The Grid framework as a suitable infrastructure for supporting Large-Scale Desktop Networked Virtual Environments

C. Bouras
Research Academic Computer Technology Institute, 61 Riga Feraiou STR, GR-26221, Patras, GREECE
Computer Engineering and Informatics Department, University of Patras
GR-26500 Rion, Patras, GREECE

E. Giannaka
Research Academic Computer Technology Institute, 61 Riga Feraiou STR, GR-26221, Patras, GREECE
Computer Engineering and Informatics Department, University of Patras
GR-26500 Rion, Patras, GREECE

Abstract: *The maturation of the Internet, the familiarization of the users with this powerful means of communication and mainly the increment of the available bandwidth, which is presented the last few years, led to the idea of extending small scale virtual environments to larger applications that could support a tremendous number of concurrent users. However, the introduction of Large Scale Virtual Environments brings along a series of problems that are mainly related to the serious increment of the transmitted information. Neither the traditional architectures, algorithms originally designed for the NVEs, nor the improvement of the network seem sufficient to support this new kind of applications. A solution to this problem seems to be the adoption of the GRID. In this paper we present an approach, which describes the way that a Desktop Networked Virtual Environment could be efficiently supported and maintained under a GRID infrastructure.*

Keywords: Large-Scale Virtual Environments, GRID

1. Introduction

The vast evolution of both the hardware and the software capabilities, which has taken place the last decade, led to the emergence of a newborn trend, which focuses on the user-system interaction [1]. Virtual Reality constitutes the basic representative of this new trend. In particular, there are a number of categories of virtual reality systems, each of which aims at the simulation of either real or imaginary worlds, based on the available means and the extent of realism that tries to achieve. One of these basic categories is the Networked Virtual Environments (NVEs).

The first NVEs originally developed could support a limited number of users, on the order of 250-500 users and were defined as Small-Scale NVEs [2]. The maturation of the Internet, the familiarization of the users with this means of communication but mainly the improvement of this dynamic means led to the need for the formation of larger NVEs. In particular, the last few years, the research interest turns to the direction of NVEs that could support efficiently a tremendous number of concurrent users that could reach up to 300.000 concurrent users [2].

The development of Large Scale Networked Virtual Environments introduced a series of problems, which are mainly related to the debility of the network to host and serve the demanding recourses that these applications require. In particular, these environments are accompanied by rich graphics for the representation of the provided information as well as a variety of provided services. These characteristics, in combination to the tremendous number of users that are called to support, result in the exponential increment of the resources needed for their smooth, natural and efficient operation, which aims at the achievement of a high sense of realism for the connected users. The problems and limitations that are introduced by the Large Scale NVEs are mainly related to the efficient transmission of the exchanged messages to the appropriate destinations and the optimized update of the virtual worlds so as the consistency to be maintained.

This newborn research field has already drawn increased research interest and many efforts have been made in order to face the limitations. The majority of the existent approaches tend to design mechanisms and algorithms, which are based on the extension of small-scale applications and when applied could support Large Scale Environments. However, these approaches are usually targeted to specific parts of these environments and cannot be easily applied to a wider range of these applications.

GRID technology emerges as a solution to these limitations. In particular, the advanced capabilities of this powerful field could be seen as an important potential for the support of Large Scale Networked Virtual Environments. However, in order the technology and concepts that underlie NVEs to be combined with the concepts and architecture of the GRID, algorithms and techniques should be designed and developed, which will act as the necessary “glue” for the development and efficient support of Large Scale NVEs.

This paper is engaged with the description of how a large-scale Networked Virtual Environment, could be supported under a GRID infrastructure, by combining the NVEs characteristics to the advanced possibilities that GRID technology can provide and by presenting all the methods and techniques that such an environment would require for an optimized performance in a GRID concept.

The remainder of this paper is structured as follows. Section II provides an introduction to the problem that Large Scale NVEs are called to solve. Section III is dedicated to related work on the direction of solving the problem and presents the research that has been made for the confrontation of the limitations. The next section describes some of the basic concepts of the GRID and relates its concepts to the concepts of the Networked Virtual Environments. Section V presents the way that GRID could be used in order to efficiently support the basic requirements of a Large Scale Virtual Environments. Finally, section VI summarizes and concludes the paper.

