ENSC 427 Final Project Presentation

P2P Gnutella analysis

Term: Spring 2013

Group #6

Members: Mark Ma, Daniel Frigo, Tauseef Alwaris

Overview

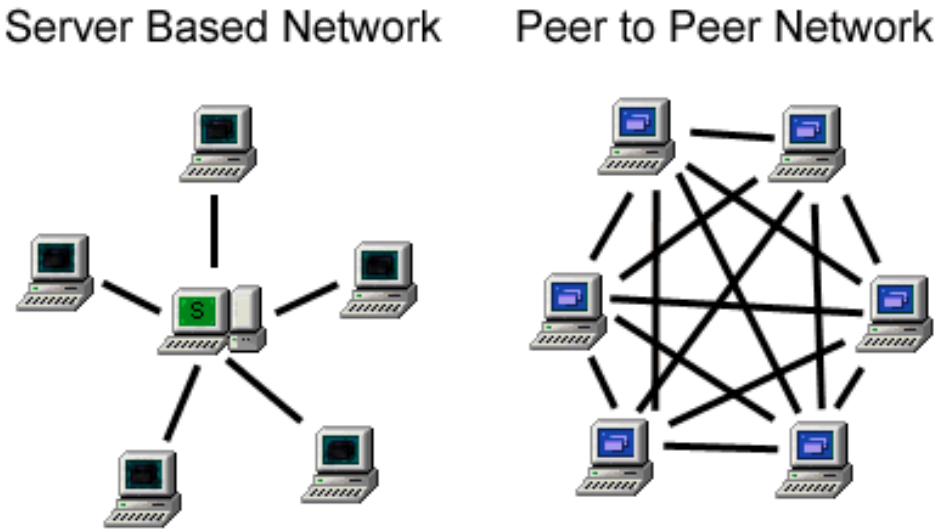
- Introduction
- Architecture
- Network Topology
- Results & analysis
- Conclusion
- Evolution in future
- References

Introduction

What is a peer to peer network?

- A network that enables users to access/share files and information from a multitude of other users rather than relying on a dedicated server
- Enable computers to act as both the client and a server rather than one of the two as seen in the traditional client based server applications

P2P Architecture



P2P Advantages

- Independent of Central server
- Easiest way of file sharing on global scale
- Clients can provide and manage resources
- Smaller setup and costs
- Capable of synchronizing with new network technologies
- More resources available due to more users

Gnutella, A P2P Network

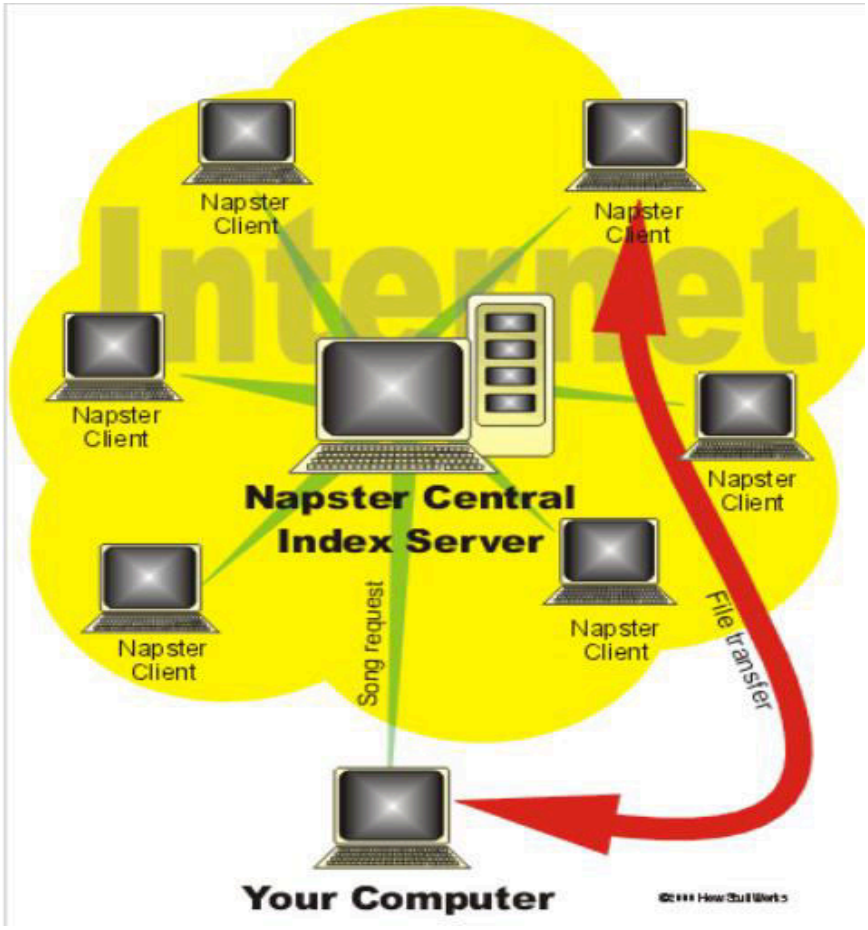
- Largest P2P network primarily used for file sharing
- Each computer becomes a network node with no central server
- Developed in early 2000, by Justin Frankel, Giancula Rubinacci and Tom Pepper
- Evolved from old P2P Napster
- Nowadays word“Gnutella” used to refer to open network protocol for file distribution

Gnutella vs Napster

- In Napster, users connecting to the network tell central server about files available for sharing
- to download a file the user would query the server which in turn would look into its database to see which (if any) machines had that file available for downloading

