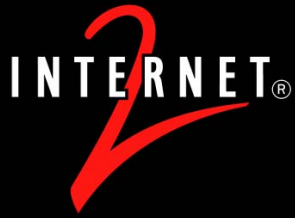




INTERNET®

www.internet2.edu



¿Puede ser P2P una herramienta para la Ciencia y la Tecnología? P2P, tecnología disruptiva...

Ana Preston

Program Manager, Internet2 International

apreston@internet2.edu

JT RedIRIS 2002

Salamanca, España

22 February 2003

Quick Overview of p2p networking

- Frameworks; Pros and Cons

Challenges for p2p

- Technical and Upper layers

Internet2 and p2p

- Highlights on U.S. approach to p2p
- Activities within Internet2

Applications

Final observations...

P2P is not a new technology

- Oct. 29 1969: first transmission UCLA → Stanford Research Inst. (SRI) [UCSB, U. Of Utah]; Peer computing status among independent computing sites
- 1970s-1980s: ARPANET completed (1978); Usenet and DNS
- 1990s: NSFnet and Internet explosion
- 1996: ICQ (bypassing DNS by offering own addressing scheme and allowing end points to be directly addressable with each other)

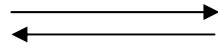
1999: Napster is born – universities notice first

2002: Napster is gone – apps continue to emerge

..” the number of file-swapping, and peer-to-peer websites, has grown by 535 percent in the past year, despite legal efforts to have them shut down.” According to the study findings, the number of p2p websites totals nearly 38,000. (Summer 2002)

<http://www.websense.com/>

simplest form..



- Models:
 - Centralized - Napster
 - Decentralized – Gnutella, Freenet
 - Controlled-decentralized | Hybrids – Morpheus, Groove
- Components: client, server, servents
- Main differences: how information is shared and how much information is shared
- Pros and Cons – hybrids seems like a better compromise

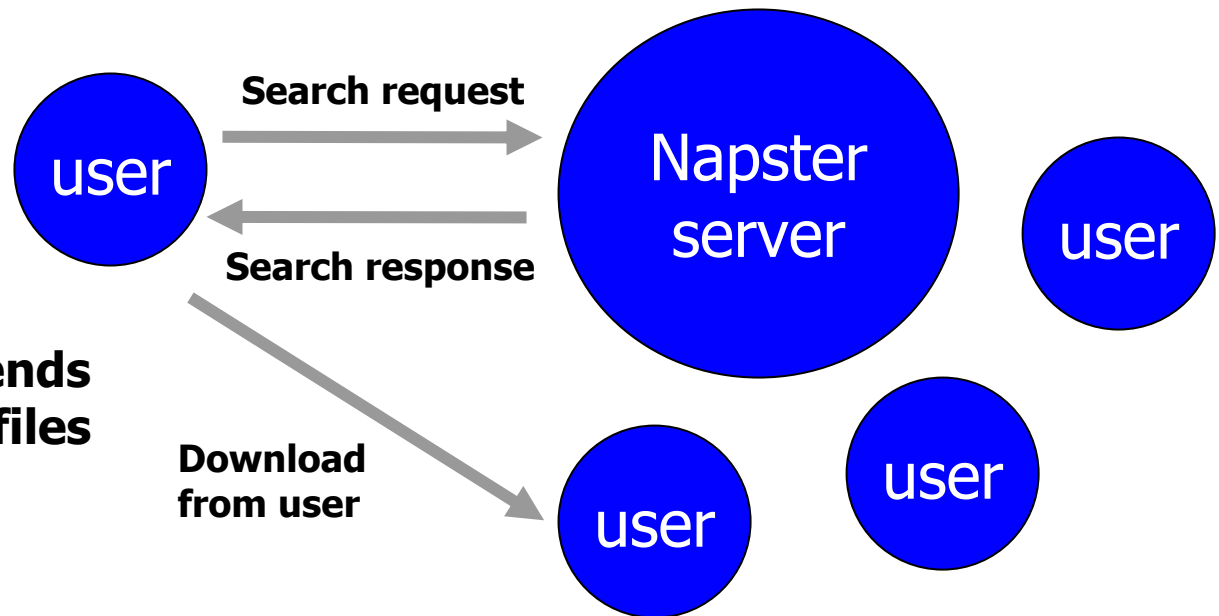
Centralized: Napster

Napster used centralized servers to keep a catalog of available files.

