

Ben Delaney
*CyberEdge
Information
Systems*

The Power of P2P

For years, Sun Microsystems has beaten on the same drum. Since 1982, their corporate motto has been “the network is the computer.” However prescient this statement seems now, until a few years ago when the Internet, and especially the Web, really took off, it seemed like a hopeless abstraction to most people. Today, however, it seems that Scott McNealy was right, and the proof has come from the most unlikely places.

Computers had only existed for a few years when people realized that linking them would multiply their power. Around the same time, it became obvious that we weren’t using some computers all the time. Because they were expensive devices, both to obtain and maintain, the industry saw these unused CPU cycles as an important waste of resources. The first time-sharing computers were built in the early 1960s to address the wasted cycles problem and to make it easier to use computers without actually being in the room. Systems from Digital Equipment and IBM were hacked together to make it possible for several people to share one expensive computer.

A descendent of these early networking attempts, peer-to-peer (P2P) networking via the Internet, is poised to become a killer app. First demonstrated as a way to share (some people say to steal) music, innovators are also using this technology in more serious ways. Let’s take a look at how P2P technology is changing the way people work and play.

New networking paradigms emerge

In the mid-1990s, AppleTalk and Microsoft’s problematic Windows for Workgroups made P2P networks possible. Neither was a perfect solution, with AppleTalk (introduced in 1984) running on a serial protocol that limited speed and Windows for Workgroups being one of the buggiest operating systems in memory. But both led the way to inexpensive networking that was ideal for small and home offices. For the first time, small offices could have some of the advantages of a LAN, such

as sharing printers and files, without the expense and overhead of a network server. Not only was the capital cost reduced considerably, but in theory at least, network management was simple and undemanding—with AppleTalk that was true.

Another new concept in networking, the Web, was a way to use the Internet to share hypertext documents among computers. It was developed in 1992 as a way for scientists at the European nuclear research lab, CERN, to collaborate. The developer of the protocols, Tim Berners-Lee, and the student who developed the first browser (in 1994), Marc Andreessen, soon became folk heroes. The Hypertext Transfer Protocol (HTTP) and the Mosaic browser, made possible an old dream of Ted Nelson’s—universally viewable and shareable hypertext documents—that he had labored on since 1960 under the name Xanadu.

One of the Web’s unanticipated benefits was getting people to think about how to use networks of hundreds of thousands or even millions of computers. As Nelson realized, such a milieu could provide a source of nearly infinite knowledge. Isn’t it likely that among the millions who cruise the Web every day we can find the answer to virtually any question? Furthering that idea, we can assume that virtually any file or bit of information that exists is on a computer hard disk on a machine connected to the Internet. Going even a step further, how many CPU cycles are wasted every day between key strokes and while computers sit idle, waiting for something to do? When we stop to consider these possibilities, the mind boggles.

It was thinking like this that helped David Gedye and Craig Kasnoff develop a way to use those wasted CPU cycles. Interested in the search for extraterrestrial intelligence (SETI), they conceived SETI@home in 1996 (see Figure 1). They raised money and found computers, and on 13 May 1999, they officially launched SETI@home from a lab on the University of California, Berkeley, campus. The system uses wasted CPU cycles by providing a lightweight client that runs in the background or as a

screen saver on PCs, Macs, and various Unix machines. The program analyzes data gathered by the giant radio telescope at Arecibo, Puerto Rico, and parses and distributes it in 500 Kbytes packages to client systems. Each client performs the same set of complicated analytical processes on the data and then returns the results to SETI@home's server, where they are verified and consolidated. As of 16 December 2000, SETI@home had exceeded a half-million years of CPU time. SETI@home had 2,662,918 participants on 7 January 2001.

There is another good way to use millions of computers and their hard disks. Napster, which is probably the most famous P2P application because of its numerous legal battles, distributes files instead of processes. Intended as a way for users to share music, Napster quickly became not just the music lover's clearinghouse but a lawsuit target for musicians, record companies, and the Recording Industry Association of America. One reason that Napster has been the focus of so much legal action is its server-based design with a central database of titles and the computers they are located on. This is an efficient way to centralize searching, but it means that the server's owner, theoretically at least, knows what's going on in the system.

