear materials manipulation, agricultural and space robotics as well as in autonomous vehicle applications. The introduction of lightweight robots is a driver of innovation when trending technologies require decreasing power consumption and low manipulator masses in view of increasing performances requirements. The aforementioned trends encourage the use of lightweight manipulators and the development of related technologies, but whenever high performances are required, a severe issue arises: vibrations. The main subject of this thesis is the development of a closedform model of three-dimensional flexible manipulator with links of general shape and the synthesis of a two-timescale control system able to improve the performances of the manipulator and substantially increase damping with respect to the classical P/PI control structure. The resulting performance are achieved by means of a controller acting on the robot joints in a

"fast" timescale, coupled with a traditional system acting on the "slow" timescale.

A compact model of a generic three-dimensional flexible body was initially developed in an objectoriented modelling framework, namely Modelica, and based on the floating frame of reference (FFR) formulation. The equations of motion have been formulated in the Newton-Euler form, allowing a complete calculation of the Coriolis and gyroscopic terms of the inertial forces for the elastic degrees of freedom. The model allows considering bodies of general shape through the calculation of shape functions, which describe the geometry, as the result of the modal analysis performed by finite elements packages. An example of link is shown in figure 1. The results achieved in the development of the generic body model have then been exploited in the development of a closed-form model of flexible manipulator.

The proposed model is based on the adoption of the spatial vector notation, which allows to combine the equations of motion of the links calculated in the previous step, leading to a multilink threedimensional model in closed form with respect to the joint angles and the modal coordinates. The motion equations can be computed for links with arbitrary shape and cardinality; hence, the model results to be highly adaptable and computationally efficient. The model has been validated by means of comparison with literature benchmarks obtained with the classical multibody approach and with experiments collected on a real manipulator. Despite its efficiency and accuracy, the closed form model is not suitable for real-time control and active vibration damping due to its inherent complexity, hence an approximated model based on the integral manifold approach has been developed starting from



1. Manipulator's link of general shape



2. Effectiveness of the damping system

the complete model previously described. The approximated model has therefore been compared with the original closedform model and with experimental data retrieved from the MERIt dataset. The subsystems resulting from the adoption of the integral manifold approximation are in fact two: a "slow" nonlinear subsystem, which can be controlled by means of classical techniques, and a "fast" linear subsystem, whose dynamic matrices exhibit a dependency on the state of the "slow" subsystem, namely the spatial configuration of the robot.

A control technique is finally

"fast" subsystems, which is based on advanced nonconvex H-infinity control. The overall control system provides better performances with respect of the classical techniques, allowing a consistent vibration damping and a substantial increase in the control bandwidth, as shown by **figure 2** where the vibrations of the manipulator are compared with respect to the standard control system. Future developments will concern controllability and observability analysis on the vibration system

and a state estimator for the

"fast" subsystem, based on

presented for the family of the

considerations related to the use of sensors, such as the strain gauges position, in conjunction with the analysis of the shape functions relating the strain to the modal variables.

CASCADE - A FRAMEWORK FOR MODELING FLUVIAL SEDIMENT CONNECTIVITY AND ITS APPLICATION FOR DESIGNING LOW IMPACT HYDROPOWER PORTFOLIOS

Schmitt Rafael Jan Pablo - Supervisor: Prof. Dr. Andrea Castelletti

Human activities, and especially dam constructions, have altered water and sediment fluxes in river basins with an unprecedented rate over the last decades. The resulting changes in river and delta processes pose a risk to river, floodplain, and coastal ecosystems and human livelihoods. While fluvial sediment transfers act on networkscales, impact assessments of dam sediment trapping (regarding downstream impacts and storage loss in reservoirs) and economic benefits focus commonly on single dam sites. Such local approaches omit the connected functioning of fluvial sediment transfers and the resulting network-scale trade-offs between dam economic benefits and cumulative sediment trapping of dams. Such analyses would require sediment data and models to evaluate future impacts of dams on network-scale sediment transport. However, neither sediment data nor numerical sediment transport models are commonly available on network scales.

