CGW Workshop '17

KRAKÓW, POLAND
OCTOBER 23-25, 2017

Proceedings

Editors: Marian Bubak
Michał Turała
Kazimierz Wiatr

Published in October 2017

by Academic Computer Centre CYFRONET AGH
ul. Nawojki 11, 30-950 Kraków, P.O. Box 386, Poland

ISBN 978-83-61433-24-8

Cover design and book typesetting by Mieczysław Pilipczuk

Preface

Traditionally, the last week of October in Krakow is the time of the annual meeting of the community of researchers, developers, and users who work in the innovative research area of e-Science addressing challenges in applications, tools, software environments, and distributed computing infrastructures of which the PL-Grid infrastructure is the most mature example. This is the main objective of the CGW Workshops initiated in 2001.

The CGW'17 Workshop was organized jointly by the Academic Computer Centre Cyfronet AGH and the Department of Computer Science AGH and it was held on the premises of the AGH (AGH aula and lecture rooms at the Department of Computer Science) from 23 to 25 October 2017. This Workshop addressed the following topics in the keynote lectures, oral sessions, and poster presentations:

- e-Science, system-level science and collaborative applications,
- data intensive applications and tools,
- models, methods and tools for collaborative applications development,
- virtual laboratories and problem solving environments,
- distributed computing infrastructures (DCI),
- knowledge in e-Science,
- machine learning,
- resource management and scheduling,
- monitoring and information management,
- software engineering aspects,
- industrial and social implications.

The following keynote speakers presented their lectures:

- Alexander Boukhanovsky, ITMO University, Russia,
- Mario Cannataro, University "Magna Græcia" of Catanzaro, Catanzaro, Italy,
- Manuel Castañón-Puga, Universidad Autónoma de Baja California, Tijuana, Baja California, México,
- Keith McCormack, The University of Sheffield, Sheffield, UK,
- Eduardo J. Simoes, The University of Missouri-Columbia School of Medicine, Columbia, USA,
- Alfredo Tirado-Ramos, The University of Texas Health Science Center, San Antonio, USA,
- Hai Zhuge, Aston University, Birmingham, UK,
- Robert Adamski, Intel Corporation,
- Ben Bennett, Hewlett Packard Enterprise.

About 30 contributed papers accepted for presentation at the Workshop provide a very good overview of the research activity in the area of e-Science and distributed computing infrastructures. The contributed papers were reviewed by the Steering Committee upon submission.

CGW'17 was also an opportunity to present scientific and technical achievements as well as to overview research in related national and European projects.

On the last day of the CGW'17, the EurValve EU project presented the keynote lecture as well as 8 contributed talks addressing all aspects of the research and software development efforts towards a personalised decision support system for heart valve diseases. The most of computer simulations of the EurValve were performed on the PL-Grid infrastructure. These sessions were concluded by a presentation of the CECM – a unique EU-funded H2020 Teaming project aiming at establishing a new *Centre for New Methods in Computational Diagnostics and Personalized Therapy*.

We are indebted to the members of the Workshop Secretariat (Robert Pająk and Lucyna Bodek) and other colleagues from the ACC Cyfronet AGH (Mieczysław Pilipczuk, Mariusz Sterzel, and Angelika Zaleska-Walterbach) for organizing the event. Special thanks go to Mieczysław Pilipczuk for his work on preparation of the Proceedings for printing in a very short time and Robert Pająk for keeping the updated CGW'17 website and for organizing all social components of the event.

We also owe thanks to the Workshop sponsors: Intel, Hewlett Packard Poland, and the Development Department of the Municipality of Kraków City for their generous support.

We kindly invite you to visit the Web page of the CGW'17 (`www.cyfronet.krakow.pl/cgw17/`) to see the poster and oral presentations as well as the photo gallery.

Finally, we would like to invite you to present your new results of research at the CGW'18.

Kraków, October 2017

Marian Bubak
Michał Turała
Kazimierz Wiatr

Organization

CGW 2017 was organized by:

- **Academic Computer Centre CYFRONET AGH** (ACC CYFRONET AGH)
- Department of Computer Science AGH (DCS)

Steering Committee

Marian Bubak	- DCS / ACC CYFRONET AGH, Kraków
Jacek Kitowski	- DCS / ACC CYFRONET AGH, Kraków
Michał Turała	- IFJ PAN / ACC CYFRONET AGH, Kraków
Kazimierz Wiatr	- ACC CYFRONET AGH, Kraków

Organizing Committee

Kazimierz Wiatr	- ACC CYFRONET AGH
Marian Bubak	- DCS / ACC CYFRONET AGH
Robert Pająk	- ACC CYFRONET AGH
Mietek Pilipczuk	- ACC CYFRONET AGH
Mariusz Sterzel	- ACC CYFRONET AGH
Michał Turała	- IFJ PAN / ACC CYFRONET AGH

Sponsors

Table of Contents

Invited Lectures

Contributed Papers

PLGrid Infrastructure: a Tool for Open Science

Jacek Kitowski[1,2], Kazimierz Wiatr[1,3], Łukasz Dutka[1], Tomasz Szepieniec[1], Mariusz Sterzel[1] and Magdalena Kruszelnicka[1], Klemens Noga[1], Robert Pająk[1]

[1] AGH University, ACC Cyfronet AGH – Competence Centre for Cloud and Grid Computing, ul. Nawojki 11, 30-950 Kraków, Poland
[2] AGH University, Faculty of Computer Science, Electronics and Telecommunications, Department of Computer Science, al. Mickiewicza 30, 30-059 Kraków, Poland
[3] AGH University, Faculty of Computer Science, Electronics and Telecommunications, Department of Electronics, al. Mickiewicza 30, 30-059 Kraków, Poland

e-mails: kito@agh.edu.pl, {k.wiatr, l.dutka, t.szepieniec, m.sterzel, m.kruszelnicka, k.noga, r.pajak}@cyfronet.pl

Keywords: IT infrastructure, computing platforms, domain-specific services, clouds, grids

1. PLGrid infrastructure status

PLGrid – a nationwide distributed infrastructure for scientific computing, created within the PL-Grid project (2009-2012) [1,2] – comprises not only high performance computing hardware, but also mass storage and dedicated tools for deployment of scientific applications on the available resources.

To lower the barriers required for the researchers to use the infrastructure, it has been extended with domain-specific environments, solutions and services, developed according to the identified needs of 27 different groups of scientists. All these tools for facilitating use of the resources developed within the PLGrid Plus (2011-2014) [1,3] and PLGrid NG (2014-2015) [1] projects, cover a wide range of specialties − including provision of the dedicated software, mechanisms of data storage, modern platforms integrating new type of tools and specialized databases – and speed up obtaining scientific results. Issues studied within nearly thirty domain grids using the PLGrid infrastructure include modelling of energy demands, drug and new material design or simulation of complex metallurgical processes.

Further, significant extension of the PLGrid infrastructure resources, together with increase of the quality of IT services offered to the users, was implemented within the PLGrid Core project (2014-2015) [1]. As a result, the users have obtained access to a set of new basic services allowing for easier integration of their solutions, specific to the selected fields of science, with the PLGrid infrastructure. These services include: uniform access to distributed data, a Platform as a Service (PaaS) Cloud or applications maintenance environment, facilitating computations using BigData paradigm.

Moreover, a great computing power and huge storage for digital data are being offered to users what significantly extends the amount of computing resources provided to the scientific community. Training, consultancy and support is widely offered to the users by ACC Cyfronet AGH as a Competence Centre for Cloud and Grid Computing.

2. Polish supercomputers on the TOP500 list

The most recent edition of TOP500 – the list of the world's fastest computers – has been published on June 19th, 2017 at the ISC'17 in Frankfurt, Germany. The full list of Polish supercomputers on TOP500 is as follows: 71 – Prometheus, ACC Cyfronet AGH, 131 – Hetman, PSNC, Polish Academy of Sciences, Poznan, 135 – Tryton, TASK, Gdańsk University of Technology, 161 – Okeanos, ICM, University of Warsaw, 275 – Bem, WCNS, Wrocław University of Science and Technology, 490 – ICM, University of Warsaw.

3. Prometheus and Zeus supercomputers

The most powerful supercomputer in the history of Poland – Prometheus – has been launched in April 2015. Thanks to the innovative technology of direct liquid cooling, Prometheus is the most energy-efficient computer in Poland and one of the most energy-efficient computers in its class in the world.

Prometheus' theoretical performance is almost 2.4 PFlops. It consists of more than 2,200 servers based on HP Apollo 8000 platform, combined with super-fast InfiniBand network with 56 Gbit/s capacity. Its energy saving and high-performance Intel Haswell latest-generation processors offer more than 53,000 cores. These are accompanied by 279 TB RAM in total, by two storage file systems of 10 PB total capacity, and 180 GB/s access speed. Prometheus has also been equipped with 144 NVidia Tesla GPGPUs. Prometheus can serve for data results analysis, numerical simulations, (big) data processing and advanced visualisations provision.

For less demanding computing tasks, supercomputer Zeus still offers its computing power. Zeus has got more than 25,000 computing cores with total theoretical computational power of 374 Tflops.

4. PLGrid for Open Science

The research portfolio carried out with the help of the Zeus and, recently, Prometheus is quite reach. Among others, it includes: prediction of 3D protein structures, study of semiconductor nanostructures and catalytically activity molecules as well as effective biosensors. Computations are also used to study the behaviour of galaxies in a wide range of electromagnetic spectrum, for nuclear magnetic resonance modelling for the purposes of structural analysis of molecular systems, antidots in quantum world, for structural characteristics of human telomeres and complexity of the financial markets.

Scientific computations do not include simulations only. Computing power is utilised by Polish scientists also within international projects like CTA, EPOS, LOFAR and LHC in CERN. With the help of dedicated software packages the supercomputers perform analyses of large and dispersed data sets as well as provide advanced visualisations.

In total, in 2016 the two most powerful supercomputers in ACC Cyfronet AGH – Zeus and Prometheus, executed 7 748 677 jobs with a total duration of 24 653 years of CPU time.

Access to the Prometheus and Zeus' resources is done via the PLGrid infrastructure dedicated to scientists. It provides services and resources from federated Polish all five academic supercomputer centres, coordinated by ACC Cyfronet AGH. It is not only an IT infrastructure with raw computational resources and storage, but also includes a variety of fine tailored tools and services. PLGrid is a unique research platform providing:

- computing resources – distributed resources of all five academic HPC centres available in PLGrid provide more than 5 PFlops,
- storage – more than 40 PB of storage space and fast scratch files systems enable big data processing and analyses,
- scientific software – vast portfolio of tools, libraries and scientific applications for research in various fields of science,
- tools for scientific collaboration – tools and services such as Stash Git repositories server or JIRA issue and project tracking solution ease scientific projects coordination and communication between researchers,
- Cloud computing – PLGrid's PaaS provides elastic solutions for computational environment which can be easily adapted to researchers' needs.

Together with the computing infrastructure, in PLGrid we also provide a selection of domain-specific tools and services, which enable researchers to perform complex, large-scale experiments and manage their results in an easy way. These tools are gathered into 27 scientific domains, dedicated to important scientific topics and strategic fields of Polish

science. They have been tailored to needs of specific scientific communities. Among others, we offer advanced tools and graphical interfaces that enable construction of dedicated environments for scientific research, building application portals, conducting virtual experiments, visualization of calculations' results, executing complex scenarios with parallel tasks, as well as supporting uniform and efficient access to data. Available tools, solutions and services can be browsed in the PLGrid Applications and Services Catalogue [4].

5. Conclusions

The above-mentioned computing resources and services are provisioned in the framework of the PLGrid infrastructure, allowing Polish scientists and their foreign collaborators to access them in a convenient manner.

All of these services are important support for researchers, as they have an impact on improving and, where possible, automating the work of research groups, what greatly accelerates obtaining research results.

Acknowledgment. This work was made possible thanks to the following projects: PLGrid Plus POIG.02.03.00-00-096/10, PLGrid NG POIG.02.03.00-12-138/13 and PLGrid Core POIG.02.03.00-12-137/13, co-funded by the European Regional Development Fund as part of the Innovative Economy program, including the special purpose grant from the Polish Ministry of Science and Higher Education.

References

1. The Polish Grid Infrastructure web site: `www.plgrid.pl`,
2. Building a National Distributed e-Infrastructure – PL-Grid. Scientific and Technical Achievements, M. Bubak, T. Szepieniec, K. Wiatr (Eds.). Springer LNCS, Vol. 7136 (2012),
3. eScience on Distributed Computing Infrastructure. Achievements of PLGrid Plus Domain-Specific Services and Tools, M. Bubak, J. Kitowski, K. Wiatr (Eds.). Springer LNCS, Vol. 8500 (2014),
4. PLGrid Applications and Services Catalogue: `apps.plgrid.pl` .

Ethnography and Computation: Towards Agent-Based Modelling from Ethnography Observations

Manuel Castañón-Puga

Universidad Autónoma de Baja California, Tijuana, Baja California, México

Ethnography is the part of anthropology that deals with the scientific description of human communities. The ethnographer mainly records stories in text field notes form, to describe scenarios which he has observed and lived with members of a community. The narratives are the first steps to studying their social and cultural systems. The Agent-Based Modelling (ABM) is a computational modelling paradigm used to approach the society members. On this talk, we examine some computational ideas starting from ethnography observations toward agent-based modelling for social simulation. We used IBM Watson services to explore ways to discover entities and relationships from ethnographic text sources. We present some experiences and results of this first approach, and we discuss the methodological issues for agent-based modelling and simulation.

Dr Manuel Castañón-Puga is a full Professor of computer sciences and computer engineering at the Universidad Autónoma de Baja California in México. He's the leader of the "Complexity and Computation" academic group and the "Computational Modeling and Complex Systems" academic collaboration network. His current research interests include multi-agent systems, hybrid-intelligent software agents, social simulation, social-inspired ICT, social-computation, computational intelligence, complexity, and software and computer engineering.

His research of modelling and simulation, agent-based simulation, hybrid-intelligent agents and multi-agent systems explores the way in which software agents could be used to describe multidimensional environments, innovation, evolution and adaptation in complex adaptive systems. He collaborates with multidisciplinary researchers and scientists to create multi-dimensional computer simulations of societies, political ideologies, trading economies and urban landscapes. His research also intends to incorporate the ideas of complexity into the mainstream of computer engineering and in particular to its instruction at the undergraduate and graduate levels.

Effect of Health Information Technologies in Glycemic Control among Patients with Type II Diabetes

Eduardo J. Simoes[1], Susan Boren[1],
Mihail Popescu[1], Diana Kennedy[1], Jesus Soares[2]

[1] University of Missouri School of Medicine, Department of Health Management and Informatics, Columbia, USA
[2] Centers for Disease Control and Prevention, Division of Nutrition, Physical Activity and Obesity

Objective. The purpose of this meta-analysis was to synthesize findings and reveal the effects of health information technologies (HITs) on glycemic control among patients with type II diabetes.

Methods. We systematically searched Medline, Cumulative Index of Nursing and Allied Health Literature (CINAHL), and the Cochrane Library for peer reviewed randomized control trials that studied the effect of mobile or potentially mobile technology on glycemic control (HbA1c). We also used Google Scholar to identify additional studies not listed in the abovementioned databases. We performed a hand search using references lists of eligible articles and of relevant systematic reviews and review articles to identify potential missed articles. We analyzed data using random effects meta-analytic models.

Results. 20 studies (25 estimates) met the criteria and were included in the analysis. Overall, HITs resulted in a statistically and clinically reduced estimated average glucose level (HbA1c %). The combined HbA1c reductions was -0.700 [Standardized mean difference (SMD) = -0.700, 95% CI (-0.916, -0.485)]. The reduction is significant across all four types of HIT intervention under review, with short message services and mobile phone-based approaches generating bigger effects [SMDs were -0.757 (-0.996, -0.517) and -0.716 (-0.941, -0.490), respectively].

Conclusions. HITs can be an effective tool for glycemic control among patients with type II diabetes. Future studies should examine HITs' long-term effects and explore factors that influence the effectiveness.

Dr. Eduardo J. Simoes is Chair and Alumni Distinguished Professor in the Department of Health Management and Informatics, University of Missouri School of Medicine. He has a Doctor of Medicine (University of Pernambuco-Brazil), Diploma and MSc in Community Health (London School of Hygiene and Tropical Medicine, University of London-England), and MPH (Emory University). Previous notable appointment: Director of the Prevention Research Centers Program at the Centers for Disease Control and Prevention, State Epidemiologist in Missouri. In the fields of public health, medicine, health informatics and epidemiology, he has published over 120 peer-reviewed articles, nine book chapters and over 30 reports; and presented 150 conference papers worldwide.

Patient-Centered Computable Phenotyping in Health Disparities Research

Alfredo Tirado-Ramos

The University of Texas Health Science Center, San Antonio, USA

Computable phenotypes are sets of computable inclusion/exclusion criteria for patient cohorts. Such criteria should be specific and objective enough that they can be turned into machine-readable queries, yet are generalized enough that they are portable between different data sources. There are a number of methods for creating and consuming computable phenotypes, such as OMOP, PCORNet Front Door, i2b2, and SHRINE, though the biggest challenge is still creating a baseline infrastructure to understand these systems well enough to use them in cutting edge biomedical research. In this talk we will discuss our experiences, lessons learned and next steps in creating a cluster of excellence in biomedical informatics research based on computable phenotyping.

As the chief and founder of the Clinical Informatics Research Division of the University of Texas Health Science Center at San Antonio, Dr. Tirado-Ramos leads a full-spectrum biomedical informatics program and explores the intersection between informatics, translational science, and clinically relevant data-centric problems including, but not limited to, computable phenotype-based research in health disparities, obesity, amyotrophic lateral sclerosis, aging, and cancer. He and his team have created and maintain an information research system for interdisciplinary collaboration between pediatric endocrinologists, cancer researchers and neurologists, creating new institutional governance frameworks along the way. He also co-directs the informatics core at the Claude Pepper Older Americans Independence Center, a National Institute on Aging award, where he works on state of the art informatics infrastructures to investigate innovative interventions that target the aging process as well as aging-related diseases, with a major focus on pharmacologic interventions. Previous to arriving at the University of Texas, he served at Emory University School of Medicine as Associate Director for the Biomedical Informatics Core at the Center for AIDS Research at the Rollins School of Public Health. He also served as Scientific Member of the Winship Cancer Institute Prevention and Control Program and Senior Member of the Research Staff at the Center for Comprehensive Informatics. His work at Emory University focused on informatics applied to clinically-relevant biomedical challenges, including the correlations between infectious disease and cancer, as well as whole genome sequencing for vaccine development research.

The Next ‘Giant Leap’ for Computing

Ben Bennett

Hewlett Packard Enterprise

This talk looks at how HPE and NASA are addressing the issues of on-board computing required for Mars travel. Moreover, similar problems back here on Earth are causing HPE to redefine the computer, as we know it.

Solving Atari Games with Distributed Reinforcement Learning

Robert Adamski

Intel Corporation

Intel collaborates with partners like deepsense.ai to make a mark on the cutting-edge research leading towards intelligent machines by providing practical machine learning tools and designs that make it much easier for scientists to track their experiments and verify novel ideas. One particular step towards achieving this was to distributing a state-of-the-art Reinforcement Learning algorithm on a large Intel Xeon cluster, allowing super-fast training of agents that learned to master a wide range of Atari 2600 games.

