Proceedings of the VIII International Conference "Distributed Computing and Grid-technologies in Science and Education" (GRID 2018), Dubna, Moscow region, Russia, September 10 - 14, 2018

STUDY OF THE INTERACTION OF THE VOLUNTEER COMMUNITY IN DISTRIBUTED COMPUTING PROJECTS

I.I. Kurochkin

Institute for Information Transmission Problems of Russian Academy of Sciences, Bolshoy Karetny per. 19, build.1, Moscow, 127051, Russia

E-mail: kurochkin@iitp.ru

In this paper discusses methods for improving the computing power of distributed computing project using the computing power of volunteers. The methods involve attracting more volunteers and retaining their interest for a long time. The results of the study on the preferences of volunteers to highlight the most significant factors affecting the interest of volunteers are discussed. The approach of index and multiparameter estimation of volunteer distributed computing projects is discussed. A modified scoring system for performing calculations in the volunteer distributed computing project is proposed.

Keywords: desktop grid, BOINC, volunteer community, volunteer distributed computing, volunteer motivation, multiparameter evaluation, scoring system

1. Introduction

Distributed computing is a way to solve large computational problems using computers combined into a computing system. Of particular interest is volunteer computing. These are distributed computing using voluntarily provided computing resources.

There are several platforms for organizing distributed computing: Globus, HTCondor, Legion, but the most common is currently BOINC [1, 2]. Software BOINC (Berkeley Open Infrastructure for Network Computing) is an open, non-commercial software for organizing distributed computing on personal computers. On the basis of the BOINC platform, about 100 volunteer (voluntary) distributed computing projects have been deployed, to which about 16 million computers are connected worldwide [2]. Most volunteer distributed computing projects are research projects of leading world universities and research organizations.

Distributed computing projects based on the BOINC platform are divided into 2 types: public projects involving volunteers and enterprise projects using the organization's existing computational capabilities. The increase in the number of computational nodes for the enterprise desktop grid systems is carried out with the help of administrative influence. For public project of volunteer distributed computing (VDC project), the main goal is to attract new volunteers and their computing power and retain the participants of the project. To develop a set of measures to attract and retain volunteers in the VDC project, it is necessary to know not only the statistical parameters, such as the number of volunteers and the computing power of their computers, but also the motivation of the volunteers [3]. It is necessary to interact with the community of volunteers to attract attention and increase confidence in the VDC project [4]. Volunteer distributed computing have a number of features that can significantly slow down the calculations:

- Heterogeneity of computational nodes of a distributed system;
- Autonomy of calculations at various nodes;
- Unreliability of connections and possible shutdown of computational nodes;
- Inconsistent time of continuous operation of the node;
- Impossibility of continuous coordination of settlements between nodes;
- The presence of errors and delays in the calculations;
- The complexity of developing computing applications for all types of nodes.

3. Volunteer motivation

Participation in voluntary computing projects does not bring volunteers who provide their computing resources, no benefit and often requires certain costs to purchase the necessary equipment and pay for electricity. In 2014-2016, a sociological study of the motivations and preferences of participants in voluntary distributed computing in Russia [5] was organized and conducted on the basis of the Centre for Distributed Computing IITP RAS.

In Russia, about 4,000 volunteers are actively involved in the VDC project; almost 650 people responded to the questionnaire, which is more than 16%. Most of them are men (97%) aged 23 to 50 years (87%) with high (80%) education (mostly (55%) technical). The main driving factors that motivate people to participate in volunteer computing projects are: awareness of their involvement in scientific discoveries, help to science and sport interest. The question was asked about how, in general, faucets have confidence in the VDC projects. The answers of the volunteers were distributed as follows: if they have access to detailed information about the project (90%), or when they can get acquainted with the publications of the project results (88%) or with links to scientific papers (65%). And almost half of the respondents (45%) have confidence if there is feedback from the developers (administration of VDC project).

3. Index and multiparameter evaluation of VDC project

The involvement of volunteers in the project activity to a large extent depends on a number of parameters characterizing the project itself and how its work is organized. Using the methods of sociology, we have developed a toolkit for interviewing volunteers and organizers of a number of VDC projects. The use of the toolkit allowed to determine the list of the most important characteristics of the VDC projects that are essential for the fliers involved in their activities, as well as to evaluate the significance of such characteristics for different projects.

