Public Computing: Reconnecting People to Science[*]

Dr. David P. Anderson[†]

March 21, 2004

Abstract

The majority of the world's computing power is no longer in supercomputer centers and institutional machine rooms. Instead, it is now distributed in the hundreds of millions of personal computers all over the world. In a few more years, other consumer devices like game consoles and television set-top boxes may comprise a large fraction of total computing power.

This change is critical to scientists whose research requires extreme computing power. Projects like SETI@home and Folding@home have attracted millions of participants who donate time on their home PCs to a scientific effort. Work is underway to create similar projects in many other areas, enabling scientific explorations that were previously infeasible.

The implications of this "public computing" paradigm are social as well as scientific. It provides a basis for global communities centered around common interests and goals. It creates incentives for the public to learn about current scientific research. Ultimately, it will give the public more direct control over the directions of science progress.

1   Introduction

Computer technology has revolutionized science. Scientists have developed accurate mathematical models of the physical universe, and computers programmed with these models can approximate reality at many levels of scale: an atomic nucleus, a protein molecule, the Earth's biosphere, or the entire universe. Using these programs, we can predict the future, validate or disprove theories, and operate "virtual laboratories" that investigate chemical reactions without test tubes.

In general, greater computing power allows a closer approximation of reality. This has spurred the development of computers that are as fast as possible. One way to speed up a computation is to "parallelize" it - to divide it into pieces that can be worked on by separate processors at the same time. Most modern supercomputers work this way, using many processors in one box.

The economic forces that shape technology favor large scale. A company can spend more to develop a CPU chip if it's going to sell a million of them. So the chips used in home computers (like the Intel Pentium and the Motorola PowerPC) have developed quickly; in fact, they have doubled in speed about every 18 months, a trend known as "Moore's Law".

In the 1990s two important things happened. First, because of Moore's Law, PCs became very fast - as fast as supercomputers only a few years older. Second, the Internet expanded to the consumer market. Suddenly there were millions of fast computers, connected by a network. The idea of using these computers as a parallel supercomputer occurred to many people independently. Two projects of this type emerged in 1997: GIMPS, which searched for large prime numbers, and Distributed.net, which deciphers encrypted messages. These project attracted thousands of participants.

In 1999, a third project, SETI@home, was launched, with the goal of detecting radio signals emitted by intelligent civilizations outside Earth [1]. SETI@home acts as a "screensaver", running only when the PC is idle, and providing a graphical view of the work being done. SETI@home's appeal extended beyond hobbyists; it attracted millions of participants from all around the world. It inspired a number of other academic projects, as well as several companies that sought to commercialize the public computing paradigm.

2   The power of public computing

Public computing can provide more computing power than any supercomputer, cluster, or grid, and the disparity will grow over time. SETI@home currently runs on about 1 million computers. This provides a processing rate of 60 TeraFLOPS (trillion floating-point operations per second). In contrast, the largest conventional supercomputer, the IBM ASCI White, provides about 12 TeraFLOPs. SETI@home's 1 million computers represents a tiny fraction of the approximately 150 million Internet-connected PCs worldwide. The latter number is projected to grow to 1 billion by 2015. Thus public computing has the potential to provide many PetaFLOPs of

computing power.

Moore's Law asserts that the speed of CPU chips doubles about every 18 months. The rate of progress is even faster for "graphics coprocessors", the chips that handle 3D graphics in PCs and game consoles. Their doubling time is about 8 months, and current graphics chips have a raw floating-point arithmetic speed many times that of their host CPU. These graphics chips are becoming more programmable and flexible, and researchers are actively investigating their use for scientific computing. Because graphics chips are integrated in modern personal computers, this trend favors public computing over other paradigms.

Most computational tasks require storage (disk space) as well as computing. Here also, public resources can provide unprecedented capacity. Today, a typical PC provides about 80 Gigabytes of storage space, which in most cases is more than is used the PC owner. If 100 million computer users were each to provide 10 Gigabytes of storage, the total would be an Exabyte (10 to the 18th power) - greater than the capacity of any centralized storage system.

3   Social aspects of public computing

Public computing is effective only if many people participate. SETI@home has been very successful in this regard; we have attracted 4.6 million participants, of which about 600,000 remain active.

People learned about SETI@home through several mechanisms. The mass media have covered SETI@home, as have Internet news forums like Slashdot [2]. SETI@home's screensaver graphics are a powerful promotional mechanism: in offices and school, where computers are seen by many people, a computer running SETI@home is a highly visible advertisement.

Who participates in SETI@home, and why? To study this question, we conducted an online poll to which about 130,000 participants have responded. Our web site allows users to create online "profiles" describing themselves; about 50,000 have done so. We created online message boards with many thousands of participants, and we have the anecdotal information of email communication with thousands of users.

Our poll indicates that 92% of SETI@home users are male, and that most of them are motivated primarily by their interest in the underlying science: they want to know if intelligent life exists outside earth. Another major motivational factor is public acknowledgement. SETI@home keeps track of the contribution of each user (i.e. the amount of computation performed)

and provides numerous web-site "leader boards" where users are listed in order of their contribution. Users can also form "teams", which have their own leader boards. The team mechanism turned out to be very effective for recruiting new participants.

