

Ilya I. Kurochkin¹, Vladimir N. Yakimets^{1,2}

DOI: 10.25045/jpit.v10.i2.09

¹Institute for Information Transmission Problems of Russian Academy of Sciences, Moscow, Russia

²The Russian Presidential Academy of National Economy and Public Administration, Moscow, Russia

¹kurochkin@iitp.ru, ²iakimets@mail.ru

ATTRACTION OF VOLUNTEERS IN PROJECTS OF VOLUNTARY DISTRIBUTED COMPUTING

Received: 21.12.2018

Revised: 27.02.2019

Accepted: 10.03.2019

The main methods of attracting volunteers to volunteer distributed computing projects are reviewed. The preferences of volunteers based on the results of sociological research are discussed. A set of measures is proposed for the successful deployment and maintenance of a volunteer distributed computing project on the BOINC platform. The applicability of the considered methods of attracting volunteers for smart city and e-governance is substantiated.

Keywords: volunteer distributed computing, BOINC, volunteer distributed computing projects, volunteer motivation, multiparameter evaluation.

Introduction

The grid system is a new computing infrastructure that provides uninterrupted access to computing power and storage resources distributed throughout the world. The grid system coordinates the use of resources in the absence of centralized management of these resources and uses standard, open, universal protocols and interfaces. Distributed computing infrastructures existed before the grid systems, but usually they are specialized systems. The grid system allows you to combine different types of resources, use different methods of interaction, and provide access to resources for an unlimited number of non-local users and unlimited scalability.

Desktop grid is a grid system that integrates non-specialized computers (personal computers, laptops, smartphones, etc.) over the Internet and uses their idle computing time to perform complex computational tasks. In the early 2000s, the desktop grid implied the unification of only personal computers. Today, the desktop grid is a heterogeneous system that combines not only unspecialized computers, but also multiprocessor computing systems with a large amount of hardware and software heterogeneity. Using desktop grid as a tool for high-throughput computing [1] is of great interest because of their inherent strengths, which include the following:

- comparatively easy deployment and support;
- low financial cost of deploying and supporting desktop grid;
- high scalability (hundreds of thousands of computational nodes);
- the ability to use computing nodes of various types and architectures;
- High potential peak performance.

However, there are also a number of shortcomings that significantly reduce the class of tasks that can be effectively solved in the desktop grid:

- Heterogeneity of nodes of a distributed system;
- Autonomy of calculations at various nodes and the impossibility of continuous coordination of calculations between nodes;
- Unreliability of connections and possible shutdown of computing nodes;
- Inconsistent time of continuous operation of the node and the difficulty of calculating long tasks;
- The need to adapt the computational problem to each type and to each architecture of computational nodes;
- The presence of errors and delays in the calculations.

These drawbacks lead to a significant reduction in the effective computational performance of the desktop grid compared with the peak, as well as to a significant effect of fine tuning the desktop grid on performance.

BOINC platform

There are several platforms for organizing distributed computing: Globus [2], HTCondor [3], Legion [4], but the most common is currently BOINC (Berkeley Open Infrastructure for Network Computing) [5, 6]. BOINC has a client-server architecture and consists of the client part and the server part. It is a universal platform for computing in various fields (mathematics, particle physics, astrophysics, telecommunications, molecular biology, medicine, etc.).

The client part can be installed on most operating systems: Windows, Linux, Mac OS and etc. The server part is designed to manage a distributed computing project. The server part of the BOINC project consists of several services (daemons) that send tasks and receive/validate results, a web server, and a database of input data and results. The BOINC platform provides many settings to improve the reliability of the results obtained in the project, including the parameters of replication and the distribution of subtasks to the desktop grid computational nodes. The server part of the BOINC project is deployed on the Linux operating system using the Apache web server and the MySQL (MariaDB). To improve the reliability of the server part of the BOINC project, it makes sense to deploy the server part in a virtual machine in the cloud. In this case, you can allocate adequate resources for the functioning of the server part of the project.

Figure 1 shows a schematic diagram of the interaction of users of the desktop grid on the BOINC platform with the server part of the voluntary distributed computing project.