Urgent Computing for Metocean Extreme Events: St. Petersburg's Flood Protection Barrier

Prof. Dr. Alexander Boukhanovsky

ITMO University, Russia

The report presents the experience of creating a flood prevention system in St. Petersburg based on the principles of urgent computing. The main models, workflow, planning methods and practical implementation, as well as the estimation of the quality of calculations are described. The system has been successfully in operation since 2011, during this period of time 16 floods were prevented.

Prof. Dr. Alexander Boukhanovsky is the Chair of High Performance Computing (HPC) Department in ITMO University. In 2005, he defended his dissertation on Concurrent Software Statistical Measurements of Spatial-Temporal Fields. Since 2006 he has been working as a Professor of Information Systems and Head of the Parallel Software lab in ITMO University. In 2007 he created the eScience Research Institute where his team has created CLAVIRE (CLoud Applications VIRtual Environment).

In recent years he attracted several grants including mega grants of the Russian Federation Government, e.g. decree #220 "on measures to attract Leading Scientists in the Russian educational institution" and decree #218 "cooperation of Russian higher education institutions and organizations implementing complex projects of high-tech industry".

His research interests are high-performance computing, computer modelling of complex systems, intelligent computational technologies, statistical analysis and synthesis of spatial-temporal fields, parallel and distributed computing, distributed environments for multidisciplinary researches, decision support systems and technologies, statistical analysis and simulation in marine sciences. He is the author of 230 publications (cited over 1000 times) and has successfully advised 23 PhD candidates.

Human-Machine-Nature Symbiosis

Hai Zhuge

Aston University, Birmingham, UK

Cyberspace is being more and more tightly linked to the physical space and socioeconomic space to emerge a cyber-physical society, where humans, machines and the natural environment interact with each other, efficiently share resources and co-evolve to emerge patterns from different spaces. The emerging cyber-physical society is providing a new environment for experiencing, understanding and thinking. Human-machine-nature symbiosis is a basic mechanism that coordinates humans, machines and natural environment to realize the harmonious development of cyber-physical society. Studying the fundamental symbiotic mechanism in cyber-physical society is a way to understand cyber-physical society and influence its evolution. Human-machine-nature symbiosis is a basic mechanism for developing cyber-physical society and investigating its influence on computing, intelligence and society. Human-machine-nature symbiosis provides a flow-driven symbiotic method for studying Cyber-Physical-Social Intelligence and managing the sustainable development of cyber-physical society.

Dr Hai Zhuge is a professor in Aston University and Chinese Academy of Sciences, a Distinguished Scientist of ACM and a Fellow of British Computer Society. His major research interest is to explore the fundamental issues on semantics, knowledge, dimension, and self-organisation in a multi-disciplinary background. One of his contributions is semantic modelling. He created the Semantic Link Network model and the Resource Space Model and integrated them as a fundamental semantic space to support semantics-based management and exploration of various resource spaces. Research has established a uniform theory and method for modelling, organising, retrieving, and managing both cyber objects and concepts so that various information services can be efficiently provided with understanding. A set of high-performance semantics-based distributed platforms was established to provide a self-organised and adaptive architecture for efficiently sharing and managing various cyber objects. His model, theory and method have been applied to many applications, including summarisation, question answering and recommendation. In recent years, he is leading research towards a new science and engineering for Cyber-Physical Society. Homepage: `http://www.knowledgegrid.net/~h.zhuge/`.

Efficient Preprocessing and Analysis of Omics Data

Prof. Mario Cannataro

Department of Medical and Surgical Sciences & Data Analytics Research Center
University "Magna Græcia" of Catanzaro
88100 Catanzaro, ITALY

e-mail: cannataro@unicz.it

Genomics, proteomics, and interactomics disciplines are gaining an increasing interest in the scientific community due to the availability of novel, high throughput platforms for the investigation of the cell machinery, such as mass spectrometry, microarray, next generation sequencing, that are producing an overwhelming amount of experimental omics data. The increasing volume of omics data poses new challenges both for their efficient storage and integration as well as for their efficient preprocessing and analysis. Managing omics data requires both support and spaces for data storing, as well as efficient, possibly parallel algorithms for data preprocessing and analysis. The resulting scenario comprises a set of methodologies and bioinformatics tools, often implemented as services, for the management and analysis of omics data stored locally or in geographically distributed biological databases. The talk describes some parallel and distributed bioinformatics tools for the preprocessing and analysis of genomics, proteomics and interactomics data, developed at the Bioinformatics Laboratory of the University Magna Graecia of Catanzaro. Tools for efficient statistical and data mining analysis of mass spectrometry proteomics data (MS-Analyzer, EIPEPTIDI), as well as gene expression and genotyping data (micro-CS, DMET-Analyzer, DMET-Miner, OSAnalyzer, coreSNP) will be briefly underlined.

Mario Cannataro is a full professor of computer engineering at the University "Magna Graecia" of Catanzaro, Italy. He is the Director of the Data Analytics research centre of University of Catanzaro. His current research interests include bioinformatics, parallel and distributed computing, data mining, problem solving environments, and medical informatics. He is a Member of the editorial boards of Briefings in Bioinformatics, Encyclopaedia of Bioinformatics and Computational Biology, Encyclopaedia of Systems Biology. He was guest editor of several special issues on bioinformatics and he is serving as a program committee member of several conferences. He published three books and more than 200 papers in international journals and conference proceedings. Prof. Cannataro is a Senior Member of IEEE, ACM and BITS (Bioinformatics Italian Society), and a member of the Board of Directors of ACM Special Interest Group on Bioinformatics, Computational Biology, and Biomedical Informatics (SIGBio).

‘EurValve’ as a Map for Pervasive Computational Healthcare

Keith McCormack

The University of Sheffield, Sheffield, UK

Healthcare is rightly the most conservative of sophisticated human endeavours: ‘better safe than sorry’. But complexity has conspired with necessity, to drive clinical practice into the arms - indeed the code-typing fingertips - of the computational medical physicist, where image processing, machine learning, 4D modelling and predictive analysis can outstrip the keenest minds and the longest memories, with the speed, accuracy and infinite data resources of computational medicine. The future is a world of continuous predictive health information, constantly updated with every step, every meal, every sneeze. The fundamentals of this integrated, pervasive HealthCare 2050 are already being established, and many can be witnessed in the EC-funded EurValve project, where the machines make the measurements, calculate the odds, and suggest the treatments.

Scheduling Challenges in a Shared Private Cloud Infrastructure

Dalibor Klusáček[1], Boris Parák[1], Lukáš Hejtmánek[2]

[1]CESNET, Prague, Czech Republic
[2]Institute of Computer Science, Masaryk University, Brno, Czech Republic

e-mails: klusacek@cesnet.cz, parak@cesnet.cz, xhejtman@ics.muni.cz

Keywords: cloud, grid, resource sharing, utilization, fairness, scheduling, OpenNebula

1. Introduction

The Czech National distributed computing infrastructure MetaCentrum is a hybrid infrastructure that enables simultaneous deployment of several types of computing paradigms, i.e., grid and cloud computing as well as "MapReduce-like" computations on a dedicated Hadoop cluster. Currently, hardware resources (nodes) can be allocated either to the cloud or the grid [1]. So far, the largest fraction of all computing resources (~65%) is semi-permanently allocated to the classic "bare metal" grid-like computations, while the remaining part (~35%) is virtualized using OpenNebula framework to dynamically run either general grid "worker nodes" or classic users' virtual machines (VMs). This approach is very useful since it allows to dynamically re-allocate computing resources among the two major domains—the grid and the cloud—significantly reducing resource fragmentation [2].

Although the shared infrastructure is beneficial compared to a setup where resources are statically and permanently dedicated to either the grid or the cloud framework, several scheduling challenges remain opened. For example, since the resources are provided for free, i.e., unlike in public clouds (Amazon, Google, etc.), users are not directly motivated to care about their resource consumption. At the same time, our resource pool is—of course—limited, therefore we cannot always guarantee that resources will be available when a (new) user wants some of them. This particular feature is not a big issue in grid, since batch computations are always time-constrained and are not so sensitive to waiting. However, the nature of the cloud service is different. Here, users usually expect (nearly) immediate start of their VMs. Furthermore, VMs lifetime is not strictly constrained, i.e., VMs often run much longer than common batch jobs. This brings some fundamental challenges. For example, (a) there should always be some reserve of free resources in the cloud node pool to allow for immediate starts of new VMs, while (b) those already occupied nodes are often underutilized by inactive/idle VMs. Also, (c) it is not easy to guarantee fairness in our cloud, because well-known approaches like grid's fair-sharing do not suit very well to the nature of cloud computation. Together these somehow contradictory requirements cause that cloud resource demands grow in time while the overall CPU utilization in the cloud partition is rather low [2]. In this paper we briefly demonstrate some of these issues using real data from our system and present possible solutions intended to solve at least some of these challenges, e.g., fairness-related issues and (currently) low resource utilization.

2. Real-life system comparison

To demonstrate the differences among the usage patterns between the grid and the cloud we provide Fig. 1 (left and middle), which shows users' CPU usage over the time for grid jobs (left) and cloud VMs (middle). Clearly, the time span a given set of resources is dedicated to a given user is very constrained in the grid, while in the cloud the situation is the opposite. Apparently, users are not very motivated to dispose their old VMs, and those resources are often allocated for a given user for several months. At the same time (see Fig. 1 right), the

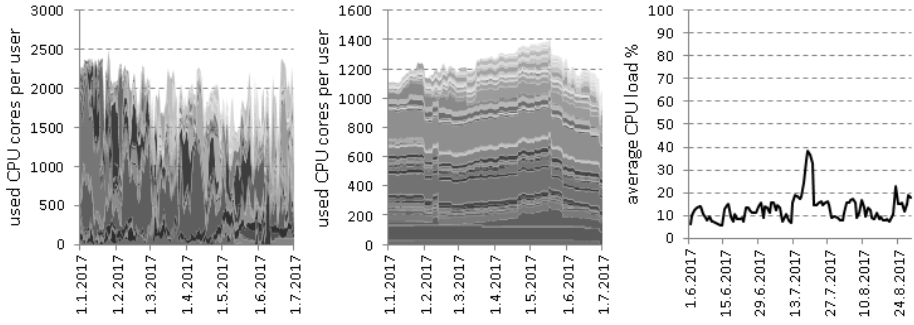


Fig. 1. Users' CPU usage in grid (left) vs. cloud (middle). Average CPU load in the cloud (right).

actual CPU load of cloud VMs is rather low (13% on average), meaning that those resources allocated for VMs are heavily underutilized. This is not the case in the grid, where the average CPU load is ~80%. Also, Fig. 1 (left and middle) reveals the practical effects of (not) using fair-sharing mechanism. Grid partition prioritizes users dynamically using fair-share, which is clearly visible in the Fig. 1 (left), where users use the resources proportionally over the time. In the cloud, there is no such functionality, thus users keep their existing allocations for long time periods. This is unfair and it may impede new users.

3. Proposed solutions and future work

To approach these issues we need to refine the policies as well as scheduling mechanisms [3] used in our system. First, we have introduced a limited lifetime for a VM (3 months by default), to limit the number of "zombie" VMs remaining in the system. Next, we plan to develop a prioritization mechanism for user/VM classification. Users/VMs with low priorities will then face much more aggressive CPU overcommitting that will increase the actual load while leaving more space for high priority workloads. This classification can also include some fairness-oriented mechanism. Furthermore, "scavenge-like" grid compu-tations using short batch jobs from the grid can be used in the cloud when VMs are idle (e.g., in the night) to further utilize free CPU cycles.

More advanced VM (re)scheduling algorithm is also an interesting option with many potential benefits. However, here we are facing many practical problems, e.g., the "black box" nature of OpenNebula scheduler [3] as well as the limitations of both the OpenNebula and our infrastructure (e.g., the inability to resize or to migrate a live VM).

Acknowledgements. We kindly acknowledge the support provided by the MetaCentrum under the program LM2015042, the CERIT Scientific Cloud under the program LM2015085 and the project Reg. No. CZ.02.1.01/0.0/0.0/16_013/0001797 co-funded by the Ministry of Education, Youth and Sports of the Czech Republic.

References

1. D. Klusáček, G. Podolníková. "Scheduling Hybrid Workloads in Shared Cloud Infrastructures". In Cracow Grid Workshop, 29-30, ACC Cyfronet AGH, 2016,
2. D. Klusáček, B. Parák. "Analysis of Mixed Workloads from Shared Cloud Infrastructure". In Job Scheduling Strategies for Parallel Processing, 2017,
3. G. Podolníková, B. Parák, D. Klusáček. "Extensible and Modular Cloud Scheduler for OpenNebula". In Cracow Grid Workshop, 45-46, ACC Cyfronet AGH, 2015.

PLGrid – Biomedical Module

Dawid Dułak, Małgorzata Gadzała, Irena Roterman

Jagiellonian University – Medical College, Krakow, Św. Anny12, Poland

e-mail: myroterm@cyf-kr.edu.pl

Keywords: PLGRID, biomedical module, protein folding

1. Introduction

Biomedical module has been organized in PLGrid system. The main goal of this module is the simulation of protein folding process. The availability of this module is limited to the group of developers at Medical College – Jagiellonian University. The publication describing the successful results is the condition to make it open for external users.

Despite of 50 years long history of protein structure project, the mechanism of folding is still unknown.[1]. The top groups in this discipline participate in the world wide project called CASP (Critical Assessment of Protein Structure Prediction). The progress in protein folding is still in status of development without satisfactory results.

The model called fuzzy oil drop model (FOD) elaborated by specialists at Jagiellonian University – Medical College introduces the new ideas: two-step model of folding process [2] and incorporating the force field based on water environment to the traditionally used force fields expressing solely the internal interactions of atom-to-atom interaction [3]. The dual minimization of internal energy (pair-wise interatomic interactions) and minimization of factors expressing the influence of environment of water is introduced in this model. The main purpose of the biomedical module is to make this model available to wider spectrum of users.

2. Description of a problem solution

The main idea of FOD model is to express the external force field in form of 3D Gauss function. This function represents the well known idea of hydrophobic core. It assumes the highest concentration of hydrophobicity in the center of protein molecule and exposure of hydrophilic residues on the surface making the protein molecule soluble [3]. According to FOD model the 3D Gauss function expresses this type of distribution in the perfect form [4]. This is why the inter-residual hydrophobic interaction is compared with the expected one directing the hydrophobic resides toward the center and exposing hydrophilic residues on the surface.

The hydrophobicity distribution is the criteria for folding process together with internal energy minimization (condition for spontaneity of the folding process). The Gromacs program is implemented in the Biomedical Module for conformational changes and internal energy minimization [5].

3. Results

The examples of successful simulations as well as the examples of incorrect results will be shown on the poster. The final structures are compared with the experimentally examined ones using the parameters following the CASP project to make the classification unified with top groups in the discipline of protein structure prediction [6].

The presentation of failures is aimed on the identification of sources of failure. The examples of failure are more informative due to the presence of significant errors making the simulation path incorrect.

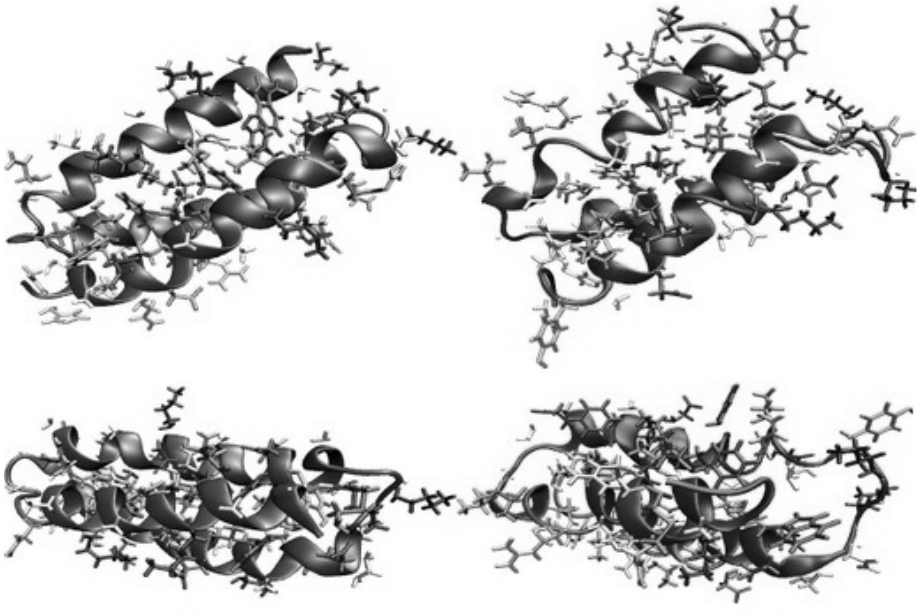


Fig. 1. Examples of graphic presentation of final results – proteins folded according to FOD model.

4. Conclusions and future work

The set of proteins (about 50) was used to test the model. Two papers are in preparation. They present the successful results as well as the failures. The later are used to search for source of discordance of result structures in respect to the experimentally determined ones. The reliable model for protein folding is of high importance in computer-aided drug design. The fast design of drug is critical for personalized medicine, which is the challenge for current medicine.

References

1. Dill KA, MacCallum JL. The protein-folding problem, 50 years on. Science. 2012 ; 338(6110): 1042-6,
2. Creighton TE. Protein folding. Biochem J. 1990;270:1–16,
3. Kalinowska B, Banach M, Konieczny L, Roterman I. Application of Divergence Entropy to Characterize the Structure of the Hydrophobic Core in DNA Interacting Proteins. Entropy, 2015, 17(3), 1477-1507,
4. Roterman I, Banach M, Kalinowska B, Konieczny L. Influence of the Aqueous Environment on Protein Structure—A Plausible Hypothesis Concerning the Mechanism of Amyloidogenesis. Entropy 2016, 18(10), 351; doi:10.3390/e18100351,
5. Kutzner C, Páll S, Fechner M, Esztermann A, de Groot BL, Grubmüller H. Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. J Comput Chem. 2015; 36(26): 1990-2008,
6. Gadzała M, Kalinowska B, Banach M, Konieczny L, Roterman I. Determining protein similarity by comparing hydrophobic core structure. Heliyon. 2017 ; 3(2): e00235.

Towards Stable Co-evolution of Deep Neural Networks and Fitness Predictors

Paweł Koperek[1], Włodzimierz Funika[1], Jacek Kitowski[1,2]

[1] AGH, Faculty of Computer Science, Electronics and Telecommunication, Dept. of Computer Science, al. Mickiewicza 30, 30-059, Kraków, Poland,
[2] AGH, ACC CYFRONET AGH, ul. Nawojki 11, 30-950, Kraków, Poland,

e-mails: pkoperek@gmail.com, {funika,kito}@agh.edu.pl

Keywords: deep neural networks, co-evolution, fitness prediction

1. Introduction

Deep neural networks (DNN) are a very powerful machine learning technique. They have numerous applications, with state-of-the-art performance reported in several domains, like visual object recognition or text processing. Unfortunately, choosing the correct network topology to model a specific problem, is not straight-forward. Using the automated methods for creating deep neural network models would greatly improve their quality and speed up creating innovative structures. The factor which limits the usability of such an approach is the time of training - the time that it takes to evaluate the model. The evolutionary methods would become applicable if only the evaluation time could be reduced. One method, which can help in such a situation, is the co-evolution of so called fitness predictors [1. In this approach, two populations of individuals are maintained. The first one improves potential so1utions to the problem, which the process is supposed to solve. The second one works on finding estimators, which can be used to perform a low-cost approximation of fitness within the first population.

