P2P networks, free riding, and counter-mechanisms

İbrahim Körpeoğlu
Department of Computer Engineering
Bilkent University

Seminar Place: Department of Electrical and Electronics Engineering
Bilkent University, Ankara
Date: October 19, 2004

Outline of the talk

- Content sharing and distribution in Internet
- P2P networks and architectures
- Gnutella architecture and protocol
- Free riding and Gnutella
- Preventing free riding in Gnutella

Outline

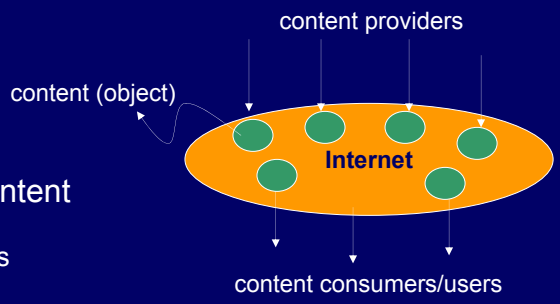
- Content sharing and distribution in Internet
- P2P networks and architectures
- Gnutella architecture and protocol
- Free riding and Gnutella
- Preventing free riding in Gnutella

Internet and accessing content

- Internet is full of interesting content (objects).
  - HTML pages
  - document files (.pdf, .doc, .ps, …)
  - music files (mp3, ….).
  - video files (.mpg, …).
  - images (.gif, .jpg).
  - programs
- There are content provides and content consumers.
- There are various ways for content consumers to access and use the content.

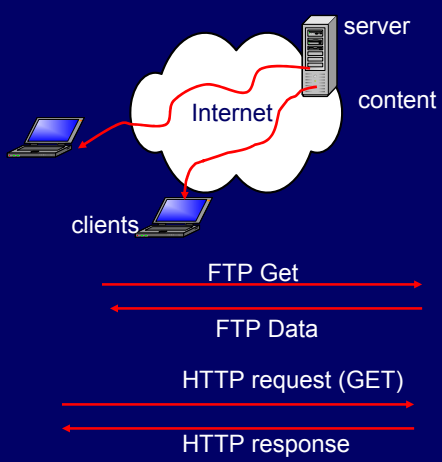
Content providers and users

- Who are the content provides
  - Companies
  - Institutions
  - Individuals
  - Governments
  - …
- Who are the content consumers
  - Individual users
  - Companies
  - Institutions.
  - Groups.
  - ….



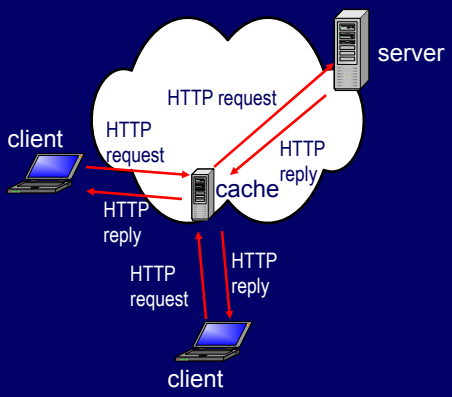
Ways to download content: pure client-server architecture

- Use client server architecture and associated protocols.
- Put content in servers.
- Clients can download the content via an Internet protocol
  - FTP, or
  - HTTP
- Servers can become bottleneck points.
  - Reliability and performance problems.



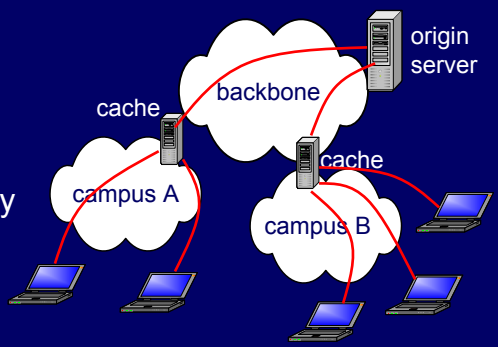
Ways to download content: web caching

- Install web caches (web proxies) in between clients and servers.
- Requests are first directed to the proxy.
  - If a requested object is in the cache, it is satisfied from there, otherwise from the origin server.
- Better performance.
- Better reliability.



Web caching: where can we install the proxies

- Proxies can be installed on network boundaries
  - A campus edge
  - A department edge
  - A country edge.
- We may have a hierarchy of caches.
- Update of caches is required when content changes.



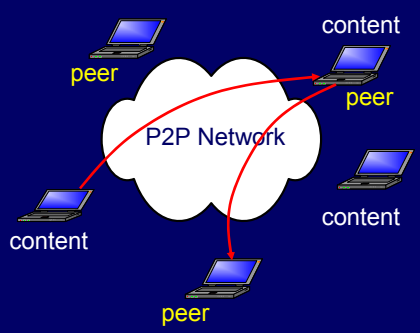
with the use of caches, we can have better response time, better scalability, and better reliability.

Outline

- Content sharing and distribution in Internet
- P2P networks and architectures
- Gnutella architecture and protocol
- Free riding and Gnutella
- Preventing free riding in Gnutella

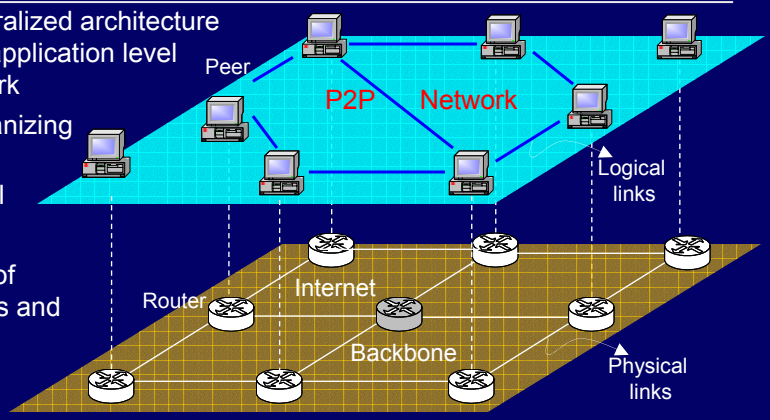
Ways to download content: P2P file sharing

- Content is stored directly on edge devices: PCs.
  - They are called peers.
- A peer can be both a client and server.
  - Call it a SERVENT
- An important problem is how to locate the content?



