Univerzita Karlova v Praze
1.lékařská fakulta

Charles University in Prague

1st Faculty of Medicine

Studijní obor: Biomedicínská informatika
Study domain *Biomedical informatics*

Autoreferát dizertační práce

Summary of the dissertation

Využití technologie GRID při zpracování medicínské informace

Utilization of GRID technology in processing of medical information

Mgr. Tomáš Kulhánek

Praha 2015

Doktorské studijní programy v biomedicíně

Univerzita Karlova v Praze a Akademie věd České republiky

Obor: Biomedicínská informatika

Předseda oborové rady: Prof. RNDr. Jana Zvárová, DrSc.

Školicí pracoviště: CESNET z.s.p.o., Ústav patologické fyziologie 1.LFUK

Školitel: Ing. Milan Šárek, CSc.

Konzultant: Doc. MUDr. Jiří Kofránek, CSc.

Disertační práce bude nejméně pět pracovních dnů před konáním obhajoby zveřejněna k nahlížení veřejnosti v tištěné podobě na Oddělení pro vědeckou činnost a zahraniční styky Děkanátu 1. lékařské fakulty.

Contents

Abstrakt (česky)

Práce se soustředí na vybrané oblasti biomedicínského výzkumu, které mohou profitovat ze současných výpočetních infrastruktur vybudovaných ve vědecké komunitě v evropském a světovém prostoru. Teorie výpočtu, paralelismu a distribuovaného počítání je stručně uvedena s ohledem na počítání v gridech a cloudech. Byla studována oblast výměny medicínských snímků a Gridový PACS systém byl propojen s existujícími distribuovanými systémy pro sdílení DICOM snímků. Další studovanou doménou byla věda týkající se lidského hlasu. Vzdálený přístup k aplikaci pro analýzu hlasu v reálném čase byl představen zároveň s úpravou protokolů pro vzdálenou plochu pro přenos zvukových nahrávek. To přináší možnost využití stávajících aplikací na dálku specialisty na hlas.

Byl studován přístup tzv. systémové biologie v oblasti lidské fyziologie a patofyziologie. Bylo přispěno k metodologii modelování lidské fyziologie pro tvorbu komplexních modelů založených na akauzálním a objektově orientovaném modelovacím přístupu. Byly představeny metody pro studium parametrů pomocí technologie počítání v gridech a v cloudech. Proces identifikace parametrů středně komplexních modelů kardiovasculárního systému a komplexního modelu lidské fyziologie lze významně zrychlit při použití cloud computingu a dobrých výsledků lze dosáhnout v rozumném čase. Tato metoda umožňuje aplikovat parametrické studie ve fyziologickém a biologickém výzkumu. Toto může zlepšit praktické použití matematických modelů a identifikaci parametrů ve zdravotní péči.

Klíčová slova: gridové počítání, počítání v cloudu, výpočetní fyziologie, odhad parametrů, výměna medicínských snímků, analýza hlasového signálu

Abstract

This thesis focuses on selected areas of biomedical research in order to benefit from current computational infrastructures established in scientific community in european and global area. The theory of computation, parallelism and distributed computing, with focus on grid computing and cloud computing, is briefly introduced. Exchange of medical images was studied and a seamless integration of grid-based PACS system was established with the current distributed system in order to share DICOM medical images. Voice science was studied and access to real-time voice analysis application via remote desktop technology was introduced using customized protocol to transfer sound recording. This brings a possibility to access current legacy application remotely by voice specialists.

The systems biology approach within domain of human physiology and pathophysiology was studied. Modeling methodology of human physiology was improved in order to build complex models based on acausal and object-oriented modeling techniques. Methods for conducting a parameter study (especially parameter estimation and parameter sweep) were introduced using grid computing and cloud computing technology. The identification of parameters gain substantial speedup by utilizing cloud computing deployment when performed on medium complex models of cardiovascular system and complex models of human physiology. This makes such kind of study applicable in order to perform identification of physiological system in reasonable time for physiological and biological research and good results are available in a reasonable time. This can improve practical usage of mathematical models in healthcare.

Keywords: grid computing, cloud computing, computational physiology, systems biology, parameter estimation, medical image exchange, voice signal analysis

List of Abbreviations

BOINC	Berkeley Open Infrastructure for Network Computing, page 7
EGI	European Grid Infrastructure, page 7
FMI	Functional Mockup Interface, page 12
FMU	Functional Mockup Unit, page 12
IaaS	Infrastructure as a Service, page 8
NGI	National Grid Initiative, page 7
RDP	Remote Desktop Protocol, page 10
WLCG	Worldwide Large Hadron Collider Computing Grid, page 7
XML	Extensible Markup Language, page 12

1. Introduction

Grid computing is usually defined as sharing computational and data storage resources across organizational boundaries which can give a user much more computational or storage capacity. Grid computing in contrast to common distributed computing focus on large-scale resource sharing. The technology under grid computing provides access to a computational resources in a federated way, while preserving some rights of the owner. Requirements, standards and architecture, were proposed and published, e.g., by Foster et al. [1, 2] and such infrastructures are currently distinguished as "service" grids. It's non-trivial task to maintain scientific grid, thus specialists from the so-called national grid initiatives (NGI) maintains and cooperates with similar grid initatives of other countries. In Europe these are coordinated, e.g., by European Grid Infrastructure (EGI). One of the largest project computed in these grid infrastructures are related to experiments of high-energy physics in order to process a large number of observed data in a reasonable time [3]. The Worldwide Large Hadron Collider Computing Grid (WLCG) was designed to store and process almost 30 PetaBytes of data per year in the period of 2009-2013 [4].

Another approach to grid computing is joining desktop computers from an individual user to form a voluntary or desktop grid. It was popularized by a project that tries to identify uncommon signals from space to search for extraterrestrial intelligence (SETI@Home)[1][5]. And general-purpose frameworks were built in order to facilitate the development of projects that use a similar philosophy of computing on desktop computers, e.g., BOINC [6] and others.

