Lab2 Report

Naseem Machlovi

The cluster needs to simulate the following items.

- 1. Create a file on three nodes.
- 2. The client can submit MULTIPLE values, one at a time, to the system and the values will be stored in the log file distributedly.
 - a. For example, the client can call the RPC, SubmitValue, such as 1, on one of the nodes.
 - b. The system needs to write one line into the file, such as write 1.
 - c. When the system ensures consistency, it sends back a success message to the client.
- 3. Simulate the basic RAFT as discussed during lecture 8.

How to run:

The script is ready to run on GCP, given you have to update the ip address in both node and client code as shown below.

In order to run it locally, uncomment the local host form the both node and client.py and uncomment the following line in run_server function in node.py file

Finally, to tun the code:

python3 node.py node1 python3 node.py node2 python3 node.py node3

python3 client.py

with QuietXMLRPCServer((self.ip, self.port), allow_none=True) as server:

```
def run_server(self):
    """Run the XML-RPC server to handle incoming requests."""

# with QuietXMLRPCServer((self.ip, self.port), allow_none=True) as server:
    with QuietXMLRPCServer(("0.0.0.0", self.port), allow_none=True) as server:

server.register_instance(self)
    # server.register_function(self.delete_log_file)
    # print(f"{self.name} is listening on {self.ip}:{self.port}")
    logging.info(f"{self.name} is listening on {self.ip}:{self.port}")
    try:
        while self.running:
            server.handle_request()
        except KeyboardInterrupt:
        print(f"{self.name} server is shutting down.")
        self.running = False
    finally:
        print(f"{self.name} has shut down cleanly.")
```

In this lab, we have implemented Raft consensus for 3 node cluster mechanism based on RPC protocol. The final implementation has been done GCP cloud using XMLRPC. For the RAFT implementation, we have followed all the protocols as mentioned above and achieved the targeted results. To start the server and they can listen other nodes on designated ports, we have implemented run server function which will be run as a thread in for each node. Each node, has ip address for all the other nodes, in order to process the raft consensus and other mandatory functions such as heart beat, voting and append logs.

To start the RAFT consensus, we start the election with election function as shown below, where we have called the request vote function, which will ask each client node to vote for the given candidate and update the votes, which will in return start the leader. The leader

will start sending the heart beat to each of the follower and incase the follower doesnot receive the heart beat before its timeout, it will start its election and ignore the heartbeat from the lower term leader otherwise we will have the leader working. The implementation has shown below.

```
munec@worker1:~/Distributed-System$ python3 node.p
y node1
node1 log file not found. Starting with an empty 1
og.
node1 is requesting votes for term 1
time out set for node1 to 3.8337828209963742
time out set for node1 to 3.8337828209963742
node1 is now the leader.
election timeout 3.296514006872829

| make content of the content of t
```

Node is the leader for term1 but the timeout for node has been achieved before receiving any heartbeat from the leader term 1 and therefore a new election started, resulting in node2 as leader.

```
request_vote(self):
"""Request votes from peers to become a candidate."""
self.votes_received = 1
self.current_term += 1
self.role = "candidate"
print(f"{self.name} is requesting votes for term {self.current_term}")
last_log_index = len(self.log) - 1
last_log_term = self.log[last_log_index].term if self.log else 0
self.election_timeout = random.uniform(2.0, 20.0)
for peer, (ip, port) in self.peers.items():
         with xmlrpc.client.ServerProxy(f"http://{ip}:{port}/") as client:
             # Pass candidate's term, last log term, and last log index
             response = client.vote(self.name, self.current_term, last_log_term, last_log_index)
             if response:
                self.votes_received += 1
       print(f"Connection to {peer} failed.")
if self.votes_received > len(self.peers) // 2:
     self.start_leader()
```

```
def heartbeat(self):
    """Send periodic heartbeats to followers."""
    while self.is_leader_flag:
        # print(f"{self.name} sending heartbeat (term {self.current term} )...")
        for peer, (ip, port) in self.peers.items():
            try:
                with xmlrpc.client.ServerProxy(f"http://{ip}:{port}/") as client:
                    client.receive heartbeat(self.current_term)
           except ConnectionRefusedError:
                print(f"Connection to {peer} failed.")
        time.sleep(self.heartbeat_interval) # Sleep based on the heartbeat interval
def receive_heartbeat(self, leader_term):
    """Process a heartbeat received from the leader."""
   with self.lock:
        if leader_term >= self.current_term:
            self.current_term = leader_term
            self.last_heartbeat_time = time.time() # Reset the election timeout
            if self.role != "follower":
               print(f"{self.name} switching to follower due to received heartbeat.")
                self.role = "follower"
                self.is_leader_flag = False
                self.in_cooldown = True # Enter cooldown after receiving a heartbeat
           print(f"{self.name} ignored heartbeat with lower term {leader_term}.")
```

