Gnutella

1. Gnutella Background
   a) Gnutella, the Name …….
   b) History Of Gnutella
   c) What Is Gnutella?
2. Gnutella, In Operation
   a) Gnutella Jargon
   b) Gnutella Descriptor
   c) Gnutella Scenario – algorithm
   d) Searching Gnutella – animations
   e) Joining a Gnutella Network
3. Gnutella Protocol…
   a) Gnutella Descriptors
   b) Gnutella Descriptor Header & payload types
4. Gnutella Clients…
   a) Current Gnutella Clients

Gnutella Name

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**.

GNU: Recursive Acronym
GNU’s Not Unix ….

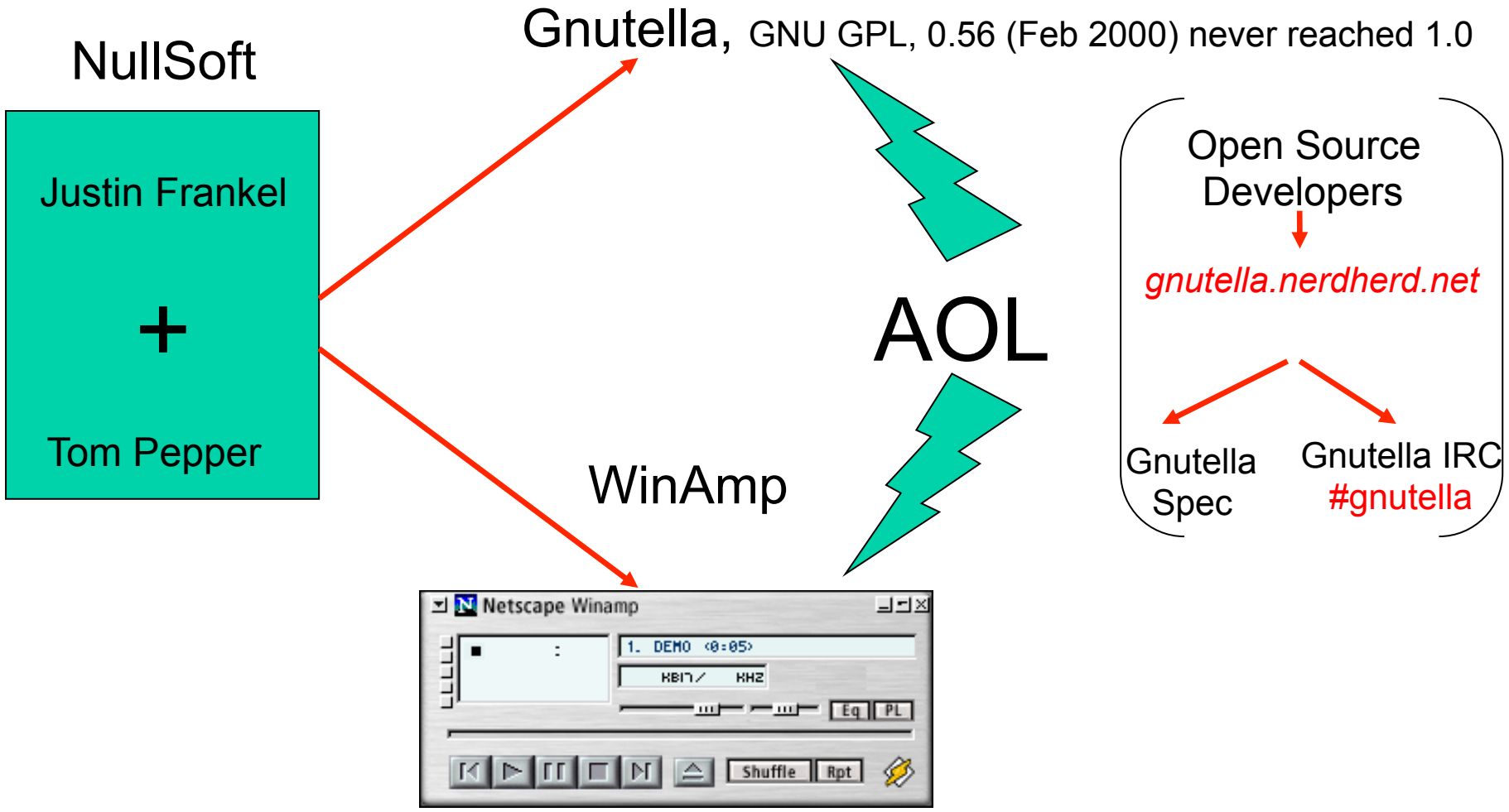
GNU

+

Nutella

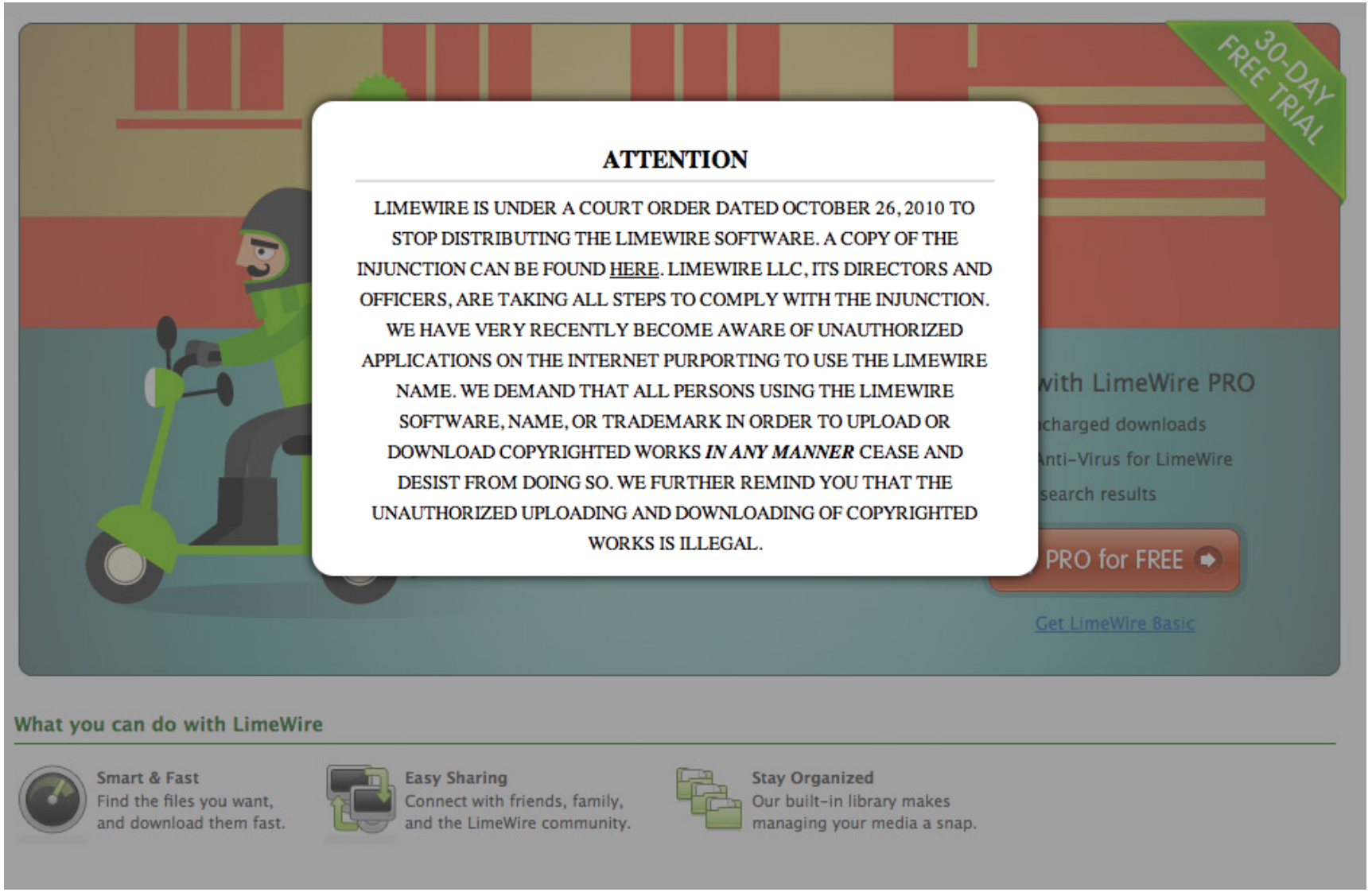
Nutella: a hazelnut chocolate spread produced by the Italian confectioner *Ferrero* ….