In recent years, the development of virtualization technologies has enhanced the availability of services that are provided by grid computing. It has additionally enabled an evolution of the so-called *cloud computing*, in which computing resources can be rapidly provisioned and released with minimal management effort or service provider interaction. This implicates

[1] http://setiathome.ssl.berkeley.edu/

important feature of cloud-computing – elasticity – ability to scale up and down computing resources when required [7]. The cloud computing is provided in several models, however, currently the scientific infrastructures offer mainly Infrastructure as a Service (IaaS), which offers the whole virtual infrastructure including virtual machine and network accessible for user per request.

With respect to technology development available in scientific infrastructures, this thesis focus not only on grid computing but also on cloud computing technology, which were available for scientific computing within grid infrastructures since 2012.

2. Hypothesis

Based on the development of technologies and the philosophy of providing them to end users, this thesis focuses on the multidisciplinary research related to grid computing, as well as to cloud computing. The hypothesis of this thesis is that the technologies that relate to grid computing and cloud computing may improve the processing of medical information in order to perform demanding tasks that are almost impossible or require onerous effort to achieve, using classical local or institutional resources.

The particular goals of this thesis were:

- To study the latest achievements in the field of exchanging medical images and possible improvements using the grid computing and cloud computing technology.
- To identify use cases in other fields of biomedicine which are suitable to utilizing the power of grid computing and cloud computing infrastructure.
- To develop and test the prototype application that utilizes grid or cloud technologies.

This thesis tries to discuss the hypothesis in different areas of biomedical research and its application which were identified during the work. (1) the exchange and processing of medical images, (2) the analysis of human voice and (3) the modeling and simulation of human physiology.

It tries to find answers to the following additional questions:

- *Is it beneficial to utilize grid computing and cloud computing technology for the processing of medical information and how do we do this?*
- *What are the limitations of processing medical information in grid or cloud?*
- *How can the grid computing and cloud computing influence the direction of biomedical research?* There was an idea that grid computing technology inspires the current architecture of distributed systems, e.g., exchanging medical images and influences the direction of information systems in hospitals.

3. Methods

From a computer science (informatics) point of view, it is assumed that the processing of medical information is, in general, a computational problem, which is understood as a task that can be solved by a computer. An algorithm is a set of operations that is used to accomplish tasks and solve problems. The important features of an algorithm are effectivity (what is the time complexity of the algorithm regarding the size of input data) and scalability (how far can an algorithm benefit from parallel computing). Grid computing and cloud computing brings a technology that enables parallel computing in a large amount of shared computers, servers or cluster of servers introduces large speedup of computation and can decrease the time of computation substantially. However, problems solvable by algorithms with exponential time complexity (e.g. NP-hard or NP-complete) can't be addressed by any large scale infrastructure [8]. Therefore, additional non-exact methods for such type of problems are used to obtain at least some solution including heuristic method (eliminate some steps or solution classes that seems to not go to optimal solution), randomization (pseudo random values are generated and statistical methods can be used to compute expected optimal value) and others.

3.1 Sharing Medical Images

The DICOM standard was used as a joint protocol to integrate grid-based system Globus MEDICUS [9] with current production system MEDIMED for sharing medical images among different hospitals [10], which is based on common distributed system with central cluster. Globus MEDICUS [9, 11] implements a DICOM Grid Interface Service (DGIS) and integrates the open-source PixelMed™ Java DICOM Toolkit[1] into a web service, communicating via the DICOM protocol. Furthermore, it forwards queries to the underlying services within Globus toolkit. The console application of the MEDIMED project can be interconnected via DICOM protocol with local PACS system and selected DICOM studies can be sent/retrieved to another institution connected to the MEDIMED project.

3.2 Voice Science

The software for parameterized Voice Range Profile (ParVRP) and Voice Range Profile in Real time (RealVoiceLab) was already developed and calibrated for selected types of microphones in an MS Windows platform by Fric et al. [12, 13]. Its implementation is carried out in an MATLAB environment, utilizing Signal Processing Toolbox[2]. It is compiled with a MATLAB Compiler and distributed as an executable. To migrate this legacy application into distributed environment, the virtualization can be used with a protocol to control an application remotely. Remote Desktop Protocol (RDP) is a proprietary protocol that is used for desktop sharing. It was primarily developed in a Microsoft Windows platform, however, today, clients and

[1] http://www.pixelmed.com/ accessed February 2015
[2] http://www.mathworks.com/products/signal/ accessed February 2015

servers exist for several other platforms. RDP itself contains the redirection of several services, e.g., audio, sound recording, drive access.

3.3 Computational physiology

A mathematical formalization of the fundamental knowledge and relation among a biological system – a mathematical model - is used as a base abstraction in order to utilize the current discoveries of the genomics and proteomics. It is also used to formalize the knowledge and construct a "Physiome Model". By definition, a model is the simplification of a complex reality. Constructing the models and integrating them into a complex entity, which can be used for further purposes, is schematically illustrated in Figure 3.1.