Within this context, the thesis sets out to answer the research questions of what would be optimal trade-offs between dam sediment trapping and economic hydropower benefits in large river systems? To answer this research question, the objectives of this thesis are 1) developing a network-scale sediment transport and connectivity model, 2) test

its application for various, data scarce river basins with various initialization strategies and evaluate the robustness of results, and, 3) apply the model for quantifying cumulative dam sediment trapping and network-scale trade-offs between sediment trapping and hydroelectric production. First, we introduce a numerical modeling framework for connected sediment transfers on the scale of large river networks. The CASCADE (CAtchment Sediment Connectivity and DElivery) framework is a parsimonious statistical framework coupling network scale, grainsize specific sediment budgets with recent concepts of sediment connectivity. Analyzing sediment connectivity in two major Asian river basins (Da River, 50570 km2 and Se Kong, Se San, Sre Pok (3S) Rivers, 82500 km2) guantified how sediment connectivity is a multi-scale, multi-domain property of river systems driven by the spatial distribution and properties of sediment supply and fluvial transport processes. Such information was so far available in qualitative terms, mostly, and only for smaller, well monitored rivers. CASCADE can be parameterized deterministically based on remotely sensed data using hydro-morphologic equilibrium considerations for large and datascarce river systems. However, we

also show how CASCADE allows to implement network-scale stochastic modeling for the 3S basin. Such stochastic modeling allows to disaggregate point observations of sediment flux and grain size into spatially distributed estimates of grain sizes and supply from many sediment sources to the river network via an inverse Monte Carlo Approach. This source information is used to model sediment flux and grain size composition in the entire river network.

One key result is that sediment transport is spatially highly heterogeneous. This heterogeneity, which is derived from the CASCADE framework and validated with other lines of evidence, is observed for both case studies. This poses the opportunity to minimize dam sediment trapping by placing dams in parts of the networks where sediment transport is naturally low compared to the hydro-power potential. To analyse impacts of hydropower developments, we add a simplified model of reservoir hydraulics to CASCADE and analyze actual and hypothetical hydropower development strategies for the 3S rivers. Developing all hydro-power dams will reduce total sediment flux from the 3S by more than 90 %. Nearly 60 % of this reduction are attributable to a single dam site. We then introduce a network scale analysis



1. Visualizing sediment connectivity for the Se Kong, Se San, Sre Pok (3S) River basin, a major tributary system of the Mekong River. Sediment connectivity was modelled for that basin applying the CASCADE modeling framework using an inverse approach for model initialization. Each curved line represents the transport of sediment from one of many sediment sources through the river network to the basin outlet. Colors from blue to yellow indicate different, increasing, sediment fluxes from each source. The 3S basin is currently subject to major hydroelectric dam developments, likely to trap the majority of sediment changing the pattern and magnitude of sediment connectivity in the 3S basin, sediment delivery to the lower Mekong River, the Tonle Sap lake, and the Mekong Delta. This will impact ecosystems and livelihoods alike.

of dam portfolios (i.e., different combinations of the proposed dam sites). The aim of this analysis is to identify optimal trade-offs between dam hydro-power production, production costs, and sediment trapping. We find that, because of the spatial heterogeneity in sediment transfers, very similar hydropower production levels can be reached for very different levels of sediment trapping in function of how dams are placed within the river network. In the basin under study, up to 70 % of the hydro-power potential could have been developed with a minor (20 %) reduction in sediment flux. Additionally, using an empirical model for hydro-power production costs, we find that there is a strong synergy between providing cheap hydro-power and reducing dam sediment trapping on network scales. To conclude, the CASCADE approach,

which was developed for this thesis, enables analyzing and quantifying connected sediment transfers from a whole-network perspective. We provide evidence for how this novel information can be used to identify optimal dam portfolios that minimize trade-offs between hydropower production and sediment trapping in dams. Such a strategic planning of hydropower portfolios is advantageous from an environmental and economic perspective, and poses major potential for reducing impacts of future hydro-power developments on large river systems.