In [2] we demonstrated that using this technique provides very promising results, however the stability of evolution needs to be improved. Even though at the end of the process, the overall fitness has been improved, in some iterations the fitness was decreasing instead of remaining at the same level or increasing. Such a behavior is not surprising, because in every iteration potentially a new fitness predictor is used to evaluate the main population. This means that the set of samples used for training the DNN can change a lot between iterations, what will result in obtaining a different fitness values for the same individual.

This problem can be addressed in different ways. In this paper we explore three of them: 1) training fitness predictors population over multiple iterations per single DNN population iteration, 2) increasing the number of individuals used as trainers of the fitness predictors population, 3) increasing the number of individuals used as trainers for the DNN population. Further we discuss the advantages and disadvantages of those methods and validate experimentally our hypotheses.

2. Description of a problem solution

In the first approach, we explore training fitness predictors over multiple iterations per single DNN population iteration. By going through more iterations of evolution, fitness predictors can align better to the trainer and discover what subset of samples provides the best estimation of the original, large dataset.

The second approach focuses on the fact that by allowing to change the trainer in each iteration, we can lose the progress of the population made by using the previous trainers. The direction set by the new trainer might be exactly the opposite as compared with the replaced one. In order to combat this, we postulate to extend the population of trainers with individuals which were elected as trainers for the previous iterations. This would prevent from rapidly changing the direction of evolution, but also could possibly slow down the progress.

Finally, we want to explore the possibility to improve the quality of fitness predictors by providing multiple training DNNs. Fitness predictors are evolved to give evaluation results as close as possible to the original dataset. When using only a single DNN to train them, we risk that it can be skewed in some way, what would promote fitness predictors which align well with that skew. By using multiple trainers we are trying to avoid a situation where a specific DNN can dominate the fitness predictors population and promote only those individuals which work well with skews manifested by this DNN.

Tab. 1 presents results obtained in our experiments. For 25 iterations we evolved a solution to a sample problem of classifying the images from the MNIST [3] dataset. First we established a baseline result (a) which demonstrates the problems experienced without using any techniques to improve the stability. Next we present the results of evolution (b-i) in the described scenarios. To measure the improvement of each experiment, we calculate the absolute value of the sum of all the negative differences between the maximum fitness values in the subsequent iterations. The lower value of this metric, the less the value of fitness is degraded in the course of evolution.

Tab. 1. Results of experiments. Fitness is understood as image recognition accuracy.

Experiment	Maximum fitness (%)			Sum of negative differences (absolute)
	First iteration	Last iteration	Improvement	
(a) Baseline	89,78	89,82	0,04	3,09
(b) 3 FP iterations	89,78	89,63	-0,15	3,3
(c) 5 FP iterations	89,78	89,67	-0,11	1,95
(d) 3 FP, 1 DNN trainer	89,78	89,68	-0,1	2,15
(e) 5 FP, 1 DNN trainer	89,78	89,84	0,06	2,76
(f) 3 FP, 3 DNN trainers	89,78	89,91	0,13	0,91
(g) 5 FP, 3 DNN trainers	89,78	90,00	**0,22**	**0,76**
(h) 3 FP, 5 DNN trainers	89,78	89,67	-0,11	**0,66**
(i) 5 FP, 5 DNN trainers	89,78	89,79	0,01	0,77

3. Conclusions and future work

Our experiments show that the best results were obtained, when the number of trainers was increased together in both populations: fitness predictors and DNNs. In test run g the fitness improvement was highest (0,22%) while the sum of negative differences was the second lowest among all the runs (0,76). In the case, where the negative differences were most strongly reduced (h – 0,66), the evolution did not improve the best solution.

In all the tests, the reduction of fitness was still observed: we were not able to fully eliminate the instability of the evolutionary process. We plan to further experiment with other techniques, e.g. by only evolving the individuals at the size-fitness Pareto front.

Acknowledgements. This research is partly supported by AGH grant no. 11.11.230.337.

References

1. M. Schmidt and H. Lipson: Co-evolving Fitness Predictors for Accelerating and Reducing Evaluations. GPTP 2006, 1, 2006,
2. W. Funika and P. Koperek: Co-evolution of fitness predictors and Deep Neural Networks, accepted for publication in Proc. PPAM 2017, LNCS, Springer,
3. Y. LeCun and C. Cortes: MNIST handwritten digit database. 2010.

Development of a Novel, SPH Solver, for Modelling of Tumor Proliferation

Bartosz Minch, Filip Koperski

AGH University of Science and Technology, Cracow, 30-033 Poland

e-mails: bartosz.minch@outlook.com, f.koperski@gmail.com

Keywords: cancer, modelling, tumor, proliferation, SPH, Navier-Stokes,

1. Introduction

The neoplastic diseases, such as cancer, are the most difficult problems of our civilization. Cancer morbidity is growing and more than 8 million people die each year [2]. Tumor growth is a complex phenomenon and we still do not know all details about this process and possible growth scenarios. The skin cancer – melanoma – is one of the most malignant tumors. Up to now, there are not many numerical models focusing on its dynamics.

Herein we consider two basic setups of growth. First, where tumor develops inside healthy tissue and second, describing tumor invading the skin surface. The main motivation of this research is to check if SPH method has the potential to model multiphase complex biological processes and show the development of cancer in the tissue, which in return would help us to better understanding its incipient evolution.

2. Model

In our SPH model the biological system, consisting of healthy and cancerous tissues, is treated as a two-phase viscous fluid. The healthy tissue serves as a base in which tumor develops. This two-phase fluid dynamics and mechanical interactions between the tumor and normal tissue phases are described by the Navier-Stokes equations, and particularly, very difficult to solve numerically the Cahn-Hilliard equation. The normal and cancerous SPH particles represent certain volume of tissue cells. We assume that the healthy cells do not divide while every tumor cell has its current life cycle stage as an attribute. We consider three stages: “division”, “quiescence” and “necrosis” [3]. Well oxygenated cells can enter the process of mitosis - a process of their division onto two equal cells that contain basic amount of oxygen. In the quiescent state the particle lose ability to enter mitosis and consumption rate is being decreased. The „necrosis” state means that the particle is dead and, subsequently, it is removed from the system. The oxygen concentration is governed by the modified Fick’s second law of diffusion, which influences the current states of SPH particles.

$$\frac{\partial c(r,t)}{\partial t} = D\nabla^2 c(r,t) - n(r,t) \tag{1}$$

We define two other attributes besides the stage of cell life-cycle and oxygen concentration. There are the cell lifetime and cumulative descriptor representing the states of the particle nearest neighbors. As shown in Fig.1, the main setup for simulations consists of a tissue cube with blood vessels. We define two alternative assumptions. First, where the vessels are made of motionless particles and are the sources of oxygen, which diffuses in the bulk tissue, and the second, where the blood capillaries are integrated into healthy particles (i. e., each healthy particle is a source of oxygen). This second assumption can mimic the reality only at the beginning stage of tumor evolution, when the blood vessels do not penetrate tumor tissue.

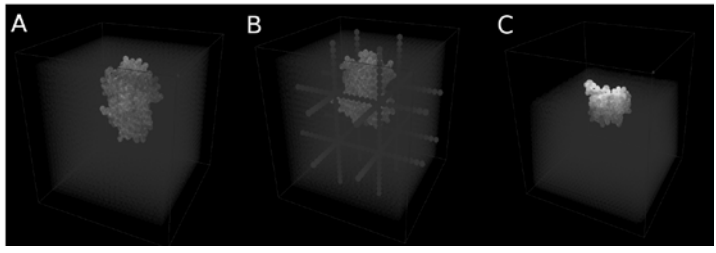


Fig. 1. The snapshot from 3-D SPH simulation. We used three configurations: where tumor is inside healthy tissue with (a) or without (b) blood vessels or where tumor is located on the skin surface (c). The colors of dots represent stages of tumor cells: green – proliferative, yellow – quiescent.

3. Results

We have conducted series of simulations for scrutinizing the effect of specific parameters (e.g., viscosity of tumor, oxygen distribution and diffusion constants) on tumor growth speed. Our results show similarities with other tumor models such as continuous and particle automata models [2]. Simulated SPH system resembles observable shapes of tumor and its fingering properties (see Fig.2) [4].

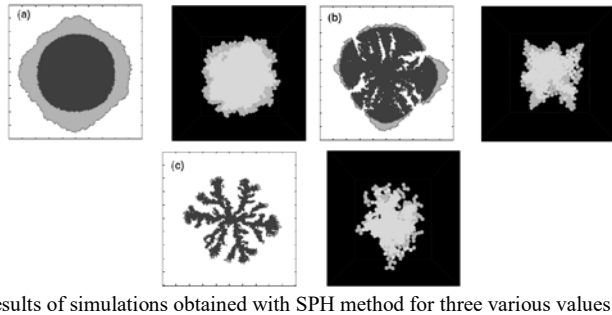


Fig. 2. 3-D results of simulations obtained with SPH method for three various values of oxygen consumption constants compared to the results of tumor 2-D dynamics from [4].

4. Conclusions and future work

The SPH method is a very promising supplement to larger and multi-scale models of cancer. However, some evolution characteristics we received were not satisfactory enough. Mainly, this is due to small spatio-temporal scale of simulation. Most of simulations were carried out by using maximum 100,000 particles. We expect that larger simulations, with a greater number of SPH particles (up to 10, 000, 000) would provide more precise and interesting results.

Acknowledgements: The work has been supported by the Polish National Science Center (NCN) project 2013/10/M/ST6/00531 entitled: Multi-scale model of tumor dynamics as a key component of the system for optimal anti-cancer therapy.

References

1. Thieulot, C., Janssen, L., & Espanol, P. (2005). Smoothed particle hydrodynamics model for phase separating fluid mixtures. I. General equations. Physical Review E, 72(1), 1 - 15. [016713],
2. J. Ferlay, I. Soerjomataram, and M. Ervik, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. GLOBOCAN 2012 v1.0, Lyon, France: International Agency for Research on Cancer, 2013. `http://globocan.iarc.fr/`,
3. V. Cristini and J. Lowengrub, Multiscale Modeling of Cancer. An Integrated Experimental and Mathematical Modeling Approach. Cambridge: Cambridge University Press, 2010,
4. P. Gerlee and A. R. A. Anderson, "Diffusion-limited tumour growth: Simulations and analysis," Mathematical Biosciences and Engineering, Math Biosci Eng. 2010 Apr; 7(2): 385–400.

Pseudo-random Numbers Generators Implemented with FPGA Technology

Marek Wróblewski, Marek Sawerwain

Institute of Control & Computation Engineering
University of Zielona Gora, Licealna 9, Zielona Gora 65-417, Poland

e-mails: M.Wroblewski@issi.uz.zgora.pl, M.Sawerwain@issi.uz.zgora.pl

Keywords: pseudo-random number generators, high-level synthesis, FPGA

1. Introduction

This article discusses the implementation of several selected modern pseudo-random generators for FPGAs, specifically for the Virtex 7 family. The implementation was done using high-level synthesis (HLS) techniques. Available HLS tools allow for the rapid implementation of multiple pseudo-random numbers generators of the same type. This allows you to achieve high performance that can be compared to the performance achieved with the SSE, AltiVEC, or AVX vector instructions used in traditional CPUs. The use of FPGA technology allows for a significant reduction in power demand needed to power the whole of pseudo-random number generator. Generating pseudo-random numbers is a very important task in the context of many computational methods, e.g. based on the Monte Carlo method [1] [2] [5], or cryptography applications [3]. FPGAs can be used as an accelerator in a PC architecture. It is also important to know the efficiency and quantity of energy needed to efficiently generate large amounts of value from pseudorandom numbers generators. The use of a high level synthesis method allows for the rapid implementation of generators in the target FPGA.

2. Description of a problem solution

The developed PRNGFPGLIB library offers access to PRNGs implemented on FPGA of the following types: LFSR113, GM19, GM31, GM61, MRG32k3a, MT19937, as well as XORSHIFT, MWC256, CMWC4096.

Used for implementation in this work is board VC707 with Virtex 7 XC7VX485TFFG1761-2 has exactly 485,760 logical units. Programming of FPGAs is done mainly in VHDL and Verilog languages. However, there is the possibility of so-called high-level synthesis, e.g. in C/C++. The generators were implemented with Vivado HLS [4] and Vivado 2016.4.r. The expression #pragma HLS INLINE off, used means that the function will not be embedded in the main function, which makes it easier to analyze the generator's performance in the Vivado HLS environment. The changes to be made in the code involve the implementation of some arithmetic operators. Although Vivado HLS supports operations on all basic types of integers including the remainder operator with modulo division. Unfortunately this is just the operation, offering quite a low performance. For example, in GM generators, modulo division by parameter g is implemented as function modullint_by_gm31_g. The use of multiplication and bit shift significantly improves performance, and ullint is the redefinition of the unsigned long long int type name. In the GM and MT [6] generators for loops use unroll loop optimization technology contribute to increased performance, for example for the GM19 generator, unroll loop expansion can be done manually or use the #pragma HLS UNROLL expression.

3. Results

Although individual generators have different requirements for FPGA hardware resources, the implementation of a single generator requires only about 0.2 W of power. In case of CMWC4096 generator the large consumption of LUT is dictated that some of these resources will play the random-access memory role. In addition, a similar test was performed using the 2.4 GHz Intel Core i7-3517U mobile processor. Achieved times were about 2.5 times longer in the case of the GM family, while in the case of the other 1.5 to 2 times. For the GM61, the basic version needed as much as 6786 clock cycles. Performance was improved with the unroll of the for loop, the quantity of clock timings dropped to 3329. The biggest gain from optimization was obtained by replacing the instruction of modulo operation with its own realization. The result was a result of 145 clock cycles. Two modifications made the acceleration 46 times. During the process of generating numbers the consumption of each generator was 0.242 W. Although they are simpler in implementation and have lower periods, they have achieved very high productivity of 100 million and even 300 million PRNs per second. What is the effect of manual optimization of code implementing PRNG created directly in VHDL language.

Tab. 1. The performance of a single generator described by the number of clock (termed clks per num.) necessary to determine the next pseudorandom number and the quantity of random numbers generated per one second (termed PRNs/sec.) assuming that we will use a 200 MHz clock.

Generator	GM19	GM31	GM61	MRG32k3a	MT19937	LFSR113	XORSHIFT	MWC256	CMWC4096
Clks / num.	65	160	145	99	9242	1	6	8	6
PRNs/sec.	3,076,923	1,250,000	1,379,310	2,020,202	13,497,728	200,000,000	33,333,333	25,000,000	33,333,333

4. Conclusions and future work

Performance depends on the type of generator, but the ease of implementation of multiple generators allows you to scale performance to your needs. It should also be emphasized that for the LFSR113 or XORSHIFT generators, the best results were obtained in terms of performance that could be achieved by manual coding in VHDL. They are as efficient as their CPU counterparts, with significantly lower clock speeds and the amount of power they need [7]. Further development of implemented generators is naturally possible. Using GPU solutions [8], you can provide varieties with streams to facilitate the use of the generator in a parallel environment. The achieved energy efficiency of the system is very important in the era of green computing.

References

1. A.M. Ferrenberg, D.P. Landau, Y.J. Wong: Monte Carlo simulations: Hidden errors from "good" random number generators, Phys. Rev. Lett. 69 3382 (1992),
2. J. Wiśniewska, M. Sawerwain, W. Leoński: High performance computing and quantum trajectory method in CPU and GPU systems. Journal of Physics: Conference Series, 574(1) 012127 (2015),
3. P. L'Ecuyer: Good parameters and implementations for combined multiple recursive random number generators. Operations Research, 47(1) 159-164 (1999),
4. M. Pietron, P. Russek, K. Wiatr: Loop profiling tool for HPC code inspection as an efficient method of FPGA based acceleration, Int. J. Appl. Math. Comput. Sci., 2010, 20(3) 581-589 (2010),
5. A. Dąbrowska-Boruch, G. Gancarczyk., K. Wiatr: Implementation of a RANLUX based pseudo-random number generator in FPGA using VHDL and Impulse C, Computing and Informatics, Slovak Academy of Sciences, 32(6) 1272-1292 (2013),
6. M. Matsumoto, T. Tishimura, M. Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator, ACM Trans. on Mod. and Comp. Simul. 8(1) 330 (1998),
7. J.Y. Lee, G.D. Peterson, R.J. Hinde, R.J. Harrison: HASPRNG: Hardware Accelerated Scalable Parallel Random Number Generators. Comp. Phys. Comm., 180(12) 2574-2581 (2009),
8. L.Yu. Barash, L.N. Shchur: PRAND: GPU accelerated parallel random number generation library: Using most reliable algorithms and applying parallelism of modern GPUs and CPUs, Comp. Phys. Comm., 185(4) 1343-1353 (2014).

Parallel Independent Component Analysis Algorithm - Performance Comparison for EEG Signal

Anna Gajos[1], Grzegorz M. Wójcik[1], Przemysław Stpiczyński[2]

[1] Department of Neuroinformatics, Institute of Computer Science,
Maria Curie-Sklodowska University, Akademicka 9, 20-033 Lublin, Poland
[2] Institute of Mathematics, Maria Curie-Sklodowska University,
Plac Marii Curie-Sklodowskiej 1, 20-031 Lublin, Poland

e-mails: {agajos, gmwojcik}@umcs.pl, przemyslaw.stpiczynski@umcs.lublin.pl

Keywords: independent component analysis, parallel computing, EEG signal

1. Introduction

Independent Component Analysis (ICA) is iterative algorithm with unknown number of steps. It is time-consuming but it is also the most often used method for cleaning EEG signal from artifacts. The algorithm is designed to separate independent sources of signals from each other [1, 2, 3]. Most EEG manufacturers do not provide ICA implementations in their tools. An example is the Electric Geodesic EEG System with 256-channelled caps. The included Net Station program allows to export data to another format and external processing, but it is not possible to return. For this reason, we have developed our own software that allows to decode data, apply ICA and return to the NetStation [4]. We also decided to use the capabilities of the new generation of Intel processors to speed it up and then compare performance for the two selected architectures.

2. 2. Implementation and data setup

Our ICA implementation was based on the fastICA version available from it++ library. It is, however, written in C language. Unlike it++, it does not use the reduction of matrix dimensions and it works on all available data.

Appropriate data arrangement prevents so-called cache misses. In addition, parts of the code that used the whole signal were included in parallel blocks. We also used array notation and built-in function from Intel Cilk Plus C and C++ extensions to provide effective code vectorization.

We prepared 4 sets of data, containing 1, 10, 100 and 1000 seconds, each containing a record of 1000 samples per second.