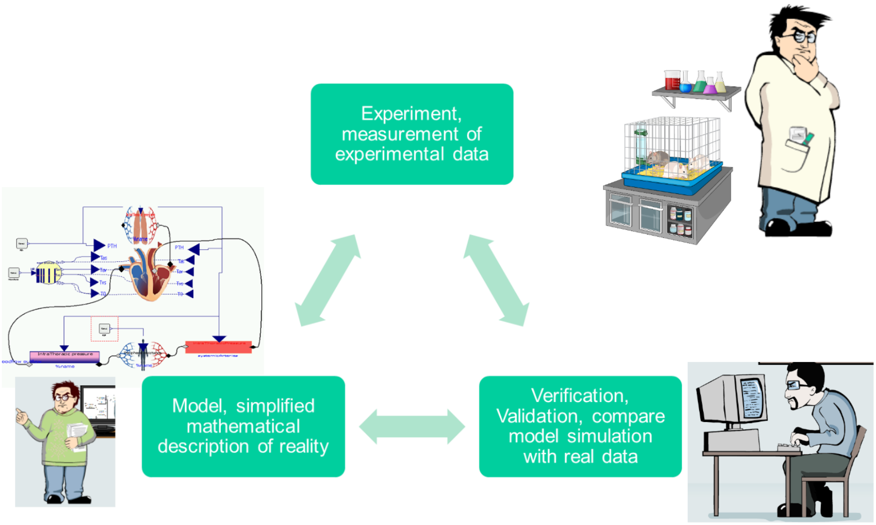

Figure 3.1: Schematic illustration of the scientific process. The experiments produce data that are interpreted and a hypothesis is formalized as a model. Validation compares the model simulation with the experiment, if the model satisfies the criteria – if it is in agreement with real experiments – then the validated model can be used for other purposes.

There are used several technologies in order to implement a formalized model of physiology. Within the work of this thesis a Modelica language was choosen, because it was identified as robust to maintain most complex models of human physiology in understandable way [23].

Usually, some knowledge of the system - the structure - is available and unknown coefficients (parameters) remain unknown. Once the model is formalized and constructed, a further problem is to estimate the model parameters so that the model reproduces a real world system. Without any further knowledge about the model, the problem of parameter estimation (or system identification) was shown to belong to the *NP-complete* problems [25], which implies that the best known exact algorithm solving this problem has exponential time complexity. The heuristic methods (evolution strategies), randomization methods (Monte-Carlo method) and others are commonly used in order to find at least some parameter estimation in a reasonable time. Therefore, in further work of this thesis an evolution strategy, genetic algorithm, was choosen as most robust for common models and system was proposed and implemented, which integrates this algorithm implemented in MATLAB environment and model simulation implemented in Modelica language.

The specific model of a studied system that is implemented in Modelica can be simulated in some Modelica tool. Or can be exported into a standard Functional Mockup Unit (FMU). Functional Mockup Interface (FMI) defines FMU as a standardized XML metadata description, packaged together with a binary library .DLL (or .SO), following a standardized API, published by Blochwitz et al. [26][3]. This API can be used to get/set values of model variables and to simulate the model.

4. Results

The pilot virtual infrastructure dedicated for research purposes, as proposed by the author of this thesis (3.), was established to consolidate and share

[3] `https://www.fmi-standard.org/` accessed February 2015

resources among different projects.

4.1 Medical Image Sharing

The pilot infrastructure of several servers was installed in several institutions in Prague, Czech Republic. Globus Toolkit and Globus MEDICUS were installed on them, the system connected with MEDIMED project integrates classical production system to share medical images with grid-based PACS system via the DICOM protocol. The grid-based system was tested with about 1300 DICOM records and enhanced with simple DICOMViewer available as web application. The grid-based solution allows to store large set of data records and manage replicas. The standard protocol to transfer data files gridFTP allows to effectively transfer parts of the files to the desired location from existing replicas within grid infrastructure to a desired location where an image processing can be performed. The grid-based solution brings robustness against the problems like single point of failure or bottleneck, where current systems of sharing medical images may suffer with such problems.

4.2 Remote Voice Analysis

RDP protocol was customized and support to transfer sound recording was implemented. A client plugin available for the Linux "rdesktop" application as well as for the Windows default "tsclient" application were customized to initiate sound recording and the raw data obtained from sound device is transferred to the custom RDP channel. The server plugin writes the data from the custom channel directly to the file in WAV format and concurently provides samples to the analytical application for real-time processing.

The default sound recording features of RDP protocol version 5.2 and 7.0 degrades sound quality transferred to the server application, furthermore, some samples of the sound were lost and sound become garbled or scratchy. The sound quality using the custom RDP channel is without loss

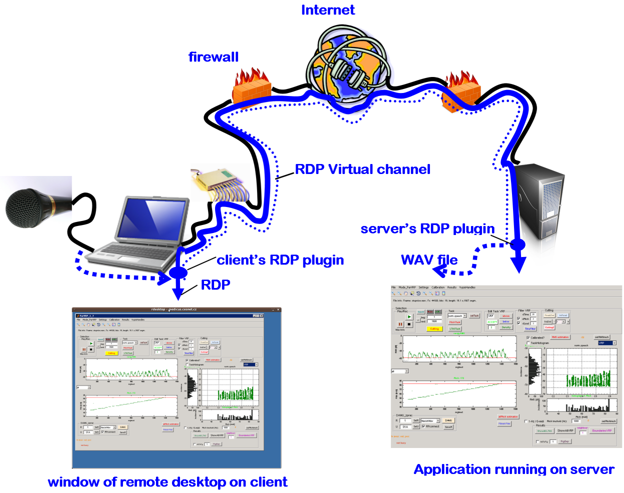

Figure 4.1: Architecture of a system for remote voice analysis and RDP plugins for sound recording redirection.

of information and acceptable for further analysis. Additionally, the remote application with custom RDP plugin was packaged as a virtual machine template and can be provisioned on cloud computing infrastructure in case of the need.

The application allows to record the voice of patient, the voice signal is transfered to remote application where it is processed and analysed, the results are visualized in real-time. The application is now used by several voice therapists and voice pedagogues in different areas of the Czech Republic and Slovakia to analyze the voice non-invasive and to see e.g. the progress of the voice training methods.

4.3 Computational Physiology

The proposed architecture of the system for parameter estimation (Figure 4.2) was influenced by the need of some interactivity and for the overall accessibility for users, which is fulfilled by the web UI.