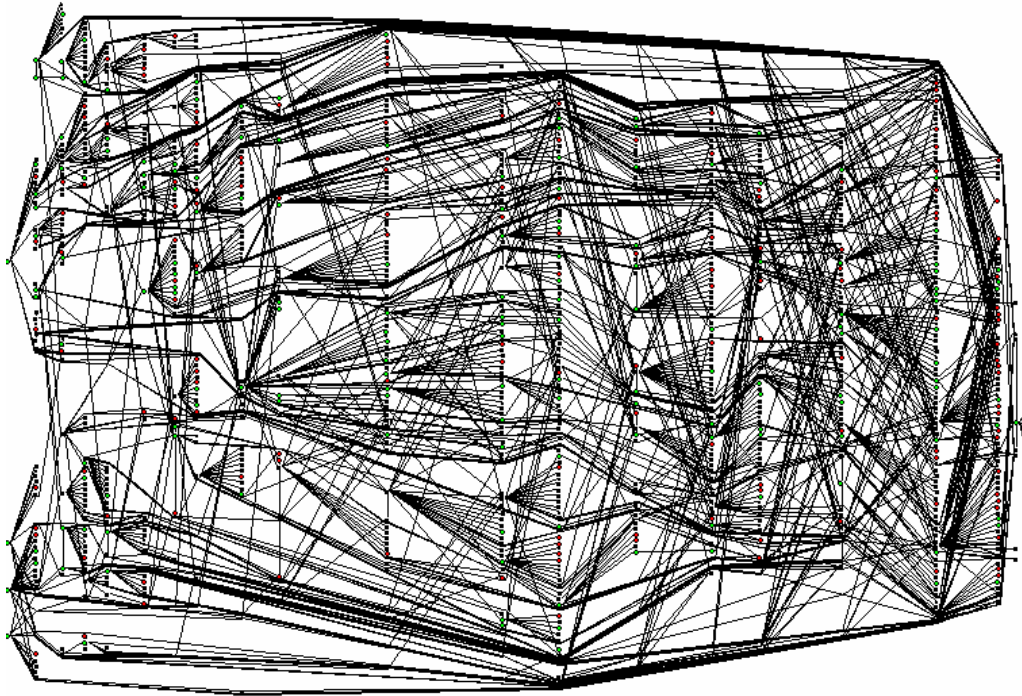
**1. User sends out request
Napster searches
central database**

**2. The central server sends
back a list of available files
for download**

**3. Requesting user downloads
the file directly from another
Napster user computer**

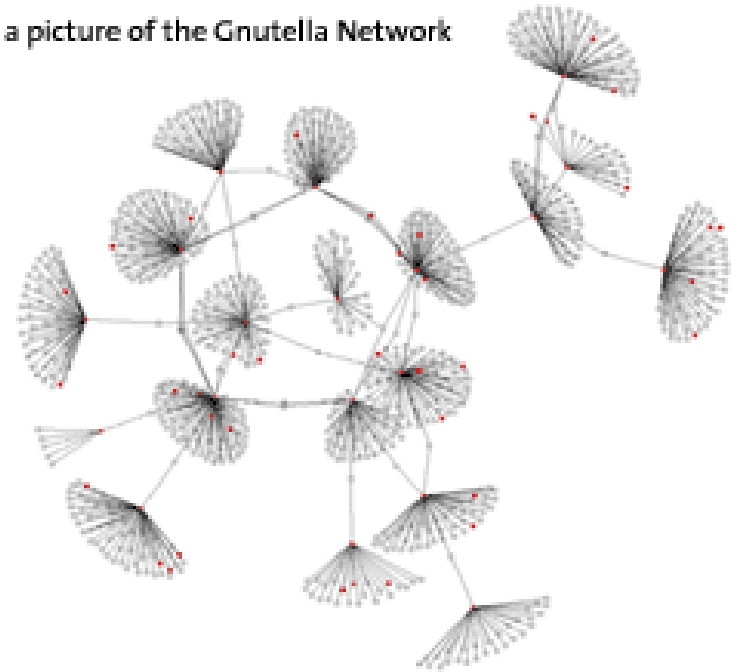


Decentralized: Gnutella network



***Partial Map of the
Gnutella Network***
<http://dss.clip2.com>

a picture of the Gnutella Network



<http://www.limewire.com>

- See also ***gnuTellaVision: Real Time Visualization of a Peer to Peer Network***

<http://www.sims.berkeley.edu/~rachna/courses/infoviz/gtv/paper.html>

Another example: Freenet

Somewhat similar to Gnutella but...

as file passes through 'vine-like' framework,
the file makes a copy of itself at each point
along its route

Implemented encryption to hid the originating
point of the file

vision of open source project is to allow all
information, copyrighted or not, to be
distributed anonymously and untraceable in a
p2p network



Centralized vs Decentralized: pros and cons

Centralized:

Pros:

- More effective, comprehensive searches
- Access is controlled

Cons:

- System has single points of entry; one fails could bring whole system down
- Broken links, out of date information.

