Generalization aspects in the virtual laboratory system

M. Lawenda, N. Meyer, T. Rajtar, M. Okoń, D. Stokłosa, D. Kaliszan, M. Stroiński

Poznań Supercomputing and Networking Center

Abstract:

In the paper some theoretical study on virtual laboratory aspects are discussed. Possibilities of the VL system creation to control many laboratory apparatus in a remote way are considered by the authors. The main research goals on which the authors want to focus their attention are: laboratory framework and dynamic measurement scenarios.

1   Introduction

Virtual laboratories (VL) are now in the scope of activities of research groups and companies which are working on new solutions for the Grid computational environment [1,2]. Continuous technological progress gives us new capabilities. The growth of computers performance, network throughput and software possibilities allow the creation of new applications which will ease people's work, extend their abilities and simplify the creation process.

The main reason for creating such laboratories is the remote access to the scientific equipment, the facility and acceleration of the education process, exchanging opinions in research groups and the management of the experiment results. This idea is especially attractive for the experimental sciences and technologies, particularly physics, chemistry, structural biology, experimental medicine, radio astronomy and also for widely taken engineering [4,5]. Although the virtual laboratory idea is expensive to introduce, it is still the solution of the future for research environments with limited equipment resources. It is the only way to widely open the science for society demands in education and professional specialization in new technologies. The realization of the virtual laboratory environment will allow the scientists and engineers to work on their projects by remote events simulation, interpretation of experimental data and, in some cases, to run real experiments in a customized laboratory.

First, we should define what the virtual laboratory is. Virtual laboratory is a heterogeneous, distributed environment, which allows scientists all over the world to work on a common group of projects. This environment should allow conducting experiments with the usage of physical devices, doing simulation using the computational application, and communication between users working on the same topic. Similarly to typical laboratory tools and research techniques depending on the specific field of science, virtual laboratories can benefit from some collaboration techniques as tele-immersion, but they are not mandatory.
In order to meet these goals, the virtual laboratory should specify some requirements. The most important requirements follow:

- the possibility of experiments performing (real and computational),

- the possibility of comparing the achieved results in real experiment and simulation, which can help adjust the simulation parameters properly,
- time slot reservation for experiments execution; especially for experiments that require supervision or can be observed by more people on-line,
- load balancing – e.g. experiments should be run on the less loaded sites,
- sharing experiments results – the result should be stored in a database and available for other scientists (with the possibility to restrict access rights),
- library of electronic papers, reports and books grouped by a specific topic.
- communication with the physical laboratory staff, which can be necessary to help run some specific experiments, e.g. setting up and operating devices to experiments,
- communication with other people working on similar problems – using chat, audio and video conferences tools, etc.

Authors research work is focused on the virtual laboratory aspects which have significant influence on creating a general system which will allow taking control on many different devices. In the virtual laboratory system there are many factors which are important in our work. The most important are: laboratory framework, dynamic measurement scenarios and taking into consideration the human factor.

The Virtual Laboratory system is developed in Poznań Supercomputing and Networking System [11] in collaboration with the Institute of Bioorganic Chemistry [12] and the Radioastronomy Department of Nicolaus Copernicus University [13].

The correctness of our concept will be proved by putting into practice two different laboratories. One of them is the Virtual Laboratory of Nuclear Magnetic Resonance Spectrometer, the other - the Virtual Laboratory of Radioastronomy.

The present state of the art in virtual laboratories development discipline was discussed in [10].

In chapter 2 we have discussed the concept of the virtual laboratory framework. Chapter 3 deals with the idea of dynamic measurement scenarios as a way to speed up experiment execution series.

2 Virtual laboratory framework

Due to the increasing popularity of virtual laboratories and the necessity of building more and more modern installations of laboratories, there is a need to develop a conception of the general framework for those systems [3]. The virtual laboratory framework allows creating a new instance of laboratory which can be later adapted to some laboratory device [9].

