DETERMINATION OF REPLICATION PARAMETERS IN THE PROJECT OF THE VOLUNTARY DISTRIBUTED COMPUTING NETMAX@HOME

ОПРЕДЕЛЕНИЕ ПАРАМЕТРОВ РЕПЛИКАЦИИ В ПРОЕКТЕ ДОБРОВОЛЬНЫХ РАСПРЕДЕЛЕННЫХ ВЫЧИСЛЕНИЙ NETMAX@HOME

PhD, Kurochkin I.[1]
Institute for Information Transmission Problems RAS[1], Moscow, Russia

Abstract: *The project of the voluntary distributed computing NetMax@home on the BOINC platform was developed. At a project definition phase to public access some features of functioning of the project were revealed. Determination of replication parameters and modification of the application allowed to achieve increase of efficiency of calculations in the project.*
KEYWORDS: VOLUNTARY DISTRIBUTED COMPUTING, BOINC, GRID-SYSTEM, LOAD BALANCING SYSTEM, REPLICATION

1. Introduction

Use of supercomputers imposes serious financial obligations on researchers as it is necessary to buy time of high-performance cloud system or to buy the supercomputer, and also to organize and support supercomputer infrastructure. In this regard the concept of the distributed computing and use of grid-systems can solve a problem with big computing complexity and do without essential financial expenses.

The BOINC platform (Berkeley Open Infrastructure for Network Computing) is an open non-commercial software for the organization of the voluntary distributed computing on personal computers. BOINC has client-server architecture and consists of client part and server. Is a universal platform for calculations in various areas of science (mathematics, molecular biology, medicine, astrophysics, telecommunications, etc.) . The client part can be established on all widespread operating systems: Microsoft Windows, Linux, Mac OS, Solaris, FreeBSD, etc. The server part is intended for management of the project – the distributed computing on a certain computing task. Projects of the distributed computing on bases of the BOINC platform share on 2 types:

1. public projects with participation of volunteers;
2. closed (internal) projects with use of the available computing means.

Features of control of parameters of public projects of the voluntary distributed computing will be considered further.

2. Prerequisites and means for solving the problem

The mathematical model of functioning of a telecommunication network is developed [1, 2]. Application for simulation of mathematical modelling of the functioning of a telecommunication networks was created. Using the application and analysis of results revealed that need a large quantity of experiments with different sets of parameters of the initial network. In view of the large number of similar calculations with different initial parameters to reduce the time a series of experiments distributed version of the application was developed.

Use of the international project of the voluntary distributed computing NetMax@home allowed to involve a large number of computing capacities [3]. Despite successful carrying out preliminary numerical experiment, the following general restrictions were revealed:

• heterogeneity of nodes of the distributed system, and as a result different speed of calculation;

• autonomy of calculations on various nodes;

• changeable time of continuous work of nodes and impossibility of calculation of long tasks;

• impossibility of continuous coordination of computing between nodes;

• existence of mistakes and delays at calculations;

• need of fine control of system of load balancing.

Owing to these restrictions, and also rather big time of performance of some subtasks on nodes (till 30-40 hours) efficiency of use of computing nodes decreased [3, 4].

Correctly picked up parameters: the quantity of tasks and number of computing knots which one task is given, are capable to reduce considerably time of calculation and to increase quantity of successfully complete tasks.

3. Solution of the examined problem

To determine some parameters of load balancing system (tasks sending system) in the project of the voluntary distributed computing it was decided to carry out mathematical modelling of process of task send to nodes of grid-system. In order that mathematical modelling was most approached to the real working systems, data from the international projects of the distributed computing of Gerasim@home [5], SAT@home [6], NetMax@home [3] were used.

