Volunteer Computing: Science@home

STS.034 Final Project (Draft 12/06/2015)

Yijun Jiang

It is not my first time writing something on a computer. In the past, when my fingers were busy typing, my $800 Dell desktop computer was in fact very idle: among its four processors, only 6% of total computational power was utilized, half by me and the other half by the operating system. However, things have changed. As I am writing this article, my computer is devoting a great deal of the remaining power to analyze radio signals from spinning neutron stars, hoping to find traces of gravitational waves, which are ripples in the curvature of spacetime predicted by Albert Einstein almost a century ago but have never been directly detected to this day.

No, I am not an astrophysicist and, just like many of you, I know little about things happening deep in the universe. Nor am I a computer scientist who is good at organizing computational resources. What I have done is simply participating in a volunteer computing project named *Einstein@home*, whose client software allows me to donate my idle CPU to frontier astrophysics research.

Science for Everyone

Ever since the time of Galileo when modern science originated, new discoveries reshaping our understanding about the world have been constantly made by the minority of elites with brilliant minds. As a result, in the view of most people, science belongs to Newtons, Darwins and Einsteins. Even in the recent years when research is often done by collaboration rather than individual work, frontier science still seems far away from ordinary people who are not associated with the academia.

However, time has come when more and more scientists realize that their own communities are becoming incapable of handling larger and larger research projects. One of the biggest challenges is the severe conflict between exploding demand of computational power and limited resources of modern workstations. The Large Hadron Collider (LHC), for example, produced on average 0.8 GB data per second in 2012, which could fill up 15,000 DVDs in a single day. Many other projects, from simulating climate change to calculating protein structures, are faced with the same calculational power bottleneck. The traditional solution by supercomputers, unfortunately, is often infeasible to research groups both monetarily and technologically.

On the other hand, however, PCs are so widespread that, as predicted in 2013 by Gartner, Inc., 2 billion of them would be in use by 2014. More importantly, modern computers are state of the art and continuously upgrading. Sony's PlayStation 4s, for example, is already one part per 20,000 as fast as the best supercomputer in the world, and by March 2015, over 20 million of them had been sold. The full computational power of PlayStation 4s alone is equivalent to over a thousand supercomputers! In the meantime, however, the majority of PCs are working well below their capabilities. Most daily jobs on computers only occupy CPUs at less than 10%, not to mention the full waste of resource when the machines are left idle.

Volunteer computing came into being in this context. In a volunteer computing project, a large computational task is split into tiny chunks, or work units, which can be handled by PCs. With the help of the Internet, scientists distribute these work units to people who volunteer to run them on their own computers. Volunteers do not need any professional knowledge on the science involved in the project. All they have to do is to install a client that automatically downloads new work units, performs calculation by idle CPU power, and sends the results back to the server maintained by specialists.

For scientists, volunteer computing provides them with a cheap solution to the computational power bottleneck. A volunteer computing project is relatively easy to maintain because it is volunteers rather than scientists who run and upgrade the computing units and pay for the fees that come along. Moreover, volunteer computing provides a good chance for scientists to demystify their work in front of the general public. The project *Climateprediction.net*, for example, serves to educate its participants on climate science and calls for their attention on climate change issues.

For volunteers, on the other hand, one of the greatest motivations is their interest in relevant science topics, some of which, including climate change and development of new drugs, are closely related to everyday life of ordinary people. This can be seen from the popularity of such projects in comparison with projects on more fundamental science. Personal recognition by contributing to real science is also an important reason why people volunteer. Also, many projects have credit systems that award volunteers after a certain amount of work. Volunteers or teams of volunteers can enhance their reputation in volunteer computing forums by earning more credits.

Rapid Development

The first volunteer computing project was launched by George Woltman in January 1996. The project, known as *Great Internet Mersenne Prime Search* (GIMPS), is aimed at hunting for what mathematicians call Mersenne primes – prime numbers with a specific form and are extremely rare. And at that time, only 34 of them are known to humans. Within merely one year of work by volunteers, one of them, Joel Armengaud, discovered the 35th Mersenne prime on his own computer. With 420,921 digits, this number broke the record of the largest prime ever known.

The success of GIMPS started the era of volunteer computing. In 1999, University of California, Berkeley launched the project *SETI@home*, which analyzes data from the world's largest single-aperture telescope at Arecibo Observatory to search for extraterrestrial intelligence. In the same year, Stanford University launched *Folding@home*, which studies how proteins are formed from gene sequences.

