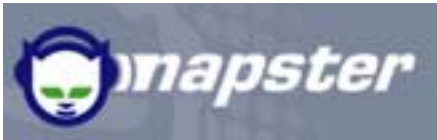


Chapter Seven: Journey to the Edge: Peer-to-Peer Computing and Content Control

Peer-to-peer computing, or P2P as it is often abbreviated, has a poster child whose impish grin seems to mock the biggest companies in the music industry.¹ Take this simple test. (A) Does Napster's logo remind you of Santa Claus, who delivers gigabytes of ripped MP3s for free? (Hint: College students look at the logo and see Santa Claus as thousands of them download tunes using the broadband connections on their campuses). (B) Does Napster's logo represent Satan, which is 'Santa' with a simple transposition of two letters?



Executives at Bertelsmann, Warner, Sony, EMI and other music companies see (B) Satan in the Napster logo. To those whose livelihoods depend on royalties, Nap-

ster with or without a logo is Satan, Beelzebub, and Belial. Shawn Fanning is a 21st century Lucifer brought to life to terrorise the recording industry. To maintain perspective, Mr Fanning was a youth of 18 when he coded Napster.

Mr Fanning did what any bright, music-loving programmer wanting to access music via the Internet could do. He cracked open his O'Reilly's Camel book, and made use of standard Internet services and snippets of code from various libraries.² He added a tight focus on MP3 files, and he attracted as many users as America Online has in one-tenth the time and for no cost.

1. Rebel, Rebel (David Bowie)

Grassroots MP3 ripping and indexing have traction. Someone with information to share converts a music track on an audio CD to an MP3 file.³ This person then clicks to a Napster server and uploads the name of the recording artist or group, the name of the song, the class or category of music to which the

-
1. For those over 30 reading this article, each of the section heads is the title of a popular song once available for download from a peer-to-peer MP3 file sharing service.
 2. The Camel book is a shorthand reference to the perl programming manuals published by Tim O'Reilly's firm in Sebastopol, California. Mr O'Reilly sponsors one of the largest peer-to-peer computing technical conferences. More information is available at <http://www.woreilly.com>
 3. The process of converting a CD audio file to the MP3 format is called ripping.

song or piece belongs, and the speed of the Internet connection for the computer on which the digital file resides. This PC connected to the Internet via the university's network then waits patiently to transfer the music to another soul somewhere in the world the Internet reaches.

The person looking for a particular song logs on to a Napster server. He or she taps in the name of the song or the artist whose music is wanted. The Napster directory and database server displays the computers with MP3 files of the desired song. A stoplight icon reports whether the server is online, busy or offline. If a machine is online, the 'customer' clicks on the title and copies the MP3 file to the computer's hard drive. The MP3 format is more space efficient than the music companies' CD audio format. A single CD-R disc can hold the equivalent of ten albums of music, or about ten hours of tunes. Unlike the offerings of Big Music (EMI, Bertelsmann et al), Napster offered different versions of tunes. A live concert version of a song might be available, even a bootleg of a demonstration version of the song. Fans of particular artists and types of music continue to see Mr Fanning as Santa Claus. Every day was Christmas when Napster delivered music without cost via the Internet. Most said, "Thank you, Santa Fanning". Others took a different view.

Napster was the first free, downloadable program that transformed the personal computers of somewhere in the range of 25 to 30 million people into servers able to exchange MP3 music files over the Internet. Napster turned the Internet into a global jukebox. The actual number of users remains hard to pin down. There is little disputing Big Music's perception that Napster users were unwilling to pay \$15.95 for a disc with the unforgettable Lawyers in Love by Jackson Brown when a free copy was a mouse click away. Napster marked the first time a teen's technology threatened a global, multi-billion dollar business. Most organisations with digital content for sale paid some attention to Napster.

Meanwhile Napster attracted headlines worldwide and described itself as "the largest, most diverse online community of music lovers in history". Big music cried "Foul" and usually added a handful of other choice, but unprintable epithets. It was not until Dr Dre and Metallica waxed eloquent about their lost royalties that those in the know knew that Napster was going to be stopped. Real money was at risk and some serious players were ready for action.

The American legal system jolted awake when Metallica and Dr Dre filed a stack of copyright case law. Think what you will about the stage presence of Metallica and Dr Dre; these people drew media attention. In March 2001, Napster executives glumly began the onerous task of blocking and removing pointers to 135,000 songs that violated copyright. Big Music declared the Napster party over. Dr Dre intoned, "Yo". Metallica just glared.