The above code shows the heart beat implementation for each leader and followers

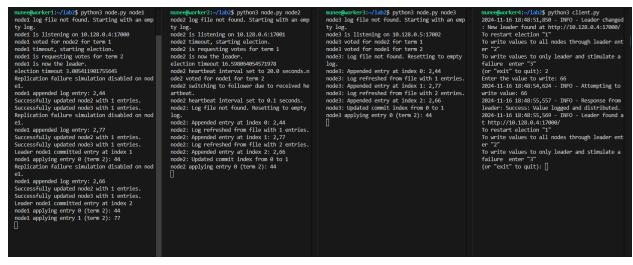
There are 3 scenarios to be simulated.

a. Leader election phase and everything goes well, NO crashes or slow followers.

```
orker1:~/lab2$ python3 node.py node1
                                                     munee@worker2:~/lab2$ python3 node.py node2
                                                                                                               e@worker3:~/lab2$ python3 node.py node3
node1 log file not found. Starting with an em
                                                    node2 log file not found. Starting with an em
                                                                                                          node3 log file not found. Starting with an em
                                                                                                          pty log.
node3 is listening on 10.128.0.5:17002
pty log.
                                                    pty log.
node1 is listening on 10.128.0.4:17000
                                                     node2 is listening on 10.128.0.6:17001
node1 voted for node2 for term 1
                                                     node2 timeout, starting election.
                                                                                                          node3 voted for node2 for term 1
                                                    node2 is requesting votes for term 1 node2 is now the leader.
                                                     election timeout 16.598064054571978
```

b. Simulate leader change in a perfect situation, which means NO left entries partially replicated (slide-8 page 54). It's basically the same as item a with another round of leader election, but picking the best candidate. This is a simplified selection due to the perfect situation. i. The leader change process can start with the client, e.g., by typing a command

We write multiple values through client on leader, which will write them on each node and will reach to raft consensus if has max write achieved, which incase here is 2 node writes.



For relection based on that we have no slow followers, as we have shown above all writes have been achieved, we passed a value 1 through client to restart the election, leading to node3 as new leader successfully.