2. The problem

A Networked Virtual Environment is the simulation generated by a computer system, which simulates a real or imaginary scene aiming at providing a high sense of realism to the connected users [8]. In particular, a NVE constitutes a computer system, which generates virtual worlds where the users can interact both with the system and the other connected users in real time. The users are connected to the Internet and working on different computers, access the same virtual scene. The simulation of the virtual scenes is realized through distributed and heterogeneous computational resources.

The evolution of the software applications and services in combination to the melioration of the network allows for the development of networked applications, which are characterized by the enhancement and combination of many advanced features. For NVEs in particular, where the achievement of high realism constitutes a key concept, the realistic and detailed representation of the provided information is of high importance. Therefore, the potentialities that technology presents in combination to the increased needs of the users result in NVEs to adopt rich representation for the information in terms of graphics and media. However, the rich representation of the provided information implies additional computational resources for the network host, which is translated in more bandwidth. The heavy demands that Large Scale NVEs place, as, apart from the rich representation, the number of the connected users increases tremendously, require more than simply more bandwidth [3].

From the above, it becomes clear that traditional architectures and algorithms adopted for the development and support of, the so-called, small-scale Virtual Environments seem insufficient to be applied to environments with an importantly larger number of users. In particular, small-scale approaches, from an architectural point of view, fall usually into one of the following architectures:

- Client-Server architectures, where the clients communicate their changes to one or more servers. These servers are responsible for the redistribution of the received information to all connected clients.
- Peer-to-Peer architectures, where the clients communicate directly their modifications and updates of the world to all connected clients.

Both approaches, when scaling into a larger number of users fail to support the virtual environment efficiently as either, the clients, the server, or both fail due to bandwidth deficiency [3].

A promising alternative for the support of the Large Scale Virtual Environments rises through the GRID technology. In particular, according to [6], GRID computing is "a hardware and software infrastructure that clusters and integrates high-end computers, networks, databases and scientific instruments from multiple sources to form a virtual supercomputer on which users can work collaboratively". The potentialities that GRID technology introduces could provide an efficient and reliable infrastructure for the support of Large Scale Virtual Environments.

To this direction, there is a number of issues for Large Scale Virtual Environments that need to be taken into account for achieving an effective and reliable system. One basic issue is the partitioning of the virtual world. This key issue is related to the optimized distribution of the client nodes within the Grid to the server-nodes [4]. In particular, the partitioning problem is related to the design and development of intelligent algorithms and techniques, which will be able to assign efficiently the processes to the appropriate task carriers, so as to maintain the consistency between the users' view and furthermore, reduce the network load within the GRID. The quantitative definition of "large" for Networked Virtual Environments applies to both the size of the virtual worlds as well as to the number of the participants. Therefore, with a simplified assumption a Large Scale Virtual Environment could be properly partitioned so as to constitute a collection of smaller scale environments. These, smaller scale environments would be then easier managed for the optimization of the performance.

However, when developing a Large Scale Virtual Environment, things are not that simple. In particular, in order the partitioning of the virtual worlds to be efficient and reliable, it should be based on intelligent algorithms and techniques that from one hand will be able to provide the transparent, for the end user, transition from one part of the world to another and on the other hand should be able to support the efficient transmission of information to all destinations under an optimized performance.

In addition, the methods developed for the optimal partitioning of the virtual space should be based on both the architectural and communication model that the platform adopts. The architectural and communication model, in terms of the protocols adopted for the transmission of certain types of information within the Virtual Environment, play an important role to the overall performance. Therefore, they should be taken into account by the partitioning methods.

Concluding, there should be an intelligent compromise of all the above into a generalized schema that will act as the interface between the GRID and the Virtual Environments' technology, so as to provide an effective and reliable system.

3. Related Work

The path for the support of Large Scale Networked Virtual Environments over the GRID draws increased interest. The research in this field is bi-directional. From one hand, GRID technology, continually evolving, tries to exploit more fields, where it could be effectively applied. One of these fields is Networked Virtual Environments. An important representative of this direction is the CrossGrid project [8], which aims to develop, implement and exploit new Grid components for interactive compute and data intensive applications. From the other hand, Networked Virtual Environments turn to the direction of larger scale applications. The adaptation of the GRID technology, for the development and support of Large Scale Virtual Environments, forms a new research field to this direction. Most of the research work in this field arises from the category of Immersive Virtual Reality. “RealityGrid” [6]is an example of such approaches.