All tests were performed on two computers:

1. Intel Xeon X5650 2.67GHz - 12 cores,
2. Intel Xeon L5640 2.26GHz - 12 cores (architecture provided with PLGrid).

For each dataset, we performed a test using 12 threads. We also did a test on a 1 thread on a Xeon X5650.

3. Results

The table shows the execution time of the application for all data configurations. By using parallel computation, the algorithm became more efficient. Although the same number of threads were used there is a noticeable difference in performance between different architectures.

Tab. 1. Results for Xeon X5650 (1 and 12 cores) and L5640 (12 cores).

Data set	Xeon X5650 (1)	Xeon X5650 (12)	Xeon L5640 (12)
1 s	7.996	3.348	7.082
10 s	23.56	7.351	19.766
100 s	238.448	68.554	165.611
1000 s	2383.238	647.969	1520.755

4. Summary

Parallelisation of algorithm and use of available architectures improved efficiency [5]. In addition, it is clear that performance depends on the architecture that is being used and the newer architecture generates much more benefits.

In the future, we plan to test the performance of the algorithm on Xeon Phi processors as well as on other architectures offered by PLGrid.

References

1. Glen D Brown, Satoshi Yamada, and Terrence J Sejnowski. Independent component analysis at the neural cocktail party. Trends in neurosciences, 24(1):54–63, 2001,
2. Arnaud Delorme, Terrence Sejnowski, and Scott Makeig. Enhanced detection of artifacts in eeg data using higher-order statistics and independent component analysis. Neuroimage, 34(4):1443–1449, 2007,
3. Anna Gajos and Grzegorz M Wójcik. Electroencephalographic detection of synesthesia. In Annales UMCS, Informatica, volume 14, pages 43–52, 2014,
4. Anna Gajos and Grzegorz M Wo ́jcik. Independent component analysis of eeg data for egi system. Bio-Algorithms and Med-Systems, 12(2):67–72, 2016,
5. A. Supalov, A. Semin, M. Klemm, and Ch. Dahnken. Optimizing HPC Applications with Intel Cluster Tools. Apress, Berkely, CA, USA, 2014.

High Level Framework for Mapping Deep Learning Neural Models to FPGAs

Michał Karwatowski[1,2], Maciej Wielgosz[1,2], Marcin Pietroń[1,2], Kazimierz Wiatr[1,2]

[1] AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Kraków, Poland
[2] ACC Cyfronet AGH, ul. Nawojki 11, 30-950 Kraków, Poland

e-mails:{mkarwat, wielgosz, pietron, wiatr}@agh.edu.pl

Keywords: artificial intelligence, deep learning, FPGA, natural language processing

1. Introduction

Deep learning(DL) models have become a powerful and efficient algorithms in a wide variety of applications. They are superseding traditional handmade solutions, as various DL algorithms are capable of extracting complex and disguised features from the data. One of the fields that greatly benefits from the progress in artificial intelligence(AI) is natural language processing (NLP) [1]. Thanks to information revolution, the amount of text created and read today is growing exponentially. Proper classification, searching and analysis of written data exceeded human capabilities. Throughout years of development, a selection of NLP algorithms facilitated this task, with neural networks on the frontier today.

AI algorithms can be very computationally demanding and many datacenters dedicate substantial resources for this task. General Purpose Graphic Processing Units (GPGPUs) usually provide the highest performance. However, efficient power usage is an important criteria and many organizations turned to Field-Programmable Gate Arrays (FPGA) as they often offer the highest performance to power ratio without the inflexibility of Application-Specific Integrated Circuits (ASIC). Designing efficient FPGA accelerator can be a challenging task, that requires in-depth knowledge about both hardware and software. Many AI researchers does not possess this combination of skills, therefore they are reaching for ready solutions to speed up their applications. There are a few frameworks that allow for fast and easy execution of DL algorithms on GPGPUs. However, GPGPU is not the best solution for all the applications. In datacenter applications GPGPUs tend to achieve higher throughput than FPGA, but on the expense of higher power requirements. When compared regarding throughput per power usage the results are not conclusive [2].

In this paper we are presenting out approach to neural networks implementation on FPGAs. The goal is to create a user friendly procedure that will allow for FPGA usage without in-depth hardware knowledge, and therefore open FPGA realm for data scientists.

2. Description of framework architecture

One of the most important goals of our approach is to facilitate the use of hardware. Therefore we chose well known and widely used API as a first user interface – Keras. It is a high level API that can be easily used to create complex DL models. Creation of a simple networks require only a few lines of code. Important feature for our approach is the possibility to export designed network along with trained weights as json and HDF5 files. Our platform of choice is currently Xilinx Virtex7 XC7VX485T, but the same solution can be easily ported to other Xilinx devices. For hardware implementation we are using Xilinx Vivado tool, additionally to make the process more flexible, for network IP generation we are using Vivado HLS. That allowed us to create a set of function templates in C++ that are used to create a specific DL networks. Data transfer between FPGA and CPU over PCIe is performed using Xillybus driver. The process of transforming Keras model to an IP core that can be used in FPGA is presented in the figure 1.

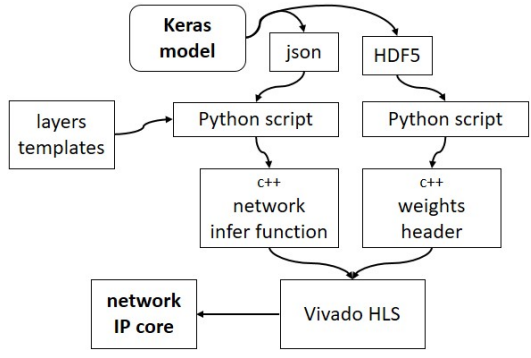


Fig. 1. Processing keras model to FPGA IP core.

In the first step user creates a network model in Keras, when accuracy is satisfying network model and weights need to be exported. Weights from HDF5 file are processed by python script to C++ header format. It is important that their values will be known during compilation as Vivado HLS will not be able to create hardcoded network otherwise. A separate header is created for each layer. Json file is used to create infer function of neural network. This is a top level function that performs all the computations on input data. As FPGA is not bound to any standard data type, at this point user can decide how the data and weights should be represented. In many situations full single/double floating point precision is not required [3]. Vivado HLS support integer type of any width (also nonstandard, like 7,13) and fixed point of any integer and fraction width. To validate generation process simulation can be performed, also to make sure that selected precision is sufficient. If results are satisfactory Vivado HLS synthesizes networks IP core. Next, user need to insert the IP core into provided Vivado design and generate bitsream. Easy to use software interface is also provided and can be used as a verification tool or a base for specific application.

3. Conclusions and future work

Proposed framework will allow data scientists to use FPGA accelerators without prior hardware knowledge. They will be able to implement the same algorithm on different platforms and compare performance to choose the most suitable one.

Framework has been tested on a few small networks and provided correct results without additional user intervention. However, many building blocks require refinement. Tested networks were small comparing to state-of-the-art solutions. Layers templates cannot perform all functionalities, like padding. Not all of the C++ code generation for Vivado HLS is performed fully automatically. Although python scripts generate the code, the whole process is not yet automatic and each step need to be started manually.

Presented framework requires many improvements, however preliminary results are promising and show that the framework can be used without in-depth hardware knowledge.

References

1. M. M. Lopez, K. Jugal: Deep Learning applied to NLP. arXiv preprint arXiv:1703.03091 (2017),
2. E. Nurvitadhi, D. Sheffield, J. Sim, A. Mishra, G. Venkatesh, D. Marr: Accelerating Binarized Neural Networks: Comparison of FPGA, CPU, GPU, and ASIC. In Field-Programmable Technology (FPT), 2016 International Conference on (pp. 77-84). IEEE,
3. M. Karwatowski, M. Wielgosz, M. Pietron, M. Staruchowicz, K. Wiatr: Comparison of Semantic Vectors with Reduced Precision using the Cosine Similarity Measure,. Proceedings of the 2017 intelligent systems conference (IntelliSys), London, UK, ISBN (IEEE XPLORE): 978-1-5090-6435-9, s. 898-904.

Cropping Input Image can Lead to a Better Training of Convolutional Neural Networks

Konrad D. Zuchniak

AGH University of Science and Technology, Kraków,
Faculty of Computer Science, Electronics and Telecommunication, Department of Computer Science

e-mails: konrad.zuchniak@gmail.com, dzwinel@agh.edu.pl

Keywords: machine learning, convolutional neural network, preprocessing, image rocognition

1. Introduction

In the advent of Big Data era, information becomes a very precious and extremely important asset. By exploring and analyzing big data, we can extract knowledge that can be used in further development of our civilization. Data classification is a fundamentaly problem of data analysis and machine learning. Currently, deep learning network architectures, including convolutional neural nets (CNN) [1] are the top classifiers. Comparing to classical but still robust machine learning tools such as SVM, the neural networks show their dominantion for big multidimensional datasets. Otherwise, they suffer overfitting problem. The most popular method that counteracts overfitting of neural networks is, so called, *dropout* [2]. This is a simple procedure, which consists in deactivation of randomly selected neurons. Another method is the *batch normalization* [3], which reduces the difference between individual data batches. Herein, we present another trick based on images cropping and we demonstrate its influence on the performance of convolutional neural networks.

2. Proposed solution

CNNs are the most effective classifiers for the analysis of big repositories of images. Herein, we propose to crop CNN input images of MxN size to (M-A)x(N-B) resolution. On the one hand, such the modification of data results in loss of information but, on the other, this way we can significantly increase the size of input data thus avoiding overtraining. The change of objects location on the image is the following benefit of data cropping. In result, the classifier is not able to assign the location of a specific object (objects) what eliminates this redundant feature. This way, cropping should increase the classifier generalization ability.

3. Results

To show the influence of cropping on the CNN classifier accuracy, we compare it with the effect of dropout employed for a simple neural network consisting of a single hidden and fully-conected layer. We investigated the effect of dropout on classification accuracy for various number of neurons N in the hidden layer and sizes S of training data. In our tests we used a few popular testing datasets. In Table 1 we present typical N/S dependence on the classifier accuracy obtained for MNIST dataset [4] ($7x10^4$ samples, 28x28 grayscale images of handwritten digits). As shown in Table 1, dropout increases the accuracy only in the case of adequately large S. For small S and large N (i.e., more neurons should be deactivated to fit the NN architecture to a given data size) the effect of dropout is the strongest one. Artificial multiplication of learning data by cropping the input images, is an another (opposite) way to escape the overtraining. To demonstrate the effect of this procedure we train the CNN on CIFAR-10 data set [5] ($6x10^4$ samples, 32x32 colour images in 10 classes) both on the original data as well as on data augmented by cropped images. In Table 2 we collect the obtained

results of training in terms of its accuracy, for various sizes of cropped images (uncropped image is 32x32 pixels).

Tab. 1. Dropout influence on the classifier accuracy.

N	S		
	10^2	10^3	10^4
10	-12.71%	-5.05%	-3.44%
10^2	-0.02%	0.47%	-0.09%
10^3	4.15%	0.58%	0.27%

Tab. 2: Relationship between the size of image fragments and the resulting classification accuracy.

Side size (px)	Original (32)	28x28	24x24	20x20	16x16	12x12
Accuracy	68.50%	74.31%	75.94%	71.48%	63.34%	51.57%

We demonstrated that by including randomly cropped images of size 24x24 pixels to the input data, we can observe significant increase of classification accuracy. It exceeds 10%, i.e., is greater than the accuracy increase due to dropout obtained for similar in size and dimensionality MNIST dataset (see Table 1). In or research we used TensorFlow [6] library. All computations were performed on supercomputer Prometheus (ACK CYFRNET Centre).

4. Conclusions and the future work

We demonstrated that increasing CNN input data size by including also cropped images, can increase classification accuracy. In the nearest future we plan to follow two directions of research:

- Developing other cropping-like data augmentation methods and their generalization. For example, audio files trimming could be done by cutting down narrower time intervals and selectively choosing the frequency,
- Searching for new methods of artificial reproduction of data built from images.

Acknowledgments. This research is supported by the Polish National Center of Science (NCN) DEC-2013/09/B/ST6/01549 grant and, partly, by PLGrid Infrastructure project.

References

1. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems (2012) p.1097-1105,
2. Nitish Srivastava, Geo rey Hinton, Alex Krizhevsky, Ilya Sutskever and Ruslan Salakhutdinov. "Dropout: a simple way to prevent neural networks from overfitting." Journal of machine learning research 15.1 (2014) p.1929-1958,
3. Sergey Ioffe and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." International Conference on Machine Learning, PMLR 37 (2015) p.448-456,
4. MNIST dataset: `http://yann.lecun.com/exdb/mnist/`,
5. CIFAR-10 dataset: `https://www.cs.toronto.edu/~kriz/cifar.html`,
6. TensorFlow library:`https://www.tensorflow.org/`.

PCJ as a Tool for Massively Parallel Data Processing

Marek Nowicki[2], Łukasz Górski[1,2], Magdalena Ryczkowska[1,2], Piotr Bała[1]

[1] ICM, University of Warsaw, Poland
[2] WMiI, Nicolaus Copernicus University, Toruń, Poland

e-mails: {faramir,lgorski,gdama,bala}@icm.edu.pl

Keywords: BigData, Java, PCJ, parallel computing, PGAS

1. Introduction

In this report we present PCJ (Parallel Computing in Java) as a novel tool for scalable data processing in Java. PCJ library is Java library based on PGAS (Partitioned Global Address Space) programming paradigm and allows for easy and feasible development of computational applications including BigData processing.

The demand on increasingly faster data processing resulted in creating dedicated tools and algorithms. One of the most widely used is MapReduce [1] model together with open-source Apache Hadoop platform [2]. The big advantage of this tool is fault tolerance and ability to keep thousands of terabytes of data on distributed file system. However, gaining high performance in some sort of problems might be a huge challenge. Because of that Apache Spark has been developed [3]. It allows to keep data in-memory and speed up analysis.

2. PCJ Library

PCJ [4] is a library that allows developing applications in pure Java language. It does not require any language extensions or special compiler. The user has to download single jar file and can develop and run parallel applications on any system with Java installed. PCJ uses PGAS (Partitioned Global Address Space) programming model, with all communication details like threads administration or network programming hidden. The communication in the model is one-sided and asynchronous. PCJ library provides necessary tools for threads numbering, data broadcast and threads synchronization. All those features make programming simpler together with high performance preserved.

The PCJ applications can run on the traditional HPC systems such as x86 clusters, multicore PCs and other systems including recent Intel KNL processors. The applications implemented with PCJ and Java scale up to hundreds thousands cores. A good example is 2D stencil code running with 196k cores of Cray XC40 at HLRS.

3. BigData processing with PCJ

The PCJ library has been compared with Apache Hadoop. The performance results confirm that applications implementation based on the PCJ library are much faster (5-500 times depending on the problem) than Hadoop, even for typical map-reduce applications such as counting of words in the file.

There are preliminary results for executing application written in the PCJ library in Apache Spark ecosystem. The PCJ based implementation of the π evaluation is more than 3 times faster. One should note that application developed with PCJ has been run on the Hadoop cluster as Spark application using Hadoop task management. Even with such setup, performance of PCJ was significantly higher.

The PCJ library has been used for parallelization of the DNA sequence search within large database containing more than 20 millions of records (52GB file) which is the key

element of the processing NGS results. The parallelization is based on the work distribution based on the partitioned of the input sequence and processing using NCBI-BLAST. The load balancing has been ensured by monitoring the execution of BLAST instances. It was demonstrated that this design allow the application to scale almost linearly (more than 90% efficiency for 32 nodes) up to 1536 cores of the HPC cluster. The performance results for Cray XC40 are similar and present at least 90% parallel efficiency for 32 nodes and 75% parallel efficiency at 128 nodes (6144 cores) [6].

4. Conclusions

PCJ library is highly scalable, easy to use tool for development of parallel applications, including BigData processing. The performance of applications implemented with PCJ is higher than for traditional tools used for data processing such as Hadoop or Spark. Moreover, the development with PCJ is much easier than in the case of other tools. It requires less libraries to use, and minimizes number of language constructs used. The resulting code is usually shorter and more readable.

PCJ applications can be developed and tested using standard Java environment, the time consuming installation of the infrastructure tools such as Hadoop is not required.

Compare to other tools, PCJ library has no fault tolerance mechanism, however, experimental version exists and will be integrated with the main release soon.

Acknowledgments. This work has been performed using the PL-Grid infrastructure. Partial support from CHIST-ERA consortium is acknowledged through NCN grant 2014/14/Z/ST6/00007. MN acknowledges EuroLab-4-HPC cross-site collaboration grant and PRACE for awarding access to resource HazelHen at HLRS (Stuttgart, Germany).

References

1. Dean, S. Ghemawat: MapReduce: simplified data processing on large clusters. *Communications of the ACM*, vol. 51 no. 1 pp. 107-113 (2008),
2. Apache Hadoop. http://hadoop.apache.org/. Accessed: 22 Sept. 2017,
3. Apache Spark. http://spark.apache.org/. Accessed: 22 Sept. 2017,
4. M. Nowicki, P. Bała. Parallel computations in Java with PCJ library In: W. W. Smari and V. Zeljkovic (Eds.) *2012 International Conference on High Performance Computing and Simulation (HPCS)*, IEEE 2012 pp. 381-387,
5. `http://pcj.icm.edu.pl` Accessed: 22 Sep 2017,
6. M. Nowicki, D. Bzhalava, P. Bała Massively Parallel Sequence Alignment with BLAST Through Work Distribution Implemented Using PCJ Library In: S. Ibrahim, Kim-Kwang R. Choo, Z. Yan, W. Pedrycz (Eds.) *Algorithms and Architectures for Parallel Processing. ICA3PP 2017. Lecture Notes in Computer Science, vol 10393*. Springer, Cham, 2017, pp. 503-512.

Anomaly Detection Service for Financial Data Streams

Paweł Gławiński[1], Marek Wojciechowski[2], Maciej Zakrzewicz[2]

[1] Softman SA, Piaseczno, Poland
[2] Poznan University of Technology, Poznan, Poland

e-mails: sprzedaz@softman.pl, {marek,mzakrz}@cs.put.poznan.pl

Keywords: anomaly detection, SOA, infrastructural service

1. Introduction

Anomaly detection [1] in versatile financial data streams is a vital business problem. Typical anomalies include credit card frauds, purchase card frauds, financial reporting fraud, insurance fraud, fraudulent claims for health care, credit applications fraud, credit transactional fraud, etc. Existing IT solutions for anomaly detection are typically implemented in the form of domain-specific, isolated built-in functions which cannot be easily shared nor expanded. As financial IT systems are usually distributed and need to be flexible and evolve over time, the cloud-based SOA approach is getting more and more popular. In SOA, software modules can be reused to reduce development time and common functionalities can be easily shared.

We work on theoretical and design frameworks to provide anomaly detection functions as SOA services. We argue that several requirements need to be addressed by such shared services: they have to concurrently handle multiple high-volume data streams (using asynchrony, caching, intermediate result materialization, background learning), deal with data model heterogeneity across the streams, support correlating independent streams, handle both expert-defined and machine-learned detection rules.