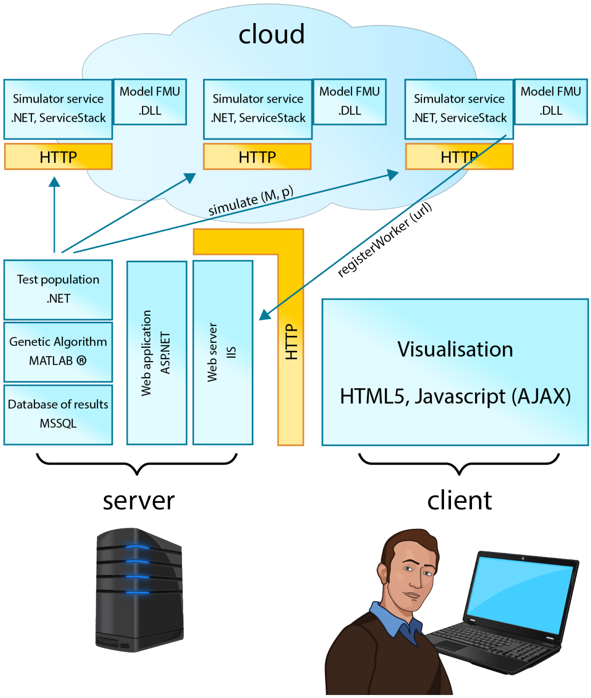

Figure 4.2: Architecture of a system that employs genetic algorithm and distributes the task *simulate* into a cloud computing environment.

The Modelica models is exported to standardized FMU and wrapped with custom code allowing, this can be called by custom application to perform simulation within needed constraints. In the time of writing this thesis, the most stable Modelica tool was Dymola version 2015[1], and most stable export was to FMU for a MS Windows platform.

[1] http://www.dynasim.se - Dymola tool, accessed March 2015

4.4 Parameter Sweep

5. Discussion

The presented solution, which is based on Globus MEDICUS, is, in general, a data warehouse, that stores one or more copies of DICOM images. However, federated files and metadata that are stored within home institutions, which only share network infrastructure to interchange the DICOM studies, seems to be a preferred and more acceptable solution by hospitals today, as published by Chervenak et al. [27]. The grid computing infrastructure seems to be suitable for research and educational purposes, but not generally acceptable for clinical use.

In the case of remote voice analysis, the remote access to an application via network protocol keeps the majority of user experience, as presented in section 4.2 and is a way how to migrate legacy application into the computing infrastructure and offer it as a service via network protocols and can be instantiated per request.

In the case of the application for parameter estimation presented in section 4.3, the computation is sensitive on communication overhead. For simple models, local high performance computing (HPC) resources are most beneficial. For medium and highly complex models, the deployment of worker nodes into a cloud computing environment is worth considering. Another challenge is an optimal size of population for genetic algorithm so the algorithm will converge to some acceptable solution in a reasonable time, as Gotshall et al. proposed a method for determining the optimum population size for a given problem [28].

The parameter sweep problem is considered as embarrassingly parallel and highly suitable for high throughput computing (HTC), which is the main focus of current grid computing infrastructures.

When porting an application to a grid environment, one of the important decision to consider is the platform of the used system. The architecture, which involves computational nodes that are deployed in a cloud computing infrastructure is influenced by the fact that the model implementation is exported from a third party tool to the standard FMU library for the MS Windows platform, as mentioned in section **??**. This determines the platform of the worker node and the virtualization - or, in the case of parameter estimation, cloud computing is utilized on a prepared platform with a MS Windows license. In the case of parameter sweep, a desktop grid computing BOINC worker and application for a MS Windows platform is only prepared for volunteers with the compatible system. To utilize the service grid infrastructure, an export of the model into a FMU library and implementation of the wrapper service must be done in the grid computing platform, which is usually a Linux based system. Another option is to use WINE[1] – a compatibility layer that is capable of running Windows applications on several POSIX-compliant operating systems, such as Linux, Mac OSX and BSD.

For smaller types of application and scientific community with their own tools, the question is, whether or not to invest on porting their tools to grid specific platform and parallel programming model. In the case of integrating with a service grid middleware or with desktop grid framework, expert knowledge is needed to configure and customize the system. This is the case for the sharing of medical images (section 4.1) and for parameter estimation and parameter sweep, which was tried with the desktop grid approach - BOINC framework (section 4.4).

Virtualization facilitates the integration effort, as presented in the case of remote analysis of the human voice (section 4.2) and in thecase of deployment of worker nodes in a cloud computing environment for parameter estimation (section 4.3).

Based on previous results and ideas, the answer to the research questions can be formulated:

- *Is it beneficial to utilize grid computing and cloud computing technology for the processing of medical information and how?*

[1] https://www.winehq.org/ WINE. Accessed March 2015

Grid computing and cloud computing can significantly speedup parameter study of medium and complex models in computational physiology. Such a speedup could influence its applicability in clinical use. For the case of sharing and processing medical images or analysis of voice signals, grid computing or cloud computing introduces technology that facilitates cooperation among a community of users from different geographically dispersed areas and facilitates the sharing of large data sets.

- *What are the limitations of processing medical information in grid or cloud?*

  Limitation are given by the effort needed to integrate or port an application carry out computation or share data. The cost of porting an application to cloud computing is reduced by virtualization technology, rather than to a grid computing environment, which needs additional work in order to adapt the application for a grid computing platform and API.

  The limitation are given by the theoretical features of algorithms too. Grid computing and cloud computing are not general solutions for hard problems (NP-complete problems). With connected with non-exact methods, a concurrent processing of many tasks may bring an acceptable non-exact solution.

- *How can the grid computing and cloud computing influence the direction of biomedical research?*

  The fact that the computation or data are processed remotely is one of the paradigm shift. The data moves from files stored in some folder to elements or objects living somewhere on server or cloud which can be shared among researchers.