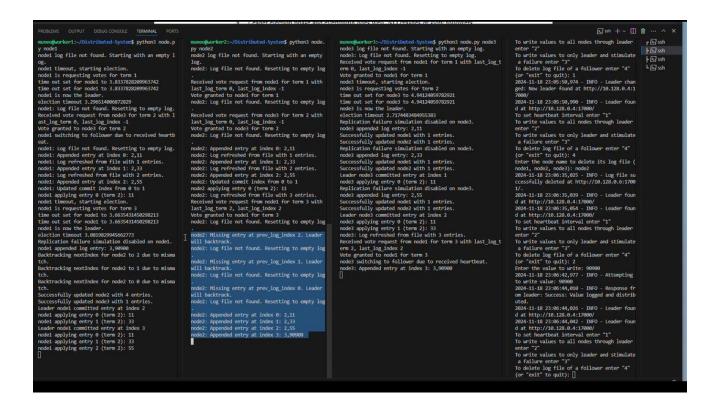
```
node3 timeout, starting election.
node3 is requesting votes for term 3
node3 is now the leader.
                                                                                                                 node2: Appended entry at index 0: 2,44
node2: Log refreshed from file with 1 entries.
el. nodel appended log entry: 2,77
Successfully updated node2 with 1 entries.
Successfully updated node3 with 1 entries.
Leader node1 committed entry at index 1 node1 applying entry 0 (term 2): 44
Replication failure simulation disabled on nod
                                                                                                                                                                                                                                                                                                                                               To restart election "1"
To write values to all nodes through leader ent
                                                                                                               node2: Appended entry at index 1: 2,77
node2: Log refreshed from file with 2 entries.
node2: Appended entry at index 2: 2,66
node2: Updated commit index from 0 to 1
                                                                                                                                                                                                                                election timeout 16.361920069177877
                                                                                                                                                                                                                                                                                                                                                To write values to only leader and stimulate a
                                                                                                                                                                                                                                                                                                                                               failure enter "3"
(or "exit" to quit): 1
2024-11-16 18:50:39,290 - INFO - Heartbeat inte
                                                                                                               node2 applying entry 0 (term 2): 44
node2 voted for node3 for term 3
                                                                                                                                                                                                                                                                                                                                               rval temporarily set to 20.0 seconds on http://
10.128.0.4:17000/. Response: None
2024-11-16 18:50:59,322 - INFO - Heartbeat inte
node1 appended log entry: 2,66
Successfully updated node2 with 1 entries.
Successfully updated node2 with 1 entries.
Leader node1 committed entry at index 2
node1 applying entry 0 (term 2): 44
node1 applying entry 1 (term 2): 77
node1 heartbeat interval set to 20.0 seconds.
node1 voted for node3 for term 3
                                                                                                                                                                                                                                                                                                                                               rval reset to original value of 0.1 seconds on http://10.128.0.4:17000/.
                                                                                                                                                                                                                                                                                                                                               2024-11-16 18:50:59,449 - INFO - Leader changed
: New leader found at http://10.128.0.5:17002/
To restart election "1"
 node1 switching to follower due to received he
                                                                                                                                                                                                                                                                                                                                               To write values to all nodes through leader ent
                                                                                                                                                                                                                                                                                                                                               To write values to only leader and stimulate a failure enter "3" (or "exit" to quit): []
 node1 heartbeat interval set to 0.1 seconds.
```

The append entries code implementation is below, which will write the inputs from the leader to each follower logs file and keep trach of indexes.

```
def append_entries(self, term, entries):
    ""Leader appends entries and attempts replication to followers."""
   if not self.is_leader_flag:
       return False
   entries = [LogEntry(term, command) if not isinstance(command, LogEntry) else command for command in entries]
   # Append new entries to leader's log and save to file
   for entry in entries:
       self.log.append(entry)
       with open(self.LOG_FILE, "a") as f:
           f.write(entry.to_string() + "\n")
       print(f"{self.name} appended log entry: {entry.to_string()}")
   for peer, (ip, port) in self.peers.items():
       if self.simulate_replication_failure and peer != "leader":
           print(f"Replication to {peer} skipped due to simulation.")
       while True:
               prev_log_index = self.next_index[peer] - 1
               prev_log_term = self.log[prev_log_index].term if prev_log_index >= 0 else 0
               entries_to_send = self.log[self.next_index[peer]:]
               if not entries_to_send:
                   break # No new entries to send
```

c. Simulate a scenario that the leader crashed (e.g., log file deleted) with an inconsistency log (e.g., missing only ONE log entry) on the other two followers. Check out slide-8 page 62. i. Again, it can be started by the client's command.

To implement this part, we deleted the logs files from of the followers file (node2) and showed the back tracking based on raft so it can updated the follower logs files. In our case, we did it for node2 and leader perform the back track to update the values as shown below.