It is a very difficult assignment to define a laboratory framework which will be appropriate to many types of devices. Finding the solution which will meet the needs of most research groups and allow performing experiments using different apparatus will consume a lot of time, money and human energy. Nevertheless, achieving this goal seems to be worth a big price and bring a lot of advantages (e.g. it will allow reducing resources which are needed to create a new laboratory) in the future.

To meet the conditions mentioned in the introduction, the whole system should have a modular architecture, which allows to increase the system possibilities (e.g. by adding new modules), easily changing the existing modules into others (e.g. with better possibilities) without the necessity of recompilation. It can be possible to build a new virtual laboratory only by defining necessary modules and adapting it to the surrounding environment. Creators do not have to care about architecture and many other important aspects of the new system. They can only focus their attention on aspects directly connected with the implemented device and organizing work for the users (e.g. defining rights lists and edge conditions, creating

laboratory groups, assuring collaboration between apparatus and the existing virtual laboratory modules).

Due to specialization and different nature of some laboratories it is necessary to distinguish two types of modules: general (can be implement in most types of instances of the virtual laboratory) and specific (can be used only with a given type of device).

Next, the conception of the general framework for the virtual laboratory will be discussed. In the virtual laboratory architecture we can distinguish four layers (Fig.1). The first of them is the Access layer. Originally it consists of tools which enable access to the laboratory resources and for presentation stored data. The most important modules are:

- Dynamic Measurement Scenarios – assure creating, controlling and execution of the operations (experiments) chain, which is submitted by the user,
- Users and Laboratories Management – tools for creating new laboratories profiles, new users accounts, management of the access rights to devices, digital library and communication media,
- Data Presentation Tool – presents experiment results collected in the data management system; here we can also find an on-line visualization tool and accounting data presentation tool,
- Teamwork Environment – allows users to communicate with one another (e.g. chat, discussion lists, audio, video),

The second is the Grid layer where services which we can call general were accumulated. They are required in most laboratories instances. There are also connections with the outside grid services e.g. with Globus services. The most important are:

- Authorization Centre – granting (or not) access to many virtual laboratory resources, certificates management, authentication tool,
- Global Scheduling – responsible for choosing an appropriate laboratory device and load balancing when possible,
- Digital Science Library – storing experiments results and electronic publications,
- Data Transport – responsible for downloading and uploading data from/to destination machine (e.g. experiment results, publication from library),
- Grid Gateway – communication with the grid broker (e.g. GRMS [14]), sending computational tasks to the grid and receiving experiment tasks from the grid systems,

On the Supervision layer specific services are gathered. These services have to be implemented taking into account a given device and their specification. Usually it consists of:

- Local Scheduling – tasks scheduling on a given device taking into consideration parameters and priority,
- Resource Monitoring – using resources control, controlling the current state of tasks,
- Users Accounting – information about the used resources,

Finally we can define the Resources layer which consists of devices to experiment execution and also the necessary software:

- Laboratory Devices – laboratory apparatus and software for experiment execution,
- Computational Servers – also software for pre- and post-processing computation,
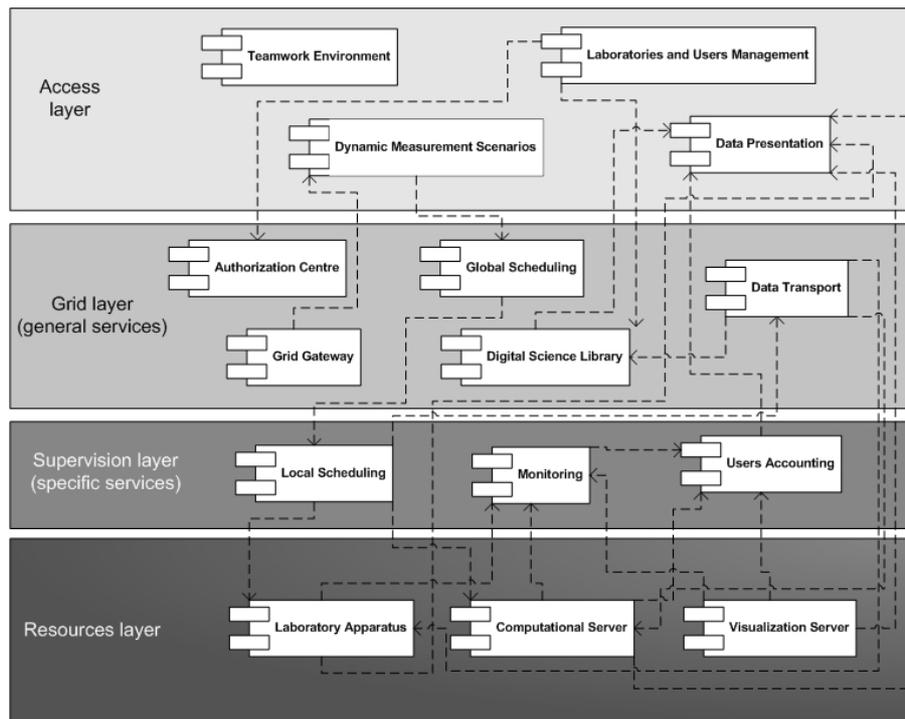- Visualization Servers – also visualization software,

Fig. 1 The framework of virtual laboratory architecture

From the connection point of view, laboratory framework can be divided into three levels (client-agent-server structure): client interface, application server and device server. Such layer architecture provides completing tasks which are assigned to the system. The tasks every level is responsible for are described below.

On the client interface level users have a possibility to submit a task using the available interface (e.g. from the WWW page). Some parts of the interface content should be specific for a particular device, in which the distinctive parameters for a given laboratory type should be taken into account. The user interface uses a standardized protocol to communicate with the application server.

Most services should be gathered on the application server. On the one hand it allows fitting the needs to the specific laboratory, and, on the other hand, decreasing the load of other servers, which are generally designed for executing other tasks. Moreover, it allows fast adapting of the laboratory framework to the new device because only few modules have to be installed outside of the application server and that is why most necessary changes can be done by the local administrator (virtual laboratory server administrator). There (on the VL server), the advanced scheduling process can be executed as well as monitoring, accounting and authorization modules can be running.

The device server is responsible for receiving tasks from the application server, job submitting, receiving results of the executed tasks (after experiment) and transporting them to the application server. The module which controls the served device must know its specificity. It uses the API device functions to submitting tasks and controlling a device.

3 Dynamic measurement scenarios

In many types of laboratories the experiment process execution consists of very similar stages. The given experimental process is often recurrent and must be executed many times by some parameters modification to achieve the required results. In this situation the measurement scenarios concept seems to be very useful.

The concept of the measurement scenarios allows defining the process of an experiment in any way, from pre-processing, through executing the experiment, to the post-processing and visualization tasks. Users are also allowed to add their own module as a part of the scenario. Defining the measurement scenario allows to spare a lot of time during computation. The user does not have to wait for the end of a given process stage to submit another one. It is made automatically.

We distinguish the following types of measurement scenarios:

- static – with a fixed number and types of stages and connections between them,
- dynamic – with a very elastic mechanism of defining stages and connections between them,

The Static Measurement Scenario (SMS) contains a simple task's execution chain. It allows connecting particular stages only between themselves. Here, the execution path is specific and the user cannot manipulate it. Initially, we divide the experiment execution process in the SMS into four steps:

- preprocessing – here the user can prepare data to use as an input source
- experiment – two types of experiments can be distinguished; the first is the real experiment executed on a device available in the Grid; the second one is a computational experiment running on the execution host.
- postprocessing – here we can process the output data from the experiment stage
- visualization – can be treated as a part of the postprocessing stage; we decided to treat it as a separate task due to specific hardware and software needs

