Peer-to-Peer (P2P): What's the big buzz about? A look into P2P and its disruptive but attractive implications –

TERENA Networking Conference - TNC 2003 and CUC2003

Zagreb, Croatia

Ana Preston, Internet2

apreston@internet2.edu

20 May 2003

Today's talk: why care about p2p?

- Defining p2p
  - Literal p2p: Frameworks
  - Pros and Cons; hybrids
- Convergence of p2p and grid computing
- Challenges for p2p
  - Technical and Upper layers
- "p2p" and advanced R&E networking
  - Highlights on U.S. approach to p2p
  - Activities within Internet2
  - Some interesting applications within R&E communities
- Final observations…Is the buzzword dead?

“Literal” P2P: nothing new at all…

- 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)

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

Turn of the century: NAPSTER!

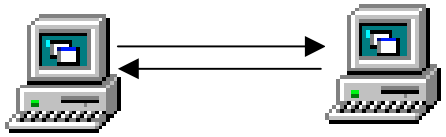
- 1999: Napster is born - universities notice first

2002: Napster is gone - apps continue to emerge

- big rush to find the next big thing in technology: “The next Napster”…. But Napster was more of a social than a technical phenomenon
- 2003: assessing the state of p2p, literally:
  - Defining p2p as computing scheme in which info is stored on many endpoints, reducing/eliminating need of central control
  - Class of apps that takes advantage of resources (e.g., storage, cycles, content, human presence) at the edges of the Internet (from *p2p: harnessing the power of disruptive technologies)*

Basics: Frameworks

- simplest form..



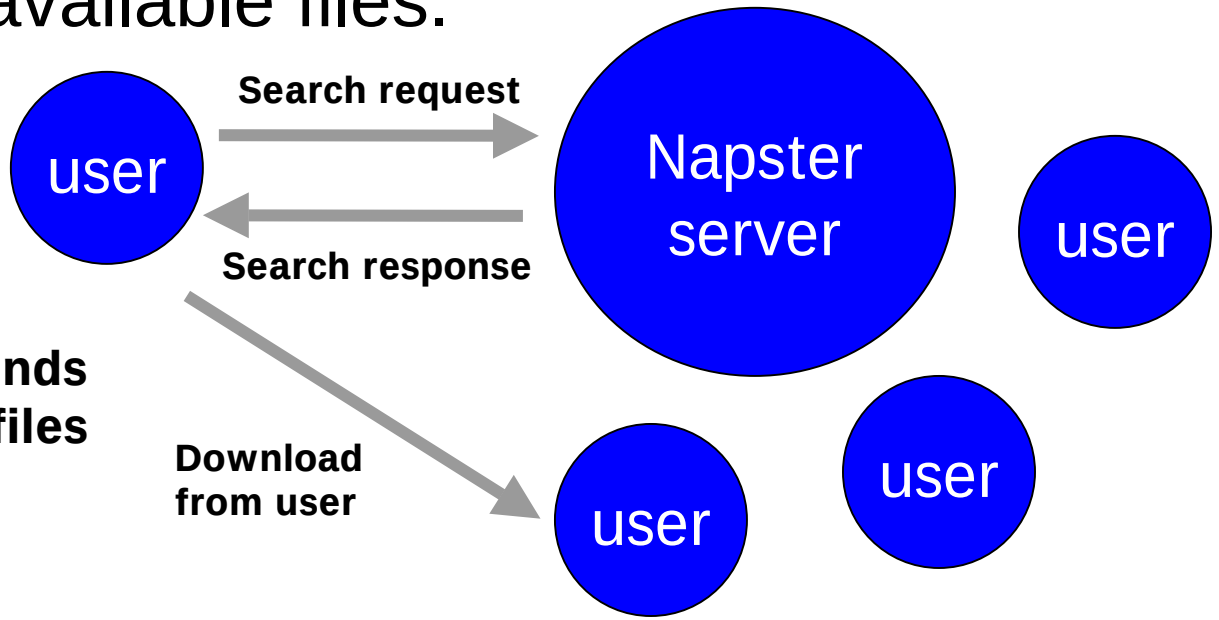
- Models:
  - Centralized - Napster
  - Decentralized – Gnutella, Freenet
  - Controlled-decentralized | Hybrids – Kazaa, Groove
- Components: client, server, servents
- Main differences: how information is shared and how much information is shared
- Pros and Cons

