Phase 3: Design and Pseudo-code for Twins

Team Name: RVP

Raj Patel, 114363611 Prince Kumar Maurya, 114354075 Vishal Singh, 114708875

Design and pseudo-code for scenario generator

Features

Operation mode

Selection of leaders

Enumeration Limits

Enumeration Order

Partition change triggers

Message drops

Liveness

Design and Pseudocode

Network partition generator design

Network partition generator pseudo code

Scenario generator

<u>Design choices for generating scenarios</u>

Design

Scenario generator pseudo code

Design and pseudo-code for Scenario Executor

Design

Network Playground pseudo code

Scenario executor pseudo code

Additional pseudo code

Changes in Diem

TestRunner

Contributions

Note for TA: For pseudo code, you may refer to the psuedo_code folder. Additionally, you may use pseudo code extension in vscode to get syntax highlighting.

Design and pseudo-code for scenario generator

Features

Operation mode

The design supports writing test scenarios onto a file in json format as well as supports directly passing the test scenarios to the test executor.

Selection of leaders

The leader selection can be done deterministically by providing leader assignment for each round or the algorithm does it on a random basis.

Enumeration Limits

The algorithm limits the enumeration by generating configurations on the fly during scenario generation. For each round assignment, a set of network partitions are chosen and nodes are placed based on the logic discussed in the pseudo-code below.

Enumeration Order

Enumeration based on partition assignments (deterministic) and randomized is supported.

Partition change triggers

Partition change triggers are performed by the networking playground based on triggers on the round numbers and the count of messages from a validator.

Message drops

All forms of message drops are supported by the form of partitions and the partition triggers.

Liveness

Liveness is ensured by the scenarios generated by the scenario generator. Additionally similar to the paper we introduce three additional rounds in the end with a majority quorum by ensuring supermajority nodes to be part of one of the partitions of the Network. This ensures that the algorithm necessarily performs a block commit.

Design and Pseudocode

Network partition generator design

The network partition generator generates all unique ways in which the network can be partitioned across all nodes (including twins).

The algorithm relies on backtracking to generate the partitions.

For example: For a set of 5 nodes the following partitions are generated.

```
[('_', '_', '_', '_', '_', '_']]
[('_', '_', '_', '_', '_', '_']]
[('_', '_', '_', '_', '_'], ['_']]
[('_', '_', '_', '_'], ['_'], ['_']]
[('_', '_'], ['_', '_', '_'], ['_']]
[('_', '_'], ['_'], ['_'], ['_']]
[('_', '_'], ['_'], ['_'], ['_']]
[('_', '_'], ['_'], ['_'], ['_']]
[('_'], ['_', '_', '_', '_'], ['_']]
[('_'], ['_', '_', '_'], ['_'], ['_']]
[('_'], ['_'], ['_'], ['_'], ['_']]
[('_'], ['_'], ['_'], ['_'], ['_']]
[('_'], ['_'], ['_'], ['_'], ['_']]
[('_'], ['_'], ['_'], ['_'], ['_']]
```

Note: "_" are just placeholders where a node is placed by the scenario generator when a partition is chosen.

This contains duplicate partitions such as [['_', '_', '_'], ['_', '_']] and [['_', '_'], ['_', '_']]. This is eliminated in the next step which filters the unique partitions.

After performing pruning the following is the set obtained.

[['_', '_'], ['_', '_'], ['_']] [['_', '_'], ['_'], ['_'], ['_']] [['_'], ['_'], ['_'], ['_'], ['_']]

This set contains all unique scenarios in which the 5 nodes can be divided.

Along with this, the network partition generator also returns a set containing partition with atleast the majority (2 * F + 1) to make the work simpler for the scenario generator. This majority will in turn be used to ensure liveness

These are the majority partition set returned by the network partition.

Network partition generator pseudo code

This can be found under psuedo_code/network_partition_generator

```
Module NetworkPartitionGenerator{
  Procedure generatePartitionRec(
     partitions <- partitions U current partition
     current partition[lastElement] <- current partition[lastElement] U node list[index]</pre>
     generatePartitionRec(index+1, node_list, current_partition, partitions)
     current partition[lastelement].pop()
     new_partition <- new_partition U node_list[index]</pre>
     generatePartitionRec(index+1, node list, current partition, partitions)
     current partition.pop()
    p.sort()
   unique partitions <- {}
   for p in partitions:
       key <- key U |np|
```

```
partitions <- unique partitions.values()</pre>
  return partitions
  current partition <- ∅
  partitions <- ∅
  generatePartitionRec(1, node_list, current_partition, partitions)
  partitions <- prune duplicate partition(partitions)</pre>
  majority partitions <- get majority partitions(partitions, F)</pre>
  return partitions, majority partitions
Function get_majority_partitions(
    majority_partitions = \emptyset
      if len(p[0]) >= (2 * F + 1):
        majority_partitions <- majority_partitions U p</pre>
```

```
return majority_partitions
}
```

Scenario generator

Design choices for generating scenarios

There are several ways in which the scenario generator can be designed to cater to the requirements stipulated to ensure liveness.