Decentralized:

Pros:

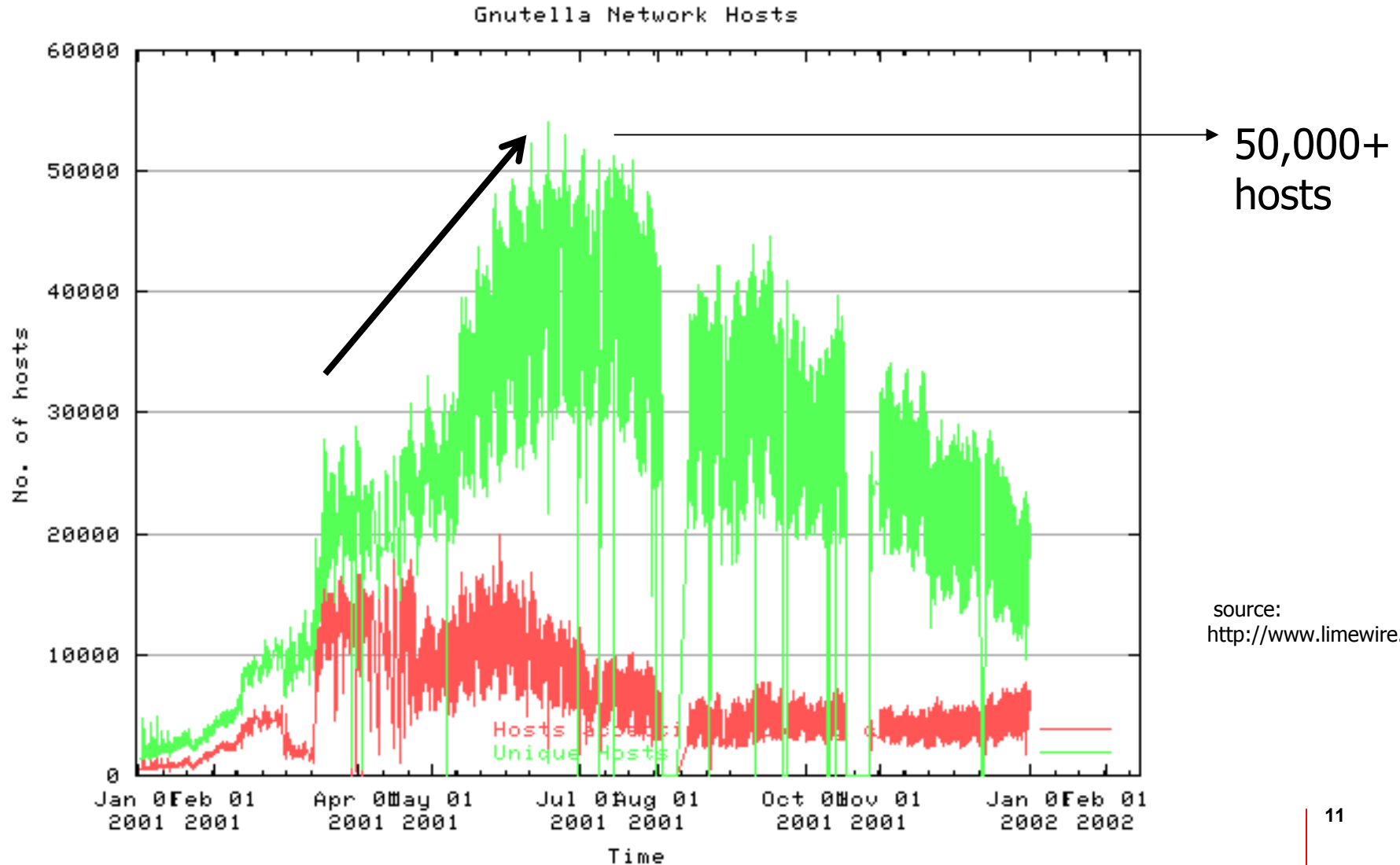
- Users speak directly to other users with no intermediate or central authority
- Isolated node failure can quickly and automatically be worked around.

Cons:

- Free loading
- Scalability
- Searches are less effective and can be slow.