Correctly picked up parameters of load balancing system and parameters of replication are capable to reduce considerably time of calculation and to increase quantity of successfully complete tasks [7, 8]. When carrying out experiments the problem of emergence of "tails"(Fig. 1) when the remained small amount of tasks calculate too long time of rather general time of calculation of tasks was revealed.
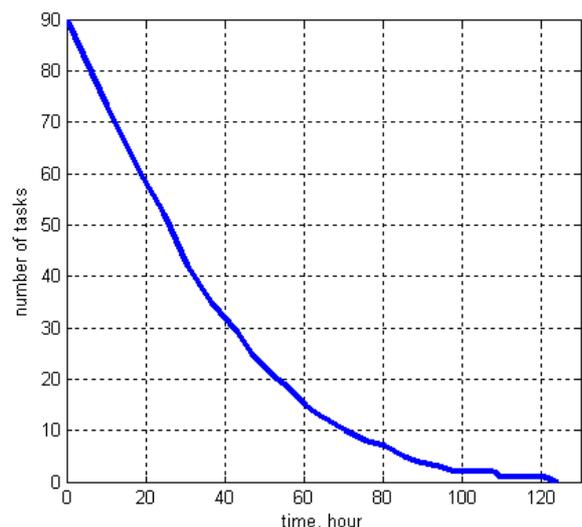

Fig. 1. 'Tail" of calculation.

When modelling the following parameters were considered: number of users (computing nodes), an interval of time of performance of one task, and also percent of refusal (error) when performing a task, number of initial copies of one task. The percent of refusal is set by sigmoidal function $F(t)$ of increase in percent of refusal depending on time of processing of a task. The interval of

time of calculation of each task is normally distributed random variable. Refusal function:

$$F(t) = a_0 + \frac{(1 - a_0)a_2}{1 + \exp(-(t - a_3)a_1)} \quad (1)$$

where a_0 – basic percent of refusal of performance of a task, a_3 – depends on an interval of values of duration of performance of tasks.

On the basis of data on primary numerical experiment in the NetMax@home project parameters of function of refusal were determined as a_0=0.15; a_1=0.7; a_2=1; a_3=10.

4. Results and discussion

Increase of computing power of grid-system possible in two ways:

1. Increase of efficiency of the existing nodes;
2. Increase in quantity of computing nodes.

Increase of efficiency of use of the existing nodes of grid-system was realized by selection of parameters of the project. During modelling of delivery of tasks to nodes of grid-system it was found out that length of "tail" makes to 15%.

As a result of processing of results of mathematical modelling, and also the conducted research [9] were defined: interval of values of duration of tasks (from 0.5 to 6 hours) and limit value of parameter of initial replication (3 copies). Use of values of parameter of initial replication more than 3 didn't give increase in successfully performed tasks more than for 3%, however reduced efficiency of grid-system not less than by 25%.
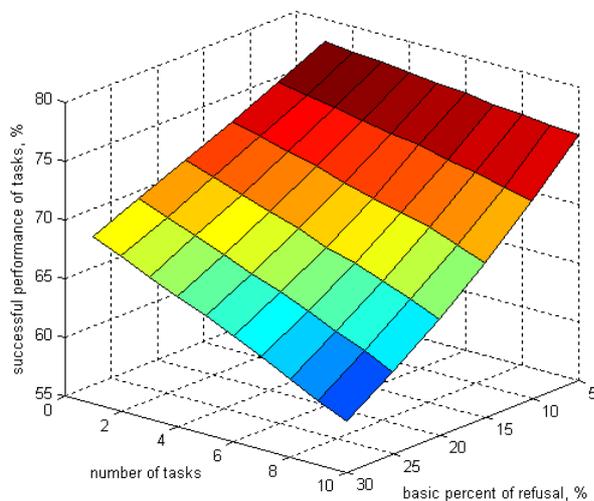

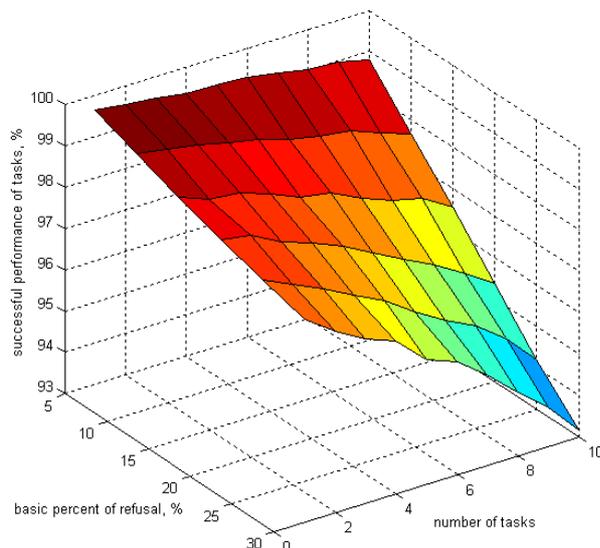Fig. 2. Percent of successful performance of tasks for 2 copies.