P2P network features

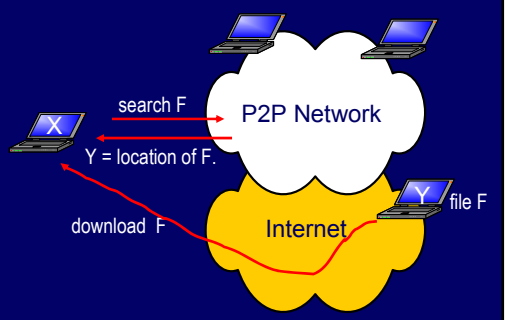
- Decentralized architecture
  - An application level network
- self-organizing network
- No central control
- Consists of autonomous and anonymous peers



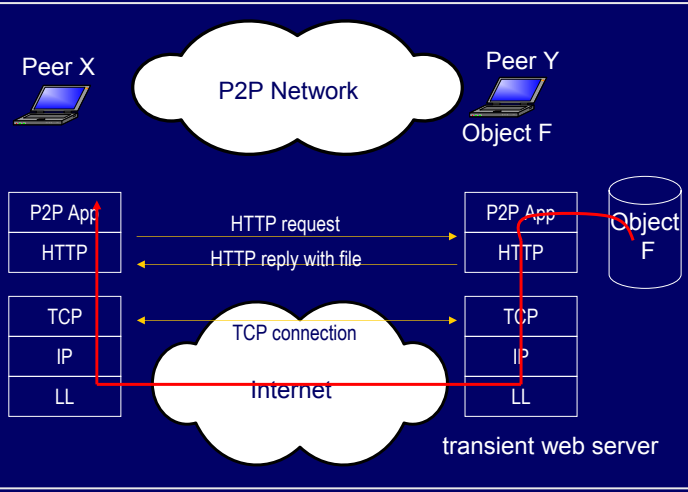
- A peer may be connected to one or more other peers

Example use of P2P services

- Peer X wants a file F.
- X invokes a search operation in the P2P network for file F.
- P2P network returns the location of file F, i.e. the address of a peer Y holding the file.
- Peer X downloads the file directly from peer Y.

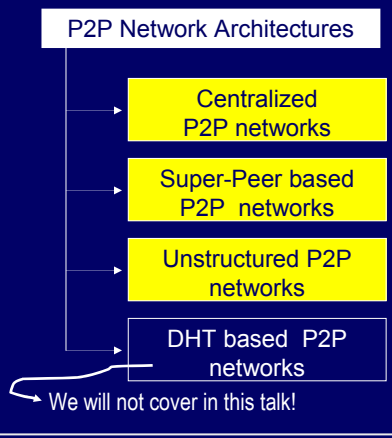


Closer look to downloads



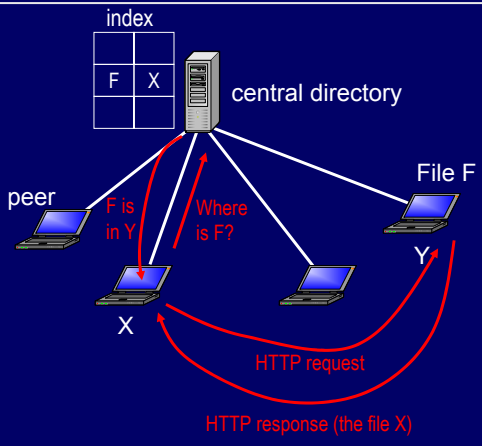
P2P network architecture and protocols

- How is a P2P network designed to perform two functions:
  - 1) locate a desired file X.
  - 2) download the desired file X.
- There are several P2P architectures that can realize these services.
- The scheme to locate the content usually directly affects the properties of the architecture.



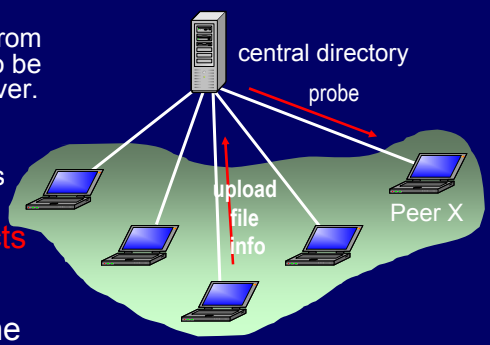
Centralized architecture

- There is a central directory server which stores location information (index) for files shared.
- A peer that wants to download a file consults to this server for the location of the file.
- Disadvantages:
  - Single point of failure
  - Performance bottleneck
  - Legal issues.



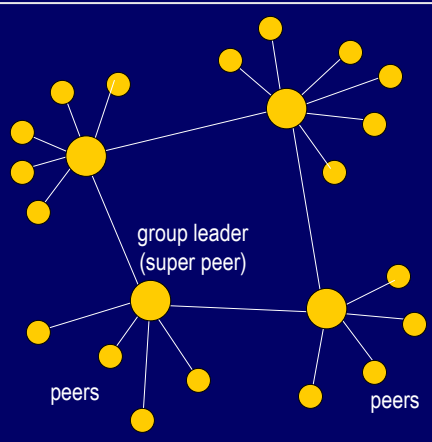
Centralized architecture: connection/disconnection to the network

- Disconnections
  - When a peer disconnects from the network, this fact has to be detected by the central server.
    - Periodic probing
  - The server then deletes all location information for files stored in this peer.
- When a new peer connects to the network, it has to upload information about its shared files to the central server.