History Of Gnutella

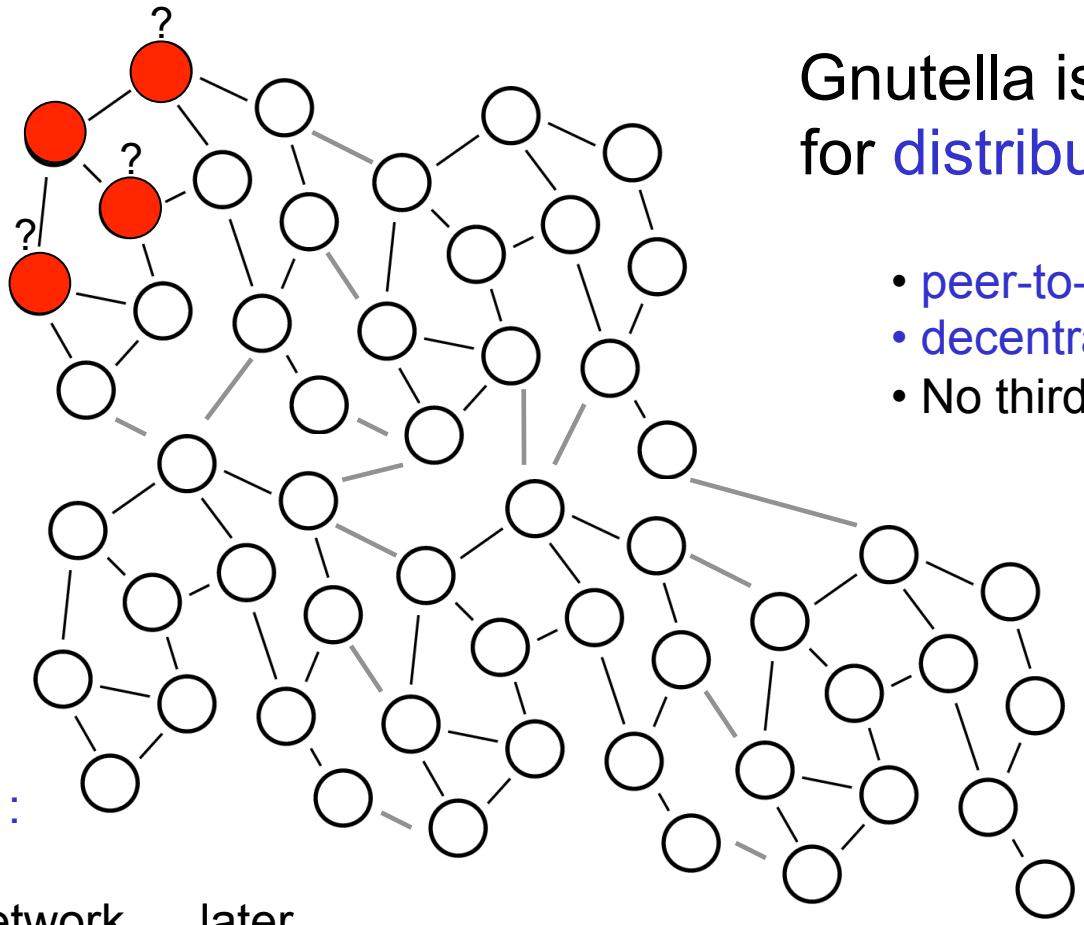


The End of Gnutella? October 2010

http://www.limewire.com



What is Gnutella?



Gnutella is a protocol for distributed search

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

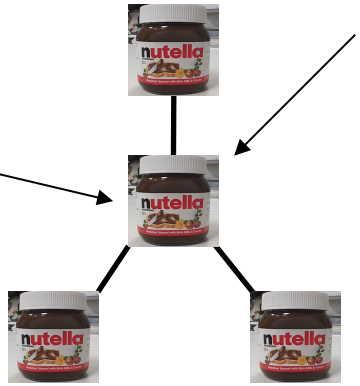
Two stages :

1. Join Network … later
2. Use Network
    1. Discover other peers
    2. Search other peers

The Jargon

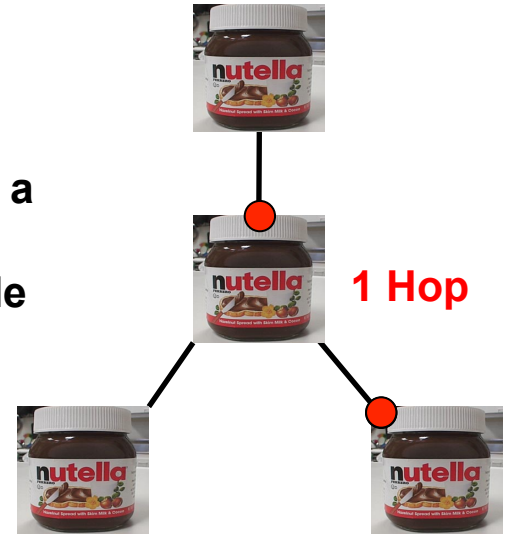
Servent: A Gnutella node.

