

# Advanced Algorithms and Programming Methods: Distributed Algorithms

---

January 7, 2019

## BROADCAST: THE FLOODING ALGORITHM

### Definition

Consider a distributed computing system where only one entity  $x$  knows some important information; this entity would like to share this knowledge with all the other entities in the system. This problem is called *broadcasting*.

Solving this problem means designing a set of rules that, when executed by the entities, will lead to a configuration in which all entities know the information; the solution must work regardless of which entity has the information at the beginning.

According to this, broadcasting requires connectivity restrictions to be solved: in particular, every entity has to be reachable from every other entity.

### Time and Message complexity

We will denote by  $M(Bcast)$  and  $T(Bcast)$  the message and time complexity.

The first message is sent by the initiator and has to reach all the entities.

Given  $d(a, b)$  the distance, i.e., the minimal number of edges between  $a$  and  $b$ , the ideal time depends on the time spent to go from the initiator to any node, thus the *eccentricity* or radius of  $G$ .

$$r(\text{initiator}) = \text{Max}\{d(x, y) : x, y \in V\}$$

Given that any node can be the initiator, we need to compute the diameter of the graph:

$$T(Bcast) = \text{diameter}(G) \leq n - 1 = \mathcal{O}(n)$$

We consider  $m$  as the number of links. We could say that there will be  $\leq 2$  messages on each link.

Let  $s$  be our initiator:

$$M(BCast) = |N(s)| + \sum_{x \neq s} (|N(x)| - 1) = \sum_x |N(x)| - \sum_{x \neq s} 1 = 2m - (n - 1)$$

The *flooding* algorithm has optimal time complexity, but the number of messages sent can be reduced! In fact,  $M(BCast) \geq m = \Omega(m)$  and  $M(Flooding) = 2m - n + 1$ .

### Assumptions

Before proceeding with the actual code, we need to make some assumptions about the solution we are providing. In particular:

- Unique initiator: the initiator is unique by definition of the problem
- Total reliability & bidirectional link: to simplify the solution
- G is connected: otherwise the problem would be unsolvable for a given initiator K

### Code

We have to consider three different states -INITIATOR, SLEEPING, DONE- and two possible events -spontaneous impulse event and receiving of a message.

```

if INITIATOR
    spontaneous impulse event
        send(I) to N(X)
        become(DONE)
    receiving(I)
        do-nothing
if SLEEPING
    receiving(I)
        send(I) to N(x) - {sender}
        become(DONE)
    spontaneous impulse event
        do-nothing
if DONE
    receiving(I)
        do-nothing
    spontaneous impulse event
        do-nothing

```

Notice that if the system is asynchronous, many different executions are possible, depending on the speed at which messages travel.

It is also worth noting that in a distributed environment the entities end their computations at different points in time, hence there is only a concept of *local termination* instead of a *global* one. No entity knows when the entire process terminates!

## PROTOCOL SHOUT

We are assuming that there is a single initiator, there are bidirectional links, there is total reliability and  $G$  is connected.

### **Preliminary: definition of Spanning Tree**

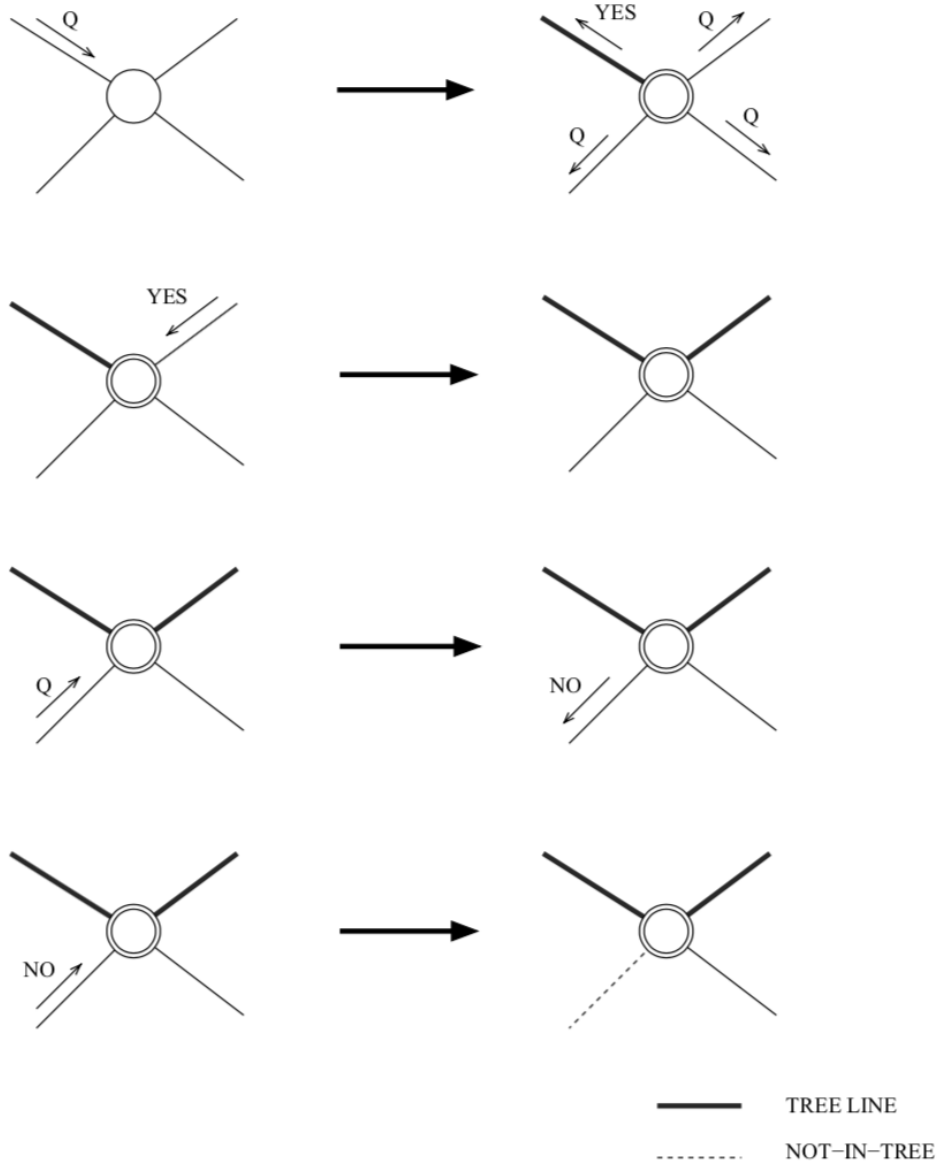
A spanning tree  $T$  of a graph  $G = (V, E)$  is an acyclic subgraph of  $G$  such that  $T' = (V, E')$  and  $E' \subseteq E$ .

