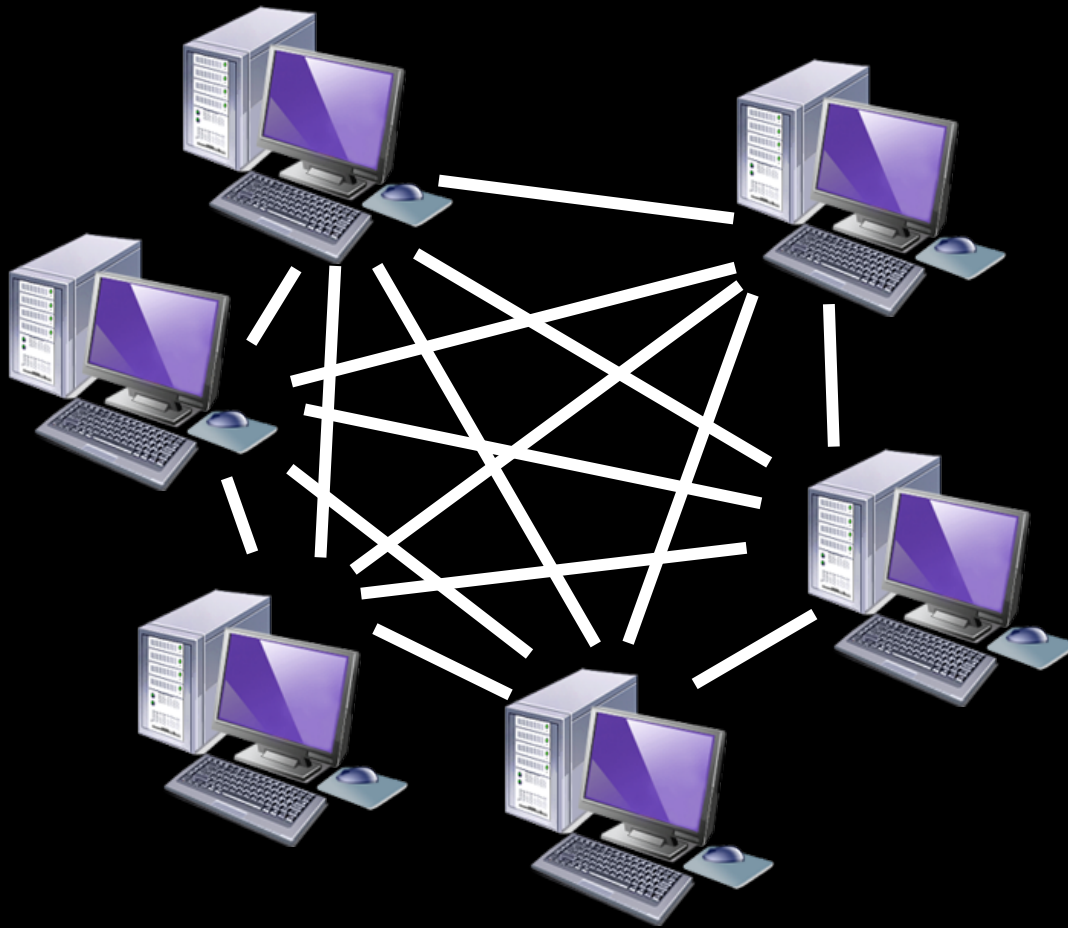


15-396

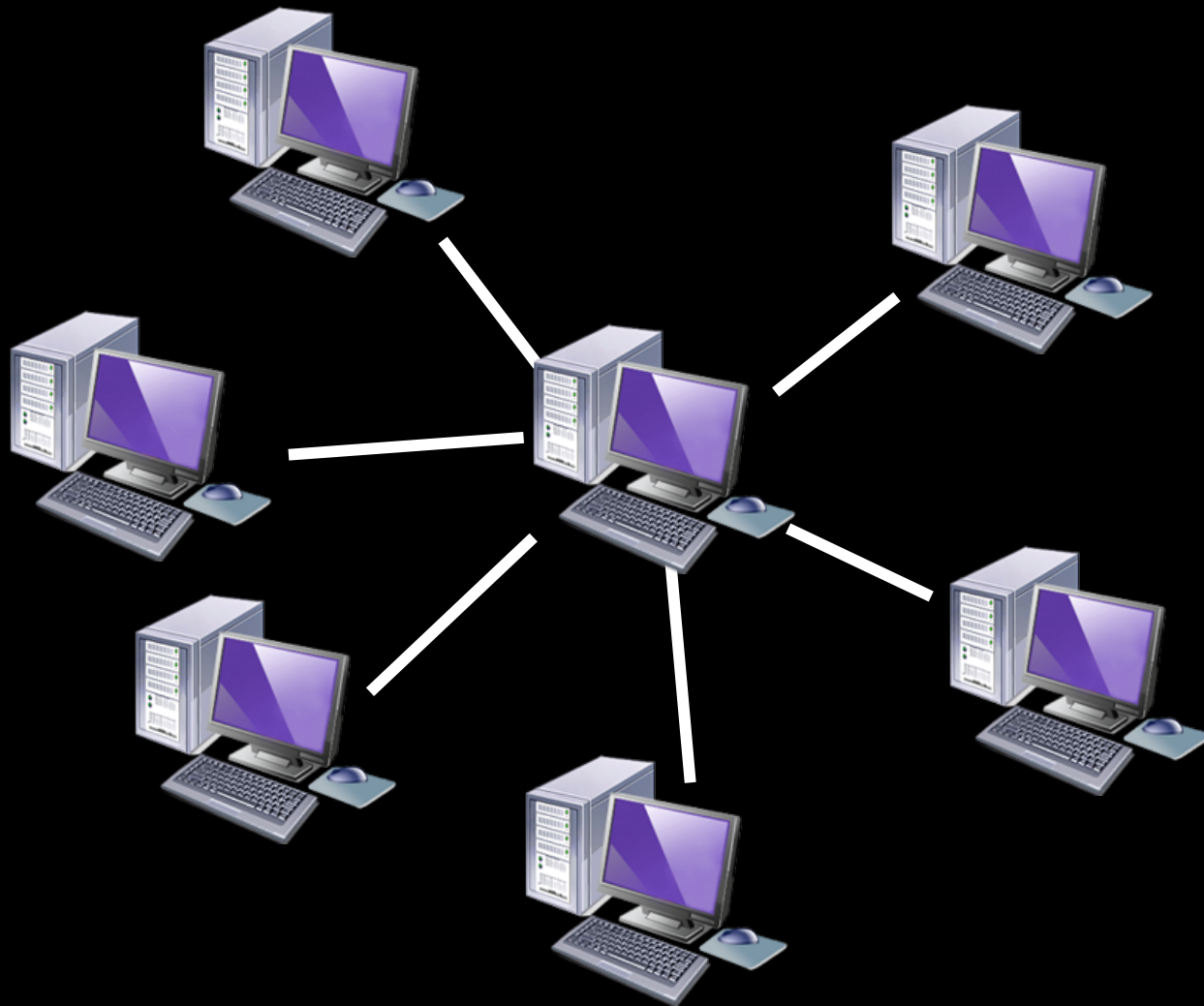
Science of teh Interwebs

P2P

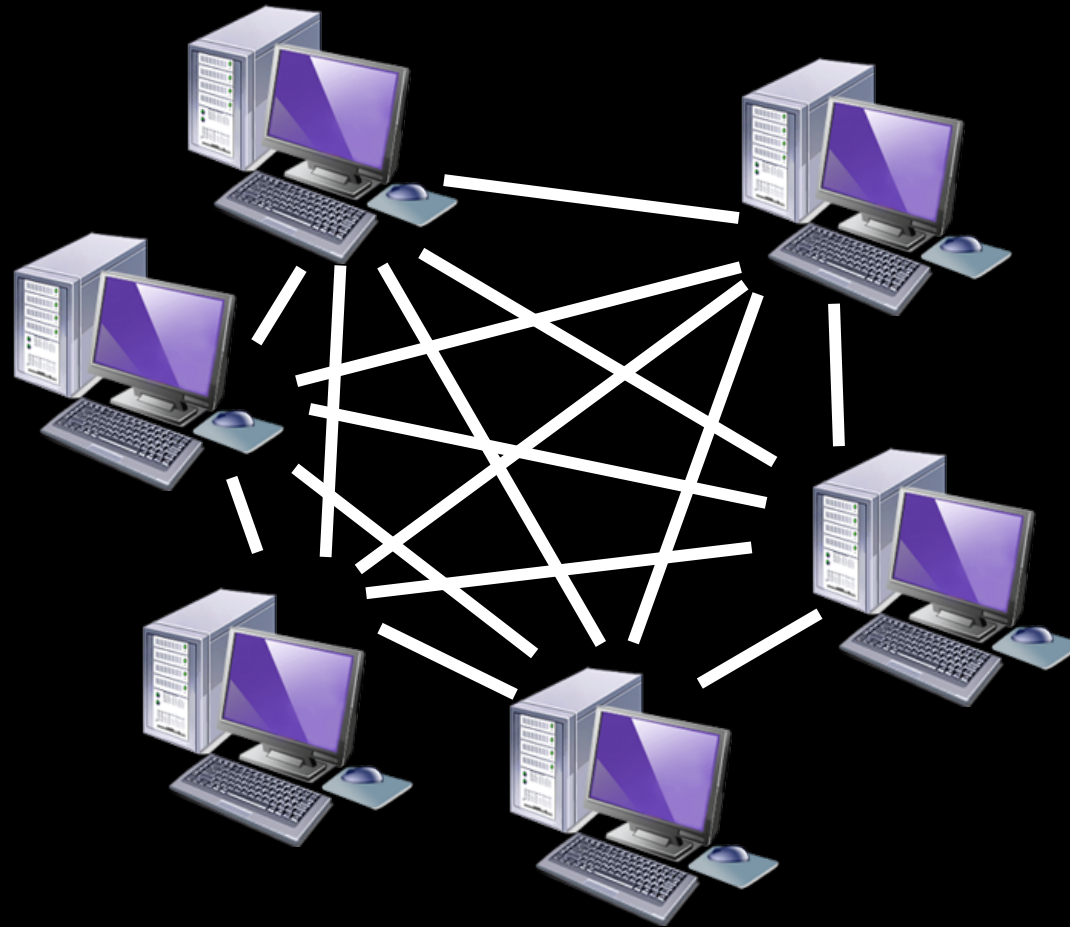
Lecture 18 (November 8, 2011)