  The research infrastructures, e.g. Integrated Structural Biology Infrastructure for Europe (INSTRUCT)[2], European Life Science Infrastructure for Biological Information (ELIXIR)[3], European Biomedical

[2] https://www.structuralbiology.eu/ accessed March 2015
[3] http://www.elixir-europe.org/ accessed March 2015

Imaging Infrastructure (Euro-BioImaging)[4] and others rely on grid-computing and cloud-computing infrastructures for science. The purpose of these initiatives is to understand high-level phenotypes from genomic, metabolomic, proteomic, imaging and other types of data. They also require multi-scale mathematical models and simulations, as noted e.g. by Hunter et al. [29] in his strategy for Virtual Physiological Human (VPH)[5].

The integration with multidimensional models of geometrical, mechanical properties and the time-dependence of the compartment's data, which is taken from medical and biological repositories, can highly improve complex models of human physiology which are based mainly on lumped-parameter approach. E.g. Itu et al. achieved parameter identification on simplified windkessel model of hemodynamics in order to study aortic coarctation, which is based on processing of MRI, and requires 6-8 minutes of computation time on a standard personal computer [30]. One of the challenge of systems biology approach, as identified by Kohl et al. [31], is to use multiparameter perturbation to identify the safe areas, e.g., for multitarget drug profile. The results presented in section 4.3 shows that the parameter study can be done on much more complex models in a reasonable time. The computation is able to become practical for clinical and further research towards patient-specific health care, in silico trails and drug discovery.

Based on the previous answers, another research question can be formulated for further research in the technology domain:

How can biomedical research influence the direction of grid-computing and cloud-computing development?

One area of discussion about this theme is how to preserve scientific data in long term in order to prevent loss of them [32, 33]. Another area of discussion is how to facilitate access to computational resources for large amounts

[4] http://www.eurobioimaging.eu/ accessed March 2015
[5] http://www.vph-institute.org/ accessed March 2015

of small scientific group, which have limited resources to port, integrate or customize their current tools and processes – to support the "long-tail" of science. The "long-tail" movement was first noted and described by Anderson [34] in the business domain. The long-tail term comes from a feature of statistical distribution, e.g., pareto distribution, where only a few (e.g., 20% – noted as head) elements have a high probability of some events (e.g., product being sold), while the rest (e.g., 80% – noted as tail) have a small probability. Thus, most businesses focus on hits (20% of products, the 80-20 rule). The expansion of the Internet and its related technologies have caused reduced sales, marketing and delivery costs for the products from the niche (80% of products) – long-tail. A strategy that focused on these kinds of products became profitable and successful, e.g., for companies such as Amazon or Apple.[34]. Cloud computing technologies seem to be customizable and may be an enabling technology to focus on long-tail science, as noted e.g. by Weinhardt et al. [35]. How to facilitate and decrease an effort to develop, customize and port domain-specific application to some distributed computing model? This problem motivated, e.g., Anjum et al. to establish "platform as a service" (category of cloud computing service model) integrating several grid computing and cloud computing standards glueing via service oriented architecture approach [36]. Complementary approach is to support consultation, training and exchange in research software development toward the domain scientists, e.g., as presented by Crouch et al. regarding the Software Sustainable Institute within United Kingdom [37].

The provenance and reproducibility of scientific results implied a need to long-term preservation of scientific data, however, if it is left on individual researcher, there is loss of data, as analysed by Vines et al. or Heidorn [32, 33].

6. Conclusion

This thesis presents the infrastructure, which, thanks to virtualization technology, joined several domain-specific tools in the field of sharing and processing medical images, performing real-time voice analysis and simulating human physiology.

A seamless integration of grid-based PACS system was established with the current distributed system in order to share DICOM medical images. Access to real-time voice analysis application via remote desktop technology brings this type of service to any computer that can connect to the Internet. A system and portal to support the analysis and building of complex models of human physiology in the phase of parameter estimation and parameter sweep was introduced. Furthermore, additional computational nodes can be flexibly joined by starting the prepared virtual machines in cloud computing deployment.

The methodology of building complex models of human physiology was contributed with the use of acausal and object-oriented modeling techniques. Methods for conducting a parameter study were shown, as well as the parameter study of complex models that gain substantial speedup by utilizing cloud computing deployment, which makes such kinds of complex studies applicable in physiological and biological research.

Bibliography

[1] Ian Foster, Carl Kesselman, and Steven Tuecke. The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *International journal of high performance computing applications*, 15(3):200–222, 2001.

[2] Ian Foster, Carl Kesselman, Jeffrey M Nick, and Steven Tuecke. The Physiology of the Grid: An Open Grid Service Architecture for Distributed Systems Integration. *Grid computing: making the global infrastructure a reality*:217–249, 2003.

[3] I. Bird, B. Jones, and K.F. Kee. The Organization and Management of Grid Infrastructures. *Computer*, 42(1):36–46, January 2009. ISSN: 0018-9162. DOI: 10.1109/MC.2009.28.

[4] Dagmar Adamova. Computing Framework for the LHC: Current Status and Challenges of the High Luminosity LHC Future. In *52 international winter meeting on nuclear physics, bormio,italy*. (January), 2014. URL: http://pos.sissa.it/archive/conferences/212/021/Bormio2014_021.pdf.

[5] David P. Anderson, Jeff Cobb, Eric Korpela, Matt Lebofsky, and Dan Werthimer. SETI@home: An Experiment in Public-Resource Computing. *Communications of the acm*, 45(11):56–61, November 2002. ISSN: 00010782. DOI: 10.1145/581571.581573.

[6] DP Anderson. Boinc: A System for Public-resource Computing and Storage. In *Grid computing, 2004. proceedings. fifth ieee/acm international workshop on*. IEEE, 2004. DOI: 10.1109/GRID.2004.14.

[7] Peter Mell and Timothy Grance. The NIST Definition of Cloud Computing, 2011. URL: http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.

[8] Michael R. Garey and David S. Johnson. Computers and Intractability: A Guide to the Theory of NP-Completeness (Series of Books in the Mathematical Sciences). *Computers and intractability*, 1979.