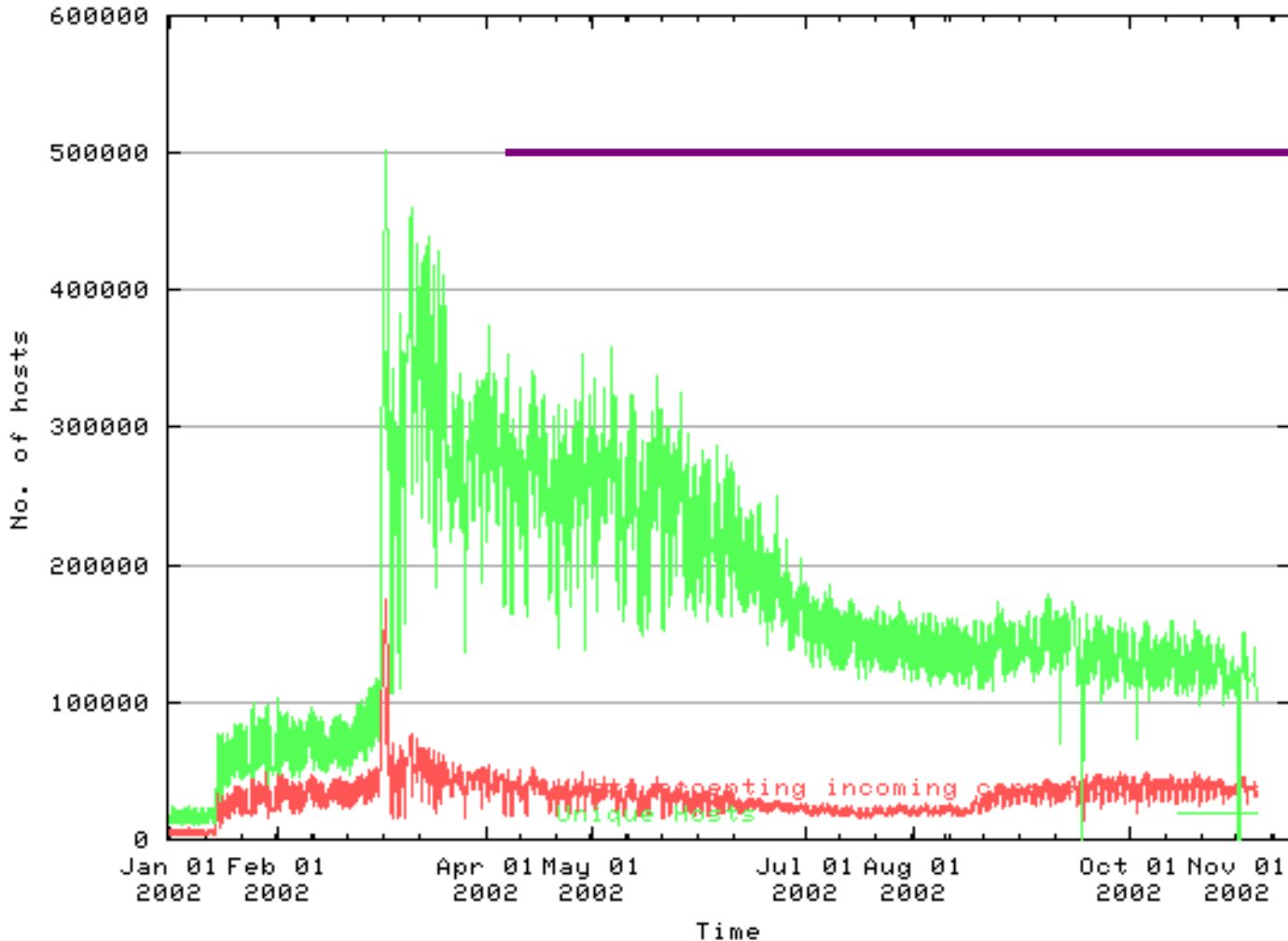
Gnutella network evolving to include “controlled decentralization” (limewire, bearshare, toadnode)

Gnutella Host Count – 2001



Gnutella Host Count - 2002

Gnutella Network Hosts (courtesy of LimeWire.com)