To increase the possibility of the jobs scenario, the Dynamic Measurement Scenario (DMS) was defined. As distinct from the SMS, in the DMS model we distinguish two types of jobs:

- experimental – a task is submitted to the laboratory devices, the specificity of this device has to be known,
- computational – a task is submitted to one of the available application servers,

Moreover, both experimental and computational tasks can be divided into: batch mode jobs (user's interaction is not needed) and interactive mode jobs (user's interaction is needed). In the DMS model, besides the definition of the tasks execution sequence, we can also define some extra connections (e.g. loops, parallel connections), conditions on the connections and different lengths of the execution paths. In fact, DMS allows defining many SMS models in one scenario. Thanks to the possibility of the conditions defining on the connections paths, the real path is determined during execution and can depend on computational results.

The creation process of the DMS model is divided into three stages:

- preparation of the connection diagram,
- writing down the dependencies diagram,
- creation of the user DMS coded in the Dynamic Measurement Scenario Language (DMSL) (written progressively during creation of the workflow diagram by a user),

On the first stage "preparation of the connection diagram", an expertise from a particular discipline is necessary. It can be done, for example, by a laboratory administrator together with a domain expert. In the figure below (Fig.2) we present an exemplary measurement scenario diagram for the Virtual Laboratory of NMR Spectroscopy:
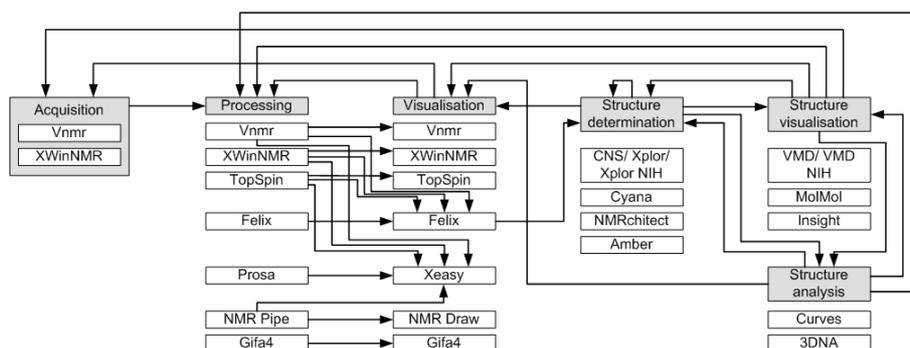
Fig. 2 An exemplary measurement scenario diagram for the Virtual Laboratory of NMR Spectroscopy

To properly write down the dependencies diagram, the system has to have knowledge about the available applications and about connections which were worked out on the previous stage. To make this operation easier for the final user, a special application can be used. An interface for an exemplary programme is presented in Fig.3.
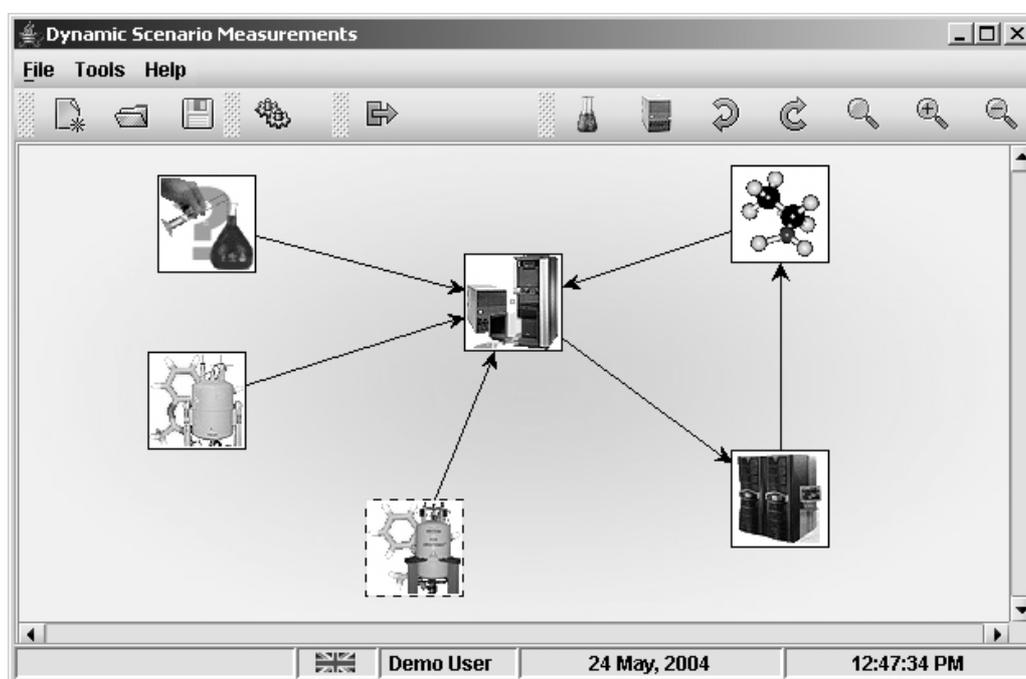

