Inter-working of Congestion Avoidance and Management

Adisak Mekkittikul

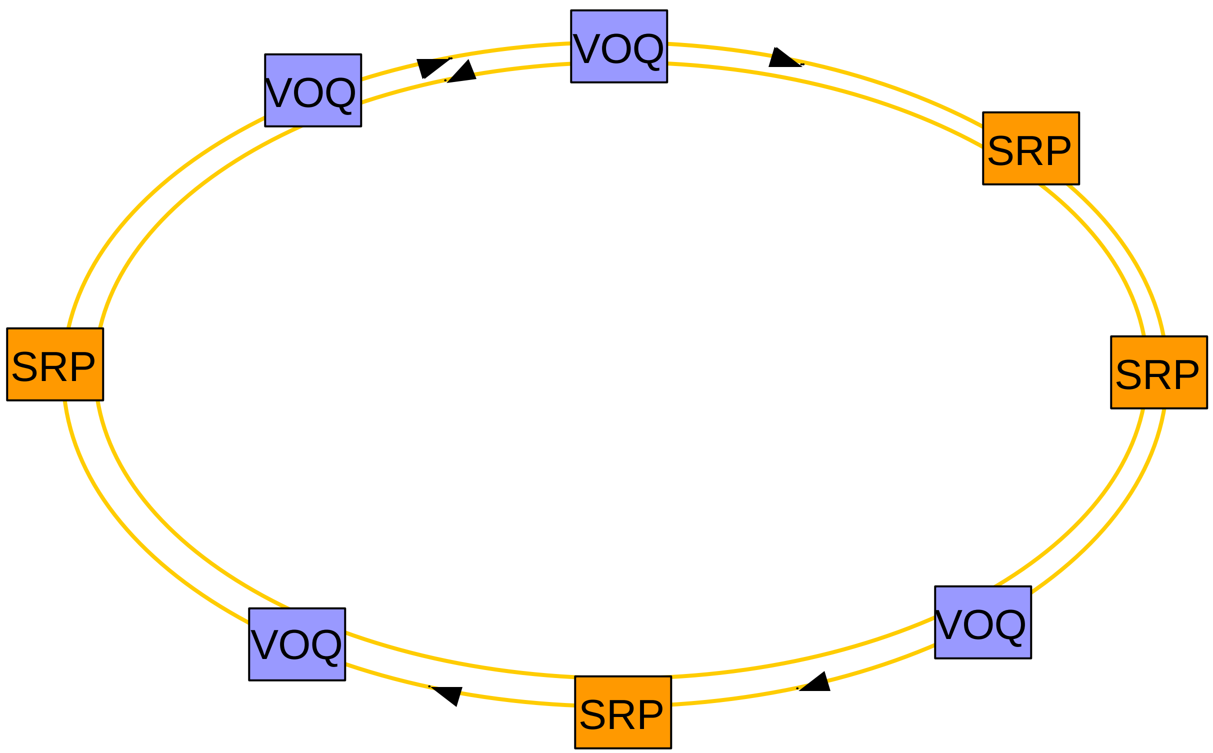
The Necessity of Inter-working

- Migration of Pre-standard deployment to full RPR
  - Minimum service interruption
  - Continue using existing equipments for useful life
- Multiple implementations of RPR MAC
- Multiple equipment suppliers
  - Alternate sourcing is crucial for carrier operation

Different Fairness Schemes Can Operate on the Same Ring?

- IF:
  - Interoperability conditions and the MAC behavior are well-defined

Case in Point: VoQ and SRP on the same ring



Two Fairness Algorithms with vast Differences: ***But they Inter-op***?

- Congestion Avoidance vs. Congestion Management
- Multiple-Choke points (VoQ, VDQ) vs. single choke point
- Single cut-through buffer vs. large transit buffers
- Rate based vs. Usage based

Inter-Operability Criteria

- SRP equal fairness properties should hold and all nodes should receive equal BW when congestion occurs.
- VoQ nodes, implementing per-destination queues, should not suffer HoL blocking.