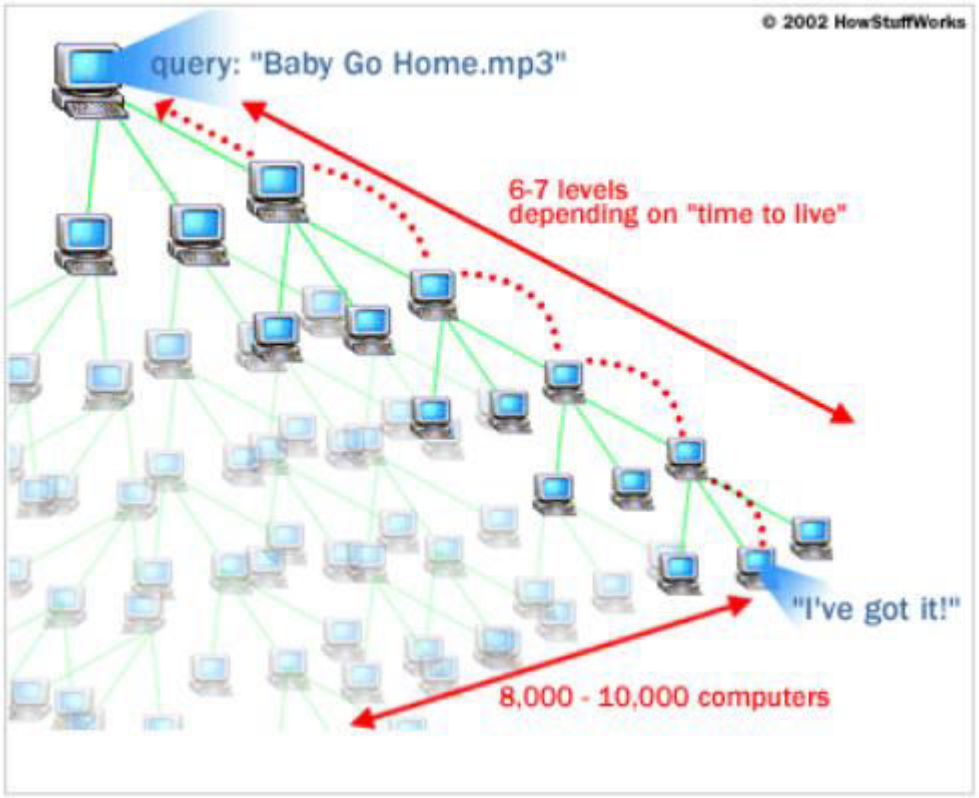
Gnutella vs Napster

Napster Architecture



Gnutella vs Napster

Gnutella Architecture



Gnutella Protocol Mechanism

- protocol defines the way in which servents communicate over the network
- consists of a set of descriptors used for exchanging data between servents and a set of rules governing the inter-servent exchange of descriptors
- Servents are defined as nodes that act as both servers and clients

Gnutella Protocol Mechanism

- Protocol defines the following 5 descriptors:

1. Ping
2. Pong
3. Query
4. QueryHit
5. Push

5 Descriptors

Ping:

- Used to actively discover hosts on the network
- A servent receiving a Ping descriptor is expected to respond with one or more Pong descriptors

5 Descriptors

Pong:

- The response to a Ping
- Includes the address of servent and the info regarding the amount of data available to network

5 Descriptors

Query:

- primary mechanism for searching
- Servent recieving query descriptor responds with a QueryHit if a match is found against its local data set

5 Descriptors

QueryHits:

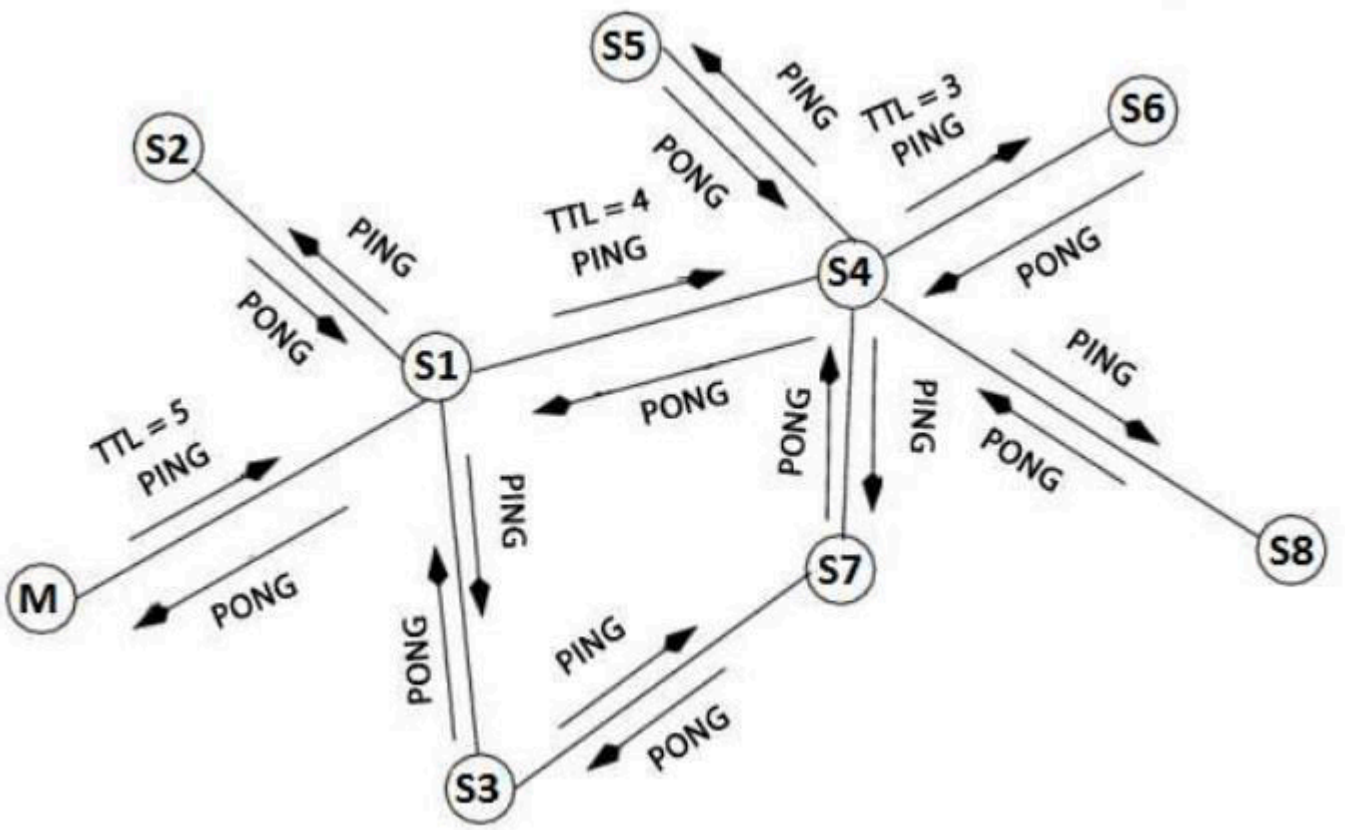
- The response to a Query
- Provides the recipient with enough info to acquire data matching the Query

