Security Aspects of Napster and Gnutella

Steven M. Bellovin
smb@research.att.com
http://www.research.att.com/~smb

Common Functions

- Share files.
- Peer-to-peer – files don’t reside on a central server.
- Each user decides which files to offer to others. Protocol supplies index and connectivity information. Data transfer is end-to-end, and does not use central server.

Napster

- Everyone connects to central server.
- Server compiles and distributes index.
- Server also provides “chat room” function – independent of file-sharing aspect.
- Protocol details reverse-engineered.

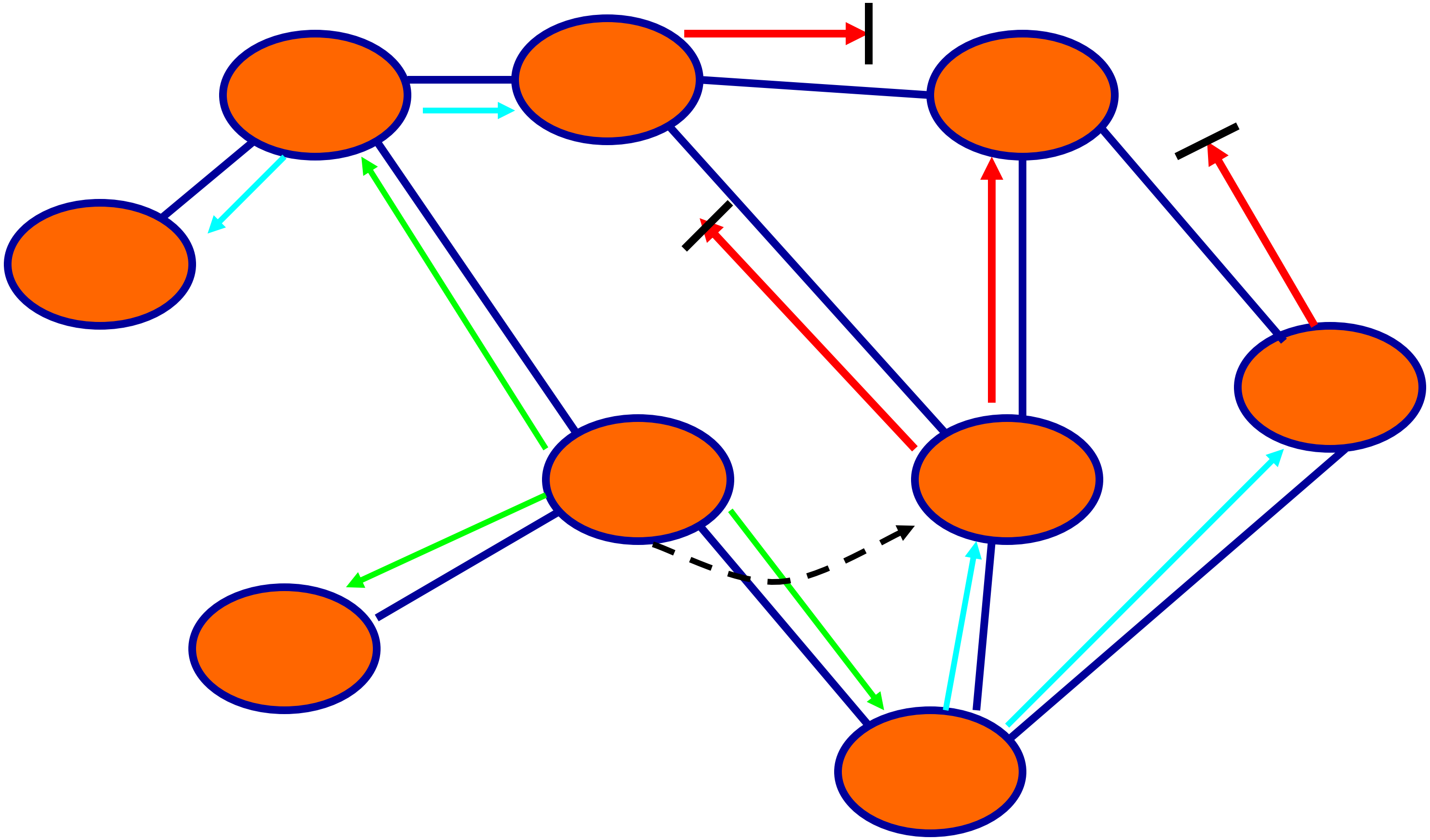
Gnutella

- No central server.
- No index.
- Users send queries to a neighbor; neighbors answer if they can, and also forward query to their neighbors.
  - Note: must know DNS name or IP address of some starting point.
- Client retrieves file directly from one answerer.
- Open protocol specification.