Each servent is both a client and a server

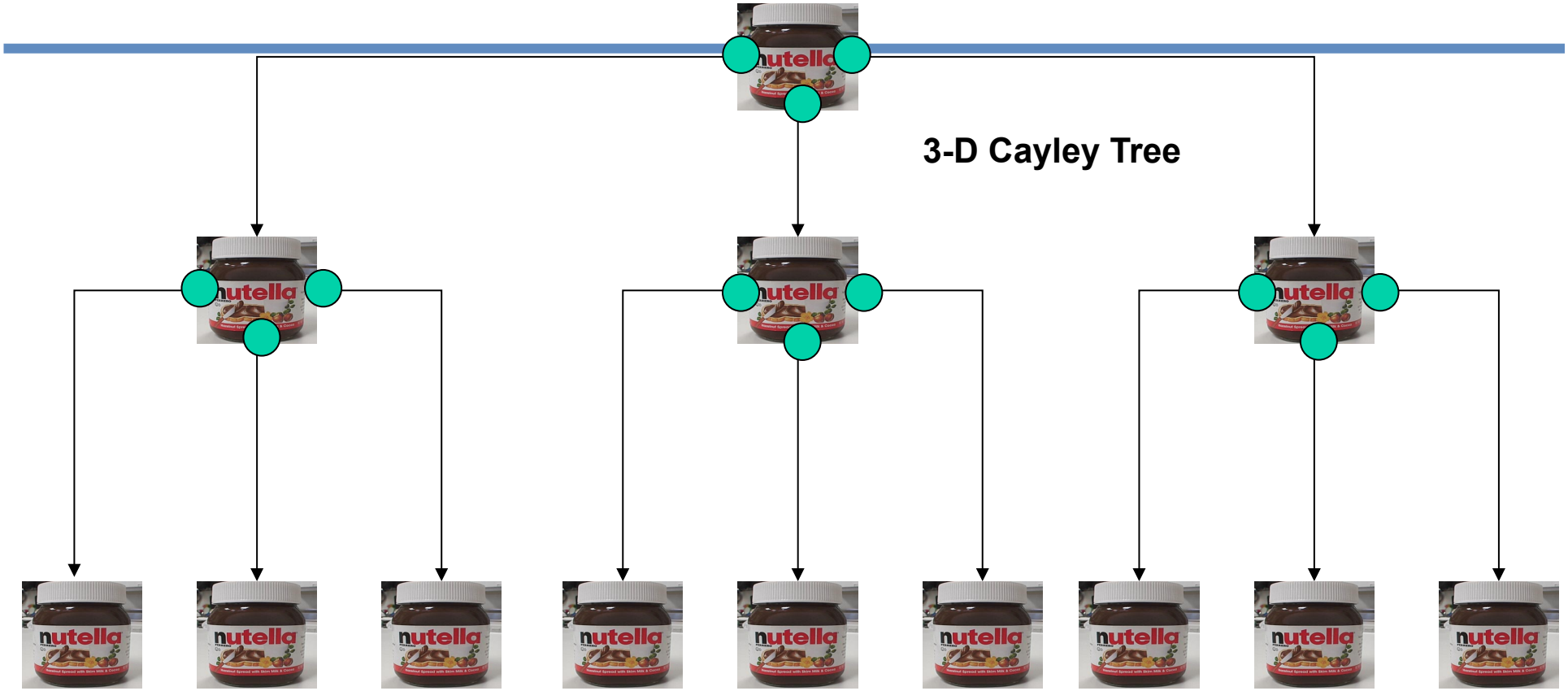


Horizon: how many hops a packet can go before it dies
(default setting is 7 in Gnutella)

Hops: a hop is a pass through an intermediate node



Searching a Gnutella Network: Broadcasting



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 + c2 +c3 + ….. ch = 3 + 9 + 27 + 81 + 243 + 729 + 2187 = 3279 Nodes

Gnutella Descriptors

- Gnutella messages that are passed around the Gnutella network

5 Descriptor Types

•**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.

•**Pong**: the response to a *Ping*.

(Each Pong packet contains a Globally Unique Identifier (GUID) plus address of *servent* and information regarding the amount of data it is making available to the network)

•**Query**: the primary mechanism for searching the distributed network. A *servent* receiving a *Query* descriptor will respond with a *QueryHit* if a match is found against its local data set.

•**QueryHit:** the response to a *Query*: contains IP address, GUID and search results

•**Push:** allows downloading from *firewalled servents*

Gnutella Scenario 1

Step 0: Join the network

Step 1: Determining who is on the network

- "Ping" packet is used to announce your presence on the network.
- Other peers respond with a "Pong" packet.
- Also forwards your Ping to other connected peers
- A Pong packet also contains:
  - an IP address
  - port number
  - amount of data that peers is sharing
  - Pong packets come back via same route

