Journal of Theoretical and Applied Information Technology
15th December 2012. Vol. 46 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A DATA COLLECTION MECHANISM OF GRID-BASED DISTRIBUTED SIMULATION PLATFORM

[1]PING KUANG, [2]HAI JIN,[3]PINGPENG YUAN, [4]WENZHI CAO
[1]Services Computing Technology and System LabSchool of Computer Science and Technology
[2]Huazhong University of Science and Technology, Wuhan, 430074, China,
E-mail: [1]hjin@mail.hust.edu.cn

ABSTRACT

With the growth of technology, more and more applications and experiments need to be run in distributed environment. Data collection as an application and experiment fact begins to become more and more important. However, data collection will confront a lot of difficulties in grid-based distributed simulation platform because the running environment of grid is different with that of the traditional simulation platform. In order to overcome these disadvantages, we present a data collection mechanism of grid-based distributed simulation platform. In order to justify the feasibility and the availability of this data collection mechanism, a series of experiments have been done. The results show that it is feasible to collect data resource in grid-based distributed simulation platform.

Keywords: *Grid computing, Simulation, Data collection, HLA (High Level Architecture), RTI (Run Time Infrastructure).*

1. INTRODUCTION

With the growth of technology, more and more applications and experiments need to be run in distributed environment. However, people can not provide corresponding distributed running environment for these applications and experiments owning to the limitation of various conditions. Under the situation, simulation method is usually used. Thus, distributed simulation technology is rapidly developed, and has become the third tool for recognizing and rebuilding the objective world besides the theory research and experiment research. The main function of grid is to realize large-scale information resource sharing. By using the virtualization technology, grid can provide the running environment for various distributed simulation applications. However, the traditional simulation technology is mainly based on a single computer, not grid. Therefore, there are a lot of technologies and methods need to be studied in a grid-based distributed simulation platform. In fact, data collection is an important function in a grid-based distributed simulation platform, which can effectively record simulation data and playback simulation process. Therefore, it is a very important research domain that how to build a data collection mechanism in a grid-based distributed simulation platform.

The rest of this paper is organized as follows. We discuss the related work in section 2. In section 3, we present a data collection mechanism of grid-based distributed simulation platform. In section 4, a series of experiments and their results analyzing are described in detail. Finally, the conclusions are drawn and the future works are described in section 5.

2. RELATED WORKS

In order to explore some efficient data collection methods for various applications, a lot of strategies and mechanisms are presented. These strategies and mechanisms can be simply shown as follows. HLA (High Level Architecture) [1] [2] has been accepted and authorized as an exact distributed simulation standard by IEEE in 2000. HLA not only separates simulation program logics from low level communication mechanism, but also simplifies the operability among simulation applications and enhances the reusability of simulation models. In [3], the issues of data collection for simulation model are discussed which are related to the aerospace and automotive manufacturing industry. In that paper, the simulation method building a data collection system and the possible solutions that can improve data collection are presented. In the web-based data collection system, each grantee is

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

provided a unique username and password to access the system. Grantees only have access to their own data [4]. In [5], a method of distributed simulation data collection with Web Services is proposed, which is combined with temporal data model. By using the method, some temporal data can be efficiently collected for HLA distributed simulation. In the web-based data collection system, each grantee is provided a unique username and password to access the system. Grantees only have access to their own data [6]. In [7], the discrete-event simulation method is developed in order to overcome the performance bottleneck of data collection. In [8], a fully distributed and federate-integrated method is proposed in order to achieve collecting the whole state. In the method, the synchronization of the federals can be achieved, especially in the replay. Based on the method, the corresponding architecture is described for data collection.

Although the above methods can provide the simulation environment for data collection and can collect and analyze the relative information for data, they will confront a lot of difficulties in grid-based distributed simulation platform because the running environment of grid is different with that of other simulation platform. In order to overcome these disadvantages, we present a data collection mechanism for grid-based distributed simulation platform.

3. DATA COLLECTION MECHANISM

In order to implement the simulation function of data collection in grid platform [9], we need to propose a standard and scalable data collection model and design some related data collection modules for data collection mechanism. The data collection mechanism can be described as follows.

3.1 Data Collection Model

In order to implement the simulation function of data collection in grid platform, it is necessary to further apply a formalization expression method to describe the behavioral characteristic of the systems. By using mathematic theory and XML, we employ the hierarchy language to support the description. The behavioral of each client is expressed as GC=<O, E, A, S>, where O is the set of the objective of client, E is the set of detectable events, A is the set of the internal action of client, S is the set of the state of data collection. Each element of GC is described as follows.

