

19. Tomáš Kulhánek et al. "Hybridní architektura pro webové simulátory (cze) Hybrid architecture for web simulators (en)". In: *sborník příspěvků MEDSOFT*. 2013, pp. 115–121. URL: http://www.creativeconnections.cz/medsoft/2013/Medsoft_2013_Kulhanek.pdf
20. Tomáš Kulhánek et al. "Identifikace fyziologických systémů". In: *sborník příspěvků MEDSOFT*. 2014, pp. 148–153. URL: http://www.creativeconnections.cz/medsoft/2014/Medsoft_2014_Kulhanek.pdf

Without relationship to the thesis

21. Marek Mateják et al. "Physiolibrary - Modelica Library for Physiology". In: *Proceedings of the 10th International Modelica Conference, March 10-12, 2014, Lund, Sweden*. Linköping University Electronic Press, 2014, pp. 499–505. DOI: [10.3384/ecp14096499](https://doi.org/10.3384/ecp14096499)
22. Jiří Kofránek et al. "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, pp. 182–188
23. Jiří Kofránek et al. *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>

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 biomedicině*Univerzita Karlova v Praze a Akademie věd České republiky*

Obor: Biomedicínská informatika

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

Školící 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.

8. Tomáš Kulhánek et al. "Distributed Computation and Parameter Estimation in Identification of Physiological Systems". In: *Proceedings of VPH Conference* (2010)
9. Tomáš Kulhánek et al. "RESTful Web Service to Build Loosely Coupled Web Based Simulation of Human Physiology". In: *Transactions of Japanese Society for Medical and Biological Engineering* 51.Supplement (2013), R–32. doi: [10.11239/jsmbe.51.R-32](https://doi.org/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, pp. 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*. Ed. by D Kuželová. Institute of Computer Science/MatfyzPress, 2009, pp. 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*. Ed. by Daniel Schwarz et al. 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 - bezztrátové nahrávání zvuku přes ip sítě". In: *sborník příspěvků MEDSOFT*. 2010, pp. 96–101
15. Tomáš Kulhánek et al. "Od výukového modelu k identifikaci fyziologického systému". In: *Mefanet 2010*. Ed. by Daniel Schwarz et al. 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". In: *Mefanet Report 04* (2011), pp. 69–72. 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*. Ed. by D Kuželová. Institute of Computer Science/MatfyzPress, 2011, pp. 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, pp. 180–184. URL: http://www.creativeconnections.cz/medsoft/2012/Medsoft_Tom_kulhanek_Tom_v_s.pdf

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." In: *Computers in Biology and Medicine* 54 (Sept. 2014), pp. 137–144. ISSN: 1879-0534. DOI: [10.1016/j.combiomed.2014.08.025](https://doi.org/10.1016/j.combiomed.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". In: *Scandinavian Journal of Clinical & Laboratory Investigation* (2014). DOI: [10.3109/00365513.2014.984320](https://doi.org/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, pp. 1–3. ISBN: 9781605583983. DOI: [10.1145/1551722.1551732](https://doi.org/10.1145/1551722.1551732)
4. Tomáš Kulhánek, Marek Frič, and Milan Šárek. "Remote Analysis of Human Voice – Lossless Sound Recording Redirection". In: *Analysis of Biomedical Signals and Images. Proceedings of 20th International EURASIP Conference (BIOSIGNAL)* (2010), pp. 394–397. URL: <http://bs2010.biosignal.cz/papers/1092.pdf>
5. Tomáš Kulhánek. "Infrastructure for Data Storage and Computation in Biomedical Research". In: *European Journal for Biomedical Informatics (EJBI)* 6.1 (2010), pp. 55–58. ISSN: 1801-5603. URL: <http://www.ejbi.org/img/ejbi/ejbi2010-1.pdf>
6. Tomáš Kulhánek et al. "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, pp. 712–715. ISBN: 9781479921317. DOI: [10.1109/BHI.2014.6864463](https://doi.org/10.1109/BHI.2014.6864463)
7. Tomáš Kulhánek et al. "Simple Models of the Cardiovascular System for Educational and Research Purposes". In: *MEFANET Journal* 2.2 (2014), pp. 56–63