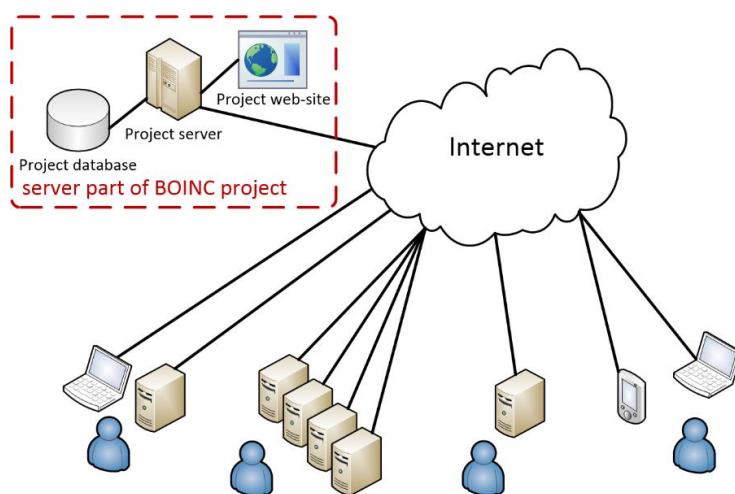


Figure 1. General scheme of desktop grid

About 80 voluntary (volunteer) distributed computing projects, to which millions of computers are connected worldwide are currently operating on the basis of the BOINC platform, [6].

Most projects have a task that can be separated into many independent subtasks. With such a partition, the calculation algorithm for each subtask is the same; only the input data set changes. This type of task is called “bag of tasks” [7], or a task with data separation. Each subtask can be calculated independently of the others, and the completion of the calculation of a numerical experiment occurs when the entire set of subtasks is performed. As an example of such tasks we can cite the following tasks: combinatorial optimization problems and exhaustive search [8], SAT-approach [9], machine learning tasks, simulation modeling and etc.

Volunteer motivation

Distributed computing projects based on the BOINC platform are divided into 2 types: public projects involving volunteers [8, 10] and enterprise projects using the organization's existing computational resources [11].

The increase in the number of computing nodes for the enterprise desktop grid systems is carried out with the help of administrative influence. For public projects of voluntary distributed computing (VDC projects), the task is to attract new volunteers and their computing power and to keep the participants of the VDC project. To develop a set of measures to attract and retain volunteers in the VDC project, it is necessary to know not only the number of volunteers and the computing power of their computers, but also the motivation of volunteers. It is necessary to identify the main factors of interest of the volunteer community in the VDC project. It is necessary to understand what exactly contributes to increasing the credibility of the VDC project.

It was decided to conduct a sociological study of the volunteers' preferences to improve the efficiency of computational experiments in the VDC projects.

Two sociological surveys were carried out at VDC projects based on BOINC platform [12]:

- a survey conducted by the authors of the platform BOINC (University of California) in 2006 (more than 30 thousand respondents, including about 500 people from Russia and Ukraine),
- a survey in 2013, organized by IBM among the participants of the World Community Grid (WCG) project, which involved more than 15 thousand respondents, including about 300 Russian-speaking [13].

In 2014, the sociological study of the motivations and preferences of participants in voluntary distributed computing in Russia was organized and conducted on the basis of the Center for Distributed Computing of the Institute for Information Transmission Problems of the Russian Academy of Sciences [14]. Let's review some of the results of the study. To survey the preferences of the Russian volunteer community, we developed a special questionnaire. It contained 26 questions, divided into the following sections:

- Experience of voluntary distributed computing;
- The interests of the volunteer;
- Attitude towards national projects;
- General information;
- Wishes for VDC projects;
- Personal identification (optional).

Each of the 15 questions of the questionnaire was half-closed, so that in addition to the suggested answers, the respondent had the opportunity to enter his or her own version.

Information support and the involvement of volunteers (crunchers) to participate in the survey were provided by the administrators of the site BOINC.RU. They managed to use almost all existing Russian-language Internet resources dedicated to the voluntary distributed computing. Due to this, a high representativeness of the survey results was ensured.

Survey results

A total of about 650 people answered the questionnaire. Most of them are men (97%) aged 23 to 50 years (87%) (figure 2) with higher (80%) education (mainly (55%) technical) (figure 3). For comparison, we note that 90% of men participated in the WCG survey, and 68% of respondents were aged 25 to 54 years.