Server-Based Network



Peer-to-Peer Network



What are the pros and cons of each?

Client-Server:

Easier to maintain

Faster Routing

P2P:

All clients provide resources (bandwidth, storage, etc.), so the system scales with itself

Robustness in case of failures

No one to blame in case of illegal content



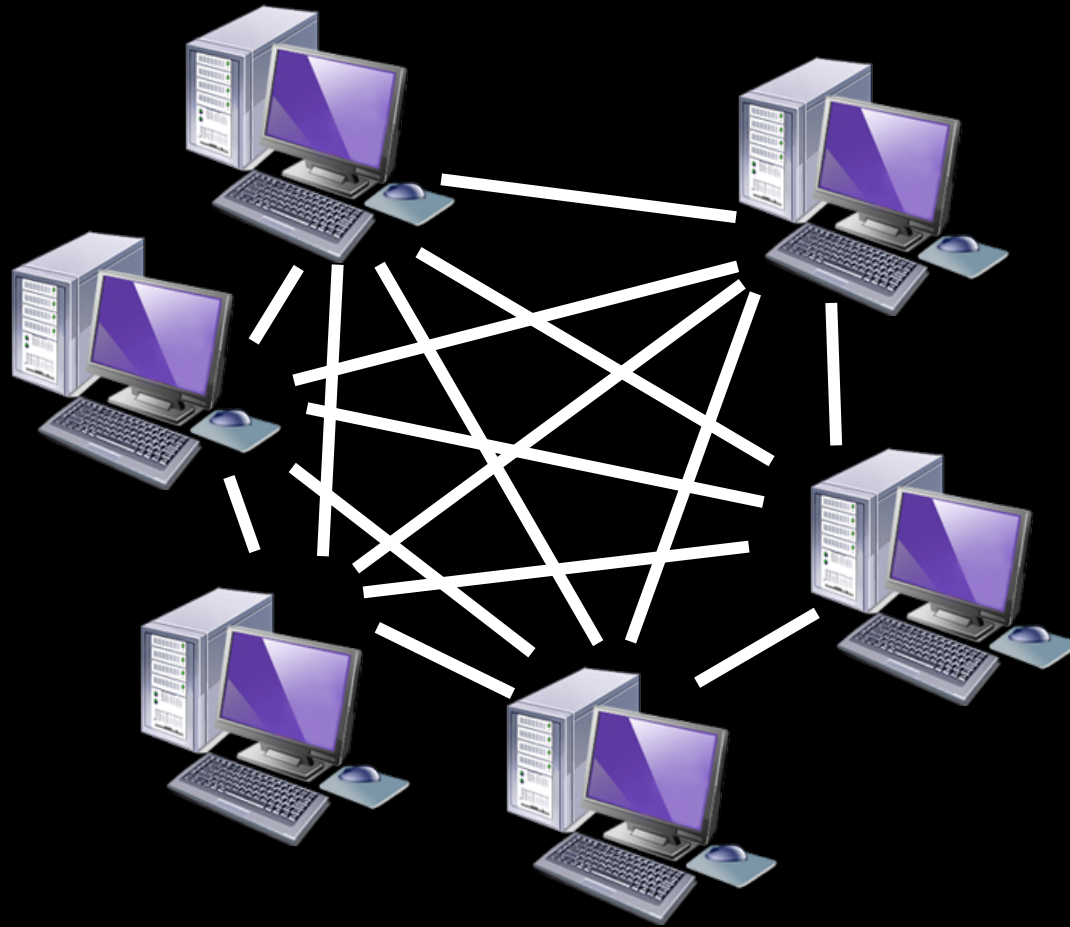
**Napster ran from June 1999
to July 2001**