O=<O_1: ProvideData; O_2: CollectData>

E=<E_1: Receive Subscribe Object Request; E_2: Receive Subscribe Interaction Request; E_3: Receive Transferred Subscription Request; E_4: Receive Client Data Request; E_5: Receive Transferred Client Data Request>

A=<A1: Subscribe Simulation Object; A_2: Publish Subscribe Transfer Interaction; A_3: Reflect Object Attribute Updates; A_4: Receive Simulation Interaction; A_5: Publish Subscribe Transfer Interaction; A_6: Parse Transferred Subscribe Interaction; A_7: Publish Client Data Request; A_8: Parse Transferred Client Data Request; A_9: Validate Request; A_{10}: Process Data Request; A_{11}: Refuse Request; A_{12}: Send Data>

S={ S_1, S_2, S_3,…, S_n }, S_i=<0: Idle; 1: Busy>, i=1,2,…,n, S_i denotes the i-th state

For the grid-based distributed simulation platform, the data collection model applied to generate optimized joint logging program is essential for completing collection of simulation data. Based on the formal description on behavioral characteristics of each client, we propose a cooperative reinforcement learning algorithm of data collection model which derives from Q-learning algorithm.

The Q-learning algorithm [10][11] has already been widely utilized for cooperative decisions in multi-client system. As a dynamic difference algorithm, it can be denoted as:

$$Q_{t+1}(S_i, a_i)=(1-x) Q_t(S_i, a_i)+x(r(S_i, a_i)+y \max(Q_t(S_i^*, a_i))) \quad (1)$$

where t denotes the t-th iteration, x is the iteration factor, $Q(S_i, a_i)$ is the value of evaluation function for client to execute action a_i at state S_i, $Qt(S_i, a_i)$ is the t-th iteration value of $Q_t(S_i, a_i)$ in term of equation (1), $r(S_i,a_i)$ the number that the client obtained reward when he executes action ai at state S_i , y is the affection factor, S_i^*={ S_1, S_2, S_3,…, S_i }, $Q(S_i^*, a_i)$ is the value of evaluation function for client to execute action ai at state Si and advance to its state S_i, $Qt(S_i^*, a_i)$ is the t-th iteration value of $Q(S_i^*, a_i)$ in term of equation (1).

In the problem space represented by a finite state set S={ S_1, S_2, S_3,…, S_n}, let all clients have the same activities set A. In light of the activity strategy AS, the given agent at current state $S_i \in S$ can obtain reward y by taking action $a_i \in A$.

Based on the reward r (S_i, $_{ai}$), client obtains its corresponding value of evaluation function Q: $S\times A\rightarrow P$ and determines the amended activity strategy according to the evaluation function value

Journal of Theoretical and Applied Information Technology
15th December 2012. Vol. 46 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Q. For the logging client in the grid-based distributed simulation platform, we assume that each client is able to self-regulate its own behavior in order to maximize its utility. Therefore, each client selects its own optimal activity strategy to form optimized joint activities. Thus, the reinforcement learning algorithm about data collection of client i is described as follow:

$$Q_{t+1}^{K}(S_i, a_i) = (1-x)\, Q_t^{K}(S_i, a_i)+x(r(S_i, a_i)+y \max(Q_t^{K}(S_i{*}, a_i{*}))) \quad (2)$$

Where, $\max(Q_t^{K}(S_i{*}, a_i))$ represents the highest value of evaluation function to execute all clients joint activities ai* at state S_i.

Considering the incomplete information among clients into account, it is difficult to calculate in Formula (2) by Nash equilibrium to construct iterative learning rule. In order to solve this question, we propose independent learning to construct the iterative learning process of each logging client. For each logging client, the iterative learning rules completely depend on its current state information and the received reward by taking its own activity strategy. For a given client, it adopts the following iterative process:

$$Q_{t+1}^{K}(S_i, a_i)=\max\{Q_t^{K}(S_i, a_i), (r(S_i, a_i^{*})+y \max(Q_t^{K}(S_i{*}, a_i)))$$
(3)

In Equation (3), let the reward value of each client by taking local action strategy equal to the reward to execute joint actions, that is:

$$r(S_i, a_i^{*})= r^{k}(S_i, a_i)= r^{i}(S_i, a_i)$$
(4)

Where, $\forall_i, k\in m$, $a_i^{*} = (a_1, a_2, a_3 ,\ldots a_i,.., a_k,.., a_m)$.