Fig. 3. Percent of successful performance of tasks for 3 copies.

The second way of increase of computing power of grid-system was realized by means of the following measures:
- the publication of new information on the project site;
- support of existence of a large number of the tasks ready to be sent;
- providing feedback with administration of the project;
- existence of checkpoints when performing tasks on computing nodes.

Modification of the application according to the revealed features was required. It is possible to carry a preliminary estimate of time of calculation of a separate task, formation of results of calculations of a certain size, desirable use of checkpoints. After a stage of testing planning of experiment was corrected and a number of steps for acceleration of computation of a totality of tasks of experiment were undertaken.

Modification of the application allowed to achieve increase of efficiency of calculations in the project of the voluntary distributed computing NetMax@home.

5. Conclusion

Control of parameters of the project of the voluntary distributed computing allows to increase not only number of volunteers (computing nodes), but also to increase efficiency of use of the existing computing nodes. It should be taken into account that each project of the voluntary distributed computing has features which have strong impact on control of parameters.

Determination of parameters of load balancing system and replication allowed to reduce computing complexity of experiments, to reduce time of calculation of "tail" and to increase efficiency of grid-system.

6. References

1. Kurochkin, Grinberg, Different Criteria of Dynamic Routing, Procedia Computer Science, Volume 66, 2015, Pages 166-173, ISSN 1877-0509.
2. Grinberg Y.R., Kurochkin I.I.: Mathematical modelling of dynamic networks of sequential filling of flows due // Problems of computations in the distributed media: Proceedings of ISA RAS, 2009, vol.46, p.233-258
3. Kurochkin I.I. The voluntary distributed computing project NetMax@home // Distributed computing and grid-technologies in science and education: book of abstracts of the 6th international conference. Dubna: JINR, 2014.
4. Francesco Asnicar, Nadir Sella, Luca Masera, Paolo Morettin, Thomas Tolio, Stanislau Semeniuta, Claudio Moser, Enrico Blanzieri, Valter Cavecchia "TN-Grid and gene@home Project: Volunteer Computing for Bioinformatics"// CEUR Workshop Proceedings. Proceedings of the Second International Conference BOINC-based High Performance Computing: Fundamental Research and Development (BOINC:FAST 2015). Vol. 1502. Technical University of Aachen, Germany, 2015. pp. 1-15
5. Vatutin E.I., Titov V.S. Voluntary distributed computing for solving discrete combinatorial optimization problems using Gerasim@home project // Distributed computing and grid-technologies in science and education: book of abstracts of the 6th international conference. Dubna: JINR, 2014.
6. Zaikin O., Kochemazov S. The Search for Systems of Diagonal Latin Squares Using the SAT@home Project // International Journal of Open Information Technologies. Vol. 3, No. 11 (2015). pp. 4-9.
7. Ivashko E., Golovin A. Partition Algorithm for Association Rules Mining in BOINC-based Enterprise Desktop Grid. Lecture Notes in Computer Science. Parallel Computing Technologies 13th International Conference, 2015, 268–272, Springer.
8. Chernov I., Nikitina N. Virtual screening in a desktop grid: Replication and the optimal quorum // 13th International Conference, PaCT 2015. Petrozavodsk, Russia, August 31 – September 4, 2015. Proceedings. — Lecture Notes in Computer Science. Parallel Computing Technologies. — Switzerland : Springer International Publishing, 2015. — P. 258–267

9. Yakimets V.N., Kurochkin I.I. The voluntary distributed calculations in Russia: the sociological analysis//In the collection: INFORMATION SOCIETY: EDUCATION, SCIENCE, CULTURE AND TECHNOLOGIES of the FUTURE Works XVIII of the joint conference "Internet and Modern Society" (IMS-2015). ITMO university, St. Petersburg, 2015. C. 345-352.