Gnutella Protocol Details

- Simple protocol: 5 messages.
  - Ping, pong, push, query, query hits.
- Uses “flooding protocol” – speak to all neighbors.
- HTTP used for actual content transfer.
- No login, no authentication, no central authority of any type.

Gnutella Topology



Common Header

- 16-byte Windows GUID
  - Clients *must* drop messages if GUID seen recently.
- Message type.
- Time-to-live (limits maximum spread of message).
- Hop count – how far away the sender is.
- Payload length.

Ping and Pong

- Used for topology discovery – ask who’s out there.
- Nodes that choose to reply with their IP address, plus the amount of data they’re sharing.
- Provides new connection points for nodes.
- But what if they lie about their IP address?

Query and Query Reply

- Query lists search terms, minimum server speed acceptable.
- Query response gives IP address, port, speed, files that satisfy query, GUID of querier.
  - Querier then connects to offerer and requests file.

Push

- Intended to bypass firewall – you can’t serve a file if you’re behind a firewall.
- If requester can’t connect, it sends a “push” command instead, with its IP address and port number.
- Offerer does an outbound connect to that host, and sends the file.

Gnutella Analysis

- Gives away topology information.
- Hard to control via firewalls.
- Unchecked IP address and port number announcements can be used to generate flooding attacks, and possibly worse.
- GUID *may* be usable to trace back Gnutella messages.

GUID Tracing

- On Windows 95, 98, NT, GUID contains the hardware MAC address, which is constant over time.
- Privacy violation – can be used to link requests over time.
- Windows 2000 (and the UNIX clients I've looked at) use random-appearing GUIDs.
  - Is there some hidden linkage?

Leakage

- Announces IP addresses.
- Appears to announce full path names.
- Announces Gnutella topology, which may (or may not) reflect real-world patterns of association.
- Can use any port number – hard to detect, hard to control outbound via firewalls.
- Nosy node can record queries, responses.

Flooding

- Pong messages contain IP addresses and port numbers – will other nodes auto-connect?
  - What if a node claims to be port 80 on www.cnn.com?
- Query/Push pair is worse – an attacker can induce many sites to try to send a large file to some arbitrary destination.
  - Similar to “FTP Bounce” attack.

Content Issues

- What if I send you fake content?
- What if I send obscene content in response to innocent queries?
- Note: falsely advertising a high-speed link can be used to attract clients.