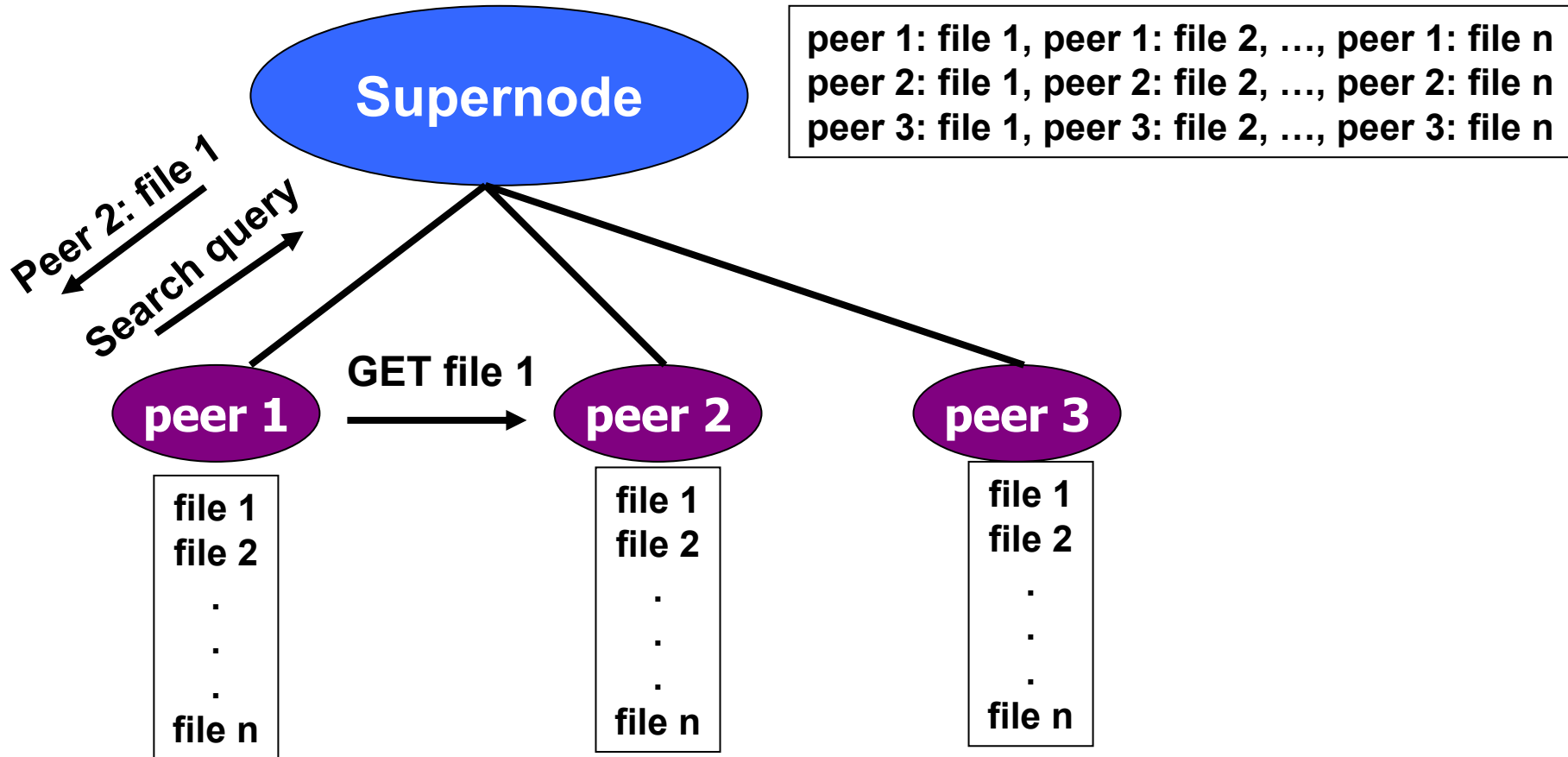
500,000
Hosts
(!)

source:
<http://www.limewire.com>

Hybrids: controlled decentralization

- Characteristics of both centralized and decentralized frameworks: User's computer may act as client, a server or a servent; there are server operators which may control which clients and/or servents are allowed to access a particular server.
- e.g., Morpheus:
 - The full gamut (not just mp3's)
 - Uses metadata (XML) to describe contents of file; easier to find things
 - Improved download performance and faster searches (faststream)
 - "No more" incomplete downloads. SmartStream: Fail-over system that attempts to locate another peer sharing same requested file, and automatically resume download where it left off at failed host.