Contents

Abstrakt (česky)	4
Abstract	5
List of Abbreviations	6
1 Introduction	7
2 Hypothesis	8
3 Methods	10
4 Results	14
5 Discussion	20
6 Conclusion	25
Bibliography	26
Publication of the author	30

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. Práce se zabývá oblastí výměny medicínských snímků a představuje propojení Gridového PACS systému s existujícími distribuovanými systémy pro sdílení DICOM snímků. Práce se dále zaměřuje na studium vědy týkající se lidského hlasu. Práce představuje vzdálený způsob přístupu k aplikaci pro analýzu hlasu v reálném čase pomocí úpravy protokolů pro vzdálenou plochu a pro přenos zvukových nahrávek. Tento dílčí výsledek ukazuje možnost využití stávajících aplikací na dálku specialisty na hlas.

Oblast lidské fyziologie a patofyziologie byla studována pomocí přístupu tzv. systémové biologie. Práce přispívá v oblasti metodologie modelování lidské fyziologie pro tvorbu komplexních modelů založených na akauzálním a objektově orientovaném modelovacím přístupu. Metody pro studium parametrů byly představeny pomocí technologie počítání v gridech a v cloudech. Práce ukazuje, že proces identifikaci parametrů středně komplexních modelů kardiovaskulárního 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 do budoucna.

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

BIBLIOGRAPHY

- [27] P Kohl et al. "Systems Biology: An Approach." In: *Clinical Pharmacology and Therapeutics* 88 (2010), pp. 25–33. ISSN: 1532-6535. DOI: [10.1038/clpt.2010.92](https://doi.org/10.1038/clpt.2010.92).
- [28] Timothy H. Vines et al. "The Availability of Research Data Declines Rapidly with Article Age". In: *Current Biology* 24 (2014), pp. 94–97. ISSN: 09609822. DOI: [10.1016/j.cub.2013.11.014](https://doi.org/10.1016/j.cub.2013.11.014). arXiv: [1312.5670](https://arxiv.org/abs/1312.5670).
- [29] P. Bryan Heidorn. "Shedding Light on the Dark Data in the Long Tail of Science". In: *Library Trends* 57.2 (2008), pp. 280–299. ISSN: 1559-0682. DOI: [10.1353/lib.0.0036](https://doi.org/10.1353/lib.0.0036).
- [30] Chris Anderson. "The Long Tail: How Endless Choice Is Creating Unlimited Demand". In: *World Journal Of The International Linguistic Association* (2006), p. 256. ISSN: 02650487.
- [31] Christof Weinhardt et al. "Cloud Computing – A Classification, Business Models, and Research Directions". In: *Business & Information Systems Engineering* 1 (2009), pp. 391–399. ISSN: 1867-0202. DOI: [10.1007/s12599-009-0071-2](https://doi.org/10.1007/s12599-009-0071-2).
- [32] Ashiq Anjum et al. "Glueing Grids and Clouds Together: a Service-oriented Approach". In: *International Journal of Web and Grid Services* 8.3 (2012), pp. 248–265. DOI: [10.1504/IJWGS.2012.049169](https://doi.org/10.1504/IJWGS.2012.049169).
- [33] Stephen Crouch et al. "The Software Sustainability Institute: Changing Research Software Attitudes and Practices". In: *Computing in Science and Engineering* 15.6 (2013), pp. 74–80. ISSN: 15219615. DOI: [10.1109/MCSE.2013.133](https://doi.org/10.1109/MCSE.2013.133).

