SDN Controller

Team : **Radiance**

Team members **:**

1. 140110047 Shachi Deshpande
2. 140050002 Deep Modh
3. 140050007 Neeladrishekhar Kanjilal
4. 140050087 Ashna Gaur

Index

2. Abstraction

Input :

1. SDN controller takes traffic request inputs from the application interface.
   This is in the form of N*N*k*m matrix, where
   N is number of nodes controlled by our controller,
   k is number of service levels,
   m is number of instances of a particular service for given pair of source and destination.
2. The traffic request matrix with specification of protocols and priorities comes from the application interface via the northbound interface.
3. The N*N connectivity matrix input will be taken as vectors from the network interface at the interval of every 20 clock cycles.
4. The values taken for various variables are:
   m=10
   k=1

The controller is scalable to larger values but in test benches it takes value as 7
N is represented by 3 bit binary and “000” is reserved for ‘-1’. Hence 7 nodes can be represented

A network graph (adjacency matrix) means the following in our context-

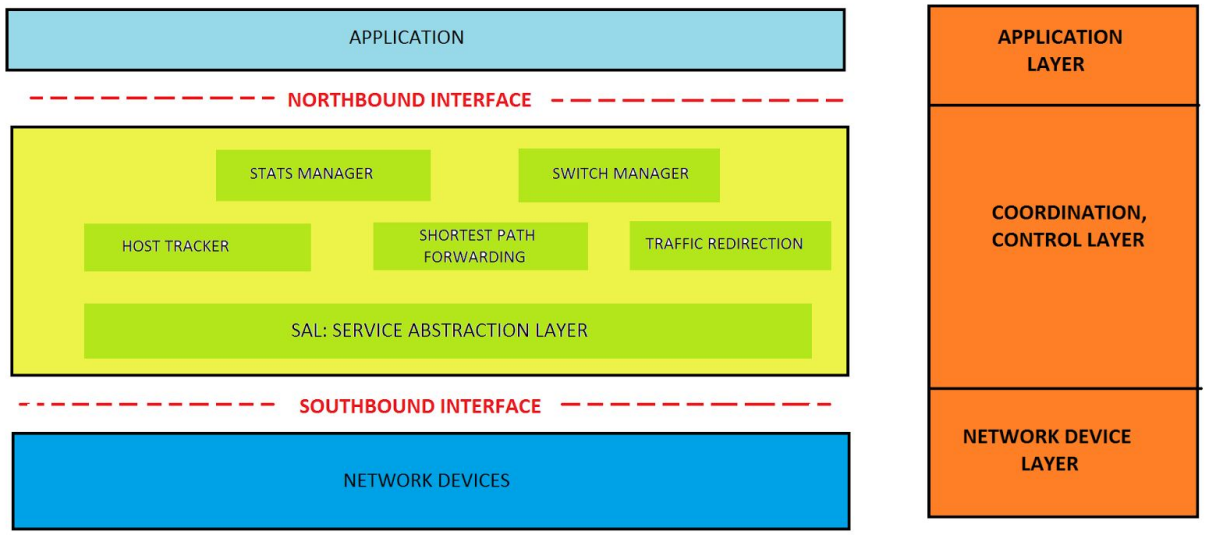
1. The nodes in network are vertices of graph.
2. If direct connectivity exists between 2 nodes in the network, then there is an edge between corresponding edges in the graph.
3. So, when path (i.e. continuous sequence of edges) exists between 2 nodes of a network, we can say that these nodes are connected in the graph.

Functionality :

Send maximum number of messages successfully.

3. High Level Architecture

3.a) High Level Block Diagram :



Functionalities:

Stats Manager: Maintaining statistics
Switch Manager: Manages switches connections with input and output ports
Host Tracker: Information related to messages of that host
Shortest Path Forwarding: The main algorithm
Traffic Redirection: Implements work conservation & priority management of requested services
SAL: A crossbar connecting the protocol plugin to the network function module

3.b) High-Level Idea of Working

1. **Computing shortest path** :
   We get the N*N connectivity matrix, and we divide the network into some predetermined number and sizes of sub-graphs.
   After every 20 clock-cycles, we will get the input of this connectivity matrix.
   Every time few new connections (or edges) are made and/or few edges are lost.
   Accordingly our graph changes.
   So, we compute the shortest path in the sub-graphs at the end of receipt of connectivity matrix every time. We keep the shortest path between every pair of nodes updated in this way.
   We plan to implement the algorithm in **higher level languages like python** and integrate it with our project. Otherwise we will implement it in VHDL itself.