Gnutella Scenario 2

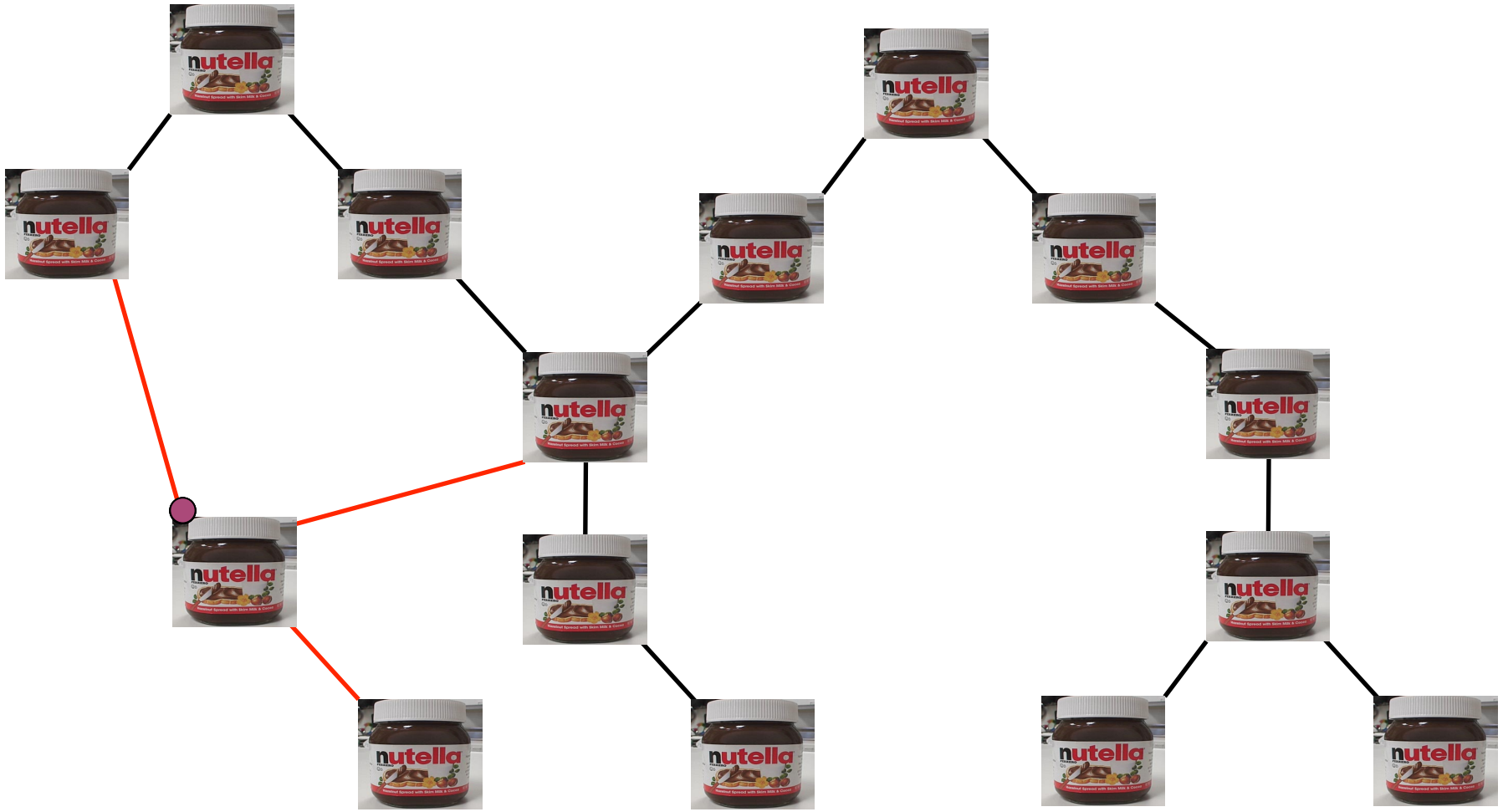
Step 2: Searching

- Gnutella is a protocol for distributed search.
- Gnutella "Query" ask other peers if they have the file you desire (and have an acceptably fast network connection).
- A Query packet might ask, "Do you have any content that matches the string ‘Homer"?
- Peers check to see if they have matches & respond (if they have any matches) & send packet to connected peers
- Continues for TTL

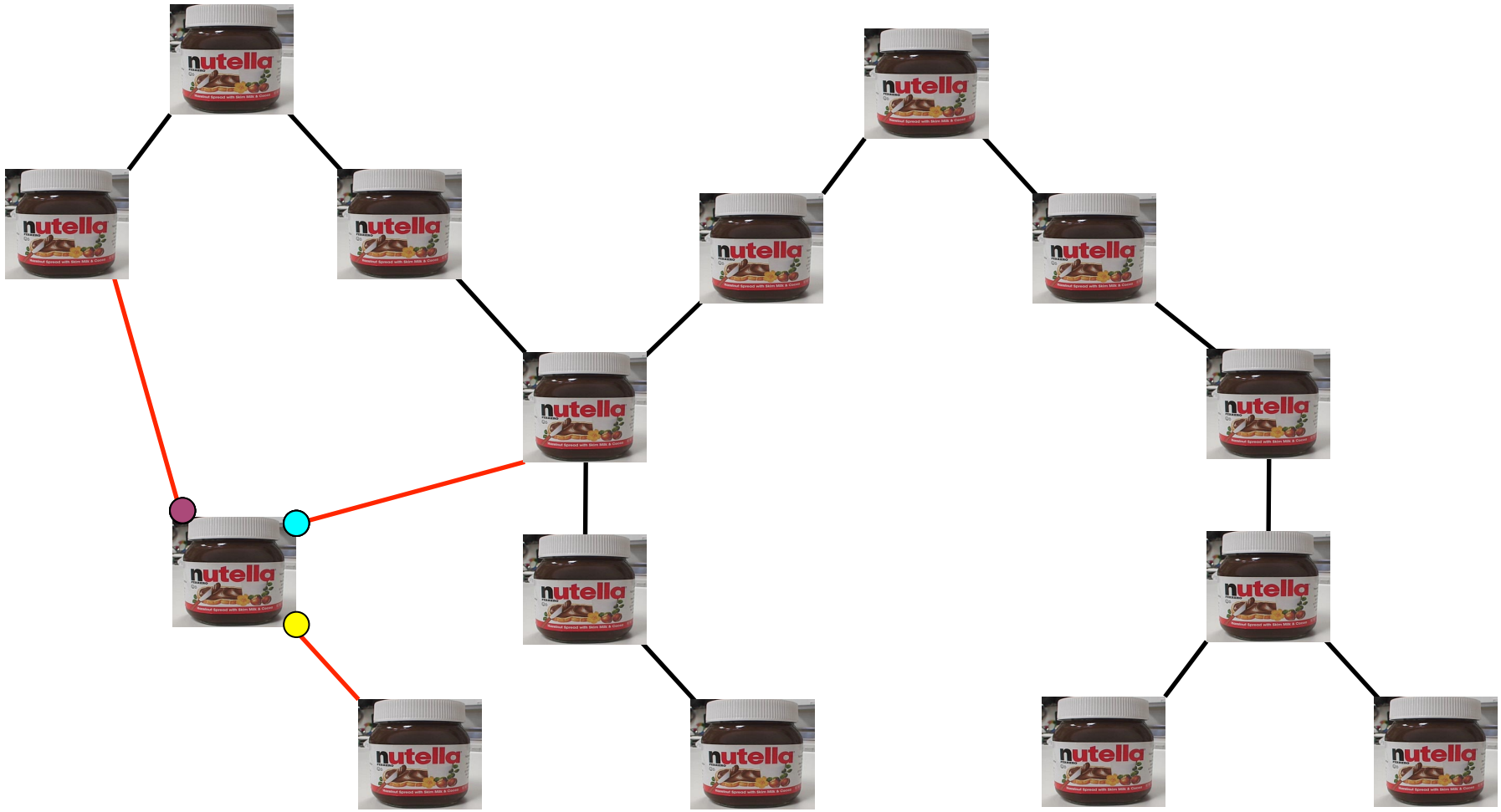
Step 3: Downloading

- Peers respond with a “QueryHit” (contains contact info)
- File transfers use direct connection using HTTP protocol’s GET method
- When there is a firewall a "Push" packet is used – reroutes via Push path