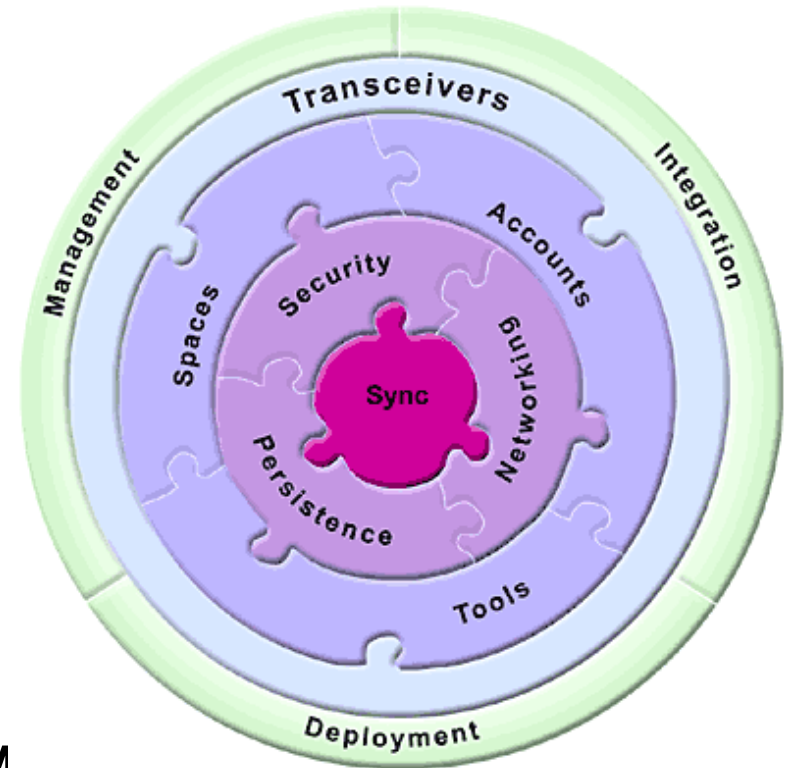
More on Morpheus



Specifically designed for workplace and controlled decentralization

For businesses that may want to use p2p environment with internal controls

Incorporates file sharing, IM, blackboard and secure environment.



Groove Architecture Diagram
<http://www.groove.net/>

Architectural:

- What scales? P2P may mean being able to choose appropriate balance between centralization and decentralization.
- What architecture to preserve end to end? [IPv6 and p2p, other]

Peer/Resource Discovery:

- Finding things in p2p systems gets harder when not dealing with widely replicated content (MP3s)
- Pressing importance on search architectures and metadata

Security/Trust:

- How secure can one feel in a decentralized network?
- Privacy and Trust, authentication and authorization
- How does a content provider (CP) establish a level of trust in a p2p framework? Will the CP trust the client to follow distribution rules? And the other way around?
- Encryption mechanisms, sandboxing, PKIish... Other ways

Standards/Interoperability:

- Lack of standards, coupled with the immense growth this has resulted in some of the development processes impacting the management issues.
- Interoperability between apps by implementing common protocols or XML-based standards.
- JXTA (<http://www.jxta.org>): an open source initiative to let existing and future computing platforms of all types and sizes interact as peers.

Resource Management:

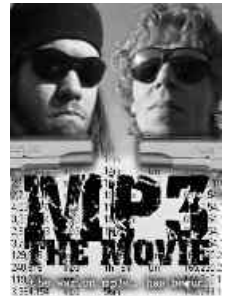
- “Network is only as strong as its weakest link”
- Bandwidth and continuously upgrading network capacity is not a long-term viable solution
- Better network awareness into some of the file sharing applications
- Maybe that p2p is forcing us to look for new solutions architectures in resource management

Other hurdles that give p2p (and us) a headache: copyright and legal

Copyright vs. new value of p2p comes from sharing information and building on it

At US universities: no clear consensus on how to approach their responsibility with US copyright law

- Consensus that universities are responsible for informing students about the law but not to police them
- “Right now, we can justify throttling p2p traffic because of copyright issues – when it becomes legit, we have a problem” (Bill St. Arnaud at p2p workshop)
- For fun, see “MP3” the movie (<http://www.filmwave.com/mp3/>)
MP3 task force will put an end to MP3s forever...



P2P and higher Ed in the US

Initially, used to just be blocking vs. not blocking b/c of liability and network performance.

