Grid Computing, 7700

Guest!

Dr Ian Taylor
itaylor@lsu.edu

Louisiana State University
LSU
P2P and Grids

1. Grid computing
   a) Globus
   b) Service-based Grid computing
   c) Grids and P2P

2. What is P2P
   a) Why P2P - history
   b) P2P definition
   c) Gnutella
   d) Scalability
Grid Computing: Globus Tools GT2

Recap: consists of four elements:

- **Resource Management**: to allocate resources provided by a Grid - GRAM
- **Data Management**: involves accessing and managing data – GridFTP, GASS
- **Information Services**: to provide information about Grid services - MDS

And of course:

- **Security**: to provide authentication, delegation and authorization
Get To Know Your Grid


• “The Grid is flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources

• The concept of Virtual Organisations
A Globus Grid

Users/Clients

Internet Routing

Users/Clients

Internet Routing

GSI

X. 509

VO

VO

Resources

Middleware (Globus)

Single Sign-on

GRAM

GridFTP

MDS

MDS

MDS

Mutual Authentication

Resources

VO
Web Service

Client

XML

Web Service Interface
(WSDL)

(SOAP) = XML + Envelope

Server
OGSA + WSRF

- Open services Grid Architecture (OGSA)
  - Web services → Grid services = "a Web service that provides a set of well-defined interfaces and that follows specific conventions"

The implementation:

WSRF - Web Services Resource Framework
  - Adding (stateful) resources to Web services
Globus V.4

Users/Clients

Internet Routing

Web Services (WSRF)

GSI

X.509

VO

Resources

Middleware (Globus)

Single Sign-on

Mutual Authentication

GRAM

GridFTP

MDS

VO
P2P Interface (XML?)

Jxta Pipes

Super Peers?

Rendezvous Nodes

Gnuella

Grouping Mechanisms

Resources

Groups

GSI?

X. 509

Authorization?

Middleware

Jxta, P2PS,
P2P and Grids?

Are the converging?

- Yes!
- **Scalability:** P2P has address scalable networks
- **Decentralized:** P2P Super peer nets
  - Many architecture are moving this way
- **File Services:** Bittorrent, GRID Torrent??
- Etc …

- Ok, so what’s P2P
What’s Exciting?

• 0.5 Billion currently ‘Connected Devices’
  - With a CPU capability more than 100 times that of an early 1990s supercomputer
  - Gartner Group - 90% of CPU power is wasted
  - Mobile Devices - 1 billion currently, estimated 1.5 billion within 5 years. Capability is increasing
  - Potential demonstrated by SETI@Home - so far used 1 million years of CPU time
  - Feb2003: press release: United Devices are using their meta-processor to help US DoD to find a cure for smallpox
  - Leveraging previously unused resources

• P2P research is concerned in addressing some of the key difficulties of current distributed computing:
  - scalability
  - reliability
  - interoperability.
Historical P2P

• Peer to Peer (P2P) - originally used to describe the communication of two peers.

• The internet started as peer to peer system e.g. ARPANET
  • goal - to share computing resources around the USA using different networks
  • UCLA, SRI, Utah and Santa Barabara
  • all had equal status – P2P

• From late 1960’s until 1994, machines were assumed to be switch on, connected and had an IP address assigned

• Then, invention of Mosaic and WWW led to a different type of user….  

![Diagram](image)

- Clients
  - dial-up modems
  - IP addresses changing
  - unpredictable

The 1990's Client-Server Internet

Late 1990’s Naptser, then Gnutella 2000 – the new P2P
Modern Peer to Peer

What is an P2P application?

P2P is a class of applications that takes advantage of resources e.g. storage, cycles, content, human presence, available at the edges of the Internet – Clay Shirky

Computers/devices “at the edges of the internet” are those:

• Operating within transient environments - computers come and go frequently
• They can be behind a firewall or NAT systems
• Have to operate outside of DNS
• Often have to deal with differing transport protocols, devices and operating systems
A P2P Network

Application

P2P Network

NAT

Firewall

TCP/IP

Bluetooth

TCP/IP

HTTP

XP

Linux

TCP/IP

Bluetooth

TCP/IP

14
Example 2: File Sharing with Napster

- Launched in May 1999, by Shawn Fanning (19) and Sean Parker (20)
- Allowed Users to MP3 Files - compression format, good quality but 1/12th original size
- April 2000 - Metallica starts law suit - Huge and long court case
- November 2000 - Napster has 38 Million members
- July 2001 - Napster ordered offline, June 2002 bankrupt

