

УДК: 004.75

Российские участники добровольных распределенных вычислений на платформе BOINC. Статистика участия

В. И. Тищенко, А. Л. Прочко

Институт системного анализа РАН,
Россия, 117312, г. Москва, проспект 60-летия Октября, д. 9

E-mail: isa@isa.ru

Получено 6 октября 2014 г.

В статье рассмотрено сообщество российских участников добровольных распределенных вычислений (ДРВ), реализуемых на открытой программной платформе BOINC. Для проведения статистического анализа активности российских участников ДРВ использованы данные, полученные при работе с API BOINC, приложением BOINC, и сайтом boincstats.com. Скрипт для получения данных и создания соответствующей базы данных с этого сайта был написан на PHP, для хранения данных, использовались базы данных MySQL.

В базе данных были аккумулированы показатели по всем российским проектам, включая архивные, что позволило рассчитать показатели, характеризующие поведение российских участников во всех проектах и командах BOINC — абсолютное и относительное количество российских участников, активность участия, количество привнесенных очков в систему, количество участников в каждом из российских проектов, заинтересованность участников в концепции ДРВ.

Показано, что позиции России в рейтинге стран очень низки и сохраняются практически на одном уровне в течение 4 лет. По мнению авторов исследования, низкие показатели поведения российских участников ДРВ, обусловлены индивидуализмом и закрытостью российских Интернет-пользователей, а также малым интересом к развитию фундаментального научного знания, научному поиску, что, возможно, связано с низким авторитетом как науки в целом, так и гражданской науки, краудсорсинга, в частности, и, соответственно, недостаточном распространении идей использования механизма добровольных распределённых вычислений для реализации исследовательских проектов.

Ключевые слова: добровольные распределенные вычисления, платформа BOINC, виртуальные сообщества

Russian participants in BOINC-based volunteer computing projects. The activity statistics

V. I. Tishchenko, A. L. Prochko

Institute of Systems Analysis, Russian Academy of Sciences, 9, 60-letiya Oktyabrya pr., Moscow, 117312, Russia

The article analyses the activity statistics of the Russian participants of volunteer computing (VC) using platform BOINC obtained by the authors. The data has been received with API BOINC and site www.boincstats.com. The script for the database was written in PHP, for data storing was used MySQL.

The database indicators were accumulated across all Russian projects, which allowed the calculation of the indicators characterizing the behavior of the Russian participants in all projects and teams BOINC — absolute and relative number of Russian participants, their activity, the number of introduced points system, the number of participants in each of the Russian project participants, interest in the concept of the VC.

It is shown that the position of Russia in the countries ranking is very low and is retained at the same level for 4 years. According to the authors, low activity of the Russian participants of the VC, due to individualism and the closure of Russian Internet users, as well as to a small interest in the development of fundamental science, scientific research. This, possibly due to the low-prestige as a science as a whole, as well as civil science, crowdsourcing, in particular. And, therefore, we can see insufficient dissemination of the ideas of using the mechanism of VC for research projects.

Keywords: volunteer computing, BOINC platform, virtual computing

Введение

На сегодняшний день BOINC (англ. — Berkeley Open Infrastructure for Network Computing) — открытая программная платформа Университета Беркли распределенных вычислений для проектов в области математики, молекулярной биологии, медицины, астрофизики и климатологии. BOINC предоставляет возможность исследователям научных проблем, требующих огромные вычислительные мощности, интегрировать свободные вычислительные ресурсы персональных компьютеров пользователей Интернета. Проект добровольных вычислений состоит из трех основных частей: сервера, веб-сайта и прикладного программного обеспечения (ПО). Процессом вычислений управляет сервер, именно на нем создаются задания для обработки. Для того чтобы подключиться к проекту, пользователь (доброволец) должен скачать и установить на свой ПК стандартный «BOINC-клиент». Подключение к проекту происходит путем указания URL сайта проекта, после чего BOINC-клиент автоматически скачивает с сервера прикладное ПО (ориентируясь на тип ОС и процессора ПК пользователя) и задания для обработки. Затем весь процесс вычислений на стороне пользователя происходит автоматически. В BOINC-клиенте существует гибкая система настроек, позволяющая эффективно задействовать свободные ресурсы ПК и распределять их между BOINC-проектами [Anderson, 2014].

