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.

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

2. 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

3. 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

4. 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

13. 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

14. 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

15. 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

16. 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

17. 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

18. 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

19. Tomáš Kulhánek and Marek Frič. Remote Analysis of Human Voice - Environment for Voice Training and ORL Medicine. In *Egi technical forum, prague*. Prague, 2012

Without relationship to the thesis

1. Jiří Kofránek, Marek Mateják, Pavol Privitzer, Martin Tribula, Tomáš Kulhánek, Jan Šilar, and Rudolf Pecinovský. HumMod-Golem Edition:

Contents

Abstrakt (česky) 4

Abstract 5

List of Abbreviations 6

1 Introduction 7

2 Hypothesis 7

3 Methods 8

4 Results 9

5 Discussion 9

Publication of the author 20

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. Gridový PACS systém byl propojen s existujícími distribuovanými systémy pro sdílení DICOM snímků. 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. Bylo přispěno do modelovací technologie pro tvorbu komplexních modelů založených na akauzální a objektově orientovaném modelovacím přístupu. Byli představeny další metody pro studium parametrů fyziologického systému a portál pro podporu tvorby a analýzy komplexních modelů fyziologie člověka ve fázi odhadu parametrů modelů. Studium parametrů v komplexních modelech získá významné zrychlení pokud se výpočet nasadí do tzv. cloud computingu. Takové nasazení aplikovatelné ve fyziologickém a biologickém výzkumu má potenciál zlepšit toto použití i ve zdravotní péči.

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

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

5. 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

6. 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

7. 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

8. 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

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

10. 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

11. 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

12. 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

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 Γ"026e30f& laboratory investigation*, 2014. DOI: 10.3109/00365513.2014.984320 **IF(2013)=2.009**

Other publications

1. 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

2. 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

3. 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

4. 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

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. 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 was introduced using customized protocol to transfer sound recording. This brings a possibility to access remotely current legacy application. The modeling methodology was contributed in order to build complex models based on acausal and object-oriented modeling techniques. Methods for conducting a parameter study were shown, especially parameter estimation and parameter sweep. Parameter study of complex models gain substantial speedup by utilizing cloud computing deployment, which makes such kinds of complex studies applicable in physiological and biological research and have potential to improve such usage in healthcare.

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

List of Abbreviations

CVS Cardiovascular System, page 21

DFT Discrete Fourier Transformation, page 16

DICOM Digital Imaging and Communication Protocol, page 11

FFT Fast Fourier Transformation, page 16

FMI Functional Mockup Interface, page 24

FMU Functional Mockup Unit, page 24

FTP File Transfer Protocol, page 14

RDP Remote Desktop Protocol, page 17

SSL Secure Sockets Layer, page 13

VPN Virtual Private Network, page 12

VRP Voice Range Profile, page 16

XML Extensible Markup Language, page 24

[14] 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.

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

[16] 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* Γ"026e30f*& information systems engineering*, 1:391–399, 2009. ISSN: 1867-0202. DOI: 10.1007/s12599-009-0071-2.

[17] 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.

[18] 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.

[9] Kristin Tø ndel, Ulf G Indahl, Arne B Gjuvsland, Jon Olav Vik, Peter Hunter, Stig W Omholt, and Harald Martens. Hierarchical Cluster-based Partial Least Squares Regression (HC-PLSR) is an Efficient Tool for Metamodelling of Nonlinear Dynamic Models. *Bmc systems biology*, 5(1):90, 2011. ISSN: 1752-0509. DOI: 10.1186/1752-0509-5-90. URL: http://www.biomedcentral.com/1752-0509/5/90.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

1. Introduction

Grid computing is usually defined as sharing computational and data storage resources across organizational boundaries. 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*, which also utilizes a virtual environment on powerful computing infrastructures. 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. It discusses its utilization in biomedical research and its application in relation to the processing of medical information.

The term "medical information" is too broad and further work in this thesis focuses on the following selected areas: (1) the exchange and processing of medical images, (2) the analysis of human voice and (3) the modeling and simulation of human physiology.

2. Hypothesis

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. 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?* When work on this thesis begun, grid computing was believed to be an answer to scalability issues, e.g., for exchanging large amounts of data or carrying out demanding long-term computation.
- *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 and may significantly decrease the time of computation of the problem. Additionally, it can increase the size of solvable problems in a

Bibliography

[1] 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.

[3] 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.

[4] 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.

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

[6] 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.

[7] 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.

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

any scientific domain.

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.

reasonable time. It was shown that those algorithms that suffer with exponential time complexity does not benefit much from technology speedup. The size of problems, which are computable by exponential algorithm in a reasonable time, will increase only slightly when speedup is introduced [1]. Therefore for such type of problems, additional non-exact methods are used including heuristic (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. Additionally, some the algorithm is not scalable and, therefore, cannot gain additional speedup when executed on more CPUs in parallel.

The methods used in further work was to identify a key

4. Results

In previous chapters, there were introduced different methods available for selected use cases in research in biology and medicine. As each of the use cases and available system was proposed on different operating system platform, different architecture and or different middleware the virtualization was utilized to build the virtual infrastructures for purposes of each project. The paper [2] *Infrastructure for Data Storage and Computation in Biomedical Research* in Appendix **??** describes result of establishing the virtualization on physical infrastructure to share computational power among different platforms.