1. Construct Database
   - Users connect to Napster Server
   - Server builds up a list of available songs and locations

2. User A searches for song.mp3

3. Server searches database. Finds song on User C's machine

4. Server informs User A of the location of song.mp3

5. User A connects to User C and downloads song.mp3
The ‘Animal’ GNU: Either of two large African antelopes (*Connochaetes gnou* or *C. taurinus*) having a drooping mane and beard, a long tufted tail, and curved horns in both sexes. Also called *wildebeest*.

Gnutella = GNU + Nutella

GNU: Recursive Acronym GNU’s Not Unix ....

Nutella: a hazelnut chocolate spread produced by the Italian confectioner *Ferrero* ....
History Of Gnutella

Gnutella, GNU GPL, 0.56 (Feb 2000)

Open Source Developers
- gnutella.nerdherd.net
  - Bryan Mayland

Gnutella Spec
- #gnutella

Justin Frankel + Tom Pepper

Gnullsoft

NullSoft

AOL
We are Geeky and Rich!

And now I am trying to be cool ... and writing songs ...
What is Gnutella?

Gnutella is a protocol for **distributed search**

- peer-to-peer comms
- decentralized model
- No third party lookup

Two stages:

1. Join Network ... later
2. Use Network
   1. Discover other peers
   2. Search other peers
Searching in Gnutella involves broadcasting a Query message to all connected peers. Each connected peer will send it to their connected peers (say 3) and so on. Typically, this search will run 7 hops. If the number of connected peers, \( c = 3 \) and the hops i.e. TTL=7 then the total number of peers searched (in a fully connected network) will be:

\[
S = c + c^2 + c^3 + \ldots \quad \text{ch} = 3 + 9 + 27 + 81 + 243 + 729 + 2187 = 3279 \text{ Nodes}
\]
Searching a Gnutella Network: From one Node
Searching a Gnutella Network: All nodes
Social Networks

• Stanley Milgram (Harvard professor) - 1967 social networking experiment
• How many 'social hops' would it take for messages to traverse through the US population (200 million)

- Posted 160 letters randomly chosen people in Omaha, Nebraska
- Asked them to try to pass these letters to a stockbroker working in Boston, Massachusetts
- Rules:
  - use intermediacies whom they know on a first name basis
  - chosen intelligently
  - make a note at each hop
- 42 letters made it !!
- Average of 5.5 hops
- Demonstrated the 'small world effect'

Proved that the social network of the United States is indeed connected with a path-length (number of hops) of around 6 - The 6 degrees of separation!

Does this mean that it takes 6 hops to traverse 200 million people??
Lessons Learned from Milgrim’s Experiment

- Social circles are highly clustered
- A few members have wide-ranging connections
  - these form a bridge between far-flung social clusters
  - this bridging plays a critical role in bringing the network closer together

For example

- A quarter of all letters passed through a local storekeeper
- A half were mediated by just 3 people

Lessons Learned

- These people acted as gateways or hubs between the source and the wider world
- A small number of bridges dramatically reduces the number of hops
From Social Networks to Computer Networks...

- There are a number of similarities to social networks
  - People = peers
  - Intermediaries = Hubs, Gateways or Rendezvous Nodes (JXTA speak...)
  - Number of intermediaries passed through = number of hops

Are P2P Networks Special then?

- P2P networks are more like social networks than other types of computer network because they are often:
  - Self Organizing
  - Ad-Hoc
  - Employ clustering techniques based on prior interactions (like we form relationships)
  - Decentralized discovery and communication (like we form neighbourhoods, villages, cities etc)
Decentralized

- Gnutella
- Freenet
- Internet routing
Centralized + Decentralized

- New Wave of P2P
- Clip2 Gnutella Reflector (next)
- FastTrack
  - KaZaA
  - Morpheus
- Email
- Like Social Networks perhaps?
The Gnutella Network Today

The figure below is a view of the topology of a Gnutella network as shown on the LimeWire web site, the popular Gnutella file-sharing client. Notice how the power-law or centralized-decentralized structure is demonstrated.