Организаторы проекта, использующие ресурсы компьютеров участников добровольных распределенных вычислений (ДРВ), учитывают ряд факторов, которые, по их мнению, способствуют его активному развитию [Зайкин и др., 2014]. Помимо технических задач это условия, мотивирующие пользователей предоставлять вычислительные ресурсы своих компьютеров на добровольной основе и, соответственно, к участию в проекте:

- ощущение причастности к важным научным исследованиям и, соответственно, получению значимых научных результатов;

- командный дух и атмосфера состязательности; участники ДРВ могут объединяться в команды по разным признакам (национальному, региональному и пр.); за выполненное задание пропорционально затраченным вычислительным ресурсам участникам проектов начисляются так называемые «кредиты». Количество «кредитов» является характеристикой, по которой команды и отдельные участники соревнуются между собой;

- информированность о командном и/или индивидуальном участии в проекте; при получении результатов обычно на сайте проекта выкладывается информация об участнике, на ПК которого был получен данный результат.

Все проекты при «запуске» проходят последовательно три статуса: «альфа», «бета» и «релиз». Статусы проектов, использующих платформу BOINC, отображаются на сайте Formula BOINC (<http://formula-boinc.org/>). На данный момент из 74 активных проектов 54 имеют статус «альфа», 8 — «бета» и 12 — «релиз». Статус «альфа» означает, что проект функционирует и находится на начальном этапе разработки.

На территории СНГ на данный момент активно действуют пять проектов добровольных вычислений на основе BOINC: четыре в России (Einstein@Home, OPTIMA@home и SAT@home, Gerasim@home) и один на Украине (SLinCA@home). Все эти проекты имеют статус «альфа».

Статистический анализ поведения российских участников добровольных распределенных вычислений

Для ознакомления со статистикой проектов, активностью индивидуальных участников ДРВ или командными очками используются сайты, на которых статистические показатели графически визуализируются при посредстве одного из приложения BOINC — API (англ. — application programming interface) [Программное обеспечение..., 2014].

С помощью API можно получить информацию о количестве суммарных очков участника ДРВ, его позиции в мировом рейтинге, его уникальном идентификаторе в системе BOINC,

а также статистику по каждому проекту (количество очков и команда, в составе которой он участвовал в этом проекте).

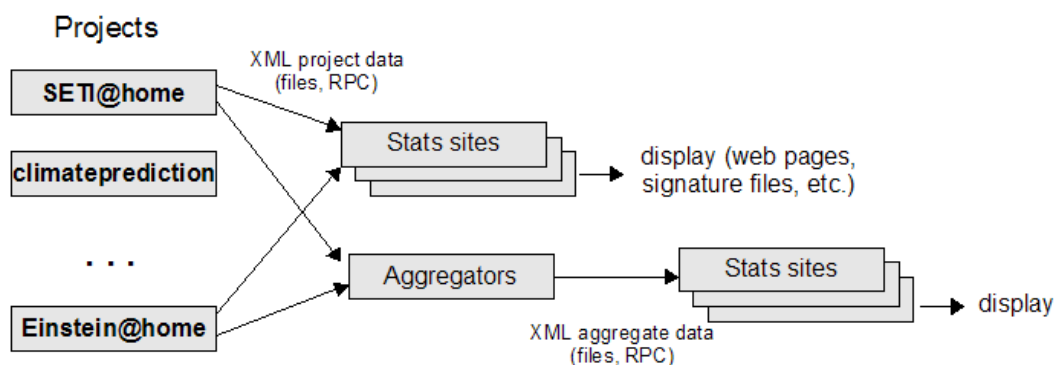


Рис. 1. Схема распространения информации через API системы BOINC

Для проведения статистического анализа активности российских участников ДРВ мы использовали данные, полученные при работе с API BOINC и с сайтом boincstats.com. Скрипт для получения данных и создания соответствующей базы данных с этого сайта был написан на PHP, для хранения данных использовались базы данных MySQL.

Таким образом, была организована база данных (см. рис. 2), содержащая данные по пользователям, которые указали в качестве своей «принадлежности» Россию. Таких пользователей мы рассматривали в качестве «российских пользователей». В базе мы аккумулировали показатели по всем проектам, включая архивные, в которых российские участники ДРВ принимали участие. Это позволило рассчитать показатели, характеризующие закономерности участия российских участников в проектах и командах BOINC.

Получив необходимые данные, можно сделать выводы о российских участниках в проектах на платформе BOINC. Всего на 1 июля 2014 года в системе BOINC находится 45 333 участника из России [Менеджер BOINC аккаунтов, 2014].