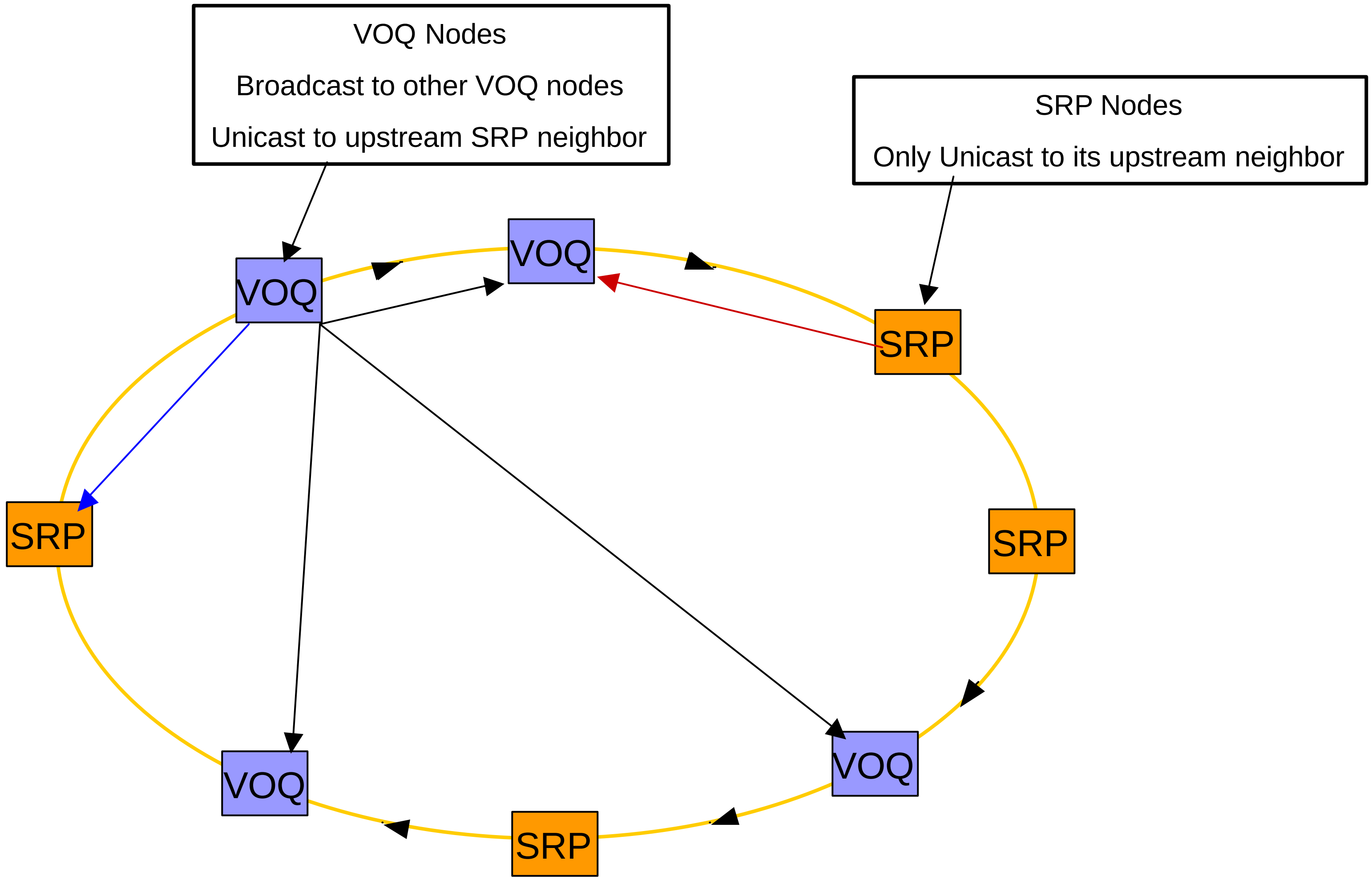
Assume no Change to SRP, what is required for the interop

- Packet format conversion
  - Most of the fields are already compatible
- Fairness message conversion
  - Translate usage to rate and vice versa
- Multiple choke points to single choke point conversion
  - SRP can understand only single choke point.

What needs to be implemented on a VoQ node

- VoQ node must understand SRP usage message from the downstream neighbor.
  - It must me able to convert the usage received in the SRP message to rcm.
- VoQ node must be able to send correct usage value to the upstream neighbor.
  - determine a single choke point from RCM message for SRP.
  - Convert the rcm value to usage for the SRP node.