Napster lost its battle. However, the war for unfettered digital music continues. Napster showed millions worldwide how to share their favourites without fooling around with trips to Tower Records, trading cash for tunes. In retrospect, for those over 55, the furore over Dr Dre's mellifluous 'Nuthin' but a 'G' Thang' is amusing, even quaint. Copyright-sensitive professionals were not amused. Quaint was not a word used when talking to Big Music, Metallica, and Dr Dre about their money.

Napster will spawn Son of Napster. Remember how in those 1950 Japanese B movies *Mothra*, *Rodan* and *Godzilla* kept coming back. Each return brought a tougher, nastier beast. Big music, the motion picture industry, professional publishers, in fact any content creator, knows that Son of Napster will rise from the piles of legal documents that buried the original peer-to-peer 'monster'.

2. Ticket to Ride (The Beatles)

Napster put peer-to-peer computing on the map. Despite the cooling of the Internet hot house, peer-to-peer development has accelerated. Those flush with confidence describe peer-to-peer computing as "edge of network services" and "the next big thing". The phrase "edge of the network" is a bit misleading, but it is catchy. The peer-to-peer architectures move functions once reserved for a centralised server to clients located anywhere and everywhere. *Edge* conveys a better image than the borderless and fuzzy anywhere of Web pages.

The basic idea for peer-to-peer services is that the server is a gateway for far too many actions that capable, powerful personal computers want to take. The server for some types of applications is little more than a chokepoint. Servers have to be watched and maintained by system administrators, whose administrative processes get in the way of what users want to do.

If Napster taught 30 million computer users anything, it is that the old-fashioned mainframe way is not for them. Some enlightened Internet users recognise that servers do perform some useful functions, but with the increasing power of the basic desktop machine, seemingly ever-increasing bandwidth, plus a growing array of wireless computing devices-servers and clients were not the way to deliver what some Internet users wanted. Peer-to-peer architectures have become the ticket to ride the Internet in an unencumbered, organic way.

Napster demonstrated that linked individual PCs can work as a functioning system and provide extremely popular services. Most peer-to-peer architectures distribute computing tasks across two or more machines. Because peer-to-peer computing is comparatively immature, even the most robust peer-to-peer systems routinely suffer performance and maintenance problems. Peer-to-peer architectures are not perfect. In some peer-to-peer architectures, serv-

ers (virtual and real) must be used to provide directory and database services. Napster used a clutch of servers to hook those looking for songs with those who had copies of the song ready for copying.

Computer scientists know that many ‘flavours’ of peer-to-peer computing will be entering the ecosystem of the Internet. There is considerable disagreement about the details of how specific services can and should be provided in a particular implementation of a specific peer-to-peer architectures. There are a handful of basics that can be found in some form in the types of peer-to-peer architectures described in this section. They are:

- *File transfer.* This is moving a binary file from Machine A to Machine B without the need for a human to do much more than click a button.
- *File sharing.* This is the set of security bits that tells another system that it is in order to read, change, copy or perform some other function on another machine offering a directory to the public or to users with specific permission.
- *Telnet services.* The basic plumbing of any Internet Protocol function resides in some form in peer-to-peer architectures.
- *Instant messaging.* America Online -type communications with provisions for voice and video.
- *Software transfer and distribution.* Mechanisms to distribute to multiple clients updates or new builds of the software required to make a peer-to-peer architecture work. New services that support automatic network-based storage and retrieval services are increasingly important.

Collaboration-Peer-to-peer networks support shared functions often intended to replicate the features of sitting in a meeting, looking at PowerPoint slides, and making changes to a spreadsheet so everyone in the meeting can see the numbers. What makes peer-to-peer applications interesting to many developers is that the network of machines can support agent-based and automated services. Without a server acting as a traffic cop, intelligent scripts can ferret out information and perform a wide range of tasks. Software can watch for new information and take specific actions when a particular event occurs in a folder on a machine in a peer-to-peer network. For example, a lover of an ‘artist’ such as Eminem will be alerted via e-mail when a new Eminem album is released. The structure of a peer-to-peer network unfolds a rich landscape which can be mined, monitored and explored in many ways. Individual peers on a network can largely control their own actions.

Each of the major architectures for peer-to-peer networks can be modified. The six core functions can be mixed and matched to meet specific needs. The next section defines briefly each major type of peer-to-peer architecture that is needed. Once the definitions are behind us, we will look at examples of each major type of peer-to-peer architecture.