The overwhelming majority of the respondents (96%) are volunteers – direct participants of the voluntary distributed computing, 3% of the respondents attributed themselves to the founders of the volunteer teams

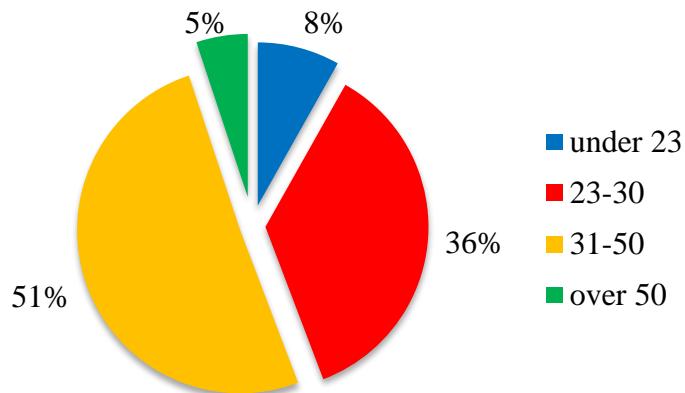


Figure 2. Age of volunteers

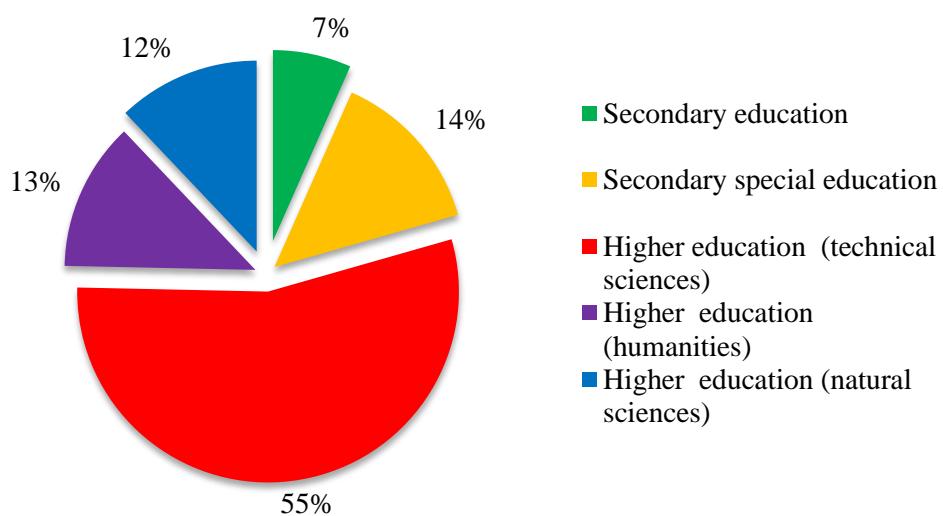


Figure 3. Education of volunteers

Figure 4 shows the distribution of the crunchers according to the length of employment of the voluntary distributed computing: half of them have been involved in voluntary distributed computing for more than 3 years, including 17% have more than 7 years of experience.

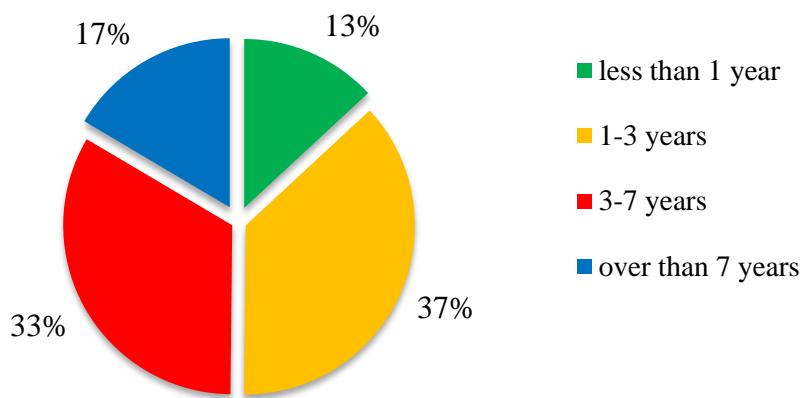


Figure 4. Involvement in voluntary distributed computing

Interestingly, in response to the question on reasons they are attracted to participate in the voluntary distributed computing (it was allowed to choose more than one answer), almost 93% of

respondents indicated their desire to help scientific research, 51% – their involvement in scientific discoveries, and 27% – just sporting interest (figure 5).

