Volunteer Computing: Science@home

STS.034 Final Project (First Draft)

Yijun Jiang

12/04/2015

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 its remaining power analyzing radio signals from spinning neutron stars, hoping to find traces of gravitational waves, ripples in the curvature of spacetime predicted by Albert Einstein almost a century ago but have never been directly detected.

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 allows me to donate my idle CPU power into frontier astrophysics research.

Science for Everyone

Since the time of Galileo Galilei 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 frontier research is done more by collaboration 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, even for general purposes, are state of the art and continuously upgrading. Sony's PlayStation 4s is already one part per 20,000 as fast as the topmost supercomputer, 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, most PCs are far from working at their full power. 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 unattended.

Volunteer computing came into being in this context. In a volunteer computing project, a large computational task is split into tiny chunks that 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 to have 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 and sends the results back to the server maintained by specialized scientists and computer engineers.

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 relatively high participation rates of such projects in comparison with projects researching 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 having 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, 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 brought about the era of volunteer computing. In 1999, University of California, Bekerley started the project *SETI@home*, which analyzes data from the Arecibo Observatory in Puerto Rico, where the largest single-aperture telescope in the world is located. This project searches for extraterrestrial intelligence (thus the name SETI). 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 volunteer computing projects developed its own program to distribute jobs to volunteers and run calculation on their computers, as well as websites for them 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 and a template for scientists to create their own volunteer computing projects. Since then, a lot more projects are 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 the thrift of volunteer computing, scientists have made substantial progress on many computation 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.

Risks and Challenges

One of the main challenges faced by many volunteer computing projects is the low volunteer rate. Today among every 10,000 PCs, only about 2.5 of them are actively participating in volunteer computing projects. The great computational potential owned by the public is far from being fully exploited. Consequently, the progress of relevant science is limited, which reduces people's further interest to participate in turn. The low participation rate also does harm to the science education attempt of many projects. Since public knowledge in science can pass around, losing one volunteer may mean losing his or her family and friends.

According to a survey on BOINC website, among all those who refuse to volunteer over 40% of people who