### **Definition**

We want to build a spanning tree in a distributed system, but we have to consider that the entities do not know  $G$ , not even its size. The only things an entity is aware of are the labels on the ports leading to its neighbors and the fact that, if it sends a message to a neighbor, the message will eventually be received.

### **Solution** *Ask-Your-Neighbors*

- The initiator  $s$  will send a message  $Q = ('Are\ you\ my\ neighbor\ in\ the\ spanning\ tree?')$  to all its neighbors
- An entity  $x \neq s$  will reply 'Yes' only the first time it is asked and, in this occasion, it will ask all its other neighbors; otherwise it will reply 'No'.
- Each entity terminates when it has received a reply from all neighbors to which it asked the question.



Example of SHOUT

**Code:** We have four states, INITIATOR, IDLE, ACTIVE, DONE; the terminal state is DONE.

```

if INITIATOR
    spontaneous impulse event
        root = true
        tree-neighbors = {}
        send(Q) to N(x)
        counter = 0
        become ACTIVE
if IDLE
    receiving(Q)
        root = false
        parent = sender
        tree-neighbors = {sender}
        send('Yes') to sender
        counter = 1
        //managing leaves and intermediate node cases
        if counter = |N(x)| then
            become DONE
        else
            send (Q) to N(x) - {sender}
            become ACTIVE
if ACTIVE
    receiving(Q)
        send ('No') to sender
    receiving('Yes')
        tree-neighbors = tree-neighbors  $\cup$  sender
        counter += 1
        if counter = |N(x)|
            become DONE
    receiving('No')
        counter += 1
        if counter = |N(x)|
            become DONE

```

### Complexity

First of all, notice that *shout* can be assimilated to a combination of *flood* and *reply*. In other words,  $M(SHOUT) = 2 \times M(FLOOD)$ . To compute the total amount of messages sent in this protocol we have to consider the messages sent by the initiator, plus the other messages Q sent, plus the amount of 'Yes' and 'No' answers.

$$\begin{aligned}
 M(SHOUT) &= Q + NO + YES \\
 &= (n-1) + 2 \times (m - (n-1)) + 2 \times (m - (n-1)) + (n-1) \\
 &= 4m - 2n + 2 = 2 \times (2m - n + 1)
 \end{aligned}$$

## SHOUT+

We might wonder that SHOUT could be improved. For example, we might avoid sending all the 'No' messages.

This is the idea behind the SHOUT+ algorithm. In fact, the code is almost the same as SHOUT, but we do not need to send the 'No' message. Of course, this change needs to be managed.

In order to do this, we only need to change the ACTIVE state.

```
if ACTIVE
    receiving(Q)
        send ('No') to sender
    receiving('Yes')
        tree-neighbors = tree-neighbors  $\cup$  sender
        counter += 1
        if counter = |N(x)|
            become DONE
    receiving('No')
        counter += 1
        if counter = |N(x)|
            become DONE
```

is replaced by

```
if ACTIVE
    receiving(Q)
        counter += 1
        if counter = |N(x)|
            become (DONE)
    receiving('Yes')
        tree-neighbors = tree-neighbors  $\cup$  sender
        counter += 1
        if counter = |N(x)|
            become DONE
```

### Complexity

The messages are much less.  $M(\text{SHOUT+}) = 2m$  while  $M(\text{SHOUT}) = 2(2m - n + 1)$

## LEADER ELECTION

### Definition

Moving the system from an initial configuration where all entities are in the same state (called *available*), into a final configuration where all entities are in the same state (called *follower*), except one which is in a different state (called *leader*).

We assume a bidirectional graph, connected and no failures.

### SATURATION TECHNIQUE

This technique is used to solve many other problems, independently from the number and positions of the initiators (there might be multiple initiators).

The states are *available*, *awake* and *processing*. The algorithm can be divided in three parts:

- Activation phase: started by all initiators, consists in a "wake-up" - all nodes are *activated*
- Saturation phase: started by the leaves, a unique pair of neighbors is identified (*saturated nodes*), the other become *processing*.
- Resolution phase: started by the saturated nodes, it varies depending on the protocol

### Code

We have four states AVAILABLE, ACTIVE, PROCESSING, SATURATED. The initial state is AVAILABLE.

```
if AVAILABLE    I haven't been activated yet
    spontaneously
        send('Activate ') to N(x)
        neighbors = N(x)
        if |neighbors| = 1    leaf case
            parent = neighbors
            send('Saturation ') to parent
            become(PROCESSING)
        else
            become(ACTIVE)
    receiving('Activate')
        send('Activate ') to N(x) - {sender}
        neighbors = N(x)
        if |neighbors| = 1
            parent = neighbors
            send('Saturation ') to parent
            become(PROCESSING)
        else
            become(ACTIVE)
```

```

if ACTIVE
    receiving(M)
    neighbors = neighbors - {sender}
    if |neighbors| = 1
        parent = neighbors
        send(' Saturation ') to parent
        become(PROCESSING)

if PROCESSING
    receiving(M)
    become(SATURATED)

```

### Complexity

$$M(\textit{Saturation}) = 2n - 2 + n + n - 2 = 4(n - 1)$$

But how can this technique be useful for our election problem?

### SOLUTION TO THE ELECTION PROBLEM

Let us assume we assign a label  $v(x)$  to each node (notice that if nodes are indistinguishable, ranking is unsolvable). Then, it is possible to:

1. Run the saturation technique
2. The two saturated nodes exchange the value and if it is different they can elect as a leader the one with smaller value.

Observations:

- If nodes do not have distinct labels, the leader cannot be elected.
- Each saturated node knows where the other one is, since it knows from which edge it has received the last message.