5 Descriptors

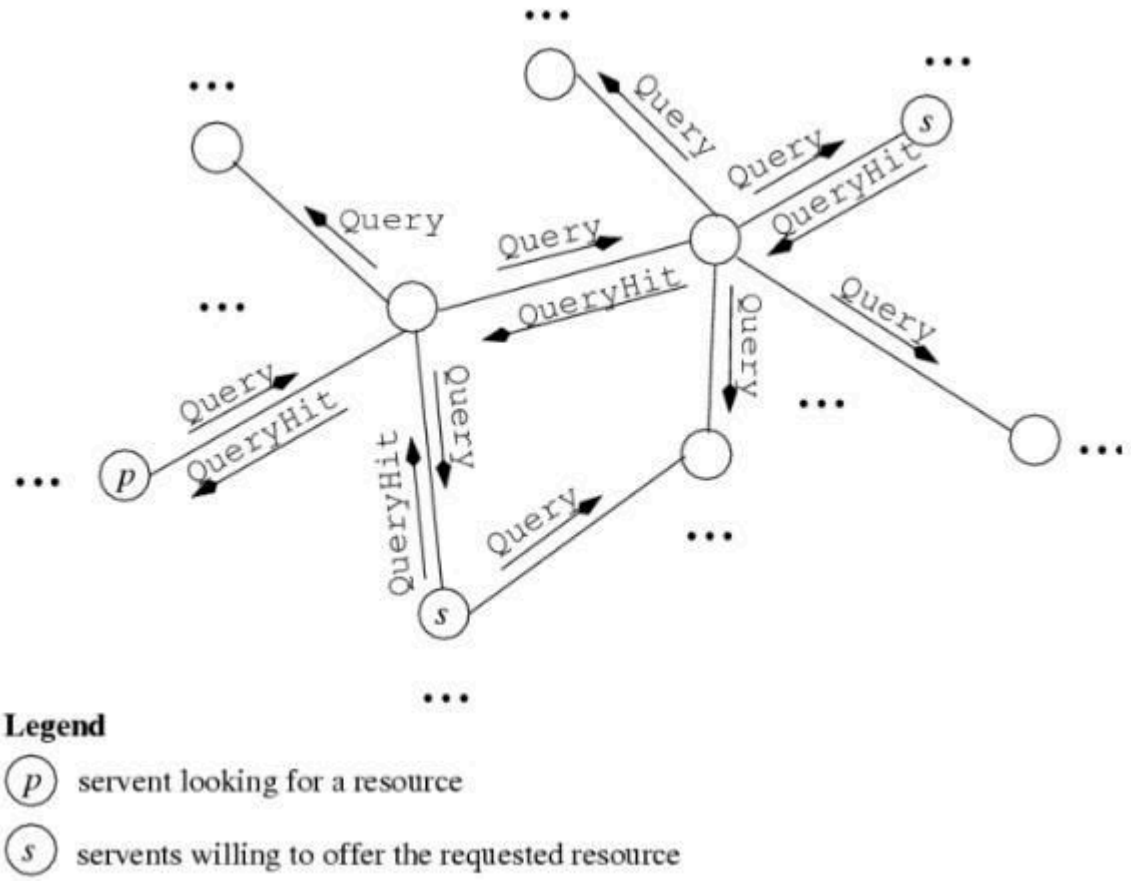
Push:

- Allows firewalled servent to contribute file based data to the network

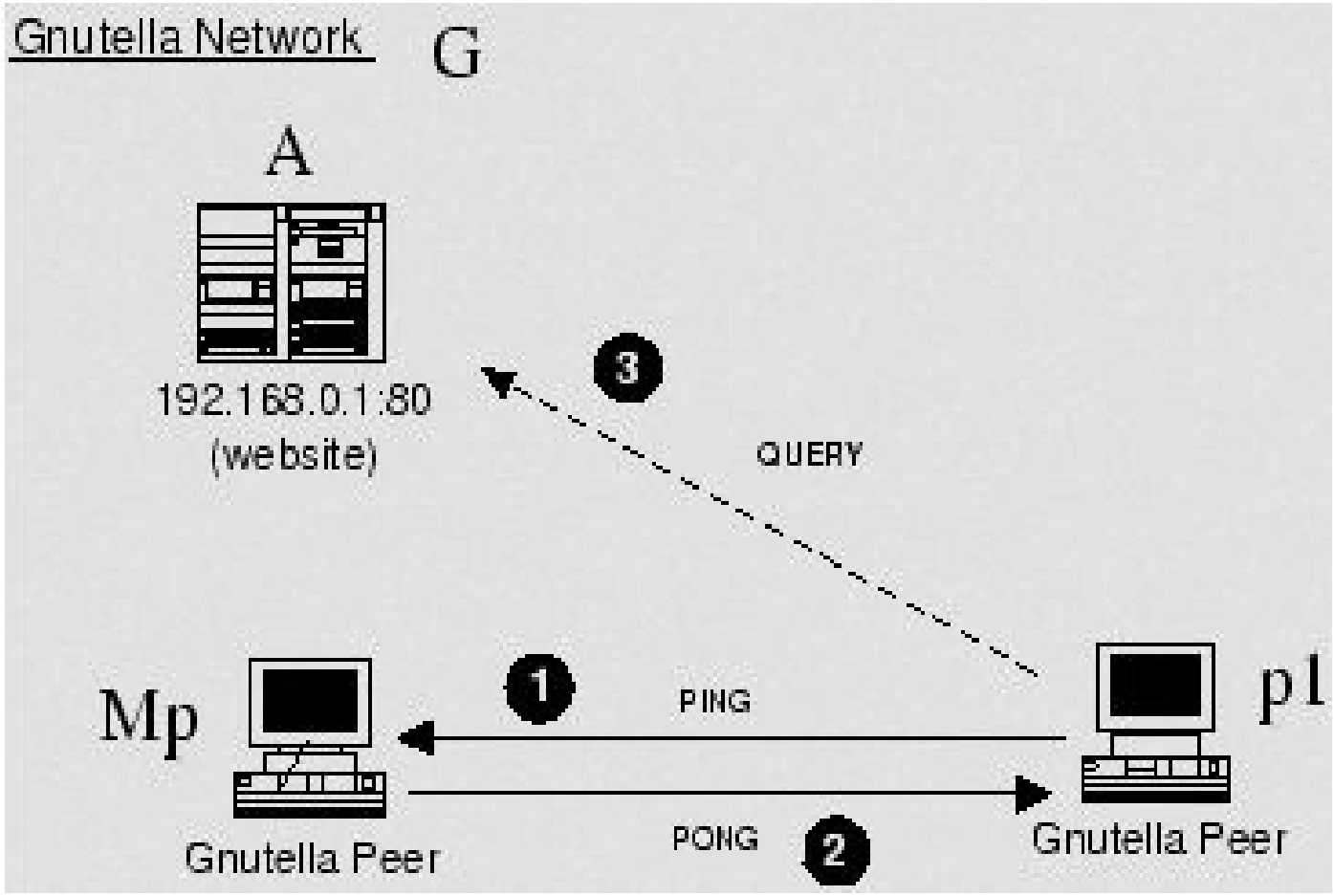
Ping Pong Mechanism



Query /QueryHit Mechanism

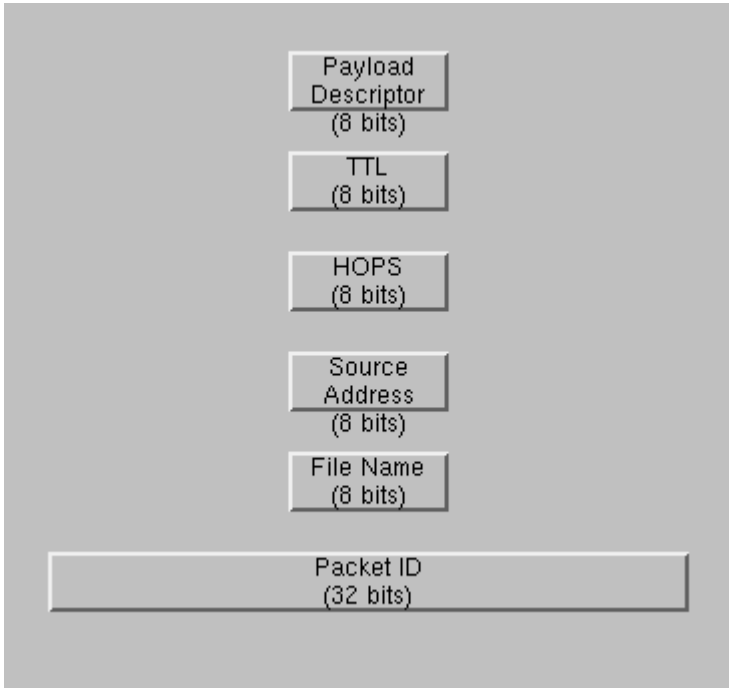


Overall Descriptors Mechanism



Network Topology

Packet Format



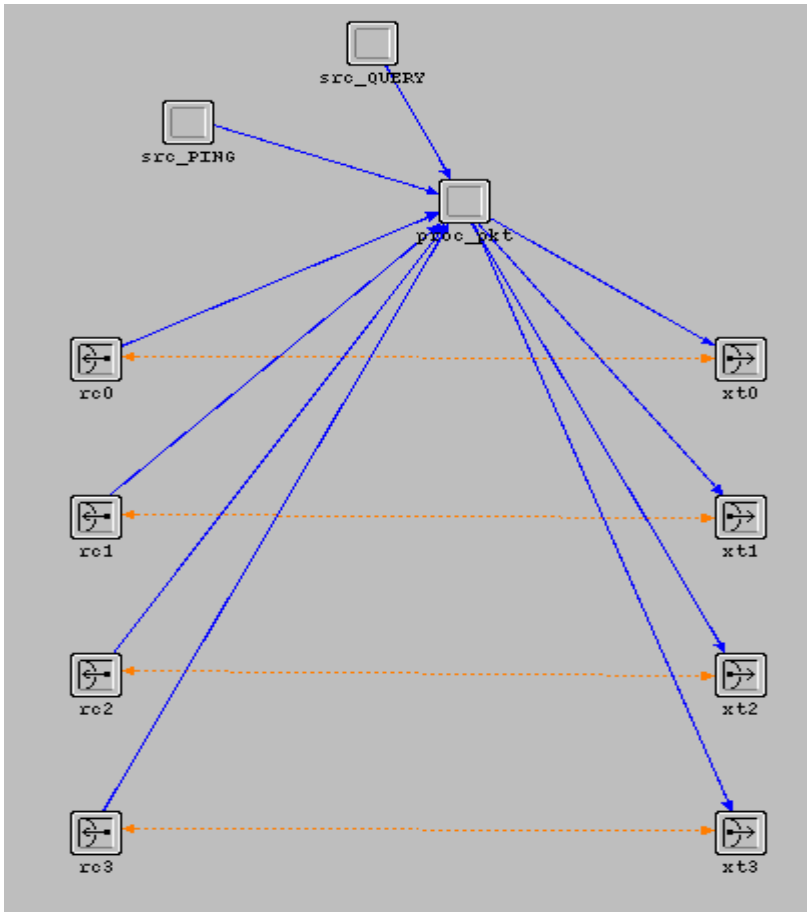
