CSCI/ECEN 5673: Distributed Systems

Spring 2017

Homework 3 Solutions

Please submit a hardcopy of your answers in class on Thursday, March 16, 2017. BBA students may submit a PDF copy of their answers via the submission link provided on Moodle. Write your answers in the space provided. Please DO NOT use any extra space. The space provided is sufficient for answering the questions.

Topics covered: Replicated state machines, Totem, Paxos, Raft.
Lecture Sets: five, six, seven and eight.

Honor Code Pledge: On my honor, as a University of Colorado at Boulder student, I have neither given nor received unauthorized assistance on this work.

Name:

1. In the Totem group communication system, explain the problem that occurs if two nodes fail in quick succession. How does Totem recover from this type of failure?

   If two processes fail during join message exchange, the construction of proc_set and fail_set result in both failed processes to be in the fail_set

   If one of the process fails during configuration phase, its failure will be detected during the circulation of the commit token. This will result in restarting the membership protocol.

Consider a distributed system with the following failure model (referred to as DS failure model):

(1) a node may suffer a crash failure;
(2) a node may send messages but not receive messages; and
(3) a node may receive messages but not send messages.

2. Explain how Raft will perform under DS failure model. What changes in Raft design will you make to tolerate failures with in the DS failure model?

Crash failure: Raft already handles this using election timeouts. A new leader election phase is started whenever a leader crashes. Follower crash doesn't impact progress in Raft as it relies on majority rule.

A node may send messages but not receive messages: If a follower fails, it will have no impact on Raft due to majority rule. If leader fails, it won't receive any new client requests, so it will send empty AppendEntry messages and the followers will receive this message and assume that the leader is working fine. So, this type of leader failure will go completely undetected and no client requests will be serviced.

Heartbeat mechanism to detect failures will not detect this type of failure. Failure detection mechanism in Raft will have to be changed to use a two-way heartbeat, wherein a follower sends a heartbeat message to the leader and the leader replies to that heartbeat message.

A node may receive messages but not send messages: If a follower fails, it will have no impact on Raft due to majority rule. If leader fails, the followers will not receive any messages from the leader and the leader election will start to elect a new leader. The leader will receive all messages during leader election, it will transition to become a follower and a new leader will be elected.

3. Can Paxos be improved under DS failure model? If it can be improved, explain how. If not, explain why not.

   Crash failure: Paxos already handles crash failures and relies on majority rule.

   A node may send messages but not receive messages: If an acceptor fails, it will have no impact due to majority rule. If a proposer fails, it won't receive any accepted/reject messages to a proposal it sends. So, its proposal will not be processed. The system will continue since any node can be a proposer.

   A node may receive messages but not send messages: If an acceptor fails, it will have no impact due to majority rule. If a proposer fails before send its proposal message, no acceptor will receive that proposal, but the system will run correctly for other proposers. If a proposer fails after sending its proposal message but before sending an accept message, the accept message will not be received by any proposers. So, this proposal will not be processed, but the system will run correctly for other proposers.