Centralized, e.g., a la Napster

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

1. User sends out request
Napster searches central database

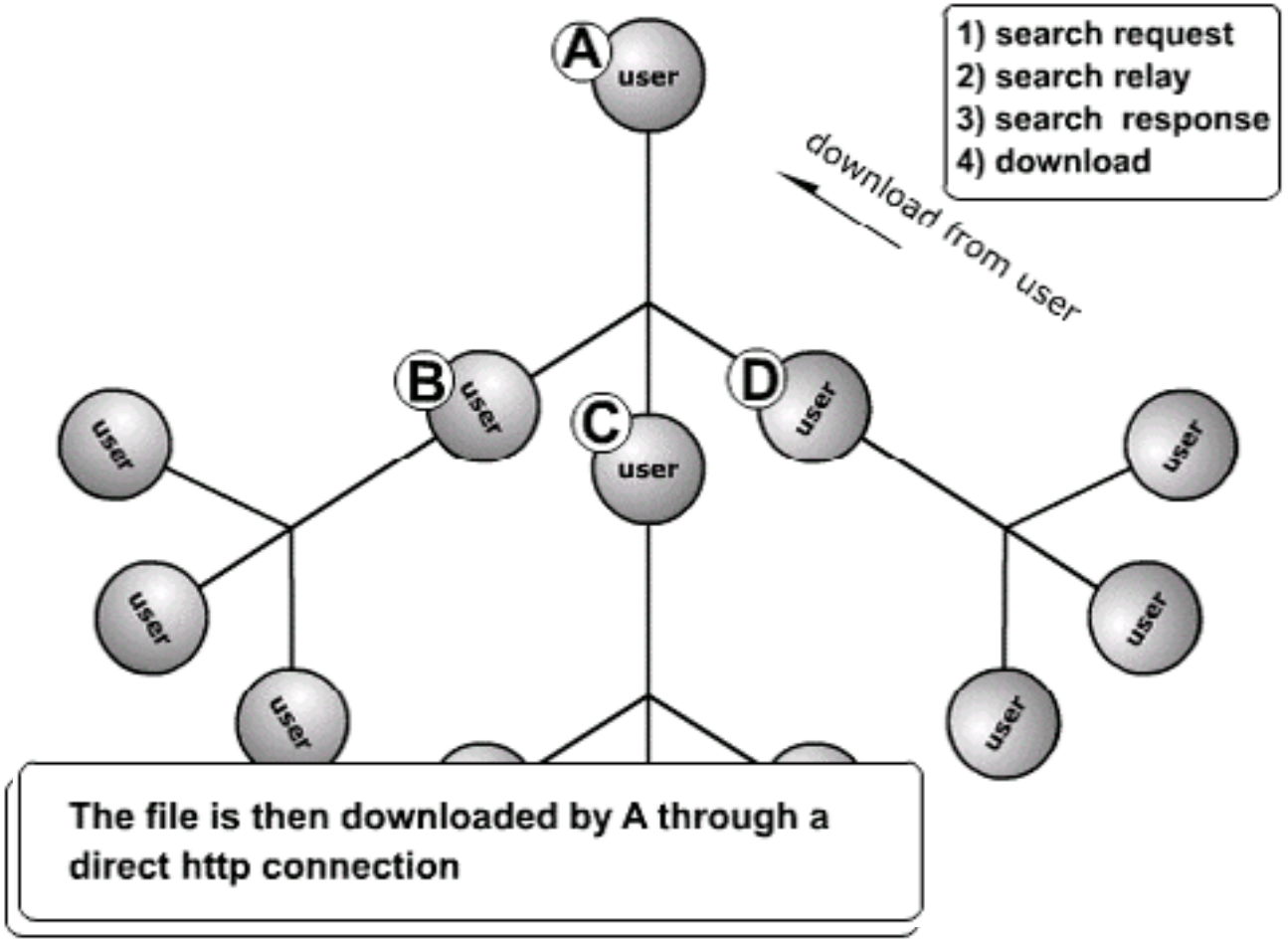
2. The central server sends back a list of available files for download