A new approach to the assessment of VDC projects was developed [6], consisting of 2 complementary parts: 1) Multi-parameter evaluation of the activities of the VDC projects from the side of the volunteers and other participants through their questioning. After processing the individual estimates given by the respondents on special scales, the group average estimates of the project activity for each of the parameters were calculated. This allowed us to graphically create a comprehensive visualization of the multidimensional “portrait” for each VDC project (Figure1).

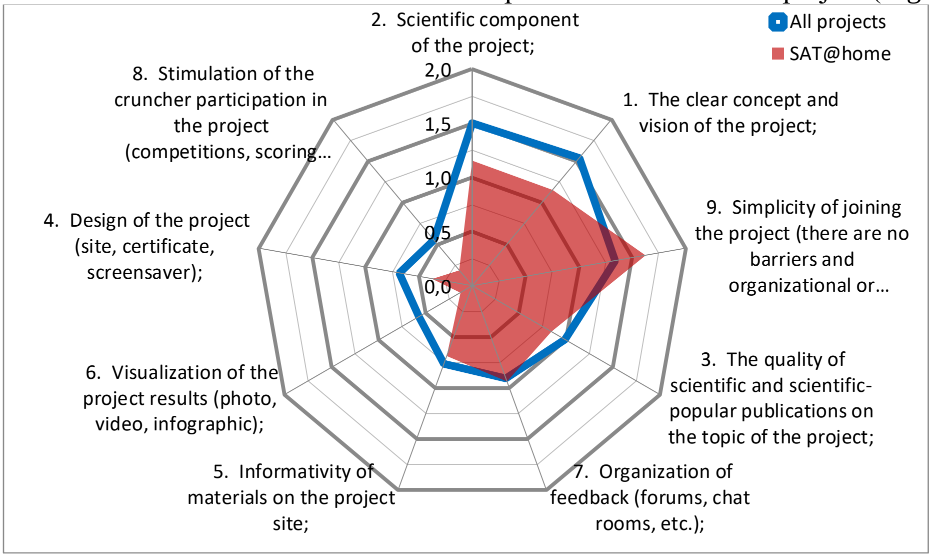

Figure 1. Multi-parameter evaluation of SAT@home VDC project

2) The calculation of the aggregate index – the YaK-index, when the average group estimates were “weighed” taking into account the coefficients of their significance (see Table 1).

Table 1. YaK-index

Project Title	YaK-index (with average weights of characteristics)	YaK-index (with individual weights of characteristics)
Folding@home	0.65	0.68
LHC@home	0.64	0.64
Gerasim@home	0.59	0.61
SAT@home	0.58	0.57

Multi-parameter assessment provided a visual visualization of the multidimensional "portrait" of the VDC project, highlighting its strengths and weaknesses. Based on such data, the project team can develop proposals for a constructive impact on the identified weaknesses of the project. Such proposals could help improve the performance of the VDC project.

4. Scoring system for VDC project

In order to maintain sporting interest among volunteers, a scoring system was introduced into BOINC, charging a certain number of credits, depending on the amount of calculations performed. Scoring or credit systems in BOINC can vary depending on the project and take into account its features, which allows you to develop the most appropriate and objective mechanisms for scoring.

To date, several different credit systems have been developed in BOINC projects. The choice of a specific implementation depends on the features of the project, the amount of calculations required for processing tasks and the degree of variation of this indicator on a set of tasks. Criteria for credit assignment:

- In proportion to the volume of calculations;
- Fixed number of credits for one task;
- In proportion to the allocated resources (not only computing);
- For the speed of task calculation;
- In addition to the main credit system, reward for quick return of the result;
- Depending on the nature of the project work (the number of results obtained, the amount of data processed, etc.).

The standard credit system is based on measuring the performance of a specific computer using special tests, as well as on the amount of CPU time spent on the task. The BOINC client, after completing the task, requests a certain number of credits from the server, which is calculated by the following formula (1):

$$Credits_{claimed} = \frac{(whetstone + dhrystone) * time_{CPU}}{1728000}, \quad (1)$$

where *whetstone* – speed of floating point calculations (FLOPS); *dhrystone* – computation speed with integers (IntOPS); $time_{CPU}$ – processor runtime in seconds.

