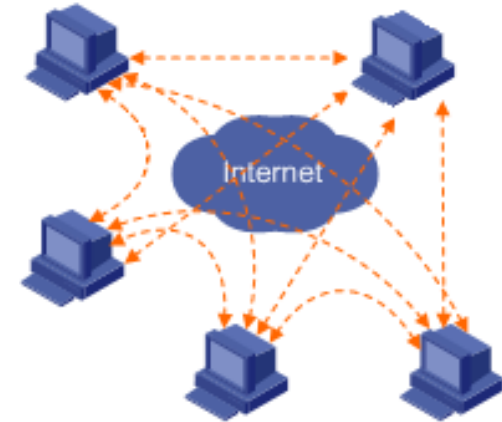


P2P Networks

by Octavio Herrera-Ruiz

P2P Definition

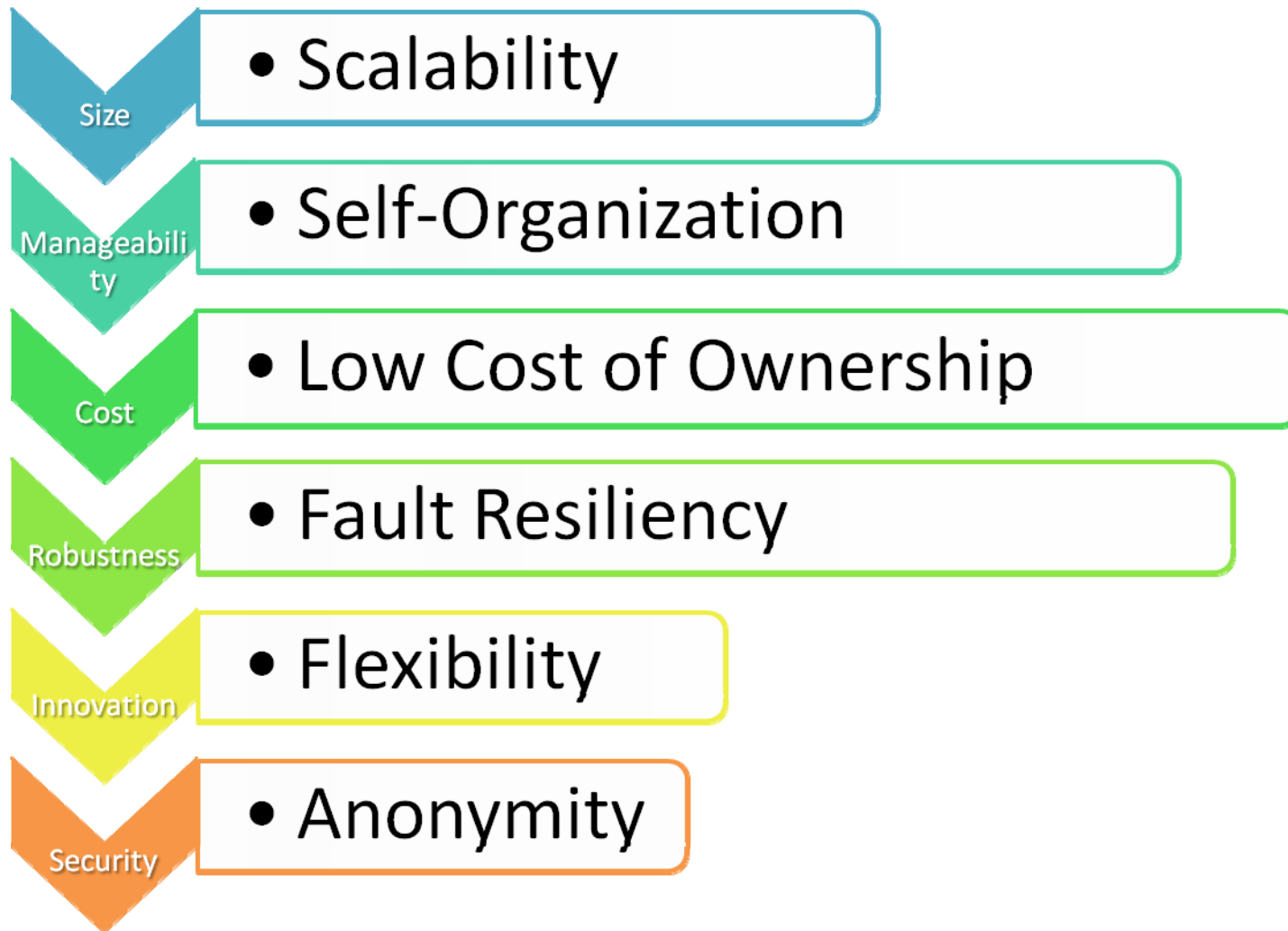


- ❖ P2P=Peer-to-Peer
- ❖ Distributed Systems
- ❖ Nodes are both: *Clients* and *Servers*
