TRILHA PRINCIPAL

Collaborative and distributed computing applied to biomedicine with the FightAIDS@Home

Gabriella Castro Barbosa Costa, *Graduate Student in Computer Science, UFJF*,
Bruno Zonovelli da Silva, *Graduate Student in Computational and Mathematical Modelling, UFJF*,
Humberto Luiz de Oliveira Dalpra, *Graduate in Information Systems, CES/JF*,
Izaías Fernando Vilarino, *Graduate in Information Systems, CES/JF*,
Mayara Mendes Paiva, *Graduate in Information Systems, CES/JF* and
Wagner Arbex, *PhD in Computer and Systems Engineering, Embrapa and UFJF*

Abstract—**Distributed computing systems are used for high performance computing tasks, taking advantage of join processing power of multiple independent computers, but they are interconnected by network. These high performance systems can be divided into two classes of distributed computing systems: computer cluster and grid computing. The FightAIDS@Home is a distributed system toward the search for solutions for the AIDS treatment, and this project is led by the Olson Laboratory, California. Nowadays, this project use computing resources of the Grid World Community Grid, which consist mostly of computers of ordinary users, using open and non-specific standard protocols and interfaces to ensure interoperability between different systems. These paper presents information from a case study about FightAIDS@Home and its implementation, using aspects of distributed systems such as grid computer and cloud computing to support implementation of collaborative computing.**

Index Terms—**Distributed computing, grid computing, FightAIDS@Home, World Community Grid.**

I. Introduction

O FightAIDS@Home is a distributed system used in the biomedicine field that helps in the research and in the discovery of alternatives for treating and curing AIDS. This project associated itself in November/2005 with the World Community Grid (WCG), a computer grid whose processing results come from the collaboration among volunteers, that donate processing time from a large number of personal and work computers from all over the world through the Internet. Participation is open to everyone and in order to become a volunteer all that is required is to copy and install the freely available project integration software, available at the WCG site, that will execute in background when the volunteer computer is idle.

In the first three years of the project, more than 409 thousand persons became volunteer and made available the processing power of over 1,1 million computers to give impulse to humanitarian research, the equivalent of 192,400 processing years for this computer grid [4].

This paper intends to explain FightAIDS@Home as a distributed system, showing the concepts of grid computing and its working in the collaborating stations. We will describe the project origin, initial motivation and its main responsible and maintainers, establishing a detailed and current view of FightAIDS@Home at the WCG, with some of its technical specifications and characteristic and also reporting on some of the possible advances and future perspectives of the WCG.

In spite of FightAIDS@Home being a distributed system, its structure and working does not characterize a cloud computing system. Nevertheless, it incorporated concepts and tools from cloud computing, such ad distribution transparency [1], a concept that is common in distributed systems like this one.

Besides, distributed and collaborative computing implemented by the WCG and used by FightAIDS@Home is also associated with the concept of e-Science, that may be defined in a simplified way as activities of intensive scientific computing that are performed in highly distributed network environments [10], as in grid computing.

This text is organized into two parts. Initially, the second and third sections explain FightAIDS@Home, presenting some of its basic aspects and the fundamentals of its organization, technique and some data on its working inside the WCG. Following, the fourth section is about the WCG, presenting data on some of the projects it hosts and how collaborative computing has been adopted over the world. The content of this paper is completed with the current introduction, that presents some observations on the problem at hand and with the conclusion, which summarizes the importance of FightAIDS@Home and of WCG as its computing infrastructure.

II. Concepts and Aspects of Dsitributed Systems in FightAIDS@Home

Corresponding author: Wagner Arbex, *arbex@cnpgl.embrapa.br*

O FightAIDS@Home is a distributed computer system hosted at the WCG (a communitarian grid computing system kept by IBM) whose goal is to help the research that seek the cure and treatment to many diseases [11], being the first project hosted at WCD, as we can see in the information provided by the Olson lab [7].

The working structure of a computational grid proposes

> to combine the processing power of many computers connected by a network in order to perform trasks that would not be possible to execute using a single computer (at least with acceptable performance) and, at the same time, incur in a cost that is smaller than the one incurred by a supercomputer of equivalent power [9].