**It was a hybrid between
P2P and centralized**

**Centralization was the
reason the music industry
shut it down**

**Without any centralized
control how do you
efficiently find a file?**

Bad Solution: Ask Everybody



The Chord Protocol

Chord consistently maps a key onto a node

Both keys and nodes will get an **m-bit** identifier

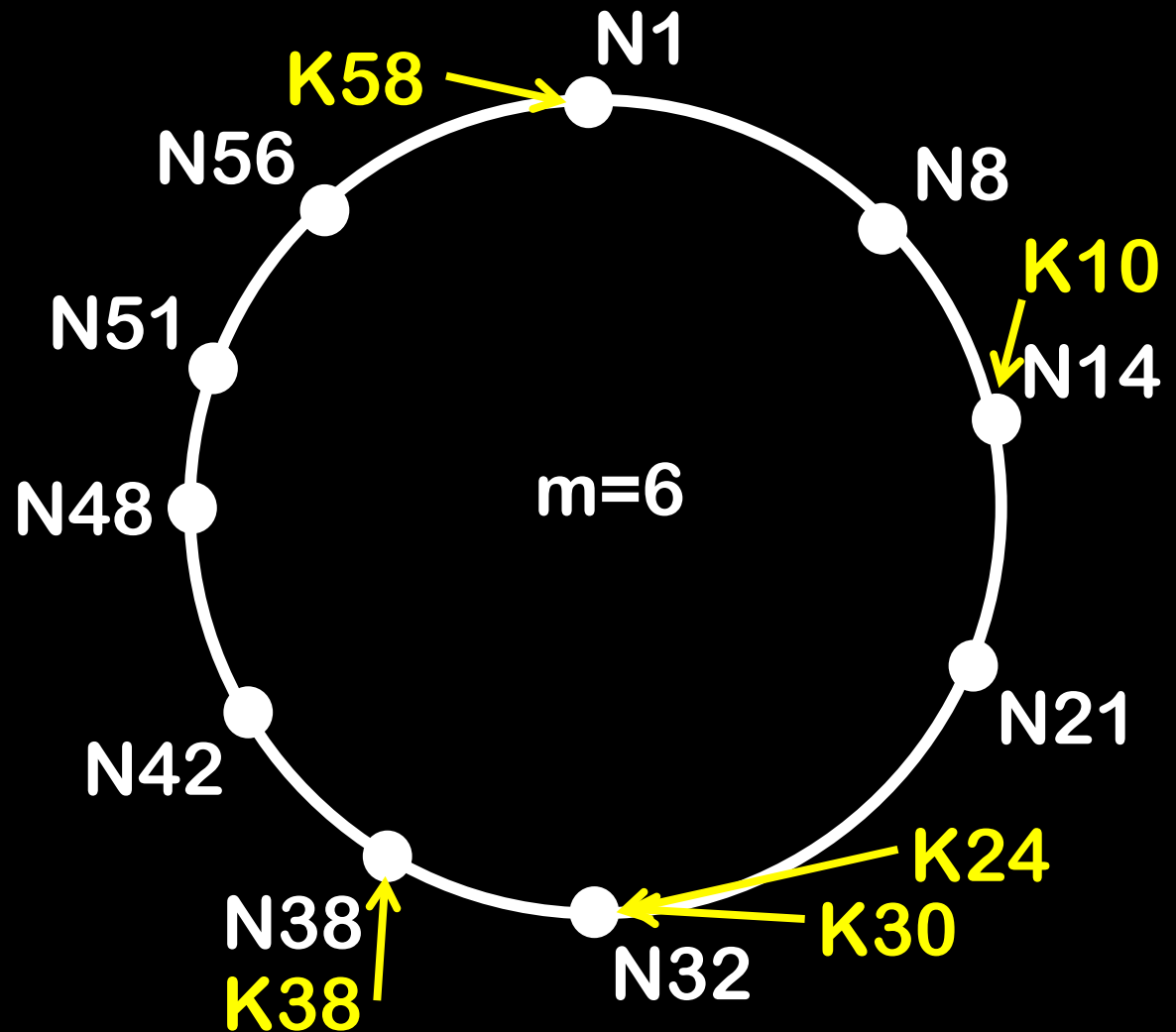
Node's identifier will be a hash of its IP address