Choice 1: Scenario generation based on gauging past progress

In this form of scenario generation, the generator lays out scenarios in which it gauges the next network partition based on the progress of the past network partition and the state of Diem processes inferred from messages. We call this type of choice to be **memory full**. In which some reasoning is required based on the previous state to determine the next state.

Consider the following choice of network partition for round 1 (R1) with the leader as node 1.

R1:

```
<u>1</u>, 2 | 4, <u>1</u>' | 3
```

To maintain progress some consensus has to be there such that nodes can transition to the next round. Let us choose one such network partition that provides a consensus state. This network partition will be chosen by some trigger.

```
R1:
```

```
1, 2 | 4, 1' | 3
1, 2, 3 | 4 | 1' -> consensus 1
```

Now nodes 1, 2, 3 will progress to the next round. For 1' and 3 to progress as well, we introduce one more network scenario such that progress gets maintained.

```
R1:
```

```
1, 2 | 4, 1' | 3
1, 2, 3 | 4 | 1'
1', 4, 2 | 1, 3 -> progress for remaining nodes consensus 2
```

R2:

```
1', 4, 2 | 1, 3 -> node 2 should be consistent
```

In cases where commits have to be performed, leader assignments have to be tracked as well.

Advantages:

- 1. This type of scenario generation keeps the scenario executor's implementation lean. The scenario executor just has to choose a network partition based on some state trigger and the round number of the current validators.
- 2. With past progress in track, specific scenarios can be generated to ensure block commits are deterministic.

Disadvantages:

1. Scenario generation becomes complicated. Gauging the next state based on the previous state is difficult. (A **DFA** (**Deterministic finite automata**) can be produced to keep track of state transition which could lower down the complexity).

Choice 2: Memoryless scenario generation

In this form of scenario generation network partitions can be laid out randomly while ensuring progress. In Diem's context a quorum is required to maintain progress hence in each round there has to be a consensus state which could drive a QC or a TC. Consider the example in choice 1. For ensuring progress for some nodes only the first consensus state is required. The scenario executor can then define partition change in such a way that progress is maintained for remaining nodes that were not part of consensus (4, 1'). This also allows for random leader assignments as even if the leader is in a minority partition progress is ensured by forming a TC.

Advantages:

1. Implementation of scenario generation is simple.

Disadvantages:

- 1. Block commits are difficult to be determined based on scenarios.
- 2. The scenario executor has to cater to ensuring the progress of non-minority nodes in some scenarios.

Design

Our design extends choice 2 discussed above. For block commits to happen, in the end, we provide 3 rounds having a scenario with the supermajority that would ensure a block commit. This would ensure the liveness property holds of the algorithm. For any safety violation to happen it would have happened prior to this point. Likewise, if the algorithm misses committing after this generated scenario it would flag a liveness violation.

The ScenarioGenerator, generates scenario in a recursive pattern. In a given round the first M - 1 network partitions are meant for finding safety / liveness violations while the last round is meant to perform some progress.

Example:

Round 1:

1.... M -1 random partitions meant for finding safety / liveness violationsMth partiton with majority consensus to maintain progress by forming QC or TC

...

Round N:

1.... M -1 random partitions meant for finding safety / liveness violations

Mth partition with majority consensus to maintain progress by forming QC or TC

Round N + 1:

Majority consensus to ensure liveness

Round N + 2:

Majority consensus to ensure liveness

Round N + 3:

Majority consensus to ensure liveness

The last three additional rounds give chance for any BFT algorithm (Diem in our case) to commit a block.

Assign leaders

Assign leaders provide leader assignments for each round except for the last three rounds where we provide scenarios for block commit to happen. Assignments can be provided for leaders to be chosen deterministically. It also supports additional types such as sequential where leader assignments happen in a round-robin fashion similar to the way in PBFT.

Round assignment

The round assignment first chooses the number of partitions to be present at each round. It chooses a random number with the help of M (max partition per round). Hence, the partition per round is between [1, M] (inclusive).

Next, it uses NetworkPartitionGenerator to generate all unique ways for partitioning N nodes. Now, for each round and each partition in that round until the penultimate partition (for that specific round), it chooses a network partition and uses a populate partition to make placement for the leader as well as other nodes in that partition.

For the final assignment, it chooses the one which forms a majority from the majority partition set and performs node placement using a populate consensus partition. This ensures that the last partition will be used for subsequent messages such that it forms a QC or TC to advance to the next round and maintain the liveness of the system.

In the end, it generates three-round scenarios which can form a quorum and can perform a block commit.