Network Topology

Link Model

- NOT SURE WHAT TO PUT HERE

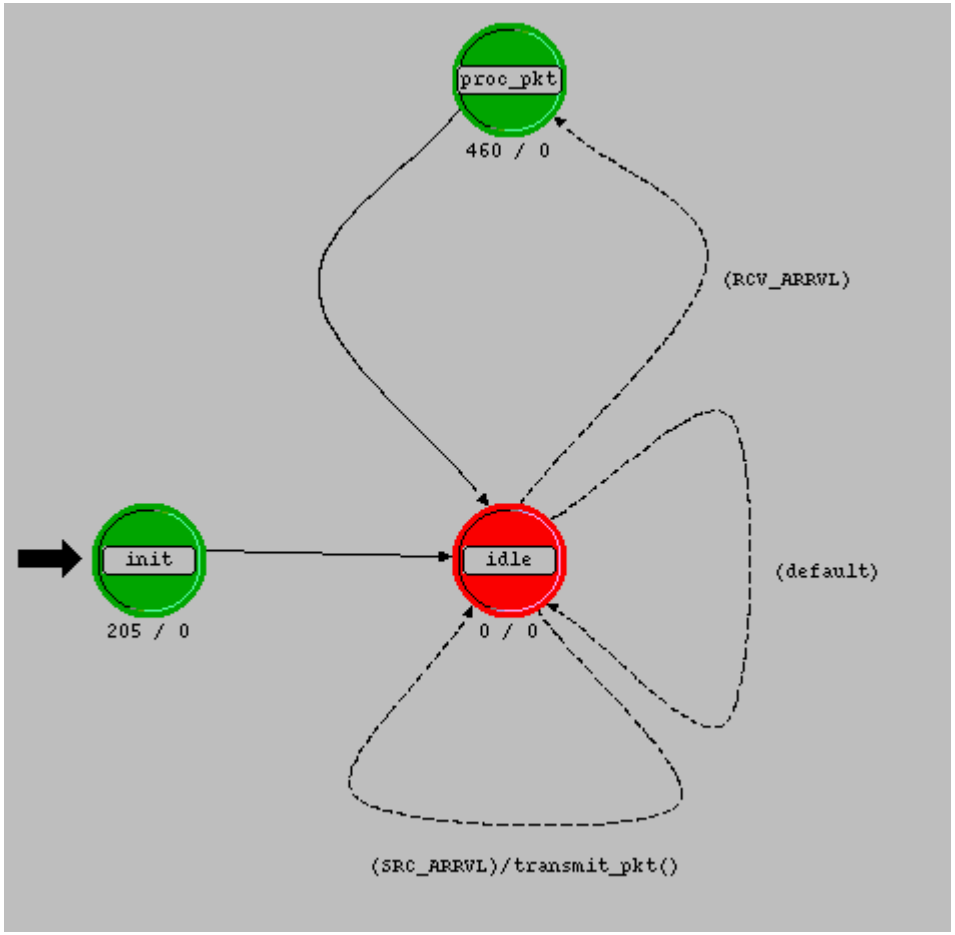
Network Topology

Node Model



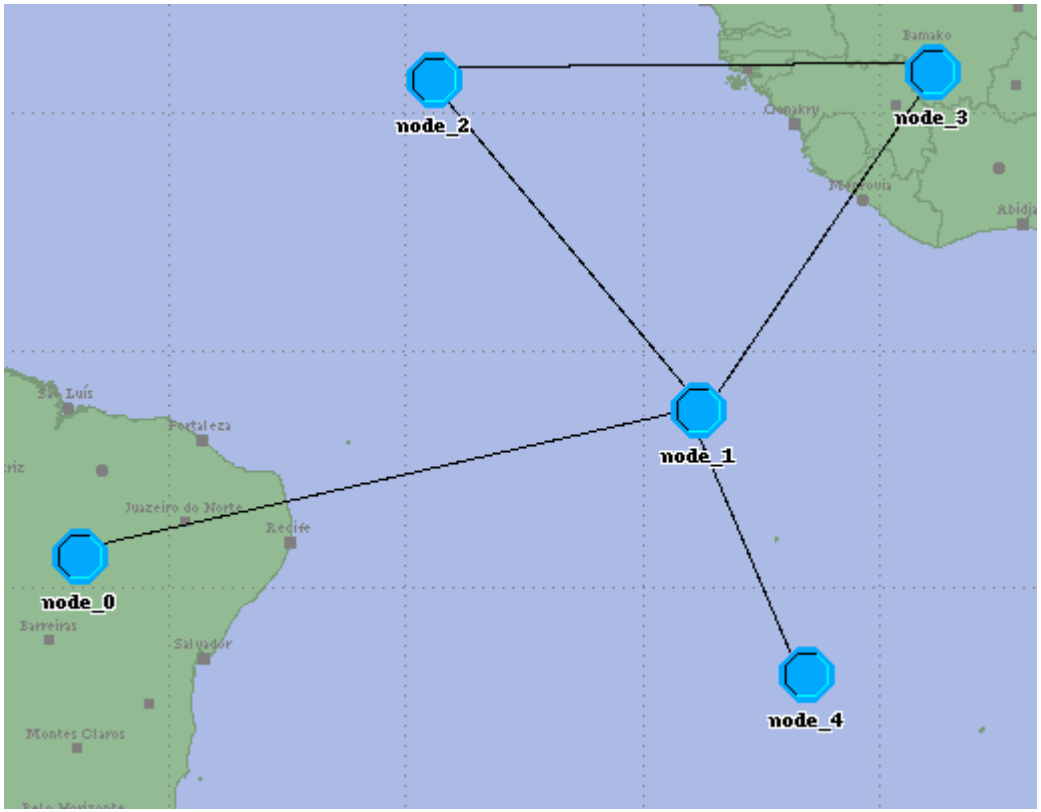
Network Topology

Process Model



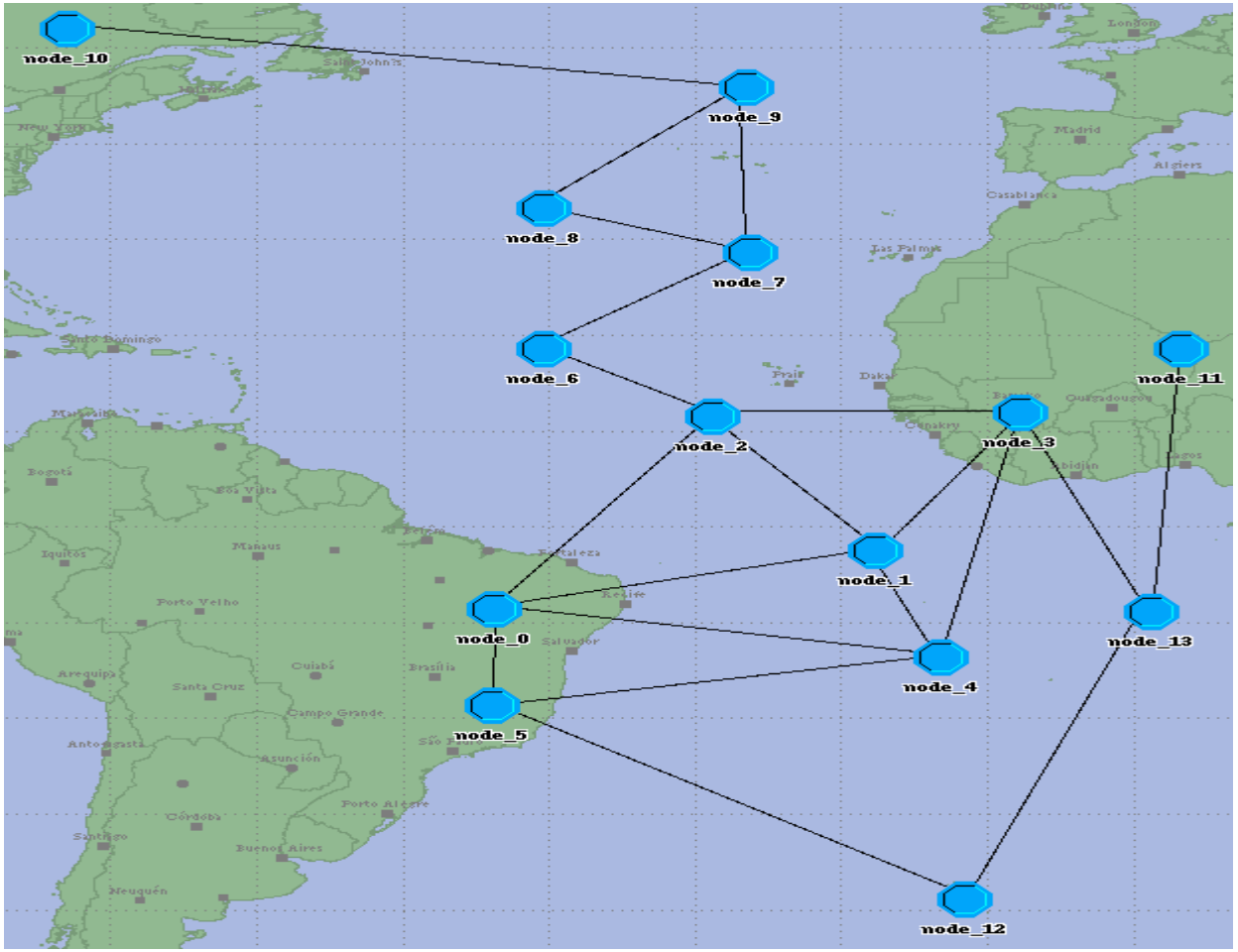
Network Topology

Network Model Simple Case



Network Topology