2. After 200 clock-cycles we get the traffic matrix. For every pair of nodes, we **create a routing string** using the pre-computed shortest paths, and associate this string along with every message when we decide the next-hop router for a message. This ensures that we are routing the message along shortest possible path. This acts as **forwarding table**(We will can it **NH Matrix**(Next Hop)) entry for the particular packet, for the concerned router in the routing string.

3. In the initial stage of our project we will keep k=1, i.e. we have only one service, and thus no different relative priorities. In later stage of project k is made greater than 1, we basically have for each router a priority sensitive forwarding.
   Using the routing string we have already determined the sequence of nodes to be followed. Now, we need to determine which information packet is forwarded initially for given node pair. This is done by accepting the incoming packets into different section of router-related memory. So, during forwarding, we give higher priority of forwarding to the corresponding memory section storing those packets.
   This functioning would be similar to **priority arbiter**, who gives more time to process of higher priority.

4. **Protection Path**- It might happen that during forwarding particular edge of graph

vanishes (i.e. direct connection between a pair of points is lost). In such case, we wait for the acknowledgement packet from receiver. If there is such a problem, the message packet fails to reach the receiver, and hence after a predetermined waiting time, we compute the shortest path between the corresponding source destination pair once again. Since it would automatically not contain the vanished edges, we can be sure that this new path, called protection path, is completely edge disjoint from the failed link, and can successfully transfer the packet.

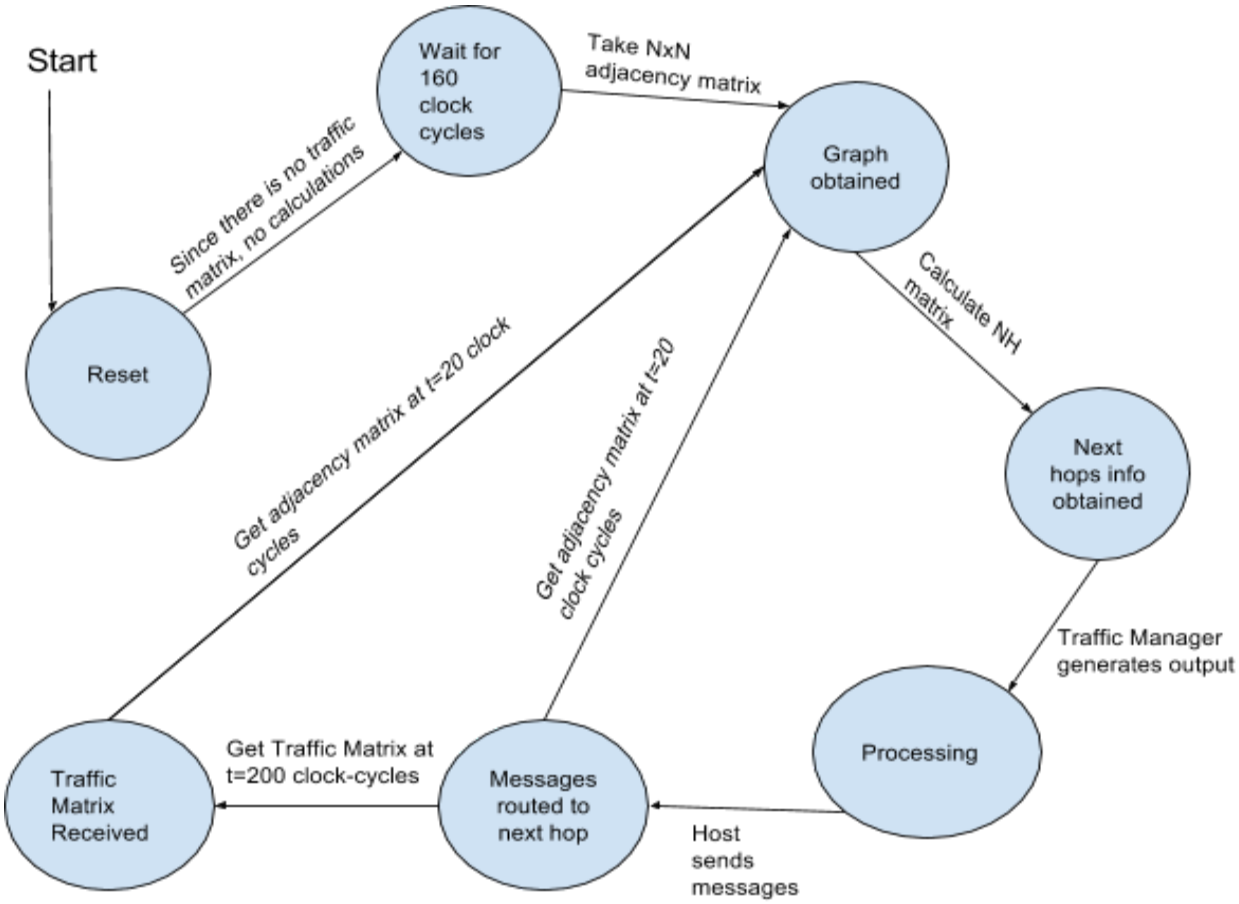
5. **Statistics Collection-** The application interface has an output corresponding to network statistics obtained (jitter, average latency for each path, and packet drop rate). For this we implement a statistics module in our code.