Then... Digital video/movie files of 200-800MB downloaded with KaZaA or similar P2P file sharing applications

Metering and allocation fair use of resources; opting for user education and cooperation, bandwidth limiting, adding capacity, additional fees to cover bandwidth costs



Internet2 and p2p: current mechanisms – technical/policy

- special purpose middleboxes deployed at the campus edge -- "QoS appliances" support a range of functionality, including: application classification, DSCP marking, shaping, differentiated queuing, and TCP window size spoofing OR
- per-user or per-subnet rate-limiting OR
- usage-based accounting, charge-back, quotas, or other policy mechanisms OR
- additional equipment and operational expenses, may also be significant non-financial impacts on both commodity and Internet2 connectivity (e.g. loss of end-to-end transparency or reduced performance).
- what is a university to do?...

Internet2 QoS Working Group:

- Campus Bandwidth Management BoF (I2 MM, Arlington, VA May 02)
<http://www.internet2.edu/qos/wg/200205-Arlington.shtml>
- QoS Appliances: Disease or Cure? (Joint Techs, Tempe, AZ, Jan 02) <http://www.internet2.edu/qos/wg/200201-Tempe.shtml>
- Campus Bandwidth Management (Fall 2002 Internet2 Member meeting)

Do we ensure the delivery of high priority traffic and let all other traffic go best effort, or do we try to shape non-desirable traffic as well?

But, quickly changing characteristics of non-desirable traffic and ability and cost associated with the products trying to keep up with it (and what is non-desirable?....)

Maybe the questions are no longer relevant...or need to evolve pro-actively...



In the US – cont.

Panels and BoFs at several Internet2 and Higher Ed conferences

Collaborative Computing in Higher Education: Peer-to-Peer and Beyond

January 30-31, 2002, ASU, Tempe, AZ

Our goal: to explore the technical and future dimensions of the fast-growing P2P services spaces and the opportunities and challenges presented for universities.

P2P does not equate to file sharing; yet whatever it is universities are seeing something happen

Still we are haunted by what to do and how to keep up...

Nobody had done this before

set the stage for academia, industry, and government to share experiences and suggest future directions for P2P and related technologies in higher education.

Workshop pages, presentations and program

<http://www.internet2.edu/activities/html/p2pworkshop.html>

Workshop Draft Summary

<http://www.internet2.edu/activities/html/p2pworkshop-summary.html>



Internet2 p2p working group

under Internet2's end-to-end performance initiative -
<http://www.internet2.edu/p2p/>

The working group will focus on best practices, trends and collaboration efforts that were seeded on Jan. workshop.

Working Group Chairs

David Futey <dfutey@stanford.edu>

Linda Roos <lroos@oar.net>

Email list

- Archives at listserv.utk.edu/p2p/archives.html
- Subscribe: send an email to listserv.utk.edu and in the body:
 - subscribe p2p <yourname>

Endpoints on the Internet exchange
information and form communities
collaboration and workflow opportunities

Edge resources (content, storage, computing
power, bandwidth, human attention...)

how these can be better utilized and
accessible in real time (presence) and

“each” with their own “identity” (autonomy at
the edge)

File sharing / Content Distribution

- Not just mp3s/media sharing ones
- Distributed searching: Used to easily lookup and share files and offer content management (NextPage)

Instant Messaging

- Jabber, IM

Distributed Computation:

- Use under utilized Internet and/or network resources for improving computation and data analysis

P2P groupware / Collaboration

- Cooperative publishing, messaging, group project management. Secure environments are offered in some products

Development Frameworks; Development tools and suites

- Project JXTA (Sun), .NET (Microsoft)

[See **2002 P2P Networking Overview** - <http://www.oreilly.com/catalog/p2presearch>

Also: http://www.openp2p.com/pub/a/p2p/2000/12/05/book_ch01_meme.html

Classifying p2p...

Another classification of p2p applications according to function and audience (Burton Group)

	Research	Commercial	
Business	JXTA .NET Jabber Open Cola Folders	NextPage WorldStreet Groove Jibe Endeavors Entropia DataSynapse	Business
Consumer	Cancer Research Genome Seq Free Haven SETI@home FreeNet Publius Mojo Nation	CenterSpan OpenCola SwarmCast ICQ AIM Kazaa, Morpheus Gnutella LimeWire Bearshare	Consumer
	Research	Commercial	

P2P is not distributed computing; similar challenges and issues from: sharing and taking advantage of resources available at endpoints and harnessing their power for computationally intensive problems

SETI@home, fightaids@home,
genome@home

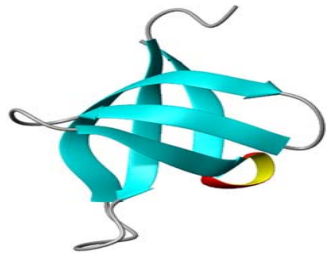
Grid computing and e-science

- Computational grids to solve/simulate real-life problems
- E-Science
- Commercial applications
- United Devices, Entropia, Avaki, etc.