Hierarchical overlay architecture (super-peer architecture)

- Location directory (index) is distributed.
- Peers are organized into groups
- Two types of peers:
  - Normal peers.
  - Group leaders (super peers)
- A group leader keeps the location information for files in its peer group.
- A peer has connection to only to its group leader.

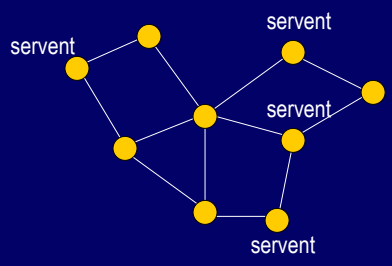


Outline of the talk

- Content sharing and distribution in Internet
- P2P networks and architectures
- Gnutella architecture and protocol
- Free riding and Gnutella
- Preventing free riding in Gnutella

Unstructured P2P architecture and Gnutella

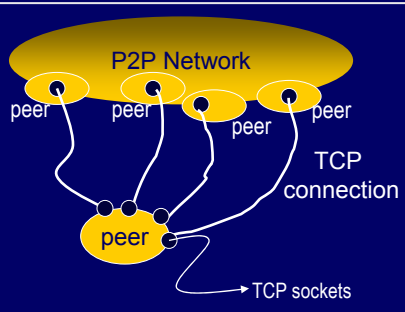
- Totally decentralized architecture
- Flat topology
  - Each peer has equal responsibility.
- A peer is both a server and client: servlet
- No directory maintained.
- Location is found by flooding a location query through the P2P network.



Gnutella is a P2P network with this architecture.

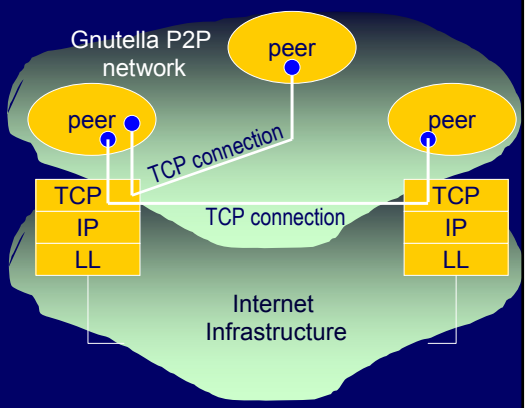
Gnutella connection (i.e. Gnutella link)

- A peer is connected to the P2P network through one or more peers (neighbors).
- A connection to another peer is done using TCP sockets.
  - Reliable delivery of Gnutella protocol messages between peers.
- Gnutella connections go over TCP connections.



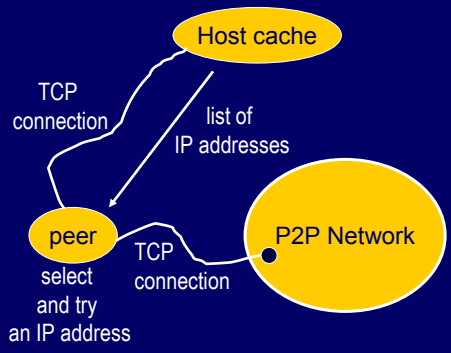
Gnutella software and TCP connections

- A Gnutella servlet running on a machine is just an application level software.
- Uses TCP to connect to other peers
- Implements HTTP protocol and uses that during downloads.



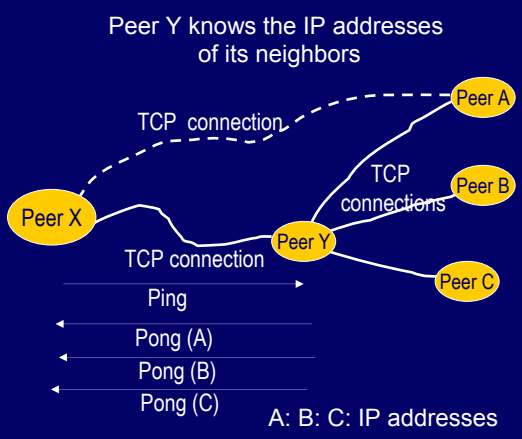
How does a peer initially know which peer(s) to connect to? Use of host caches

- A peer consults to a host cache machine.
- A host cache is a special server that knows the IP addresses of some active peers.
- A list of IP addresses of active peers can be downloaded from the host cache.



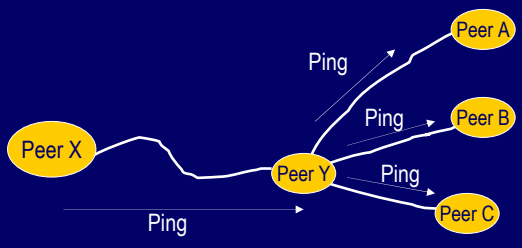
How does a peer learn more about other peers? Ping/Pong messages

- A peer can use its neighbors to learn about other peers.
- It sends Ping messages to its neighbors.
- Pong messages will be returned as replies
- A Pong message includes one IP address (of the peer or a peer).



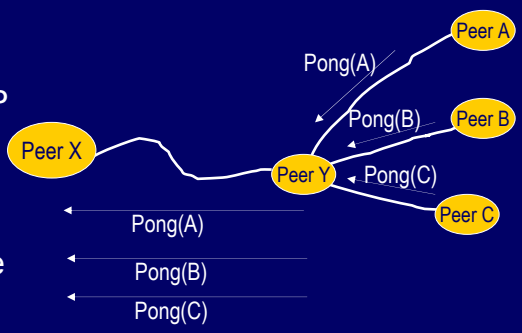
How does a peer learn more about other peers? Ping messages

- Ping messages can be relayed through more than one hop.
- A peer receiving a Ping message may forward it to all its neighbors.
- If time-to-live (TTL) value of Ping has reached to zero, no forwarding is done.

