Peer-to-Peer Networking:
Mapping the Gnutella Network

Chinhua Lin

Contents

- Why Peer-to-Peer?
- Categories of P2P systems
- Gnutella Protocol
- Network Analysis
- Potential Improvements
Why Peer-to-Peer?

- Scalability
- Reliability
- Information
- Network Bandwidth

Categories of P2P systems

- Centrally coordinated
  - Napster, SETI@home
- Hierarchical
  - DNS
- Completely decentralized
  - Gnutella, Freenet
Gnutella Protocol
- A protocol for distributed search

- A Gnutella node is called a servent
- Client/Server: every Gnutella servent is both client and server.
- Fault-tolerant: the network won’t be interrupted if a subset of servents goes offline? (to be verified)
- A virtual network (app level)

Descriptor

- Type of Descriptors
  - Ping (Group membership)
  - Pong (Group membership)
  - Query
  - QueryHit
  - Push

- Description Header

<table>
<thead>
<tr>
<th>Message ID</th>
<th>Payload Description</th>
<th>TTL</th>
<th>Hops</th>
<th>Payload Length</th>
</tr>
</thead>
</table>
Ping:
- Probe the network
- A servent should forward an incoming Ping to all its directly connected nodes
- A node should periodically PINGs its neighbors, maintains a local list to help reconnect.

Pong:
- Only to respond to a ping message
- Should be sent along the same path as Ping
- Return Port#, IP, amount of sharing data

Example of Ping/Pong

A PING with the same UID is seen; so do not respond.
- **Query**
  - min speed, search criteria

- **QueryHit**

<table>
<thead>
<tr>
<th>Number of Hits</th>
<th>Port</th>
<th>IP Address</th>
<th>Speed</th>
<th>Result Set</th>
<th>Servent Identifier</th>
</tr>
</thead>
</table>

- **Push**

<table>
<thead>
<tr>
<th>Servent Identifier</th>
<th>File Index</th>
<th>IP Address</th>
<th>Port</th>
</tr>
</thead>
</table>

---

**Example of Push**

B doesn't support incoming connections

[Diagram showing the query and push process, with a server A connecting to a cloud server B via a query and query hit, and the push attempt to B being blocked with a cross symbol.]
**File downloads**

- **QueryHit:**
  After receiving the QueryHit, initiate direct download
- **HTTP:**
  Files are downloaded outside of the virtual Gnutella network

---

**Compared with Freenet**

- **Freenet:** Queries are forwarded to another node according to a local decision.

<table>
<thead>
<tr>
<th>Gnutella</th>
<th>Freenet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sacrifices efficiency for faster searches and better worse-case guarantees.</td>
<td>High scalability and efficient search in average conditions but sacrificing worse-case performance</td>
</tr>
</tbody>
</table>
*Connect to Gnutella network*

- Before issuing any descriptors, need to connect first.
- A Gnutella servent connects itself to the network by establishing a connection with another servent currently on the network (one or more connections).
- How to connect?
  1. Obtain the address
  2. Build a TCP/IP connection
  3. Send a request string
  4. Get the response
Network Analysis

- Purpose:
  - Organizational patterns
  - Network traffic
  - Efficiency in infrastructure usage
- Growth Trends & Dynamic Behavior
- Estimated Traffic
- Connectivity & Reliability

Design of Crawler system

- A crawler acts as a servent
  - Use ping/pong
  - Collect network topology
- Client/Server crawling strategy
  - Server is responsible for:
    - manages the list of nodes to be contacted
    - assembles final graph
    - assigns work to clients
Growth Trends & Dynamic Behavior

- 7 months: Nov 2000 – May 2001
- Gnutella doesn’t scale well, but still grow by 25 times during this interval
- 40% of nodes leave in 4hrs
- 25% alive over 24 hrs

Estimated Traffic

- November 2000: 36%- Query, user-generated traffic
  55%- Ping&Pong, maintain group membership
- June 2001: 91%-Query
  8 %-Ping
- 95% of any two node pair could exchange messages within 7 hops
- Conclusion:
  - Almost all broadcast messages reach all nodes.
  - Most links support similar traffic

Connectivity & Reliability

- November 2000
- (March, May 2001)

“Super peers”: with more connections/provide more contents
it’s important to map right with physical network
A perfect mapping:

A broadcasts a message, it traverses the D-E link once.

A inefficient mapping:
Mismatch
- two experiments show the mismatch of Gnutella

- Local area network
  - autonomous system (AS)
  - traffic cross borders is more expensive
  - Most Gnutella traffic cross borders, only 2-5% of Gnutella connections link node within a AS

- Domain Name System
  - assume DNS reflects Internet Infrastructure
  - Gnutella nodes cluster independently

Potential Improvements
- from these measurements and analysis

- Security mechanism
  - Ward off denial-of-service attacks

- Using Proxy cache mechanism
  - Query-caching scheme

- Improve message forwarding method
  - Freenet,

- Mix dissemination schemes
  - Random query forwarding
Reference

- M. Ripeanu, I. Foster, and A. Iamnitchi, “Mapping the gnutella network”, IEEE Internet Computing
- Li Gong, Peer-to-Peer in Action, vol. 6, no. 1, January/February 2002
- Gnutella Protocol Specification, version 0.4; available at www9.limewire.com/developer/gnutella_protocol_0.4.pdf
- Gnutella Applications
  www.limewire.com