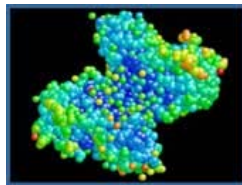
Science used to about test tubes, wet labs and big instruments

But increasingly science is moving to networks and computers; Applications that harness the power of the network at the edges

Science is more global and distributed



folding@home



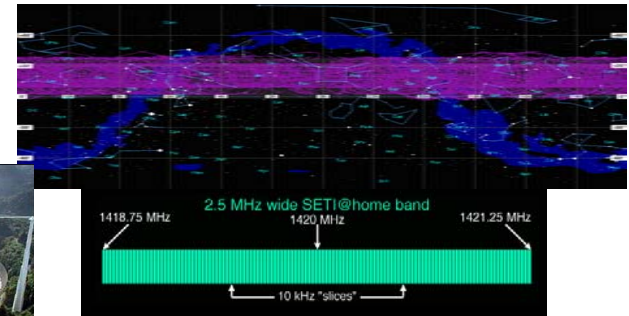
- A virtual supercomputer

<http://www.stanford.edu/group/pandegroup/Cosm/>

<http://members.ud.com/vyyc/cancer/>



Arecibo Radio Telescope



SETI@home

- running on 500,000+ PCs, ~1000 CPU years p/day
- over half a million CPU years so far
- 22 teraflops sustained 24x7

.eduCommons project

www.educcommons.org

- open system for creating, sharing, and reusing educational content and discourse to support people's learning

1. Search for educational resources
2. Download education resources from others
3. Use resources
4. Collaborate with other .eduCommons learners
5. Contribute new material to the eduCommons

the .eduCommons

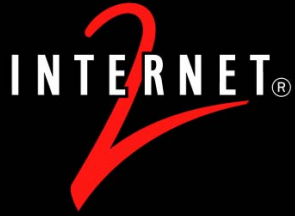
Worldwide Lexicon Project

<http://picto.weblogger.com/>

- open source
- a multilingual dictionary service for the Internet
- a simple, standardized protocol for talking to dictionary, encyclopedia and translation servers – Gnutella for dictionaries
- distributed human computation: enlists the help of internet users who are logged in but not busy.



Kevin Kallaugher -- The Baltimore Sun, The Economist (London) and the International Herald Tribune.



Regarding the copyright infringing argument...a few thoughts...

P2p disrupts traditional distribution mechanisms

Notions of copyright and intellectual property need to be put in a digital-age context (and new business models will need to be developed and implemented)

read Larry Lessig! (www.lessig.org)

Yet, a whole new paradigm in what newer generations want and expect...yet the apathetic approach..(on both sides – consumers and those that can have an effect on things....)

P2P must be disruptive...

Peer-to-peer (p2p): third generation of the Internet

1st generation: “raw” Internet

2nd generation: the Web

3rd generation: making new services to users cheaply and quickly by making use of their PCs as active participants in computing processes

P2P doing this in “disruptive” ways

P2P and web services: converging into single category: effective use of distributed resources

What standards for such network infrastructure in this space:

- Sun (JXTA) and Microsoft (.NET): both aiming at this but very differently

Beyond definition and applications: Content, choice and control

- The user becomes not only a consumer but a content provider P2P allows the end user to participate in the Internet again: original vision where everyone creates as well as consumes

- Control is at the endpoint
- A mindset change on how computing can be accomplished. P2P is representative of bigger ideological changes that are taking place

Many challenges remain for p2p but challenges are also opportunities.

Internet2 and advanced networking community:

- good test bed for basic research on some aspects of p2p; an environment that overcomes the many barriers that are holding the deployment of p2p products in current corporate environments

Resources

- www.openp2p.com (O'Reilly)
- www.peertal.com (news and companies)
- Upcoming at I2 p2p working group <http://www.internet2.edu/p2p>

P2P: Harnessing the Power of Disruptive Technologies,
edited by Andy Oram, O'Reilly Books

<http://www.oreilly.com/catalog/peertopeer/>

2001 P2P Networking Overview

The Emergent P2P Platform of Presence, Identity, and Edge
Resources

By Clay Shirky, Kelly Truelove, Rael Dornfest and Lucas
Gonze

<http://www.oreilly.com/catalog/p2presearch>

¡SI!

¡ GRACIAS !

Ana Preston <apreston@internet2.edu>



INTERNET®

www.internet2.edu