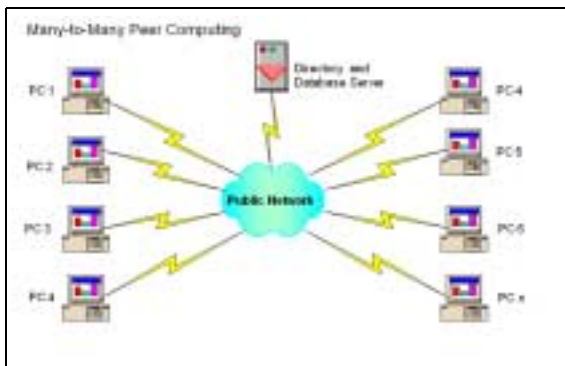
Peer-to-peer architectures perform the same type of functions that most users are familiar with. The jump into the great beyond is that peer-to-peer services are distributed over multiple computers unbounded by the reach of the cabling in an office. There are three types of peer-to-peer architectures in use at this time:

- Many-to-many computing.
- Few-to-many computing.
- Few-to-few peer computing.

The terminology is confusing enough to make differentiating them tricky. Simplistic diagrams of each of the three types of architecture accompany brief descriptions of each of these main types.

2.1 Sweet Dreams (Marilyn Manson): Many-to-Many architecture

Napster and its variants are excellent examples of many-to-many computing, the architecture that gave sweet dreams to many music lovers. The architecture is sometimes described as the Gnutella model, or public information sharing. Any number of computers can make files available to other users.



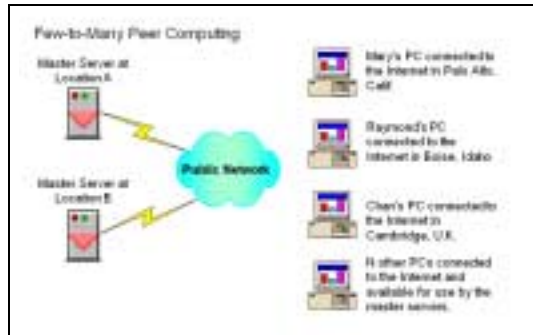
Many-to-Many

The database containing information about what files are available to others resides on a database server. The next-generation peer-to-peer architectures exploit virtual directories and database servers. Because these exist only in memory, shutting next-generation servers may be more difficult. In a peer-to-peer network, a search engine

allows any user to locate information on any machine included in the peer-to-peer system. When a match is found, the user is able to link to a specific machine and copy the required file or information. An important distinction between Napster and Gnutella is that Gnutella does not have a central directory as does Napster.

The many-to-many system does a good job of providing access to files on peers. However, when network traffic is heavy, system performance can degrade. Many-to-many peer computing is not scalable in the same way as traditional client-server architectures. Maintenance can place additional demands on the architecture.

2.2 Space Oddity (David Bowie): Few-to-Many architecture

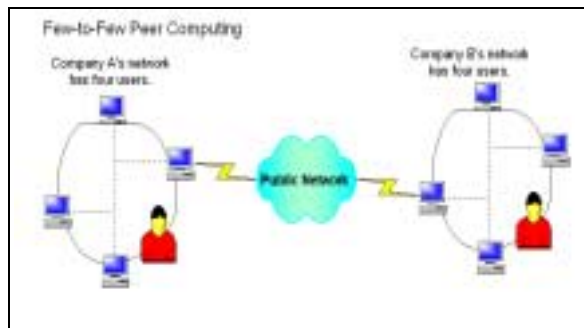


Few-to-Many

Few-to-many computing is cleverly implemented in the SETI@Home system. (SETI is an acronym for the Search for Extraterrestrial Intelligence). A small number of master servers distribute tasks to computers connected to the Internet. The master servers collect the results of the tasks or problems distributed to the individual PCs available to the master servers.

Many PCs work on specific tasks and make the results available to the master servers. This approach is also called distributed computing. Many computers work on problems that would require massive centralised computing resources to handle if the lower cost few-to-many model were not available.

The few-to-many architecture of SETI@Home has inspired a number of commercial enterprises to assemble peers. The business proposition of these companies is to sell the network's computational and storage capacity to enterprises with large-scale number crunching tasks.



Few-to-Few

2.3 My Friends (Red Hot Chile Peppers): Few-to-Few architecture

Few-to-few peer computing allows users to create a virtual 'space' without having to involve a system administrator or invoke complex commands. Any user regardless of

location and type of Internet connection can enter a secure virtual space and interact. Functions are controlled by the participants. Typical activities include shared Web browsing, text and rich media messaging, sharing files, etc. The eight users see and share as if each user were linked on a single network in a private virtual space.