Gnutella was the first of several Napster-like systems that take the central server out of the system. Oddly enough, Gnutella no longer exists. It was developed by the same people who brought WinAmp to the world: Justin Frankel and Tom Pepper at Nullsoft in March 2000. When AOL bought Nullsoft, they nixed Gnutella because there was too much legal and business risk. Several free agents have since taken up the flag, producing Gnutella clones that function in the same way and can operate together. The Gnutella client is called a Servent—both a server and a client in one package (see Figure 2). It connects directly to other users, starting with a few IP addresses supplied by the software's developers or found on many Web sites. The Gnutella system is generally impervious to monitoring or interference because there's no way to know which computers are participating and content is distributed throughout the Internet.

The next big thing

Searching for little green men and sharing music and other files is pretty good, but more critical projects are also using P2P networking's power. I first saw one at the Ars Electronica Conference, held September 2000 in Linz, Austria. The Distributed Annotation System (DAS), developed by Lincoln Stein of the Cold Spring Harbor Labo-

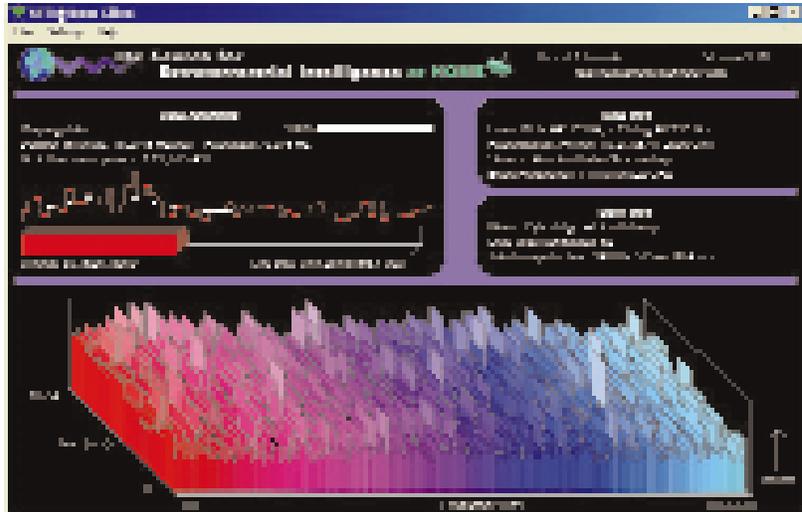


Figure 1. SETI@home has amassed more than 500,000 years of CPU time in its search for intelligence in the universe.

ratory in Cold Spring Harbor, New York, is a P2P system that shares notes regarding DNA code sequencing, or annotations. DAS is designed to solve a problem caused by the success of modern sequencing techniques—the volume of data is overwhelming researchers.

As Stein explains, “there is a lot of duplication of effort because the genome is a very, very big space.” Researchers are often working on the same regions, often using complimentary approaches. The difficulty is in correlating what multiple groups are doing, because the literature is not up to date. It takes months for papers to appear in print, and journals aren't set up to publish this large-scale, minutely detailed research.

Researchers are applying computerized sequencing technology to thousands of genomes, from viruses to human beings, which creates problems. Annotations describe particular landmarks in the

Figure 2. TodeNode is one of many Gnutella systems servents—clients-server applications that let users share any type of file.

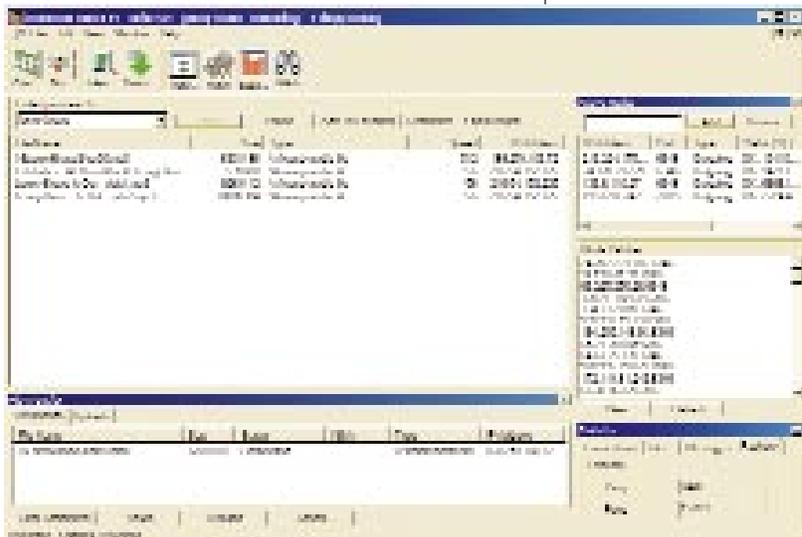




Figure 3. This view of the DAS client, Geodesic, shows two *C. elegans* sequences with differing information.

