Election Algorithms

• Many distributed algorithms need one process to act as a leader or coordinator
  – Doesn’t matter which process does the job, just need to pick one
  – Example: pick a master in Berkeley clock synchronization algorithm
• Election algorithms: technique to pick a unique coordinator
  – Assumption: each process has a unique ID
  – Goal: find the non-crashed process with the highest ID

Bully Algorithm

• Assumptions
  – Each process knows the ID and address of every other process
  – Communication is reliable
• A process initiates an election if it just recovered from failure or it notices that the coordinator has failed
• Three types of messages: Election, OK, Coordinator
• Several processes can initiate an election simultaneously
  – Need consistent result
Bully Algorithm Details

• Any process $P$ can initiate an election
• $P$ sends $\textit{Election}$ messages to all processes with higher IDs and awaits $\textit{OK}$ messages
  – If no $\textit{OK}$ messages, $P$ becomes coordinator and sends $\textit{Coordinator}$ messages to all processes with lower IDs
  – If it receives an $\textit{OK}$, it drops out and waits for an $\textit{Coordinator}$ message
• If a process receives an $\textit{Election}$ message
  – Immediately sends $\textit{Coordinator}$ message if it is the process with highest ID
  – Otherwise, returns an $\textit{OK}$ and starts an election
• If a process receives a $\textit{Coordinator}$ message, it treats sender as the coordinator

Bully Algorithm Example

(a)

(b)

(c)

(d)

(e)
Ring Algorithm

- Processes are arranged in a logical ring, each process knows the structure of the ring.
- A process initiates an election if it just recovered from failure or it notices that the coordinator has failed.
- Initiator sends Election message to closest downstream node that is alive.
  - Election message is forwarded around the ring.
  - Each process adds its own ID to the Election message.
- When Election message comes back, initiator picks node with highest ID and sends a Coordinator message specifying the winner of the election.
  - Coordinator message is removed when it has circulated once.
- Multiple elections can be in progress.

Ring Algorithm Example
Comparison of Bully and Ring Algorithms

• Assume n processes and one election in progress
• Bully algorithm
  – Worst case: initiator is node with lowest ID
    • Triggers n-2 elections at higher ranked nodes: O(n^2) messages
  – Best case: initiator is node with highest ID
    • Immediate election: n-1 messages
• Ring algorithm
  – 2n messages always

Election in Wireless Environments

• Goal: elect the best leader (e.g., node with longest battery lifetime)
Election in Wireless Networks

In the end, source a notes that h is the best leader and broadcasts this info to all nodes.

Superpeer Election

- How can we select superpeers such that
  - Normal nodes have low-latency access to superpeers
  - Superpeers are evenly distributed across the overlay network
  - There is a predefined fraction of superpeers
  - Each superpeer should not need to serve more than a fixed number of normal nodes
Superpeer Election in DHT-Based Systems

• Basic idea: Reserve a fraction of the ID space for superpeers
• If $S$ superpeers are needed for a system that uses $m$-bit identifiers, simply reserve the $k = \left\lfloor \log_2 S \right\rfloor$ leftmost bits for superpeers
• With $N$ nodes, we will have $N/2^{m-k}$ superpeers on average
• Routing to superpeer:
  – Let $m=8$ and $k=3$, send lookup request for key $p$ to node responsible for $p \text{ AND } 11100000$ which is treated as the superpeer