CONSENSUS METHODS FOR DISTRIBUTED INFERENCE IN DENSE COOPERATIVE WIRELESS NETWORKS

Soatti Gloria - Supervisor: Prof. Monica Barbara Nicoli

With the development of largescale networks, centralized algorithms revealed to be unfeasible for in-network data processing. In fact, their performance is penalized by communication/computational overhead and latency required for data aggregation/processing at the fusion center, that badly scale with the network size. In recent years, distributed algorithms have become more attractive, as they enable the network to solve complex inference problems without the support of any central unit, facilitating energy-efficient, high-throughput and low-latency network communications. The aim of distributed estimation problem considered in this thesis is to infer some application-related parameters through consensusbased methods. An example is shown in Fig. 1(a), where nodes, deployed over the space and interconnected by bidirectional links, have to estimate some unknown parameters **\theta** n a fully distributed way. The estimate is $\boldsymbol{\theta}$ computed by successive refinements of the local estimate at each node, θ , based on information exchanges, till a consensus is reached within the network, i.e., $\theta^{\uparrow} \rightarrow \theta^{\uparrow}$ as shown in Fig. 1(b). In this thesis, we provide new

consensus-based methodologies for in-network processing able to reduce the inter-node signalling and guarantee close convergence to the centralized solution even in critical cases. e.g., non-linear or underdetermined estimation settings. These algorithms are designed to improve the performance of conventional consensus approaches by accounting for the different reliabilities of node estimates. Fundamental limits and convergence properties are analytically evaluated for all methods.

The proposed techniques are used as basis to develop distributed algorithms for self-organization of dense cooperative wireless networks, in different application scenarios. Specifically, distributed inference is considered for the estimation of fundamental parameters needed for node cooperation, such as channelstate information, node locations, interference sensing, resource scheduling, and precise vehicle positioning. A first application is the distributed estimation of multilink channel-state information that is instrumental for network localization purposes, as addressed later. In static indoor scenarios. the channel response between any two nodes varies rapidly from link to link. However, some channel parameters, , representative of site-specific static/dynamic multipath fading and slowly varying in space, are common to all links of the network and can benefit from distributed estimation. A new stochastic channel model is developed to describe the relationship between the received signal strength (RSS) statistics and node locations. Particularly, the analytical study, validated on real data, proves that both the first



two moments of the RSS log-Rician distribution can be modelled as jointly Gaussian with average linearly related to the link logdistance. Thus, the statistics of the RSS moments, shared by all links, can be considered as site-specific parameters **0** to be self-learned by the nodes through distributed algorithms. The proposed methods are shown to significantly improve the performance of conventional solutions.

The non-linear relationship between the RSS statistics and the node locations is then exploited for network localization, where each node has to cooperatively infer the positions $\boldsymbol{\theta}$ of all network nodes from local measurements of the joint RSS moments, regardless of link availability. The analysis of the new methods is thus extended to non-linear scenarios, where the non-linearity of the measurement model is handled by performing linearization at each consensus step. Moreover, an experimental assessment of the proposed localization algorithm is carried out based on indoor field measurements.