- ❖ Aggregate Resources
- ❖ Nodes organized into network topologies
 - o Nodes at the edge of the Internet
- ❖ Naming space independent of the DNS



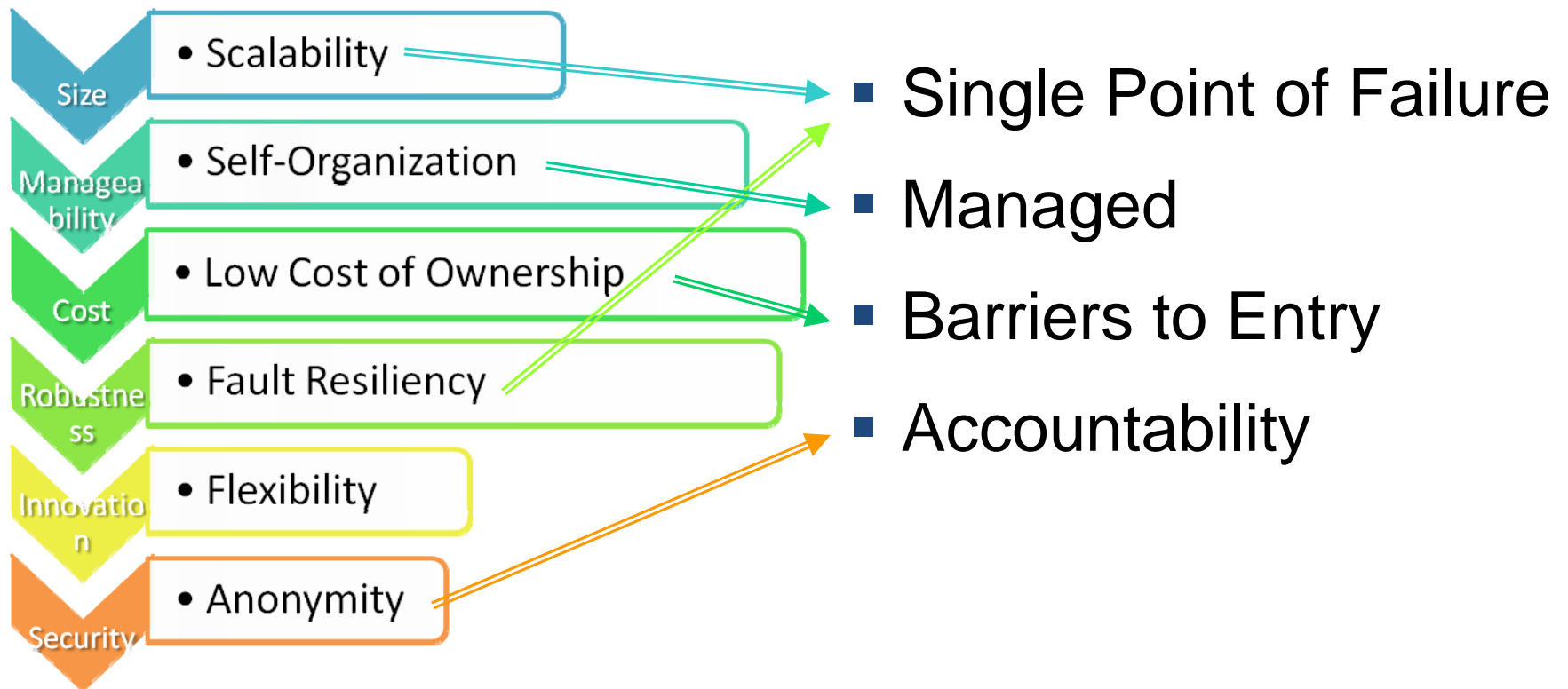
e-mail (SMTP)