The efforts and research on the field of Immersive Reality are very interesting and of great importance, as the results could assist to the application of the GRID to other fields of Virtual Reality. However, the essence of “large-scale environments” usually corresponds to desktop virtual environments, which, due to their affordable cost and hardware requirements, are more widespread among the end users. To this direction, which constitutes the area of interest for this paper, one of the most important representatives is Butterfly.net, which extends the simple client-server model by adopting multiple servers and uses the GRID technology, as the underlying infrastructure [5]. The adoption of the GRID technology as a solution to the growing needs of Desktop Networked Virtual Environments still remains a novel field of research and development.

4. GRID support to NVEs basic requirements

The concept of large-scale virtual environments has a twofold meaning· large to the surface of the virtual world and large to the number of the concurrent users. As mentioned above, each of this subsistence contributes to the increment of the system and network resources needed. Therefore, a suitable solution for the support of these applications should aim at the efficient confrontation of these parameters. Even though networked virtual environments can vary on the variety and type of the provided services, there are some basic requirements that all NVEs should efficiently support. In large-scale applications, in particular, where the processes are more demanding, intelligent methods and algorithms should be applied, for handling and confronting efficiently these processes. These basic requirements and the methods that could be applied for the performance optimization are described in the following paragraphs.

Avatar Movement within the virtual space: In the majority of NVE applications, the end user has the ability to navigate within the virtual space. In Large-scale NVEs, where the sense of space is enlarged, leading to increment of the movement possibility, intelligent algorithms should be applied for the optimum handling of these processes. In particular, techniques and algorithms, as dead reckoning, could be applied for predicting and interpolating the users’ position within the virtual space; so as to reduce the number of sent information.

Initial Transmission of the virtual space: The initial transmission of the virtual space to the end users is one of the most demanding processes, as a heavy amount of information should be transmitted over the network. In large-scale NVEs, in particularly, where the amount of sent information is much bigger, due to the size of the virtual space, algorithms could be applied for the smoother and more efficient transmission of the required information. In particular, the initial transmission of the virtual space could be coped with algorithms that perform an efficient compression to the sent packets so that the waiting times until the world is accessible to the user can be reduced as well as with the use of streaming for the geometry data.

Sending events: For the sending of the events, compression techniques could be applied on the packets, so as to reduce their sizes and facilitate their transmission over the network.

Receiving events: As in the case of sending the events, where compression algorithms could be applied, in the receiving events process, decompression algorithms could be applied for the efficient restoration of the compressed packets.

Performing Updates: The process of performing the necessary updates in a large-scale virtual environment is exponentially incremented in a NVE as the concurrent number of users is tremendously increased. For handling the increased number of the updates, methods for the merging of multiple synchronization messages could be applied. In addition, as the virtual space is importantly enlarged and based on the fact that certain messages affect certain users within the virtual environment, methods for sending the synchronization messages to a subgroup of clients, for a certain area of the Virtual World could be applied, so as to avoid supererogatory information.

Object pre-processing: The objects that constitute the virtual space and the level of detail in which they are designed, play an important role for the users perception as well as for the achievement of realism. However, the more detailed the objects representation, the more resource demanding a virtual environment is. Therefore, for the optimization of the performance without affecting the realism, algorithms for a levels-of-detail creation for the 3D objects could be applied, as well as 3D mesh compression algorithms for reducing the size of the objects. In addition, algorithms for the delayed representation of remote entities, in regard to the network latency could be applied so as to efficiently overcome possible network deficiencies.

5. Proposed Architecture

The adaptation of the traditional architectures for the maintenance of large-scale virtual environment and the support of their basic requirements will finally lead to the same state; overloading and exhaustion of resources of the respective task carrier. In particular, in the case of client-server architectures the burden would fall into the server while in the case of peer-to-peer architectures the client would need to face a possible resource starvation. One solution for facing these limitations would be the hybrid solutions, which could adopt different types of communication and connection for different types of services. The nature of large-scale virtual environments would, however, acquire a huge amount of available resources, which could be partially faced with clustering. However, clusters are comprised by a fixed number of processors and resources and are usually based on a single type of operating system and processors [7]. Thus, the adaptation of clusters for the support of large-scale Networked Virtual Environment will face scalability and possible operability limitations.