Scenario generator pseudo code

This can be found under psuedo_code/scenario_generator

```
Module ScenarioGenerator{
probability of overlap // probability of overlapping partition
probability partition has overlap // probability that a network partition contains a overlap
  parition_assignments
  partition choices <- [i for i in 1...M + 1]
  partition per round <- ∅
  for i in 1...(N + 1)
    partition per round[i] <- random.choice(partition choices)</pre>
  final assignments <- Ø
  for i in 1...(N + 1):
      final assignments[i] <- parition assignments[i]</pre>
    for j in 0..(partition per round[i] - 1)
      random.choice(partition scenarios),
```

```
final assignments[i] <- final assignments[i] U (populate conensus partition(
 majority = random.choice(major partitions)
   majority,
    random.choice(non faulty),
Function populate partition(network partition) {
 node shuffled <- random.sample(nodes, total nodes + F)</pre>
 no of partitions <- |network partition|</pre>
```

```
for i in 0...|network partition|:
    if random.uniform(0, 1) < probability_partition_has_overlap:</pre>
nodes shuffled <- random.sample(nodes, total nodes + F)</pre>
majority_partition <- network_partition[0]</pre>
```

```
twin nodes = Ø
  for j in 0...|network_partition[i]|
    network_partition[i][j] <- nodes_shuffled.pop</pre>
  return ⊥
pending assignments = \emptyset
```

```
pending assignments <- pending assignments \mathsf{U} i
if type == "sequential":
 for i in 1....(N + 1):
else if type == "random":
  for i in 1....(N + 1)
   if i ∉ assignments:
      final assignments[i] <- assignments[i]</pre>
```

Design and pseudo-code for Scenario Executor

Design

The scenario executor based on the mode of operation of the scenario generator either reads the json file of the configuration or else directly receives the configuration. The scenario executor performs the task of spawning the validators, initialization of the network playground, running the scenarios, and performing verification and validation for safety and liveness.

Spawning the validators

Spawning of validators is done similar to the way it is done in phase 2 of the implementation. Once, the spawning of 3 * F + 1 validator is done. It further spawns additional F nodes with configuration as i % (3 * F + 1) validator is done.

1). Configuration means the cryptographic key which is used uniquely identifies the validator. While spawning it also provides the leader assignments to each node so that the `getLeader` method overridden in the validator can make use of the leader assignments provided per round.

Network playground

The network playground is the glue that ties in the entire execution logic and ensures progress for nodes that belong to a minority in a particular round. In order to do this, the selection of network partitions is based on the highest seen round by the validators. Meaning if a message sent by a higher round node is received at a node in the lower round then in-order for progress the network partition for the lowered ordered node is chosen based on the round number of the higher round node. This is illustrated in the example below:

R1: <u>1</u>, 2, 3 | 4, <u>1'</u> R2: 2, <u>1'</u>, 4 | 3, <u>1</u>

In round 2 when 2 sends a message to 1' and 4. Although, 1' and 4 belong to a prior round (Round 1 in our case). The network playground will choose round 2 as the network partition as the proposal message is sent for R2 and not R1. Similar logic would tie in whenever other types of message exchanges are being done. Hence, whenever differing round scenarios are encountered the network playground would ensure that progress happens.

Another setting that network playground has to handle is partition change in a given round itself. We perform this by the count of messages for a particular round. Formally, we maintain a state that keeps track of messages passed by each validator at each round. If the message passed exceeds the available partitions at a round we pick the last partition which will be the majority partition that will ensure progress from that round. For example:

R1: 1, 2 | 3, 4 | 1' 1, 2, 1' | 3, 4 1, 2, 3 | 4, 5

In the example above when the first message is passed for that round view (the round view is the highest seen round for that message) then it will choose the first partition, next for the second message it will choose the next partition and for subsequent messages, the last partition is chosen.

This will make sure that there are some sets of nodes that will make progress for sure. And the subsequent nodes can eventually make progress.

Running the scenarios

For running the scenarios the network playground will drive the communication and transition. Thus while running, the round number passed by the messages can be used to interpret the termination state. More formally, when the round number of any message exceeds R + 3 (total rounds defined + three extra rounds) then we trigger termination of nodes.

Verification

Checking for safety violations

For checking safety violations one can check the ledger state to see for any differing commits. Differing commits are indicative of a bug in the algorithm which allowed such commits to happen.

Other safety violations can lead to no blocks to commit which would be indicative of a liveness violation.

Checking for liveness violations

For checking liveness violations one can check the ledger state to see if consistent block commits are present at atleast 2 * f + 1 nodes (excluding the other twin node when checking for consistency). Then the ledger state is consistent. If in case the scenarios are laid out such that a block commit should happen, then the same has to be reflected in the ledger. No block commit in such a scenario is indicative of liveness violation.