## MINIMUM FINDING WITH SATURATION TECHNIQUE

Each entity  $x$  has in input its value  $value(x)$ . At the end each entity should know whether it is the minimum or not.

### Code

The states are AVAILABLE, ACTIVE, PROCESSING, MINIMUM, LARGE. The starting state is AVAILABLE.

```
if AVAILABLE
    spontaneously
        send('Activate ') to N(x)
        min = v(x)
        neighbors = N(x)
        if |neighbours| = 1 //leaf
            M = ('Saturation ', min)
            parent = neighbors
            send(M) to parent
            become(PROCESSING)
        else
            become(ACTIVE)
    receiving('Activate')
        send('Activate ') to N(x) - {sender}
        min = v(x)
        neighbors = N(x)
        if |neighbors| = 1 //leaf
            M = ('Saturation ', min)
            parent = neighbors
            send(M) to parent
            become(PROCESSING)
        else
            become(ACTIVE)
if ACTIVE
    receiving(M)
        min = min{min, M}
        neighbors = neighbors - {sender}
        if |neighbors| = 1 //leaf
            M = ('Saturation ', min)
            parent = neighbors
            send(M) to parent
            become(PROCESSING)
```

```

if PROCESSING
    receiving(M)
        min = MIN{min, M}
        notification = ( 'Resolution ', min)
        send (Notification) to N(X) – parent
        if v(x) = min
            become MINIMUM
        else
            become LARGE

```

### Complexity

$M(\text{minimum}) = 2 \times (n - 1) + n + n - 2$  where  $2 \times (n - 1)$  is the cost for the activation;  $n$  the cost for saturation and  $n - 2$  the cost for notification.  $M(\text{minimum}) = 4n - 4 = \mathcal{O}(n)$

## RANKING PROBLEM IN ARBITRARY NETWORKS

### Definition:

Given a graph  $G(V, E)$  rank it from the lowest value to the largest.

### Solution

In an arbitrary network, this problem could be solved in the following way:

- Find a spanning tree
- Use saturation and the minimum finding to find a starting node
- Do ranking (Centralized or De-centralized)

### CENTRALIZED RANKING

This problem could be solved in the following way:

- Build a spanning tree using SHOUT algorithm
- Elect a leader (minimum) using saturation technique
- The leader knows the minimum, it sends in that direction a ranking message.
- Every node knows the minimum in its subtrees, it can then forward the ranking message (ranking, minimum) in the right direction.
- When the node to be ranked receives the message it sends up a notification and update message (*new-minimum*) that will travel up to the leader.

### Complexity

$$M(\text{rank\_centr}) = 2 \times (n - 1) + 2 \times (n - 2) + \dots + 2(1) = n \times (n - 1)$$

## DE-CENTRALIZED RANKING

### Solution

The starter node sends a ranking message of the form *(first, second, rank)* in the direction of the first. The first has the smallest value, the second the second smallest known so far. If no value is indicated (or the value is  $\infty$ ) it means that the smallest in the corresponding subtree is unknown (at the moment).

The ranked node attempts to send a ranking message to the next node to be ranked. The second variable of the rank message is updated during its travel and the minimum values on the links of the tree are also updated.

### Complexity

$$M(rank\_decentr) = 2 \times (n-1) + (n-2) + (n-3) + \dots + 1 = (n-1)n/2 + (n-1) = (n-1)(\frac{n}{2} + 1)$$

### Code

- Build a spanning tree using SHOUT
- We elect a leader (e.i. minimum using saturation technique)
- Jump from the min to the second one and so on using messages of the form *(first, second, #rank)*. After having elected the minimum, we know the second smallest element and so we send a message with *(secondValue,  $\infty$ , 2nd)*: this message travels across the spanning tree to find the third element. After the second receives the message, set the value of the subtree as  $\infty$  because there is nothing to rank in that subtree - and so on.

## ELECTION PROBLEM

The election problem cannot be generally solved if the entities do not have different identities.

Consider a synchronous system where the entities are unique, have the same state and are anonymous. At each moment, they are doing the same thing, receive the same messages, move to the same states.

In such a system, it is impossible to elect a leader, because the entities are indistinguishable. That is why we need entities with different identities.

Notice that with distinct IDs, minimum finding is an election.

### Election in trees

To each node  $x$  is associated a distinct identifier  $v(x)$ .

A simple algorithm:

- Execute the saturation technique
- Choose the saturated node holding the minimum value

### Election in rings

A ring is a sparse network (few arcs) such as trees, but it's not symmetric. Every entity has two neighbors.

## ALL THE WAY

Each ID is fully circulated in the ring, hence each entity sees all identities.

Assumptions:

- Two versions: unidirectional and bidirectional
- Local orientation
- Distinct identities

All the entities can be the initiators. The bidirectional version assumes that the ring is oriented. The protocol works even without knowing the dimension of the ring and having messages that arrive in a non-FIFO order.

Ideally: each node sends its value; if it receives a message, it forwards it and it keeps track of the minimum value seen so far.

### Code

The states are ASLEEP, AWAKE, FOLLOWER, LEADER; the starting state is ASLEEP and the terminal states are FOLLOWER and LEADER.

*// notice that INITIALIZE and CHECK are "macros" / functions*

INITIALIZE

```
count = 0
size = 1
known = false
send('Election', id(x), size) to right
min = id(x)
```

CHECK

```
if count = ringsize
    if min = id(x)
        become(LEADER)
    else
        become(FOLLOWER)
```

if ASLEEP

*spontaneously*

```
INITIALIZE
become(AWAKE)
```

*receiving('Election', value, counter)*

```
INITIALIZE
send('Election', value, counter+1) to other
min = min{min, value}
count = count + 1
become(AWAKE)
```

```

if AWAKE
    receiving('Election', value, counter)
        if value  $\neq$  id(x)
            send('Election', value, counter+1) to other
            min = min{min, value}
            count+=1
            if known
                CHECK
        else
            ringsize = counter
            known = true
            CHECK

```

### Complexity

Each entity crosses each link, so the complexity is  $\mathcal{O}(n)^2$ ; the size of each message is  $\log(id)$ .