- [18] M. Hofmann. "On the Complexity of Parameter Calibration in Simulation Models". In: *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology* 2.4 (Oct. 2005), pp. 217–226. ISSN: 1548-5129. DOI: [10.1177/154851290500200405](https://doi.org/10.1177/154851290500200405).
- [19] T Blochwitz et al. "The Functional Mockup Interface for Tool Independent Exchange of Simulation Models". In: *Proceeding of 8th Modelica Conference, Dresden*. 2011.
- [20] Javier Fernandez de Canete et al. "Object-oriented Modeling and Simulation of the Closed Loop Cardiovascular System by Using SIM-SCAPE." In: *Computers in biology and medicine* 43.4 (May 2013), pp. 323–33. ISSN: 1879-0534. DOI: [10.1016/j.compbiomed.2013.01.007](https://doi.org/10.1016/j.compbiomed.2013.01.007).
- [21] W van Meurs. *Modeling and Simulation in Biomedical Engineering: Applications in Cardiorespiratory Physiology*. McGraw-Hill Professional, 2011.
- [23] Ann L. Chervenak et al. "A System Architecture for Sharing De-identified, Research-ready Brain Scans and Health Information Across Clinical Imaging Centers". In: *Studies in Health Technology and Informatics* 175 (2012), pp. 19–28. ISSN: 09269630. DOI: [10.3233/978-1-61499-054-3-19](https://doi.org/10.3233/978-1-61499-054-3-19).
- [24] Stanley Gotshall and Bart Rylander. "Optimal population size and the genetic algorithm". In: *Proceedings On Genetic And Evolutionary Computation Conference* (2000).
- [25] Peter Hunter et al. "A Vision and Strategy for the Virtual Physiological Human: 2012 Update". In: *Interface Focus* 3.2 (Feb. 2013). URL: <http://rsfs.royalsocietypublishing.org/content/3/2/20130004.abstract>.
- [26] Lucian Itu et al. "Non-invasive Hemodynamic Assessment of Aortic Coarctation: Validation with in Vivo Measurements". In: *Annals of Biomedical Engineering* 41.4 (2013), pp. 669–681. ISSN: 00906964. DOI: [10.1007/s10439-012-0715-0](https://doi.org/10.1007/s10439-012-0715-0).

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
DICOM	Digital Imaging and Communication Protocol, page 10
EGI	European Grid Infrastructure, page 7
FLOPS	Floating-Point Operations per Second, page 8
FMI	Functional Mockup Interface, page 14
FMU	Functional Mockup Unit, page 14
HPC	High Performance Computing, page 8
HTC	High Throughput Computing, page 8
IaaS	Infrastructure as a Service, page 8
MTC	Many Task Computing, page 8
NGI	National Grid Initiative, page 7
NP-complete	Nondeterministic Polynomial - complete, page 10
PACS	Picture Archiving and Communication System, page 11
RDP	Remote Desktop Protocol, page 11
WLCG	Worldwide Large Hadron Collider Computing Grid, page 7
XML	Extensible Markup Language, page 14

BIBLIOGRAPHY

- [10] Ioan Raicu, Ian T. Foster, and Yong Zhaoyo. “Many-task computing for grids and supercomputers”. In: *2008 Workshop on Many-Task Computing on Grids and Supercomputers, MTAGS 2008*. 2008. ISBN: 9781424428724. DOI: [10.1109/MTAGS.2008.4777912](https://doi.org/10.1109/MTAGS.2008.4777912).
- [11] 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)”. In: *Computers and Intractability* (1979).
- [12] S. G. Erberich et al. “DICOM Grid Interface Service for Clinical and Research PACS: A Globus Toolkit Web Service for Medical Data Grids”. In: *International Journal of Computer Assisted Radiology and Surgery* 1.ii (2006), pp. 100–102. ISSN: 18616410. DOI: [10.1007/s11548-006-0013-0](https://doi.org/10.1007/s11548-006-0013-0).
- [13] K. Slavicek et al. “MEDIMED - Regional Centre for Medicine Image Data Processing”. In: *2010 Third International Conference on Knowledge Discovery and Data Mining* (Jan. 2010), pp. 310–313. DOI: [10.1109/WKDD.2010.133](https://doi.org/10.1109/WKDD.2010.133).
- [14] Stephan G Erberich et al. “Globus MEDICUS – Federation of DICOM Medical Imaging Devices into Healthcare Grids.” In: *Studies in Health Technology and Informatics* 126 (2007), pp. 269–278. ISSN: 0926-9630.
- [15] Marek Frič. *Parametrizovaný fonetogram obecných řečových a hlasových projevů - ParVRP*. Praha, 2007. URL: <http://zvuk.hamu.cz/vyzkum/dokumenty/TL12x.pdf>.
- [16] Marek Frič, Tomáš Kulháněk, and Jaroslav Hrb. *Systém pro vzdáleně přístupnou analýzu hlasu RealVoiceLab*. Praha, 2012. URL: <http://zvuk.hamu.cz/vyzkum/dokumenty/TL46x.pdf>.
- [17] Jiří Kofránek, Marek Mateják, and Pavol Privitzer. “Hummod – Large Scale Physiological Models in Modelica”. In: *Proceedings 8th Modelica Conference, Dresden, Germany*. 2011, pp. 713–724.

