#ASLI ENGINEERING

FloodSet Algorithm for Distributed Consensus

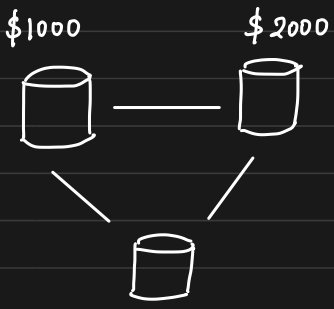
SWIPE

BY

ARPIT BHAYANI

FloodSet Algorithm : Distributed Consensus

Reaching consensus is extremely important in any distributed network.



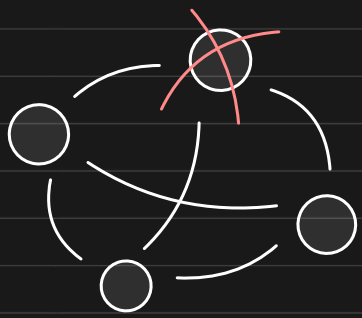
eg: we cannot have two datanodes in a cluster such that one thinks price = $1000 while the other thinks price = $2000

Depending on which node the request hits, the user would see the corresponding value, giving an inconsistent view

Somehow, the nodes need to agree on one value.

Achieving distributed consensus

- is easy when No failures
- is impossible when network unreliable
- and tricky when unreliable process



complete n-node graph

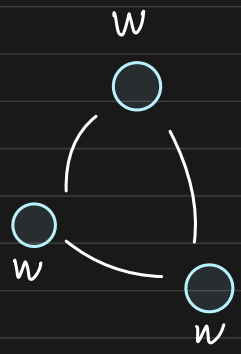
Problem Statement

All nodes / processes start with some value $v \in V$. All non-faulty nodes / processes are required to agree to one $v_x \in V$ or use default v_0

1. No two processes decide on different values
   either v_x or default v_0
2. if all processes start with $v \in V$ then they decide on v
   not converging to some other value
3. all non-faulty processes eventually decides

FloodSet Algorithm

Core Idea: Keep track of all the values seen so far and use some decision rule at the end to pick one.



Every node maintains a set W that would hold all the values seen so far. $W \in V$

if we assume at max f nodes would fail, then the FloodSet algorithm runs for $f+1$ rounds giving chances for f processes to fail.

After $f+1$ rounds, the processes that survive decides the same value.

Every node will start with $W = \{v\}$

↑ value that it holds and starts with

In each round, every node broadcasts W in the network. When a node receives W from others, it updats its own W by doing a set union.

After $f+1$ rounds, every node i will have all possible values from the other nodes

Every node then decides

if its W contains 1 element, pick that v_x

if its W contains >1 elements, pick default v_0

If all nodes start with the same value, v

only v will be sent across

and after $f+1$ rounds at all nodes $W = \{v\}$

hence, everyone decides $v_x = v$

Alternative decision strategy

Depending on the usecase, we may choose any decision strategy

1. pick the smallest one
2. pick the newest one

} so long as we have total ordering of the values

$1000, 9:00:00 am ← Total ordering on timestamp
$2000, 9:00:01 am
$1500, 9:00:02 am ← nodes deciding on the latest one

Complexity Analysis

Floodset requires $f+1$ rounds and in each round every node sends its W to every other node

Communication Complexity = $O((f+1)n^2)$