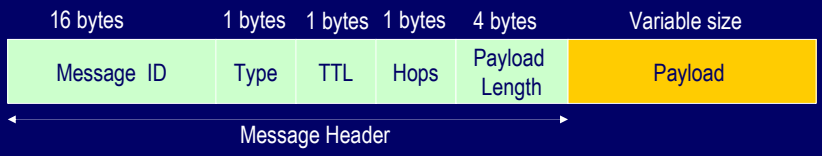


How does a peer learn more about other peers? Pong messages

- Pong messages will follow the reverse path of Ping messages.
- Pong message includes also a port number besides the IP address.
- The (port number, IP address) pair can be used to establish a TCP connection to the machine holding that IP address.



Message Format of Gnutellla

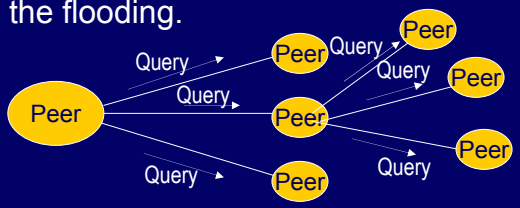


- Each Gnutella message starts with a header.
- Payload is different for each message type:
  - Ping message
  - Pong message
  - Query message
  - Query Hit message
  - Push message

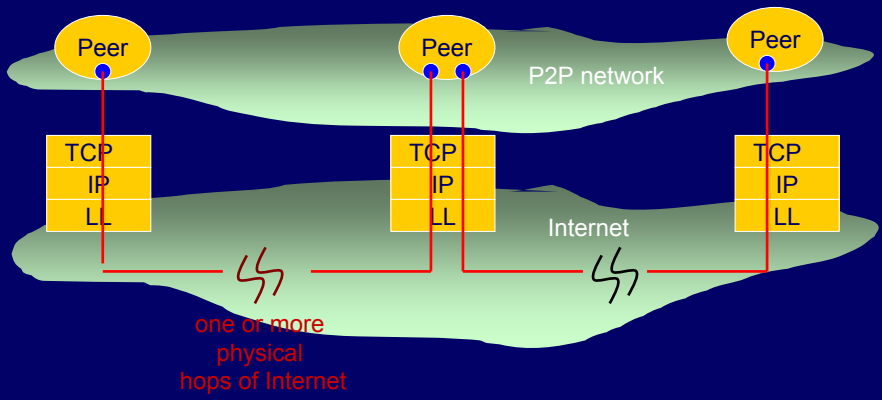
Querying

Message ID	Type			Length		
	0x80	TTL	0	X	Min Speed	(Search) Keywords

- A peer prepares a Query message and sends it to each of its neighbors.
- Each neighbor forwards the Query to their neighbors.
- TTL limits the flooding.



Querying: vertical view



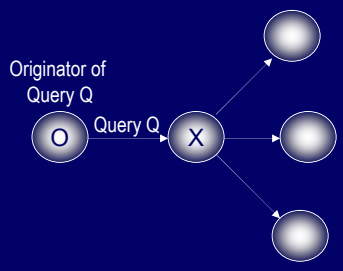
Query message generation at a peer

Generate (Query Q)

Create an empty Query message Q;

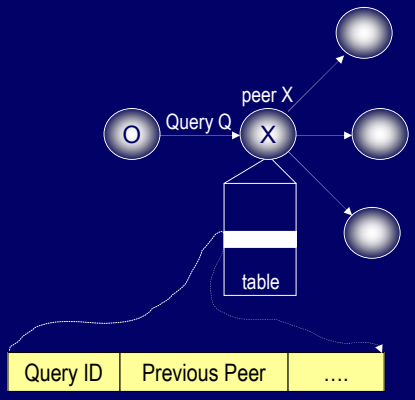
Q.QID = select a unique 16 byte number ;
Q.TTL = a predefied value, say 7;
Q.Hops = 0;
Q.Type = 0x80;

Q.Minspeed = minimum required speed of an answering peer.
Q.Keywords = search keywords received from the user.
Create an entry for query Q in the query table.
Send the query to all the neighbors.



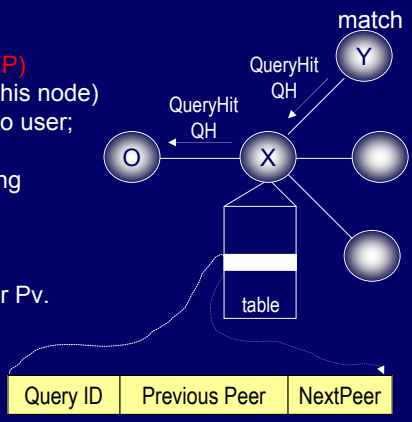
Query message processing at a peer

Process (Query Q received from a peer P)
If (Q.QID has been seen earlier)
    Drop the Query Q.
Else
    Create an entry E in table of queries
        E.QID = Q. QID;
        E.PreviousPeer = P;
    Local search for match.
    If (match found)
        Send back Query Hit
    Decrement Q.TTL
    increment Q.Hops
    If (Q.TTL == 0)
        Drop the Query Q.
    Else
        Send the Query Q to all neighbors excluding the query sender.