Sending BW Messaging



Receiving BW Messaging

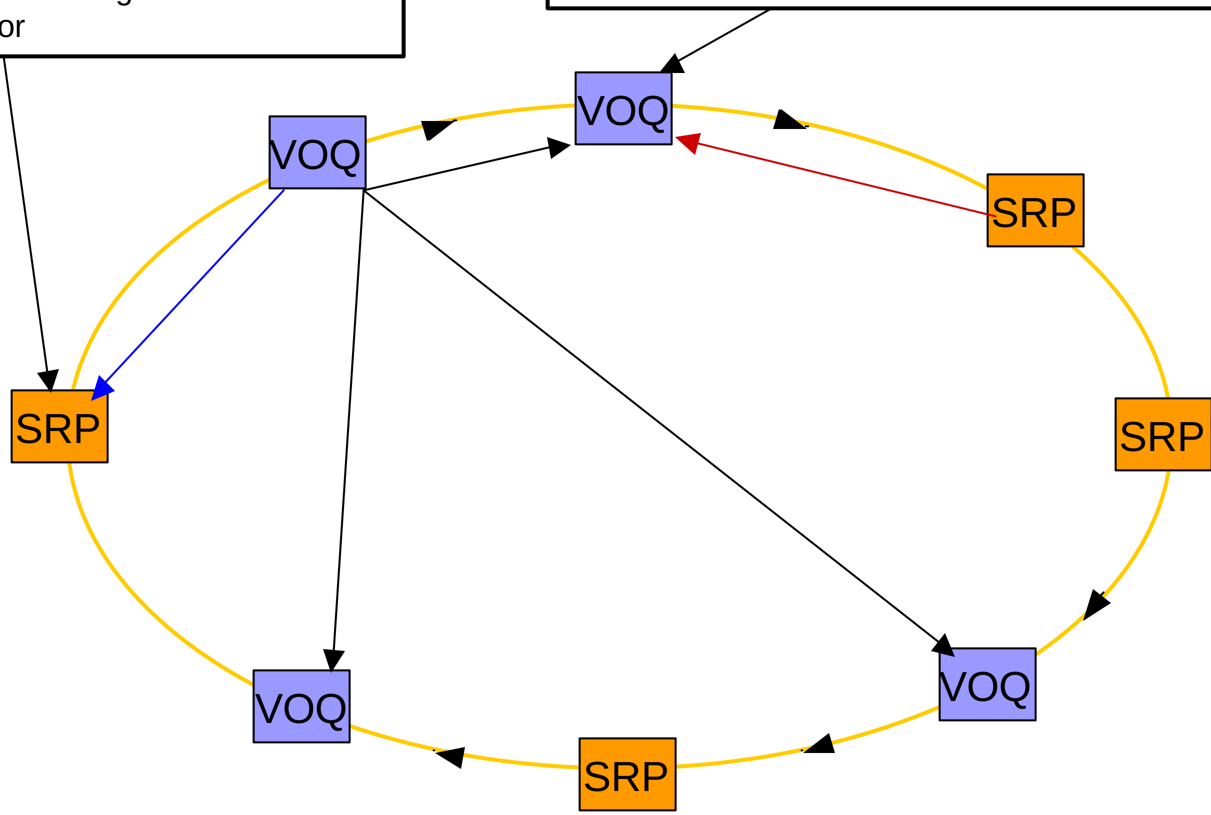
SRP Nodes

Ignore Broadcast messages from VOQ nodes

Receive only unicast message from downstream neighbor

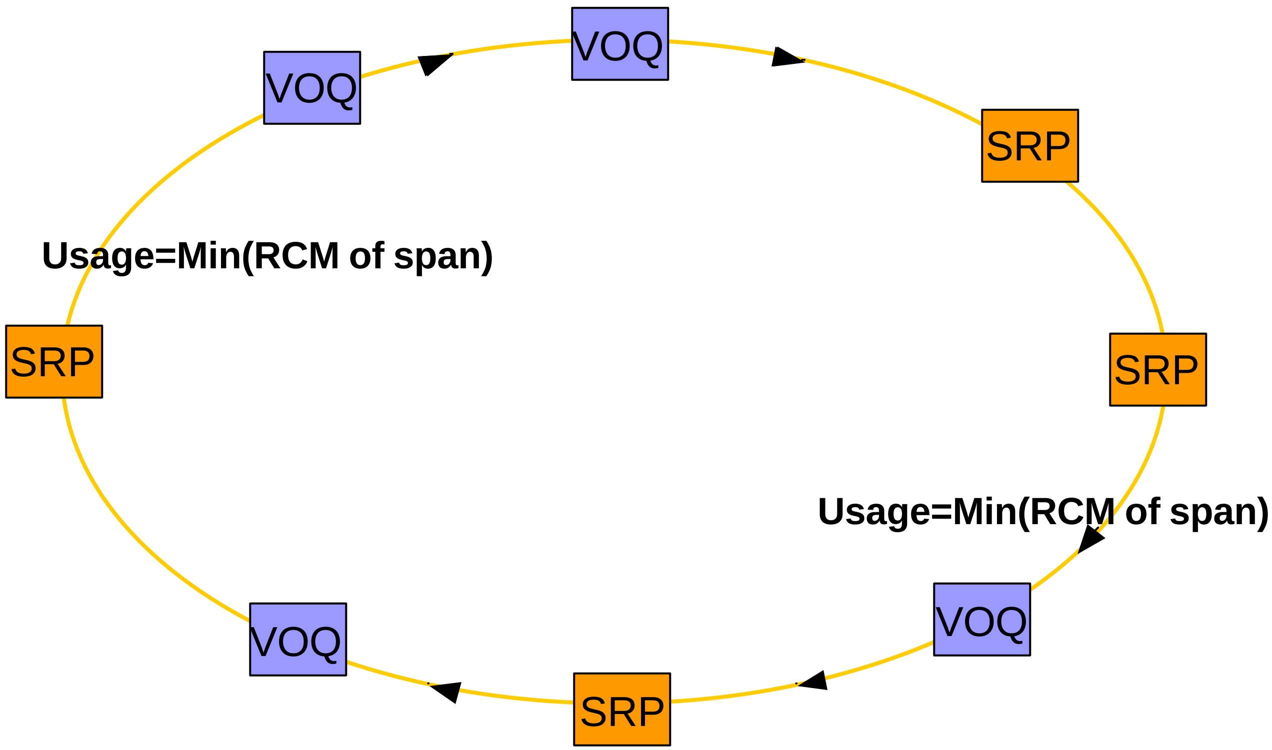