    a) We collect the difference between received time and sent time for each packet delivered for respective source destination pair and maintain this record to get **average latency**. This difference can be achieved by means of a stopwatch module inside statistics module.

    b) Also, the same information can be useful to calculate network **jitter** as variance of latency of packet delivery between each source destination pair.

    c) **Packet drops** for each source destination pair are maintained in 2 ways- we increment total packet drops when every time a router is full and message packet forwarded to that particular router gets dropped. We also increment packet drop number when suddenly a connection link fails (as described in point 4) and the packet gets lost. The answers to average latency, jitter and and packet drops appears as output of this statistics module.

3.c) State Machine Diagram



This presents a high level state machine diagram, which defines the functioning of the blocks with respect to current state.

Explanation :

1. In the 1st 200 cycles, since we don't have a traffic matrix, we won't do any forwarding or act upon the adjacency matrices.

2. At t=160 clocks, we begin processing adjacency matrices for the 1st time.

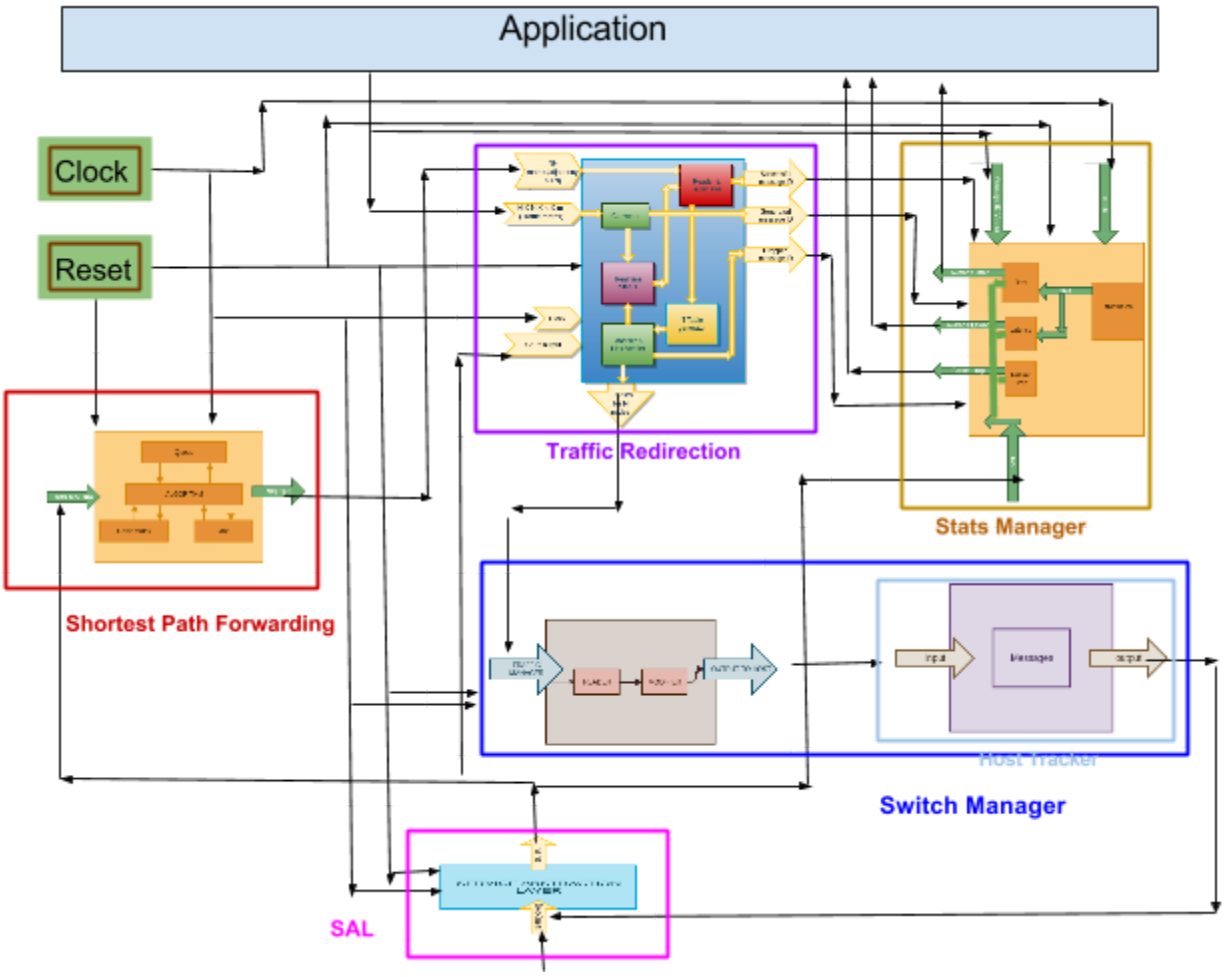
3. At t=200 clocks, we obtain the Traffic Matrix as well, and thus we can calculate the next-hop router by our algorithm block.