Figure 4. Researchers use the DAS client, Geodesic, to select sequences to compare from the *C. elegans* genome.

sequence of bases that make up a particular DNA molecule. In doing so, many scientists work on the same DNA, coming up with different landmarks, or start and stop points for sequences. Labs keep these sequencing data in databases. Because different labs have sequenced and annotated different parts of genes, there might be many different descriptions of the same sequences. Also, the databases aren't all compatible. Just as astronomers had to agree on what to call the objects they discovered and how to describe their type and location, DNA researchers are now working toward a common nomenclature.

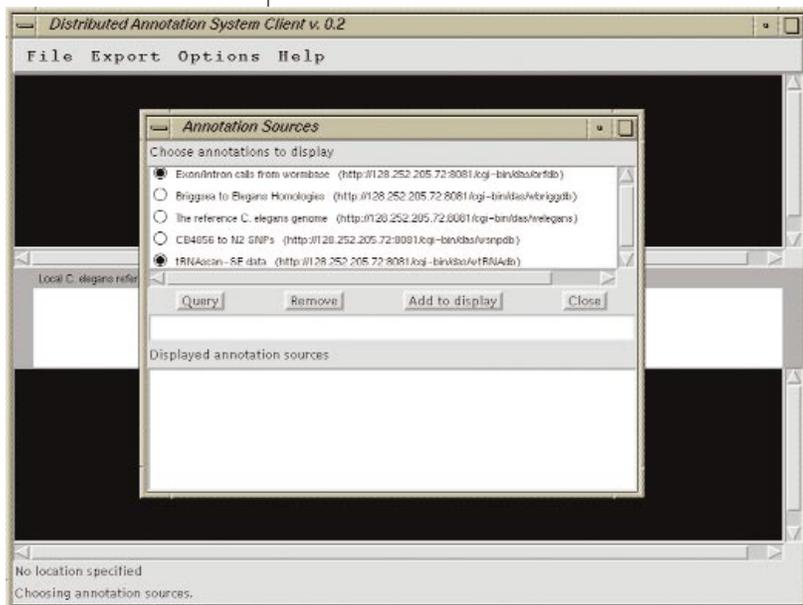
Stein's solution, the DAS, is a client application that can query a reference server that has a data-

base of an entire genome, using standardized Extensible Markup Language (XML). The DAS also knows which laboratories have annotations for which parts of the genome and offers searchers those records as they need them. The client sequence browser integrates the annotations and displays them in graphical or tabular form (see Figure 3). Because the system relies on XML and HTTP protocols, it works across the Web like a browser (see Figure 4). Stein and his colleagues have demonstrated the prototype system on the Web (<http://stein.cshl.org/das>) with 100,000 nucleotides from the human genome, 50,000 human genes, 19,000 genes from the worm *C. elegans*, and 40,000 mutations of maize. Scalability is a big concern, but in testing so far, DAS is doing well.

Business users are going to feel the impact of P2P soon, too. Ray Ozzie, the creator of Lotus Notes software, has a new company called Groove Networks (founded in October 1997) that's intended to facilitate collaboration "at the edge of the Internet." The system includes instant messaging, live voice, file sharing, pictures, threaded discussion, free-form drawing, outlining, and video functions in the client. Groove Networks designed it to be easily extensible and to enable enterprise-level communications or one-to-one file sharing. With \$60 million in startup funding and Ozzie's experience with Notes, Groove Networks looks like serious business.

One other aspect of P2P has perked up the ears of investors around the world. Several companies have devised schemes to broker unused CPU cycles on the open market. Entropia, based in San Diego, California, offers its clients "transparent dynamic scalability from one to thousands of processors, including real-time resource type and location reconfiguration, processor fault tolerance, and complete security for network traffic and data." They do this by providing a "free" client that uses idle time on members' computers, much like SETI@home.