Scalability and scalability transparency are attributes of WCG, as in all distributed systems. Therefore, the nodes that make the computational grid can be either permanent or temporary. Hence, the same happens with FightAIDS@Home. Besides, grid computational systems can be formed to execute a specific task and then disbanded. The "entry" or "exit" of a node in the grid can be done simply by activating of deactivating a software in each computer.

In order to contribute with the projects hosted at the WCG as FightAIDS@Home itself, all it takes is to copy and install the project integration software. After its installation, this software will use idle processing time in the host computer to integrate it to the computational grid in order to perform simulations, projections and calculations.

III. Main Aspects of FightAIDS@Home

O FightAIDS@Home uses distributed computing in order to help the research for the discovery of AIDS treatment. It executes a software called AutoDock, that codifies physical, chemical and biochemical characteristics for the forecasting of flexible target molecules that are the possible candidates (related to the proteic structure of the drug) that could associated to the protease structure using, for instance, clivage or peptidic bonds. The goal is to choose the best target molecules that come from the protease and test them in the wet lab [7].

This proposal to predict candidate molecules and proteic structures is called molecular modeling through structural homology and uses experimentally determined proteic structure to try to predict the conformations of other proteins that have similar amino acid sequences. The similarity between the protein with the solved structure (the mold protein) and the one we intend to model (the target protein) is a determinant of the quality of the structural models generated as well as it applicability in the development of drugs and pharmaceuticals.

According to the WCG [16], in April 8th, 2011, FightAIDS@Home surpassed the equivalent to 132 thousand processing-years and has generated 133,5 millions results. The daily average processing time, up to that date, is equivalent to 67 processing years, generating daily a little less than 68 thousand results.

A. Project organization

FightAIDS@Home is coordinated by professors Arthur Olson, from the Molecular Biology Department at the Scripps Research Institute, in La Jolla, and Rik Belwe, from the Cognitive Science Department at the University of California, in San Diego, and its structure is financed by the National Institute for Health (NIH) in order to develop new therapeutical approaches based on better knowledge of HIV structural biology [6].

Occupying important spots in the project team are dr. Alex Perryman, specialist int he use of protein structure information ; Stefano Forli, PhD in pharmaceutical sciences; Dallakyan Sargis, responsible for the hardware and software environment for the Molecular Graphics Laboratory and chief developer of Python Molecular Viewer (PMV); Garret Morris, co-author of AutoDock and specialist in computational chemistry and molecular modeling. Also a participant member, IBM is the technical responsible for WCG infrastructure [7].

FightAIDS@Home is one of the six active projects at the WCG and Table I shows the amount of processing time (years equivalent) of each of these six projects. Besides the active projects, WCG still has six other intermittent projects and six finalized ones.

TABLE I
LIST OF ACTIVE PROJECTS AT THE WCG [13]

Project	Approx. processing time (years)
FightAIDS@Home	132.000
Human Proteome Folding	86.000
Help Conquer Cancer	73.500
Help Cure Muscular Dystrophy	41.200
Computing for Clean Water	14.000
The Clean Energy Project	3.600

Among the goals of FightAIDS@Home, these researches and their teams intend to create new drugs, increase the understanding of AIDS structural biology and to stidy the mechanisms of resistance to several drugs used by the HIV to 'escape' from the action of currently used drugs [7].

B. Aspects and technical specifications

As all projects hosted at the WCG, FightAIDS@Home is based on a collaborative and distributed computing structure, where the system nodes, made available by volunteers, download and install a software agent whose execution in started with the activity of screen saver. Each node receives a small data set, called work units (WU), to process. Each WU is sized at 512 KB [18] and each WU can be transmitted to three different nodes [4].

When the nodes are idle, their screen saves will start and so will the installed software agents which will process the WUs. At the end of this processing, the results are forwarded to the FightAIDS@Home servers.

The results for the processing of the same WU sent to different nodes are compared, in order to validate the results found by each node [4].

In order to execute the software agent, the node must have Internet connection, 250Mb RAM memory, 100Mb free disk space and optionally, graphic exhibition capacity, if the user wishes they are shown. There are Linux, Mac and Windows versions of the agent available for download [17].