4. After that the processing is triggered, which basically informs each router which message to forward at each port, using output of traffic manager.

5. This results in the routers actually forwarding the messages. After that, at 20X clock cycles, we go to the previous state of obtaining graphs. At t=200X, we get both a new traffic matrix, and a new adjacency matrix, and go to the previous state of obtaining graphs. Now this cycle of states repeats over and over every 200 clock-cycles.

4. Description of functionalities

4.a) Flow of functionalities Diagram



Each of the block is explained on the following pages.

4.b) Algorithm Block :

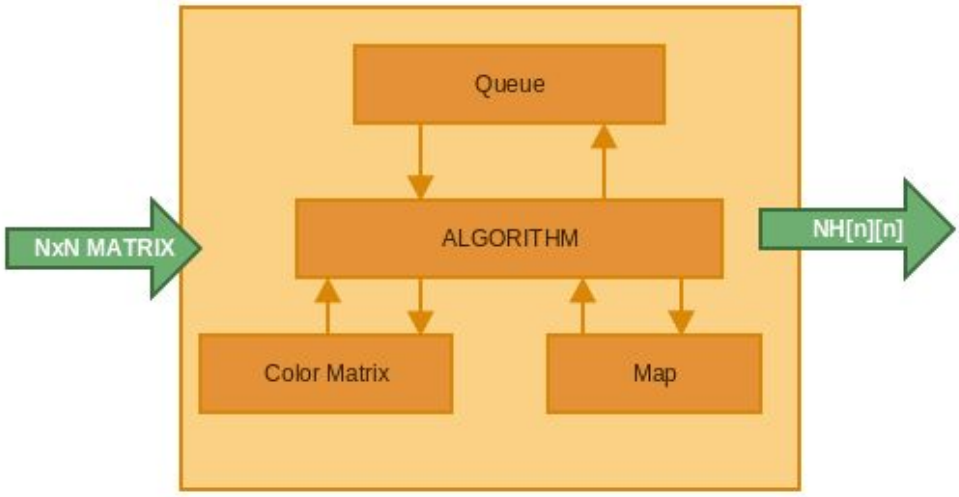
Store : NxN adjacency matrix here

SHORTEST PATH BLOCK:
Implements the path finder algorithm. The functionality is clearly described in the algorithm section.
Assumptions: No special assumptions. If the graph is connected, we get proper shortest paths via BFS and get the appropriate next-hop router.:

Input : adjacency matrix A[n] [n]

Output : NH[n] [n]
where

1. n is number of nodes.
2. NH [i] [j] is node number of next-hop of shortest path from node i to node j ;
   -1 if no such path.



Note : Algorithm computes and stores NH matrix within 20 clock cycles.

Algorithm :

1. Let i be any node.
2. Apply BFS on graph given by matrix A with starting node i.
3. Every node j in BFS tree is reachable by node i.
4. Shortest path from i to j in graph is same as path the path from i to j in BFS tree.
5. NH[i] [j] is the child of i that is ancestor of j.

Pseudo Code :

```
For each i;
        queue q // DFS queue, initially empty
        initialize NH[ i ] [ j ] to -1 for j = 1 to j = n

        color [ 1,2,...,n ] // color[ i ] is color of i-th vertex, initially -1
        M is map from color to node.

        integer c = 0; // color : unique to each child of i
        add node i in q and mark node i visited.

        for every child w of i
                color[ w ] = c;
                increment c by 1
                add key c with value w in M
                add node w in q and mark node w visited
        pop first element of q // remove node i from queue

        while q is not empty
                p is the first element of q
                for every child w of p
                        if w is not visited
                                color[ w ] = color[ p ] ;
                                NH[ i ] [ w ] = M[ color[ w ] ] ;
                                add node w in q, mark node w visited
                pop p from q
```