However, as David Anderson, leader of *SETI@home*, pointed out, each of the early projects developed its own program to distribute jobs and to run calculations, as well as its own website for volunteers to communicate and track progress. This requires advanced skills in computer science, which is not possessed by general scientists. In order to popularize volunteer computing, Anderson's group established the *Berkeley Open Infrastructure for Network Computing* (BOINC) project in 2002, which serves as a platform for scientists to create their own volunteer computing projects. Since then, a lot more projects have been released online, covering astrophysics, cryptography, molecular biology, environmental science and many other fields. Up till now, there have been more than 60 projects on the BOINC platform. Today, more than 450,000 active PCs are participating in BOINC-based projects, creating a computational power equivalent to a third of the world's fastest supercomputer.

Thanks to volunteer computing, scientists have made substantial progress on many computationally intensive fields over the past decade. In just two years searching for binary radio pulsars, the *Einstein@home* project was able to find two more of them, while previously only eight were known. GIMPS has discovered 14 new Mersenne primes, the largest of which having more than 17 million digits. Results by *Climateprediction.net*, *Rosetta@home* as well as many other projects have been published on major publications such as Nature. These results are available to volunteers on the projects' homepages.

Challenges and Future

A big problem in many volunteer computing projects is the low volunteer rate. Despite of the rapid development of volunteer computing, among every 10,000 PCs today, only 3 to 4 of them are actively participating in various projects. Obviously, the great computational potential owned by the public is far from being fully exploited. Consequently, the progress of relevant research is limited, which in return reduces people's further interest to participate. What is worse, the low participation rate obstructs the science education attempt of many projects. Since scientific knowledge of an individual can quickly spread to his or her circle, losing several volunteers may mean losing a whole community.

To break this vicious circle, current volunteer computing projects must be improved in two directions: attracting new volunteers and retaining existing ones.

According to a survey on BOINC website, among all those who refuse to volunteer, over 40% of people are concerned about security issues of these projects. It can happen that a hacker breaks into a volunteer computing project and takes control over the job distributing channel, for example, to hijack volunteers' CPU power for illegal use or to spread malware to volunteers' PCs. In fact, this scenario is already taken into account by experts, and serious efforts have been made to insure the security of volunteer computing systems. However, these efforts are rarely informed to volunteers and potential volunteers. In order to gain the trust of general public, volunteer computing projects must make sure that such information is readily available. A security statement or certificate on the downloading page, for example, can help alleviate people's concerns.

Another barrier for beginners to get started is the more or less unfriendly software. 36% of respondents who tried BOINC but never became a volunteer find the software too hard to use. Volunteer computing scientists may not believe this, for in order to let the general public get hands on the projects, the client interfaces have been already greatly simplified. However, it cannot be denied that compared with many commercial softwares, volunteer computing softwares usually lack effective tutorials for beginners. Considering the science background built into these softwares, such tutorials should only be more detailed. It is important that volunteers are given enough information so that they feel comfortable operating these softwares.

Retaining existing volunteers is just as important as recruiting new ones. A study of the project *Climateprediction.net* shows that roughly half of the participants are unable to finish even the first 1.4% of a work unit, which takes about 12 hours on a typical PC. Only one in four volunteers can stick on to the completion of their work units. In the long run, about 40% of participants who responded to the BOINC survey give up and never volunteer again.

The major feedbacks from these drop-outs are about the side effects to their computers, including slowing down the machine and causing computer to overheat. Although users can choose the CPU occupation rate in their software interfaces, for the general public who are usually limited in computer expertise, clear instructions are necessary on balancing between active volunteering and computer performance.

To maintain their interest in a project, volunteers should be given real time feedback. This includes not only a well-functioning credit system which encourages them to keep volunteering, but also an effective way to present them with the scientific achievements, so that they obtain more sense of participation. Currently many projects display their calculational processes graphically as screensavers. However, this is far from enough, for what volunteers need is the significance of their work. They are certainly more willing to know what disease may be tackled by a pharmaceutical project, rather than to watch molecules dancing randomly on their screens.

If these challenges are properly addressed, the future of volunteer computing will be unlimited. Exponentially increasing number of people will volunteer, not only on their PCs, but also on tablets and smart phones. The exploding computational power that comes along will support more and larger computationally intensive research programs. For individual research groups, fully utilizing volunteered resource saves expenditure and improves efficiency. For the general public, exposing to frontier research improves their scientific awareness. Finally, for the global society, volunteer computing provides promising solutions to many major challenges today, such as global warming, health issues and energy crisis.