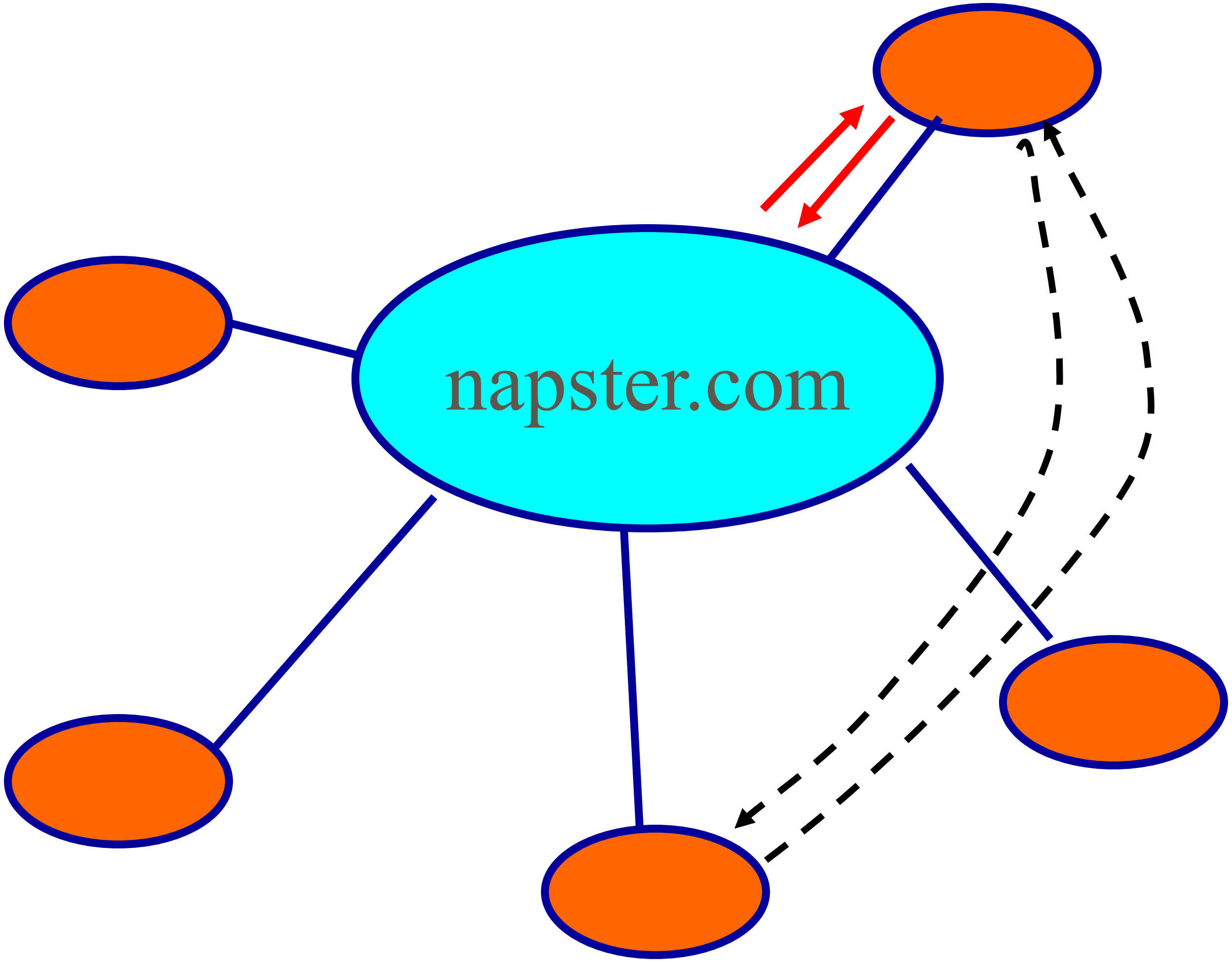
UI Issues

- Gnutella can be used to share arbitrary files.
- Some UIs provide an easy way to open files.
- Is this mechanism safe? How does it decide how to open a file? If done wrong, this is as dangerous as email attachments.
  - Can I get a .EXE or a .VBS file when I asked for an MP3?
- Again, fake line speed announcements can be used to attract clients.

Napster Protocol Details

- Complex client/server protocol with central site.
- Users can register, log in, etc.
  - Registration message includes age, income, and education…
  - Central site can bounce users, ban them, etc.
- Different message groups for chat rooms, searching/browsing, upload/download.
- File transfer is direct, and doesn't go through napster.com's site.

Napster Topology



Searching and Indexing

- Client sends search or browse requests to central site.
  - Can browse some other user's files.
  - Response come back from central site.
- Only explicitly-shared files should be retrievable.
- Only handles MP3.
  - "Wrapster" can package other file types in MP3 envelope.

Chat Rooms

- Conversations among users.
  - Nominally moderated.
- All traffic flows too/from central site.
  - Central site not working that well right now – there are several servers that don’t share status information.
- Multiple topics, etc.
- Clients can have “hot lists” of their friends.
  - Privacy issues?

File Transfers

- Transfer request goes to central site.
- Data transfer is direct.
    - Client and server both notify central site of status, to support load limits.
- Clients can use any port numbers.
- Firewall bypass mechanisms – reverse who does active connect.

UI Issues

- Less opportunity for auto-exec of nasty programs.
  - What if Wrapster functionality becomes common?
- Is browsing more intrusive than query/response?

Napster Analysis

- Much harder for clients to lie – can’t give fake IP addresses, port numbers, etc.
- Central site can exert much more control.
- Privacy issues – central site knows (almost) all.
- Fake content and fake line speed attacks still apply – but in theory, are more traceable.

Napster versus Gnutella

- Napster is more centralized – easier to monitor and control, for good or bad purposes.
- Gnutella can *probably* scale further *if* better topology reconstruction algorithms are developed.
- Only Gnutella can easily share arbitrary files – but that’s a likely growth direction for Napster.
- Gnutella is probably the style of the future – avoid central sites.

Implementation Concerns

- Both can have bugs, including buffer overflows – and bugs are the biggest cause of security problems.
  - Some Gnutella clones are poorly written.
- Both have direct user-to-user communication – can raise privacy issues.