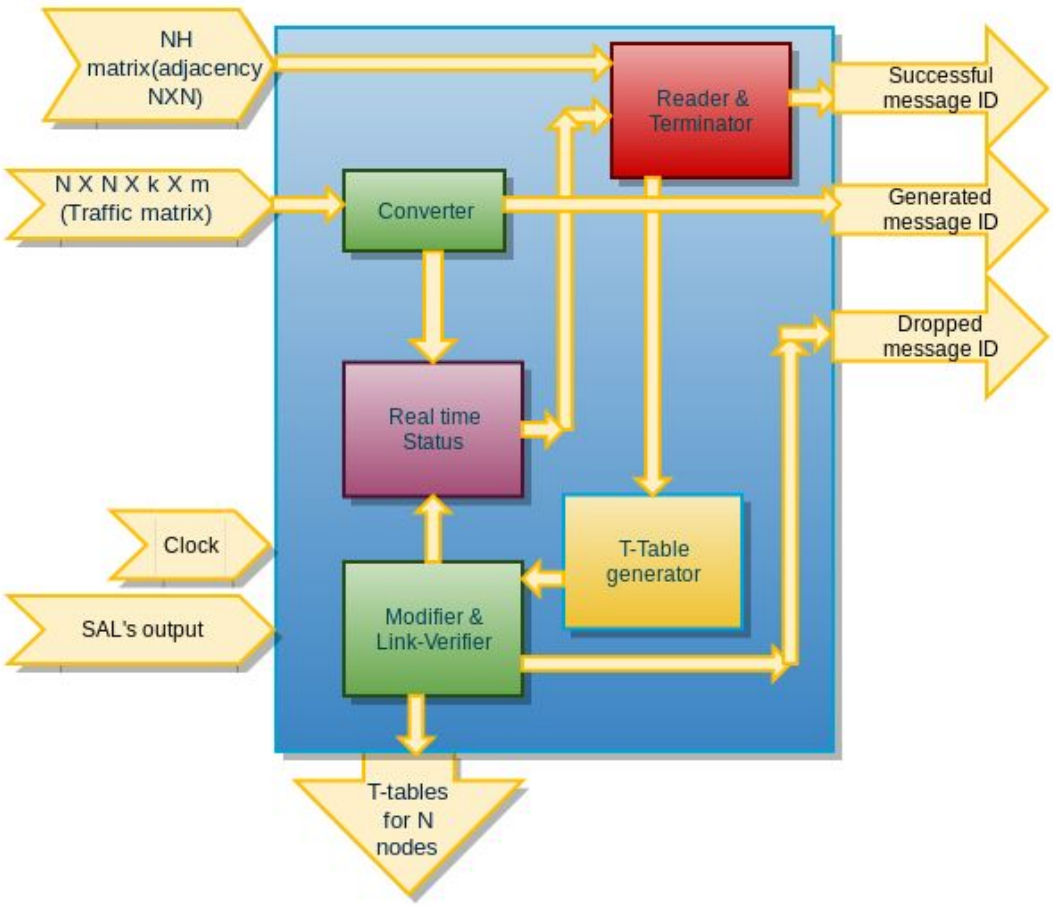
Analysis : O(n^2 * log n)

4.c) Traffic Controlling Module :

Store : NH matrix here

Input :

1. All packets available in every Node.
2. Every packet contains the Source-Node , Destination-Node.
3. Every Node has its own forwarding table corresponding to every Destination-Node which is shortest path.(NH matrix)



Assumption:

Every packet takes exactly '**1**' clock cycle to reach to its corresponding next hop node.

Expectation:

All the packets that can be “Sent” in the corresponding clock cycle from every node such that there is never a case when there is a packet available at the node and the corresponding next hop node wire is empty.

Working-Out:

1) The key Idea is to maintain a temporary table(**T-table**) corresponding to every Node having a single column and ‘p’ rows(p = number of immediate neighbours). Where every row is dedicated to every next hop node(Need to verify for existence of link from SAL(***‘Link-Verifier’***)).

2) ***Real Time Status***:
    a) This has N vectors(integer) where each vector corresponding to every node and the integer it contains represents the message IDs of the messages contained in it.
    b) It also assigns IDs to every message uniquely and gives the ID

3) ***Converter***: Its job is to take in the Traffic Matrix from standard input, Assign distinct **ID**s to every packet (and give it at output) and append it to the ‘Real Time Status’ data for continuing the packet sending method.