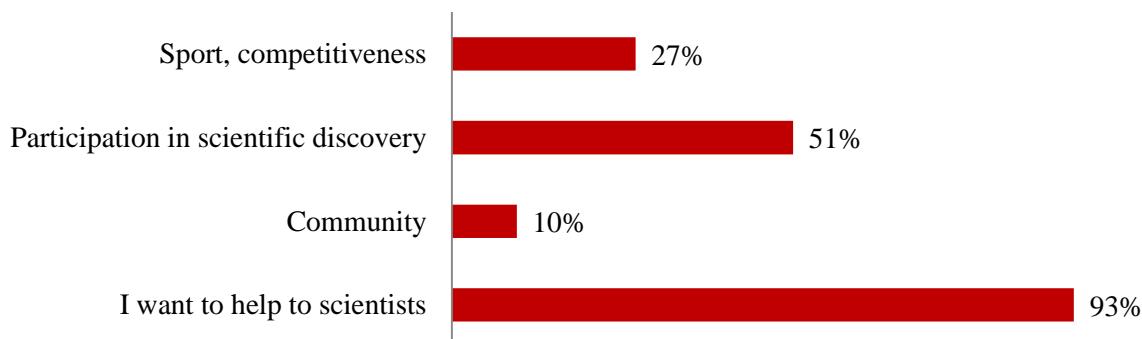


Figure 5. The reasons for participation in the voluntary distributed computing

It was revealed that the capacities available to the volunteer for the VDC were distributed as follows: 1 or 2 computers of standard configuration – 76%; 1 computing server or many computers – 17%; a lot of servers or computers – 7%. And 26% of crunchers use only CPU for computing (figure 6).

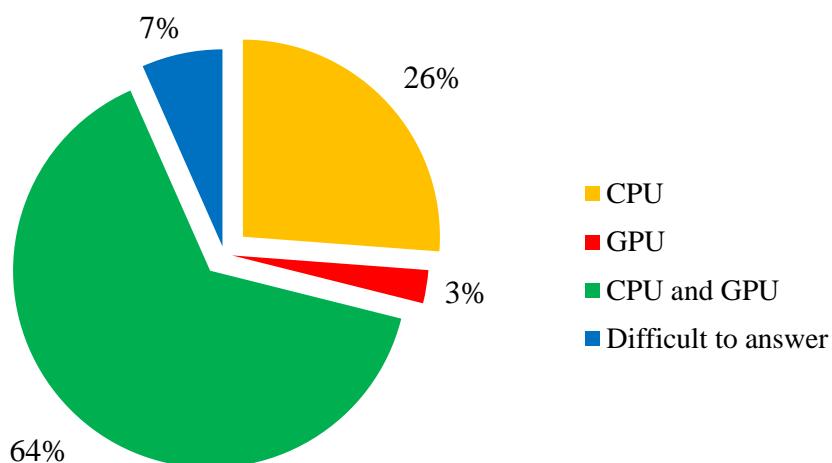


Figure 6. What volunteers use to computing

The data on the areas of the VDC projects in which the crunchers are involved in are interesting (figure 7). Two areas of knowledge dominate here: astronomy (57%) and biology / medicine (57%). A considerable number of respondents (44%) participate in projects from the field of mathematics and physics. If we also take into account that 17% of the survey participants chose the answer option “any subject”, it becomes obvious that, while carrying out the VDC in the chosen area, the fliers are interested in work in projects of other directions.

The results of a sociological survey allow us to make some generalizing conclusion regarding Russian volunteer community: these are people with a high level of education and qualifications (in various fields of knowledge); they consciously approach the choice of both the VDC project itself and the issues of interaction and relations with its organizers (scientists), and also prefer to feel as participants in research, to be aware of the problems and successes of the project they have joined; These are partners for project organizers. You can build the following averaged portrait of a Russian volunteer: this is a man (97%); aged from 23 to 50 years (87%), with a higher (80%) technical education (55%), having 1-2 computers (76%), using the BOINC platform (96%) to assist science (93%).