Let the rule for updating strategies be:

$$AS_0^{K}(S_0)=a_0\in A$$
(5)

$$AS_{t+1}^{K}(S_i)=\begin{cases} AS_t^{k}(S_i) & S_i \neq S_0 \\ \max(Q_t^{k}(S_i,a_i)) = \max(Q_{t+1}^{k}(S_i,a_i)) & \end{cases} \quad (6)$$

Equation (5) and (6) indicate that a client will keep its activity strategy unless Q values are updated. Let each client takes this policy, a joint strategy of all clients can be obtained as $AS{*}(S_i)=(AS^{1}(S_i), AS^{2}(S_i), \ldots, AS^{m}(S_i))$,where m denotes the number of clients.

3.2 Data Collection Module

In grid-based distributed simulation platform, the simulation supporting environment of data collection needs to combine the grid technology and HLA standards. Before the whole simulation system begins to collect data, the auto deployment of simulation service, the auto configuration of RTI environment and the auto assembly of simulation service are necessary. During the whole simulation system collects data, the communication of grid services, the monitoring of the simulation procedure and the auto recording of the simulation data are also necessary. In addition, in order to ensure the post running function, the replaying procedure of simulation and the procession procedure of data collection are necessary. Based on these requirements and our proposed data collection model, we develop a data collection system of grid based distributed simulation platform. In the data collection system, there includes three main modules, which are related to the function of data collection. The three modules are the simulation application module of client, the data collection module of server and the data collection module of client. The architecture of whole data collection system is illustrated as Figure 1.

4. EXPERIMENTS AND RESULT ANALYSIS

4.1 Series Of Experiments

In order to test the validity and availability of the data collection mechanism of grid-based distributed simulation platform, we build two systems. In the two systems, we use ChinaGrid to construct a distributed environment. We adopt the simulation method and our proposed data collection mechanism in the first system. In the second system, we only use the simulation method and do not adopt our proposed data collection mechanism. The settings of each experiment are shown in Table 1.

In our experiments, the task of client 1 is to collect the request information and the service information about management service, the task of client 2 is to collect the request information and the service information about record service, the task of client 3 is to collect the request information and the service information about playback service, the task of client 4 is to collect the request information and the service information about grid portal. By using the grid monitor in the two different systems, we can get the total number and the average speed that each client collects data information in [0, 100s]. The results are shown in Table 2 and Table 3, respectively.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

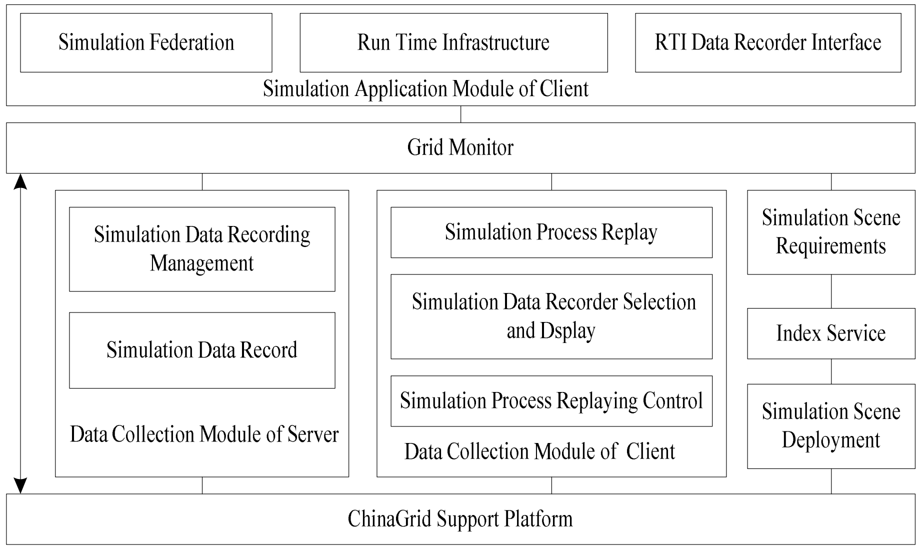

Figure 1 Architecture Of Data Collection System

Table 1. Software And Hardware Deployment

Services	Node	OS	CPU (GHz)	Memory (MB)
Manage-ment	client 1	WinXP	Intel CPU 1.70	1024
Record	client 2	WinXP	Intel CPU 1.70	1024
Playback	client 3	RedHat 9.0	Intel CPU 1.70	1024
GDSP Portal	client 4	RedHat 9.0	AMD 2500+	1024

Table 2. The State That Different Client Collects Data Information In The First System

Node	Total number (MB)	Average speed (MB/s)
client 1	1200	12.0
client 2	1580	15.8
client 3	1356	13.56
client 4	2179	21.79

In addition, we can also get the data loss ratio of each experiment in two different systems. The results are shown in Table 4.