Bibliography

- [1] Ian Foster, Carl Kesselman, and Steven Tuecke. "The Anatomy of the Grid: Enabling Scalable Virtual Organizations". In: *International Journal of High Performance Computing Applications* 15.3 (2001), pp. 200–222.
- [2] Ian Foster et al. "The Physiology of the Grid: An Open Grid Service Architecture for Distributed Systems Integration". In: *Grid Computing: Making the Global Infrastructure a Reality* (2003), pp. 217–249.
- [3] I. Bird, B. Jones, and K.F. Kee. "The Organization and Management of Grid Infrastructures". In: *Computer* 42.1 (Jan. 2009), pp. 36–46. ISSN: 0018-9162. DOI: [10.1109/MC.2009.28](https://doi.org/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 et al. "SETI@home: An Experiment in Public-Resource Computing". In: *Communications of the ACM* 45.11 (Nov. 2002), pp. 56–61. ISSN: 00010782. DOI: [10.1145/581571.581573](https://doi.org/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](https://doi.org/10.1109/GRID.2004.14).
- [7] Peter Mell and Timothy Grance. "The NIST Definition of Cloud Computing". In: (2011). URL: <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>.
- [8] Georg Hager and Gerhard Wellein. *Introduction to High Performance Computing for Scientists and Engineers*. 2010, p. 356. ISBN: 9781439811924. DOI: [10.1201/EBK1439811924](https://doi.org/10.1201/EBK1439811924).
- [9] John Levesque and Gene Wagenbreth. *High Performance Computing: Programming and Applications*. CRC Press, 2010, p. 244. ISBN: 1420077066.

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 initiatives 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.

Applications that are computed within a grid or cloud infrastructure can be characterized by the quantity of tasks being performed, the size of the input data and the communication that needs to be carried out between concurrent tasks. And using such characterization three main categories of the application model are recognized.

- The term High Throughput Computing (HTC) is used for computation in which tasks take a long time. These are relatively loosely coupled and resources are used over a long period of time. Performance or capacity is usually mentioned in operations or CPUs per month or year. Grid computing focus mainly on HTC.
- The High Performance Computing (HPC) is usually characterized as a small number of tasks which need to communicate quite often. The tasks are relatively tightly coupled and can take shorter time than HTC. Performance is measured in operations per second (FLOPS) [8, 9]. The grid computing or cloud computing infrastructure can involve HPC servers or clusters.
- Many Task Computing (MTC) aims to bridge HTC and HPC. While the computation usually takes a shorter amount of time, the data exchange is in MB rather than in GB and it involves computing much more heterogeneous problems, which are not "happily" parallel [10].

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.

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. The grid-based solution brings robustness against the mentioned problems.

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. This connects voice therapists and voice pedagogues in different areas of the Czech Republic and Slovakia to analyze the voice in non-invasive way and to see e.g. the progress of the voice training methods.

A system and portal was introduced in order to support the analysis and building of complex models of human physiology in the phase of parameter estimation and parameter sweep. Furthermore, additional computational nodes can be joined flexibly by starting the prepared virtual machines in cloud computing deployment.