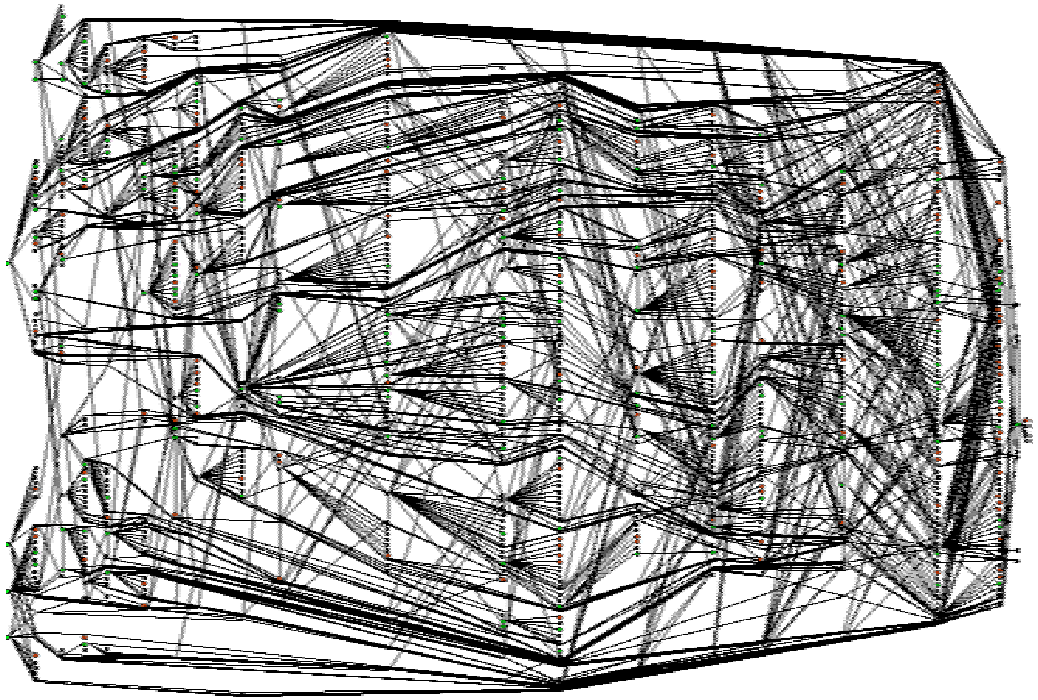
3. Requesting user downloads the file directly from another Napster user computer