We successfully developed and validated a universal cloud-based SOA infrastructural service to seamlessly integrate anomaly detection rules (based on JBoss Drools and Weka) into SOA business systems[1]. While incorporation of business rule engines [2] and data mining algorithms [3][4] into SOA solutions has been previously considered in the literature, in this paper we overview some of the architectural and design concepts required to address the aforementioned specific requirements of an Anomaly Detection Service.

2. Description of the problem solution

An overview of our Anomaly Detection SOA Service is shown in Fig. 1. Business objects are delivered to SOAP Web Service interfaces as XML documents. JAXB converts the XML documents into Java objects, which are sent to a throttling JMS queue. The objects in the queue are periodically propagated by the Controller to the embedded JBoss Drools rule engine, which then executes business anomaly detection rules on the received objects. Anomaly detection rules are designed by a business user using a visual rule editor (part of our solution). The rules can be based on the expert's knowledge or rely on statistical models obtained through machine learning using Weka. When anomalies are detected, new business objects are created and delivered to an output JMS queue. The Dispatcher delivers them to external consumers (SOAP Web Services). A database repository (MongoDB) is used as a persistence store to protect the state of the anomaly detection service in case of failures.

The system handles four types of rules: (1) simple rules based on the current business object only; (2) aggregation rules based on moving window aggregates calculated from collections of business objects; (3) calendar rules based on schedules; (4) learning rules based

[1] Supported by POIG.01.04.00-14-061/12

on models learnt from business objects received recently. For the rules of type 2 we had to extend the out-of-the box Drools engine with aggregate materialization functionality to achieve satisfactory performance of the system on large amounts of data. For the rule of type 4, due to immaturity of data mining libraries dedicated to data streams, we integrated the Weka data mining library to build prediction models (J48 decision trees and Naïve Bayes) and implemented a solution for keeping the prediction models up to date. The general idea was to build a new prediction model in the background while scoring the incoming data with a previously built model and replacing the old model with the new one when ready.

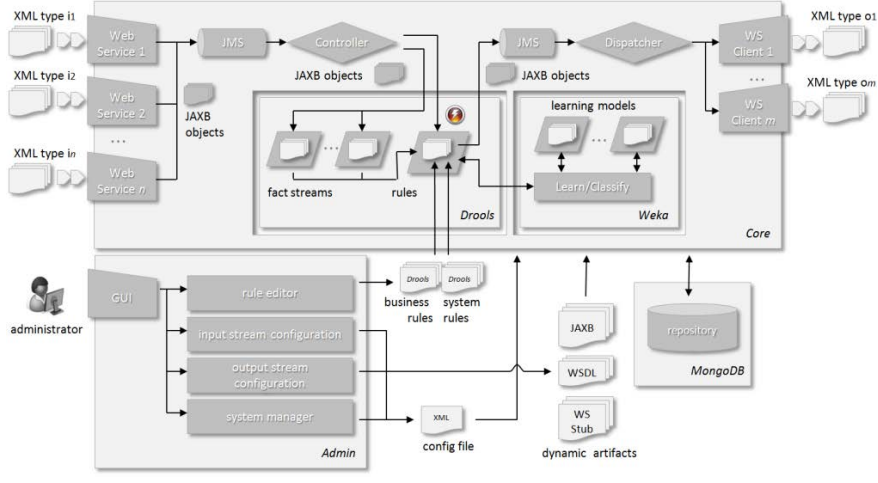


Fig. 1. System architecture.

3. Results

To validate the concept of aggregate materialization, we have performed a series of computational experiments on data mimicking the expected real-life workload and an anomaly detection rule involving a moving average of invoice totals. With the aggregate materialization turned on, the system stabilized on constant, satisfactory throughput with the increasing number of daily invoices per customer per day. Without materialization, performance degraded quickly to the level making real-time processing infeasible.

4. Conclusions

We have presented the architecture of an Anomaly Detection SOA Service targeted at financial applications. Its key features are: asynchronous processing, four types of anomaly detection rules, aggregate materialization, and off-line learning of discovery-based business rules. Our prototype system has been validated in a real-life environment.

References

1. M. Agyemang, K. Barker, and R. Alhajj: A comprehensive survey of numeric and symbolic outlier mining techniques. Intelligent Data Analysis 10(6), 2006, pp. 521-538,
2. B. Hakizumwami: Building Enterprise Services with Drools Rule Engine, OnJava, 2007,
3. A.A.A. Esmin, D.A. Pereira, M.R. Pereira, and D.L. Araújo: SMINER - a platform for data mining based on service-oriented architecture. IJBIDM 8(1), 2013, pp. 1-18,
4. L. Wang, Q. Wang, and L. Ni: The Analysis and Design of SOA-Based Financial Data Mining System. In Proc. of IHMSC, 2011.

Concept of Decentralized Access Control for Open Network of Autonomous Data Providers

Łukasz Opioła[1,2], Michał Wrzeszcz[1], Łukasz Dutka[1], Renata G. Słota[2], Jacek Kitowski[1,2]

[1]AGH University of Science and Technology, Academic Computer Centre Cyfronet AGH, Kraków Poland
[2]AGH University of Science and Technology, Faculty of Computer Science, Electronics and Telecommunications, Department of Computer Science, Kraków, Poland

e-mails: {lopiola,wrzeszcz,rena,kito}@agh.edu.pl, lukasz.dutka@cyfronet.pl

Keywords: distributed systems, decentralized systems, access control, data synchronization

1. Introduction

Thanks to constant technological advances, numerous distributed systems are created that push towards globalization of resources (knowledge, data, etc.) and services. This trend is especially visible in data access systems and eScience [1] methodology, saying that scientists from different institutions should use their collective resources to tackle big research problems. However, while ready for cooperation, the data providers typically want to regain their autonomy. For that reason, an open network of cooperating providers must be created.

In this paper, we focus on two crucial aspects of such network: authorization and authentication infrastructure (AAI) and synchronization protocol (SP). AAI is required for data providers to verify the requests that they receive (i.e. access control). It must be decentralized and support authority delegation as data providers are independent and do not trust each other. By SP we mean a protocol for exchanging metadata of users, resources, infrastructure etc, that is required for cooperation. As the network of providers is open, it is cumbersome to maintain consistency of data and avoid conflicts. The aspects of AAI and SP are strongly interconnected, as AAI requires metadata of users, resources and infrastructure, while SP must be coupled with security mechanisms that ensure safety of metadata. The metadata is vastly scattered and sparsely replicated as every provider synchronizes only the required fragments.

The desired features of AAI and SP subsystem can be summarized as: open, decentralized, ensuring data privacy, scalable and with peer-to-peer discovery on demand.

There are some AAI candidates for the backbone for such system. SAML and X.509 infrastructure, while very mature, are often over-complicated and high-maintenance. Macaroons are lightweight bearer tokens offering decentralized authorization and delegation with contextual confinements. When SP is considered, block-chain is a popular choice among peer-to-peer frameworks, but it cannot be directly used to store sensitive data and is computationally expensive. Distributed Hash Tables (DHT) constitute another class, but they cannot ensure safety and consistency of data. In general, most synchronization protocols assume that peers belong to the same deployment or organization / subject.

Onedata [2] is a globally distributed data management platform that tries to provide unified, borderless data access to files stored in autonomous data centers. This paper outlines a conceptual solution for AAI and SP that is being implemented and evaluated in that system.

2. Description of a problem solution

Our proposed concept is based on a simple, yet powerful assumption - every user in the system is a carrier of trust. It means that he is able to freely choose data providers, and by asking for their services, he puts his trust in their authenticity and good intentions. For that reasons, users should be educated and aware of the consequences following their decisions. The process of establishing cooperation starts with an environment that has a couple of

completely autonomous data providers, each having its own users. One of the users is invited by his colleague from another provider to a shared data set. The providers exchange metadata of user and the data set using SP. AAI ensures that providers can safely release the required information. The providers remain oblivious to users and resources managed by other providers. Users put their trust in providers, and providers can always verify authenticity and authorization of requests. This way, a safe cooperation of independent providers can be achieved.

To realize that model in Onedata, we propose the idea of zones. Every zone is a service that oversees and mediates in cooperation of a group of providers. It's essentially an AAI center and SP server. Providers choose a zone that they trust and rely on it in context of request authorization and data synchronization. Each zone manages a certain subset of user and resource data accumulated from underlying providers. This ensures easy and safe cooperation across single zone. Users can reach across different zones by asking for services or resources of providers subject to other zones. In such case, the zones communicate with each other every time an authorization check must be performed or during data synchronization. Every piece of information is managed by its authoritative zone, i.e. where it originated.

In our model, the aforementioned use-case looks as follows (see Fig. 1). User A, operating in provider A (subject to zone A) generates an invite token, which contains information about its origin (zone A hostname), and passes it to user B. user B contacts his zone (B) with the token. Zone B retrieves the information about token's origin and hence discovers zone A. Zone B contacts zone A, which verifies token's authenticity and accepts user B as a member of data set. Zones synchronize required data and present it to providers, which can now cooperate in maintaining the data set.

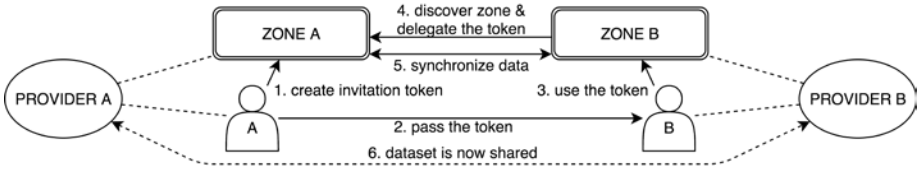


Fig. 1. Cooperation in data access of autonomous providers in *Onedata*.

3. Results

The proposed solution is being implemented in *Onedata*. Currently, cooperation within a single zone is supported, with AAI based on macaroons and SP based on publish-subscribe model. During the presentation, the concept and implementation details will be discussed.

4. Conclusions and future work

There is a demand for global data access systems that unite autonomous data providers. Onedata tries to achieve it using a novel concept of decentralized data synchronization and access control model. Future plans include introduction of data redundancy to increase failure resistance and possibility to transfer information ownership between different zones.

Acknowledgements: ŁO is grateful for AGH-UST grant no. 15.11.230.321, RS and JK are grateful for AGH-UST grant no. 11.11.230.337.

References

1. P. Hinrich, P. Grosso, and I. S. Monga. "Collaborative Research Using eScience Infrastructure and High Speed Networks." Future Generation Computer Systems 45.C (2015): 161,
2. Wrzeszcz, M., Opioła, Ł., et. al.. Effective and Scalable Data Access Control in Onedata Large Scale Distributed Virtual File System. Procedia Computer Science (2017), 108, 445-454.

Actor-Based Tensor Network Simulation

Bartosz Błaszków[1], Mateusz Front[1], Katarzyna Rycerz[1], Piotr Gawron[2]

[1] AGH University of Science and Technology, Institute of Computer Science AGH,
Department of Computer Science, al. Mickiewicza 30, 30-059 Kraków, Poland
[2] Polish Academy of Sciences, Institute of Theoretical and Applied Informatics,
Quantum Systems of Informatics Group, Bałtycka 5, 44-100 Gliwice, Poland

e-mails: blaszkow@student.agh.edu.pl, 1683ab@gmail.com, kzajac@agh.edu.pl,
gawron@iitis.pl

Keywords: tensor networks, contraction, actor model

1. Introduction

In this paper we describe the Tensia environment supporting analysis and operations on tensor networks [1]. The main goal of Tensia is to determine an optimal order of operations (called contractions) between tensors in such a network performed in parallel and possibly distributed environment. The actual calculation engine is based on the actor model [2] that supports parallel execution.

2. Tensia architecture and functionality

Tensors from tensor network may be contracted in any order, but it affects the computational complexity of the process. There are existing algorithms that given a tensor network can provide an optimal contraction order, but none of them takes into consideration that contractions may be done in parallel. That's why we designed and developed the algorithm for determining an optimal operations order basing on contraction cost estimation. Since the problem is NP-hard, an implementation puts a strong emphasis on optimization.

With the optimal calculation order figured out, we can create a tree data structure (which we call computation tree), where leafs are the tensors from original network and each node represents the result of contracting its children (see Fig 1.).

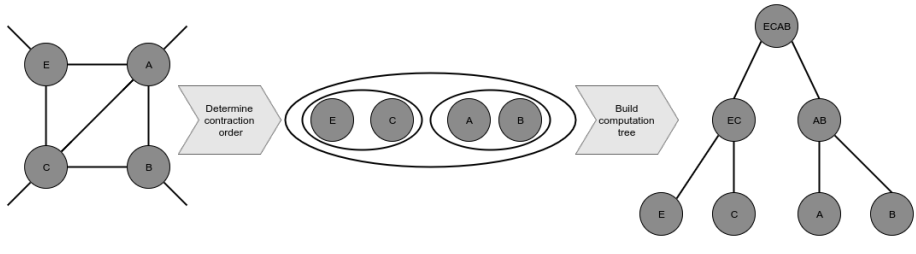


Fig. 1. An example of how tensor network can be transformed into a computation tree by Tensia.

In Tensia the computation tree is built from top to bottom using actors. Each one represents a node from a tree and spawns children responsible for providing tensors that will be contracted. Then the contraction phase begins in a bottom-up manner. Each node, after it receives the tensors to operate on, performs the actual computation utilizing low-level BLAS implementation [3] and passes the result 'up'.

3. Implementation and results

The Tensia framework was implemented using Akka's Actor System [4]. Additionaly, ND4J library [5] was used for integration with BLAS implementation. The contraction order algorithm was implemented in C and integrated with the whole system by Java Native Interface. The preliminary results showed that the algorithm determining contraction order for parallel computation is a promising solution that can be used by the Akka-based framework managing parallel operations on tensor networks.

4. Conclusions and future work

We plan to use the results from the presented work for application of tensor network formalism in different fields of science, in particular simulation of continuous-time stochastic automata networks [6], quantum walks in image segmentation [7] and hyperspectral image analysis [8].

Acknowledgements. The research presented in this paper has been partially supported by NCN grant number 2014/15/B/ST6/05204

References

1. R. Orus, A Practical Introduction to Tensor Networks: Matrix Product States and Projected Entangled Pair States, Annals of Physics 349 (2014) 117-158, arXiv:1306.2164,
2. C. Hewitt, Actor Model of Computation: Scalable Robust Information Systems, arXiv:1008.1459,
3. Home page of reference BLAS `http://www.netlib.org/blas/`,
4. Documentation for Akka Actor Systems: `https://doc.akka.io/docs/akka/snapshot/scala/general/actor-systems.html`,
5. ND4J project website: `https://nd4j.org`,
6. B. Plateau, W. J. Stewart. Stochastic Automata Networks Computational Probability Kluwer Academic Press, pp. 113--152, 1997,
7. Grady, Leo. 2006. „Random walks for image segmentation". IEEE transactions on pattern analysis and machine intelligence 28 (11): 1768–1783,
8. Biamonte, Jacob, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, i Seth Lloyd. 2016. „Quantum Machine Learning".

Particle Automata Model of Heterogeneous Melanoma Progression

Marta Panuszewska

AGH University of Science and Technology, Kraków, Poland

e-mail: panuszew@student.agh.edu.pl

Keywords: melanoma, modeling, particle automata, heterogeneous cancer

1. Introduction

The incidence of invasive melanoma and its mortality rate is rapidly rising since at least 1975. Due to depleting ozone levels in the atmosphere, the natural Earth's protection against UV radiation is diminished, leading to constant rise in the number of diagnosed patients. Computer modeling is the most appropriate way for understanding melanoma growth dynamics. However, the tumor and surrounding skin is a complex, multiscale system, with many coupled microscopic and macroscopic factors that have to be accounted for. Therefore, it is impossible to fully simulate and control melanoma growth scenario. Moreover, increasing number of model parameters results in its overfitting what decreases or completely disables its predictive power. On the other hand, the numerical cancer model allows for investigating and identification of the most crucial tumor growth factors and possible scenarios of its proliferation. Herein we describe and evaluate Particle Automata model (PAM) [1] of melanoma in scrutinizing genetic and epigenetic melanoma heterogeneity, which is a key component in cancer progression and drug resistance, as it provides population diversity and tumor robustness [2].

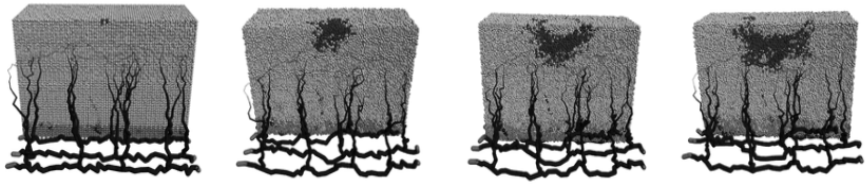
2. Description of a problem solution

Our modelling approach is based on the Particle Automata paradigm [1], in which behavior of the full network of interacting objects is determined by every object, which behaves like a finite-state automation. The states of a single object is determined by its neighbors, internal mitosis cycle, the oxygen and TAF (tumor angiogenic factors) concentration fields, which evolve in time. In the skin model we distinguish two main agents: blood vessels sections and skin cells. The blood vessels sections are tube-like shapes, connected to each other forming complex vascular structure (see Fig. 2). The skin cells are represented by spheres, organized in layers (see Fig.2) according to the tissue they represent: hypodermis (bright yellow), dermis (dark yellow) and epidermis (red) [3]. As shown in Fig.2, among dermal cells there are shown also brown cancer cells. After setup of initial conditions, simulation proceeds in discrete time steps. In each time step multiple calculations are performed. Firstly, we calculate blood flow through the blood vessels. All the vessels that contain blood deliver oxygen, which diffuses through the tissue. Next, the forces between cells and vessels are being calculated. We assume that, the cells and vessels repel with each other if they are too close and attract otherwise. Next, the total forces acting on every particle are calculated and their updated positions are computed by integrating numerically the Newtonian laws of motion. We introduce heterogeneity model assuming that in each timestep every cancer cell has a small chance to mutate and produce more voracious clone. Once that occurs, there are no more mutations until the end of simulation.

3. Results

All the simulations were divided into two main groups: homogeneous and heterogeneous. Simulations were run on very similar (mostly identical) skin tissue models, but with different

parameters of cancer tissues. All the simulations were starting with initial setup with around 20 melanoma cells in the middle (see Fig. 1a) and were ended in one of three following cases: either the tissue reached homeostasis or the tumor has taken more than 2/3 of available horizontal of vertical space (see Fig. 1d). On the diagrams below (see Fig.2) there are presented differences between homogeneous and heterogeneous melanoma growth. While on Fig.1 both types of heterogeneous melanoma are alive and coexisting, on Fig.2 mutated cells were more adjusted to the environment and caused original cells to die.



(a) Step 0. (b) Step 60 000 (104 days). (c) Step 100 000 (174 days).(d) Step 120 000 (208 days).

Fig. 1. Heterogeneous tumor growth.