$M(Alltheway) = \mathcal{O}(n)^2$  messages

$T(Alltheway) \leq 2n - 1 = \mathcal{O}(n)$

Observations:

- The algorithm also solves the data collection problem
- It works for both unidirectional and bidirectional cases

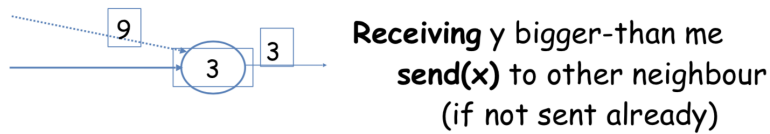
ANOTHER SOLUTION: AS FAR (AS IT CAN)

### Idea

It is not necessary to send and receive messages with larger IDs than the IDs the current node has already seen.

### Assumptions

- Unidirectional/bidirectional ring
- Different id's
- Local oriented



AsFar example

### Code

The states are ASLEEP, AWAKE, FOLLOWER, LEADER; the starting state is ASLEEP, the terminal states are FOLLOWER and LEADER.

### Unidirectional version

```

if ASLEEP
    spontaneously
        send('Election', id(x)) to right
        min = id(x)
        become(AWAKE)
    receiving('Election', value)
        send('Election', id(x)) to right
        min = id(x)
        if value < min
            send('Election', value) to right
            min = value
        become(AWAKE)
if AWAKE
    receiving('Election', value)
        if value < min
            send('Election', value) to right
            min = value
        else
            if value = min
                NOTIFY
    receiving(NOTIFY)
        send(NOTIFY) to other
        become(FOLLOWER)
NOTIFY
    send(NOTIFY) to right
    become(LEADER)
  
```

## SYNCHRONOUS SYSTEMS

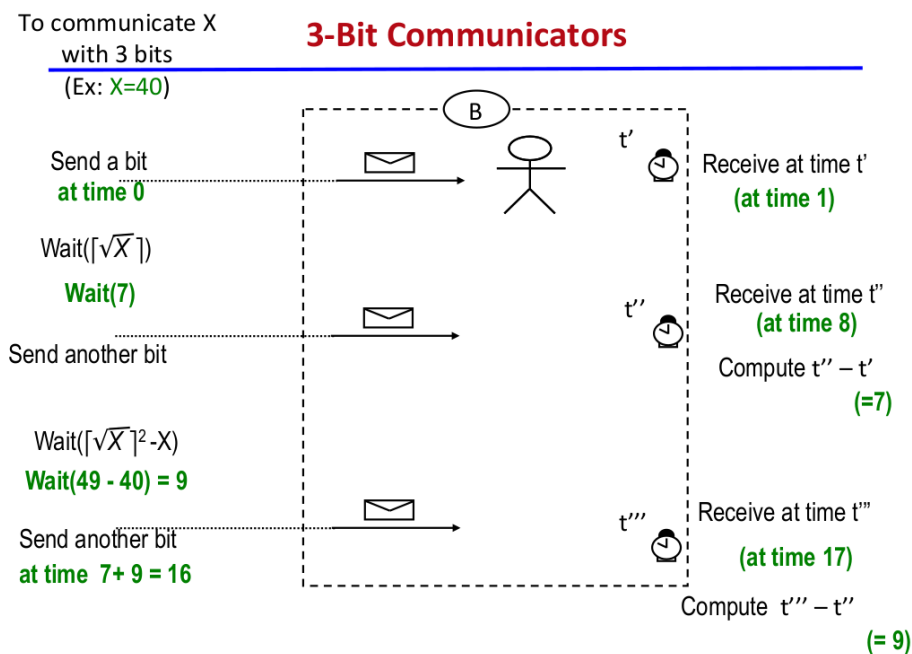
We consider as synchronous system all those systems in which all local clocks are incremented by one unit simultaneously; in other words all local clocks 'tick' in the same moment. Notice that this assumption does not mean that all the clocks have the same value, but just that their values are incremented at the same time.

The first part of this section is about calculating how much time will it take to send a message in a synchronous system.

### 2-BIT COMMUNICATORS

You want to send the number  $X$ . Send a bit, wait  $X$  seconds, send another bit. The receiver just needs to compute the difference between the times in which he receives the two messages: that is number  $X$ .

### 3-BIT COMMUNICATORS



3-bit communicators

**Time complexity**

$$\mathcal{O}(n)^2.$$



### Proof

We send a number  $K$  which is the smallest integer that satisfies the condition:

$$K \in \mathbb{N} : (K-1)^2 < X \wedge K^2 \geq X$$

This means that  $|K^2 - X| < |(K+1)^2 - K^2| = 2K - 1$ .

Therefore we have that the waiting time  $C : K^2 - C = X$  is bounded, in particular  $0 \leq C \leq 2K$ .

The time complexity is, therefore,  $\mathcal{O}(K + 2K) = \mathcal{O}(K) = \mathcal{O}(\sqrt{X})$ .

### K-BIT COMMUNICATORS

In general, for a  $k$ -bit communicator, we have that we use  $k$  bits to send the message and the time complexity is  $\mathcal{O}(k \cdot X^{\frac{1}{k}})$ .

### MIN-FINDING AND ELECTION

**General idea:** messages travel at different speed.

- Knowledge of  $n$  is not necessary
- Unidirectional version
- Synchronous

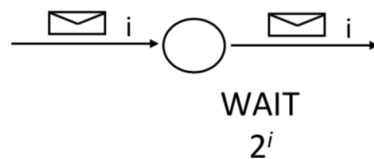
There are two ways of eliminating ids, one is like in AsFar, large ids are stopped by smaller ids; the other is that small ids travel faster so to catch up with larger ids and eliminate them.

**Important:** Each message travels at a speed which depends on the identity it contains.

Speed is assumed to be unitary, the same for every message. How can we change it? Introducing appropriate delays.

When a node receives a message containing  $i$ , it waits  $f(i)$  ticks. When a node receives its own id, it becomes the leader and sends a notification message around the ring. This message will not be delayed.

Another example could be done with  $f(i) = 2^i$ . When a node receives a message containing  $i$ , it waits  $2^i$  ticks.



Graphical example of  $2^i$  delay

When a node receives its own id, it becomes the leader and sends a notification message around the ring. This message will not be delayed.

