



Computers and Society

File Sharing

Chris Brooks

Department of Computer Science

University of San Francisco



Copyright

- Meant to protect individual works from being copied
- Applies to:
 - Written works (Books, articles, web pages ,etc)
 - Visual works (Movies, television, art, photographs, etc)
 - Audio works (music, performances, etc)
 - Software
- Copyright is intended to allow a creator to profit from his/her creative works.

Rights of the copyright holder

- A Copyright provides the following exclusive legal rights:
 - The right to reproduce and sell copies of their work.
 - the right to produce derivative works
 - The right to perform or display the work publically
 - The right to sell or deny these rights to others.
- In a nutshell, copyright allows you to specify how your work should be used, who may distribute it, and who may sell it.

The Role of the Music Industry

- Typically, an artist is represented by a label.
 - The label may own some or all of the rights to the artist's work
- That label is responsible for promoting and distributing an artist's work.
- When an artist is signed, he/she receives an advance against sales to record and tour.
- An artist then receives royalties from the sale of the recording.
 - The advance must be paid back first.
- Many people feel that major labels are exploitative, and that too little of the sale of music goes to the artist.

File sharing

- File sharing apps such as Napster, Gnutella, and BitTorrent have changed the landscape for media companies.
- Made it possible for artists and creators to easily self-distribute.
- Made it very easy to pirate music, movies, software.

First-generation applications

- So-called first-generation file-sharing applications use a centralized name server.
- All user addresses and file indexes are kept there.
 - This can be used to build an application-level address-resolution protocol.
- Napster is the canonical example of this.
- Most IM programs also work this way.
- Central server is used to search.
- Files are transferred between peers.

Weaknesses

- Central server provides a bottleneck.
- Difficult to implement in a closed or dynamic environment.
- Scalability can be a problem.
- (also, legal issues in the case of Napster)

Second-generation models

- Second-generation models remove the central server.
- All file information is distributed.
- Advantage:
 - Potentially more scalable
 - Legal issues avoided.
- Disadvantages:
 - Search is more complex
 - Discovery of peers more complicated.

Example: Gnutella

- Gnutella is an example of an open P2P system.
- Many clients use the same protocol
 - LimeWire, BearShare, Gnucleus, etc
- Gnutella uses no centralized server - all information is stored at the peers.

Example: Gnutella

- The Gnutella network is completely decentralized.
- When a peer comes onto the network, it sends a ping to all known peers.
- Those peers respond with a pong, which identifies all their known peers.
- Your peer can also remember previous connections.

Example: Gnutella

- When a peer wants to perform a search, it sends a query to all known nodes.
 - This request is then forwarded on to nodes one level away, who forward it to all nodes they know about, and so on.
 - This is known as query flooding.
 - If a hit is found, the node containing the file contacts the searcher and download begins.

Example: Gnutella

- Problems with Gnutella:
 - Searching is unreliable - the network is often partitioned.
 - Search based only on keywords - this produces naming issues.
 - If I want “Lost S1, Ep 3”, what do I search for?
 - Wastes bandwidth forwarding searches to all nodes.
 - Does not take advantage of network structure
 - Some peers may have better bandwidth, more information, or a more central network structure.

Hierarchical p2p

- Kazaa is an example of a hierarchical p2p system
- When a user enters the network, it connects to a meganode.
- This meganode forwards queries to all other nodes connected to it.
- It also collects file information for each client.
- Periodically, it also exchanges information with other meganodes.
- Again, searches typically do not reach the entire network.
- Higher-bandwidth nodes act as meganodes, process more traffic.

Hierarchical P2P

- This allows peers with greater resources to do more.
- Searches don't need to be propagated more than is necessary.
- Kazaa is a closed network, so most of its inner workings have been figured out through reverse engineering.

Issues with file distribution

- File integrity - how do you guarantee you're getting the file you think you're getting?
 - md5 or SHA hash of file contents
- How can files be shared anonymously?
 - Virtual addresses and randomized routing can hide identity.
- How to distribute large files?
 - Download chunks simultaneously from many peers at once.

Example: BitTorrent

- BitTorrent is a p2p network specifically designed for sharing large files.
 - Example: videos from the Jan 2005 tsunami
- Key ideas:
 - File is broken into lots of small pieces that can be independently downloaded
 - Users can upload pieces to other users while downloading.
 - ‘Free riders’ are punished, sharers are rewarded.

Example: BitTorrent

- Someone who wants to share a file creates a .torrent file.
 - File name, size, hash of each block.
 - Address of a tracker.
- The torrent is downloaded and the peer registers with the tracker, which provides a list of available peers and seeds.
- The peer begins requesting blocks, starting with the rarest available block.
- As it finishes receiving a block, it begins uploading those blocks to other peers.
- BitTorrent uses a tit-for-tat approach to sharing
 - Those four peers who have shared the most data in the last 10 seconds are unchoked.

Example: BitTorrent

- Difficulties with BitTorrent
 - Works best with files that are widely copied on the network.
 - In practice, files appear and disappear.
 - There is no permanent archive, or incentive for users to keep old files.
 - How to find torrents?
 - No central torrent repository
 - No metadata standard for searching and indexing torrents.
 - Since blocks are not downloaded sequentially, a partial file is not useful.

Legal Alternatives

- iTunes - pay per song, DRM included
 - Can now pay more for DRM-free
 - Apple is considering providing subscription access
- Rhapsody - subscription model
- eMusic - fixed number of songs per month. DRM-free.
- Amazon - pay per song, DRM-free.

Exercise

- Consider this problem from each of the following points of view. If you were representing this community, what do you think their concerns might be? What solutions might you propose to help address these?
 - Major label recording artists (Metallica, Madonna, etc)
 - Independent, “struggling” recording artists.
 - Publishing companies
 - Consumer advocates
 - “Remix” artists - DJs, rappers, etc.
 - Record labels
 - Music fans

Information goods

- So what's so unusual about information goods?
- Why is it so difficult to figure out how to buy and sell digital information?

Low marginal cost

- Information goods have high first-copy cost, but low or zero marginal cost.
 - It costs \$300 million to produce the first copy of Lord of the Rings, but the second copy is almost free.
 - This means that there's an incentive to be a large producer.
 - Increasing returns to scale

Nonrival

- Information goods are *nonrivalrous*
 - This means that the amount one person consumes has nothing to do with the amount available to others.
 - This has issues for sellers
 - Traditional price competition is based on *scarcity*
 - If there is a limited number of widgets available, those who want the widgets more will pay more.
 - Prices will rise or fall until the number of buyers matches the number of widgets available.
 - If there's an unlimited number of widgets available, no one will want to pay more than the lowest price.

Nonrivalry

- If we follow this plus low marginal cost to its natural conclusion, no one will want to pay more than zero for information.
- So how do you encourage people to buy what they can get for free?
- One solution: provide additional value.

Nonexcludable

- Information goods are also nonexcludable.
- This means that a producer can't keep consumers from getting the good without paying for it.
- This means that information can be thought of as a *public good*, like police, or streetlights, or roads.
- DRM is a way to try to provide excludability.
- Other approaches include indirect taxes.
 - The UK has a tax on televisions.

Experience

- Information goods are experience goods
 - This means that, until you consume it, you don't know how much it's worth to you.
- This makes it very hard for consumers to accurately determine how much they're willing to pay for information.
- Recommendations, reviews, reputation all become important value-added services.

Economic models

- How to sell information?
 - Per-item
 - Bundling
 - Value-added service
 - Advertising
 - Public subsidy