Decentralized, e.g., a la Gnutella

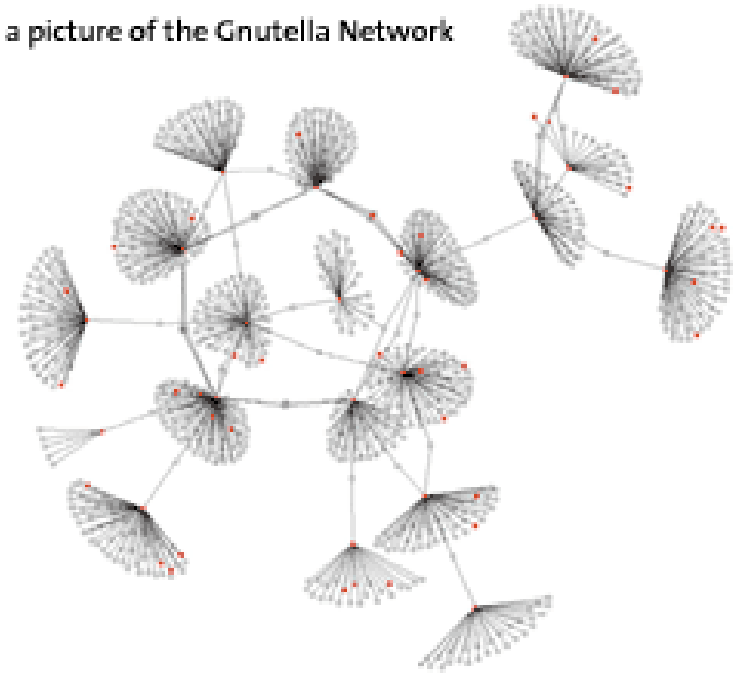
Source: www.limewire.com



Looking at Gnutella network



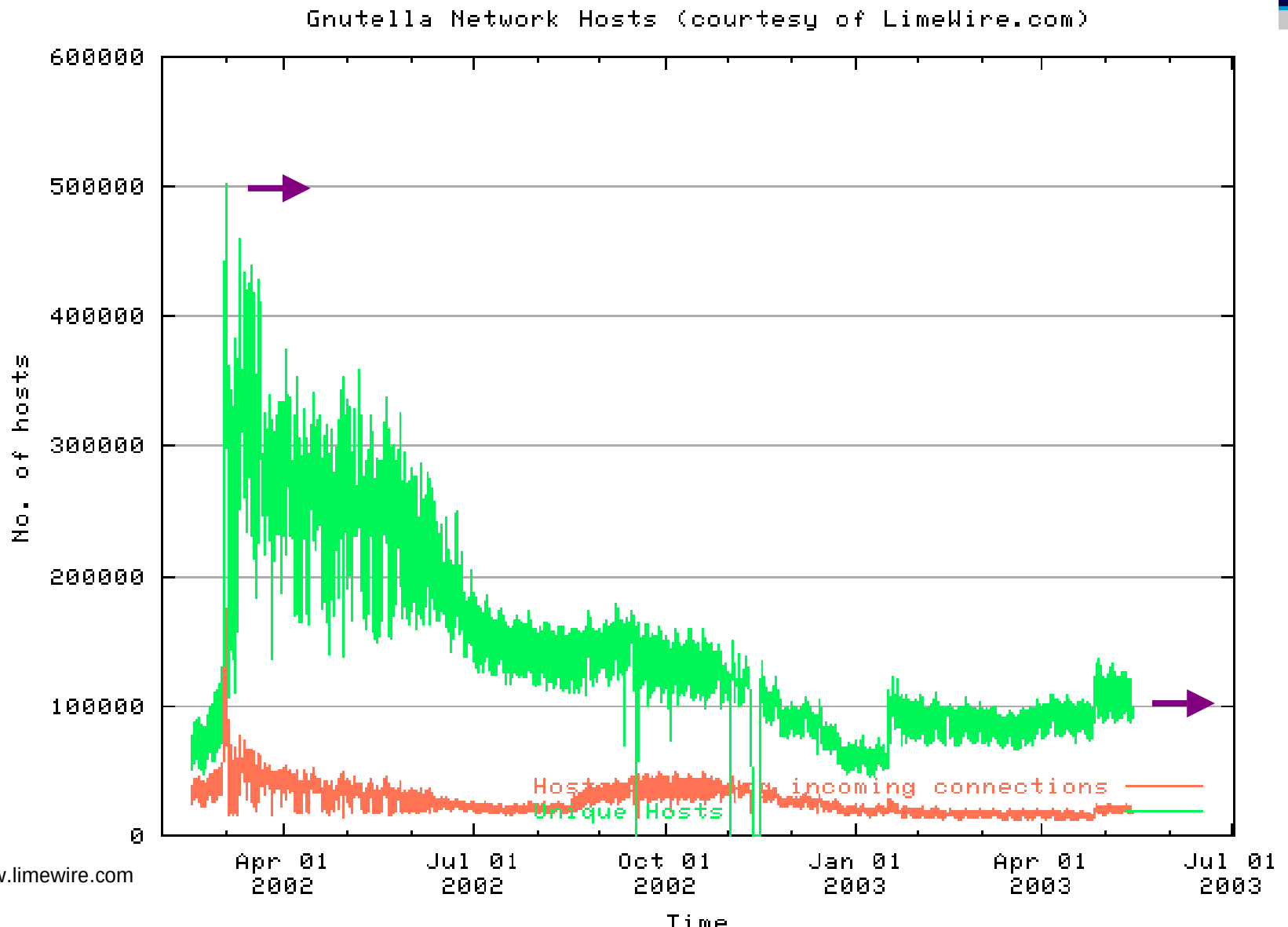
Partial Map of the Gnutella Network
http://dss.clip2.com



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

Gnutella Host Count (historical) - 2003



source:
http://www.limewire.com

Another decentralized 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

http://www.freenet.org

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.