P2P Features



P2P vs Client Server

Client Server

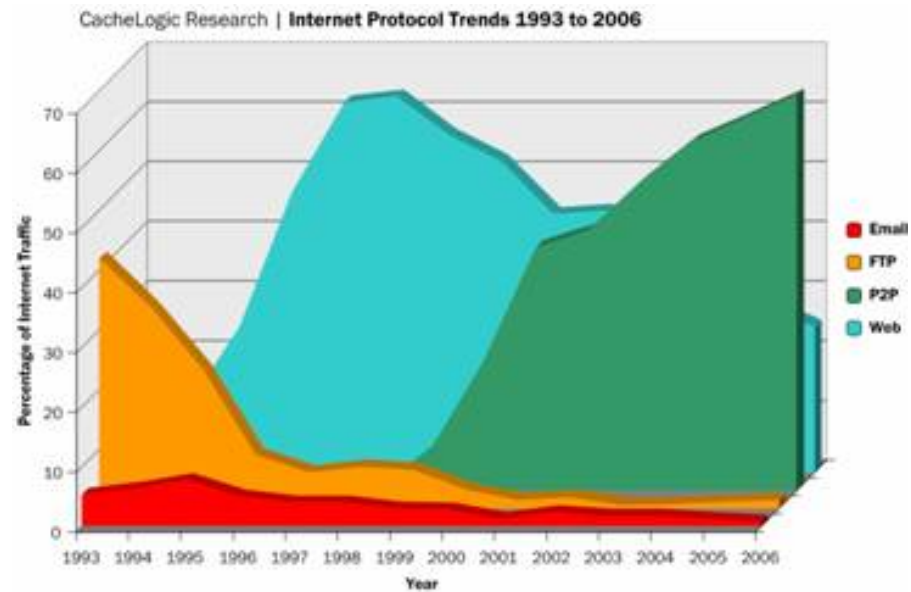


P2P vs Grid

Grid: large-scale coordinated use and sharing of geographically distributed resources