Fundamental for self-organization of the network is a cooperative sensing of the shared spectrum and related allocation of the timefrequency (TF) network resources. We consider a simplified setting with two coexisting networks, where the nodes of a secondary network, acting as interconnected sensing devices, have to detect the complete interference pattern $\boldsymbol{\theta}$ of a primary network in a distributed way. Cooperation among secondary users (SUs) is crucial, as each SU may have partial visibility due to limited sensing range or shadowing/fading problems. The consensus strategy is here nested within a decision-directed procedure for Bayesian detection of the spectrum occupancy. The distributed approach provides the estimate of the complete interference pattern to each SU. regardless of local incomplete visibility, reaching the centralized solution. Moreover, an ad-hoc measurement campaign has been carried out to highlight a practical case study of unlicensed 2.4GHz spectrum sharing. The proposed method has been developed under the DIWINE Eu project. Distributed resource management based on the cooperatively sensed interference level is then discussed. A new distributed scheduling algorithm is designed where multiple node-to-node links, sharing the same spectrum region, have to coordinate to selfadapt their TF allocation so as to guarantee the quality of service (QoS). In particular, each node autonomously inflates/deflates its allocation – based on the sensed

PhD Yearbook I 2017

415

need/release of TF resources to the neighbors. Nodes optimize the TF allocation by iterated local adjustments, till the QoS requirements are satisfied. As final contribution, the analysis of distributed inference is extended to dynamic scenarios with moving nodes. A self-organizing vehicular network is considered where vehicles can improve their Global Positioning System (GPS) location accuracies by sharing information through vehicle-to-vehicle (V2V) communication links. An innovative cooperative positioning framework is proposed where, in contrast to conventional methods based on V2V ranging, a set of passive features are used as common noisy reference points that are cooperatively localized by the vehicles and implicitly used to augment the vehicle location accuracy. A distributed message passing (MP) algorithm is designed to solve the positioning problem, integrating a Kalman filter to track vehicle dynamics. However, as features are non-cooperative objects, the consensus method is combined with MP to enable cooperation between vehicles for the distributed inference of the features' positions θ , which implicitly improve the GPS-based vehicle location accuracy.

interference level and the OoS

requirements – by signalling the

Teimourikia Mahsa - Supervisor: Prof. Mariagrazia Fugini

Safety and security are two riskdriven aspects that are usually tackled separately. The importance of considering safety and security as inter-dependent aspects is highlighted in the recent literature. Co-engineering safety and security together as cyber-security has a direct or indirect effect on critical systems. This need is more into evidence with the advent of new technologies including the *Internet* of Things (IoT) and Smart Work *Environments (SWEs)*, where it is essential to re-define the safety and security concepts. Internet of Things (IoT) is not a myth anymore and it is already at the pick of the hype cycle and is being increasingly adopted in different application areas including smart spaces. One of the areas where IoT technology is being implemented is the "Industry *4.0*" which we refer to as *Smart* Work Environments (SWEs). As with all new technologies, SWEs introduce various issues and opportunities. Researchers have already started the struggle to tackle the unresolved issues while taking advantage of vast variety of new opportunities that emerge in this field.

In the traditional risk management techniques adopted in manufacturing and industrial environments physical controls are in place to avoid risks related

to safety. However, in SWEs these controls are not enough. The need for co-engineering safety and security arises in SWEs since, safety hazards might lead to security threats and vice versa. Also, security mechanisms should be flexible enough to facilitate risk treatment. Moreover, security mechanisms may also be used in collaboration with safety management systems to protect the safety of the workers in addition to protecting the access to sensitive physical resources in the SWEs.

On the one hand, as more devices and tools are getting integrated in the IoT technology, the organizations adopting the Industry 4.0 become more and more vulnerable to the security threats. And hence new approaches should be proposed to protect the sensitive and critical resources in the SWEs. On the other hand, the IoT technology provides the chance to acquire ambient and monitoring data to be exploited for identifying and treating risks related to safety. While this is an advantage to protect persons' safety, the security rules should allow the treatment of risks, when necessary, by adapting to the safety-related contexts. In this thesis, the safety and security of risk-prone SWEs are

tackled. Starting with safety, the SWE is studied to explore different components that should be used in an efficient risk management procedure. A run-time risk management methodology is proposed that exploits an automated risk assessment process that is developed considering the commonly adopted risk assessment techniques in the industry. To extract the safety knowledge in a computer-readable way, an ontology is designed and developed. Coming to security, a risk-adaptive Access Control (AC) model based on Attribute-Based Access Control (ABAC) is developed considering hierarchical safetyrelated contexts. Upon receiving risk descriptions, based on the designed meta-rules, the AC system adapts the security rules to allow risk treatment. To manage dynamically adapted security rules, a conflict analysis algorithm is proposed to find conflicting rules that might lead to unpredictable results at design time.