Few-to-few architectures can provide useful, task-oriented services without the cost and complexity of a Lotus Notes / Domino installation. Unlike the Internet, few-to-few architectures can accommodate server functions by creating a 'virtual server' using the resources of each machine in a particular virtual space. For this reason, organisations are likely to take a close look at few-to-few systems of peer computing.

3. Into the Great Wide Open (Tom Petty & The Heartbreakers): Gnutella

Many-to-many computing is the digital equivalent of Spanish conquistadors standing on the shores of the New World. The opportunities were limitless, and the resistance mounted by the indigenous people was slight. Peer-to-peer computing is personified in the media by one Gene Kan; he has been one of the most visible drivers behind the Gnutella portal that is located at <http://gnutella.wego.com>.

But the motive force of Gnutella was Justin Frankel, the programmer who created Winamp, one of the first, hugely popular 'skinable' programs for listening to music via the Internet. Winamp gave the users unprecedented control over what to play and what type of interface best suited each user's personal tastes. The 'skins' available for Winamp can make the Winamp interface look like a zebra-skinned settee or a control panel from the starship Enterprise.

Mr Frankel sold his company to America Online (soon to be the owner of EMI and Warner recording properties), and while an employee of AOL he and a handful of programmers polished Gnutella in their spare time in early 2000. Without alerting AOL executives, Mr Frankel released the program as open source code. Once released, anyone could download the code, make changes and use the software to facilitate many-to-many online services.

Mr Frankel seemed unconcerned about the machinations of corporate America in general and AOL in particular. He soon found himself the focal point of AOL corporate scrutiny. America Online removed Mr Frankel's Nullsoft Web page that provided access to the Gnutella source code. But the action was a textbook case study of locking the digital barn after the digital horse had galloped into the great wide open Internet. Almost immediately other programmers took Mr Frankel's packet of code, disassembled it, rebuilt it, and in a matter of months made available the original code base plus six or seven important variants by the summer of 2000. Today one can see Gnutella as one of those programs that is playing a role similar to that played by WordStar in

word processing or Lotus 1-2-3. Although better programs come along, WordStar, Lotus 1-2-3 and Gnutella change fundamental behaviour.

Anyone wanting to use Gnutella must download software. For example, Gn00a can be used as it is. For the adventuresome and Big Music's squads of programmers, download Gn00b, a version that one can use to build a purpose-built Gnutella application.¹

A Gnutella application plays a dual role. Once a user installs Gnutella, the personal computer is both a client responding to requests from others in the Gnutella 'space' and a server providing services to other users also in the 'space'. (A 'space' is the virtual meeting room in which a group of users gathers). The essence of Gnutella's many-to-many model is that a separate server is not needed for most functions. Peers provide the types of services normally assigned to a server in a traditional client-server network.

The Gnutella space is a virtual environment in which the peers can interact as if they were one system. Anyone with the Gnutella software running is a peer in this type of architecture. Security and access are not features implemented with the type of stringency one would find in a corporate network or in America Online's instant messenger environment. There are some similarities to the 'party line' telephone connection. Anyone on the line could talk and exchange information.

After the software has been installed on a personal computer, the user connects to hosts that act as entrance points to Gnutella peers. The hosts usually provide only the IP address and port data and do not share files. One can locate Gnutella hosts by clicking to such portals as Gnutelliams and looking at the list of gateways now online.² Because there is no central database containing look up tables for mapping users to data, Gnutella affords some user anonymity. The Gnutella software implements a procedure to tokenise and then hides the identity of the user generating a query.

When a connection is established, machines on the network communicate via messages. Each peer receives and sends messages. Gnutella supports a number of standard Internet and special purpose messages, including:

- Internet pings that ask, "Are you online?"
- 'Pongs' or messages that say, "Ready to send and receive". A pong contains the IP address of the ponger, the port to use and data about the

1. *These programs are available at the Gnutella Developer home page, <http://gnutelladev.wego.com/>, managed by Gene Kan.*

2. *Gnutelliams is a useful Gnutella portal offering a directory of Gnutella client downloads for Windows, Linux/Unix, Java and Macintosh. It is at <http://www.gnutelliams.com>*

files available to peers. Peers forward pongs in order to allow others in the space to locate machines and files.

- Queries are messages that say, “I am looking for Mozart files”. These are forwarded to all the machines in the space. Each query has a unique identifier, but the originator of the query is not included in the message.
- Answer messages are replies to queries. These messages include the IP address, port and any other information required to complete the file transfer. These responses include a unique identification string associated with the replying peer. Answers are not broadcast to all peers in the space. The answer is sent backwards to a specific peer along the path taken by the original query. It is therefore nearly impossible to trace query responses in a Gnutella system.