Looking closely, we can see that the execution is a procedure [2] that starts with each node receiving a set of tasks and data to be processed, that is the WUs. In spite of the fact that each WU has a standard initial size, the tasks and WUs are calculated according to the capacity of the node. For instance, the servers never send to a node a task and an amount of data bigger than the node configuration can treat (as a function of the node's RAM memory).

The next step, after the processing itself, is to return the data found by each node to the servers, together with reports on those results, when the latter is validated.

Strategies of processing and validation [19] of FightAIDS@Home are based on Simple Validation - Type 1 in separate or together with Redundant Computations.

In order to fully understand processing and validation policies for simple validation, it is necessary to understand the execution policy for the redundant processing strategy.

In redundant processing, two copies of the same WU are sent to two different nodes and when both of them return their results they are compared in order to assure that they are identical and then considered valid and accepted. If there is any disagreement in the results, then the same data set is sent to other nodes until two of them achieve the same results.

The same procedure is done if there is any interruption and the result is not returned to the servers. That is, a result is considered valid and accepted in redundant processing if at least two different results on the same data source found by two different nodes are equal. This strategy establishes a high level of trust on the results itself and on the nodes used in processing. In this strategy, "points" are attributed to the nodes in order to grade them as "reliable" - something that will be considered in simple validation.

In simple validation, a single copy of the WU is sent to a single reliable node. For a node to be considered reliable, it must participate in the project long enough for the platform to trust it and also must return good results in that time period.

If the node selected to receive a WU is not reliable, then a copy of the same WU is distributed to another node and the rules and procedures from redundant processing apply. Besides, as a precaution, the code executing in each node in specific moments can verify together with the servers if the WU processing is going according to expectations.

In order to establish and maintain a trust relationship with the participating nodes, the reliable ones are randomly sampled in order to have their results checked twice, as in redundant processing, so that they can be "scored" again and keep its trusted node status.

The processing and validation strategies are used in several applications. For instance, simple validation (the primary strategy for FightAIDS@Home) is also used in other research projects, such as the Discovery Dengue Drugs. Similarly, FightAIDS@Home itself uses redundant processing that originally is a validation and processing strategy for the Help Conquer Cancer project.

Besides Simple Validation - Type 1 and Redundant Computations strategies, there is still Simple Validation - Type 2 that, differently from the first two strategies, may generate different results every time a WU is processed, due to the characteristics of the application executing the WU in this strategy. Simple Validation - Type 2 is not used in FightAIDS@Home.

Control, communication and validation modules that negotiate with the node that is receiving the WU serve the projects mentioned above, as well as the processing strategies.

IV. Current picture and Future of the WCG

GIVEN the usage of grid computing, together with the infrastructure, a "virtual system" of huge processing capacity is created for systems that are hosted at WCG, allowing for high processing rates for those systems [5].

A. Current picture

The five countries that contribute the most to WCG processing time are USA, Japan, France, Canada and Germany, with Brazil standing at 11th place (Table II). Processing time means the contribution from each country from the beginning of the project, that is, what "used" in processing time from each participating country since its beginning.

TABLE II
List of countries and execution time given to WCG [14]

#	Country	Approx. Processing Time (years)
1	USA	153.000
2	Japan	33.000
3	France	22.000
4	Canada	20.000
5	Germany	17.500
6	United Kingdom	16.000
7	Netherlands	11.500
8	Australia	10.300
9	Spain	6.700
10	Taiwan	6.000
11	Brazil	5.300

As to the current number of volunteers to the project, Brazil stands at the 7th place in the ranking, as presented in Table III, where we can see the top ten contributing countries.

As seen in Figure 1, African countries that are those that suffer the most with the HIV virus are those with

TABLE III
List of countries and number of volunteers to the WCG [15]

#	Country	Approx. number of collaborators
1	USA	65.400
2	France	27.700
3	Germany	24.700
4	Japan	19.300
5	United Kingdom	14.800
6	Canada	11.400
7	Brazil	10.100
8	Italy	7.300
9	Australia	5.900
10	Spain	5.700

less resources invested in the collaborative effort in which FightAIDS@Home is included.