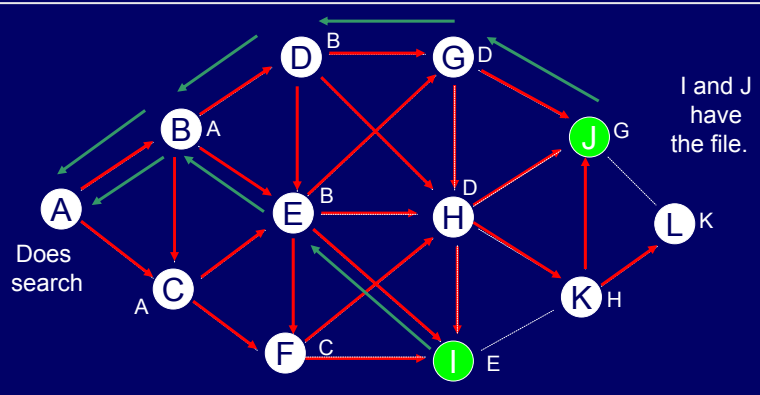


Query Hit message processing at a peer

Process (QueryHit QH received from a peer P)
If (the corresponding Query is generated at this node)
display the results in the Query hit to user;
Else
Search for the table entry E matching
the QH.QID
Pv = E.PreviousPeer;
E.NextPeer = P.
Send the Query Hit to previous peer Pv.



Gnutella search in action



I and J have the file.

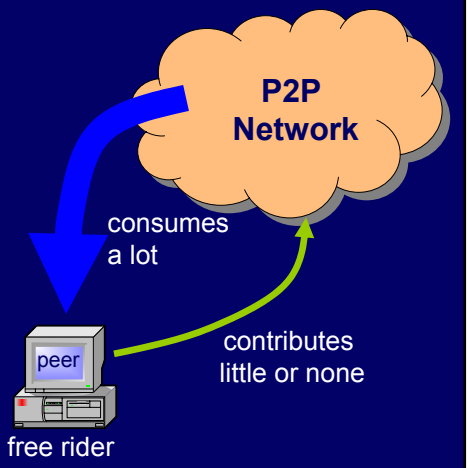
A will get two replies.

Outline of the talk

- Content sharing and distribution in Internet
- P2P networks and architectures
- Gnutella architecture and protocol
- Free riding and Gnutella
- Preventing free riding in Gnutella

What is free riding?

- Free riding means exploiting P2P network resources, but not contributing to the P2P network at desirable levels.
- A free rider may do search and download operations, but:
  - May not share any content
  - May not share any interesting content.
  - May not forward protocol messages, etc.

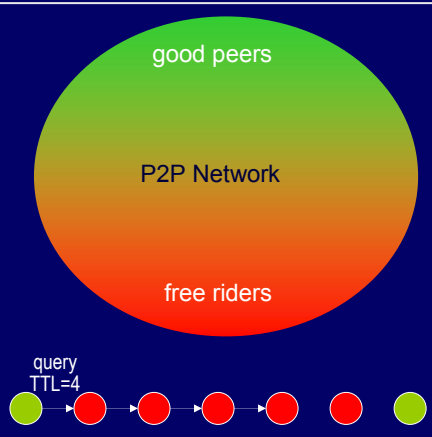


Why peers free ride?

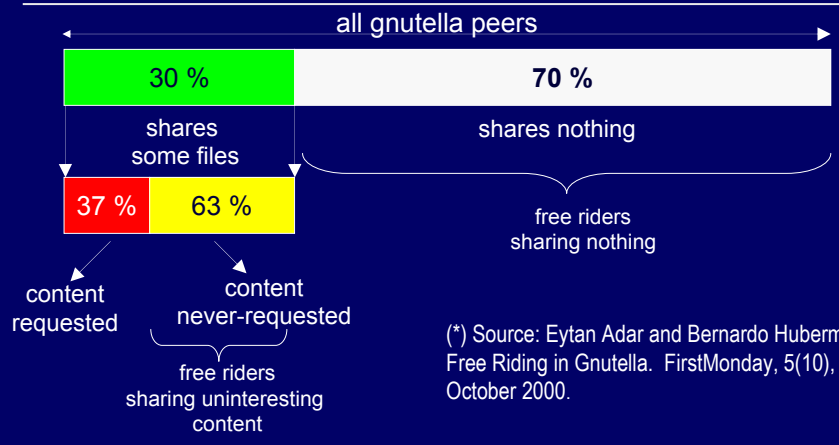
- The peer may not want to share and contribute because:
  - the peer has security concerns.
  - the peers doesn't want to consume its network connection's bandwidth.
  - the peer doesn't want to consume its local storage and processing resources.

Why free riding is not desired?

- Good peers will be overloaded.
- Available content will decrease in time if only few people share.
- Fault tolerance will degrade
  - If good peers get down, content will not be reachable.
- Search horizon will effectively shrink if too many free riders exist in the network.



Free riding in Gnutella*



(*) Source: Eytan Adar and Bernardo Huberman, Free Riding in Gnutella. FirstMonday, 5(10), October 2000.

About only 10 % of peers provide content that is searched for and downloaded.

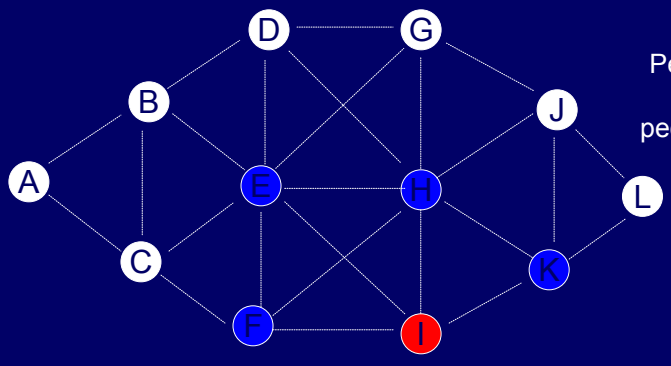
Outline of the talk

- Content sharing and distribution in Internet
- P2P networks and architectures
- Gnutella architecture and protocol
- Free riding and Gnutella
- Preventing free riding in Gnutella
  - A scheme to detect and discourage free riders.