[9] S. G. Erberich, M. Bhandekar, M. D. Nelson, A. Chervenak, and C. Kesselman. DICOM Grid Interface Service for Clinical and Research PACS: A Globus Toolkit Web Service for Medical Data Grids. *International journal of computer assisted radiology and surgery*, 1(ii):100–102, 2006. ISSN: 18616410. DOI: 10.1007/s11548-006-0013-0.

[10] K. Slavicek, O. Dostal, M. Javornik, and M. Drdla. MEDIMED - Regional Centre for Medicine Image Data Processing. *2010 third international conference on knowledge discovery and data mining*:310–313, January 2010. DOI: 10.1109/WKDD.2010.133.

[11] Stephan G Erberich, Jonathan C Silverstein, Ann Chervenak, Robert Schuler, Marvin D Nelson, and Carl Kesselman. Globus MEDICUS – Federation of DICOM Medical Imaging Devices into Healthcare Grids. *Studies in health technology and informatics*, 126:269–278, 2007. ISSN: 0926-9630.

[12] Marek Frič. Parametrizovaný fonetogram obecných řečových a hlasových projevů - ParVRP. Praha, 2007. URL: http://zvuk.hamu.cz/vyzkum/dokumenty/TL12x.pdf.

[13] Marek Frič, Tomáš Kulhánek, and Jaroslav Hrb. Systém pro vzdáleně přístupnou analýzu hlasu RealVoiceLab. Praha, 2012. URL: http://zvuk.hamu.cz/vyzkum/dokumenty/TL46x.pdf.

[14] A C Guyton, T G Coleman, and H J Granger. Circulation: Overall Regulation. *Annual review of physiology*, 34:13–46, 1972. ISSN: 0066-4278. DOI: 10.1146/annurev.ph.34.030172.000305.

[15] Robert L Hester, Radu Iliescu, Richard Summers, and Thomas G Coleman. Systems Biology and Integrative Physiological Modelling. *The journal of physiology*, 589(Pt 5):1053–60, March 2011. ISSN: 1469-7793. DOI: 10.1113/jphysiol.2010.201558.

[16] Robert L Hester, Alison J Brown, Leland Husband, Radu Iliescu, Drew Pruett, Richard Summers, and Thomas G Coleman. HumMod: A Modeling Environment for the Simulation of Integrative Human Physiol-

ogy. *Frontiers in physiology*, 2(April):12, January 2011. ISSN: 1664-042X. DOI: 10.3389/fphys.2011.00012.

[17] Erik Butterworth, Bartholomew E. Jardine, Gary M. Raymond, Maxwell L. Neal, and James B. Bassingthwaighte. JSim, an Open-source Modeling System for Data Analysis. *F1000research*, 2, May 2014. ISSN: 2046-1402. DOI: 10.12688/f1000research.2-288.v2.

[18] P J Hunter. The IUPS Physiome Project: A Framework for Computational Physiology. *Progress in biophysics and molecular biology*, 85(2-3):551–69, January 2004. ISSN: 0079-6107. DOI: 10.1016/j.pbiomolbio.2004.02.006.

[19] Tommy Yu, Catherine M Lloyd, David P Nickerson, Michael T Cooling, Andrew K Miller, Alan Garny, Jonna R Terkildsen, James Lawson, Randall D Britten, Peter J Hunter, and Poul M F Nielsen. The Physiome Model Repository 2. *Bioinformatics (oxford, england)*, 27(5):743–744, 2011. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btq723.

[20] Michael Hucka, ABBJ Finney, Benjamin J Bornstein, Sarah M Keating, Bruce E Shapiro, Joanne Matthews, Ben L Kovitz, Maria J Schilstra, Akira Funahashi, and John C Doyle. Evolving a Lingua Franca and Associated Software Infrastructure for Computational Systems Biology: The Systems Biology Markup Language (SBML) Project. *Systems biology*, 1(1):41–53, 2004.

[21] Nicolas Le Novère, Benjamin Bornstein, Alexander Broicher, Mélanie Courtot, Marco Donizelli, Harish Dharuri, Lu Li, Herbert Sauro, Maria Schilstra, Bruce Shapiro, Jacky L Snoep, and Michael Hucka. BioModels Database: A Free, Centralized Database of Curated, Published, Quantitative Kinetic Models of Biochemical and Cellular Systems. *Nucleic acids research*, 34:D689–D691, 2006. ISSN: 1362-4962. DOI: 10.1093/nar/gkj092.

[22] Jiří Kofránek and Jan Rusz. Restoration of Guyton´s Diagram for Regulation of the Circulation as a Basis For Quantitative Physiological

Model Development. *Physiological research / academia scientiarum bohemoslovaca*, 59(6):897–908, January 2010. ISSN: 1802-9973.

[23] Jiří Kofránek, Marek Mateják, and Pavol Privitzer. Hummod – Large Scale Physiological Models in Modelica. In *Proceedings 8th modelica conference, dresden, germany*, 2011, pages 713–724.

[25] M. Hofmann. On the Complexity of Parameter Calibration in Simulation Models. *The journal of defense modeling and simulation: applications, methodology, technology*, 2(4):217–226, October 2005. ISSN: 1548-5129. DOI: 10.1177/154851290500200405.

[26] T Blochwitz, M Otter, M Arnold, and C Bausch. The Functional Mockup Interface for Tool Independent Exchange of Simulation Models. In *Proceeding of 8th modelica conference, dresden*, 2011.

[27] Ann L. Chervenak, Theo G M Van Erp, Carl Kesselman, Mike D'Arcy, Janet Sobell, David Keator, Lisa Dahm, Jim Murry, Meng Law, Anton Hasso, Joseph Ames, Fabio MacCiardi, and Steven G. Potkin. A System Architecture for Sharing De-identified, Research-ready Brain Scans and Health Information Across Clinical Imaging Centers. *Studies in health technology and informatics*, 175:19–28, 2012. ISSN: 09269630. DOI: 10.3233/978-1-61499-054-3-19.