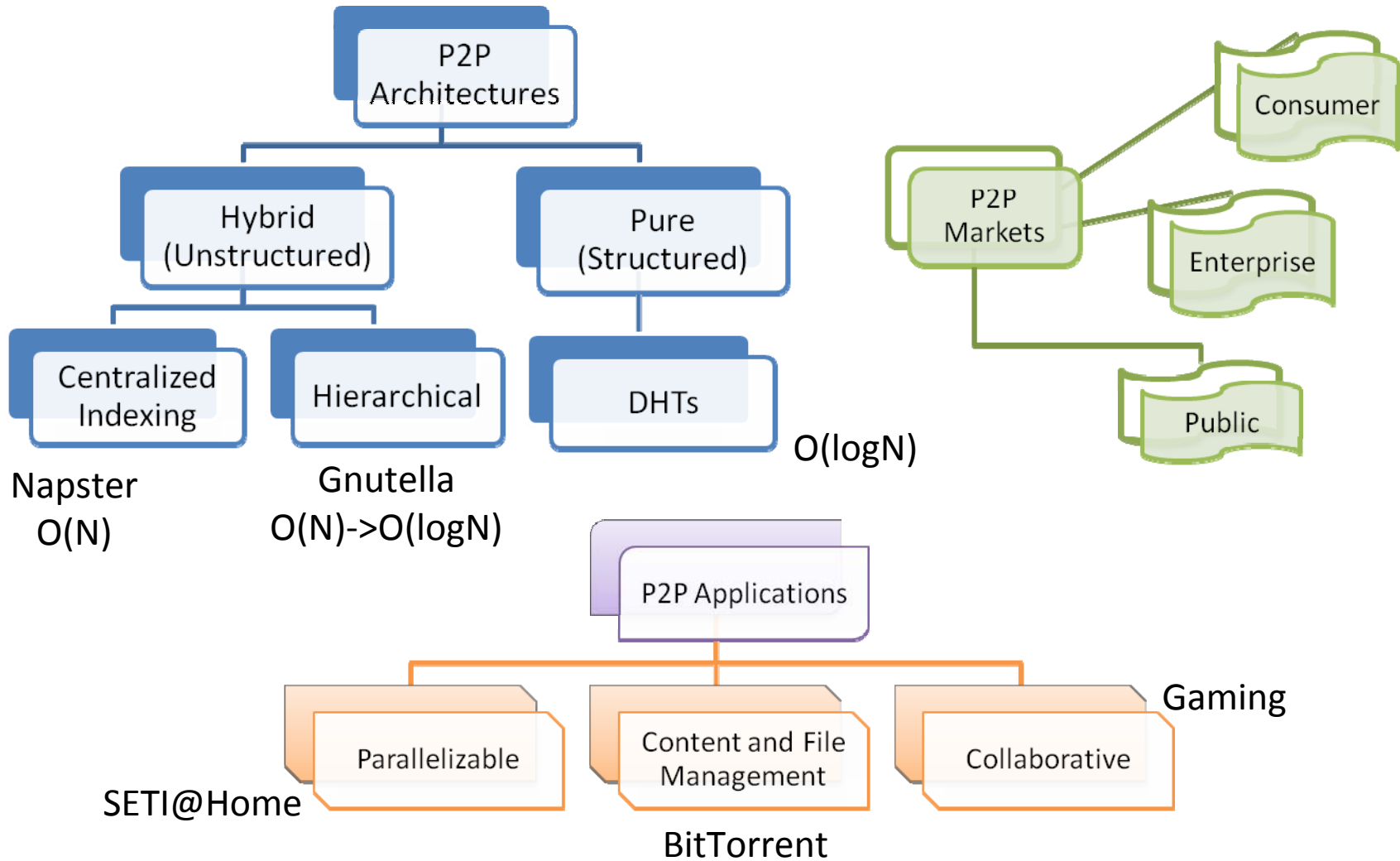
- *Node Availability*:
Random (**Low**)
- Best-effort
- Autonomy &
Anonymity
- *Node Availability*:
Server-like
- High performance
- Accountability &
Access Control