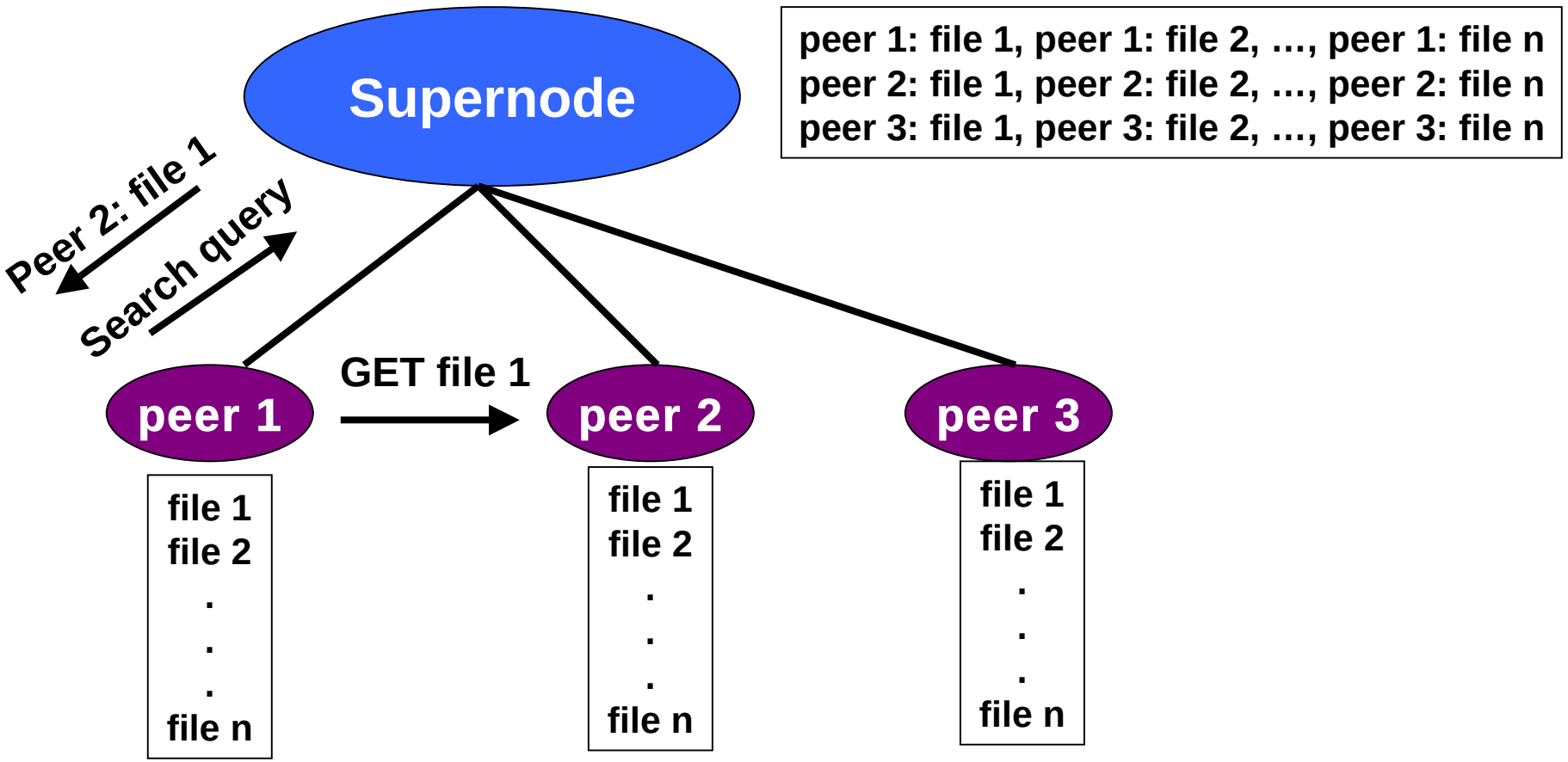
  Cons:

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

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 would locate another peer sharing same requested file, and automatically resume downloads where it left off at failed host.