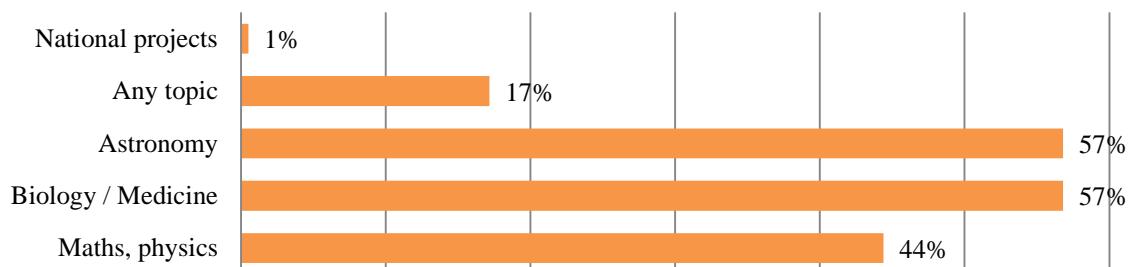


Figure 7. Topics of the VDC projects

The question was asked about how, in general, fliers have confidence in the VDC projects. The answers of the fliers 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 articles (65%) (figure 8). Half of the respondents (45%) have confidence if there is feedback from the developers of VDC project.

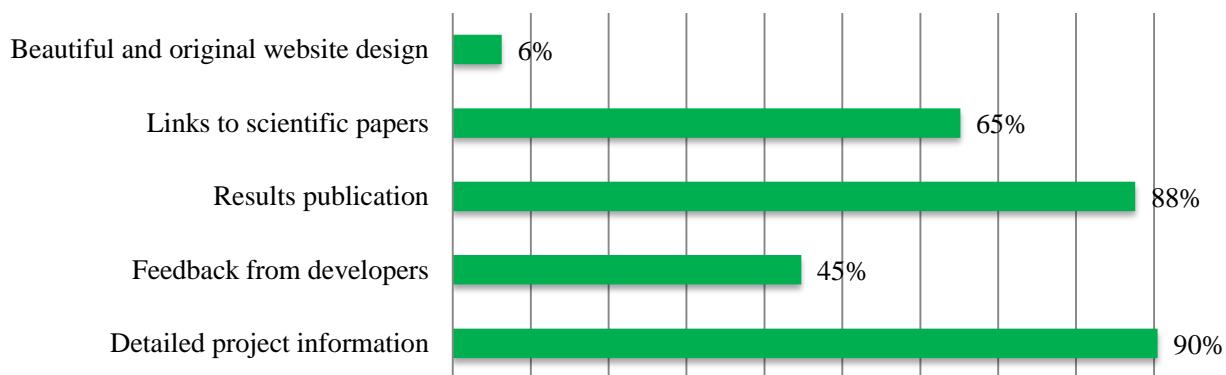


Figure 8. Trust in the VDC project

The preferences of the Russian volunteers, which they rely on, while participating in various VDC projects, were revealed. Thus, more than half of the respondents - 54% of the crunchers – are ready to participate in the implementation of national projects if they are provided with detailed information on goals and objectives, at the same time 34% are generally ready to cooperate without preliminary conditions (figure 9).