P2P Traffic



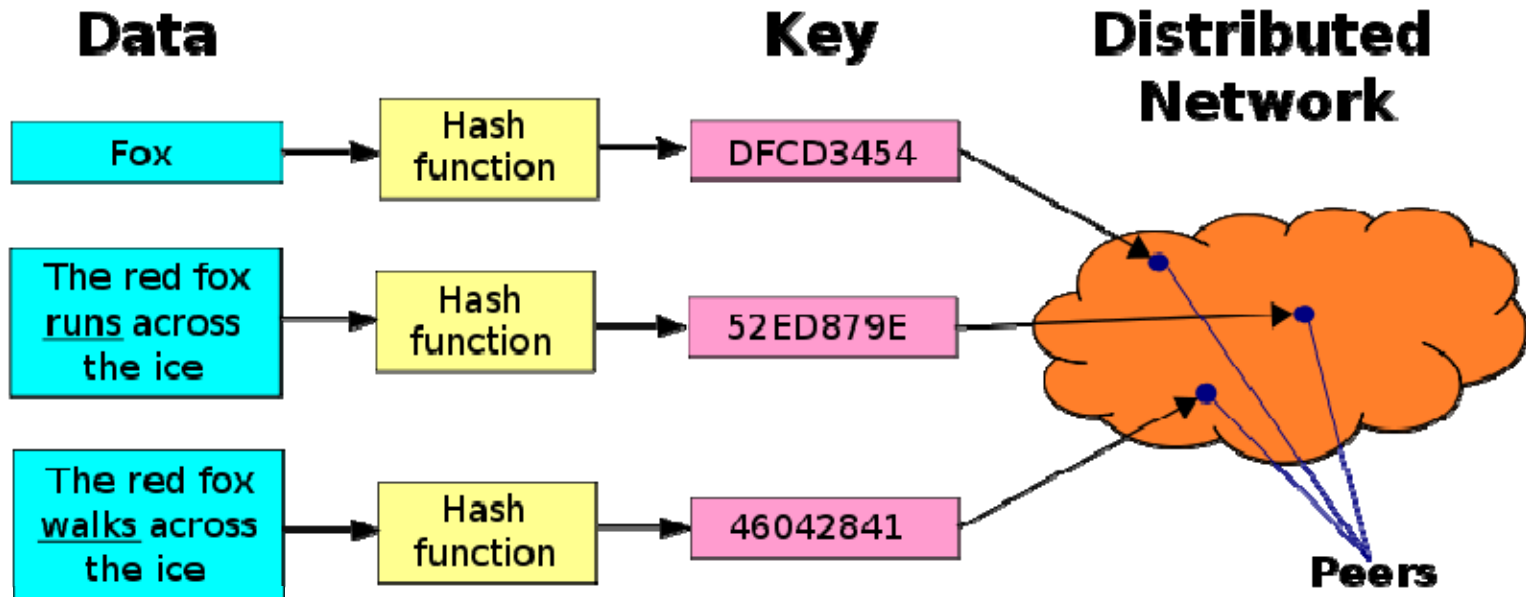
Source: Mihajlo A. Jovanovic, Fred S. Annexstein, Kenneth A. Berman, Scalability Issues in Large Peer-to-Peer Networks A case study of Gnutella. (<http://www.medianet.kent.edu/surveys/IAD06S-P2PArchitectures-chibuike/P2P%20App.%20Survey%20Paper.htm>)