For instance, according to the WCG [16] in April 8th, 2011, the first country to appear in the list of processing time "donation" to the WCG is South Africa, in the 31st position, which is equivalent to less than 1.200 processing years. Besides, a number that may surprise many is the small participation of China, standing at the 20th place, with the processing time equivalent to less than 2.600 years.
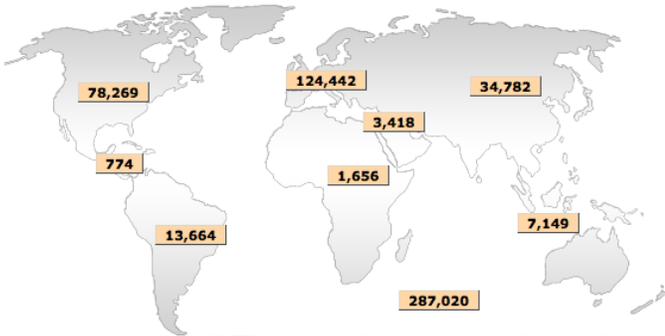
Fig. 1. Participation of subcontinental regions in the WCG in terms of collaborators [12].

The last project record at the WCG [8], up to April 5th, 2011, accounts for 34 finished experiments in FightAIDS@Home. The 35th is undergoing (with 85% executed) and experiment 36 is about to start.

B. Future Perspectives

When IBM and the members of the main science, education and social assistance in the world launched the WCG is November, 2004, they claimed as their mission "to create the largest public grid computing system with humanitarian goal " [3], by using collaboratively idle processing time, once that, according to the same source, computer users use between 10% and 15% of the processing capacity of their equipments.

The acceptance and the success of this project can be verifies based on the numbers it amassed in its first years.

Two years after its launch, 250.000 persons in the whole world were already participating in the WCG and 500.000 collaborators were enrolled. In this time period, the equivalent of 72.000 distributed and collaborative processing years were used [3]. In the following year, it grew even bigger, for more 409.000 volunteers were added to a grand total of 1.1 million computers to help impulse humanitarian research, the equivalent of 192.400 processing years for this grid computing system [4].

In spite of those good numbers, the WCG understood that its work was only beginning and its consulting committee, even in the beginning of the project, already performed the task of continuously search form new research project that could benefit from this grid computing system. In this sense, one can realize that the proposal of new researches from the WCG go according to the common sense of world needs, given that its future projects shall continue to fight infectious diseases but also to fight hunger and natural disasters.

V. Final Thoughts

ALL the advances performed and the search of results that will certainly help in the creation of newer and better AIDS treatments would not be possible without using the grid computing supported project, because thousands of years would be required from a single computer to amass all the processing power currently available for the WCG.

FightAIDS@Home is an initiative that uses the concepts of Distributed Systems and of volunteer collaborative computing in the search for continuous advancements in the fight against AIDS. The goal is to find new discoveries in order to help Human health (for it is a worldwide issue), exploring a huge processing capacity that is currently inactive in millions of machines all over the world and is currently becoming a beacon of hope for the future of mankind.

References

[1] W. Arbex, M. V. B. da Silva, M. F. M. Guimarães, R. F. Tagliatti, L. G. de Andrade, M. N. M. Muniz, and L. A. V. de Carvalho, "Storage as a service and utility computing for bioinformatics computing environment: aspects of cloud computing applied to scientific computing," in *Anais do IV Encontro Acadêmico em Modelagem Computacional*, Trabalho classificado como Trabalho Destaque em Ciência da Computação no IV Encontro Acadêmico em Modelagem Computacional do Laboratório Nacional de Computação Científica, Laboratório Nacional de Computação Científica, Petrópolis: LNCC, 2011.

[2] Berkeley Open Infrastructure for Network Computing. (2011). How BOINC works. Access date: 15 july 2011, Berkeley Open Infrastructure for Network Computing - BOINC, [Online]. Available: http://boinc.berkeley.edu/wiki/How_BOINC_works.

[3] IBM, *Como funciona a World Community Grid*, http://www-05.ibm.com/pt/ibm/ccr/HIW_02132007_trPTG.pdf, 2007. [Online]. Available: http://www-05.ibm.com/pt/ibm/ccr/HIW_02132007_trPTG.pdf.

[4] —, "World community grid," Agenda Sustentável, São Paulo, Tech. Rep., 2008. [Online]. Available: http : / / www . agendasustentavel . com . br / images / pdf/001208.pdf.

