HY590.45
Modern Topics in Scalable Storage Systems

Kostas Magoutis
magoutis@csd.uoc.gr
http://www.csd.uoc.gr/~magoutis
Paxos – learning chosen value

\[
N', v \\
\text{prepare } N' \\
\text{value } v' \\
\text{prepare } N'
\]
Paxos – propagate chosen value
Paxos – everyone learns outcome
Example
Example (contd.)
Lamport: implementing a state machine

• How to run multiple instances of Paxos
  – Assume the existence of a distinguished proposer (leader)
  – A leader will run Paxos for a number of instances
  – The leader may crash, at which point there may be gaps in the chosen instances (1-134, 138, ..)
  – A new leader will try to fill in those slots or propose no-op
  – As soon as gap fills, commands can be executed

• Multi-Paxos
  – New leader: execute phase 1 for infinitely many instances
  – Acceptors can respond with reasonably short messages
  – Cost of Paxos effectively the cost of executing phase 2
Multi-Paxos

- New leader @N
- Learn accepted values for past instances

If a majority has not accepted anything for instances > I

Skip prepare phase until a propose is rejected!
Multi-Paxos

Servers play all roles

Replicas write to disk prior to sending ACK