Network Model Complex Case



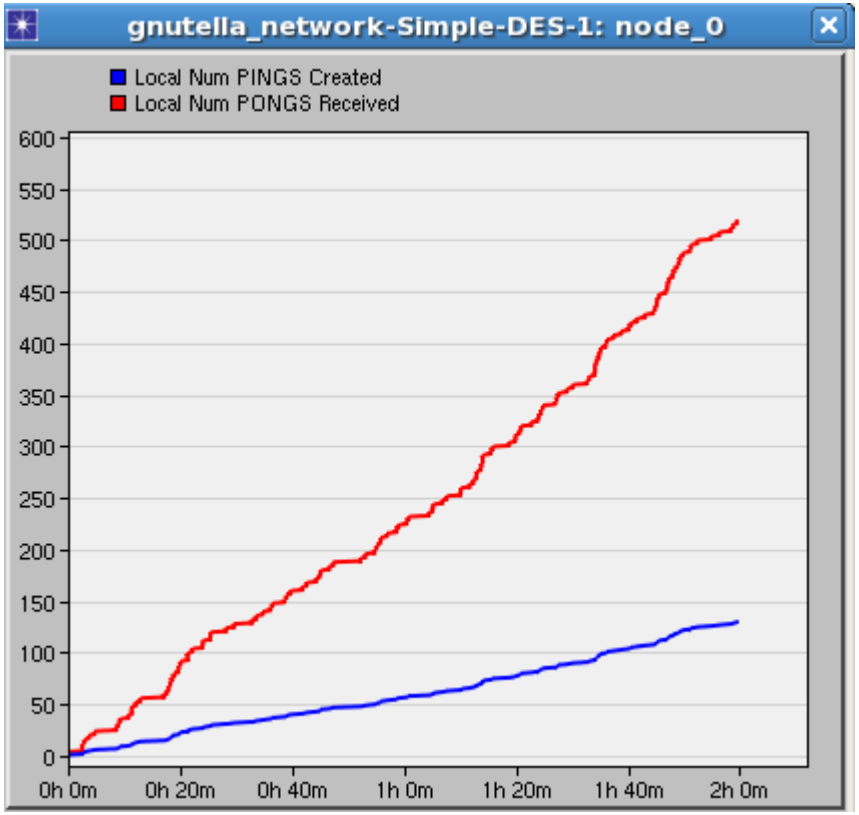
Simulations and Results

Network Model Simple Case

- TTL = 3
- 1 PING generated every 60 seconds
- 1 QUERY generated every 30 seconds
- Nodes assigned 10 files randomly from a total of 100 different files.
- Nodes 0 and 4 search for 4 of these files