[28] Stanley Gotshall and Bart Rylander. Optimal population size and the genetic algorithm. *Proceedings on genetic and evolutionary computation conference*, 2000.

[29] Peter Hunter, Tara Chapman, Peter V Coveney, Bernard de Bono, Vanessa Diaz, John Fenner, Alejandro F Frangi, Peter Harris, Rod Hose, Peter Kohl, Pat Lawford, Keith McCormack, Miriam Mendes, Stig Omholt, Alfio Quarteroni, Nour Shublaq, John Skå r, Karl Stroetmann, Jesper Tegner, S Randall Thomas, Ioannis Tollis, Ioannis Tsamardinos, Johannes H G M van Beek, and Marco Viceconti. A Vision and Strategy for the Virtual Physiological Human: 2012 Update. *Interface focus*, 3(2), February 2013. URL: http://rsfs.royalsocietypublishing.org/content/3/2/20130004.abstract.

[30] Lucian Itu, Puneet Sharma, Kristóf Ralovich, Viorel Mihalef, Razvan Ionasec, Allen Everett, Richard Ringel, Ali Kamen, and Dorin Comaniciu. Non-invasive Hemodynamic Assessment of Aortic Coarctation: Validation with in Vivo Measurements. *Annals of biomedical engineering*, 41(4):669–681, 2013. ISSN: 00906964. DOI: 10.1007/s10439-012-0715-0.

[31] P Kohl, E J Crampin, T A Quinn, and D Noble. Systems Biology: An Approach. *Clinical pharmacology and therapeutics*, 88:25–33, 2010. ISSN: 1532-6535. DOI: 10.1038/clpt.2010.92.

[32] Timothy H. Vines, Arianne Y K Albert, Rose L. Andrew, Florence Débarre, Dan G. Bock, Michelle T. Franklin, Kimberly J. Gilbert, Jean Sébastien Moore, Sébastien Renaut, and Diana J. Rennison. The Availability of Research Data Declines Rapidly with Article Age. *Current biology*, 24:94–97, 2014. ISSN: 09609822. DOI: 10.1016/j.cub.2013.11.014. arXiv: 1312.5670.

[33] P. Bryan Heidorn. Shedding Light on the Dark Data in the Long Tail of Science. *Library trends*, 57(2):280–299, 2008. ISSN: 1559-0682. DOI: 10.1353/lib.0.0036.

[34] Chris Anderson. The Long Tail: How Endless Choice Is Creating Unlimited Demand. *Word journal of the international linguistic association*:256, 2006. ISSN: 02650487.

[35] Christof Weinhardt, Arun Anandasivam, Benjamin Blau, Nikolay Borissov, Thomas Meinl, Wibke Michalk, and Jochen Stöß er. Cloud Computing – A Classification, Business Models, and Research Directions. *Business & information systems engineering*, 1:391–399, 2009. ISSN: 1867-0202. DOI: 10.1007/s12599-009-0071-2.

[36] Ashiq Anjum, Richard Hill, Richard McClatchey, Nik Bessis, and Andrew Branson. Glueing Grids and Clouds Together: a Service-oriented Approach. *International journal of web and grid services*, 8(3):248–265, 2012. DOI: 10.1504/IJWGS.2012.049169.

[37] Stephen Crouch, Neil Chue Hong, Simon Hettrick, Mike Jackson, Aleksandra Pawlik, Shoaib Sufi, Les Carr, David De Roure, Carole Goble, and Mark Parsons. The Software Sustainability Institute: Changing Research Software Attitudes and Practices. *Computing in science and engineering*, 15(6):74–80, 2013. ISSN: 15219615. DOI: 10.1109/MCSE.2013.133.

Publication of the author

Related to the thesis

in IF Journals

1. Tomáš Kulhánek, Jiří Kofránek, and Marek Mateják. Modeling of Short-term Mechanism of Arterial Pressure Control in the Cardiovascular System: Object-oriented and Acausal Approach. *Computers in biology and medicine*, 54:137–144, September 2014. ISSN: 1879-0534. DOI: 10.1016/j.compbiomed.2014.08.025 **IF(2013)=1.475**

2. Marek Mateják, Tomáš Kulhánek, and Stanislav Matoušek. Adair-based Hemoglobin Equilibrium with Oxygen, Carbon Dioxide and Hydrogen Ion Activity. *Scandinavian journal of clinical & laboratory investigation*, 2014. DOI: 10.3109/00365513.2014.984320 **IF(2013)=2.009**

Other publications

3. Tomáš Kulhánek and Milan Šárek. Processing of medical images in virtual distributed environment. In *Proceedings of the 2009 euro american conference on telematics and information systems new opportunities to increase digital citizenship - eatis '09*. ACM, 2009, pages 1–3. ISBN: 9781605583983. DOI: 10.1145/1551722.1551732

4. Tomáš Kulhánek, Marek Frič, and Milan Šárek. Remote Analysis of Human Voice – Lossless Sound Recording Redirection. *Analysis of biomedical signals and images. proceedings of 20th international eurasip conference (biosignal)*:394–397, 2010. URL: http://bs2010.biosignal.cz/papers/1092.pdf

5. Tomáš Kulhánek. Infrastructure for Data Storage and Computation in Biomedical Research. *European journal for biomedical informatics (ejbi)*, 6(1):55–58, 2010. ISSN: 1801-5603. URL: http://www.ejbi.org/img/ejbi/ejbi2010-1.pdf

6. Tomáš Kulhánek, Marek Mateják, Jan Šilar, and Jiří Kofránek. Parameter Estimation of Complex Mathematical Models of Human Physiology Using Remote Simulation Distributed in Scientific Cloud. In

Biomedical and health informatics (bhi), 2014 ieee-embs international conference on, 2014, pages 712–715. ISBN: 9781479921317. DOI: 10.1109/BHI.2014.6864463

7. Tomáš Kulhánek, Martin Tribula, Jiří Kofránek, and Marek Mateják. Simple Models of the Cardiovascular System for Educational and Research Purposes. *Mefanet journal*, 2(2):56–63, 2014