The number of total credits is calculated as the average $Credits_{claimed}$ value of all nodes that performed this task. This system has a number of significant drawbacks. It does not always guarantee the objectivity of scoring, and is also vulnerable to “cheating” and is platform dependent. In newer versions of BOINC, this system has been improved by the following changes: by testing, the maximum possible performance of specific equipment is calculated ($peak_{FLOPS}$); waiver of processor time in favor of measuring normal time; introduction of several stages of the normalization of credits (cross-version normalization, host normalization). These innovations allowed projects to charge credits more objectively.

The system of charging a fixed number of credits is most convenient if all tasks in the project require approximately the same number of calculations. This system unambiguously solves the problem of objective scoring, but in most existing projects it is not appropriate due to large variations in the complexity of tasks. This mechanism is used in projects WUProp@home, SAT@home.

The reward **system for quick return** is based on the prediction of the number of required operations (FLOP) for each task, which allows you to assign a specific fixed number of credits for each task. In addition, this system involves a reward for the rapid execution of tasks – Quck Return Bonus (QRB). This scoring system is used in Folding@home. The total number of credits assigned to the user is calculated by the following formula (2):

$$Credits = base * max\left(1, \sqrt{k\frac{deadline}{elapsed\_time}}\right), \quad (2)$$

where *base* – a fixed number of credits for the current task, established on the basis of a preliminary calculation on the reference machine; *deadline* – maximum execution time for this task; *elapsed_time* – the actual time of the task; *k* is a coefficient set depending on the importance of the task (at Folding@home project it is 0.75).

A similar scoring system is used in the GPUGrid project. The number of credits for one task is directly related to the number of FLOPs required to process this task. All tasks are subject to a deadline that does not exceed 5 days. In addition to the basic number of credits, there is an additional reward in the form of an increase in credits by 50% in case the task is completed faster than 24 hours, and by 25% in case of completion within 48 hours.

At the IITP RAS, it became necessary to develop an alternative scoring system for the NetMax@home project, based on the SZTAKI BOINC infrastructure deployed. The input data of the scoring mechanism is a set of results that pass validation checks, as well as a set of their parameters, which include the name of the task or work unit (WU), time sent to the user, time of receipt of the processed task by the server. This validator must meet the conditions for an objective distribution of credits between users, as well as being resistant to the substitution of task processing time, elapsed processor time, number of processor cores and other parameters affecting the mechanisms for assigning credits. After analyzing the existing mechanisms for scoring, it was decided to assign credits according to the following system (3):

$$Credits = \begin{cases} base * k * (QRB + 1.0), dt < qrbThreshold \\ base * k, dt \geq qrbThreshold \end{cases}, \quad (3)$$

where, *Credits* is the total number of credits, *base* is a fixed number of credits for a task, *k* is the base credit ratio (all users are the same), *QRB* is the Quick Return Bonus factor, *dt* is the elapsed time to process the task, *qrbThreshold* is the time threshold for *QRB*.

6. Conclusion

The combined use of methods to attract volunteers will quickly increase the computational capacity of the VDC project. Monitoring of the index and multiparameter evaluation will allow making timely updates to the VDC project to maintain the interest of the volunteer community in it.

7. Acknowledgement

This work was supported by the Russian Foundation for Basic Research (grants No. 18-29-03264, 18-57-06003).

References

[1] D.P. Anderson “BOINC: a system for public-resource computing and storage”, Grid Computing, IEEE, 2004.

[2] The server of statistics of voluntary distributed computing projects on the BOINC platform. http://boincstats.com. (date of access: 30.08.2018).

[3] Clary, E. G., Snyder, M., Ridge, R. D., Copeland, J., Stukas, A. A., Haugen, J., & Miene, P. (1998). Understanding and assessing the motivations of volunteers: a functional approach. Journal of personality and social psychology, 74(6), 1516.

[4] Posypkin M., Semenov A., Zaikin O. (2012). Using BOINC desktop grid to solve large scale SAT problems // Computer Science, 13(1), 25.

[5] Yakimets V.N., Kurochkin I.I. Voluntary distributed computing in Russia: a sociological analysis. Collection of scientific articles of the XVIII Joint Conference "Internet and Con-temporary Society" (IMS-2015), St. Petersburg, June 23, 2015, St. Petersburg: ITMO University, 2015. Sc.345-352. ISBN 978-5-7577-0502-6. (in Russian)

[6] Yakimets V.N., Kurochkin I.I. Analysis of results of the rating of volunteer distributed computing projects // Russian Supercomputing Days 2018, September 24-25, 2018, Moscow, Russia: Proceedings of international conference, MSU, 2018, pp.893-908.