In time  $2^i n + n$  the smallest id  $i$  traverses the ring. Let the second smallest be  $i+1$  (waiting time  $2^{i+1}$ ). How many links does it have the time to traverse while the smallest id goes around?

$$(2^i + n)/2^{i+1} \approx n/2$$

We can generalize that by thinking about the  $j^{th}$  id: in time:  $2^j n + n$ , has traversed  $(n/2^j)$  links. While:

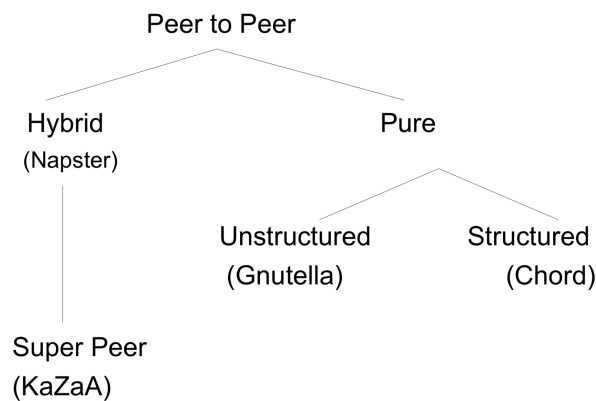
$$totalMessages = \sum_{j=1}^{n-1} = n \sum_{j=1}^{n-1} 1/2^j = O(n)$$

So we can say that the total time is:

$$totalComplexity = O(2^i n)$$

## PEER-TO-PEER

### ARCHITECTURE

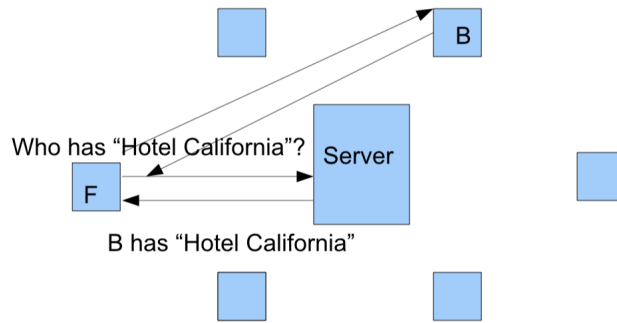


### p2p architecture

### SEARCH

### HYBRID P2P

**Napster** presents a centralized lookup. The peers send metadata to the look-up server. When a resource request arrives, the server returns the list of the peers that store the resource. Data are then exchanged among peers.



Search in hybrid p2p systems

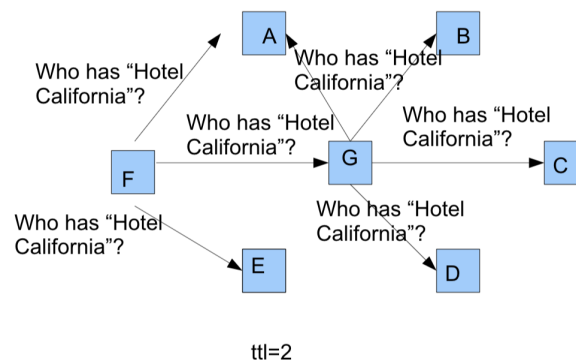
#### UNSTRUCTURED PURE P2P

**Gnutella** is a pure p2p without a central server. Peers have an initial set of addresses known for the first connection; peers are equally treated, no matter which bandwidth they have and how many files they share. Each peer provides both the files and sends/replies to routing requests. Each peer is both client and server.

Each peer knows a subset of neighbouring peers, there are some cache servers that maintain as many peer addresses as possible; when the application starts, it contacts one of these cache servers that will add the p2p network to the new peer.

When the search starts, requests from a peer R are sent from neighbour to neighbour (PING). The message stops either when the resource is found or after a limited and predefined number of steps (TTL = time to live). If the resource has been found, the address of the peer P that stores it, sends to R (PONG). R will directly contact P and will download the resource.

**Problem:** hard to control/regulate



PING with Gnutella



## SUPER PEER (ZAZAA)

It behaves like a pure and a hybrid p2p system.

- There are different servers, the super peers supervises the subnetworks (subtrees)
- The super peers execute the queries for the leaves (e.i. they work as servers)
- The super peers are also operate as normal peers
- The peers exchange information directly
- Combine the characteristics of hybrid and pure systems

**Search?** We have to do some questions:

- Good numbers of leaves for each super peer?
- How should super peers connect together?
- The system is made more reliable k redundant

## BIT TORRENT

Many users can simultaneously download the same file without delay too much each other. Files are not downloaded from the same server but they are divided into different portions. Each peer that makes a request, offers already downloaded portions to the other peers, thus contributing to the downloading of the other peers.

We concentrate on the efficient fetching and not on the search. There is a single publisher and many downloaders. The same file is distributed among many peers.

## INFORMATION STORAGE

The main questions are: where to store information in a p2p system? How to find it?

Parameters:

- System scalability: limit the communication overhead and the memory used by the nodes with respect to the number of nodes in the system.
- Robustness and adaptability: in the presence of failures and changes

There could be many approach in which we can store our information.

- Centralized approach: we send a request to the server. We use  $O(N)$  memory on the server, search  $O(1)$  step to reach the server
- Decentralized approach: we send a request to all our neighbors (different optimizations, e.i. TTL). We use  $O(1)$  memory and we use  $O(n^2)$  steps for the search.

Another approach is to have **distributed hash tables** in the system. It's a good compromise for what regard the complexity of the operations. We use  $O(\log(N))$  steps to find the informations and  $O(\log(N))$  entries in the routing table of each node.

Technique:

- Adaptability: it is simple to insert (assing thge information) and remove nodes (reassignment to the neighbouring nodes)
- Balances information on the nodes (makes routing efficient)

Peers and data are mapped in the same address space through hash functions. For nodes the hash of the IP; for data hash of the content (title, ...)

## CHORD

**Data distribution:** data are distributed among peers using a precise algorithm. Data are replicated to improve availability.