P2P Taxonomies



DHT

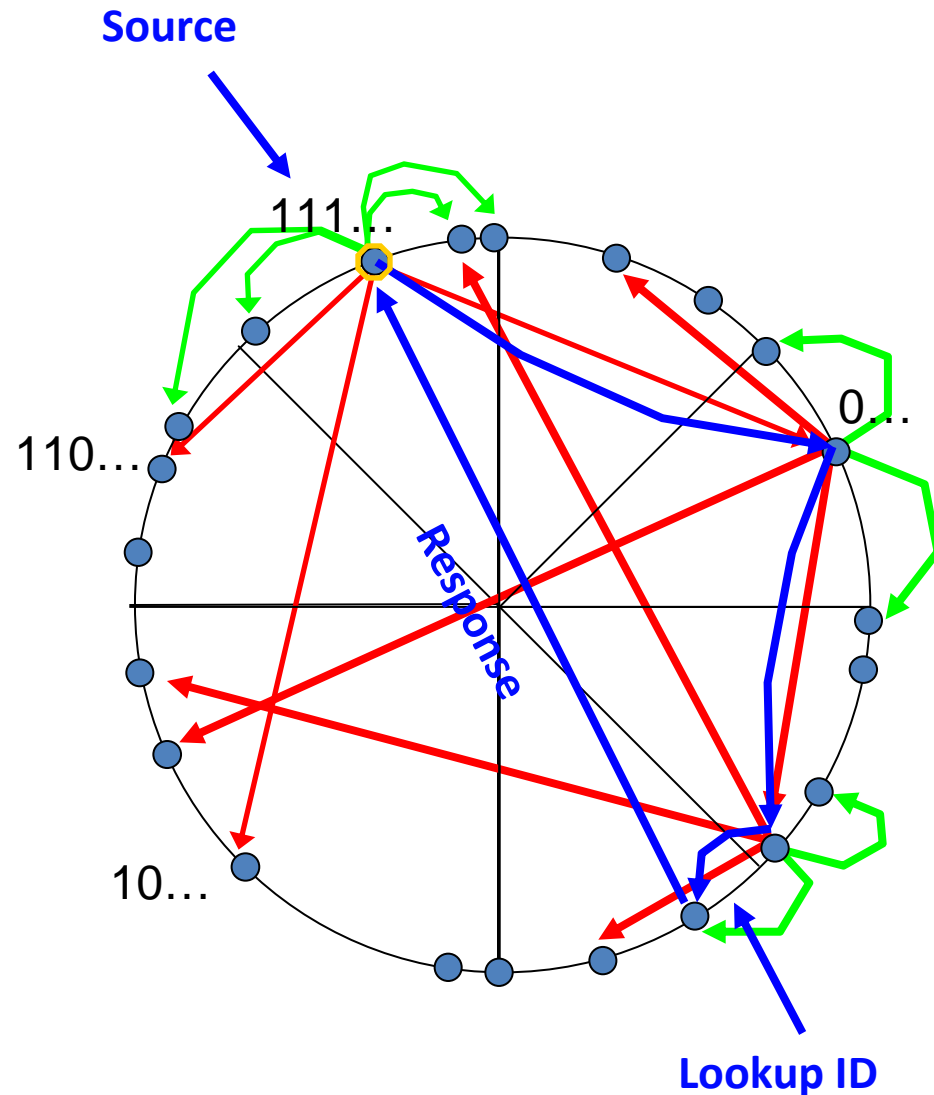
- Distributed Hash Tables



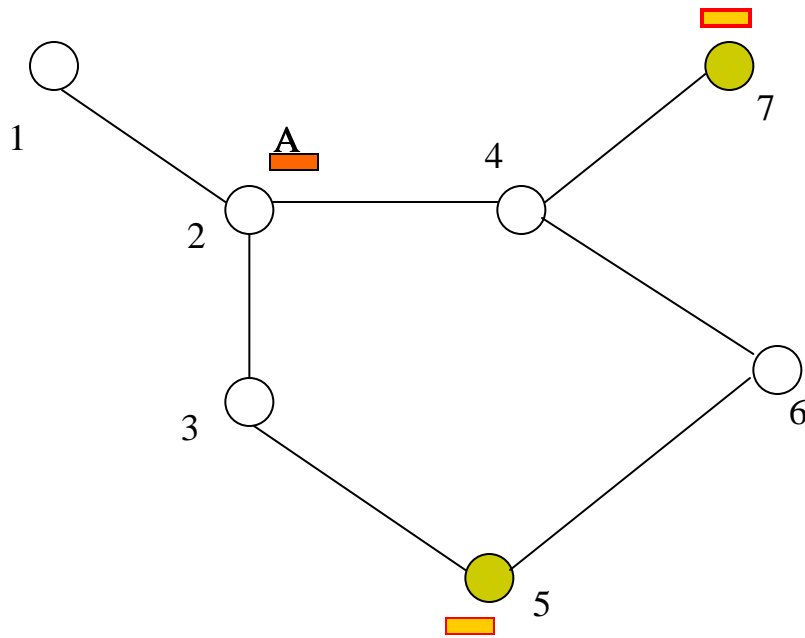
Source: Wikipedia

How Does Lookup Work?