Get and *push* messages are requests for file transfer from the peer with the data to the peer wanting the data. When a peer with the requested data is behind a firewall, Gnutella allows that information to be pushed through the firewall to the requesting peer. This is necessary because firewalls are set up to prevent certain types of file transfers. For unlucky Gnutella users who find that both the peer requesting data and the peer providing data are behind firewalls, the file transfer is not possible.

In order to keep the Gnutella space from flooding available bandwidth with infinite rebroadcasts of messages, time-out mechanisms are baked into Gnutella. Once sent, each message has a time-to-live function which decrements to zero. A zero TTL value causes the message to die.

At the same time as university programming classes were digging into Gnutella and many-to-many applications in class and in extracurricular hack sessions, Gene Kan emerged as the spokesman for the Gnutella movement. In March 2001 he sold his InfraSearch technology to Sun Microsystems. The InfraSearch service was online for a short time. A user entered a search word or words and the results from geographically separate servers were displayed. A virtual index held the pointers to the objects on each participating server.

As president of XCF Ventures, Mr. Kan has been tapped by the media as the principal spokesman for Gnutella. He has a high profile and presents the polished presence necessary to navigate the treacherous straits between hostile legal forces and paranoid multinational content companies. Gnutella variants are used by thousands of Internet users to locate and download pirated software, digitised motion pictures, pornography, and, of course, MP3 files. To get a feel for the traffic in Gnutella fuelled services, query a newsgroup for the string ‘warez’ and follow a handful of postings.

The many-to-many environment demonstrates asymmetrical sharing. Although there are thousands or millions of people on a system, the majority of the content is provided by a small percentage of users. According to Eytan Adar and Bernardo A. Huberman, “almost 70 percent of Gnutella users share no files, and nearly 50 percent of all responses are returned by the top one percent of sharing hosts. Furthermore, we found out that free riding is distributed evenly between domains, so that no one group contributes significantly more than others, and that peers who volunteer to share files are not necessarily those who have desirable ones”.¹

A glimpse at the future of Gnutella sparkles off the exterior of the most recent release of Bear Share, now in version 2.2.0. This is a Gnutella-based application from Free Peers, whose marketing collateral says, “Bear Share provides a simple, easy-to-use interface combined with a powerful connection and search engine that puts thousands of different files in easy reach”.²



1. Eytan Adar and Bernardo A. Huberman, *Free Riding on Gnutella* http://www.firstmonday.dk/issues/issue5_10/adar/index.html, October 2000
2. For the software, click to <http://www.bearshare.com>. For gateway information, click to <http://www.bearshare.net>

The most recent build provided the screen shot of the initial connection screen. Once a user locates a host, tabs provide point-and-click access to upload, search, and monitor functions. The interface, documentation and support provided for Bear Share are better than for some commercial software applications. Tabs, icons and an intuitive layout allow users to load the program and begin looking for specific information without referring to online help or the manual.

Among the features in this release of the program are greater code stability, support for multiple active searches, active monitoring and updating of shared directories, and reverse lookup for computer addresses, among others. The Florida-based company generates revenue from advertising and the operation of various hosting services.

Gnutella improved on the Napster model by eliminating to a large extent the centralised server that made Napster vulnerable to its critics. Many-to-many applications work when large numbers of people participate. More data and people using the system increase the likelihood that the desired information will be available from someone on the network. But success has a price.

The number of users on the system creates two problems. First, under load, many-to-many systems are sluggish. With the increased number of users, visibility and buzz increase. The heightened visibility and the availability of information increase scrutiny of the content on the system. Copyright is the main point of contention in Gnutella-type systems.

Content company executives may want to ponder the words of Rob Lord, one of the programmers who developed Gnutella, "We didn't get into this 'space' cuz we're internet gold-seeking cuckoos. We're legitimate nihilistic media terrorists as history will no doubt canonise us".¹

4. Just Can't Get Enough (New Radicals): low cost supercomputing

At any one time, millions of computers are sitting idle and connected to the Internet. Why not harness the computational capabilities of these computers, creating one of the world's largest and least expensive supercomputers? This is the question that Berkeley computer scientists, among others, asked and answered. Distributed computing is the process of linking computers over the Internet and putting their idle processing power to work to create a 'virtual supercomputer'. Many computer scientists cannot get enough computing horsepower to do their work. Using a peer-to-peer architecture offers one way to get access to calculating and storage resources.