**Assignment:**

- Each peer has an ID (hash of the IP)
- Each resource has a key (hash of the title, ...)
- We use long keys to avoid collision
- The peers stores resources with keys similar to one of the peers
- Given a key of a resource the peer sends the request to the peer with the most similar key

Using Chord we have to use a consistent hashing:

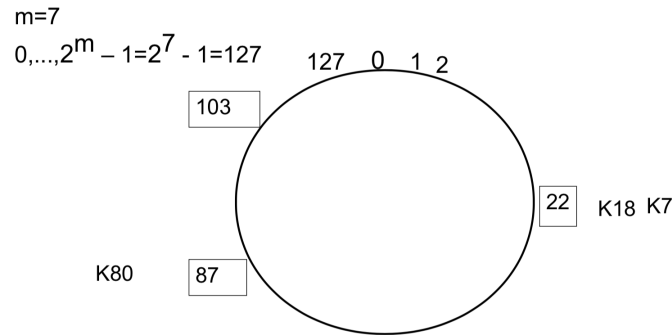
- To assign the keys to the peers
- There should be load balancing so that with very high probability each peer receives the same number of resources/keys
- We ask for key updates when a peer connects/disconnects from the network
- The insertion of the N-th key with high probability will require the movement of  $1/N$  other keys

**Advantages:**

- We maintain consistent hashing although the information is not stored in all the nodes
- The system is simple, correctness and performance are easy to prove
- It takes at most  $O(\log(N))$  to reach the destination

- A peer requires a table of  $O(\log(N))$  bits for an efficient routing, but the performance decrease when the tables are not updated
- The insertion/removal of a node generates  $O(\log^2(N))$  messages

**Idea:** The logical space is circular  $0, \dots, 2^m - 1$ , ring (mod  $2^m$ ). A resource with key  $K$  is stored in the node which is the closest successor from  $K$  (successor( $K$ )).



A key is stored at its successor: node with next higher ID

Example of how Chord assign ID

**Successor:** the successor of peer 22 is 87, of 87 is 103, of 103 is 22.

**Routing:** each node stores only its successor. If the resource is not on this node, then the query is sent to the successor.

It requires  $O(1)$  of memory,  $O(N)$  for search: in the worst case we have to visit the whole ring. The failure of a node blocks the procedure.

This algorithm in the worst case it requires too many operations (we maintain other routing informations).

Note that this extra information is not required for the correctness of the procedure but it helps speeding up the search. This protocol works given that the value of the next successor is correctly maintained/updated.

The best solution is that each node stores  $m$  values. The search of  $k$  is sent to the furthest known predecessor of  $k$ . We store more values of the close nodes and less of more distant nodes, routing is more precise close to a node. It requires  $O(\log(N))$  of memory and  $O(\log(N))$  for the search.

#### Finger table

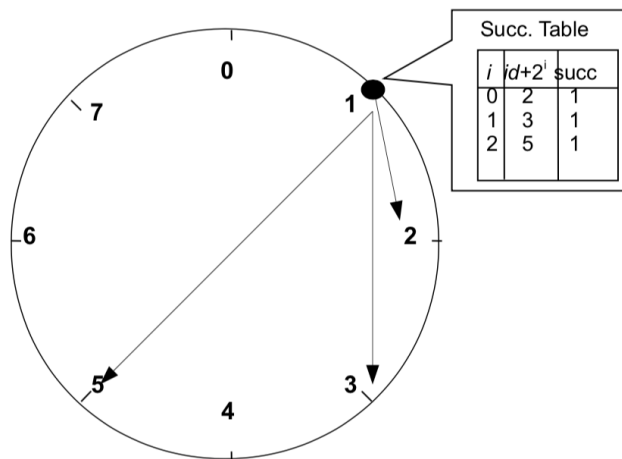
Each node maintains a finger table (the  $i$ -th entry of the finger table of  $x$  is the first node that succeeds equals  $(x + 2^i) \bmod 2^m$ ) and the predecessor node. An item is identified by id is stored on the successor node of id.

#### EXAMPLES

Here below it will be provided some examples of how this algorithm works.

Assume  $m=3$ ,  $2^m=8$

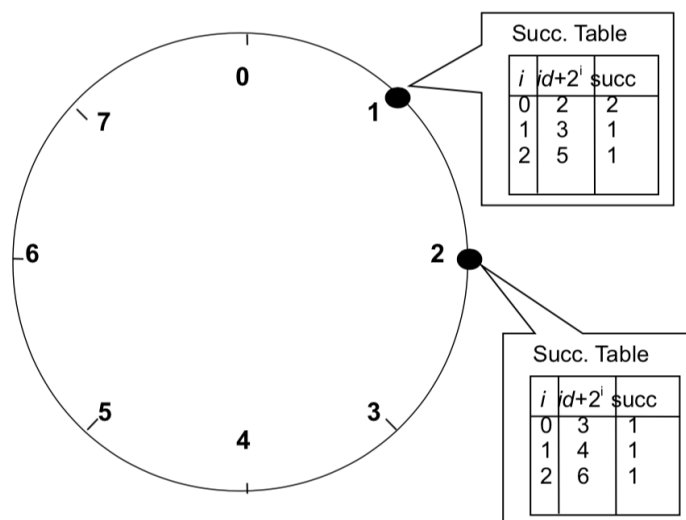
- Node  $n1:(1)$  joins  $\rightarrow$  all entries in its finger table are initialized to itself (no other peer is there)