Simulations and Results

- *PINGS created vs. PONGS received*

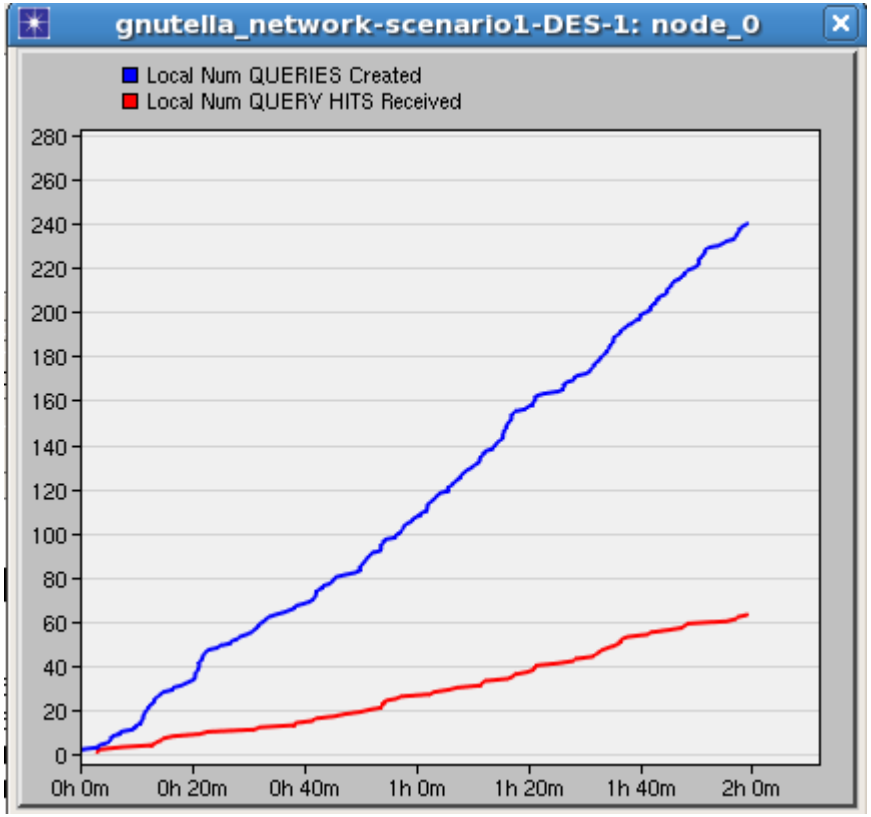


Simulations and Results

- 130 PINGS created by node 0
- 520 PONGS received
- 4 nodes found in the network

Simulations and Results

- *QUERIES created vs. QUERY HITS received*



Simulations and Results

- 240 QUERIES created by node 0
- 62 QUERY HITS received
- About 25% of queries successful

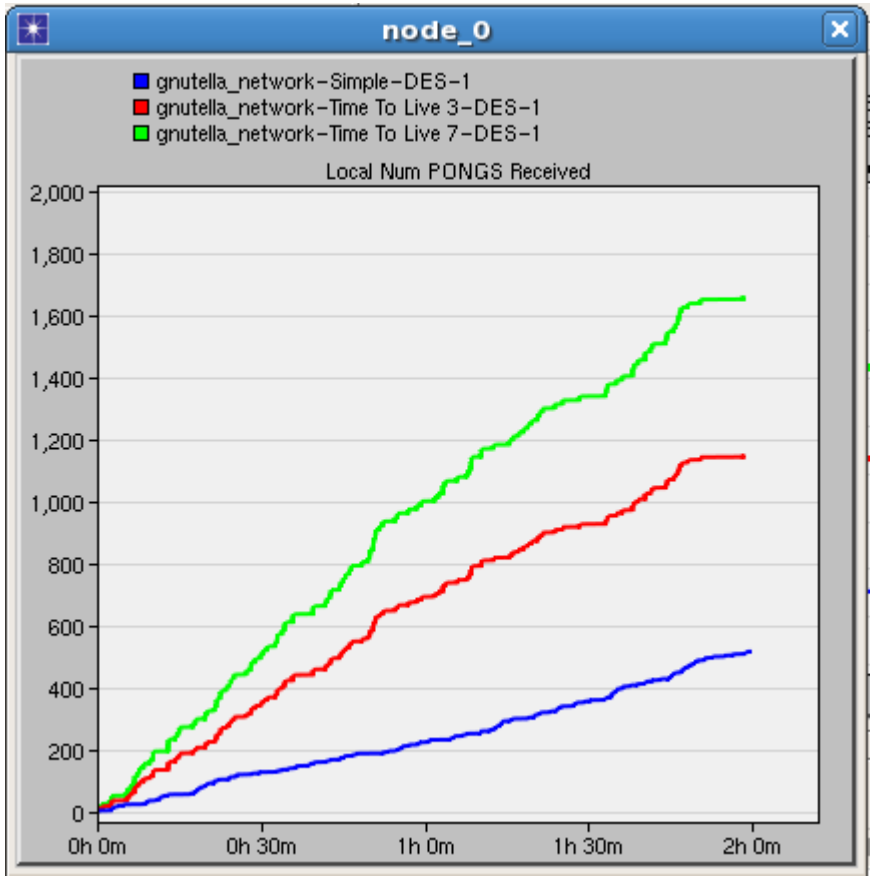
Simulations and Results

Network Model Complex Case

- TTL = 3 and 7
- 1 PING generated every 60 seconds
- 1 QUERY generated every 30 seconds
- Nodes assigned 10 files randomly from a total of 500 different files.
- Nodes 0, 4, 10, and 11 search for 10 of these files