8. Tomáš Kulhánek, J Šilar, Marek Mateják, Pavol Privitzer, Jiří Kofranek, and Martin Tribula. Distributed Computation and Parameter Estimation in Identification of Physiological Systems. *Vph conference*, 2010

9. Tomáš Kulhánek, Filip Ježek, Marek Mateják, Jan Šilar, Pavol Privitzer, Martin Tribula, and Jiří Kofránek. RESTful Web Service to Build Loosely Coupled Web Based Simulation of Human Physiology. *Transactions of japanese society for medical and biological engineering*, 51(Supplement):R–32, 2013. DOI: 10.11239/jsmbe.51.R-32

10. Tomáš Kulhánek and Milan Šárek. Virtualizace a integrace v gridovém pacs systému (cze) virtualization and integration in grid pacs system (eng). In *Mefanet 2008*. MSD Brno, 2008. URL: http://www.mefanet.cz/res/file/mefanet2008/prispevky/25_kulhanek.pdf

11. Milan Šárek and Tomáš Kulhánek. Nové směry medicínských aplikací sdružení CESNET. in *Sborník příspěvků medsoft*, 2009, pages 145–148

12. Tomáš Kulhánek. Virtual Distributed Environment for Exchange of Medical Images. In *Proceeding of phd. conference 2009 was published as: doktorandské dny '09*. D Kuželová, editor. Institute of Computer Science/MatfyzPress, 2009, pages 62–64. URL: http://www.cs.cas.cz/hakl/doktorandsky-den/files/2009/sbornik-dd-2009.pdf

13. Tomáš Kulhánek, Marek Frič, and Milan Šárek. Vzdálený přístup k virtuálním výukovým a výzkumným aplikacím - podpora foniatrických vyšetření. In *Mefanet 2009*. Daniel Schwarz, Jaroslav Majerník, Ladislav Dušek, Stanislav Štípek, and Vladimír Mihál, editors. MSD Brno, 2009. URL: http://www.mefanet.cz/res/file/mefanet2009/prispevky/kulhanek_full.pdf

14. Tomáš Kulhánek, Marek Frič, and Milan Šárek. Vzdálená analýza lidského hlasu - bezeztrátové nahrávání zvuku přes ip sítě. In *Sborník příspěvků medsoft*, 2010, pages 96–101

15. Tomáš Kulhánek, Jan Šilar, Jiří Kofránek, Marek Mateják, Pavol Privitzer, and Martin Tribula. Od výukového modelu k identifikaci fyziologického systému. In *Mefanet 2010*. Daniel Schwarz, Martin Komenda, Jaroslav Majerník, Stanislav Štípek, Vladimír Mihál, and Ladislav Dušek, editors. MSD Brno, 2010. URL: http://www.mefanet.cz/res/file/mefanet2010/prispevky/kulhanek-full.pdf

16. Tomáš Kulhánek. From Educational Models Towards Identification of Physiological Systems. *Mefanet report*, 04:69–72, 2011. URL: http://www.mefanet.cz/res/file/reporty/mefanet-report-2011.pdf

17. Tomáš Kulhánek. Infrastructure for Data Storage and Computation in Biomedical Research. In *Proceeding of phd. conference 2011 was published as: doktorandské dny '11*. D Kuželová, editor. Institute of Computer Science/MatfyzPress, 2011, pages 90–94. URL: http://www.cs.cas.cz/hakl/doktorandsky-den/files/2011/sbornik-dd-2011.pdf

18. Tomáš Kulhánek, Marek Frič, and Jaroslav Hrb. Vzdálená analýza lidského hlasu v reálném čase. In *Sborník příspěvků medsoft*, 2012, pages 180–184. URL: http://www.creativeconnections.cz/medsoft/2012/Medsoft_Kulh\'{a}nek_Tom\'{a}\v{s}.pdf

19. Tomáš Kulhánek, Marek Mateják, Jan Šilar, Pavol Privitzer, Martin Tribula, Filip Ježek, and Jiří Kofránek. Hybridní architektura pro webové simulátory (cze) Hybrid architecture for web simulators (en). In *Sborník příspěvků medsoft*, 2013, pages 115–121. URL: http://www.creativeconnections.cz/medsoft/2013/Medsoft_2013_Kulhanek.pdf

20. Tomáš Kulhánek, Marek Mateják, Jan Šilar, and Jiří Kofránek. Identifikace fyziologických systémů. In *Sborník příspěvků medsoft*, 2014, pages 148–153. URL: http://www.creativeconnections.cz/medsoft/2014/Medsoft_2014_Kulhanek.pdf

Without relationship to the thesis

21. Jiří Kofránek, Marek Mateják, Pavol Privitzer, Martin Tribula, Tomáš Kulhánek, Jan Šilar, and Rudolf Pecinovský. HumMod-Golem Edition: Large Scale Model of Integrative Physiology for Virtual Patient Simulators. In *Proceedings of world congress in computer science 2013 (worldcomp'13), international conference on modeling, simulation and visualisation methods (msv'13)*, 2013, pages 182–188

22. Marek Mateják, Tomáš Kulhánek, Jan Šilar, Pavol Privitzer, Filip Ježek, and Jiří Kofránek. Physiolibrary - Modelica Library for Physiology. In *Proceedings of the 10th international modelica conference, march 10-12, 2014, lund, sweden*. Linköping University Electronic Press, 2014, pages 499–505. DOI: 10.3384/ecp14096499

23. Jiří Kofránek, Pavol Privitzer, Martin Tribula, Jan Šilar, Tomáš Kulhánek, Marek Mateják, and Filip Ježek. Virtual Patient Simulator, CZ Utility model. Application number: 2014-30329. Issue number: 27613. 2014. URL: http://isdv.upv.cz/portal/pls/portal/portlets.pts.det?xprim=10082946