A distributed scheme against free riding

- Designed for Gnutella
- Works in a distributed manner. No central control and management
- Based on monitoring protocol messages and updating counters.
- A proposal for:
  - Detecting free riders.
  - Applying discouraging counter-actions against free riders.

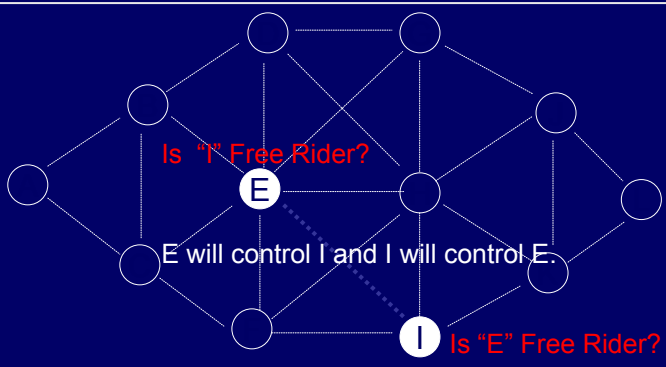
The approach



Peer I will control its neighboring peers E, F, H, and K.

The approach is based on having each peer control its direct neighbors.

The approach

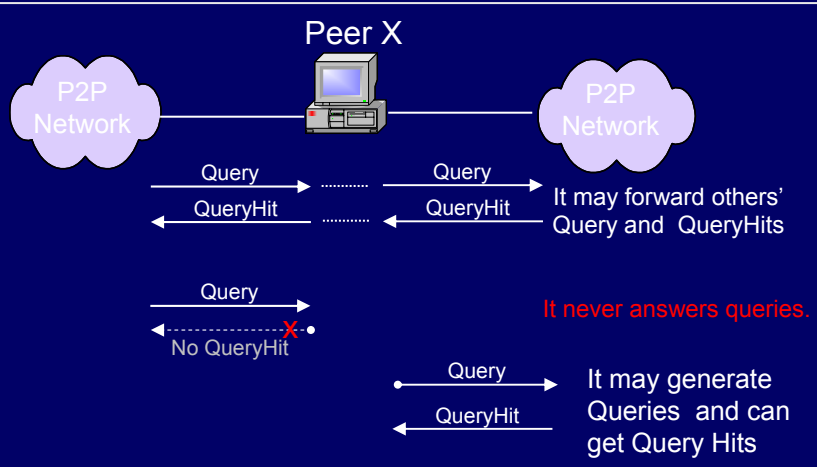


Neighboring peers will control each other.

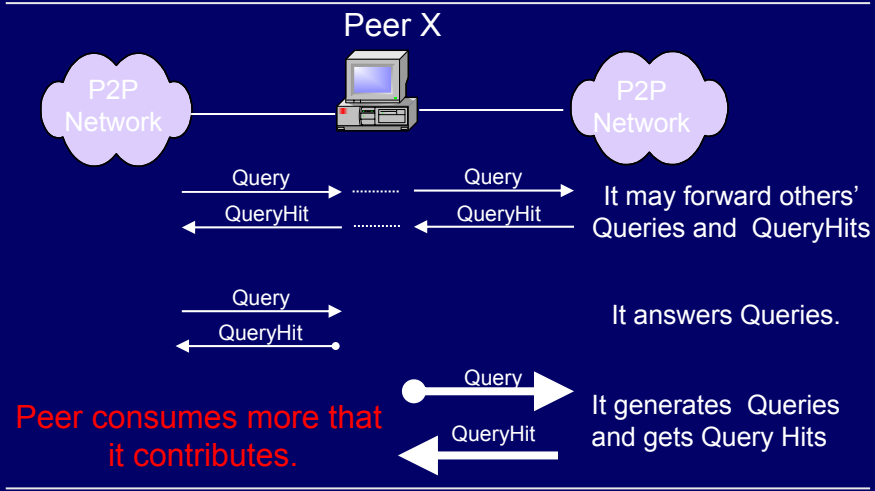
Free Riding Types

- We classify free riding behavior into three groups.
  - A free rider peer can behave as a:
    1) Non-Contributor
    2) Consumer
    3) Dropper
- A free rider may exhibit one or more of these behaviors.

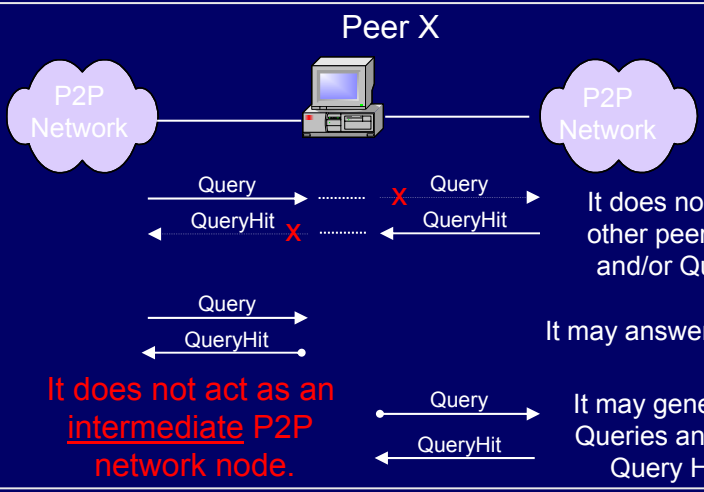
What is a non-contributor?