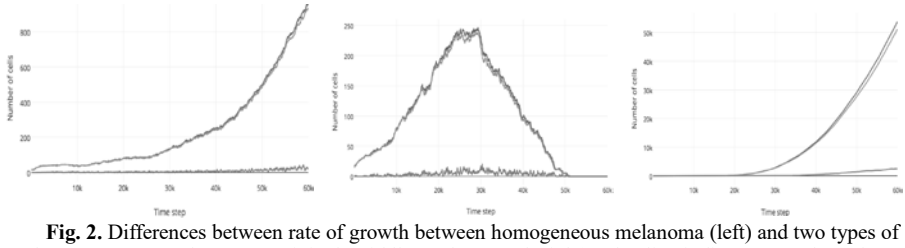


Fig. 2. Differences between rate of growth between homogeneous melanoma (left) and two types of heterogeneous melanoma – original (middle) and mutated (right). Blue line indicates a total number of cells, orange - cells alive and green - dead cells.

4. Conclusions and future work

The results show that it is possible to simulate heterogeneous melanoma in the environment described above. It was possible to obtain a stable tissue with heterogeneous melanoma growing in time. There are many possible directions of research that could still be taken in the subject of heterogeneous melanoma modeling with the use of PAM, such as: introducing various models of heterogeneity, and testing various drug therapy scenarios for selected drug resistance mechanisms.

Acknowledgements: The work has been supported by the Polish National Science Center (NCN) project 2013/10/M/ST6/00531 entitled: Multi-scale model of tumor dynamics as a key component of the system for optimal anti-cancer therapy.

References

1. Witold Dzwinel, Rafal Wcisło, David A Yuen, and Shea Miller. Pam: Particle automata in modeling of multiscale biological systems. ACM Transactions on Modeling and Computer Simulation (TOMACS), 26(3):20, 2016,
2. Henry HQ Heng, Steven W Bremer, Joshua B Stevens, Karen J Ye, Guo Liu, and Christine J Ye. Genetic and epigenetic heterogeneity in cancer: A genome-centric perspective. Journal of cellular physiology, 220(3):538–547, 2009,
3. Witold Dzwinel, Adrian Kłusek, Rafał Wcisło, Marta Panuszewska, and Paweł Topa. Continuous and discrete models of melanoma progression simulated in multi-gpu environment. In PPAM 2017: Proceedings of International Conference on Parallel Processing and Applied Mathematics, Lublin, 10-14 September 2017.

Container-Based Architecture for Resilient and Reproducible Scientific Workflows

Michał Orzechowski, Bartosz Baliś

AGH University of Science and Technology,
Faculty of Computer Science, Electronics and Telecommunications
Department of Computer Science, Kraków, Poland

e-mails: {morzech, balis}@agh.edu.pl

Keywords: application containers, scientific workflows, reproducibility, resilience

1. Introduction

Reproducibility and resilience are among key requirements for scientific workflows and both require workflow state persistence and recovery mechanisms [1]. In recent years, the rapid rise in popularity of container technologies had a significant impact on the simplicity of installation, deployment, and execution of scientific applications. While some workflow engines already support running containerized scientific workflows as well as running the workflow execution software itself from within the containers [2,3], the advantages of containers for workflow state persistence and recovery have not been fully investigated. We propose a container-based architecture for execution of scientific workflows, focusing on mechanisms for workflow state management. The proposed architecture reduces data footprint related to state transfer and storage, and facilitates complex workflow execution scenarios including reproducibility, fault tolerance, live migration, and smart re-runs.

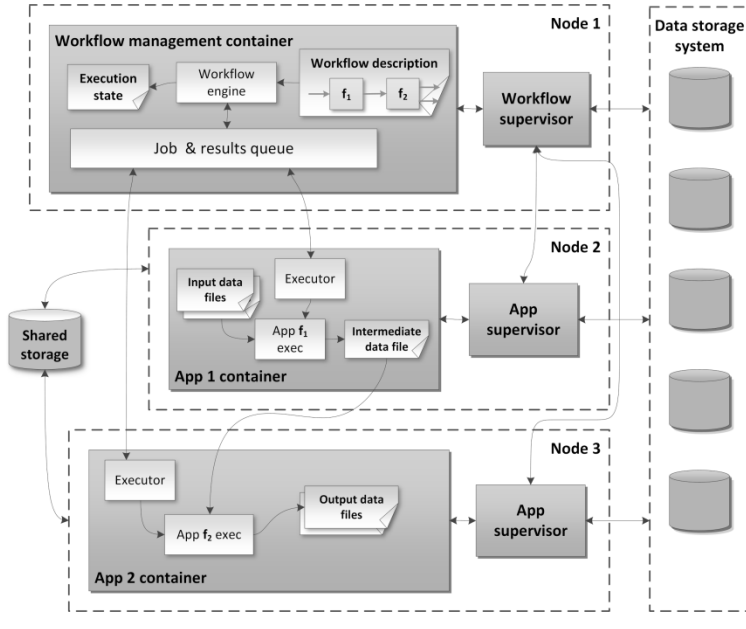


Fig. 1. Proposed architecture for resilient and reproducible scientific workflows.

2. Proposed architecture and results

In order to evaluate the advantages of workflow state management using containers, we propose a container-based architecture for distributed workflow execution (Fig. 1) that incorporates the sidecar container pattern, by adding supervisor containers to workflow management and application containers. The supervisors manage the state of components running within those containers, monitor them and apply fault-tolerance strategies in case of errors. The managed state incorporates the state of runtime environments, intermediate data, output data, and the internal state of the workflow execution engine. All these state data can be gathered without the need to tamper with the workflow engine or workflow code and execution. The only assumption is that the workflow engine has the capability to save and recover its internal execution state, e.g. using a file.

We have implemented the proposed architecture by containerizing the HyperFlow workflow management system [4] with Docker container technology and used the Kubernetes container orchestrator to run containerized sample workflows. The state management system is currently in the proof-of-concept phase, implemented mainly using Kubernetes container management internal mechanisms.

3. Conclusion

The proposed container-based architecture and mechanisms for workflow state management can be used to significantly simplify automation of complex workflow execution scenarios such as 1) *reproducibility*: upon successful completion of the workflow its full execution state (including the runtime environment) is captured and can be used to reproduce the same execution in the future; 2) *smart-rerun*: as the state of each execution step of the workflow is captured (including intermediate data), one can replace some components of the workflow and re-run it, in which case the workflow engine will either reuse already computed data, or repeat the execution if necessary; 3) *live migration*: if one decides to migrate the workflow to a different computing infrastructure, its persisted execution state can be used to restore the execution environment and resume the workflow after migration is complete; and 4) *fault tolerance*: upon failure caused by the infrastructure of software error, workflow execution can be resumed as soon as the cause of the failure is removed. The architecture described in this paper is a basis for a resilient execution environment for smart levee monitoring and decision support system in the ISMOP project (`www.ismop.edu.pl`) [5]. Future work will focus on running more complex workflows and detailed tests of each of the execution scenarios.

Acknowledgments. This work is partially supported by the ISMOP project (PBS1/B9/18/2013); and by the AGH University of Science and Technology Statutory Fund no. 11.11.230.337.

References

1. E. Deelman, et al. "Workflows and e-Science: An overview of workflow system features and capabilities." Future generation computer systems 25.5 (2009): 528-540,
2. J. Stubbs, et al. "Endofday: A Container Workflow Engine for Scalable, Reproducible Computation." Proc. 8th International Workshop on Science Gateways, 2016,
3. W. Gerlach, et al. "Skyport: container-based execution environment management for multi-cloud scientific workflows." Proc. 5th Int. Workshop on Data-Intensive Computing in the Clouds. IEEE Press, 2014,
4. B. Balis. Hyperflow: "A model of computation, programming approach and enactment engine for complex distributed workflows." Future Generation Computer Systems, 55:147-162, 2016,
5. B. Balis, T. Bartynski, M. Bubak, D. Harezlak, M. Kasztelnik, M. Malawski, P. Nowakowski, M. Pawlik, B. Wilk. "Smart levee monitoring and flood decision support system: reference architecture and urgent computing management." Procedia Comp. Science, 108:2220-2229, 2017.

Minimal Computational Models for Characterization of Heart Valve Interventions: Preliminary Evaluation of Model Personalization Process

Krzysztof Czechowicz[1] and D. Rod Hose[1,2]

[1] Mathematics and Modelling in Medicine, Dept. of Infection, Immunity and Cardiovascular Disease, The University of Sheffield, UK
[2] Insigneo Institute for in silico Medicine, Sheffield, UK

e-mail: k.czechowicz@sheffield.ac.uk

Keywords: 0D modelling, 3D modelling, optimisation, patient specific

1. Introduction

Valvular Heart Disease currently affects 2.5% of the population, but as it is a disease of the elderly it is on the rise in our society: the population beyond 85 is expected to double by 2028 [1] with a concomitant increase in the number of cases. The treatment for severe cases is heart valve replacement, and the timing and nature of the intervention is crucial to obtain the best outcome. The EurValve project aims to help clinicians to make the best decision by creating a physiology-based decision support system (DSS). At the core of the DSS is a minimal-complexity computational model that is able to represent key cardiac parameters under rest and exercise conditions, pre- and post-intervention. The motivation for adoption of the minimal model is that its parameters can be personalised using minimal clinical data that is collected in standard clinical pathways. It is also important for the anticipated clinical use scenario that both the model personalisation operation and model execution in a predictive capacity be rapid, taking no more than a few minutes, and robust.

2. Methods

The model on which EurValve is based is illustrated in Figure 1. At its core is a lumped-parameter model of the systemic circulation, consisting of the left ventricle, aortic valve, systemic circulation, left atrium and mitral valve. This minimal model, published by Shi to the open-access CellML repository, is a gross simplification of the model developed by Shi and Korakianitis [3]. There is renewed interest in the use of models of this type to characterise individual patient physiology [2], [4]. The ventricle and atrium are represented by variable elastance models and the systemic circulation by a 3 element Windkessel model. Even this minimal model contains 23 input parameters that might, in principle, be personalised to represent the individual. The valves are represented by polynomial characterisations of the pressure-flow relationships computed from 3D computational fluid dynamics (CFD) of the individual patient valves, segmented from medical image data. The image segmentation is performed using a service developed by Philips, and the CFD using ANSYS Fluent. An alternative and rapid valve characterisation using a reduced order model (ROM) is also offered in the EurValve work flow. Few of the model input parameters are directly available from clinical measurements. However several of the output parameters, for example integral flow measures (cardiac output) and systolic and diastolic systemic pressures and chamber volumes, are available. In principle at least a subset of the input parameters can be tuned to produce the observed outputs. The key parameters to tune have been identified based on a sensitivity analysis conducted by the project partners at the Technical University of Eindhoven, and the remaining parameters are based on published literature values. For the preliminary study reported in this paper, to demonstrate the process, just four parameters (left ventricular maximal elastance, distal resistance, capacitance, stressed blood volume) are personalised

using four measured parameters (mean and diastolic arterial pressure, systolic and diastolic left ventricular volume). Many optimisation processes are possible: for the analysis reported here the process uses a Genetic Algorithm with few generations to provide an initial estimate for a Nelder-Mead simplex operation (both operated in Matlab) to converge rapidly and robustly to an optimal solution. Separately, EurValve is examining the operation of an unscented Kalman filter approach for the cases which have time-resolved data available.

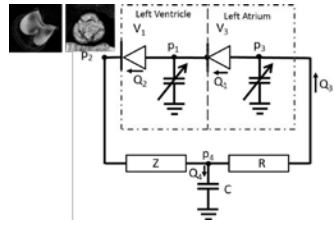


Fig. 1. Schematic of the 0D model. Z and R represent the proximal and distal resistance respectively, Q_i, p_i, V_i is the flow, pressure and volume respectively, where i is the location of the calculation. The valve coefficients (in this illustration for the aortic valve) are computed by 3D computational fluid dynamics.

3. Preliminary results: process stability

The stability of the method was tested on representative artificial data. Baseline values were taken from the literature and a population of 110 'digital patients' was generated using a normal distribution with a standard deviation of 25% and cut of at +- 50% of the initial values. 5 of the runs failed to converge during the optimisation step. Further inspection of the failed cases showed that the difference between the mean and diastolic pressures was below 0.1mmHg, meaning that those series were unphysiological and such a combination is not to be expected in real applications.

4. Conclusions and future work

It has been demonstrated that a very simple minimal lumped parameter model can be personalised to reproduce representative temporal distributions of pressure and flow in simulations including representations of aortic valve stenosis, the latter based on CFD simulations. The process has been automated and it is robust in operation. The next steps are to operate on real patient data and to compare the results with other optimisation methods, and to integrate Reduced Order Models as an alternative to the CFD process.

Acknowledgements. This work is supported by the EU project EurValve *Personalised Decision Support for Heart Valve Disease* H2020 PHC-30-2015 689617. This research was supported in part by the PL-Grid infrastructure.

References

1. J.L. d'Arcy, B.D. Prendergast, J.B. Chambers, S.G. Ray, B. Bridgewater: Valvular heart disease: the next cardiac epidemic. Heart, 2011, 97(2), pp. 91-93,
2. C.E. Hann, J.G. Chase, T. Desaive, C.B. Froissart, J. Revie, D. Stevenson, B. Lambermont, A. Ghuysen, P. Kolh, G.M. Shaw: Unique parameter identification for cardiac diagnosis in critical care using minimal data sets. Comput. Methods Programs Biomed. 2010, 99(1), 75-87,
3. Y. Shi, T. Korakianitis: Numerical simulation of cardiovascular dynamics with healthy and diseased heart valves, J. Biomech 2006, 39 (11), pp. 1964-82,
4. K. Sughimoto , F. Liang , Y. Takahara, K. Mogi, K. Yamazaki, S. Takagi, H. Liu: Assessment of cardiovascular function by combining clinical data with a computational model of the cardiovascular system. J Thorac Cardiovasc Surg. 2013, 145(5), pp. 1367-72.

Influence of Similarity Measures in Case-Based Reasoning for the Treatment of Valvular Heart Disease

Hélène Feuillâtre[1,2], Vincent Auffret[1,2,3], Miguel Castro[1,2], Hervé Le Breton[1,2,3], Mireille Garreau[1,2], Pascal Haigron[1,2]

[1] INSERM U1099, Rennes, 35000 France
[2] Université de Rennes 1, LTSI, Rennes, 35000 France
[3] CHU Rennes, Service de Cardiologie et Maladies Vasculaires, Rennes, 35000 France

e-mails : {helene.feuillatre, miguel.castro, mireille.garreau, pascal.haigron} @univ-rennes1.fr, {vincent.auffret, herve.lebreton} @chu-rennes.fr

Keywords: Case-Based Reasoning, Similarity measure, Transcatheter Aortic Valve Implantation

1. Introduction

Case-Based Reasoning (CBR) makes the assumption that past experiences may be useful in solving similar current problems. In the case of Transcatheter Aortic Valve Implantation (TAVI), the CBR could help practitioner to plan the procedure, i.e. to take the most convenient decision about the vascular access and the prosthesis. The CBR is composed of four steps: Retrieve, Reuse, Revise and Retain [1]. In this work we present the CBR process implemented in the EurValve project and focus on the case retrieve issue. Different similarity measures (SMs) can be used to find the *k* closest cases in the retrieve step [2]. Our objective is more precisely to analyze the influence of different definitions of SM on the CBR performance.

2. Method

The retrieve step is mainly based on the computation of a metric to assess the similarity between cases. In our CBR (Fig. 1), the other steps (Reuse, Revise and Retain) are realized through a graphical user interface (GUI) in order to leave the final choice for the decision making to the practitioner. In the retrieve step, defining a convenient SM is essential.

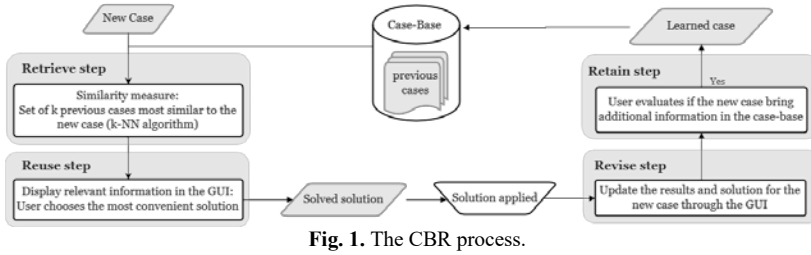


Fig. 1. The CBR process.

Generally a standard weighted heterogeneous SM is used in association with the *k*-nearest neighbor algorithm to retrieve similar cases in the case-base. The Heterogeneous Euclidean-Overlap Metric (HEOM) [3] uses the overlap metric for qualitative attributes and the normalized Euclidean metric for quantitative attributes. El-Fakdi et al. [4] have shown the feasibility of designing a CBR for TAVI but they not focus on investigating SMs. A classical definition of heterogeneous SM with attributes selection was used. In this work we define two additional weighted heterogeneous SMs (WHSM1 and WHSM2), where different normalized

metrics are considered according to the attribute type (Euclidean distance for quantitative data, Hamming distance for Boolean data and similarity matrix for ordinal data). In WHSM1, all case attributes are taken into account. In WHSM2, only relevant attributes are selected according to the decision to take.

3. Results

A leave-one-out cross validation is performed to evaluate the SMs in a case-base of 69 patients. These cases have no missing attributes. The Fig. 2 illustrates the performance of the four SMs for the decision about the vascular access (right or left transfemoral, left subclavian, transapical and transaortic). The performance represents the percentage of cases where the correct decision appears at least once into the *k* most similar cases. With WHSM2, the percentage of cases reaches almost 86% when *k=5*. Results show that using a dedicated SM with metrics adapted to the selected attributes improves the retrieve step.

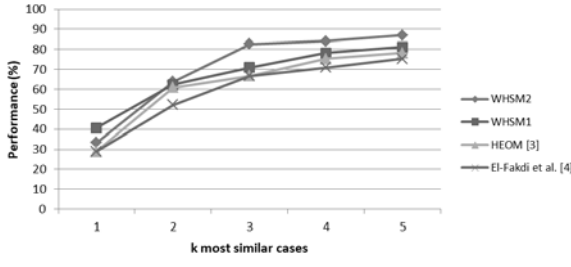


Fig. 2. Performance (%) of different similarity measures in CBR.

4. Conclusions and future work

We have shown objectively that the definition of SMs can influence the retrieved cases and consequently the decision taken through the CBR. Analysing the performance of CBR is a critical issue. In our work, SMs are defined from standard clinical attributes. They can be improved by using clinical decision trees for attributes selection and weights determination. Learning approaches could be also considered to estimate the weights (by using evolutionary algorithms for example). High level attributes could be defined thanks to additional descriptors to better identify similar shapes. These descriptors could be based on dedicated anatomical maps or statistical shape models. Moreover, the management of case-base, the easy access to a centralized and secured database and its updating with respect to the amount of information contained in a new case represent also critical issues.

Acknowledgments. This work is supported by the EU project EurValve *Personalised Decision Support for Heart Valve Disease* H2020 PHC-30-2015 689617.