1. *<http://www.nullsoft.com/>, retrieved on March 14, 2001. From this site one may download BlorpScript, a GPL'd PHP image/content browsing system and AAmazing, a nifty Winamp visualisation plug-in.*

The most visible outcome of a scientist's need for inexhaustible computational horsepower is SETI, or Search for Extraterrestrial Intelligence.¹ The SETI Institute serves as an institutional home for scientific and educational projects relevant to the nature, distribution and prevalence of life in the universe. The institute conducts and encourages research and related activities in a large number of fields including, but not limited to, all science and technology aspects of astronomy and the planetary sciences, chemical evolution, the origin of life, biological evolution, and cultural evolution. The SETI@Home project is not directly sponsored by the SETI Institute. But the institute has benefited from the publicity generated by the SETI@Home project.

SETI@Home project allows individuals to make their computer available for crunching the data returned from monitoring devices. A user downloads a software bundle that installs as a screen saver on the individual's computer. Participants include individuals, academic institutions and various corporations, including Lehman Brothers, the French telephony company Alcatel, and the upscale consulting outfit Booz, Allen & Hamilton. Once a personal computer is set up to participate in SETI@Home, servers at Berkeley copy a chunk of data collected with the Arecibo Radio Telescope in Puerto Rico, as part of Project SERENDIP. The SETI Institute is a major supporter of the SERENDIP search.

The idea behind SETI@Home is to take advantage of the unused processing cycles of personal computers. When the participants' computers are idle, the SETI@Home applet downloads a 300 kilobyte chunk of SERENDIP data for analysis. The results of this analysis are automatically sent back to the SERENDIP team and combined with crunched data from SETI@Home participants in more than 200 countries. In the two years of SETI@Home operation, no sign of extraterrestrial life has been found. However, the distributed system has received about 300 million results from users around the world.

The limitations of this system are that only a 2.5 megahertz piece of the observed spectrum is analysed by SETI@Home. The data processing does not occur in real time so that interesting signals are flagged and then analysed further by project engineers. The advantage of this scheme is that it permits looking for a variety of signal types that the current SERENDIP processing lacks processing capacity to analyse.

Other academics have tapped distributed computing to tackle other computationally-intense problems, including prime numbers, Fermat numbers and optimal Golomb rulers. An organisation called Distributed.net has focused on cryptography since 1997.

1. More information is available from <http://setiathome.ssl.berkeley.edu/>

Among the commercial enterprises developing or offering products and services for few-to-many architectures are:

- *Applied Meta Computing* has government customers such as NASA and the US Defense Department as well as clients from the Fortune 500. The company has received what it calls “a significant investment” from Polaris Venture Partners, an early stage venture firm with over \$1.2 billion under management.¹
- *Data Synapse* offers its Web Proc product to financial institutions, banks, broker deals and others with high-demand computational requirements.²
- *Distributed Science* is a result of merging DCypher.Net into Distributed Science. The company provides both computational and network storage services. The company aggregates excess computational and storage capacity and sells it to its customers. Distributed Science’s customers do not have to make massive investments in additional hardware. At the end of 2000, Distributed Science said it had signed up 145,000 personal computer users with more than half (about 72,000) from outside the United States. Each computer user makes a personal computer’s idle time available to Distributed Science.³
- *Entropyia*, another distributed computing start up, obtained \$7 million in funding from Mission Ventures and Silicon Valley Bank’s San Diego Technology Group. Entropia focuses on enterprise distributed computing. The firm licenses its Entropia 2000 Enterprise Server. The software enables customers to get the same power and scalability of its Internet technology on intranet local and wide area networks. Entropia offers application software integration services.⁴
- *Parabon Computation* offers the Frontier distributed computing product and for-fee services aimed at biotechnology, financial and pharmacology research. A person wanting to participate in a Parabon project can download the Pioneer applet. Parabon pays individuals to participate in its projects.⁵
- *Popular Power* has started paying people for making their personal computers available for influenza virus research. Brian Behlendorf, a co-

1. <http://www.appliedmeta.com>

2. <http://www.datasynapse>

3. www.distributedscience.com and ‘Distributed Science Spreading the PC’, Bob Savage, *LocalBusiness.com*, http://www.localbusiness.com/Story/0,1118,LAX_537355,00.html, December 13, 2000

4. <http://www.entropyia.com>

5. <http://www.parabon.com/>

founder of the open-source Apache Web server project, is one of the investors in the company.¹

- *United Devices*, founded by SETI@Home founder David Anderson, secured \$13 million in venture capital funding from Softbank Venture Capital, Oak Investment Partners, and others. The company's major customer is Exodus Communications. United Devices provides software tools to allow those requiring distributed networks to build, enable, deploy and support an Internet-distributed computing project.²

For enterprises looking for a low-cost way to build a distributed computing application, TurboLinux sells EnFuzion for about \$400. EnFuzion lets corporations farm out computing tasks across their computers using any operating system. Companies exploring distributed applications include J.P. Morgan, Procter & Gamble, and Motorola.