[5] —, ([2008]). World Community Grid. Access date: 1 jun. 2011, [Online]. Available: http://www-05.ibm.com/pt/ibm/ccr/world_community_grid.html.

[6] Olson Laboratory. (2011). Resarch team. Access date: 1 abr. 2011, [Online]. Available: http : / / fightaidsathome.scripps.edu/team.html.

[7] —, (2011). What is FightAIDS@Home? Access date: 1 abr. 2011, [Online]. Available: http : / / fightaidsathome.scripps.edu/.

[8] —, (2011). What is FightAIDS@Home? Access date: 8 abr. 2011, [Online]. Available: http : / / fightaidsathome.scripps.edu/status.

[9] C. E. M. da Silva. (2005). Grid computing. Access date: 21 set. 2010, Hardware.com.br, [Online]. Available: http://www.hardware.com.br/termos/grid-computing.

[10] Wikipedia. (2011). e-Science. Sep. 20, 2011, [Online]. Available: http://en.wikipedia.org/wiki/E-Science.

[11] World Community Grid. (2010). About us. Access date: 21 set. 2010, World Community Grid, [Online]. Available: http://www.worldcommunitygrid.org/about_us/viewAboutUs.do.

[12] —, (2011). By geography. Access date: 8 abr. 2011, World Community Grid, [Online]. Available: http : / / www . worldcommunitygrid . org / stat / viewStatsByCountryAT.do?sort=members.

[13] —, (2011). By projects. Access date: 1 jun. 2011, World Community Grid, [Online]. Available: http://www.worldcommunitygrid.org/stat/viewProjects.do.

[14] —, (2011). Country statistics. Access date: 8 abr. 2011, World Community Grid, [Online]. Available: http : / / www . worldcommunitygrid . org / stat / viewStatsByCountryAT.do?sort=cpu.

[15] —, (2011). Country statistics. Access date: 8 abr. 2011, World Community Grid, [Online]. Available: http : / / www . worldcommunitygrid . org / stat / viewStatsByCountryAT.do?sort=members.

[16] —, (2011). FightAIDS@Home project statistics. Access date: 8 abr. 2011, World Community Grid, [Online]. Available: http://www.worldcommunitygrid.org/stat/viewProject.do?projectShortName=faah.

[17] —, (2011). Getting started. Access date: 1 jun. 2011, World Community Grid, [Online]. Available: http://www.worldcommunitygrid.org/help/viewTopic.do?shortName=minimumreq.

[18] —, (2011). Overview: AfricanClimate@Home. Access date: 1 jun. 2011, World Community Grid, [Online]. Available: http://www.worldcommunitygrid.org/help/viewTopic.do?shortName=acah.

[19] —, (2011). Overview: points. Access date: 1 jun. 2011, World Community Grid, [Online]. Available: http : / / www . worldcommunitygrid . org / help / viewTopic.do?shortName=points.

Gabriella Castro Barbosa Costa is graduated in Information Systems at the Higher Studies Center at Juiz de Fora and is currently a graduate student at the Computer Science masters program from the Federal University of Juiz de Fora.

Bruno Zonovelli da Silva is graduated in Information Systems at the Higher Studies Center at Juiz de Fora and is currently a graduate student at the Mathematical and Computational modeling Masters program from the Federal University of Juiz de Fora.

Humberto Luiz de Oliveira Dalpra is graduated in Information Systems at the Higher Studies Center at Juiz de Fora and is currently a trainee at CEOsoftware Information Systems Company.

Izaías Fernando Vilarino is graduated in Information Systems at the Higher Studies Center at Juiz de Fora.

Mayara Mendes Paiva is graduated in Information Systems at the Higher Studies Center at Juiz de Fora and is currently a systems analyst at Solucionar Computer Systems.

Wagner Arbex (corresponding author) has a PhD in Computer Science and Systems Engineering from the Federal University of Rio de Janeiro and is nowadays an analyst at the Brazilian Agricultural Research Corporation (Embrapa) and guest lecturer in the masters program of the Federal University of Juiz de Fora. E-mail: *arbex@cnpgl.embrapa.br*.