References

1. A. Aamodt and E. Plaza, “Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches,” AI Communications, no. 1, pp. 39–59, 1994,
2. Y. Avramenko and A. Kraslawski, “Similarity concept for case-based design in process engineering,” Computers & Chemical Engineering, vol. 30, no. 3, pp. 548–557, Jan. 2006,
3. D. R. Wilson and T. R. Martinez, “Improved heterogeneous distance functions,” Journal of artificial intelligence research, vol. 6, pp. 1–34, 1997,
4. A. El-Fakdi, F. Gamero, J. Meléndez, V. Auffret, and P. Haigron, “eXiTCDSS: A framework for a workflow-based CBR for interventional Clinical Decision Support Systems and its application to TAVI,” *Expert Systems with Applications*, vol. 41, no. 2, pp. 284–294, Feb. 2014.

Towards Measuring Activity Levels with a Smart Home in a Box

Ryan McConville, James Pope, Raul Santos-Rodriguez, Robert Piechocki, Ian Craddock

University of Bristol, UK

e-mails: {ryan.mcconville,james.pope,enrsr,r.j.piechocki,
ian.craddock}@bristol.ac.uk

Keywords: smart home in a box, activity level, activity metrics, indoor localisation

1. Introduction

The EurValve Smart Home in a Box (SHiB) (J. Pope, 2017) is designed to be a scalable and easy to deploy system to collect data from patients in their home environment. The system is to be deployed in the homes of 60 heart valve surgery patients at three different periods of time, prior to, a short time after, and a considerable time after, surgery. The kit consists of a wearable device that is worn on the wrist which broadcasts 25Hz accelerometer readings using Bluetooth Low Energy (BLE) to four small gateways, each one positioned in rooms of interest (livingroom, bedroom, kitchen and a patient chosen room). The Received Signal Strength Indicator (RSSI) value for the wearable is also logged by each gateway. Each gateway connects to a 4G router which periodically sends the collected data to a central server for analysis. Using the accelerometer data we can recognize the activities of patients in their home, such as walking, lying down or sitting. Further we add context to this by localizing where patients are inside their house to better understand their behavior. The objective is to provide information of interest to clinicians about the activity levels of patients to enrich self-reporting mechanisms.

2. Active patient recognition

As an effort towards measuring patient activity levels, we illustrate the methodology by analyzing the data of a single patient over the course of 4 full days. The patient deployed and setup the system which included a straightforward calibration process that captures how the patient walks, sits and lies down. Each activity was performed for two minutes in a different room: calibration for walking was carried out in the kitchen, sitting in the livingroom, and lying in the bedroom. A patient chosen activity was carried out in the fourth room. Using the data from this calibration process a machine learning model was trained on each activity. This model is used to predict whether a patient is currently active (i.e. not sitting or lying). Further another machine learning model is trained using the RSSI values obtained from the wearable to gateway communication. RSSI can be used as a means of estimating the distance between a transmitting and receiving device, in this case the wearable and a gateway. This model is be used to predict which room the patient is in. A rolling 5 second mean of RSSI values was calculated and the predictions below a threshold t (here t=0.99) were discarded, as it was assumed the patient had not changed room since the last prediction above t. This is to account for the variability in RSSI.

3. Results

We first discuss the indoor location prediction for the patient using the EurValve system installed in their own home and calibrated by themselves. As we can see in Fig. 1 insights into the behavior of the patients daily living can be seen from the time spent in each room and the rooms they transition between, as well as the frequency of these transitions. Periods of activity

are also plotted. We can see bouts of activity throughout the day corresponding to room transitions, as well as during the night as the patient is sleeping. In general, clinicians can get a high level overview of patient behavior from these plots for completeness.

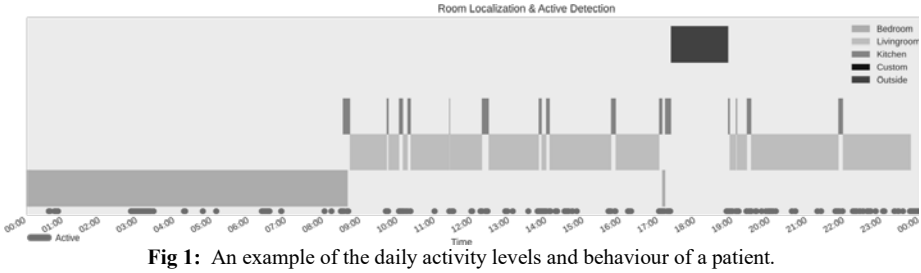


Fig 1: An example of the daily activity levels and behaviour of a patient.

In Table 1 we provide a number of metrics calculated from the predictions of our Random Forest model on whether the patient is active or not. We report the mean, min and max of each metric, computed on a daily basis. These measures can then used to understand the true activity levels of patients at home, as opposed to relying on self reporting. We can see wide variation in the amount of time active per day, as well as days where the patient has very little continuous activity and days where there are long periods of inactivity. By tracking these metrics the clinician may be able to better understand the condition of the patient.

Tab. 1. The mean/min/max of daily metrics for a single patient.

Active	**In bed**	**Outside**	**Continuous Active period**	**Continuous Inactive period**	**Room transfers**
100/5/157 minutes	8:57/8:30/9:45 hours	3:43/1:32/5:54 hours	27/1/41 minutes	177/33/299 minutes	27/19/31

4. Conclusion and future work

We have provided some initial analysis working towards automatically deriving metrics of clinical interest regarding the activity levels of a patient using a wrist worn wearable as part of a SHiB deployed in their home. Key challenges we face include the limited labels we have for each activity, as well as the reduced set of activities we have prior knowledge of.

Acknowledgments. This work is supported by the EU project EurValve *Personalised Decision Support for Heart Valve Disease* H2020 PHC-30-2015 689617.

References

1. J. Pope, R. McConville, M. Kozlowski, X. Fafoutis, R. Santos-Rodriguez, R. J. Piechocki, I. Craddock: SPHERE in a Box: Practical and Scalable EurValve Activity Monitoring Smart Home Kit. in Proc. IEEE Workshop on Networks of Sensors, Wearable, and Medical Devices, 2017.

Uncertainty in Model-Based Treatment Decision Support: Applied to Aortic Stenosis

Roel Meiburg, Marcel C.M. Rutten, Frans N. van de Vosse

Eindhoven University of Technology, Den Dolech 2, Eindhoven, 5612AZ Netherlands

e-mails: {r.meiburg, m.c.m.rutten, f.n.v.d.vosse}@tue.nl

Keywords: sensitivity analysis, uncertainty quantification, data assimilation, decision support

1. Introduction

Mathematical methods have been applied to gain insight into pathophysiology of the cardiovascular system as early as the 1900's [1], and are maturing to the point where clinical application may be feasible [2] in the form of treatment outcome prediction. However, several points must be considered. Firstly, these models must be tailored to describe patient-specific haemodynamics, which is done by tuning model input parameters. Due to the limited available data clinically, the most important model parameters to tune must be identified, and because of the noisiness of clinical data, the tuning procedure must be robust against measurement uncertainty. Finally, this method must quantify the uncertainty on the predicted model outputs. The estimation method presented here will be applied to the clinical problem of patient condition assessment prior to a minimally invasive aortic valve replacement to treat aortic stenosis.

2. Materials & Methods

The presented framework consists of a three-part process: parameter prioritisation; parameter estimation; and treatment prediction. It will be applied to a simple lumped parameter model of the circulatory system, where the heart is described via a single-fibre model [3], and the closed circulation with a Windkessel model [4].

Parameter prioritisation is based on a variance-based global sensitivity index, the Sobol index, which relates the input parameter variance to output parameter variance [5]. These indices are calculated via the Polynomial Chaos Expansion (PCE) method, which first computes a meta-model of orthogonal polynomials, after which output variance and Sobol indices can be determined analytically. [6]

Parameter estimation is performed via the Unscented Kalman Filter (UKF) [7], which is able to handle limited, noisy clinical data while also estimating model parameters. Furthermore, it shows the parameter variation during the cardiac cycle, which can be interpreted as a measure parameter uncertainty.

Finally, the PCE method is reused with the personalized parameter ranges to produce an estimate of the output of interest as well as output variance after *in silico* treatment.

3. Results

The framework has been successfully applied to data generated by a mock-loop [8], which mimics the full circulation and ventricles. The single-fibre model is also used to model the heart function. Figure 1 shows the pressure in the pump which mimics the left ventricle as measured by a pressure wire, the Kalman estimate and the result of the estimated parameters. Figure 2 shows an example of parameter evolution and resulting estimated range. Quantifying the estimated parameter uncertainty leads to imperfect estimated outputs, as shown in Table 1.

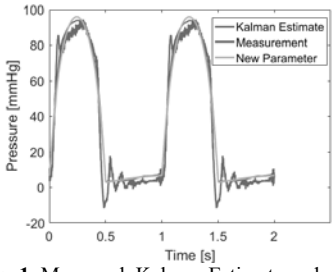


Fig. 1. Measured, Kalman Estimate and model prediction of left ventricular pressure.

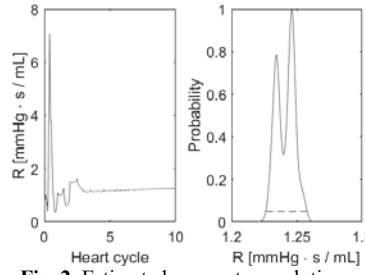


Fig. 2. Estimated parameter evolution and parameter probability range.

Tab. 1. Measured outputs and predicted values.

Parameter Name	Measured Value	Predicted Value +- Standard deviation
Systolic Pressure [mmHg]	88	80 +- 5
Diastolic Pressure [mmHg]	54	63 +- 5
Maximum Aortic Valve Pressure drop [mmHg]	92	75 +-5
Mean Aortic Valve Pressure Drop [mmHg]	15	18 +- 1
Cardiac Output [L/min]	1.7	1.8 +- 0.64

4. Conclusions and future work

We have created a framework to use mathematical modelling in a clinical setting, which aims to overcome the challenges of limited, noisy clinical data, provide quantitative measures of treatment outcome, while staying computationally tractable.

In the future, a more comprehensive model which is able to more accurately describe patient haemodynamics is required. Also, the connection between the not-normally distributed input parameter estimates and output uncertainty quantification requires improvement, such as applying moment-independent uncertainty propagation methods.

Acknowledgments. This work is supported by the EU project EurValve *Personalised Decision Support for Heart Valve Disease* H2020 PHC-30-2015 689617. This research was supported in part by the PL-Grid infrastructure.

References

1. F. Otto, "Die Grundform Des Arteriellen Pulses." Zeit Bio 37: 483–526,
2. L. Speelman et al. "Patient-Specific AAA Wall Stress Analysis: 99-Percentile Versus Peak Stress." Eur J Vasc Endovasc Surg 36, no. 6: 668–76,
3. Bovendeerd, P.H.M., et al. "Dependence of Intramyocardial Pressure and Coronary Flow on Ventricular Loading and Contractility: A Model Study." Ann Biomed Eng 34, no. 12: 1833–45,
4. G.N. Jager, et al. "Oscillatory Flow Impedance in Electrical Analog of Arterial System: Representation of Sleeve Effect and Non-Newtonian Properties of Blood." Circ Res 16, no. 2: 121–33,
5. I.M. Sobol′, "Global Sensitivity Indices for Nonlinear Mathematical Models and Their Monte Carlo Estimates." Math Comput Sim, Second IMACS Seminar on MC Methods, 55, no. 1–3: 271–80,
6. G. Blatman, and B. Sudret. "An Adaptive Algorithm to Build up Sparse Polynomial Chaos Expansions for Stochastic Finite Element Analysis." J Prob Eng Mech 25, no. 2: 183–97,
7. E.A. Wan, and R. Van Der Merwe. "The Unscented Kalman Filter for Nonlinear Estimation." In Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373), 153–58, 2000,
8. S. Schampaert, et al. "Modeling the Interaction Between the Intra-Aortic Balloon Pump and the Cardiovascular System: The Effect of Timing." ResearchGate 59, no. 1.

Shape-Driven Dynamic Valve Segmentation for Cardiac TEE Ultrasound Images

Matthias Lenga, Tilman Wekel, Juergen Weese

Philips Research Laboratories Hamburg

e-mail: matthias.lenga@philips.com

Keywords: dynamic segmentation, shape-constrained models, deformable models, aortic valve, mitral valve

1. Introduction

The aortic and mitral valve area are of high interest for various applications such as planning implant procedures, outcome prediction, characterizing stenosis, or enabling CFD-based flow simulations. While recent segmentation approaches are quite promising for CT data, many of those fail for Echo images due to the much higher level of noise. However, there is an urgent need for enabling this technology for the ultrasound domain because it is cheap, fast, and minimally invasive. In this paper we show, how the valve regions can be robustly segmented in TEE images based on a shape-driven and multi-resolution approach. The core element of the pipeline is a shape-constrained deformable model that is defined in two different resolutions. First we localize and adapt a very coarse model to a single image. Secondly, we successively adapt the much more detailed model that comprises complex and dynamic representations of the aortic and mitral valve. It is shown how this technique is applicable to - and even benefits from time series data.

2. Methods

The core of our processing is a model-based segmentation framework that basically consists of the algorithm and the model that represents the prior information about the respective anatomy[1]. The **algorithms** are independent of the application. All the application depended parameters are encoded in the model, i.e. which components are available, how detailed the segmentation should be, how the image looks like. Details of the heart and valve anatomy may be captured by additional post-processing steps that are configured by the model and that depend on the model-based segmentation result.

A segmentation task generally consists of four steps:

1. Localisation of the heart in the image
2. Rigid and affine adaption of the mesh to the image
3. Deformable adaptation of the mesh to the image
4. Post-processing: Valve refinement and measurements

The **model** itself represents the prior knowledge about the anatomy that is going to be segmented. One or more triangulated “mean” meshes represent the geometry as well as the topology of the anatomy, how it is subdivided into sub parts such as the ventricles or the valves and how these parts typically look like at different resolutions. Boundary features are also a part of the model and describe how the respective image region around each triangle typically looks like. Both elements, mean mesh and set of features need to be trained based on a sufficiently large training dataset. The TEE model that is used for the presented purpose

consists of two resolution levels as it can be seen in Figure 1. The coarse model is used to localize the heart and to roughly adapt the shape of the chambers.

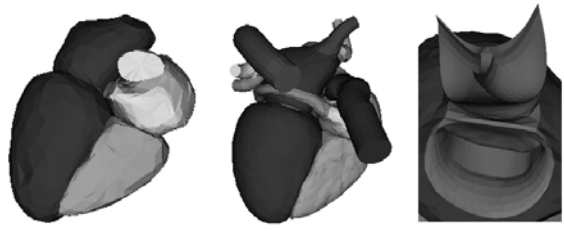


Fig. 1. Visualization of the model. Coarse geometry to roughly localize and adapt initial shape (left). Complex model (center) with detailed valve geometry (right).

3. Preliminary results

The proposed pipeline enables a fully automatic segmentation with minimal user parametrization. The quality of the segmented valves improves drastically if the state is known prior to the processing (open or closed). Figure 2 shows some preliminary results for visual inspection.

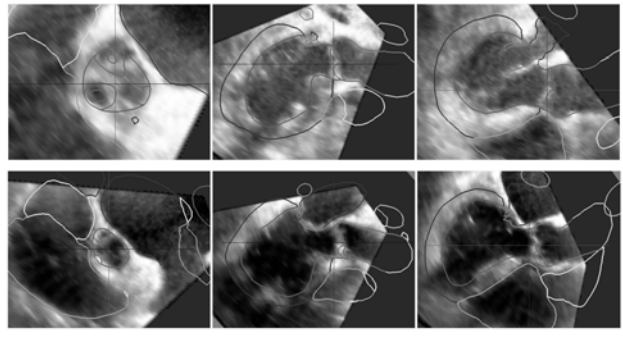


Fig. 2. Preliminary results for two different patients (row - wise) showing AV and MV regions.

4. Conclusions and future work

In future work we will try to improve the segmentation quality by incorporating more training data on the one hand, and further improve the topology of the model on the other hand. Another focus will be to robustly derive the valve state from the data directly without any additional user input.

Acknowledgements. This work is supported by the EU project EurValve *Personalised Decision Support for Heart Valve Disease* H2020 PHC-30-2015 689617. This research was supported in part by the PL-Grid infrastructure.

References

1. Weese, Jürgen, et al. "Shape constrained deformable models for 3D medical image segmentation." Biennial International Conference on Information Processing in Medical Imaging. Springer, Berlin, Heidelberg, 2001.

Architecture for Managing and Querying Collaborative Datasets

Daniel A. Silva Soto[1,2,3], Steven M. Wood[1,3]

[1] Scientific Computing, Royal Hallamshire Hospital, Beech Hill Road, Sheffield, UK
[2] Medical School, The University of Sheffield, Beech Hill Road, Sheffield, UK
[3] Insigneo Institute, The University of Sheffield, Mappin Street, Sheffield, UK

e-mails: d.silva@sheffield.ac.uk, steven.wood@sth.nhs.uk

Keywords: Data publication, data federation, data access, DICOM, SQL

1. Introduction

The EurValve research project aims to implement and validate a model-based decision support system for aortic valve disease [1]. As part of the multidisciplinary project, different partners collect and contribute data towards a virtual, larger dataset. To this end, an integrated data management system with a wide range of APIs supporting access for all classes of user was built, providing tools to clinical users which allow the de-identification of sensitive data before publication for the project. The system allows groups of users to define their own data models for their evolving needs - such as data scientists using machine learning to infer missing data values. Data access is managed so users have full control over their own datasets but only read access to others - e.g. engineering partners consume clinical data and produce subject specific biomarkers. Finally, it provides a transparent layer for running queries across all of the datasets in a project domain, provided they have been annotated correctly.

2. Solution design

Data publication is the process of collecting data from a variety of sources, structuring it into tables (that may be relationally linked), and making the data and schema accessible. The Data Publication Suite (DPS) is our solution for data publication, allowing processing data from many sources (e.g. relational databases, DICOM (image) stores, CSV files) and aiding structuring data from unstructured sources (e.g. loose files). Tables can be linked through related fields and fields can be semantically annotated and processed (e.g. anonymised).

Different project partners contribute to the virtual dataset by predicting values in different ways (physics simulation, data inferred from machine learning) before or after they are measured clinically. The virtual dataset is an amalgamation of data sourced from a number of providers. Traditional federated strategies involve multiple datasets that are joined by common keys. Our virtual dataset uses a preference rank to “stack” data columns and pick the appropriate data values across different datasets (e.g. a data value that has been clinically measured is preferred over a population average.) Virtual datasets have two properties: the ranked datasets and the fields by which they are joined. The ranked datasets are composed of a (single) root dataset and an ordered list of datasets which are federated and stacked to it.

Users may perform queries using a custom query language and our query execution API. We offer two types of queries, on top of traditional SQL queries. Federated queries are distributed across distinct databases. They allow the most common features of the SQL query syntax but execute across geographically disparate resources. The join strategy for these resources is defined in the virtual dataset definition. Stacked queries always return the requested data values and their provenantial dataset. A stacked query is first analysed to determine what sub-query needs to be executed against each of the sources of the virtual dataset. Sub-query results are scanned for empty entries, which are filled in with data values if available in the stacked datasets. Finally, user-defined filters are applied on the aggregate table of results. The processing of stacked queries is shown in Figure 1.