The applications for few-to-many computing are focused on computational or storage intensive tasks. This technology has some interesting commercial applications. The principal benefit from the technology is reducing the capital cost associated with adding specialised computers and storage systems. Costs for system upgrades cannot easily be driven to zero, but more efficient use of computing resources is one of the primary benefits of distributed architectures. The downside is that optimising distributed system performance is still a challenging task. Schedulers and caching systems offer some performance improvements, but the field is comparatively new. Distributed computing has tremendous potential in engineering, scientific and research applications.

5. Chemicals Between Us (Bush): Groove.net

Few-to-few peer computing is a what might be called a 'hybrid architecture'. It blends one or more virtual servers that can perform security and performance functions with Gnutella-type peer-to-peer computing. The potential of distributed computing can be tapped if those participating require access to specific computational tasks or need specialised storage services greater than any single user's hard drive capacity. It allows the fluidity and openness of the many-to-many architectures but in a more controlled, tightly structured environment. The essence of this architecture is that a group of people working on a project can bond and interact in a secure virtual space.

The principal problem with the many-to-many model is security – or lack of it. The hybrid architecture allows log in, authorisation, usage tracking and other services to be handled by a virtual server running necessary authorisa-

1. <http://www.popularpower.com>

2. <http://www.uniteddevices.com/home.htm>

tion services. Once a user is logged in, the few-to-few peer network looks and feels like a Gnutella session or an America Online chat room.

The leader of the few-to-few architecture is Ray Ozzie, the developer of Lotus Notes. After leaving IBM, Mr Ozzie conceptualised a software environment that would exploit Internet services, yet provide the specialised functionality that enterprises and commercial organisations require. Mr Ozzie and his team worked for about three years to get Groove to the market. Groove is similar to a house decorated in Early American and Modern furniture. Handled with restraint and good taste, the mix is at once familiar and trendy. Groove has Notes-like functions, but it has the throttle touch of a Gnutella applet.¹

Groove is a system that brings together people, messages, discussions, documents and applications in the context of a shared project or activity. The application includes a range of collaborative tools, including a threaded discussion, chat, a notepad, a sketchpad, a calendar and a file archive. One can emulate to a degree Groove functions by cobbling together a suite of applications for a Notes and Domino environment. One can provide similar functionality by integrating Internet-accessible services from such companies as vJungle, building the missing bits from vJungle's OpenEx tools. Groove is one of the first companies to offer the hybrid service with security and control features baked in, not left out or glued on as an afterthought.

The Groove server is actually a bundle of code called the "Groove Relay". No single machine performs dedicated Relay functions. Instead, each peer



1. <http://www.groove.net> and <http://www.groovenetworks.com>

executes relay functions as required. Purists may argue that Mr Ozzie has not created a hybrid peer-to-peer architecture. For those concerned with security, it is important to understand that the implementation of a virtual relay server solves in part many of the security challenges associated with Gnutella, for example. Groove can be configured to integrate with users' existing directory and certification systems.

The Groove software is available without charge. Click to www.groove.com to download the application that allows users to create Groove shared spaces. Groove users, even though invited to a shared space with security services running, could delete information from a shared space. They could invite other users to the shared space, so information expected to be read by one set of eyes is now shared. They could even attach a file such as an '.eve' file with a virus using the Groove files tool. Alternatively, a user could attach a file containing a virus into the files tool. In this case, the virus would not fire until someone launched the attachment. In that case, the virus could affect the Groove user's personal computer.

Groove supports spontaneous collaboration. There is no need for the complex set-up required for audio conferencing. Groove accommodates users on local area networks behind different corporate firewalls, on dial-up connections with dynamic IP addresses, behind cable-modems, and so on. Regardless of connection, a Groove user can create a shared space and invite other Groove users into it to work on a project.

When Groove members create information with the Groove tools, the data reside on the machines of all invited group members. The data are synchronised when a user or a group of users make changes. The biggest drawback to Groove is the quirky terminology the company uses. Instead of 'voice chat', Groove prefers 'transceiver'. Groove's peculiar jargon impedes a new user's understanding of the services and features.