Liveness also implies progress hence the logs must contain incremental changes in round numbers along with the ledger states.

There are two approaches for verification and validation.

- 1. Online
- 2. Offline

Online verification: In online verification, the scenario executor performs verification and validation on the fly. In this, it checks for any safety and liveness violation based on the ledger and logs generated.

Advantages:

- 1. Any safety or liveness violation can be detected immediately.
- 2. It usually reduces the execution time by detecting bugs early.

Disadvantages:

1. Could be computationally expensive as it makes periodical checks.

Offline verification: In offline verification, the scenario executor performs liveness and safety checks at the last when the execution is either done or the execution exceeds a certain threshold. The threshold is a guesstimate of how long it takes for a set number of rounds to execute. (As there are no network delays this can be estimated quite accurately by taking the upper bound for each timeout to occur multiplied by the total number of rounds).

Advantages:

- 1. Would be relatively less computationally expensive.
- 2. Simple to implement.

Disadvantages:

1. Execution time is relatively higher as it waits for the upper bound time to reach in case of a liveness bug.

In our design, we employ offline verification as it is much more simple to implement.

Network Playground pseudo code

This can be found under pseudo code/network_playground

```
Module NetworkPlaygroud{
total nodes, // count of nodes excluding twin (4 * F + 1)
 Func Main() : EventLoop{
     loop : Wait for next event M; Main.start event processing(M)
     Procedure start event processing(M) {
         onReceive(M)
Procedure onReceive (M)
  senderID <- M.senderID</pre>
  receiverID <- M.receiverID</pre>
  ScenarioExecutor.OnReceive(M)
         send(M, M.payload, nodeID)
       if same partition(senderID, twinNodeID)
```

```
send(M, M.payload, twinNodeID)
if same partition(senderID, receiverID)
  send(M, M.payload, receiverID)
if same partition(senderID, twinReceiverID)
  send(M, M.payload, twinReceiverID)
for partition in current_partitions[partition_id]:
    if (sender ∈ partition) && (receiver ∈ partition):
```

Scenario executor pseudo code

This can be found under psuedo_code/scenario_executor

```
Module ScenarioExecutor{
num nodes, //total number of nodes excluding twin
twin nodes, //nodes for which to create twins
round partitions, //Configuration of Partition for each round
round leaders //Leader for each round
id // file id of Configuration
reached timeout <- False
round partitions, round leaders = load from json(id)
Function start()
     round partitions, round leaders = load from json(id)
   NetworkPlaygroud <- NetworkPlayground(round partitions, number of nodes)
   F \leftarrow |twin nodes| / 2
   Keys <- generateKeys(num nodes)</pre>
  np <- NetworkPlayground()</pre>
     Nodes <- spawn(i, Keys[i % num nodes], round leaders, np)</pre>
```

```
terminate(i)
OnReceive (M) {
  current round=M.round;
  reachedTimeout <- True</pre>
Function verify(){
  SafetyCheck() &&
  LivenessCheck()
Function SafetyCheck() {
      ledger = twin_ledger
    ledger = getLedger(node) or ledger
        other ledger = getLedger(node)
```

```
return False
    return False
Function LivenessCheck() {
```

Additional pseudo code

Changes in Diem

Can be found under psuedo_code/diem

```
Module Diem {
id
keys
leader_assignments
NetworkPlaygroudNode // reference of network playground for relaying messages
broadcast <- -1
Module LeaderElection{
Message {
Procedure sendMessage(payload, round, receiver) {
 M <- Message(payload, max(round, PaceMaker.current round), id, receiver)
 send(NetworkPlaygroudNode, M)
Module Mempool{
```

```
Procedure get_transactions() {
    // payload contains the message payload
    // round contains the max of round of original message or current round
    // receiver is the receiver (could be a node or broadcast type)
    current_transaction <- current_transaction + 1
    return str(id) + "-" + str(current_transaction)
    }
}</pre>
```

TestRunner

Can be found under psuedo_code/test_runner

```
Module TestRunner {
Function RunTests() {
  for config in configurations:
      config.N, // total number of rounds
      config.probability_of_overlap // probability of overlapping partition
      config.probability_partition_has_overlap // probability that a network partition contains a
    leaders = generator.assign_leaders(config.type, config.assignments)
      config.parition_assignments
```

```
// emphasis being a file is saved by save_to_json with name as id
// this file can be picked up at anytime by scenario executor
executor = ScenarioExecutor(
    config.total_nodes,
    config.twin_nodesleaders,
    round_assignments,
    leaders,
    config.N,
    id,
    mode
)
executor.start()
executor.verify()
}
```

Contributions

Pseudo codes:

Raj Patel, 114363611 Scenario generator Prince Kumar Maurya, 114354075 - Network Playground Vishal Singh, 114708875 - Scenario Executor

The design, design decisions and documentation were done mutually by all team members.