The methodology of building complex models of human physiology was contributed with the recommendation and implementation 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. This makes such kinds of complex studies applicable in physiological and biological research in future.

One area of discussion about this theme is how to preserve scientific data in long term in order to prevent loss of them [28, 29]. 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 [28, 29].

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 [30] 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.[30]. 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. [31].

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 [32]. 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 [33].

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 are:

- 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 [11]. 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 standard¹ was used as a joint protocol to integrate grid-based system Globus MEDICUS [12] with current production system MEDIMED for sharing medical images among different hospitals [13], which is based on common distributed system with central cluster. Globus MEDICUS [12, 14] implements a DICOM Grid Interface Service (DGIS) and integrates the open-source PixelMed™ Java DICOM Toolkit² into a web service, communicating via the DICOM protocol. Furthermore, it forwards queries to the

¹DICOM: <http://dicom.nema.org/> accessed January 2015

²<http://www.pixelmed.com/> accessed February 2015

infrastructure for Europe (INSTRUCT)², European Life Science Infrastructure for Biological Information (ELIXIR)³, European Biomedical 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. [25] 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 [26]. One of the challenge of systems biology approach, as identified by Kohl et al. [27], 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 trials 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?

²<https://www.structuralbiology.eu/> accessed March 2015

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

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

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 In-

underlying services within Globus toolkit. The console application of the MEDIMED project can be interconnected via DICOM protocol with local Picture Archiving and Communication System (PACS) 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. [15, 16]. Its implementation is carried out in an MATLAB environment, utilizing Signal Processing Toolbox³. 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 servers exist for several other platforms. RDP itself contains the redirection of several services, e.g., audio, sound recording, drive access, etc.

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

³<http://www.mathworks.com/products/signal/> accessed February 2015

Figure 3.1.

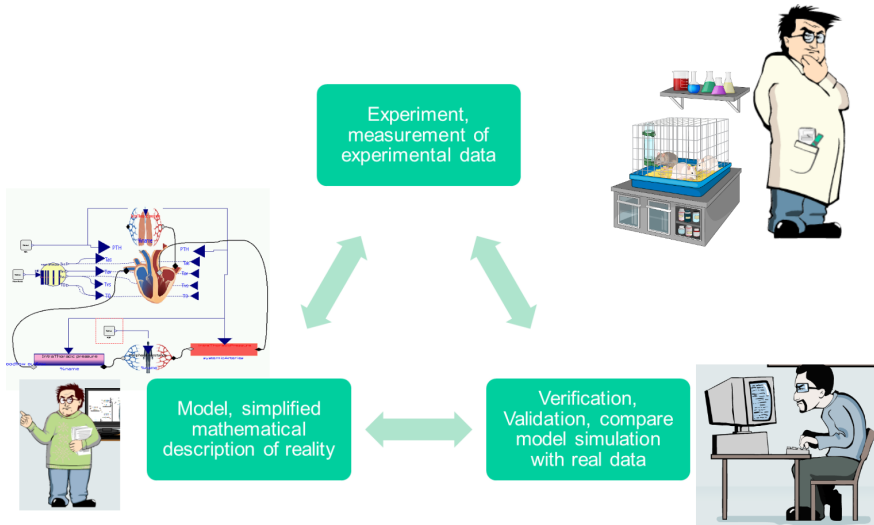


Figure 3.1: Schematic illustration of the scientific process. The experiments produce data that are interpreted and a hypothesis is formalized as a model. Validation compares the model simulation with the experiment, if the model satisfies the criteria – if it is in agreement with real experiments – 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 chosen, because it was identified as robust to maintain most complex models of human physiology in understandable way [17]. Additionally, a library Physiobrary developed by Matejak et al. (21.) was enhanced and used especially within the hydraulic domain components as listed in Table 3.1.