- Assign IDs to nodes
 - Map hash values to node with closest ID
- Leaf set is successors and predecessors
 - All that's needed for correctness
- Routing table matches successively longer prefixes
 - Allows efficient lookups



Gnutella Operation



1. Node 2 initiates search for File A
2. Broadcast to all its neighbors
3. Neighbors forward the request (random walk)
4. Nodes that have file initiate reply
5. Reply is back propagated
6. P2P download

Research in P2P

- **Models/Tools**
 - Topologies
 - Node Behavior (on/off)
 - Simulation
- **Algorithms/Architectures**
 - Search
 - Security
 - Fairness
- **Measurement**
 - Heterogeneity
 - Size
 - IP aliasing and firewalls
- **Applications**
 - Commercial
 - Academic
- **Others**

Churn in P2P Networks

- **Churn**
 - Continuous and unsynchronized arrival and departures of nodes
- **Effects of churn**
 - Routing
 - Increased maintenance cost
 - Routing table inconsistencies
 - Delayed lookups
 - Content
 - Intermittent availability

Churn in P2P Networks

Alleviating Mechanisms

- Routing
 - Periodic maintenance
 - Dynamic timeouts
- Content
 - Redundancy
 - Bundling
 - Incentives

Limiting Factors

- Bandwidth
- Storage Space

1990 → 2005:

Disk increased by 8000-fold
while bandwidth increased
only 50-fold [2]

Redundancy Maintenance Cost

Erasure Coding

$$C_c = \left(1 + \frac{L_c}{k}\right) \frac{F}{\delta}$$

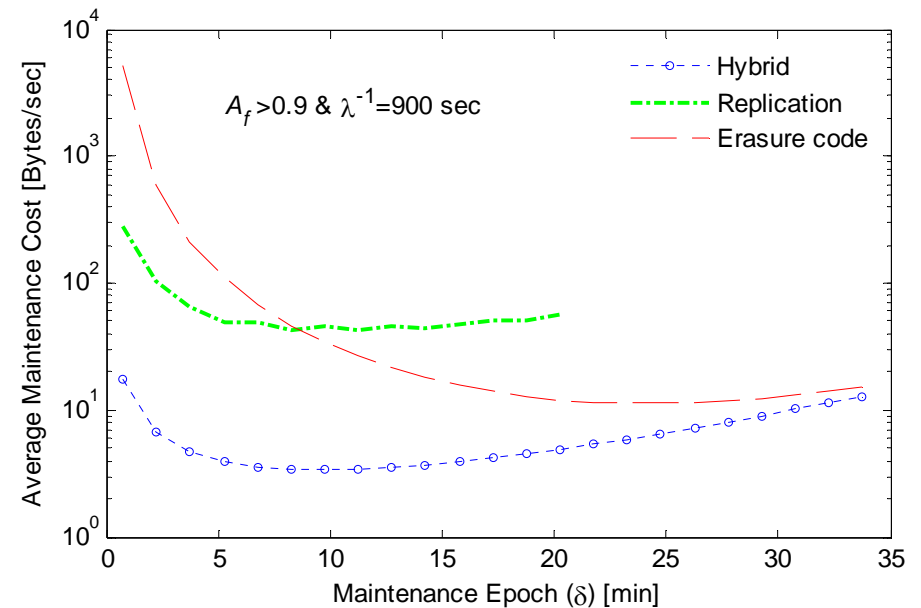
- L_i = # Segments lost
- K = # blocks needed to restore file
- F = File Size
- δ = Maintenance interval

Hybrid

$$C_h = \left(\frac{L_h}{k}\right) \frac{F}{\delta}$$

Replication

$$C_r = (L_r) \frac{F}{\delta}$$



Questions...

Thanks!

References

- [1] “Handling Churn in a DHT”, by Sean Rhea, Dennis Geels, Timothy Roscoe, and John Kubiatowicz, USENIX Annual Technical Conference, June 29, 2004
- [2] “High Availability, Scalable Storage, Dynamic Peer Networks: Pick Two” by Charles Blake and [Rodrigo Rodrigues](#). In *Ninth Workshop on Hot Topics in Operating Systems (HotOS-IX)*, (Lihue, Hawaii), May 2003, pp. 1-6.