Simulations and Results

PONGS received by node 0

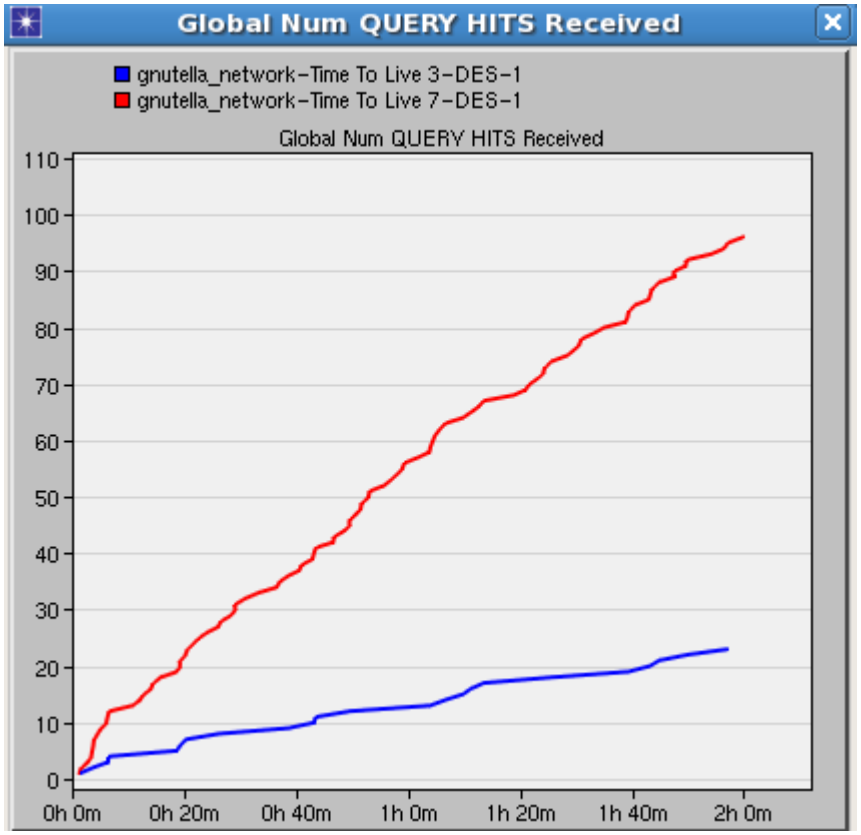


Simulations and Results

- Again 130 PINGS were created
- Simple case finds 4 nodes
- Complex case TTL = 3 finds 9 nodes
- Complex case TTL = 7 finds 13 nodes

Simulations and Results

- *Total number of QUERY HITS received for TTL = 3 and TTL = 7*

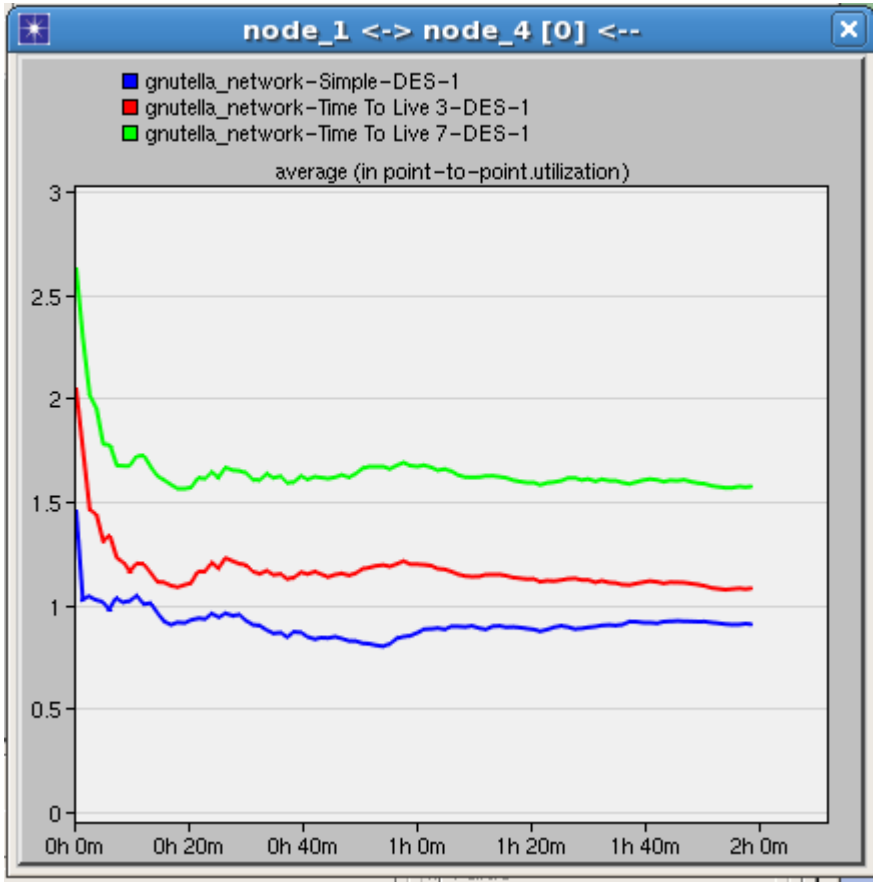


Simulations and Results

- Total of 996 QUERIES created
- Simple case did not receive any HITS
- TTL = 3 about 2.3% queries successful
- TTL = 7 about 9.6% queries successful

Simulations and Results

- *Link utilization for the 3 scenarios*



Advantages

- protocol allows users to transfer files between each other instead of having to search through a main server
- Distributed processing with low traffic
- network can work all the time since no key piece of equipment that should it fail would cause the network to be disrupted

Advantages

- transfer of data is available under most of firewall systems
- Flexibility in Query processing
- Low maintenance costs
- Easy access

Disadvantages

- Does not provide quality service due to overloading and downloading failures
- Unwillingness of file sharing by the peers
- Low downloading speed
- Lack of new features and services
- High Bandwidth usage
- Not guaranteed to find a file

Future Work

- Could change the structure of internet from web centric to purely distributed data model
- Expected to grow continuously with higher and higher number of nodes
- Will have to go through many changes in order to provide fully reliable results
- User ratings to be released to encourage file sharing

References

- [1] "How Gnutella Works" Internet: http://computer.howstuffworks.com/file-sharing1.htm [Mar. 23, 2013]
- [2] "The Annotated Gnutella Protocol Specification v0.4 (1)" Internet: http://rfc-gnutella.sourceforge.net/developer/stable/ [Mar. 23, 2013]
- [3] "Gnutella" Internet: http://en.wikipedia.org/wiki/Gnutella [Mar. 20, 2013]

References

- [4] Fabrizio Cornelli, Ernesto Damiani, Sabrina De Capitani di Vimercati, Stefano Paraboschi, Pierangela Samarati. "Choosin Reputable Servents in a P2P Network" Internet http://www2002.org/CDROM/refereed/68/ [Mar 22, 2013]

- [5] Gene Kan. "Gnutella" in Peer-To-Peer: Harnessing the Benefits of DisruptiveTechnology, 1st ed., Andy Oram, Ed. Sebastopol, CA: O'Reilly & Associates, Inc., 2001, pp.94-122

- [6] "Gnutella Nework" Inernet: http://alumni.cs.ucr.edu/~csyiazti/courses/cs260-2/project/html/img42.png [March 22, 2013]

- [7] Igor Ivkovic, "Improving Gnutella Protocol: Protocol Analysis and Research Proposals" Internet: http://www.cs.cornell.edu/people/egs/615/gnutella.pdf [March 22, 2013]