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 [18], which implies that the best known exact algo-

When porting an application to a grid environment, one of the important decision is the platform of the used system, which is sometimes hard to involve, e.g., within the thesis there were available code from a third party tool for the MS Windows platform only. This can determine the platform of the worker node and the virtualization - or, in the case of parameter estimation, cloud computing is utilized on a prepared platform. In the case of parameter sweep, a desktop grid computing BOINC worker and application for a MS Windows platform is 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. An option can be to use WINE¹ – a compatibility layer that is capable of running Windows applications on several POSIX-compliant operating systems, such as Linux, Mac OSX and BSD. This would allow utilizing traditional service grid infrastructure for, e.g., parameter sweep application having an advantage not to maintain the desktop grid BOINC server.

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.

Virtualization facilitates the integration effort, as presented in the case of remote analysis of the human voice (section 4.2) and in the case 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:

¹<https://www.winehq.org/> WINE. Accessed March 2015

puter. Workunits are generated by a tool within BOINC framework.

Parameter sweep method can enhance the ability to perform identifiability and uncertainty analysis of general complex models and can deliver results of explored parameter space in a reasonable time.

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. [23]. 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 keeps the majority of user experience via network protocol, as presented in section 4.2. It is a way how to migrate legacy application into the computing infrastructure and how to offer it as a service via network protocols. Cloud computing allows to instantiate the virtual machine with such service on demand.

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 to estimate optimal size of population for genetic algorithm in order to optimize computational time and reduce suboptimal results, as proposed by Gotshall et al. [24].

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.






Icon	Description
	Acausal hydraulic connectors – the MODELICA tool generates "Kirchhoff law" analogy for all non-flow variables $p_1..p_n$ (pressure) and "flow" variable $q_1..q_n$ (flowrate): $p_1 = p_2 = \dots = p_n$ $\sum_{i=1}^n q_i = 0$
	Hydraulic Resistor–characterized by G–conductance parameter (reciprocal value of resistance $G = 1/R$) and defined by relation among quantities from both hydraulic connectors, q–flowrate and $(p_{out} - p_{in})$ –pressure gradient: $q = G * (p_{out} - p_{in})$
	Elastic compartment–characterized by parameters: C–compliance (reciprocal value of elastance $C = 1/E$), V_0 –unstressed volume, p_0 –external pressure. The relation among p–pressure, V–volume and q–flowrate are: $p - p_0 = \begin{cases} 0 & \text{if } V < V_0 \\ \frac{V - V_0}{C} & \text{otherwise} \end{cases}$ $\frac{dV}{dt} = q$
	Valve is characterized by g_{on} – inflow, g_{off} – backflow conductance and P_{knee} forward threshold pressure. The relation for dp pressure gradient and q flowrate between the two connectors are: $dp = \begin{cases} pass/g_{on} + P_{knee} & \text{for } pass > 0 \\ pass + P_{knee} & \text{otherwise} \end{cases}$ $q = \begin{cases} pass + P_{knee} & \text{for } pass > 0 \\ pass \times g_{off} + P_{knee} \times g_{off} & \text{otherwise} \end{cases}$
	Inertia element is characterized by the I–intertance parameter and the relationship between pressure and solution flow of the two connectors: $q_{out} = -q_{in}$ $\frac{dq_{in}}{dt} = \frac{p_{in} - p_{out}}{I}$

Table 3.1: Icon and description of hydraulic components of Physiobrary (21.). These are used, e.g., in order to model cardiovascular system.

rithm 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 chosen 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. [19]⁴. 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 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. Current systems of sharing medical images may suffer from the problem of single point of failure or bottleneck. The grid-based solution brings robustness against the mentioned problems.

⁴<https://www.fmi-standard.org/> accessed February 2015

compl.	name	S(10)	S(20)	S(30)	S(40)	S(50)	S(60)	S(80)	S(160)
high	HumMod [17]	10	20.4	24.8	35.4	41.8	49.8	64.0	122
medium	Meurs2011[21]	8.70	16.6	24.4	29.6	32.9	38.7	55.9	53.0
low	Matejak2014[22]	7.50	11.8	12.5	15.4	15.7	14.7	16.7	12.6

Table 4.1: Comparison of model scalability and complexity. Speedup S on 10 CPUs till 160 CPUs of parameter estimation, using cloud computing cluster on 1-6 virtual machines, each 10 CPUs (2x5-core Intel E5-2620 2GHz, 1Gbit/s Ethernet.) (resp. 5-10 virtual machines, each 16 CPUs on physical hardware 2x 8-core Intel E5-2670 2.6GHz). Genetic algorithm configured with a population 120 (resp. 640) individuals for 10 (resp. 20) generations. Speedup estimated from measuring the serial computation on 1 CPU.