4) Now we do the following steps :
    a) Go through all the packets in the Node and see their Destination Node.(***‘Reader’***) If they are already at their destination remove them from the Real Time Status and output the IDs(***‘Terminator’***).

    b) We get their corresponding next hop node from the NH matrix taken in input.

    c) If the row corresponding to the next hop is empty(or if the packet it contains has priority less than the priority of the packet in view) we put the packet in view in that row. (Thus we achieve **applications/services prioritization**)

    d) Once we have traversed through all the packets in a Node what we get is the list of packets to be sent in the corresponding clock cycle. (Thus we achieve **Work conserving memory controller**)(***‘T-table generator’***)

e) This we do for every node. The corresponding listed packets are removed from the memory of their contained Nodes and appended to the memory of their next hop nodes(***'Modifier'***)

   i) Here we implement our logic for packet dropping. In case the number of packets in the Node is already 10 then we decide to **Drop** that packet(output the IDs corresponding to them) (This can be improvised by making it wait for a constant number of clock cycles instead of dropping it. Note that this does not affect the work conserving scenario as that wire cannot be used by any other node.
   ii) If the Link to a next hop is destroyed for a particular Node(Info obtained from SAL) then the packet is supposed to buffer till a new graph is obtained at the 20th clock cycle(***'Link-Verifier'***)

Example:

1. Consider the following representation of a packet(M). M = (message id , next hop node number , application priority , etc.)

2. Consider a node(node number 3) who has 3 possible next hops(immediate neighbours). Suppose it initially has the following packets and its T-Table is empty.

3. (1,4,1),(2,4,1),(3,5,1),(4,9,1),(5,4,2),(6,9,1),(7,9,1),(8,9,1)

4. After processing we will have the following T-Table for this node:

(5,4,2)
(4,9,1)
(3,5,1)

Which means that from this node we must send these three packets to their respective next hops and retain the other packets in the memory(**buffering**).

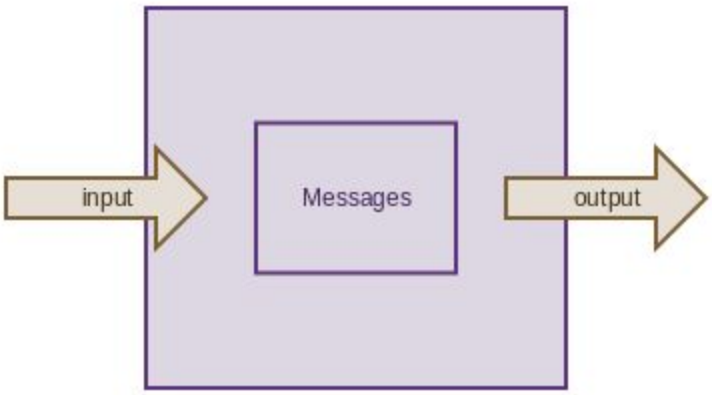
Note that for packets with same next hop and same priority application we are using our first-in-first-out logic.

4.d) Host Tracker :

Store : All the messages at every Node

Input : Messages contained at every Node and its current status

Output : Starting Node, Destination Node and the Current Node where the Message is present for each and every Message



Description :

Input will be taken from Traffic Manager output.
It keeps an account of message ID's present at every node.
This thus stores the position status of every message ID with respect to its routing string.
For every message ID, we get the node number where that message is present, as an output.

Assumption:

We assume that we get the status of each message ID from the traffic manager correctly at each clock cycle.

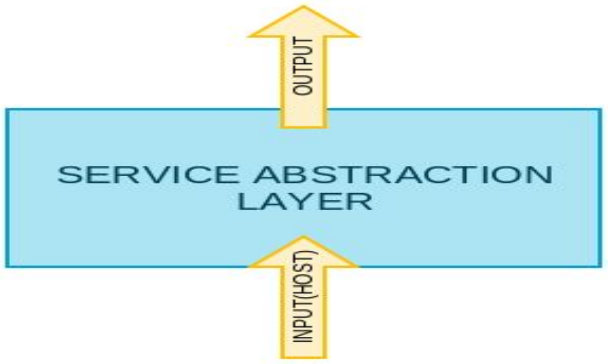
Note:The host tracker is combined with the Traffic Module in the code as its functionality was being satisfied in it.

4.e) Service Abstraction Layer : Top Module

Store : Real Time Status of each and every link between Nodes

Input : The connectivity of the links at every Node and instances of new link Generation

Output : The Adjacency matrix at every 20 clock cycles according to the status of the Links. Also information of New links created or Old links destroyed is given at running clock cycle.