4.2 Results Analysis

From Table 2 and Table 3, we find that the total number and the average speed that each client collects data information in the first system are bigger than that in second system. Moreover, the data loss ratio of each client in the first system is smaller than that in the second system in the two different experiments from Table 4. All these show that the function of data collection in the first system is well than that in the second system. The main reason is that in the first system, we adopt our proposed data collection mechanism and simulation method. But, we do not adopt our proposed data collection mechanism and simulation method in the second system.

Table 3 The State That Different Client Collects Data Information In The Second System

Node	Total number (MB)	Average speed (MB/s)
client 1	997	9.97
client 2	1230	12.3
client 3	1025	10.25
client 4	1756	17.56

Table 4 The Data Loss Ratio In Different Experiment

Node	The first system (%)	The second system (%)
client 1	17.8	35.7
client 2	20.5	45.3
client 3	16.4	33.9
client 4	13.2	27.8

5. CONCLUSIONS AND FUTURE WORKS

In this paper, a data collection mechanism has been proposed to overcome the drawbacks of existing data collection mechanisms in grid-based distributed simulation platform. The architecture of whole data collection system is presented and the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

key techniques involved in the design and implementation process are discussed in detail. Based on the formal expression, we combine with grid technology and HLA technology to increase data access performance, which effectively supports the simulation analysis. Furthermore, the data collection system adopts our proposed data collection model and our designed system modules. Experiment results show that the data collection system is feasible to collect data resource in grid-based distributed simulation platform.

Although the mechanism is feasible to collect data resource in grid-based distributed simulation platform, we don't analyze the performance of data collection. Therefore, in the future works, we will research the performance evaluation method of data collection in order to improve the data collection performance in grid-based distributed simulation platform.

REFERENCES

[1] W. G. Wang, Y. P. Xu, X. Chen; Q. Li; W. P. Wang, "High level architecture evolved modular federation object model", Journal of Systems Engineering and Electronics, 20 (3), (2009) 625-635

[2] A. Mccoy, P. Ward, S. Mcloone, and P. Delaney, "Multistep-ahead neural-network predictors for network traffic reduction in distributed interactive applications", ACM Transactions on Modeling and Computer Simulation, 17(4), (2007) 16-25

[3] N. Robertson and T. Perera, "Automated data collection for simulation", Simulation Practice and Theory, 9 (2002) 349-364

[4] R. Atanda, L. P. Mattia, and A. Benton, "Developing a data collection system", Evaluation Program Planning, 28 (2005) 335-339

[5] Z. B. Wu, "Temporal data collection for distributed simulation based on web services", 2nd International Conference on Networking and Digital Society (ICNDS), 2010, IEEE Press, Wenzhou, China, May 30-31. 303 – 307

[6] R. Atanda, L. P. Mattia, and A. Benton, "Developing a data collection system", Evaluation Program Planning, 28 (2005) 335-339

[7] A. Ingemansson and J. Oscarsson, "Discrete-event simulation and automatic data collection improve performance in a manufacturing system", Proceedings of the Winter Simulation Conference, IEEE Press, Orlando, FL, 2005, 1441-1445

[8] Emrah Kaya, Fatih Erdoğan Sevilgen, "A fully distributed data collection method for HLA based distributed simulations", Proceedings of the Summer Computer Simulation Conference, July 2009, 337-347

[9] M. Z. Xiong, D. Chen, H. Jin, S. Wu: Multi-path based Algorithms for Data Transfer in the Grid Environment. New Generation Computing. 28(2), (2010) 129-136

[10] K. Lazhar, T. Farid, B. Kamel, and A. Y. Amur, "Mobile robot navigation based on Q-learning technique", International Journal of Advanced Robotic Systems, 8(1), (2011) 45-51

[11] W. Ying and C. W. D. Silva, "Sequential Q-learning with kalman filtering for multirobot cooperative transportation", IEEE/ASME Transactions on Mechatronics, 15(2), (2010) 261-268