5. Discussion

The result presented in section **??** is an example how a standard format and protocol DICOM is utilized to integrate current production system in order to exchange medical images (MEDIMED [3]) and a grid-based solution (Globus MEDICUS[4]). Remote Desktop Protocol (RDP) is a key standard for protocol in order to integrate the application of voice analysis[5] into a

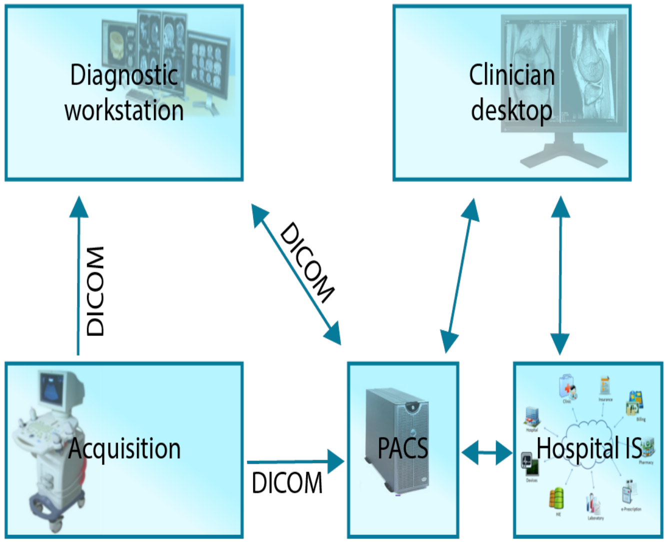

Figure 3.1: The typical workflow of a medical image in a hospital. Data acquisition is made by modalities (magnetic resonance, ultrasonography, X-ray radiography, etc.). By using the DICOM format and protocol, it can be directly transferred and visualized by diagnostic workstation. With the metadata filled by an expert physician, the image is stored in PACS. Other desktops within the hospital can retrieve the image and review the report. The hospital information system may be involved in other workflows and communicate with other formats and standards (e.g., HL7).

remote environment, which is accessible via the Internet. This is presented in section **??**. In the case of parameter estimation, a key factor is the standard Functional Mockup Interface (FMU)[6], which allows the control and simulation of a physiological model in a customized tool that is not related to modeling tool. This is presented in section **??**.

The selection of a joint element increases the chances of reusability of such a system in future development, when requirements usually change and the reconstruction of a system or architecture is needed. For example,

based on already collected records in history.

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 [13, 14]. Another area of discussion is how to facilitate access to computational resources for large amounts 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 [15] 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.[15]. 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. [16]. 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 [17]. 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 [18]. The ideas above applies for potential influence and requirements given by tructure for Biological Information (ELIXIR)[3], European Biomedical Imaging Infrastructure (Euro-BioImaging)[4] and others, which technologically 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. [10] 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 [11]. On of the challenge of systems biology approach, as identified by Kohl et al. [12], is to use multiparameter perturbation to identify the safe areas, e.g., for multitarget drug profile. The results presented in section **??** 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.

Additionally, it is a business strategy of several new ventures in order to collect anonymized patient records from clinicians, to gradually improve diagnostic methods and to provide reciprocal services for supporting clinical or therapeutic decision, as presented in section **??**. For example, Fetview[6] is an startup company, which one aim is to support fetal healthcare and gradually improve diagnostic methods, which is

[3] http://www.elixir-europe.org/ accessed March 2015
[4] http://www.eurobioimaging.eu/ accessed March 2015
[5] http://www.vph-institute.org/ accessed March 2015
[6] http://fetview.com/ accessed March 2015

CHAPTER 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. Thus, in their further development, the authors of Globus MEDICUS followed a way of federation of medical images that are stored within home institutions, as published by Chervenak et al. [7]. 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 **??**. Such service can be deployed on any web server and the occasional need to educate or perform a higher number of analysis concurrently can be satisfied with cloud computing deployment. The application process for sound signal which is currently analyzed by Fast Fourier Transformation algorithm quite effectively. Another challenge is to analyze a sound signal connected with high-speed video or videokymography methods, which need to transfer, process and store larger amounts of data.

In the case of the application for parameter estimation presented in section **??**, 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 [8].

The parameter sweep problem is considered as embarrasingly parallel and highly suitable for high throughput computing (HTC), which is the main focus of current grid computing infrastructures. Tøndel et al. introduced methods to statistically map variation of large number of parameters and to reduce drastically the number of simulations required for parameter study [9], however only non-complex models focusing on specific phenomenon were considered.

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 **??**) and for parameter estimation and parameter sweep, which was tried with the desktop grid approach - BOINC framework (section **??**). Virtualization facilitates the integration effort, as presented in the case of remote analysis of the human voice (section **??**) and in thecase of deployment of worker nodes in a cloud computing environment for parameter estimation (section **??**).

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

Based on previous results and ideas, the answer to the questions from the section **??** can be formulated:

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

  Grid computing and cloud computing may significantly speedup parameter study of medium and complex models in computational physiology. Such a speedup might 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.

  From a programming model point of view, limitation are given by the theoretical features of algorithms and the problems to be solved. Grid computing and cloud computing are not general solutions for hard problems (NP-complete problems), as discussed in section **??**. However, 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 research infrastructures, e.g. Integrated Structural Biology Infrastructure for Europe (INSTRUCT)[2], European Life Science Infras-

[2] https://www.structuralbiology.eu/ accessed March 2015