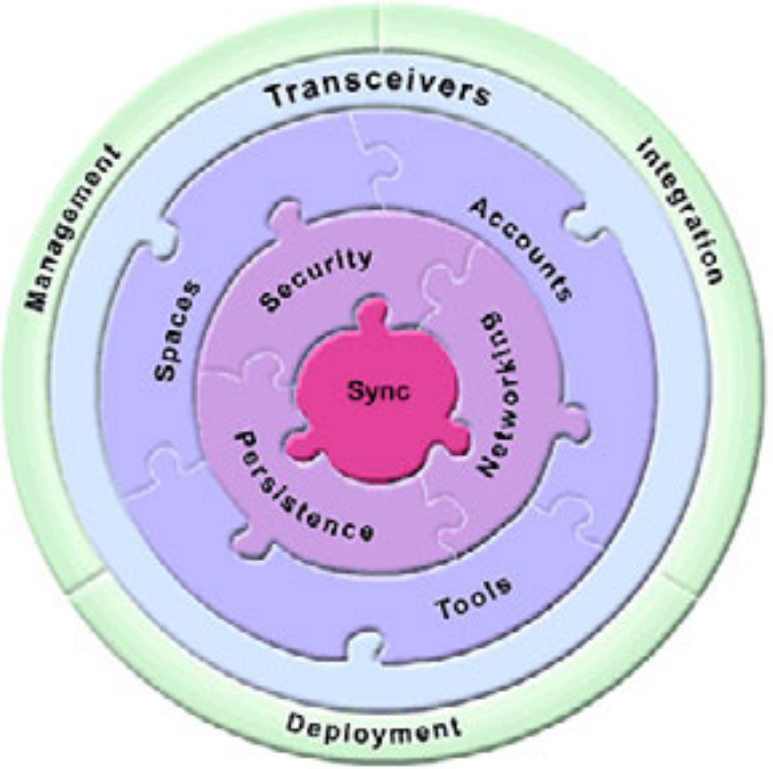
More on Morpheus



Source: *Morpheus Out of the UnderWorld* by Kelly Truelove
http://www.openp2p.com/pub/a/p2p/2001/07/02/morpheus.html

Another example: Groove

- 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/

Key application/function areas

- 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, Onobee
- Distributed Computation:
  - Use under utilized Internet and/or network resources for improving computation and data analysis (SETI@home, United Devices, Intel p2p program, entropia)
- P2P groupware / Collaboration
  - Cooperative publishing, messaging, group project management. Secure environments are offered in some products (Grove, Engenia, Onobee)
- Development Frameworks; Development tools and suites
  - Project JXTA (Sun), .NET (Microsoft)
- Recommend: O'Reilly p2p directory: companies, projects and initiatives
  http://www.openp2p.com/pub/q/p2p_category

[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	

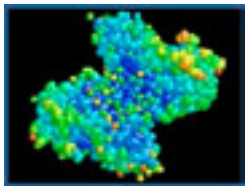
Towards apps and functionality

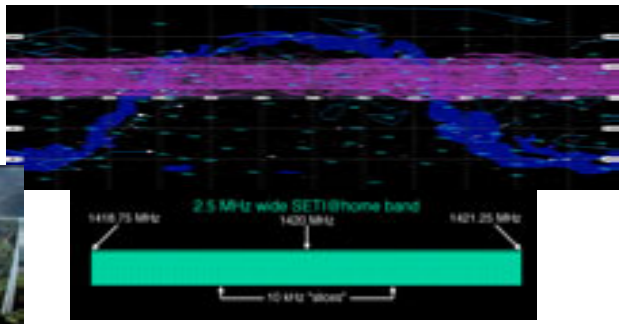
- Endpoints on the Internet exchange information and form communities
- collaboration and workflow opportunities
- Edge resources (content, storage, computing power, bandwidth, human attention…)
- how these can be better utilized and accessible in real time (presence) and
- “each” with their own “identity” (autonomy at the edge)

Distributed Computing

- 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
- Applications that harness the power of the network at the edges

folding@home





SETI@home

P2P and Grid Computing

On Death, Taxes, and the Convergence of peer-to-peer and grid computing
Ian Foster (Univ. of Chicago), Adriana Iamnitchi (Argonne National Laboratory)

http://people.cs.uchicago.edu/~anda/papers/foster_grid_vs_p2p.pdf

Important commonalities:

- Pooling and coordinated use of resources within distributed (virtual) communities
- constructed as overlay structures that coexist, but need not correspond in structure to underlying organizational structures.