Searching a Gnutella Network: From one Node



Searching a Gnutella Network: All nodes



Discovering Peers

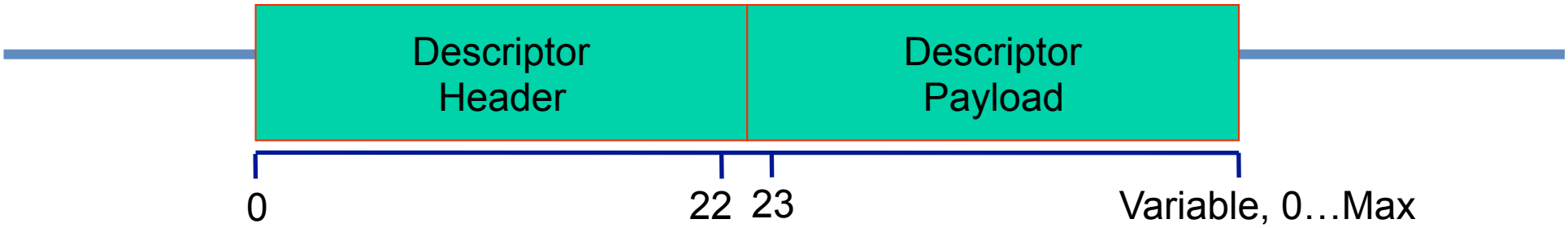
- In the early days, used ‘out of bounds’ methods:
  - IRC (Internet Relay Chat) and asked users for hosts to connect to
  - Web pages – users checked a handful of web pages to see what hosts were available.

Users typed hosts into the Gnutella software until one worked…….

- Host Caches: e.g. *GWebCache* is used to cache Gnutella hosts. Basically a HTTP server which responds to queries for IP addresses.

- Dynamically: by watching PING and PONG messages noting the addresses of peers initiating queries.

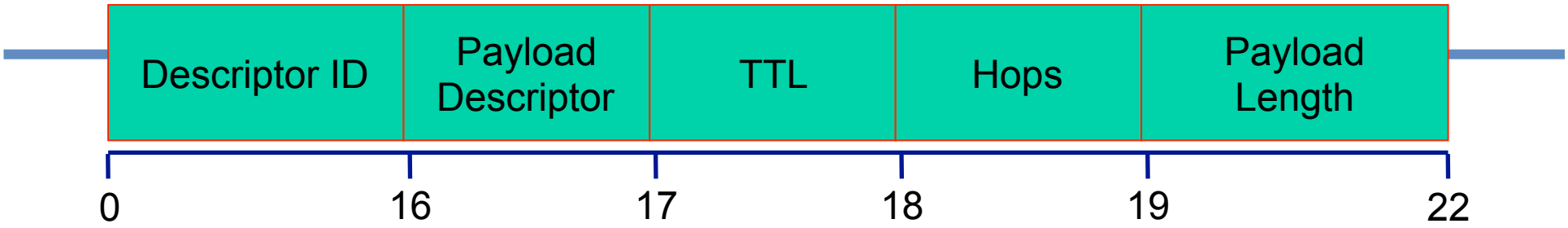
Gnutella Descriptors



Descriptor Types

- **Ping:** to actively discover hosts on the network.
- **Pong**: the response to a *Ping (*includes the GUID, address of a connected *servent* and information regarding the amount of data it is making available to the network)
- **Query**: search mechanism
- **QueryHit:** the response to a *Query* (containing GUID and file info)
- **Push:** mechanism for *firewalled servents*

Gnutella Descriptor Header



- **Descriptor ID**: a unique identifier for the descriptor on the network (16-byte string)

- **Payload Descriptor**: *0x00 = Ping: 0x01 = Pong: 0x40 = Push: 0x80 = Query: 0x81 = QueryHit*

- **TTL:** *Time To Live or Horizon.* Each *servent* decrements the TTL before passing it on - when TTL = 0, it is no longer forwarded.

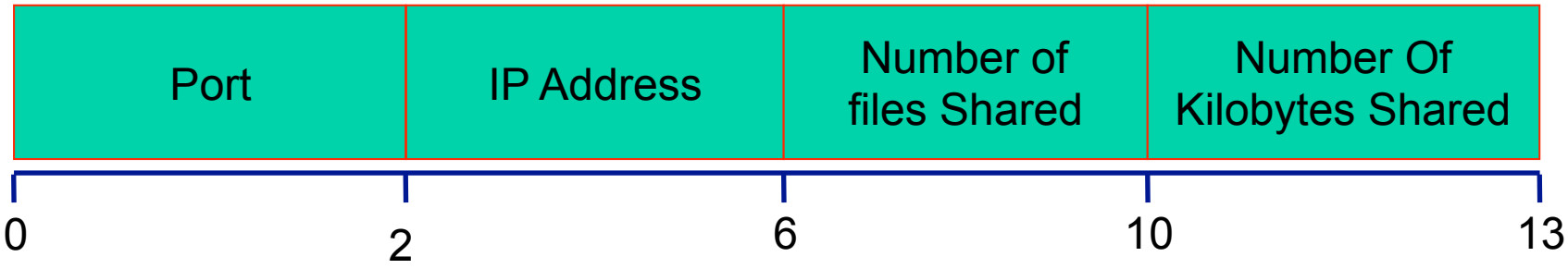
- **Hops:** counts the number of hops the descriptor has traveled i.e. hops = initial TTL when TTL expires