However, while Entropia presents a philanthropic face to its "members," it's clients hear a different story. Entropia is a for-profit company. It donates time to research, such as the current evaluation of AIDS drugs, but makes no promises about how much time such pro bono projects will get as paying customers start using the service. Only time will tell. It should be a good business model if they can sign up enough members, because there's no compensation promised other than, "By joining Entropia, you and your computer will play an important role in the unfolding of the next history-making phase of the Internet."



This argument might not be as compelling when people realize that companies such as Popular Power, of San Francisco, California, are paying people for those excess cycles. Popular Power offers the same philanthropic argument and is starting business, as is Entropia, with nonprofit research. But Popular Power promises that participants in contract jobs will get an as-yet undisclosed payment for the work their computer does. It seems likely that this model will become more popular and could become one of the major industries of the 21st century.

Join the club

New business opportunities spawn advocacy groups, Web sites, and conferences. (Can a magazine be far behind?) P2P is no exception, and at least two advocacy groups, numerous Web sites, and two conferences already exist.

The Grid Forum (<http://www.gridforum.org>), founded last year, encourages collaboration and resource sharing among P2P system developers and users. Its mandate includes, “enabling the coordinated use of geographically distributed resources in the absence of central control, omniscience, strong trust relationships” and “enabling communities (virtual organizations) to share resources as they pursue common goals.” It also hopes to provide infrastructure for P2P, including a certificate authority, security policy recommendations, and protocol standards. It sponsored a conference, the Grid Forum Global Grid Forum 1 (GGF1), on 4–7 March 2001 in Amsterdam.

Another group interested in promoting and standardizing P2P is the Peer to Peer Working Group (<http://www.peer-to-peerwg.org>), which is “a consortium for the advancement of infrastructure standards for peer-to-peer computing.” Founding members include Alliance Consulting, Bright Station PLC, Entropia, Intel, J.D. Edwards, and Science Communications among others.

The O’Reilly Network, an off-shoot of the technical publishing house, has started the Peer-To-Peer DevCenter (<http://www.oreillynet.com/p2p>) with information for P2P application developers and a P2P directory with links to developers and tools. O’Reilly is also organizing a new conference to explore the technical, legal, and business dimensions of peer-to-peer computing. The first O’Reilly Peer-to-Peer Conference was 14 February 2001, and at least one announcement made there produced waves in the P2P community. Bill Joy, chief scientist at Sun Microsystems, gave the opening keynote speech, announcing a new Sun initia-

tive intended to add impetus to the P2P community. Joy announced JXTA (pronounced juxta), an open-source framework for P2P interoperability, that could rival Sun’s Java in importance.

Although Joy didn’t provide many details, he said that JXTA will probably include at least four mechanisms:

- piping from peer to peer,
- creating groups of peers and groups of groups,
- providing monitoring and metering services, and
- creating a security layer.

He also said that Sun hopes to release the first version of JXTA in April and will sponsor a JXTA developers conference this spring. (Contact Sun at jxta@sun.com for more information.)

The arguments for P2P connectivity and distributed computing are so compelling it’s hard to imagine this technology not taking off. Ted Nelson and Scott McNealy were right—computer power isn’t measured only by CPU speed and the amount of memory but in the number of connections. Distributed resources (be they processors, memory, hard disks, or sensors) are connecting via the Internet to create something entirely new—a global computer of unlimited processing power. Where this will lead is anyone’s guess. If you are looking for ways to make money, put your unused CPU cycles to work, find obscure music or information, collaborate with coworkers or friends, or contribute to worthwhile research efforts, P2P has benefits for you. P2P is catching on fast, and even as I wrote this, there were many systems that I didn’t have room to include. The field is changing fast—in Internet time—and by the time you read it will have evolved further. Despite this rapid rate of change, there is one certainty: P2P is here to stay and will have a dramatic impact on how we use our computers in the future. **MM**

Readers may contact Delaney at CyberEdge Information Systems, email ben@cyberedge.com.

Readers may contact Multimedia at Work editors Tiziana Catarci at the Dept. Information Systems, Univ. of Rome “La Sapienza,” Via Salara 113, 00198 Rome, Italy, email catarci@dis.uniroma1.it, and Thomas Little at the Multimedia Communications Lab, Dept. of Electrical Eng., Boston Univ. 8 Saint Mary’s St. Boston, MA 02215, email tdcl@computer.org.