This is the point where GRID comes to support these demanding applications. In particular, as presented in [7], the Grid infrastructure could support large-scale environments as it allows the distribution of workloads among multiple types of operating systems and processors with a dynamic number of resources. For understanding the proposed architecture for the support of large-scale NVEs over the Grid we should first describe the basic layers [7] of the Grid Infrastructure, which are displayed in Figure 1.

The base layer of the Grid is the network. Above that layer there is the Grid Resources layer, which is comprised by the hardware parts and resources available, such as supercomputers and storage systems. These could be connected to the network, which assures the connectivity of the available resources in the Grid Environment. On the top of the Grid Resources lies the Grid Middleware that constitutes the intelligent part of the Grid. In particular, the Grid Middleware is comprised of all the necessary tools for the communication and procession of the available elements. The highest part of the Grid Infrastructure is what the end users participating in the system will see. This layer is comprised of the application and services that run on the top of the Grid.

Based on the above, for supporting large-scale Networked Virtual Environments over the Grid, the proposed architecture is presented in Figure 1. The key concept is to split the processes, which realize the basic requirements into independent threads. Each of these threads is responsible for the completion of certain tasks that could be executed in parallel without causing inconsistencies in the state of the virtual environment. Thus, the processes and the algorithms for the optimization of the performance for a large-scale networked virtual environment, which were described in the previous sub-section, should be “transferred” and implemented to the GRID middleware layer. The benefits of this approach in regard to the basic requirements are the described in the following paragraphs.

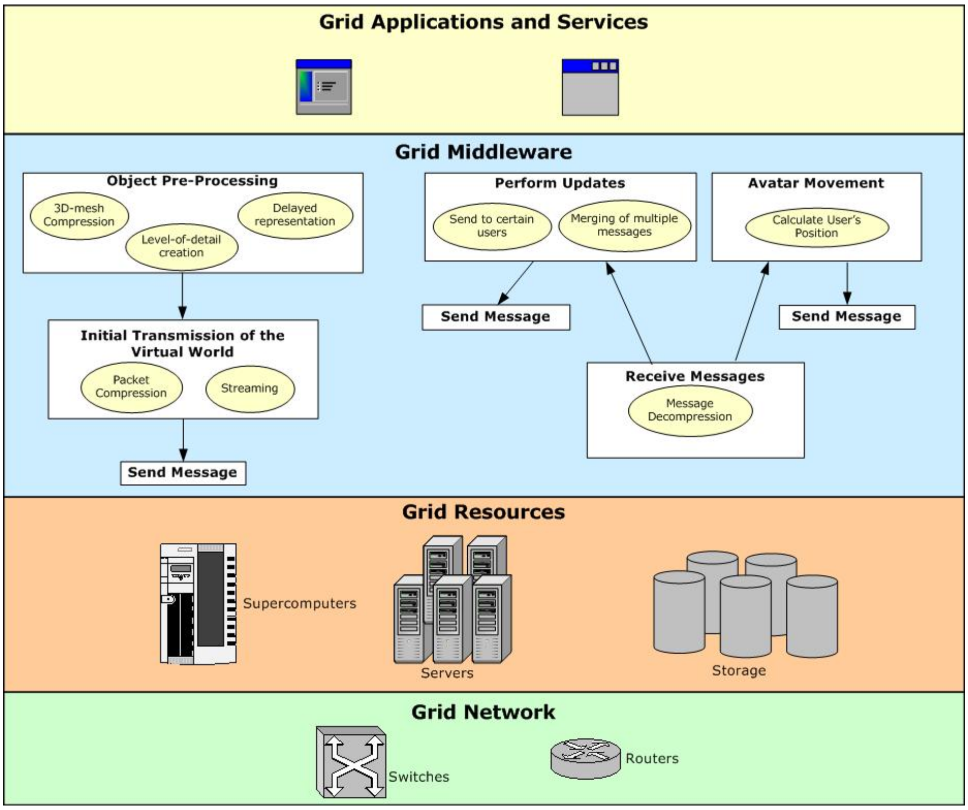

Figure 1: GRID Architecture and Networked Virtual Environments

First of all, the Grid infrastructure allows a new approach on the maintenance of the shared state of the Virtual Environment. In particular, the state of the virtual environment could be distributed among multiple threads. This distribution could be based on various factors and could concern the assignment of dedicated threads for the storage of the state for specific areas of the virtual environment, for specific objects within the environment of even specific users. This approach offers a high level of scalability at the cost of additional communication of the respective threads within the Grid Environment.