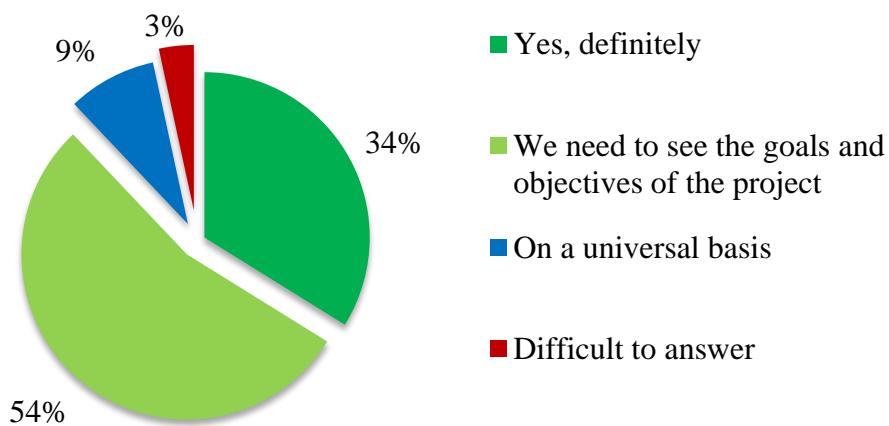


Figure 9. Attitude towards the national project

To maintain sporting interest among the volunteer community, there is a scoring system in BOINC platform that gives a certain amount of credits, depending on the calculations made. Currently, there are several different scoring systems in BOINC projects. The choice of a specific scoring system depends on the project features, the amount of computation required to process the tasks, and the homogeneity of the computational complexity of the tasks. Criteria for scoring system:

- 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.).

Volunteers have different opinions about the best and fair scoring system (figure 10).

In addition, some VDC projects involve the accrual of various virtual prizes for user contributions to the computing power of the project. These virtual prizes have the form of special images (badges) displayed on the project's web page near the user name. They symbolize various achievements in the field of calculations, for example, the total amount of calculations performed, the average daily figure, the time of participation in the project.

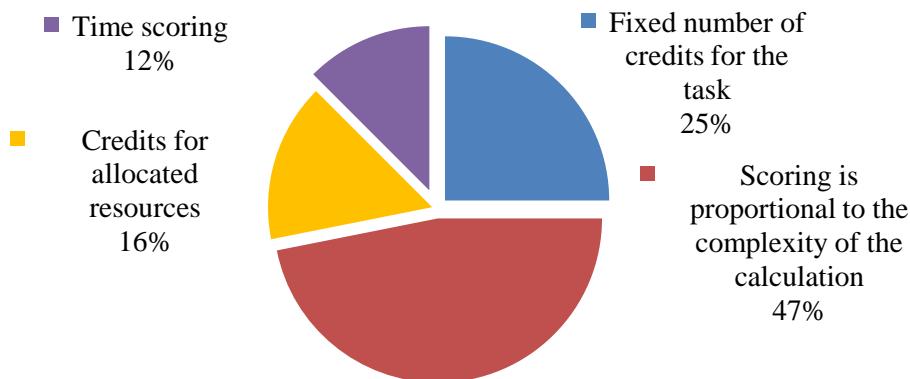


Figure 10. Scoring system types in VDC project

Let's present generalized recommendations of the volunteers to developers of the new MBC projects (figure 11). The two most important of them are an understandable brief description of the project (81%) and the publication of actual news about it (77%). The following three wishes are to constantly maintain the site (59%), follow the availability of new tasks (57%), answer questions of the fliers (53%). And the last thing is to create application versions for different platforms (47%).

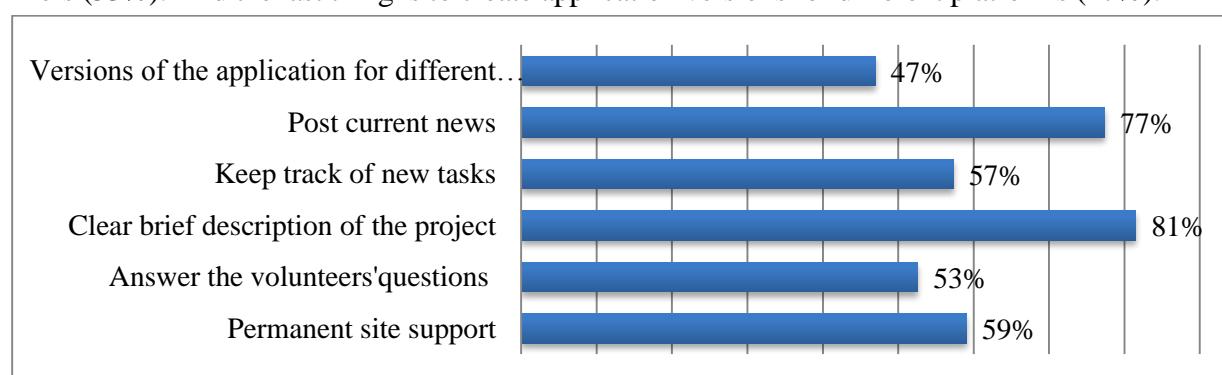


Figure 11. Recommendations to developers of the new VDC project

VDC project improvement

The following steps are necessary for increase in interest of crunchers in the VDC project:

- The publication of new information on the project website;
- Support of availability of a large number of the tasks (work units) ready to be sent;
- Support of back coupling with administration of the project;
- Availability of checkpoints in case of execution of work units on computational nodes.