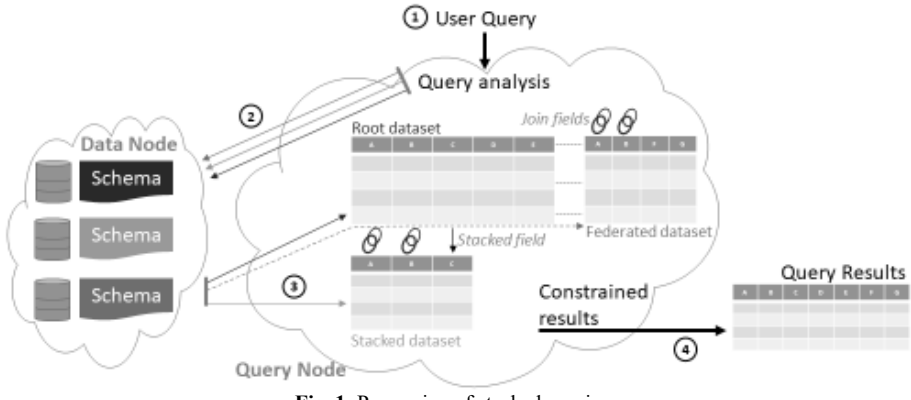


Fig. 1. Processing of stacked queries.

3. Results

The DPS is a C# application for which a number of plugins have been developed, each specialising in a particular data source, including: MS-SQL, MySQL, Access, Excel and other application-specific files. Dataset schemas are defined using JSON. Single datasets define table-field relationships for one or more tables. Field attributes include column name, display name, annotation, as well as distinct values for some fields. Virtual datasets list individual datasets and where they are hosted.

Data services are REST based written in C# and developed using the ServiceStack framework. The API is complemented by a web based query builder which features a drag-and-drop interface for building complex queries. Queries can be executed or saved for repeat use; results are previewed in the browser or downloaded in CSV format for external analysis. This is a client-side application built using HTML and Javascript; advanced functionality, such as pivot tables, are provided using open source Javascript libraries.

All data services sit behind an Apache web server, running CentOS in a VMWare virtual machine (8GB RAM, 100GB storage and 4 VCPUs) and securely hosted at The University of Sheffield [3]. Access control is provided through integration with the project's URL based security service. Access to datasets (both schema retrieval and data querying) is handled using JSON Web Tokens (JWT).

4. Conclusions and future work

We currently have three classes of users consuming data through the platform: HPC scripts pull data for use in complex modelling processes, engineers explore the data and extracting CSV files for use in tool development and prototyping, and data scientists explore machine learning techniques to infer missing data in the clinical collections. All of these users are interacting either through the UI or service layers and doing so with minimal support from the data management team.

Acknowledgments. This work is supported by the EU project EurValve Personalised Decision Support for Heart Valve Disease H2020 PHC-30-2015 689617.

References

1. EurValve project website: `http://www.eurvalve.eu/`,
2. M. Bubak, A. Belloum, S. Koulouzis, P. Nowakowski and D. Vasyunin. Data Management Services for VPH Applications. Workshop on Cloud Services for File Synchronization and Sharing, CERN, November 17-18, 2014,
3. EurValve Portal is deployed at `https://valve.cyfronet.pl`.

Advanced Security Services for Computer Simulation Research in Medicine

Jan Meizner[2], Marian Bubak[1,2], Tomasz Bartyński[2], Tomasz Gubała[2], Daniel Harężlak[2], Marek Kasztelnik[2], Maciej Malawski[1], Piotr Nowakowski[2]

[1] Department of Computer Science, AGH University of Science and Technology,
[2] ACC Cyfronet AGH, Kraków, Poland

e-mails: {bubak,malawski}@agh.edu.pl,
{j.meizner,t.gubala,d.harezlak,m.kasztelnik,p.nowakowski}@cyfronet.pl

Keywords: security, AAA, OpenID, encryption, AES, storage systems, medical data

1. Introduction

In this paper we present the Advanced Security Services for computer simulation based research, developed in the scope of the EurValve project [1]. We focus on the task of providing authentication, authorization and accounting functionality, as well as meeting challenges related to processing sensitive medical data. Our goal is to reference our previous work in the field [2] as well as describe further progress, using a fully integrated solution as a usage example. We conclude our paper with a description of the platform validation process, extending its security level with a POC tool which we have also developed.

2. Advanced Security Services

One of the key tasks in the project was to address three critical security-related aspects:

- Authentication, authorization and accounting (AAA);
- Data security during processing and storage;
- Mechanisms to ensure data cannot be recovered given reasonable time and resources, after being deleted (e.g. from decommissioned hard drives).

To address those issues, we have created a dedicated security services, as shown in Fig. 1.

3. Results – security as part of the EurValve Platform

The AAA abstraction is provided by:

- A Single Sign-On (SSO) authentication mechanism enabling integration with Identity Providers – currently PLGrid [3] OpenID – which facilitates centralized user management. In addition, a fallback mechanism based on local accounts (password-protected) has been developed (steps 1-2 in Fig. 1),
- An authorization solution based on a digitally signed JSON Web Token (JWT) [4] which is used to query a Policy Decision Point (PDP) for user access to external services (steps 3-4),
- A Policy Enforcement Point (PEP) developed as an NGINX web server script, to secure access to project services (steps 5-8),
- A mechanism which enables service owners to register services and manage access permissions via the Portal GUI or a REST API (steps A-B).

To provide data storage and processing security we have developed a mechanism capable of encrypting data using the Advanced Encryption Standard (AES) [5]. This component was

integrated into the File Store [6], which is used to store all binary data (BLOBs) via WebDAV. As there is currently no known mechanism capable of cracking AES 256 in reasonable time, this solution will prevent data from disclosure if it cannot be securely deleted. Currently, all data is encrypted (steps 1-6); however the encryption step may be bypassed if needed (steps A-B).

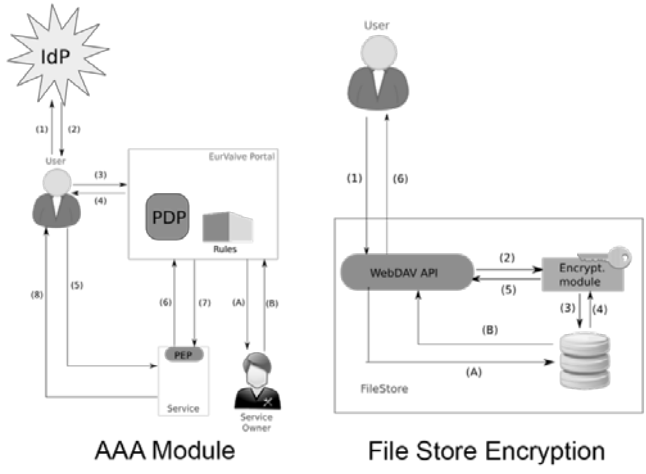


Fig. 1. Security components integrated into the EurValve Platform.

4. Conclusions and further work

The described solution has been fully integrated into the production environment and validated with a series of benchmarks, proving that it provides sufficient performance given the required amount of data. Additionally, we have developed a POC module based on the Filesystem in Userspace (FUSE) mechanism, which enables dispersal of chunks of data across multiple locations. In the future, we will considering extending this mechanism and provide additional features related to accounting and alerting.

Acknowledgments. This work is supported by the EU project EurValve, Personalised Decision Support for Heart Valve Disease H2020 PHC-30-2015 689617. This research was supported in part by the PL-Grid infrastructure.

References

1. EurValve project website: `http://www.eurvalve.eu/`,
2. Meizner J., Bubak M., Malawski M., Nowakowski P. (2014) Secure Storage and Processing of Confidential Data on Public Clouds. In: Wyrzykowski R., Dongarra J., Karczewski K., Waśniewski J. (eds) Parallel Processing and Applied Mathematics. PPAM 2013. Lecture Notes in Computer Science, vol 8384. Springer, Berlin, Heidelberg,
3. Bubak M., Kitowski J., Wiatr K., EScience on Distributed Computing Infrastructure: Achievements of PLGrid Plus Domain-specific Services and Tools, vol. 8500 (2014) Springer,
4. The JSON Web Token (JWT) Standard, `https://jwt.io/`,
5. Advanced Encryption Standard (AES), Federal Information Processing Standards Publication 197, Nov. 26, 2001,
6. Bubak M., Harężlak D., Wood S., Bartyński T., Gubała T., Kasztelnik M., Malawski M., Meizner J., Nowakowski P. (2017) Data Management System for Investigation of Heart Valve Diseases, CS3 Workshop on Cloud Services for File Synchronisation and Sharing, Amsterdam, 30 Jan - 1 Feb 20, 2017 (available at: `https://indico.cern.ch/event/565381/contributions/2402656/attachments/1403448/2143286/CS3-MBubak-eurvalve-Jan-2017.pdf`).

EurValve Model Execution Environment in Operation

Marian Bubak[1,2], Tomasz Bartyński[2], Tomasz Gubała[2], Daniel Harężlak[2], Marek Kasztelnik[2], Maciej Malawski[1], Jan Meizner[2], Piotr Nowakowski[2]

[1]Department of Computer Science, AGH University of Science and Technology, Kraków, Poland
[2]ACC Cyfronet AGH, Kraków, Poland

e-mails: {bubak,malawski}@agh.edu.pl,
{t.gubala,d.harezlak,m.kasztelnik,j.meizner,p.nowakowski}@cyfronet.pl

Keywords: personalized medicine, problem solving environment, security, flow simulations, segmentation

1. Introduction

The main goal of investigations in the EurValve project [1] is to develop a decision support system (DSS), clinically applicable for efficient treatment of valvular diseases. Simulations carried out in the process of developing this DSS require a dedicated problem solving environment which we refer to as the Model Execution Environment (MEE). MEE combines a set of complex modeling tools in a pipeline, facilitating evaluation of medical prospects and outlooks for individual patients. The research environment involves both interactive and batch processes, some of which are invoked manually, while others should be automated as they base on similar, recurring computations. A typical pipeline invokes computations on a cluster, fetching data for local analysis, launching an interactive cloud service to process selected data, and running a batch of jobs on the resulting datasets.

2. Model Execution Environment

The architecture of the MEE is shown in Fig. 1. To enable integration of tools in a pipline, a dedicated EurValve portal was created, with four main components. A GUI provides access to tabular (ArQ) and binary data (File Store). Portal users can browse data, upload new datasets or modify existing ones. Via the Atmosphere cloud platform [2, 3] a user can initiate and manage virtual machines running in private or public clouds. Computing may be delegated to the PLGrid infrastructure where the main resource is the Prometheus cluster [4], delivering 2.4 PFlops, 279 TB of RAM and 10 PB of storage capacity. Users have access to many commercial toolboxes, including Matlab and ANSYS (used by the EurValve 0D Model and 3D blood flow simulations created in the scope of the project). The security service provides a dedicated UI and REST APIs, where user/group permissions can be defined and checked

3. Results: Patient Case Pipeline

The tools presented above are applied in the execution of a Patient Case Pipeline, which orchestrates data and calculations for a specific patient in the following way:

1. For every user a dedicated File Store data structure is created where input data can be uploaded, along with a dedicated output directory where calculation results are stored.

2. For every patient case many pipelines can be executed. A pipeline execution involves a set of calculations performed on patient data. Each execution may use different inputs (e.g. to test the behavior of a model when incomplete data is available) and different model versions. Calculation results are stored in a dedicated File Store space.

3. Multiple computational models are integrated in the pipeline scope, including Segmentation (Philips), Parameter Extraction (Philips), Blood Flow Simulation (USFD), Parameter Optimization (USFD), 0DModel (USFD), Computational Fluid Dynamics (USFD)

and Reduced Order Model calculations (ANSYS), together with additional data handling, conversion and processing steps which can be arranged into various types of pipelines.

4. Results from different pipeline executions may be compared to identify differences between outputs – for example to distinguish which model yields better results, or how results differ depending on input scans.

5. A sample pipeline supported by the Portal is composed of the following steps: Segmentation, Parameter Extraction, Patient ROM, Parameter Optimization, 0DModel Suite (four model versions) and, finally, Uncertainty Analysis.

6. The pipelining system supports model versioning based on git branches and tags. Users can select which version should be started on the Prometheus supercomputer.

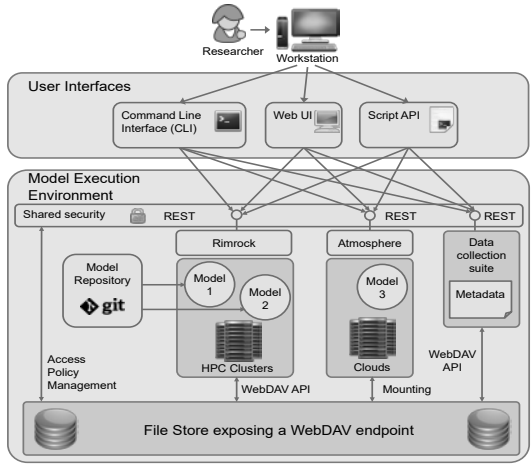


Fig. 1. Architecture of the Model Execution Environment.

4. Conclusion and future work

MEE is fully operational and used by the EurValve community. We plan to extend the patient case scenario by integrating additional data sources, for example enabling additional user parameters to be retrieved from the Data Store, and adding additional computational facilities delivered by other project partners. We also intend to provide additional automation features for existing pipelines, enabling fine-grained control over the configuration and execution of individual pipeline steps in the context of specific patient data.

Acknowledgments. This work is supported by the EU project EurValve Personalised Decision Support for Heart Valve Disease H2020 PHC-30-2015 689617. This research was supported in part by the PL-Grid infrastructure.

References

1. EurValve project website: `http://www.eurvalve.eu/`,
2. M. Kasztelnik, E. Coto, M. Bubak, M. Malawski, P. Nowakowski, J. Arenas, A. Saglimbeni, D. Testi, A.F. Frangi: *Support for Taverna Workflows in the VPH-Share Cloud Platform*, Computer Methods and Programs in Biomedicine, 146, July 2017, 37–46,
3. P. Nowakowski, M. Bubak, T. Bartyński, T. Gubała, D. Harężlak, M. Kasztelnik, M. Malawski, J. Meizner: *Cloud computing infrastructure for the VPH community*, Journal of Computational Science, Available online 21 June 2017,
4. Prometheus - `http://www.cyfronet.krakow.pl/computers/15226,artykul, prometheus.html`,
5. MEE website - `http://dice.cyfronet.pl/projects/details/EurValve`.

Centre for New Methods in Computational Diagnostics and Personalized Therapy

CECM Project Consortium
http://dice.cyfronet.pl/projects/details/CECM

e-mail: bubak@agh.edu.pl, t.gubala@cyfronet.pl, malawski@agh.edu.pl

Keywords: Teaming, centre of excellence, personalized medicine, computational medicine, business plan, H2020, Polish Smart Specialization Strategy

CECM – *the Centre for New Methods in Computational Diagnostics and Personalized Therapy* - is a unique EU-funded project, bringing together ACC Cyfronet AGH and five international partners. This will be the only project in Małopolska, and one of only three in Poland, to be carried out under the umbrella of Horizon 2020 Teaming for Excellence [1].

The aim of the project is to develop a detailed business plan for a proposed Centre of Excellence in the area of innovative medical diagnostics and personalized therapy, aided by advanced computer simulations. The plan will be submitted for the second stage of the Teaming competition in November 2018. If approved, the European Commission will allocate up to 15 million Euro to fund the Centre over a period of 7 years, with a similar amount contributed by public and private institutions in Poland.

The main objectives of the Centre will be:

- Development of new computation-based solutions for diagnostics and therapy in daily healthcare.
- Systematic involvement of regional biomed businesses, specialising in technologies and services for personalised medicine, in high-profile research projects and clinical adoption of their outcome.
- Development of education initiatives to train knowledge workers with the skills in data analytics, simulation, and HPC/Big Data, to respond to the growing demand for skilled workforce in medical devices and bio-engineering.
- Strong advancement of algorithms, models and technologies involved in personalised medicine, including design of holistic, replicable, generic framework for simulation-based Decision Support Systems (DSS) creation.

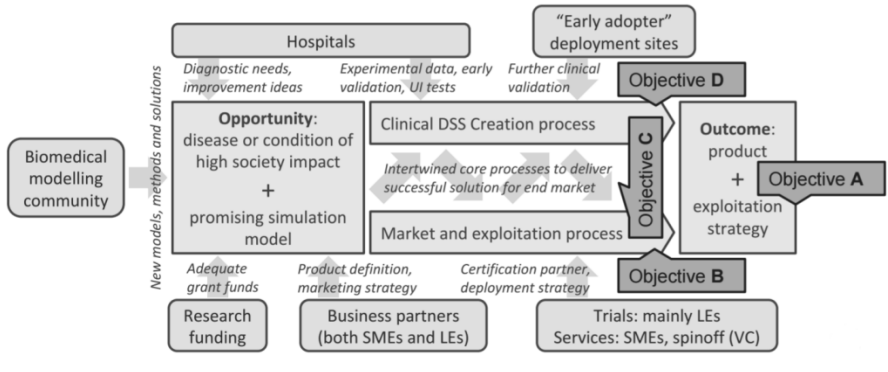


Fig. 1. The value chain in the new Centre of Excellence.

The value chain of the new Centre of Excellence is presented in Fig. 1. The Centre will operate in accordance with the Polish Smart Specialization Strategy [2,3], with the goal of bringing together international research and business communities. Principal performance indicators will include, i.a., the number of new solutions based on computational models introduced to clinical practice, the number of new innovative marketable products and services, as well as the number of frequently cited scientific publications, patents and grants obtained by the Centre.

The new Centre of Excellence will be located in Kraków – a national and European hotbed of scientific and business activity. The city's universities train a large number of future medical and IT specialists, its teaching hospitals are highly regarded by the scientific community and the number of local life science enterprises is on the increase.

The Centre will directly expand the regional research potential by fostering new scientific collaborations, offering world-class postgraduate training opportunities and facilitating knowledge transfer to newly emerging high-tech SMEs. The Centre's impact will also be felt beyond Małopolska through advances in modern medicine and the corresponding improvements in healthcare quality.

The Centre's business plan will be co-developed by:

- ACC Cyfronet AGH – experts in simulations and provisioning IT infrastructures for science,
- Klaster LifeScience Kraków – a Key National Cluster,
- University of Sheffield and Insigneo Institute – experts in *in silico* modeling in clinical practice,
- Forschungszentrum Jülich – experts in high-performance computing and large-scale data analysis for science and the industry,
- Fraunhofer ISI – experts in systemic solutions and medical innovations,
- National Centre for Research and Development.

Acknowledgments. The project is funded in the framework of H2020-WIDESPREAD-2016-2017 TEAMING PHASE 1: CSA. The Project Coordinator is the National Centre for Research and Development. In Phase II the project will be coordinated by ACC Cyfronet AGH.

References

1. EC H2020 - Spreading Excellence and Widening Participation webpage `http://ec.europa.eu/programmes/horizon2020/en/h2020-section/spreading-excellence-and-widening-participation`,
2. National Smart Specialisation Strategy website, `https://rio.jrc.ec.europa.eu/en/library/national-smart-specialisation-strategy-0`,
3. Regional Innovation Strategy 2020, Department of Economy Development, Małopolska Marshal Office, June 2016 .

Author Index