The pre-processing of the objects that constitute the virtual environment is a highly demanding process, especially when compression is also involved. However, this process concerns each individual object and is realized before the placement of the object in the virtual environment. Therefore, the pre-processing methods described above could efficiently be distributed within the Grid avoiding additional complexity or risks for the consistency of the virtual world.

The initial transmission of the Virtual Environment is another resource demanding process, especially for large-scale environments where both the size of the world and the number of the objects is massive. However, with the Grid environment, where dedicated threads could support certain areas of the virtual world for example, the size of the transmitted information is reduced and in parallel, exploiting the resources that Grid offers, streaming algorithms can effectively be adopted.

The performing of the updates as well as the avatars' navigation within a Large-scale Virtual environment mainly affects the network traffic as both the number of the objects and users is tremendously increased. The ability of dedicated threads along with the resources available pleads, once again, for the adaptation of the Grid infrastructure.

The adaptation of the proposed architecture, which is based on a high level of parallelism, allows and enables the use of intelligent algorithms. The application of the above-mentioned algorithms to the GRID level could confront the performance boundaries that both the increased size of the world and concurrent number of users introduce, ensuring a more efficient resource distribution and leading to an optimized network performance.

6. Conclusions and Future Work

The wide adaptation of Virtual Reality for the development of a wide variety of interactive applications and the users' familiarization with this emerging technology led to the need for large-scale applications. The concept of large-scale refers both to the size of the virtual world as well as to the number of the concurrent users. The initial efforts being made for moving to this new trend, mainly target to the expanding of traditional architectures, originally applied to small-scale applications. However, the majority of these applications seemed unable to overcome the demanding needs in resources and bandwidth that large-scale environments introduce. At the same time another new emerging technology made its presence for supporting demanding networking applications, which required a large amount of recourses and small execution time. GRID technology formed a promising trend and soon applied to a variety of applications.

The main focus of this paper concerns Desktop Networked Virtual Environment, which are more popular to the majority of the end users and presents a theoretical infrastructure that could allow the adaptation of GRID technology to support large-scale desktop virtual environments. To this direction, this paper, based on the basic requirements that a general Networked Virtual Environment should support, makes an approach that suggests the way that GRID could be adopted for supporting Large-Scale applications. The main concept of this approach is transforming the basic processes into independent threads, each of which carries certain tasks within the Virtual Environment, so as to achieve a high-level of parallelism. This parallelism could be then exploited by the GRID infrastructure. From the above, it becomes clear the GRID could form an efficient infrastructure for the support of Large-Scale Networked Virtual Environments.

Our future work is to apply the proposed architecture to an implemented Networked Virtual Environment, called EVE [9]. In particular, platform EVE adopts client-multi server architecture for supporting multiple, concurrent, multi-user virtual environments. Therefore, the next planned step is to re-design the processes of EVE so as to form independent threads, each of which will undertake certain type of functionality. While re-designing the code, its compatibility with the software of the GRID middleware (e.g. Globus) should be taken into account.

7. References

[1] Earnshaw R., Guedj R., Van Dam A., Vince J., "Frontiers of Human-Centred Computing, Online Communities and Virtual Environments", ISBN 1-85233-238-7, Spriner, 2001

[2] Zyda, M.J., "Networking large-scale virtual environments", Computer Animation '96, Geneva 3-4 June 1996, Page(s): 1-4

[3] Don McGregor, Andrzej Kapolka, Michael Zyda and Don Brutzman "Requirements for Large-Scale Networked Virtual Environments," Proceedings of the 7th International Conference on Telecommunications ConTel 2003, Zagreb, Croatia, 11-13 June 2003, pp. 353-358

[4] Pedro Morillo, Marcos Fernández, N. Pelechano, "A Grid Representation for Distributed Virtual Environments", European Across Grids Conference 2003: 182-189

[5] "Butterfly.net: MMG technology platform for online games", available at http://www.butterfly.net/

[6] RealityGrid Project, http://www.realitygrid.org/

[7] Force 10 Networks, http://www.force10networks.com

[8] The CrossGrid Project, http://www.eu-crossgrid.org

[9] EVE (Educational Virtual Environments) prototype, http://ouranos.ceid.upatras.gr/vr/