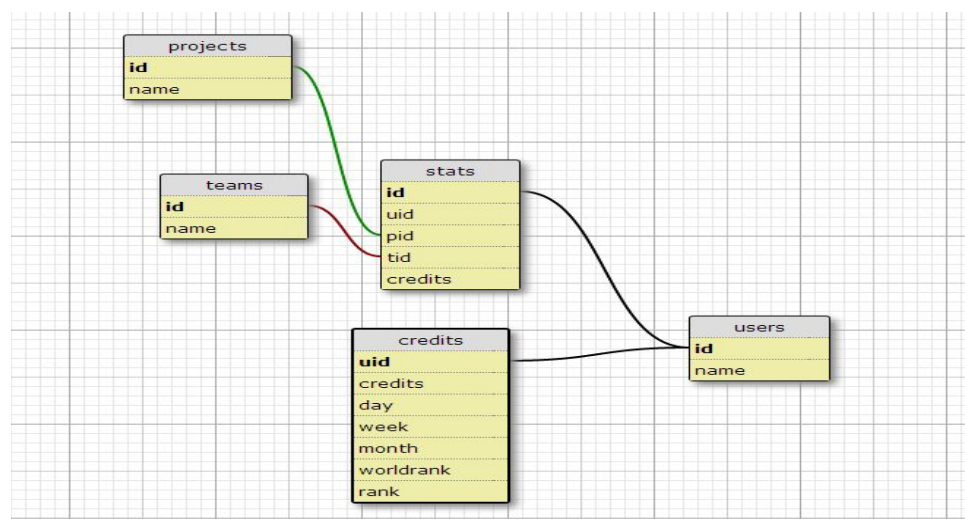


Рис. 2. Структура базы данных «Российские участники BOINC»

Рассмотрим полную статистику BOINC и выделим вклад и участие российского сегмента на 2014 год.

Как видно из таблицы, российские участники ДРВ составляют около 1.6 % от общего числа и приносят 2.5 % очков в систему. По очкам «за последний месяц» наблюдается похожая картина — 2.5 %. Если рассматривать позиции России в рейтинге стран, то она сохраняется уже

4 года. Россия (российские участники в совокупности) в рейтинге стран в настоящее время занимает 11 место по набранным очкам.

	Общая BOINC-статистика	Российский сегмент	%
Суммарные очки	2 117 324 186 166	53 691 650 954	2.5
Участники	2 816 118	45 333	1.6
Очков за последний месяц	56 540 311 021	1 434 955 070	2.5

На первый взгляд такое отставание от участников из США, Германии, Великобритании представляется естественным. Вместе с тем оно вызывает некоторое удивление, поскольку известно, что еще в сентябре 2011 года Россия вышла на первое место в Европе по числу уникальных пользователей, обогнав Германию [Тищенко, Жукова, Попков, 2014]. Еще более интересно, что Россия отстает по очкам в полтора раза от Польши, Франции, Австралии и Чехии. Представляется, что этот показатель, с одной стороны, свидетельствует об атмосфере закрытости и индивидуализма, царящей в среде российских интернет-пользователей, недостаточном распространении идей гражданской науки и краудсорсинга, а с другой, характеризует недостаточность распространности концепции распределенных вычислений и, соответственно, добровольных распределенных вычислений в России. Еще более низкий показатель у России — так называемый показатель количества очков «на душу населения». По этому показателю Россия занимает одно из последних мест в общем рейтинге из-за низкой активности участников ДРВ.

Страна	Население	Всего участвовало	Активно участвует
<i>США</i>	293 млн чел.	674 000 чел.	>75 000 чел.
<i>Германия</i>	82 млн чел.	293 000 чел.	>18 000 чел.
<i>Польша</i>	38 млн чел.	46 000 чел.	3100 чел.
<i>Чехия</i>	10 млн чел.	33 000 чел.	3400 чел.
<i>Япония</i>	127 млн чел.	61 000 чел.	>9 000 чел.
<i>Россия</i>	140 млн чел.	>45 000 чел.	4500 чел.

Естественно, что существуют проекты, в которых российские участники внесли более 40 % очков. Это непосредственно российские проекты SAT@home и Gerasim@home. Проект SAT@home был запущен 29 сентября 2011 года двумя российскими институтами — ИДСТУ СО РАН (Иркутск) и ИСА РАН (Москва). Проект Gerasim@home стартовал в 2008 году. Сейчас в нем принимают участие 450 пользователей из России.