What is a consumer

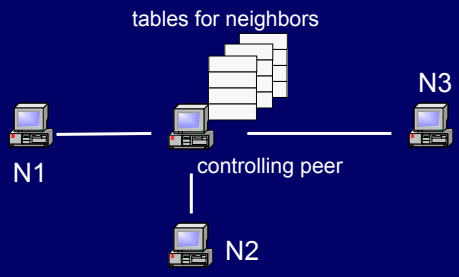


What is a dropper

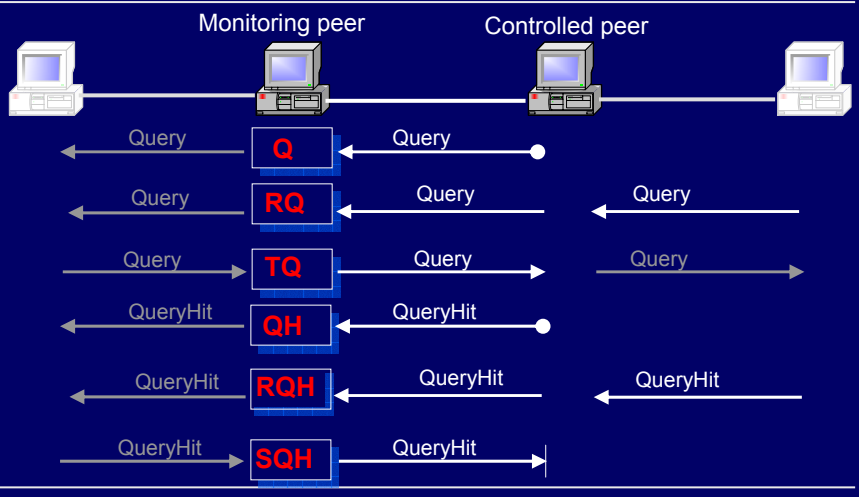


How do we decide about free riding and its type in the context of Gnutella?

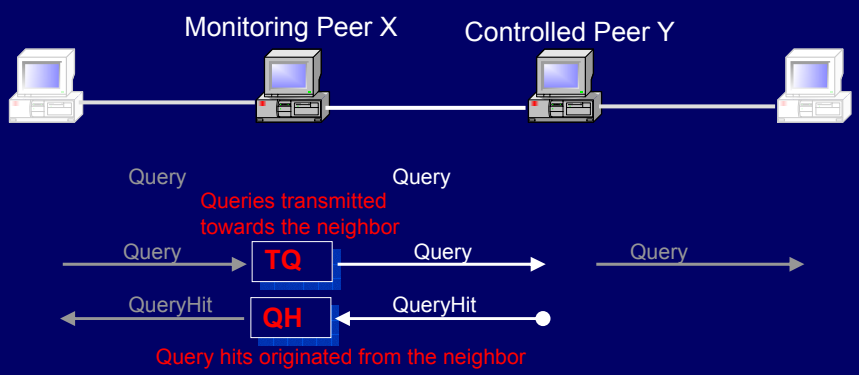
- Count the protocol messages received from and sent to neighboring peers.
- Update counters in tables maintained for each neighbor.



Counters

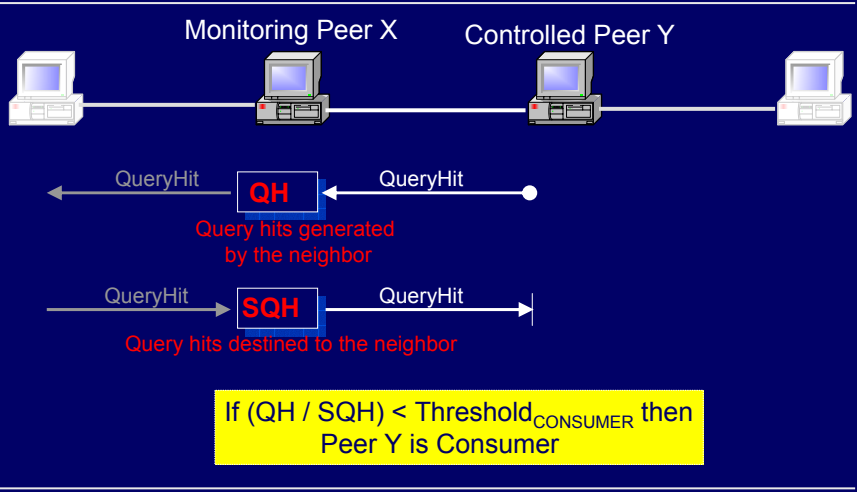


Detecting a non-contributor



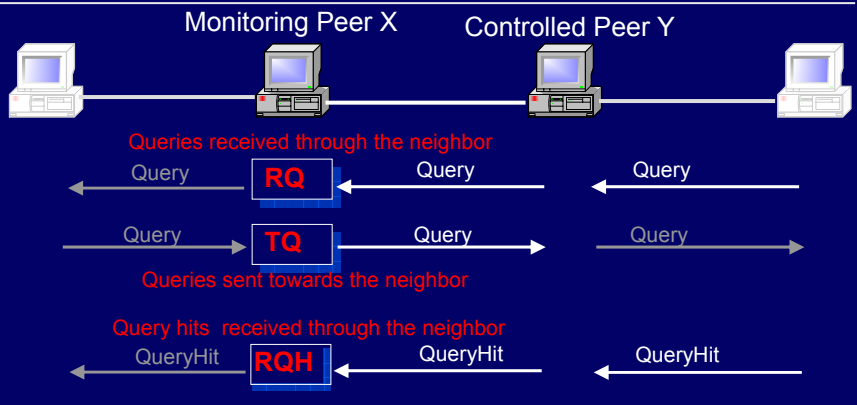
If (QH / TQ) < $\text{Threshold}_{\text{NONCONTRIBUTOR}}$ then Peer Y is Non-Contributor