Invitations to join a 'space' are issued using messages. A person who is invited to a Groove space must install the Groove application. Once the Groove applet is installed, the appropriate data are copied to the new user's machine. Each Groove member receives a complete copy of the shared-space data. The copying takes place in the background. Because of the shared space data for a particular Groove collaboration, a context for the messages and data are readily available to any user. This differs from the contextless snippets of electronic mail that often require one or more rounds of telephone tag to resolve and considerable inductive effort.

Anyone in a Groove shared space has the same information and data view as other members. The Groove shared space formed in this manner is secure. Each Groove space can be extended at any time to any person with whom

members need to collaborate. No system administrator or coding is required.

The Groove relay brokers connections for devices behind firewalls and devices that talk with Network Address Translation conventions. Groove provides store-and-forward service so shared-space members can go offline, then reconnect and be resynchronised automatically.

Groove is a new type of product. Like any new network-centric tool, users must embrace it. At the time of this writing (March 2001), the Groove software is available for free. Groove may be a harbinger of how Internet operating systems will operate.

Alternatives to Groove will be coming with increasing frequency in the months ahead. The impact of Gene Kan and InfraSearch on Sun Microsystems will almost certainly lead to a product or toolset that supports Groove-type services. Java and the Solaris operating system are attractive conduits for secure, collaborative peer-to-peer applications.

Microsoft's Sharepoint Portal Server promises to deliver much of the same functionality for Windows 2000 environments. Coupled with Microsoft's 'dot Net' initiative, the millions of programmers familiar with Microsoft languages will be able to snap together peer-to-peer applications by dragging and dropping widgets from a toolbar. The impetus behind this new type of peer-to-peer service is the human need to communicate with the Internet packet as the digital morpheme. One can see mobile devices, notebook computers and deskbound professionals entering virtual, secure spaces for a wide range of activities.

6. Living On The Edge (Aerosmith)

Peer-to-peer computing is an essential component of the pervasive network. Each of the principal peer-to-peer architectures summarised in this section can support a wide range of applications. What is the impact of peer-to-peer computing on information companies producing electronic content?

First, many information companies will make extensive use of peer-to-peer technology in their own organisations. The collaborative power of the few-to-few architecture and the potential cost savings of the few-to-many architecture are easy to explain in terms of costs and benefits. As pervasive computing becomes easier and always-on connections become a reality in major cities, peer-to-peer functions slipstream easily into many workflows. For example, a discussion about a marketing campaign or a new product development project can exercise peer-to-peer systems in a comfortable, intuitive way. It makes sense to share information among team members. It makes sense to perform complex computing and storage tasks using all available resources.

Second, peer-to-peer computing is going to accelerate the development of

new types of information products and services. It is not clear if the drivers of innovation will be the established firms, or upstarts such as Napster. Digitised video, digitised motion pictures, and digitised anything lend themselves to systems similar to those of Napster and Gnutella. This fact has not been lost on legions of programmers under the age of 13 who see no good reason not to make it possible to share games, images, books, videos and music. Portals, print magazines, books and entirely new types of search-and-retrieval mechanisms are already flowing into the market space blasted into Internet users' consciousness with Napster.

Third, commercial enterprises have two short term tasks: *Job One* for commercial information operations will be to understand, monitor and discover how to deal with what promises to be an continuing flow of software that can sidestep digital rights and for-fee distribution systems. The task will be neither easy nor rewarding. Large publishing companies are poorly equipped to deal with the technical innovations of a high school student in Boise, Idaho, or a college student in Osaka, Japan.

Job Two is to recognise that peer-to-peer technology offers new ways to integrate content into professionals' and consumers' everyday activities. A term for this new publishing model is 'in-phase distribution'. The idea is that peer-to-peer technology, combined with pervasive network connections, allows an individual to access facts, numeric data, and information at the precise moment the data are required. 'In phase' information implies a much closer blend of predictive statistics, software agents, and event triggers. Doctors making their rounds want to have access to colleagues who are knowledgeable about a particular patient. At the precise moment information is needed about drug interaction, the physician will be able to shift from data collection, to collaboration, to retrieving on-point information about what medication change to make. Similar opportunities will exist in sales, legal, technical and financial arenas.

In-phase communications means that the delivery of access when it is needed and in the format required by a specific work or leisure context points to new revenue opportunities. Napster demonstrated that peer-to-peer technology can galvanise an industry largely unchanged by the advent of the Internet. After the recording industry, peer-to-peer technology will make its impact felt upon motion pictures, professional publishing, anywhere content and needs can be matched in real time.

The edge of the network is fast becoming the centre of content delivery innovation. For many information innovators under the age of 25, Journey's song 'Don't Stop Believing' is the perfect background to inventing a new information distribution system.