Payload Length: next descriptor header is located exactly *Payload Length* bytes from end descriptor header

Gnutella Payload 1 – Ping Descriptor

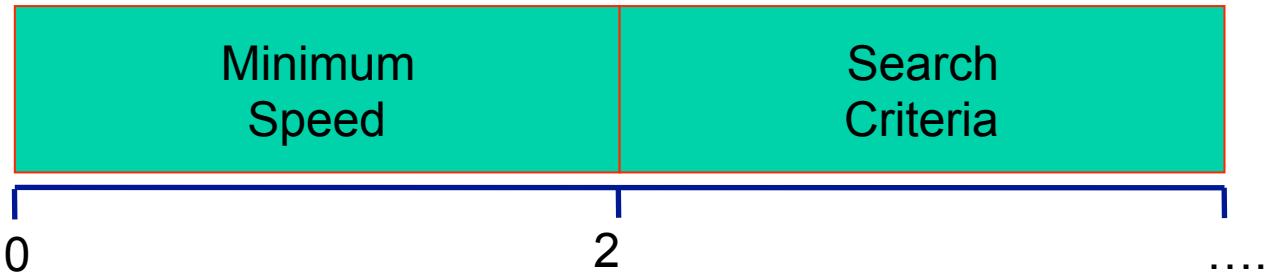
- **Ping descriptors:**
  - no associated payload
  - = zero length
- A Ping is simply represented by a ***Descriptor Header*** whose:
  - **Payload_ Length** field is 0x00000000.
  - ***Payload_Descriptor*** field = 0x00

Gnutella Payload 2 - Pong



- **Port:** port which responding host can accept *incoming* connections.
- **IP Address:** IP address of the responding host (big-endian)
- **Number of Files Shared**: number of files responding host is sharing on the network
- **Number of Kilobytes Shared**: kilobytes of data responding host is sharing on the network.

Gnutella Payload 3 - Query



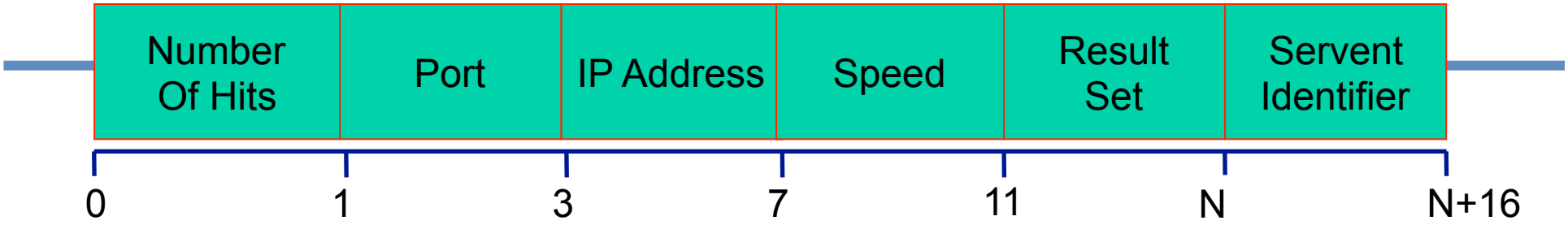
• **Minimum Speed:** minimum speed (in kb/second) of *servents* that should respond to this message.

  • A Servent receiving a Query descriptor with a minimum speed field of n kb/s should only respond with a QueryHit if it is able to communicate at a speed >= n kb/s

• **Search Criteria:** A nul (i.e. 0x00) terminated search string - maximum length is bound by *Payload_Length* field of the descriptor header.

  • e.g. “myFavouriteSong.mp3”

Gnutella Payload 4 - QueryHit

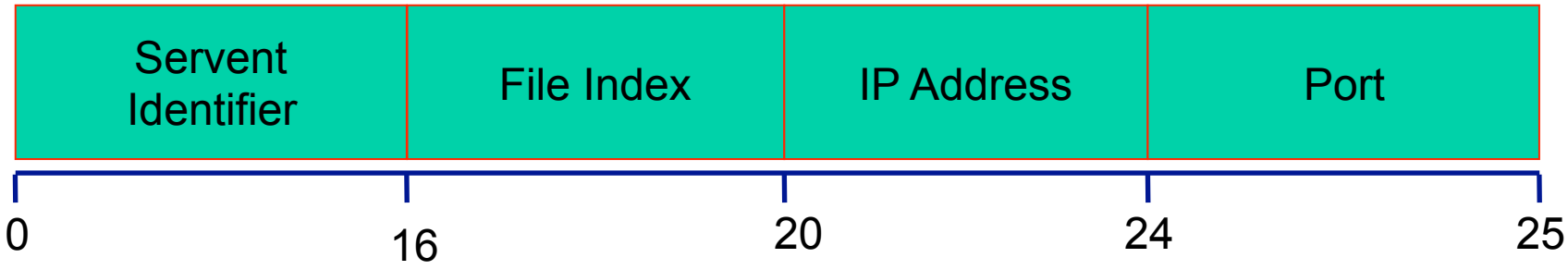


- **Number of Hits:** number of query hits in the result set
- **Port:** port which the responding host can accept incoming connections
- **IP Address:** IP address of the responding host (big-endian)
- **Speed:** speed (in kb/second) of the responding host
- **Result Set:** set of *Number_of_Hits* responses to the corresponding Query with the following structure:

- **File Index:** ID of file matching the corresponding query - assigned by the responding host
- **File Size:** size (bytes) of this file
- **File Name:** name of the file (double-nul (i.e. 0x0000) terminated)

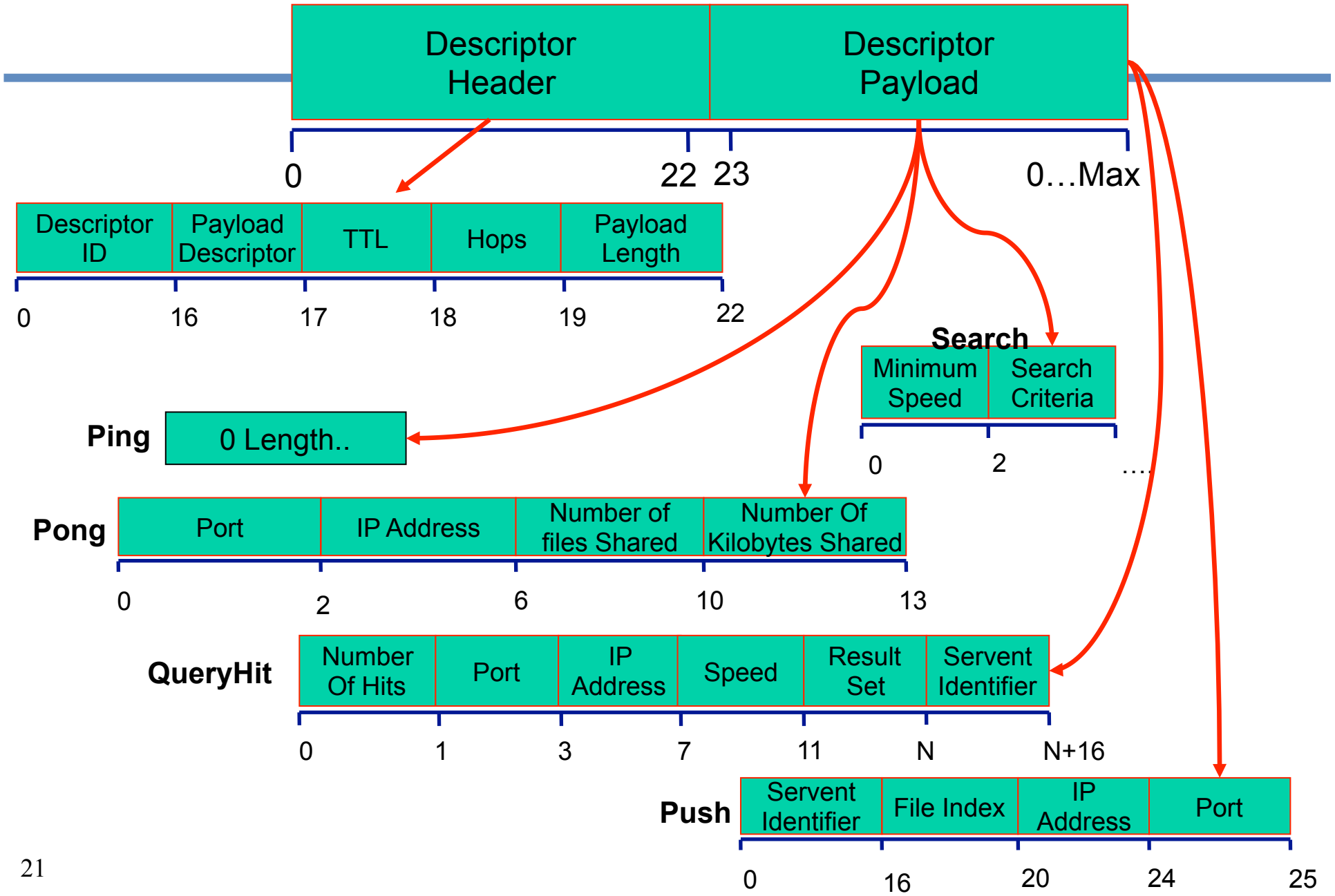
- **Servent Identifier:** servent network ID (16-byte string), typically function of servent's network address - instrumental in the operation of the ***Push Descriptor*** ….

Gnutella Payload 5 - Push



- **Servent Identifier:** target *servent* network ID (16-byte string) requested to push file (with given index *File_Index*)
- **File Index:** ID of the file to be pushed from the target *servent*
- **IP Address:** IP address of target host which file should be pushed (big-endian forma)
- **Port:** port on target host which file should be pushed

Gnutella Descriptor



Closing Remarks

1. Gnutella Background
   - ✓ the name, history, what is it?

2. Gnutella, In Operation
   - ✓ Organizing a Gnutella Network
   - ✓ Searching Gnutella for peers and files
   - ✓ Peers are discovered by IRC, GWebCache message monitoring

3. Gnutella Protocol
   - ✓ Gnutella Descriptors consists of a header and a payload
   - ✓ There are 5 types of payload: Ping, Pong, Query, QueryHit, Push

4. Things to Know
   - ✓ Gnutella scenario – joining, discovering and searching Gnutella networks
   - ✓ Know difference with Napster