Definitions: P2P and Grid Computing

P2P	Grids
Class of apps that takes advantage of resources (e.g., storage, cycles, content, human presence) at the edges of the Internet.	sharing environments that are implemented via the deployment of a persistent, standards-based service infrastructure that supports the creation of, and resource sharing within,distributed communities. Resources (e.g., computers, storage, sensors, SW apps, and data) are connected through the Internet and are owned by various administrative organizations and shared under locally defined policies. - Virtual organizations (VOs) can be a set of individuals and/or institutions defined by such rules. - Middleware software to provides basic services for security, monitoring, resource management, etc.

P2P and Grid Computing – cont.

P2P	Grids
Communities develop rapidly via unsophisticated but popular services Unstable environment drive design requirements Communities may not act cooperatively •few providers many consumers •highly variable behavior Integration of simple resources via protocols designed to provide specific vertically integrated functionality. •No simple standard service infrastructure; little adoption and no interoperability •Promising: JXTA, XtremWeb and BOINC	relatively sophisticated services and apps, connecting small number of sites into collaborations engaged in complex scientific applications. •Scaling: as system scale increases, have to think on autonomic configuration and management issues. •Commercial/modern grids aiming at creating and operating required infrastructure while dealing with issues around trust, accountability, etc. earlier grid communities had to deal less with this (b/c of self-contained nature) •Resource availability tends to be higher and more uniform (more powerful – more resource intensive - and perhaps regulated; better connected) •Core base/architecture upon which all services adhere to a set of standard interfaces and behaviors - Open Grid services architecture

Many challenges for p2p remain

- Architectural:
  - What scales? What architecture to preserve end to end?
- Peer/Resource Discovery:
  - Finding things in p2p systems gets harder when not dealing with widely replicated content (e.g., MP3s)
  - Pressing importance on search architectures and metadata
- Resource Management -> bandwidth:
  - “Bandwidth and continuously upgrading network capacity is not a long-term viable solution; need better “network awareness”
- Security/Trust:
  - 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?
  - Vulnerabilities, worms, viruses…
  - Encryption mechanisms, sandboxing, PKIish…Other ways

P2P challenges, cont.

- 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
- Upper layers: copyright issues
  - At US universities: no clear consensus on how to approach their responsibility with US copyright law
  - Agreement 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)

P2P at universities: technical/policy mechanisms

- blocking vs. not blocking; user education and cooperation
- Metering, bandwith limiting, and allocation fair use of resources:
  - per-user or per-subnet rate-limiting
  - special purpose middleboxes deployed at the campus edge -- "QoS appliances" support a range of functionality, BUT:
    - How to REALLY 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? What is non-desirable?
    - But quickly changing characteristics of non-desirable traffic and ability and cost associated with the products trying to keep up with it …
- Thinking about models where beyond just adding capacity:
  - fees are implemented to cover bw costs; usage-based accounting, charge-back, quotas, or other policy mechanisms

Have added many headaches (cost, operation, legal)

Internet2 and p2p

- Panels and BoFs at several Internet2 and Higher Ed conferences
- Internet2 QoS Working Group on Campus Bandwidth Management
- Collaborative Computing in Higher Education: Peer-to-Peer and Beyond (January, 2002, Arizona)
  - aimed at exploring the technical and future dimensions of the fast-growing P2P services spaces (beyond file sharing) and the opportunities and challenges presented for universities.
  - http://events.internet2.edu/agenda.php?session_event_id=94
- Internet2 p2p working group – http://p2p.internet2.edu