MODELLING AND VERIFICATION OF EVOLVING CYBER-PHYSICAL SPACES

Tsigkanos Christos - Advisor: Prof. Carlo Ghezzi

Computing and communication capabilities are increasingly being embedded into physical spaces thus blurring the boundary between computational and physical worlds; typically, this is the case in modern cyber-physical systems, like smart buildings or smart cities. Conceptually, such composite environments, hereafter called cyber-physical spaces (CPSp), can be abstracted into a topological model where computational and physical entities are connected in a discrete, graph-like structure. Like any other software-intensive system, a CPSp is highly dynamic and typically undergoes continuous change, a fact we emphasize by the notion of evolving CPSp. This brings a manifold of challenges as dynamics may affect safety, security, or reliability requirements of the overall space-dependent system. Requirements may predicate about relational structures spanning both spaces, while dynamics of the CPSp may also imply changes that span both cyber and physical constituents. Thus, a first challenge is one of modelling; a way to build precise representations which will serve as descriptions of the space and a specification of requirements, using mathematical concepts and language. Assuming an evolving environment and certain requirements expressed

in terms of structures in the CPSp, to check if the system satisfies the requirements, the system in terms of its evolution must be considered. Thus a second challenge is one of analysis of evolving cyber-physical spaces. Analysis refers to formally verifying if requirements in the form of a logic-based specification are satisfied. However, analysis outcomes may include cases where requirements are violated, perhaps in future CPSp evolutions. In this case, certain counteraction logic must be programmed in the system, able to handle changes in the CPSp environment that may lead to violations. This refers to self-adaptation, the ability of the system to monitor and analyze its environment, and plan accordingly counteractions that it can enact in an an autonomous fashion, to prevent, mitigate or circumvent requirements violations. Formally modelling space and its dynamics as well as supporting reasoning about various properties of evolving space, is a crucial prerequisite for engineering dependable evolving CPSp, e.g. to assure requirements satisfaction or to trigger correct adaptation. Thus, we proposed a methodology and technical framework which support modelling of evolving cyberphysical spaces and reasoning about their spatio-temporal properties. We utilised bigraphs

as a formalism for modelling CPSp as well as primitives of change, giving rise to a bigraphical reactive system (BRS) consisting of rewriting rules with both local and global application conditions. Formal reasoning facilities feature logic-based specification of properties and according model checking procedures, in both spatial and temporal fragments. In particular, regarding theoretical foundations, we extended the bigraphical rewriting approach to be equipped with additional non-local application conditions. We integrated several fundamental techniques to support reasoning about complex spatio-temporal properties of a BRS-based model of evolving space. Locally bounded spatial properties were expressed as bigraphical patterns, and bigraph matching was used as a fundamental technique to locate the points in space where such formulae hold. Concerning checking of global spatial properties, we interpreted a bigraphical model as a closure space, paving the way for adopting a spatial logic for closure spaces along with corresponding model checking procedures. Similarly, concerning checking of temporal properties, a variety of state-transition models were obtained from a BRS, serving as the underlying evaluation model for model checking with a temporal

logic. To facilitate adoption, we proposed an abstract syntax for bigraphs and presented a canonical mapping which maps bigraphs to typed graphs in a natural way.