List of tasks in case of deployment and providing of the project:

1. Technical expenses:
 - 1.1. Creation of the computing application with checkpoints;
 - 1.2. Creation of the generator of input data;
 - 1.3. Creation of the validator and aggregator of results;
 - 1.4. Server part deployment;
 - 1.5. Fine setup of parameters of a server part of BOINC project.
2. Organizational expenses:
 - 2.1. Registration of the domain;
 - 2.2. Modification of the standard website of the project;
 - 2.3. General description of scientific and administrative group of the project;
 - 2.4. The organization of competitions in the project.
3. Interaction with crunchers (attraction of new resources and holding of available):
 - 3.1. Popular scientific description of a scientific component of the project;
 - 3.2. The regular popular scientific description of the made numerical experiments;
 - 3.3. The regular publication on the website of the project of the received results;
 - 3.4. Interaction with volunteer community;
 - 3.5. Blogging of the project and publication of links to scientific papers.
4. Additional expenses:
 - 4.1. Development and deployment of scoring system;
 - 4.2. Development setup of badges system;
 - 4.3. Design of the website and information website of the project;
 - 4.4. Personalization of the received results when the volunteer can learn that he reads at present;
 - 4.5. Visualization of results, if it possible.

Interaction with the virtual community

The results of the survey allowed us to find out the motivation of volunteers and their attitude towards various aspects of VDC projects. Since volunteers are a vivid example of a virtual community, it is possible to apply this approach towards other virtual communities. As a first approximation, it is possible to use some of the results of this survey: the attitude to the national VDC project, the importance of personal results and the availability of virtual prizes, the need to introduce a scoring system and a list of expectations from project organizers. This information can be used to organize successful interaction with virtual communities in smart city and e-governance projects. In addition, the method of multiparameter evaluation of projects will be useful for these areas [15].

Conclusions

The study is devoted to an extremely important topic, in terms of the development of distributed computing - the attitude of the Internet community towards the problems of voluntary distributed computing. The presented results of processing, in fact, the first multivariate sociological survey conducted in Russia, made it possible to determine the preferences and wishes of the fliers for various VDC projects, which creates prerequisites for improving the efficiency of projects and outlines the prospects for implementing and supporting voluntary distributed computing for science.

Acknowledgement

This work was supported by the Russian Foundation for Basic Research (grant No. 18-57-06003).