4.4 Adair-based Oxygen Binding to Hemoglobin

The parameter estimation was used to compute parameters of newly constructed model of hemoglobin integrating O_2 , CO_2 and H^+ binding based on theoretical principles, which were verified on the parameter estimation algorithm system described above and noted as the low complex model Matejak2014 published as (2.). The author of this thesis implemented the model in Modelica language and estimated parameters of dissociation constants of the chemical reaction of oxygen binding to different forms of hemoglobin comparing the simulated saturation curves of O_2 with the experiments published in scientific literature.

4.5 Parameter Sweep

The desktop grid BOINC framework was customized and a system was established for parameter sweep application. The established project, *Physiome@home*, and its project web page, <http://physiome.lf1.cuni.cz/ident3/physiome>, manages workunit tasks which are downloaded and executed by BOINC workers. The worker application consist of a packaged model that is exported as FMU for a Windows platform and of a universal preconfigured wrapper application provided by BOINC framework, which integrates generic application with a BOINC manager on the desired volunteer com-

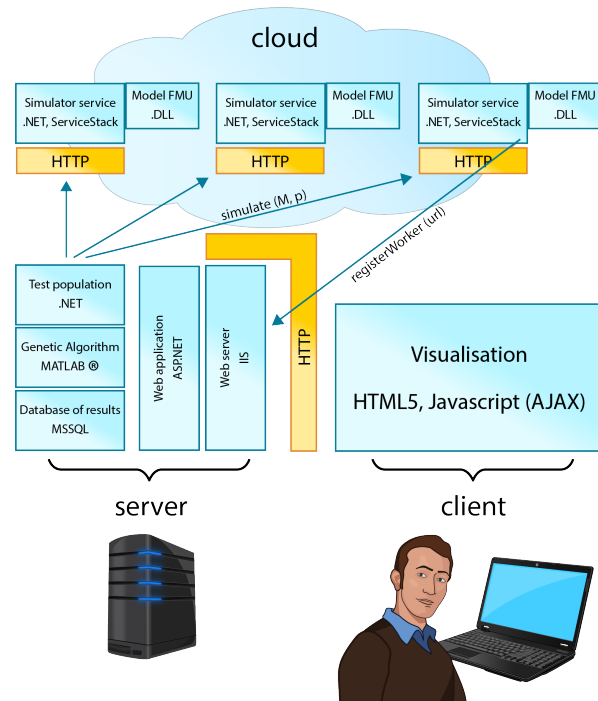


Figure 4.3: Architecture of a system that employs genetic algorithm and distributes the task *simulate* into a cloud computing environment.

estimation was obtained after 200 generations with population of 640, which implicates that the computation time can be reduced from four days to 47 minutes in the case of complex model and from 14 hours to 15 minutes in the case of medium complex model.

The parameter estimation on low complex model suffers with major network overhead. Thus, such low complex models can be effectively identified on local cluster with comparable time of computation.

4.2 Remote Voice Analysis

RDP protocol was customized and support for transferring sound recording was implemented. A client plugins were customized for the Linux "rdesktop" application as well as for the Windows default "tsclient" application. The plugin initiates sound recording from a sound device and creates a custom RDP channel. A raw data obtained from the sound device is transferred to the custom RDP channel. The server plugin writes the data from the custom channel directly to a file in WAV format and offers sound samples to the analytical application for real-time processing. The schematic architecture of the system is in Figure 4.1.