Utilizing these foundations, we subsequently considered systematically engineering applications that support both devising adaptive behaviours at run time as well as formal verification of requirements of CPSp in early stages of design. In the context of adaptive security systems that aim to protect critical assets in the face of changes in their operational environment, incorporating an explicit representation of this environment's topology at run time enabled reasoning on unexpected threats that can

arise from potentially malicious agents, through adaptive security policies that prevent, circumvent, or mitigate possible security requirements violations. Regarding verification of requirements of CPSp in early stages of design, we considered applications in the construction industry, where building information models are the de facto standard for specifying complex information about building infrastructures, a representation which we proposed to be also extended for specification of cyber-physical entities. We proposed formal static and dynamic semantics of such descriptions making it possible to support many forms of advanced analyses as typically performed in software engineering. Static

semantics aim to broadly support reasoning about latent qualities of a design. Dynamic semantics aim to deal with the dynamism that a space exhibits when additionally considering the ways it may change along with entities inhabiting it. We investigated applicability and realizability using a disaster scenario involving operation of autonomous unmanned aerial vehicles in a smart city environment as a case study, utilising complex relations in the space along with dynamics specifications augmented with spatial application conditions, and evaluated verification procedures for spatio-temporal properties in experiments using city environments of varying model sizes, sourced from domain models.

DOMAIN-SPECIFIC MODELING AND CODE GENERATION FOR CROSS-PLATFORM MOBILE AND IOT-BASED APPLICATIONS.

be available at any time, in any

place, for everyone and for any

device. So far the development

of front-end and user interfaces

for IoT systems has not played a

relevant role in research. On the

contrary, user interfaces must be

a key part in the IoT ecosystem

Umuhoza Eric - Supervisor: Prof. Marco Brambilla

Nowadays, mobile devices constitute the most common computing device. This new computing model has brought intense competition among hardware and software providers who are continuously introducing increasingly powerful mobile devices and innovative OSs into the market. In consequence, cross-platform and multi-device development has become a priority for software companies that want to reach the widest possible audience. However, developing an application for several platforms implies high costs and technical complexity. Currently, there are several frameworks implementing different methodologies for cross platform application development (Web, Hybrid, Interpreted and Cross Compiled): examples include PhoneGap (Cordova), Appcelerator Titanium, and Xamarin. Nevertheless, these approaches still require manual programming, which yields to high risks of errors, inconsistencies, and inefficiencies. On the other side, with the advent of the Internet of Things era, homes, cities and almost everything is becoming smart.

This technology is extending

very rapidly, but it still needs to

solve many problems that arise

when a technology requires to

because they can play a key role in acceptance of solutions by final adopters. The intelligent things connected together can cooperate and exchange information, but their ultimate goal is to provide value to people. Such value can be perceived only through appropriate user interfaces, which visualize information. let user navigate the information, and interact with the sensors and devices, by setting properties, collecting data and regulating their behavior. My research proposes to face the challenges of mobile and IoTbased applications development by exploiting abstraction, modeling and code generation in the spirit of model-driven development engineering. My research covered all development phases, from design to execution, of mobile and IoT-based applications by applying modeldriven development (MDD)

> approach. When following MDD approach, the application is expressed through models from which the final code is (semi)

automatically obtained. The main contributions of this work include: (1) At design time: modeling languages and design methodology for mobile and IoT-based applications. defined a platform independent modeling language for mobile applications, named Mobile IFML. The Mobile IFML has been designed extending a OMG standard for user interactions modeling, Interaction Flow Modeling Language. The mobile language has been extended to accommodate the IoT-specific requirements. I defined a set of design patterns allowing the modeling of recurrent problems in user interaction modeling for both mobile and IoT-based applications.

(2) Implementation: This research explored different approaches to generate the code for mobile and IoT-based applications from the models describing those applications. I conducted a comparative study to identify the best trade-off between different code generation strategies. (3) At Execution Time: User Behavior Analysis framework. To enhance the understanding of how the users consume the provided content I proposed a model-driven engineering approach that blends designtime information and runtime execution data of the application for generating insightful analyses of user behavior. Thanks to the proposed information fusion. new kinds of analysis can be performed, which highlight the role and meaning of visualized data in the application. New analytics results can be generated based on the displayed objects, their categorization and their properties. Results of the analysis can be immediately displayed directly on the visual models of the applications, thus making it immediate for a designer to spot possible problems or advantages of a given user interaction design choice.