Key's identifier will be a hash of the key

The Chord Ring

A ring with positions numbered 0 to $2^m - 1$

Key k is assigned to node successor $(k) = \text{node}$ whose identifier is equal to or follows the identifier of k



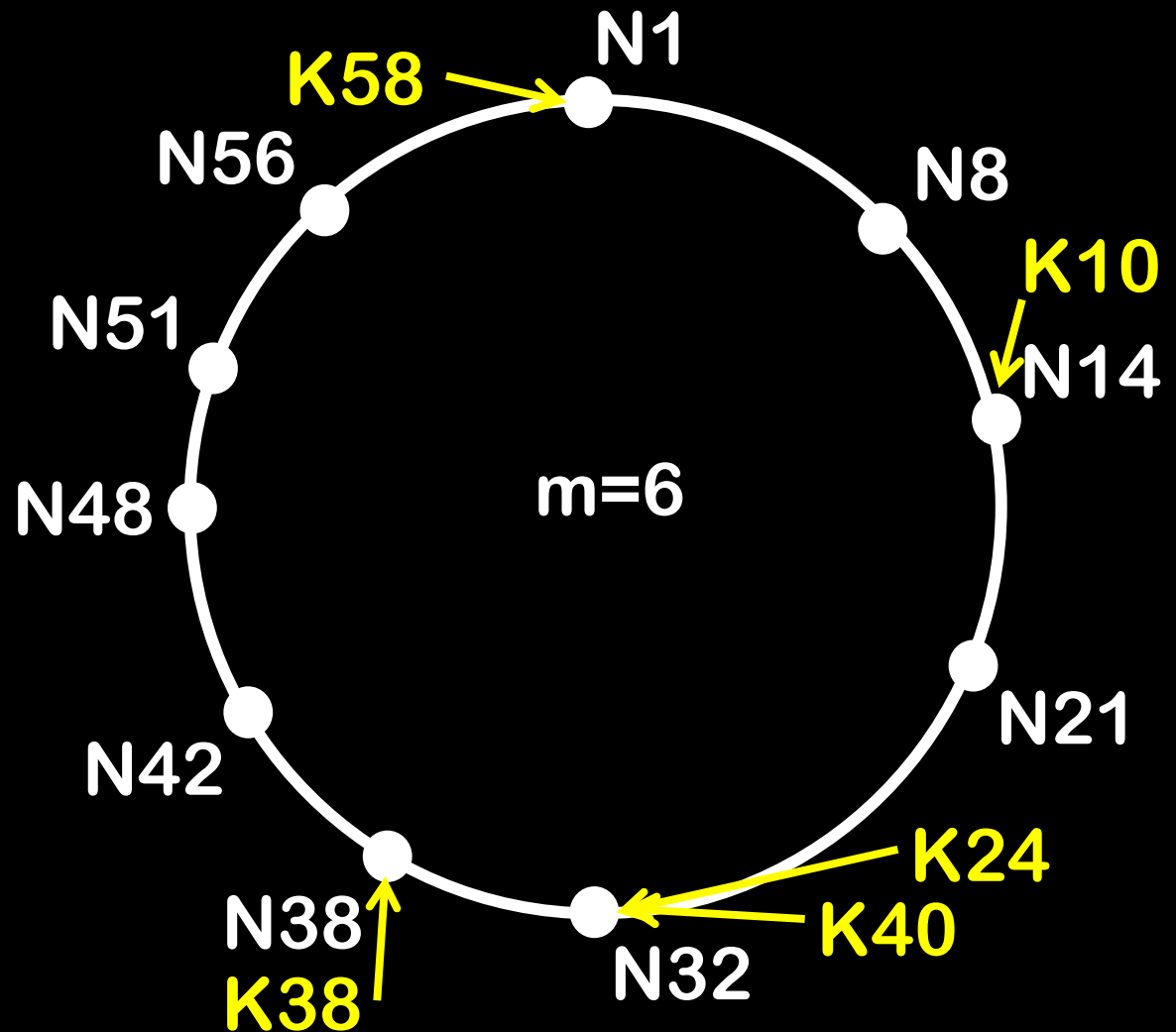
If there are N nodes and K keys, roughly how many keys is each node responsible for? K/N