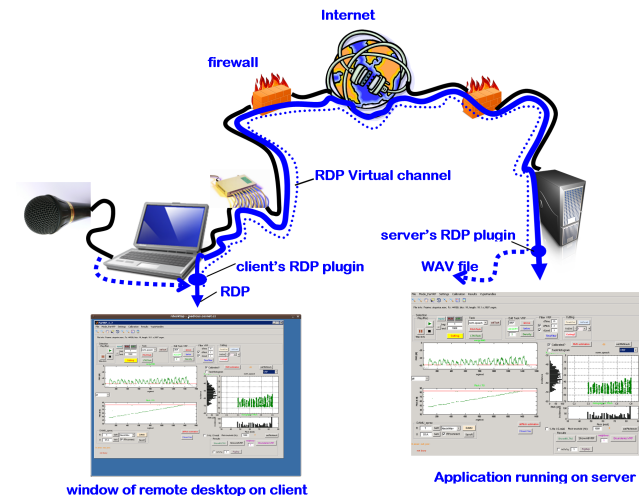


Figure 4.1: Architecture of a system for remote voice analysis and RDP plugins for sound recording redirection.

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 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 transferred to remote application where it is processed and analyzed, 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 in non-invasive way in order to see e.g. the progress of the voice training methods.

4.3 Computational Physiology

4.3.1 Modeling Methodology

Within the thesis, the modeling methodology was improved in an area of modeling cardiovascular system in a complex integrative way, which can be used for research and educational purposes. A set of recommendation was published: (1) to use acausal connector (special purpose class where no causality (what is input and output) is defined), (2) utilize object-oriented features in order to separate pure model from the specific experiment, (3) combine textual and diagram notation in order to express mathematical equations/component relations. This lead to more exact and understandable model for domain experts. The table 3.1 contains definition of basic components of hydraulic domain used to model cardiovascular system. Figure 4.2 shows an implementation of the model published originally by Fernandez de Canete [20] in Modelica language published in (7.). Such medium complex models are important for further studies. Methodology and it's usage in education is described in publication by the author of this thesis in (1.) and (7.).

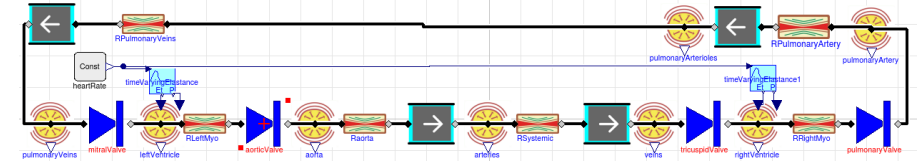


Figure 4.2: Implementation of the model of cardiovascular system [20] in Modelica language using components of PhysiLibrary (21.). The connected components is compiled by the Modelica tool with the equation defined in Table 3.1. In this case the 209 equations (133 are trivial and 76 are non-trivial equations) are generated and causality is solved by the tool.

4.3.2 Parameter Estimation

The proposed architecture of the system for parameter estimation (Figure 4.3) was influenced by the need of some interactivity and for the overall accessibility for users, which is fulfilled by the web UI.

The Modelica models is exported to standardized FMU and wrapped with a RESTful webservice implemented in .NET ServiceStack framework¹ allowing remote control of simulation. In the time of writing this thesis, the most stable Modelica tool was Dymola version 2015², and most stable export was to FMU for a MS Windows platform. Several RESTful web services packaged in a virtual machine template was instantiated in scientific cloud. The overall performance and speedup estimation were tested against the Modelica implementation of complex physiological model HumMod [17], the Modelica implementation of a model of hemodynamics of the cardiovascular system, originally published by Meurs [21] and the model of O₂, CO₂ and H⁺ binding on hemoglobin, published by Matejak et al. and contributed by the author of this thesis (2.)

To summarize the results from Table 4.1, the low complex model scales up to 40 CPUs with a speedup of 15. The medium complex model scales up to 80 CPUs with a speedup of 56 and complex model scales up to 160 CPUs (and probably more) with a speedup of 122. Practically, good parameter

¹<https://servicestack.net> accessed April 2015

²<http://www.dynasim.se> - Dymola tool, accessed March 2015