Description :

The input to this is the south-bound traffic activity.
We basically know if suddenly a link fails, via SAL input.
Also, if there is packet drop, or there is packet delivery, this is intimated to SAL as input, and SAL processes this information giving it as its own output.
This output becomes input to all modules of SDN controller

Assumption:

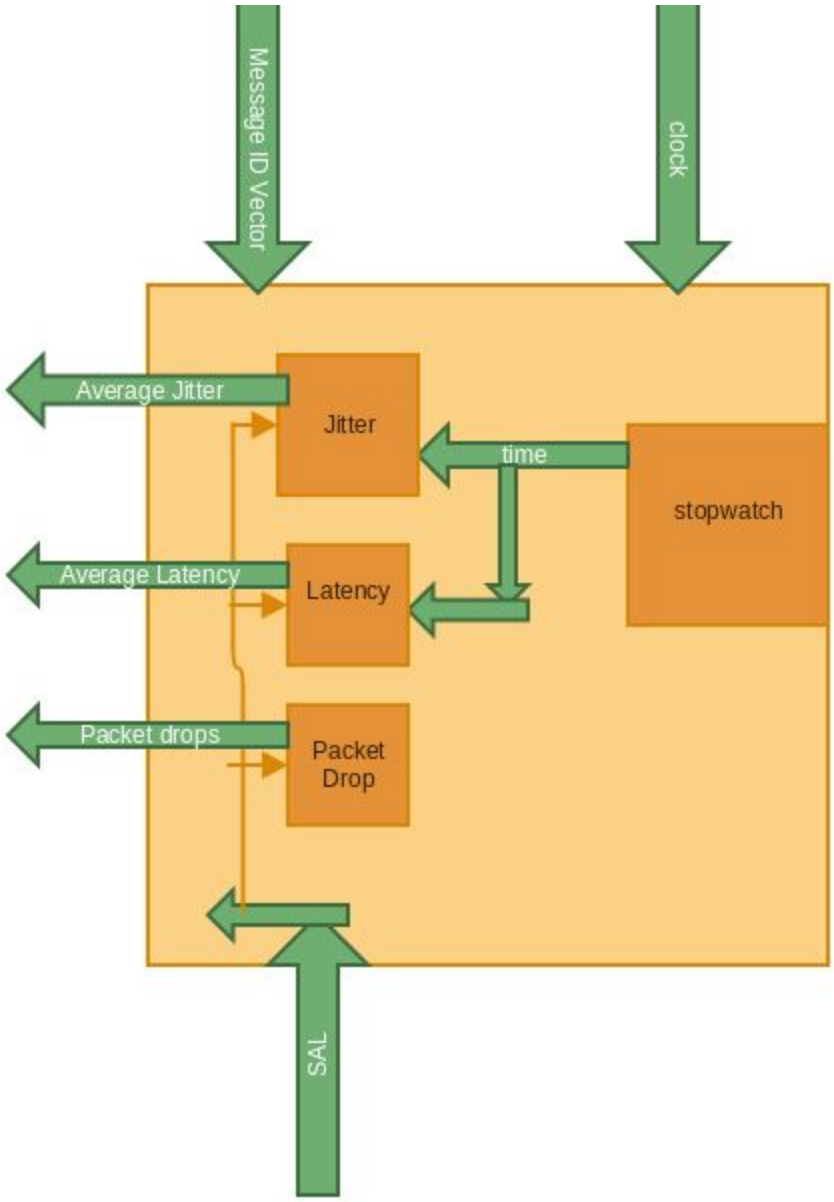
We assume that we are able to monitor the status at each router (regarding what packets a router has received at every clock cycle, what packets are dropped at a clock-cycle, and what packets are received finally at destination).

4.f) Statistics Manager :

Store : Time of traversal for every packet and total number of packets dropped.

Input : Message IDs of new messages added to request, reached destination and dropped

Output : Average Jitter, Average Latency and Total number of packets dropped.



Description :

1. This module takes in the message ID vector, clock and SAL output as its input.
2. SAL output relevant to this module simply consists a signal such that
    a. Signal indicates event when there was packet delivery at a particular destination node with corresponding message ID.
    b. Signal also indicates event whenever a router drops packet of a message ID because of queue length limitation.

3. So, whenever a destination node ‘x’ receives packet from source node ‘y’, the SAL informs the statistics module that this packet delivery is done.
4. The jitter and latency modules will attach the corresponding time at which the packet was received to that particular message ID entry.
5. Then jitter and latency is calculate in O(1) time for each such delivery.
6. Similarly, drop counter is incremented in the packet drop module whenever there is packet-drop at any router in O(1) time.