When an $(N+1)^{\text{st}}$ node joins or leaves the network, responsibility for $O(K/N)$ keys changes hands

Each node knows the IP address of its successor

Linear Search for a Key

If each node knows the location of its successor, you can perform linear search

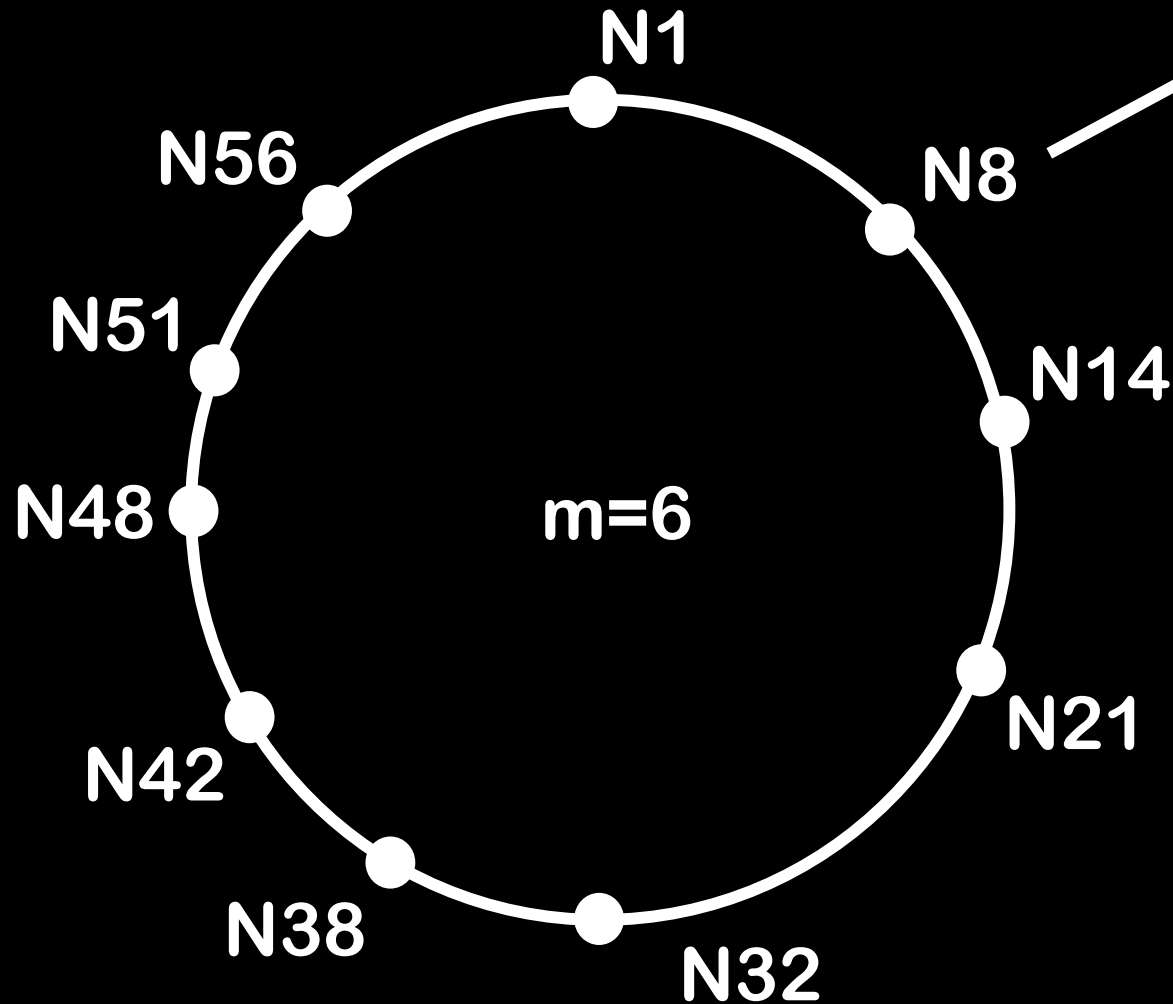


Faster Search

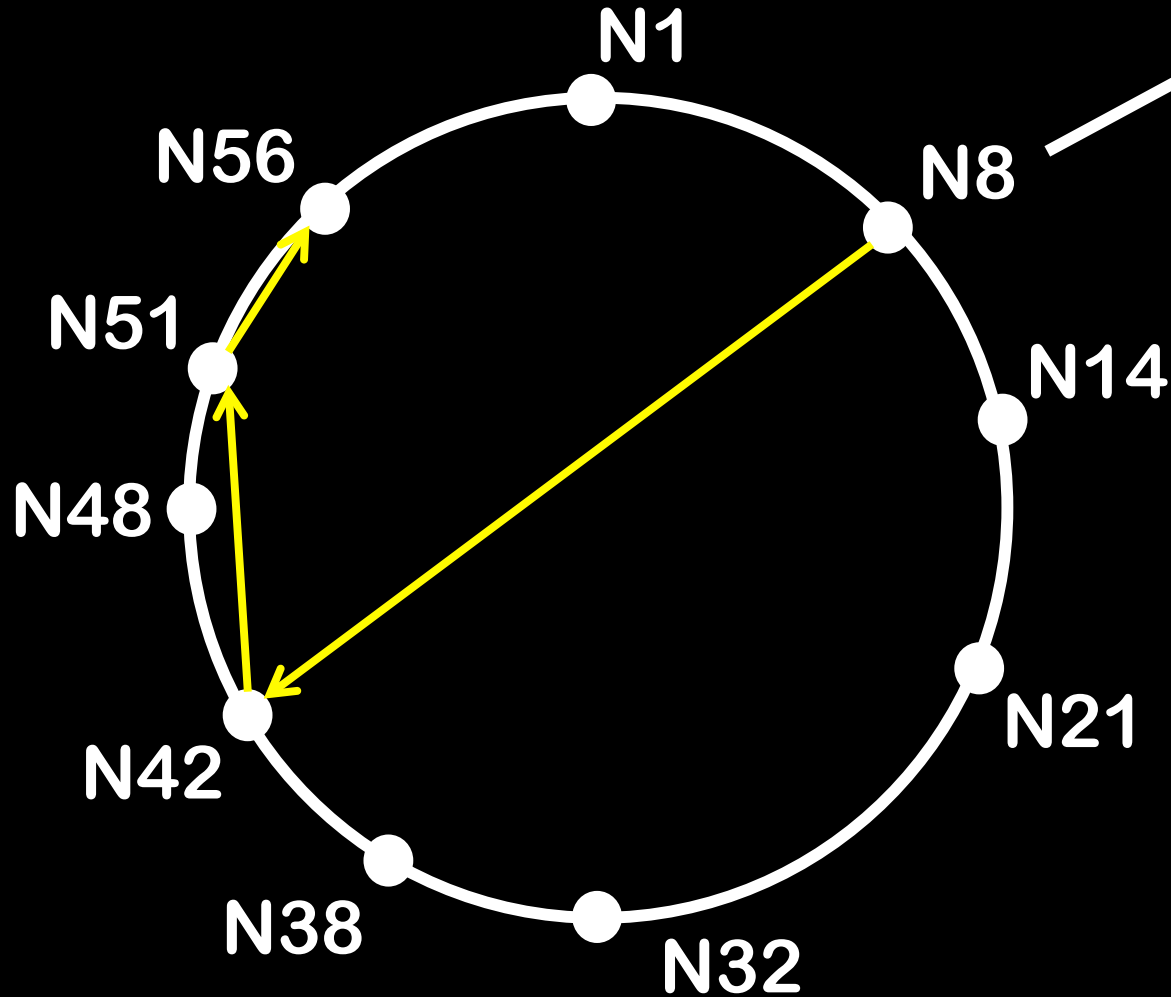
Each node will keep a “finger table” with up to m entries