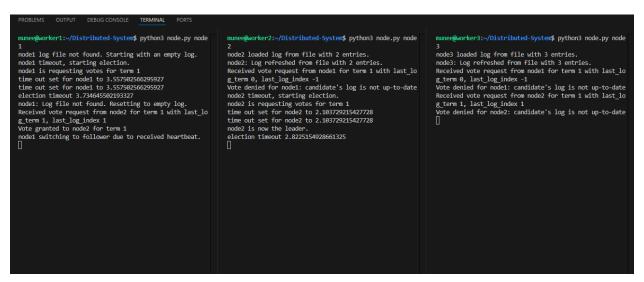


The code snippet for the above method is shown below, which update the logs based on current log in the node based on leader logs

```
def receive_append_entries(self, term, prev_log_index, prev_log_term, entries, leader_commit):
    """Follower receives and appends multiple log entries from the leader, ensuring consistency."""
    # Refresh in-memory log from file before processing entries
   self.refresh_log_from_file()
   with self.lock:
       if term < self.current_term:</pre>
           return False # Reject entries from an outdated leader
       # Update term and reset role if in a new term
       if term > self.current_term:
           self.current_term = term
           self.is_leader_flag = False
       # Reset election timer on heartbeat
       self.last_heartbeat_time = time.time()
       if prev_log_index >= len(self.log):
           print(f"{self.name}: Missing entry at prev_log_index {prev_log_index}. Leader will backtrack.")
        if prev_log_index >= 0 and (len(self.log) <= prev_log_index or self.log[prev_log_index].term != prev_log_term):
           print(f"{self.name}: Log mismatch at index {prev_log_index}. Truncating to resolve conflict.")
           self.log = self.log[:prev_log_index] # Truncate to remove conflicting entries
with open(self.LOG_FILE, "w") as f: # Rewrite log file
               for entry in self.log:
                    f.write(entry.to_string() + "\n")
       new_index = prev_log_index + 1
        for entry_str in entries:
           entry = LogEntry.from_string(entry_str)
            # Truncate if there's a conflicting entry
           if new_index < len(self.log) and self.log[new_index].term != entry.term:</pre>
               self.log = self.log[:new_index]
                print(f"{self.name}: Truncated conflicting entries from index {new_index}.")
                with open(self.LOG_FILE, "w") as f: # Rewrite log file
                   for log_entry in self.log:
                        f.write(log_entry.to_string() + "\n")
           # Append new entries if beyond current log length
            if new_index >= len(self.log):
               self.log.append(entry)
                with open(self.LOG_FILE, "a") as f:
                    f.write(entry.to_string() + "\n")
                print(f"{self.name}: Appended entry at index {new_index}: {entry.to_string()}")
            new index += 1
```

Optional BONUS: Simulate the crashed node rejoins the cluster. In this case, the leader needs to synchronize with the latest log files. The situation is similar to item c (slide-8 page 62), but with multiple rounds.

To achieved this, we modified the vote function, which reject the vote request from any candidate whose log is out of date based on it current log. We modified the logs of node1 by deleting it and when the new election phase start, it vote request was denied by both the node2 and node3 as shown below thus, choosing node2 as leader as it has the next request inline.



```
def vote(self, candidate, term, last_log_term, last_log_index):
    """Vote for a candidate if the candidate's term is greater than the current term
    and the candidate's log is at least as up-to-date as this node's log.""
    with self-lock:
    self.refresh_log_from_file()
    print(f"Received vote request from (candidate) for term (term) with last_log_term {last_log_index {last_log_index}")

    if term < self.current_term:
        print(f"Vote denied for (candidate): candidate's term (term) is less than current term (self.current_term)")
        return False

my_last_log_term = self.log[my_last_log_index].term if self.log else 0

if (last_log_term < my_last_log_term) or (last_log_term == my_last_log_term and last_log_index < my_last_log_index):
        print(f"Vote denied for (candidate): candidate's log is not up-to-date")
        return False

if (term > self.current_term) or \
        (term == self.current_term and not self.voted_for) or \
        (self.voted_for == candidate):
        self.current_term = time.time() # Reset the election timeout
        print(f"Vote denied for (candidate): already voted for {self.voted_for} in term {self.current_term)")
        return True

print(f"Vote denied for (candidate): already voted for {self.voted_for} in term {self.current_term)")
        return False
```