References

1. Foster I., Kesselman C. (Eds.). The Grid 2: Blueprint for a new computing infrastructure. Elsevier, 2003, 748 p.
2. Foster I., Kesselman C. Globus: A metacomputing infrastructure toolkit // The International Journal of Supercomputer Applications and High Performance Computing, 1997, vol.11, no. 2, pp.115–128.
3. Thain D., Tannenbaum T., Livny M. Distributed computing in practice: the Condor experience // Concurrency and computation: practice and experience, 2005, vol.17, issue 2–4, pp. 323–356.
4. Grimshaw A.S., Wulf W.A. The Legion vision of a worldwide virtual computer. Communications of the ACM, 1997, vol.40, no.1, pp.39–45.
5. Anderson D. P. BOINC: A system for public-resource computing and storage / In Grid Computing Proceedings. Fifth IEEE/ACM International Workshop on 2004, November, pp.4–10.
6. The server of statistics of voluntary distributed computing projects on the BOINC platform. <http://boincstats.com>. (date of access: 30.08.2018).
7. Da Silva D.P., Cirne W., Brasileiro F.V. Trading cycles for information: Using replication to schedule bag-of-tasks applications on computational grids / In European Conference on Parallel Processing, 2003, pp.169–180.
8. Vatutin E. I., Titov V.S. Voluntary distributed computing for solving discrete combinatorial optimization problems using Gerasim@home project / In Distributed computing and grid-technologies in science and education: book of abstracts of the 6th international conference. Dubna: JINR, 2014, pp.60–61.
9. Posypkin M., Semenov A., & Zaikin O. Using BOINC desktop grid to solve large scale SAT problems // Computer Science, 2012, vol.13, no.1, p.25.
10. Zaikin O.S., Posypkin M.A., Semenov A.A., Khrapov N.P. Experience in organizing volunteer computing: a case study of the OPTIMA@home and SAT@home projects // Vestnik of Lobachevsky State University of Nizhni Novgorod, 2012, no.5–2, pp.340–347.
11. Chien A., Calder B., Elbert S., & Bhatia K. Entropia: architecture and performance of an enterprise desktop grid system // Journal of Parallel and Distributed Computing, 2003, vol. 63, no.5, pp.597–610.
12. Clary E.G., Snyder M., Ridge R.D., Copeland J., Stukas A.A., Haugen J., Miene P. Understanding and assessing the motivations of volunteers: a functional approach // Journal of personality and social psychology, 1998, vol.74, no.6, p.1516.
13. Webpage of World Community Grid project. 2013 Member Study: Findings and Next Steps www.worldcommunitygrid.org/about_us/viewNewsArticle.do?articleId=323 (date of access: 25.09.2018).
14. 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 Contemporary Society" (IMS-2015), St. Petersburg, June 23, 2015, St. Petersburg: ITMO University, 2015. Sc.345-352. ISBN 978-5-7577-0502-6. (in Russian)
15. Yakimets V.N., Kurochkin I.I. Multiparameter and Index Evaluation of Voluntary Distributed Computing Projects. In: Alexandrov D., Boukhanovsky A., Chugunov A., Kabanov Y., Koltsova O. (eds). Digital Transformation and Global Society (DTGS 2018) // Communications in Computer and Information Science, vol.858. Springer, Cham, 2018, pp.528–542.

UOT 004.75

Kuroçkin İlya I¹, Yakimets Vladimir N.^{1,2}

¹Rusiya Elmlər Akademiyasının İnformasiyanın Ötürülməsi Problemləri İnstitutu, Moskva, Rusiya

²Rusiya Xalq Təsərrüfatı və Dövlət Xidməti Akademiyası, Moskva, Rusiya

¹kurochkin@iitp.ru, ²iakimets@mail.ru

Könüllü paylanmış hesablamalar layihələrinə könüllülərin cəlb edilməsi

Könüllü paylanmış hesablamalar layihələrinə könüllülərin cəlb edilməsinin əsas metodlarına baxılır. Sosiooji tədqiqatların nəticələri əsasında üstün tutulan könüllülərin müzakirəsi aparılır. BOINC platforması əsasında könüllü paylanmış hesablamalar layihəsinin uğurlu yerləşdirilməsi və müşayiəti üçün tədbirlər kompleksi təklif edilir. Smart city, e-gov üçün baxılan metodların tətbiqi əsaslandırılır.

Açar sözlər: könüllü paylanmış hesablamalar, BOINC, könüllü paylanmış hesablamalar layihələri, könüllülərin motivasiyası, çoxparametrlı qiymətləndirilmə.

УДК 004.75

Курочкин Илья И.¹, Якимец Владимир Н.^{1,2}

¹Институт проблем передачи информации РАН, Москва, Россия

²Российская академия народного хозяйства и государственной службы, Москва, Россия

¹kurochkin@iitp.ru, ²iakimets@mail.ru

Привлечение добровольцев в проекты добровольных распределенных вычислений

Рассматриваются основные методы привлечения добровольцев в проекты добровольных распределенных вычислений. Приводится обсуждение предпочтений добровольцев на основе результатов социологических исследований. Предлагается комплекс мер для успешного разворачивания и сопровождения проекта добровольных распределенных вычислений на платформе BOINC. Обосновывается применимость рассматриваемых методов привлечения добровольцев для smart city, e-gov.

Ключевые слова: добровольные распределенные вычисления, BOINC, проекты добровольных распределенных вычислений, мотивация добровольцев, многопараметрическая оценка.