The i^{th} entry of node n will contain the address of successor($n+2^i$)

The i^{th} entry of node n will contain the address of successor($n+2^i$)



Lookup(54)



$N8+1 = N14$
 $N8+2 = N14$
 $N8+4 = N14$
 $N8+8 = N21$
 $N8+16 = N32$
 $N8+32 = N42$

With high probability the number of nodes that must be contacted to find a successor in an N -node network is $O(\log N)$

Assume node n wants to resolve query for key k

Let p be the node that contains k

We will analyze the number of steps to reach p

Let i be such that p is in $[n+2^{i-1}, n+2^i]$

Node n will contact the smallest node in this interval. Call this node f .

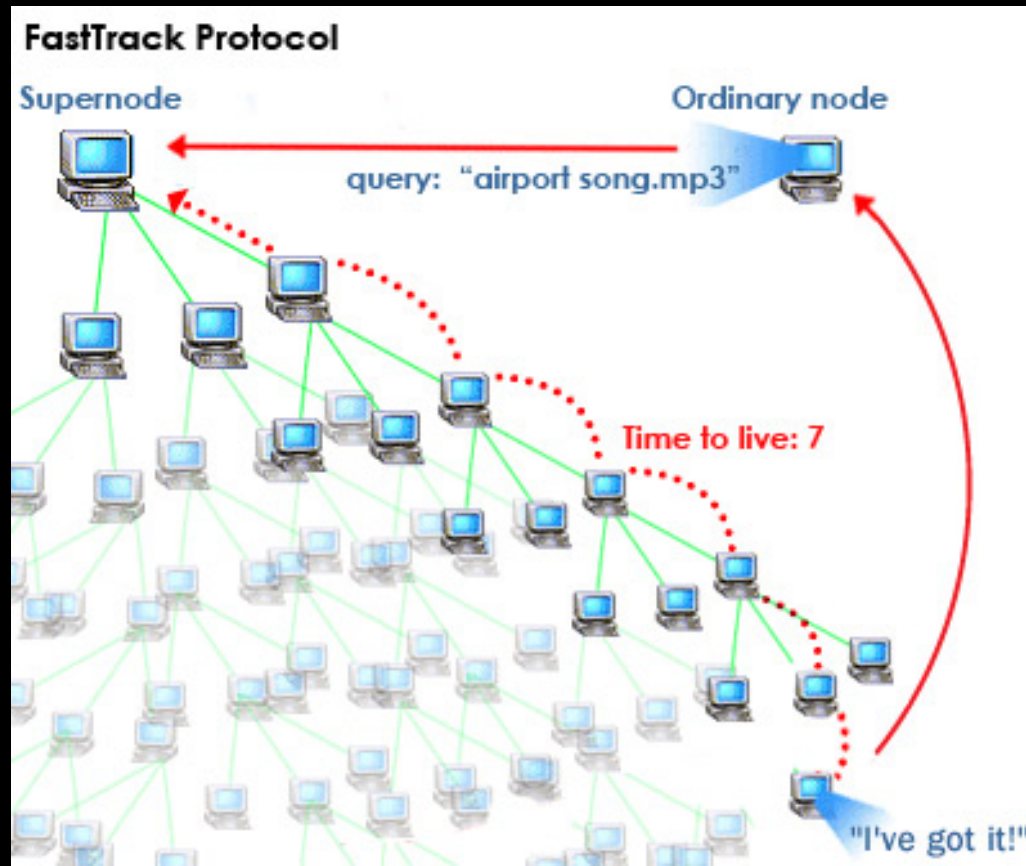
Fact: f is closer to p than to n .

Therefore, in one step, the distance to p decreases by at least half

**Other decentralized approaches
have emerged**



Kazaa: FastTrack



Two types of nodes: Regular and Supernodes

Vanuatu



**With the fully decentralized approach,
the RIAA started suing KIDS!**

<http://www.youtube.com/watch?v=3740wTqNJol>

g2g

ttyl