Internet2 p2p working group http://p2p.internet2.edu

- Under Internet2's end-to-end performance initiative
- Applications for the R&E communities (identification and cataloging)
- Security
- Resource management:
  - http://cbm.internet2.edu ; jointly with QoS wkg group, focuses on the need to for QoS, traffic routing, campus administration of local domains, and overall bandwidth management
- Information Resources: archive of documents, presentations, FAQs, meetings, and other materials of interest
  - See pages for joining mailing list archives at http://listserv.utk.edu/archives/p2p.html

END-TO-END PERFORMANCE INITIATIVE
>Home
>About Us
>Staff
>Contact

RESOURCES
>Projects
>Tools
>Library
>Presentations
>Glossary
>Forum
>Notebook

WORKING GROUPS
>Peer-to-Peer Working Group
>>Applications
>>Security
>>Rsc Mgt
>>Info Rscs

Peer-to-Peer Applications

Here is a list of P2P Applications for the research and education communities. This is an initial attempt to collect and identify such applications. So, please the form below to submit any P2P applications we may have missed in this list.

Research Applications

- Intel® Philanthropic Peer-to-Peer Program [http://www.intel.com/cure/] -- "Intel's Philanthropic peer-to-peer program links millions of PCs to crea computing resource for researchers in a wide variety of fields. For example, Leland Stanford Junior University is focusing on protein folding and rela diseases such as Alzheimer's, ALS and Parkinson's."
- SETI@home [http://www.seti.org/science/setiathome.html] -- "The SETI@home project links and uses donated computer processing capacity to a data collected from a radio telescope located in Puerto Rico."
- Worldwide Lexicon Project [http://picto.weblogger.com/] -- "The worldwide lexicon project is an open source initiative to create a multilingual dictio service for the Internet, and to create a simple, standardized protocol for talking to dictionary, encyclopedia and translation servers throughout the Similar to the SETI@Home project. While SETI@Home taps the idle CPUs of millions of personal computers, the worldwide lexicon enlists the help internet users who are logged in, but not busy. Think of this as distributed human computation."

Educational Applications

- eduCommons [http://www.educommons.org/]-- The eduCommons is an open system for creating, sharing, and reusing educational content and disc to support people's learning. [more info]
- Edutella [http://edutella.jxta.org/] -- "Edutella is a peer-to-peer service for the exchange of educational metadata. Edutella lives on top of the Sema Web framework as a distributed query and search service."

General Applications

- Chord Project [http://www.pdos.lcs.mit.edu/chord/] -- "The Chord project aims to build scalable, robust distributed systems using peer-to-peer ide

Using p2p in R&E community

- General Applications
  - Chord Project
  - Groove Networks
  - JXTA
  - LOCKSS
  - The Metadata3 Project
  - OceanStore Project
  - Piazza
  - Publius
  - Tapestry
- Research Applications
  - Intel® Philanthropic Peer-to-Peer Program
  - SETI@home
  - Worldwide Lexicon Project
- Educational Applications
  - eduCommons
  - Edutella
  - Lionshare

Please send information of other peer-to-peer applications to George Brett
ghb@internet2.edu

Development / testing of apps

LIONSHARE - http://lionshare.sourceforge.net

- An academic p2p system for distribution of academic materials within university for collaboration among faculty, students, departments (right now Pennsylvania State University) but eventually multiple universities
- in process of adding authorization – desirable to make a federated p2p application between the Internet2 member community and outside users - a "shib'd" application

Onobee (p2p multi-collaborative tool)

- Initial testing
- created a multi-peer environment on an ad hoc basis.
- Working on developing guidelines and testing methodologies

Final observations

- 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
- 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

Final observations – cont.

- 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; a whole new paradigm in what newer generations want and expect
- Many challenges remain for p2p but challenges are also opportunities.

More information

- Internet2 p2p working group, see http://p2p.internet2.edu
- Resources
  - www.openp2p.com (O'Reilly)
  - www.peertal.com (news and companies)
- *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
- *On Death, Taxes, and the Convergence of Peer-to-Peer and Grid Computing*, Ian Foster and Adriana Iamnitchi, University of Chicago and Argonne National Laboratory
  http://people.cs.uchicago.edu/~anda/papers/foster_grid_vs_p2p.pdf

The end…

Questions?

My contact information:

Ana Preston <apreston@internet2.edu>


INTERNET®
www.internet2.edu