In parallel to the main topic of this thesis, I investigated on the supposed barriers to the adoption of model-based development approaches. In that context the issues of modeling effort (many software practitioners consider

model-based development approaches time-consuming, hence prefer to avoid using models which are believed as excessive, superfluous and unnecessary artifacts) and modeling language complexity and usability have been tackled. I conducted a research with the aim of revealing whether it is actually modeling or designing that dominates the effort in the creative process of modeling software systems. The experiments showed that more than 60% of the effort is spent in designing. This means that the fault of supposedly unproductive processes should not be blamed on modeling, but to the (anyhow necessary) time devoted to thinking about the problem and identifying the solution. Regarding modeling language usability, I conducted a research to understand how the designed languages are actually used and

how they fit the users' need. This investigation showed that available languages are either too complex (e.g.: BPMN) with respect to user need or do not exactly fit the domain (e.g.: UML) and this happens mainly because of the lack of right involvement of endusers in the language development process. I proposed a usercentered approach allowing the adaptation of existing modeling languages to the user needs through a language simplification process. The proposed simplification process relies on a set of steps that encompass the selection of the language elements to simplify, generation of a set of language variants for them, measurement of effectiveness of the variants through modeling sessions performed by end-users, and extraction of quantitative and qualitative data for guiding the selection of the best language refinement.

GRAPHENE ANTENNA DESIGN FOR TERAHERTZ SENSING AND COMMUNICATIONS

Verri Valentina - Supervisor: Prof. Gian Guido Gentili

The rising demand of higher bandwidth for the next generation wireless communication systems produces the extension of the operating frequency of the systems to the millimeter/ sub-millimeter waves of the electromagnetic spectrum. In this context, new solutions such as advanced physical layer solutions and new spectral bands are developed to support these high data rates and the Terahertz band communication is considered as a key of wireless technology. Terahertz frequency (THz) band is defined as a portion of the electromagnetic spectrum, which extends from 0.1 THz to 10 THz and occupies a large regime of the electromagnetic spectrum between the infrared and microwave bands. This band, known as a "Terahertz band gap", has remained untouched for a long time by the scientist and researchers, because of the unavailability of powerful sources and detectors. In the last few years, the semi-conductor technology has grown quickly and its effect on the research in this band gap has been considerable. Infact the Terahertz spectrum finds its various applications in the medical science, imaging of hidden items, time domain spectroscopy, defense applications, earth and space science, semiconductor

wafer inspection and air pollution checking. Its unique radiation characteristics, such as the penetration through different materials, the high resolution and non-ionization, allow to the THz band to get a such broad application area. In this context, new transceiver architectures that are able to operate at THz band frequencies and to exploit the very large available bandwidth are required. Furthermore these transceivers show high power, high sensitivity and low noise figure in order to overcome the very high pathloss at THz frequencies. Besides these devices, ultra-broadband and multi-band antennas are necessary to enable multi-Gbps and Tbps links in the THz band. Moreover, in order to overcome the very high path loss of the THz band channel, new advanced antenna systems, such as very large antenna arrays, are developed. To satisfy these requirements,

different technologies, such as CMOS technology, Gallium nitride, indium phosphide, metamorphic and graphene technologies are considered. One of the most recent alternatives to develop THz band compact transceivers is based on graphene: the use of graphene for antennas and other electromagnetic devices