В то же время если же проанализировать данные за «последний месяц», то мы обнаружим, что акцент участия и, соответственно, предоставления вычислительных мощностей сделан не на российские проекты.

Название проекта	Очки
Collatz Conjecture	268 110 637
MilkyWay@home	257 472 010
GPUGRID	157 612 925
PrimeGrid	90 584 571
Einstein@Home	8 825 750
SETI@Home	59 671 842
World Community Grid	56 244 972
Asteroids@home	29 961 120
Moo! Wrapper	28 914 024
Rosetta@Home	20 087 165
theSkyNet POGS	7 993 382
POEM@HOME	7 028 420
LHC@Home Classic	6 935 433
SAT@home	6 274 717
Gerasim@Home	6 189 205

Очки российских участников в различных проектах (май 2014 года)

Российские проекты SAT@home и Gerasim@home по количеству очков и, естественно, по объему участию российских пользователей находятся на 14 и 15 месте соответственно.

Для того чтобы оценить вклад каждого участника, один из авторов работы провел так называемое включенное исследование и подключил свой компьютер к платформе BOINC. Дата присоединения — 3 марта 2014 года, проекты — «SAT@home» и «World Community Grid». Почти за 3 месяца было набрано 50 000 очков в системе BOINC. За это время в рейтинге российских пользователей удалось подняться с нуля на 30 000 мест, текущая позиция (май) ~12 000. За последний месяц профиль поднялся примерно на 3 000 позиций.

Исходя из того, что данный результат достигнут на компьютере, который использовался только в дневное время, можно сделать предположение, что в системе есть 12 000 участников, которые «позволяли» работать BOINC у себя на компьютере около трех месяцев и более. Ниже приводится статистика по российским пользователям и их суммарным очкам. Правая колонка — какое количество пользователей попадает под данный критерий.

Количество очков	Пользователей
<100 000	36000
>100 000	9300
>1 000 000	2700
>5 000 000	880
>20 000 000	300
>100 000 000	70
>500 000 000	18

Распределение российских пользователей по показателю «Суммарные очки»

Следующая характеристика поведения российских участников, которую мы анализировали, — количество участников в каждом из проектов (включая уже закрытые). Всего на сайте BOINC заявлено 135 проектов. Примерно в половине проектов принимало участие больше 100 человек. Больше тысячи участников приняло участие только в 16 проектах.

Российских участников	Количество проектов
<50	60
50–200	31
200–1000	28
1000–5000	11
>5000	5

Количество российских участников в различных проектах

Исходя из того, что многие участники включены в одни и те же проекты, можно предположить, что коммуникации среди российских участников развиты достаточно хорошо. Существует российский портал www.boinc.ru, который помогает объединить участников различных команд. На форуме имеются разделы (форумы) 12 российских команд, и при появлении «интересного» соревнования команд подключение членов команд происходит достаточно организованно. Помимо форумов коммуникации между участниками ДРВ проходят в социальных сетях. Так, например, в группе в социальной сети «ВКонтакте» [Группа..., 2014] состоят 6 000 пользователей. С учетом того, что всего российских пользователей немногим более 45 000, а активных — 4 500, с помощью соответствующей информационной политики по распространению материалов в сетях можно получить значительный прирост активности российских участников ДРВ.

В этой связи было интересно проанализировать зависимость количества проектов, в которых «работает» российский участник платформы BOINC, и количество его суммарных очков как характеристики активности на портале. Мы предположили, что с увеличением активности пользователя (количества очков) увеличивается и количество проектов, в которых он принимает участие (его «заинтересованность» в концепции добровольных распределенных вычислений).