$$1+2^0=2, 1+2^1=3, \\ 1+2^2=5$$

First assignment with Chord

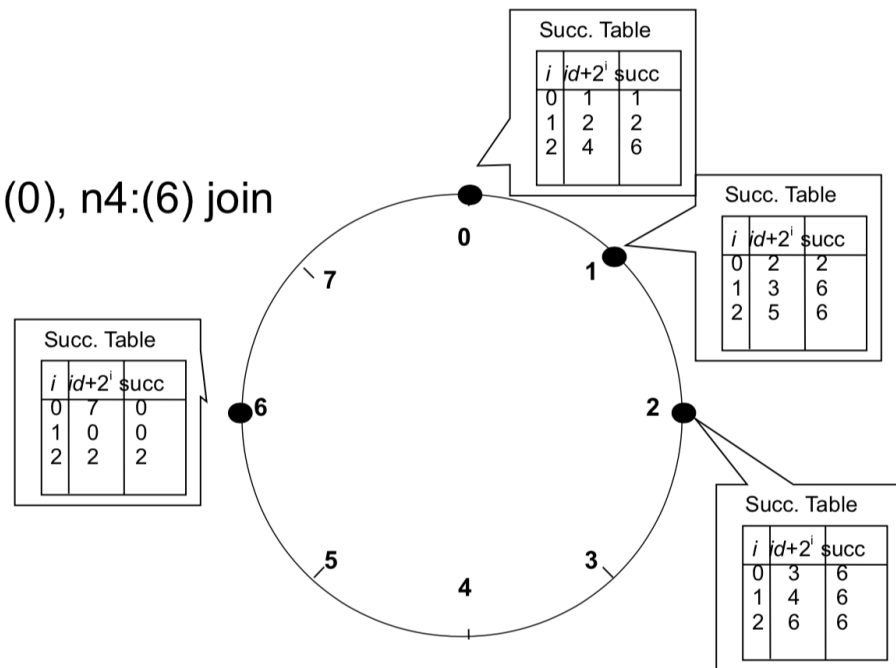
Node  $n2:(2)$  joins



Second node joins



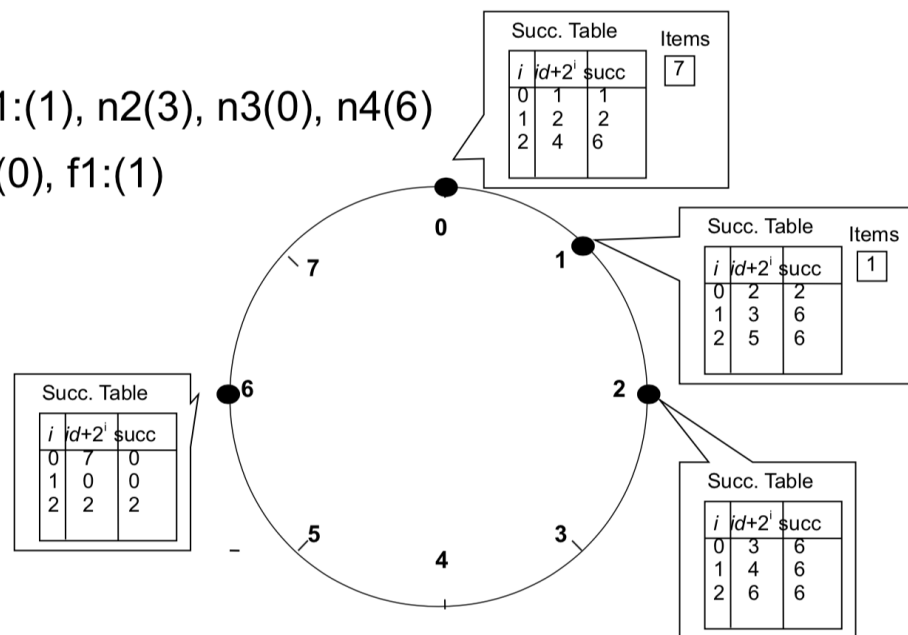
Nodes n3:(0), n4:(6) join



Third and forth joins

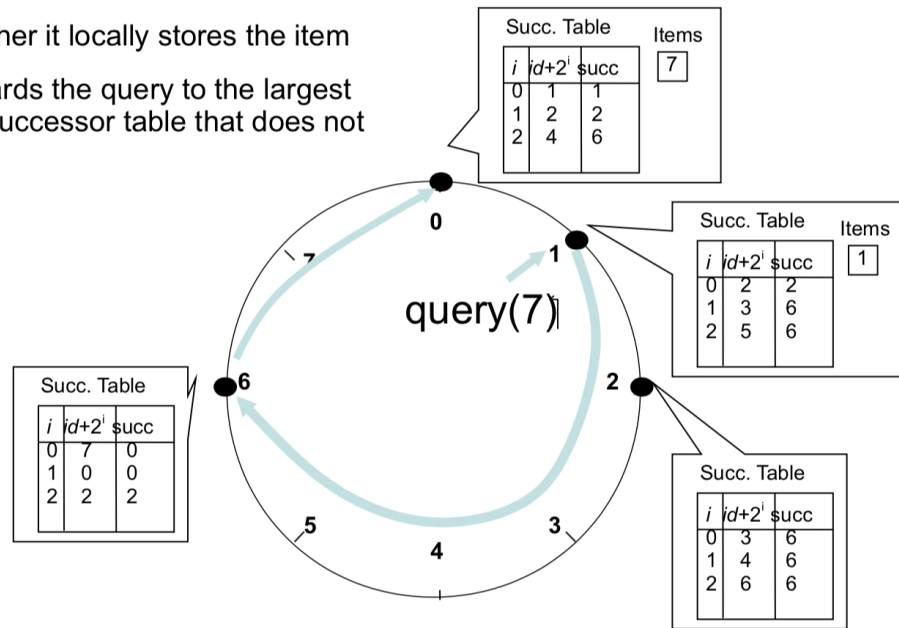
Nodes: n1:(1), n2(3), n3(0), n4(6)

Items: f7:(0), f1:(1)



Item assignment

- Upon receiving a query for item  $id$ , a node
- Checks whether it locally stores the item
- If not, it forwards the query to the largest node in its successor table that does not exceed  $id$



Routing after having received a query

## MOBILE AGENTS AND THE BLACK HOLE PROBLEM

**Definition:** Consider that we have some mobile agents that can travel across the graph and could communicate together with a whiteboard and they also have some storage, the problem is to find a malicious node (black hole) that makes disappear our mobile agents. Each agent knows the number of nodes.

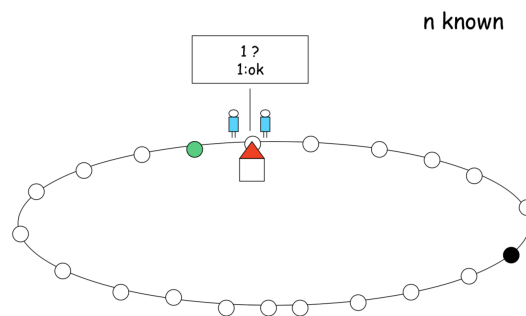
### CAUTIOUS WALK

We can consider our graph as a set of ports and in every port there is a traffic light. If the next port has not been controlled, the traffic light is red, going to the next node we activate it (yellow light), and when we come back from that node, that edge (and the next node) is safe (not the black hole).