Fig. 3 An interface for an exemplary programme for creating the dependencies diagram

In case of a laboratory where the processing software is well known, it seems to be easy. The situation will be more complex when we want to define the DMS model for a wider range of virtual laboratories and we cannot *a priori* describe all possible execution models. To solve this problem we have defined four files:

- basic – a file with information about possible applications parameters and possible connections with another applications; a list of default output files for a source application and a list of input files for a destination application; an additional condition to connect two given applications e.g. special data file conversion,
- connection – a file with information about connections which were set up by a user between nodes (applications),
- passage – a file with information about passage conditions which describe each connection link between applications; the condition has to be met to go farther

along a given path. We can also define Boolean operations like "or" (default) and "and" (conditions for all links have to be met) between links,

- arrangements – a file with parameters which were set up to a given application by a user

We can say that the first file is an input file which has to be delivered to the application, and the rest of them are output, created by a user during the DMS definition. Files are coded using a special language which we call the Dynamic Measurement Scenarios Language (DMSL) according to an XML standard.

4 Conclusions

In the paper some important aspects which can influence the generality of developed virtual laboratories are discussed by the authors. The most significant are: laboratory framework, dynamic measurement scenarios and taking into consideration the human factor.

Virtual Laboratories are undoubtedly the technology of the future. They are perceived as a remedy for problems connected with access or buying costs of expensive and rare devices. Current technologies connected with VL are in their first stadium of development. New created services allow running simple experiments, which can be managed by changes in a limited parameter pool.

The base services for virtual laboratories are currently under development. The hardware (e.g. SGI Reality Center [6]) and software (CAVERN G2 [7]) solutions, which allow remote communication and results presentation, are being developed. However, most of the currently developed laboratories are rather low advanced constructions, created for a small group of specialized tasks.

There is a lack of a universal toolkit for creating virtual laboratories. Developed solutions are assigned just for specific tasks and their structure is strongly conditioned by problems for which they are created. There is a demand to create a universal structure which could be adapted to specific types of laboratories.

The universal virtual laboratory framework allows sparing a lot of human energy and money during the development stage. It is possible to build many different types of laboratory systems which provide and adapt functionality to specified needs. Designers of a new laboratory have to focus their attention only on a few aspects connected with adaptation to a new data set, and a specific scheduling algorithm.

To prove the assumptions presented in the paper, and as an exemplary implementation, it is planned to make the NMR Spectrometer (Bruker Avance 600) and radiotelescope (diameter 32m) available to the Grid users [15].