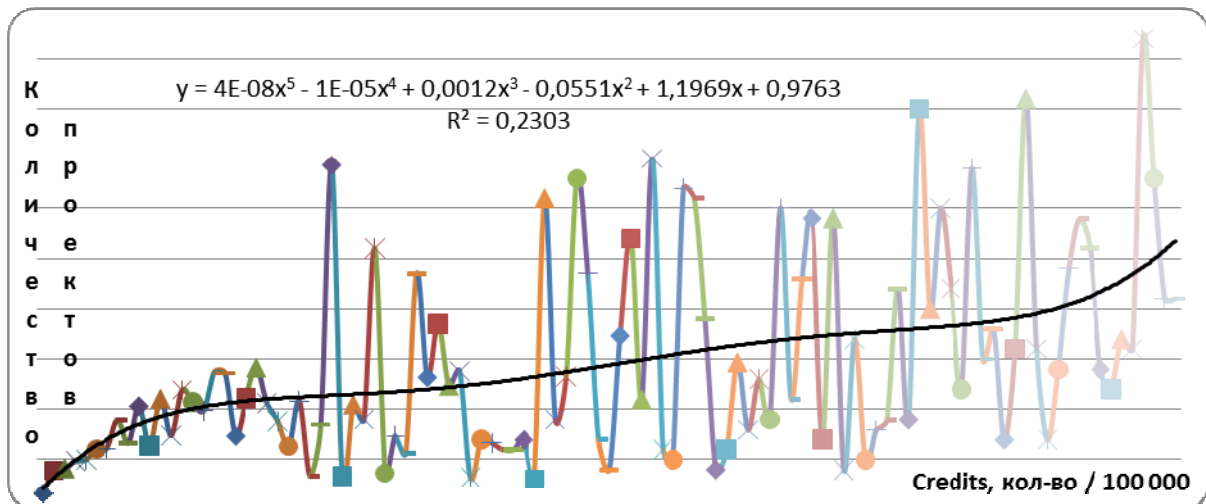


Рис. 3. Показатель среднего количества проектов для каждого диапазона очков с интервалом 100 000 (по оси абсцисс — количество проектов, по оси ординат — очки (кредиты))

Действительно, чем активнее участник (чем больше у него общее число очков), тем большее количество проектов он поддерживает, предоставляя мощности своего компьютера. В целом заметное увеличение наблюдается начиная с 500 000–1 000 000 очков.

Как видно из приведенных статистических данных, добровольные распределенные вычисления в нашей стране не очень популярны, несмотря на то что высокоскоростной Интернет становится все более доступным и распространенным. И дело не в технологических проблемах. Как отмечают в своем исследовании среди участников проекта SETI@home Oded Nov с соавторами [Nov, Arazy, Anderson, 2014], важнейшим из мотивов, влияющих на участие в системе добровольных распределенных вычислений, оказывается информация о научных публикациях результатов исследовательских проектов. Те участники добровольных распределенных вычислений, которые были ориентированы на развитие проекта как способа решения фундаментальной научной проблемы и видели результаты своего труда, демонстрировали лучшие показатели в предоставлении вычислительных мощностей своих компьютеров — активность, постоянство, время подключения и т. п.

Что касается российских пользователей Интернета, то, очевидно, их в меньшей степени интересуют развитие фундаментального научного знания, сама атмосфера научного поиска. Конечно, все это требует дальнейшего эмпирического исследования, подтверждения или опровержения этого предположения и высказанных ранее замечаний.

Список литературы

Группа «Программа BOINC и проекты распределенных вычислений». URL: <http://vk.com/boinc> (дата обращения: 08.08.2014).

Заикин О. С., Посыпкин М. А., Семенов А. А., Храпов Н. П. Опыт организации добровольных вычислений на примере проектов OPTIMA@home и SAT@home. URL: <http://omega.sp.susu.ac.ru/books/conference/PaVT2012/full/112.pdf> доступен 08.08.2014.

Менеджер BOINC аккаунтов. URL: <http://boincstats.com/ru/bam/>, <http://boincstats.com/ru> (дата обращения: 08.08.2014).

Программное обеспечение с открытым исходным кодом для организации добровольных распределенных вычислений и распределенных вычислений в сети. URL: <http://boinc.berkeley.edu/trac/wiki/GraphicsApi> (дата обращения: 09.08.2014).

Тищенко В. И., Жукова Т. И., Попков Ю. С. Сетевые взаимодействия. Предмет исследования и объект моделирования. — М.: ЛЕНАНД, 2014.

Anderson D. P. Public Computing: Reconnecting People to Science // Presented at the Conference on Shared Knowledge and the Web, Residencia de Estudiantes, Madrid, Spain, Nov. 17–19 2003. URL: <http://boinc.berkeley.edu/boinc2.pdf> (дата обращения: 08.08.2014).

Nov O., Arazy O., Anderson D. Scientists@Home: What Drives the Quantity and Quality of Online Citizen Science Participation? PLOS One, April 1 2014. URL: <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0090375> (дата обращения: 08.08.2014).