brings significant benefit such as extreme miniaturization, dynamic tuning, and even transparency and mechanical flexibility. This new material has recently attracted huge interest in various research areas due to its unique electrical, mechanical, thermal, chemical, electronic and optical properties. Graphene, isolated for the first time in 2004 by A. K. Geim and K. S. Novoselov from the University of Manchester, is a novel twodimensional material that consists of carbon atoms arranged into an honeycomb lattice. This material is considered as a natural High Impedance Surface (HIS) due to its high resistivity in an ultra-wide band and can easily reach the THz band. The propagation features above the graphene surface are defined by its conductivity that depends by other parameters such as the selected substrate and the applied voltage. Graphene also presents fascinating properties including high mobility, extraordinary high thermal conductivity and strength that allow graphene to be a good choice for the THz applications. Graphene offers a new approach to the THz communications thanks to its ability to support the propagation of Surface Plasmon Polariton (SPP) waves in the Terahertz frequency band: these can be manipulated by changing

the bias voltage applied to the graphene layers since it acts directly on the surface conductivity of the material. Thanks to its monoatomic thickness. graphene can be modeled as an infinitely thin surface of complex conductivity whose behaviour depends by the electro-chemical potential. The behaviour of the complex surface conductivity is controlled by the electro-chemical potential µc changing the applied voltage corresponds in changing of the electro-chemical potential. Using this feature it is possible to realize reconfigurable devices where the unique tunable parameter is represented by the electro-chemical potential. Development of the next generation wireless communication network requires systems with broadband capabilities in high mobility environment. Integrated systems with dimensions of a few micrometers can not be achieved by only reducing the dimensions of the classical metallic antennas because these structures present low mobility of electrons and high resonance frequency. For these reasons, thanks to its ability to support SPP, graphene is considered as a promising technology for this rising field. A graphene-based nano-patch antenna with size of a few micrometers resonates in the THz band, lower frequency respect to metallic antenna of the same size and with better performances. Patch antennas are becoming increasingly useful because they can be printed on the circuit board: they are low cost, low profile antennas and they are

easily fabricated. In these studies graphene patch antennas are analyzed: starting from a single layer dipole design, the double layer is considered in order to improve the performances. The double layer design presents better performances, infact with the same antenna structure and size it is possible to reach higher frequencies and good performances. The effect of the applied voltage on the mutual coupling between two graphene patch antennas is considered: it affects the radiation pattern and the input impedance. A trade off among different factors, such as the mutual coupling, gain and resonance frequency of antenna is necessary to match the targets. Some applications with Frequency Selective Surface (FSS) are studied: it shows different reflection and/ or transmission properties as function of frequency. A tunable graphene Jerusalem Cross (JC) FSS is used to find different reflection and/or transmission properties as function of the frequency only by changing the electro-chemical potential. A beam reconfigurable antenna for THz applications based on a switchable graphene FSS is studied: the radiation pattern changes with the presence of the graphene FSS and the beam direction changes depending on the voltage applied to the graphene FSS structure reaching a good gain. A graphene Split Ring Resonator (SRR) implemented in an antenna structure is studied: a degradation in the reflection coefficient is found when the SRR array is added to the patch antenna, but the bandwidth of

the antenna and the gain increase depending on the applied voltage. In order to increase the radiation properties of the antenna structure, the lens is used. In this structure it is possible to control both the operating frequency and the bandwidth and by using an extended lens it is possible to improve the antenna gain and directivity.

Another application studied is the phase shifter: it is an important component of the microwave, mm-wave and THz systems and it becomes especially important when implemented in phased array antenna. The phase shifter consists of an High Impedance Surface HIS placed on the side on the Dielectric Rod Waveguide DRW and changes the propagation constant of the wave in the DRW resulting in a phase shifter. The HIS is realized with graphene and the resonance frequency is tuned by changing the graphene electro-chemical potential: around HIS resonant frequency the phase shifts from π to $-\pi$. The phase shifter is realized by using the DRW because it is advantageous in the sub-THz region due to smaller losses compared to any other waveguides. A DRW with high permittivity material (Si, *εr*=11.68) and cross section 1.0 mm x 0.5 mm is considered and the excitation of the DRW is realized with a metal waveguide WR-10.