References:

[1] Lawenda M., „Laboratorium Wirtualne i Teleimersja", Poznań Supercomputing and Networking Center, RW nr 34/01, (in polish)

[2] Bubak M., Madajczyk J., Meyer N., „Laboratorium wirtualne z wykorzystaniem infrastruktury HPC/HPV", (in polish)

[3] Afsarmanesh H., Benabdelkader A., Kaletas E.C., Garita C., and Hertzberger L.O., "Towards a Multi-layer Architecture for Scientific Virtual Laboratories" In: Bubak, M., Afsarmanesh, H., Williams, R., Hertberger, B., (Eds.), Proc. Int. Conf. High Performance Computing and Networking, Amsterdam, May 8-10, 2000, Lecture Notes in Computer Science 1823, 162-176, Springer, 2000

[4] Khettry D., Xian-He Sun, "A Windows-NT virtual collaboratory for technical computing", Fifth NASA National Symposium on Large-Scale Analysis, Design and

Intelligent Synthesis Environments. Williamsburg, VA, USA. 12-15 Oct. 1999. Elsevier. Advances in Engineering Software, vol.31, no.8-9, Aug.-Sept. 2000

[5]   Nicholas P., Fushman D., Ruchinsky V., Cowburn D., "The Virtual NMR Spectrometer: a computer program for efficient simulation of NMR experiments involving pulsed field gradients"

[6]   Reality Center Overview - http://www.sgi.com/realitycenter/overview.html

[7]   Open Channel Foundation – CAVERN G2 - http://www.openchannelsoftware.org/projects/CAVERNsoft_G2/

[8]   Virtual Reality based Simulation Laboratory – http://www.hlrs.de/organization/vis/projects/vrslab.html

[9]   Lawenda M., Meyer N., Rajtar T., "General framework for Virtual Laboratory", THE 2ND CRACOW GRID WORKSHOP, Cracow, Poland, December 11-14, 2002

[10] Adamiak R. W., Gdaniec Z., Lawenda M., Meyer N., Popenda Ł., Stroiński M., Zieliński K., „Laboratorium Wirtualne w Środowisku Gridowym", Polish Optical Internet - PIONIER 2003, 9-11 April 2003, (in polish)

[11] Poznań Supercomputing and Networking System, Web site pages http://www.psnc.pl/

[12] Institute of Bioorganic Chemistry Polish Academy of Sciences, Web site pages: http://www.ibch.poznan.pl/

[13] Radioastronomy Department of Nicolaus Copernicus University, Web site pages: http://www.astro.uni.torun.pl/

[14] GridLab: A Grid Application Toolkit and Testbed, Web site pages: http://www.gridlab.org/

[15] Virtual Laboratory, Web site project pages: http://vlab.psnc.pl/

Author(s):

Marcin, Lawenda, M.Sc.
Poznań Supercomputing and Networking Center, Supercomputing Department
Noskowskiego Str. 10, 61-704 Poznań, Poland
lawenda@man.poznan.pl

Norbert, Meyer, PhD.
Poznań Supercomputing and Networking Center, Supercomputing Department
Noskowskiego Str. 10, 61-704 Poznań, Poland
meyer@man.poznan.pl

Tomasz, Rajtar, M.Sc.
Poznań Supercomputing and Networking Center, Supercomputing Department
Noskowskiego Str. 10, 61-704 Poznań, Poland
ritter@man.poznan.pl

Marcin, Okoń, M.Sc.
Poznań Supercomputing and Networking Center, Supercomputing Department
Noskowskiego Str. 10, 61-704 Poznań, Poland
hawky@man.poznan.pl

Dominik, Stokłosa, M.Sc.
Poznań Supercomputing and Networking Center, Supercomputing Department
Noskowskiego Str. 10, 61-704 Poznań, Poland
osa@man.poznan.pl

Damian, Kaliszan, M.Sc.
Poznań Supercomputing and Networking Center, Supercomputing Department
Noskowskiego Str. 10, 61-704 Poznań, Poland
damian@man.poznan.pl

Maciej, Stroiński, PhD.
Poznań Supercomputing and Networking Center, Supercomputing Department
Noskowskiego Str. 10, 61-704 Poznań, Poland
stroins@man.poznan.pl