Some SETI@home participants attempt to "cheat" - to get credit for computation not actually performed. Even more problematic are users who intentionally return incorrect results, essentially vandalizing the computation. These problems can be addressed by doing computation redundantly, and comparing the results.

SETI@home participants have contributed more than CPU time. Volunteers have translated the SETI@home web site into 30 languages, and have developed many kinds of add-on software and ancillary web sites. We believe that it is important to provide channels for this sort of contribution.

Various "communities" have formed around SETI@home. There is a single worldwide community, which interacts through the SETI@home web site. There are also national or language-specific communities, with their own web sites and message boards. The SETI@home user group in Germany has had conventions for several years. At least three couples have met and married through SETI@home communities.

4   Technical aspects of public computing

Conducting a public computing project requires adapting an application program to various platforms, implementing server systems and databases, keeping track of user accounts and credit, dealing with redundancy and error conditions, and others tasks too numerous to list here.

We are currently developing software called Berkeley Open Infrastructure for Network Computing (BOINC) that solves or helps solve most of these problems. BOINC makes it fairly easy and cheap to convert an existing application to a public computing project. BOINC projects are autonomous; each one maintains its own servers and databases, and does not depend on others. Participants can register with multiple projects, and can control how their resources are shared (for example, a user might devote 60% of his CPU time to studying global warming, and 40% to SETI).

Several BOINC-based projects are in progress, including SETI@home, a biochemistry project called Folding@home [4], and a climate study project called Climateprediction.net [3]. BOINC is a complement to Grid systems that support resource sharing within and among institutions, but do not support public computing [5].

5   Applications of public computing

To be amenable to public computing, a task must be divisible into independent pieces whose ratio of computation to data is high (otherwise the cost of Internet data transfer may exceed the cost of doing the computation centrally). Many types of computations have these properties:

1. Complex physical systems have a random and chaotic component. Their outcome is probabilistic, not exact. Studying the statistics of this outcome requires running large numbers of simulations with different random initial and boundary conditions. These simulations can be run in parallel.

2. There is an evolving field of "random algorithms" [ref] that provide approximate solutions to exact problems. These often involve random trials that can run in parallel.

3. "Genetic algorithms" are applicable to many areas. This approach involves creating a population of approximate solutions to a problem, and using the mechanisms of natural selection to approach an optimal solution.

4. Models of physical systems often have large numbers of underlying parameters whose optimal values are not known, and which combine nonlinearly. Exploring such parameter spaces requires large numbers of independent simulation runs. More generally, "Monte Carlo" algorithms involve large numbers of independent computations, corresponding to sampling in a high-dimensional space.

5. Applications that involve analyzing large amounts of data, such as data from a radio telescope (e.g., SETI@home) or from a particle accelerator, have inherent parallelism. The limiting factor is the computation-to-data ratio.

6. Some medical projects involve searching a set of millions or billions of molecules (for example, searching for potential drugs). These tasks are easily parallelized. Similarly some genetics projects involve matching a set of proteins with a DNA sequence; again, this is easily parallelized.

6   Conclusion

Carl Sagan observed that the general public's attitude toward science is increasingly one of alienation and even hostility [7]. Public computing may

help to reverse this trend. If computer owners can donate their resources to any of a wide range of projects, they will study and evaluate these projects, learning about their goals, methods, and chances of success. This process might be further encouraged by the creation of "decision markets" in which the public can make virtual bets or investments based on the outcome of science projects, analogously to political decision markets [8].

Because computer owners can contribute to whatever project they choose, the control over resource allocation for science will be shifted away from government funding agencies (with the myriad factors that control their policies) and towards the public. This has its risks: the public may be easier to deceive than a peer-review panel. But it offers a very direct and democratic mechanism for deciding research policy.

If a scientist has an idea for a computation, but finds that it will take a million years of computer time, the normal reaction is to toss the idea in a wastebasket. But public computing makes such ideas feasible: SETI@home has used 1.5 million years of CPU time. Scientists can now resurrect and reconsider these discarded ideas.

REFERENCES

1. D. P. Anderson, J. Cobb, E. Korpela, M. Lebofsky, and D. Werthimer. SETI@home: An experiment in public-resource computing. Communications of the ACM, Nov. 2002, Vol. 45 No. 11, pp. 56-61. See also http://setiathome.berkeley.edu

2. http://www.slashdot.org

3. http://climateprediction.net

4. http://folding.stanford.edu

5. http://www.globalgridforum.com/

6. R. Motwani and P. Raghavan. Randomized Algorithms. Cambridge University Press, 1995.

7. C. Sagan. The Demon-Haunted World: Science As a Candle in the Dark. Random House, 1996.

8. R. Forsythe, T. A. Rietz, and T. W. Ross. Wishes, expectations, and actions: A survey on price formation in election stock markets. Journal of Economic Behavior and Organization, 39:83–110, 1999.