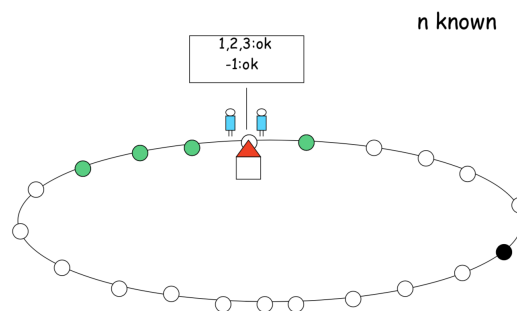
What about if the configuration of the graph is a ring?

We need two agents!

The agents, starting from the home base, must explore using cautious walk. The agents must explore disjoint areas otherwise they could both disappear. So they finally explore step by step all the ring until one of the two agents disappear of they finish the graph and the last node that has not been marked is the black hole.

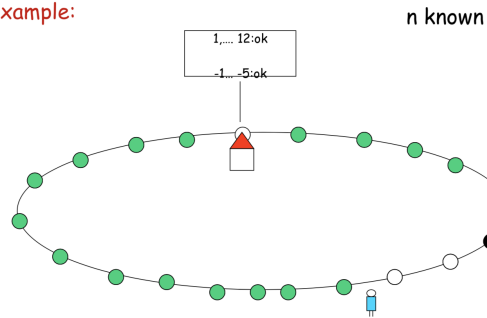


First step of the left agent



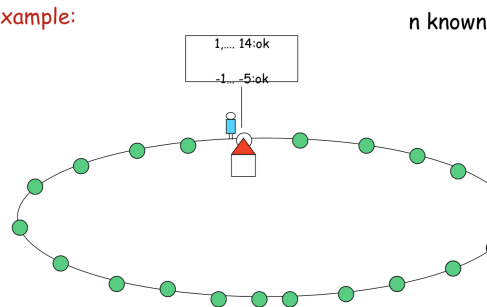
First step of the right agent

Example:



Black hole for the right agent

Example:



Left agent complete the ring and knows where is the black hole (for the whiteboard)

But what if also the intruder is mobile?

### THE INTRUDER CAPTURE PROBLEM

Let's put some rules, the intruder:

- can moves from a node to a neighbouring one
- can moves arbitrarily fast
- cannot cross a node guarded by an agent
- is invisible to the agents
- can permanently see the position of the other agents

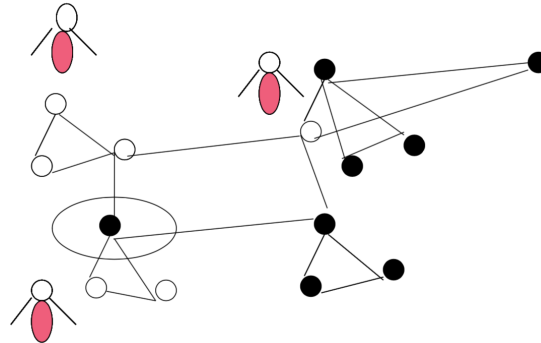
Initially the agents are located at the homebase and form a team while in the end the agents capture (surround) the intruder.

The intruder capture problem is equivalent to the decontamination problem.

As said initially the agents are located at the homebase and form a team. The whole network

is contaminated (except the homebase). An agent cleans a node when it enters in it. At the end the whole network must be clean.

**Contamination rule:** a node is contaminated if it is not protected by an agent and at least one of its neighbors is contaminated (remember that the intruder is fast).

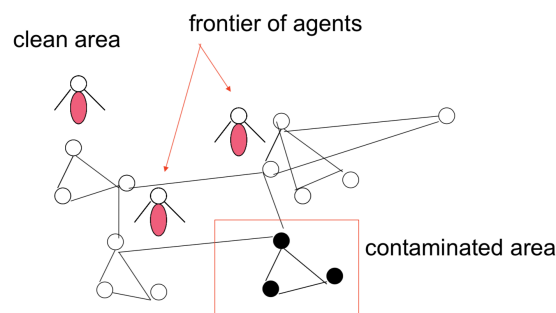


Contaminated node

**Decontamination process:** As said several times, initially the agents are located at the homebase and form a team. The whole network is contaminated (except the homebase). An agent cleans a node or an edge when it enters (or traverse the edge) in it, a node should not be re-contaminated. A node becomes contaminated if it is not protected by an agent and at least one of its neighbors is contaminated.

### CONTIGUOUS MONOTONE STRATEGIES

We consider the situation where agent move only to neighbouring nodes (contiguous) and no recontamination can occur(monotone).



Contaminated node

### Complexity issues:

- Number of agents

- Number of moves
- Time (sync or async)
- Memory (agents, nodes(whiteboard))

Strategies are not necessarily connected and monotone (i.e. agents can jump. We have to minimize the number of searchers).

There are two model of visibility, one with visibility of the other neighbors (clean, contaminated, guarded) and the other in which agents have only local knowledge.

#### DECONTAMINATING A MESH

We have to be in this situation:

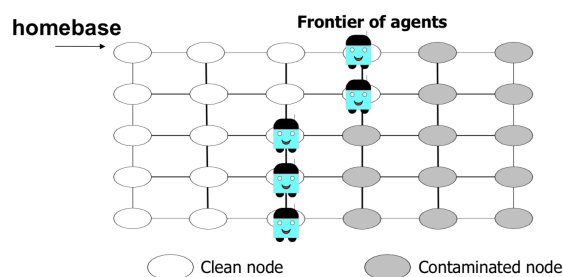
- Asynchronous system
- Node + agent storage  $O(\log(n))$  bits
- Models (visibility, local)

We consider a  $m \times n$  mesh topology. We have two possible strategies:

- With synchronizer: Searching agents do not have visibility power. The synchronizer is an agent that coordinates the moves.
- Agents with visibility: Visibility power means agent can see their neighbouring nodes, agents can move independently, the synchronizer is not required.

The main idea behind the strategy is to:

- Start from the homebase
- Contiguously clean the contaminated network by maintaining a vertical barrier of agents (to avoid recontamination, works asynchronously)
- Move one column at the time



Contaminated node

We can perform the search of a node or of an edge moving the barrier. (see slides)