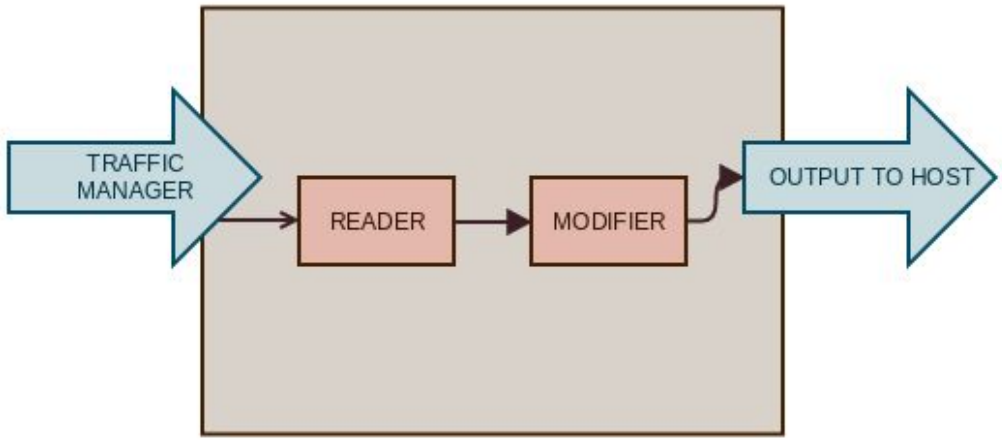
Assumption:

No special assumptions. As long as this module is getting inputs at the correct clock tick event, it will calculate the correct statistics.

4.g) Switch Manager (coded in the traffic module):

Input : Packets that are to be sent along with their respective traversing node and next-hop node

Output : Change the location of the packets according to the input along the links that are already existing. Also if a link is destroyed then send the negative acknowledgment to the source node of message



Description :

Takes input from Traffic Manager at each clock cycle and gives as output to update the message status with respect to routing string at different nodes into the host tracker.

Assumption :

We are getting the message position status from the traffic manager at every clock cycle for every message ID.

5. Plan for testing

Generation :

Generating 45-50 test cases using random function in python, which consists of:

1. NxN matrices
2. NxNxkxm matrix for messages

Expected Output :

1. We will collect the statistics and observe the trends for messages which reached , number of messages lost, time statistics.
2. If these matches real time statistics then our model is working fine.

Corner cases:

Case 1 :

Every node has number of messages to be sent equal to saturation limit.

Case 2:

Every message has high priority or every message has a common destination

6. Further planning

1. Increment k from k=1.
2. Show how message is being transmitted.
3. If possible we would like to simulate the network by network simulation tools and record data during the simulation. For this we need to consult our network's instructor and if time allows would continue with this implementation.

Final Stage Working

7.Contribution of Work:

Stats Manager: Shachi
Algorithm: Deep
Traffic Module, Switch Manager, Host Tracker: Neeladrishekhar and Ashna

Service Abstraction Layer, Combining, Testing, Debugging: All member.

Milestones Achieved :

1) Any level of of priority (k) [Will vary the input style accordingly]
2) Number of nodes and buffer strength of each node can be varied. All the variables need to be changed in the package and generic values in the respective Components accordingly.
3) Ability to display all the message packets currently in transit via real_time_T_Table signal.
4) Ability to show all the packets currently existing at every node in real_time_status signal.

8. OUTPUT

NextHop: type array of array of std_logic_vector

Here NextHop(i)(j) represents the next hop node from i'th node to j'th final destination.

Real_Time_Status:

Real_Time_Status(i)(j): represents the packet at node i's => j'th memory block.

Real_Time_Status(i)(j).cur : represents the node its currently present at(which is 'i').

Real_Time_Status(i)(j).dst : represents the destination node the packet is targeted at.

Real_Time_Status(i)(j).src : represents the source node the packet started at.

Real_Time_Status(i)(j).idN: represents the id of the packet .

Real_Time_Status(i)(j).k : represents the priority level **(application)**of the packet.

Real_Time_T_Table:

Real_Time_T_Table(i)(j) : represents the packet at node i destined to node j and will go to its next hop.

Real_Time_T_Table(i)(j).nxt : represents the next hop of that packet.

Real_Time_T_Table(i)(j).cur : represents the node it is currently present at.

Real_Time_T_Table(i)(j).idN : represents the id of the packet.

Real_Time_T_Table(i)(j).idx : represents the memory position of this packet in its current node.

Latency: integer - average end to end delay over all nodes

Jitter: integer - average variation of latency over all nodes

Total_Dropped: integer - total packets dropped

Latency_Status: array of array of integer

Latency_Status(i)(j) - represents the average latency of path from node i to node j.

Jitter_Status: array of array of integer

Jitter_Status(i)(j) - represents the average jitter of path from node i to node j.