Detecting a consumer



If (QH / SQH) < $Threshold_{CONSUMER}$ then Peer Y is Consumer

Detecting a dropper

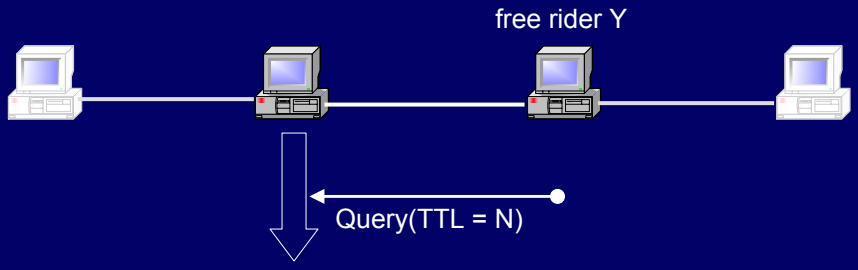


If ([RQ + RQH] / TQ) < $Threshold_{DROPPER}$ then Peer Y is Dropper

Actions

- We define three levels of discouraging counter-actions that can be applied against neighboring free riders.
- Level 1: Decrement the TTL value in Queries
  - Decrease the search horizon of the free rider
- Level 2: Ignore Queries partially
  - Free rider will not be able to search through this neighbor
- Level 3 actions: Disconnect the peer.
  - Free rider has to try to connect through another peer if it still wants to use the P2P network services

Level 1 action: decrement TTL



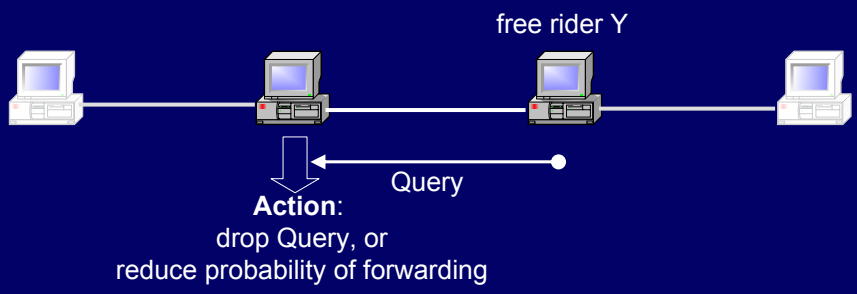
Action
decrease TTL by M

Query(TTL = N-M)

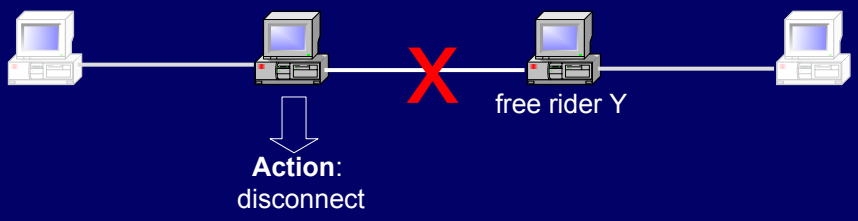
M can be 1 or more.

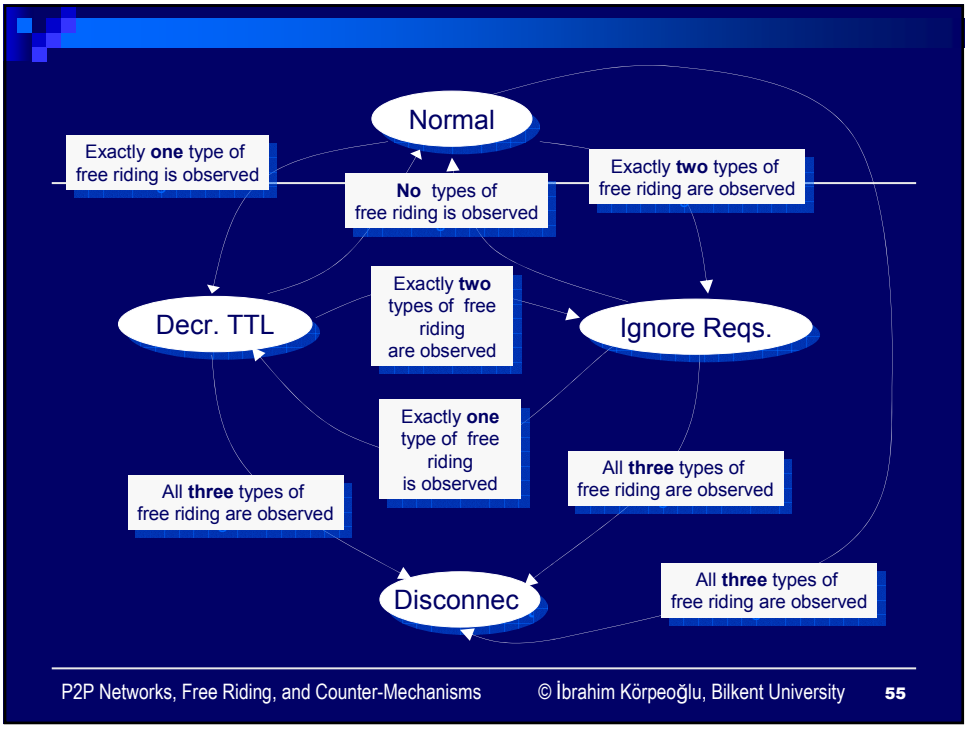
This will cause the horizon for the neighbor to shrink

Level 2 action: ignore requests



Level 3 action: disconnect peer





Summary and Future Work

- Great amount of free riding is observed in Gnutella network.
- Free riding has adverse affects on the proper operation of P2P networks.
- We described a distributed and measurement based scheme to:
  - Detect free riders in a Gnutella network
  - Take discouraging actions against free riders in a Gnutella network.
- We are currently working on a simulation tool that is simulating free riding on a Gnutella network model.
  - The simulator will be used to evaluate the described scheme.