VOQ Nodes

Receive both broadcast messages from the other VoQ nodes and unicast message from downstream SRP neighbor



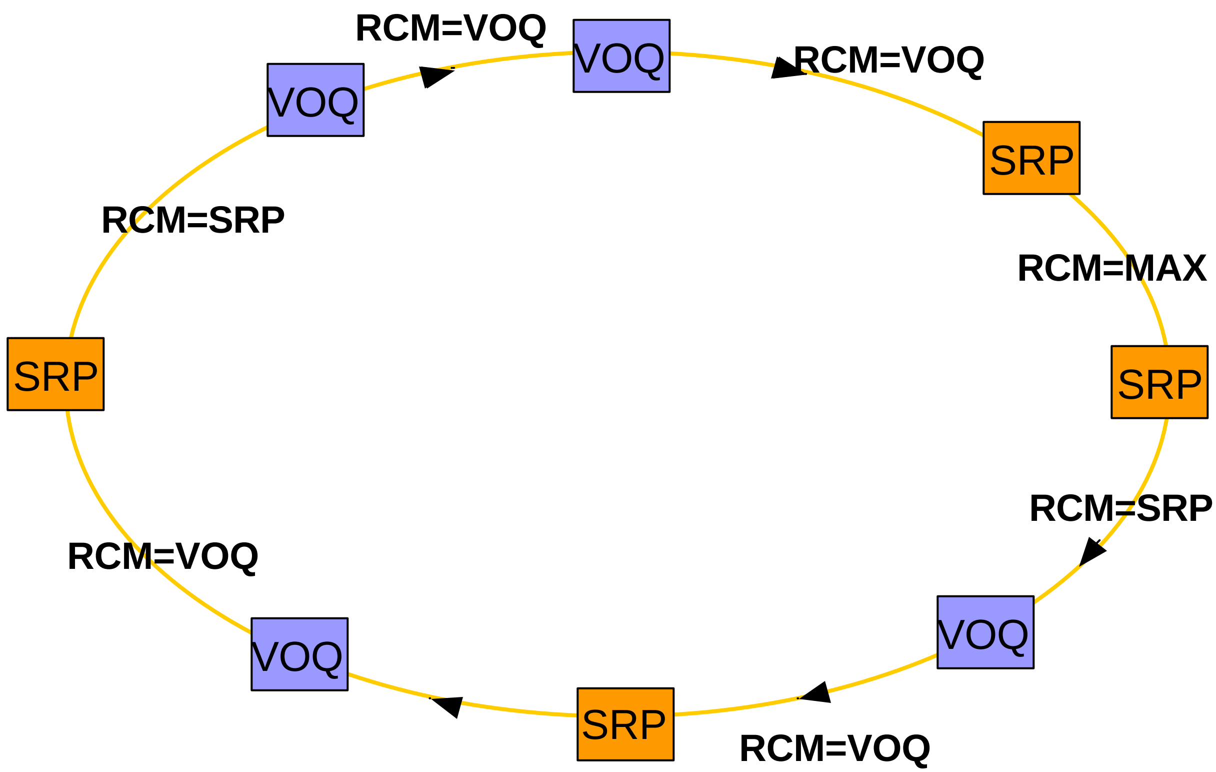
Usage View (SRP node)

See only the most congested node



RCM View (VOQ node)

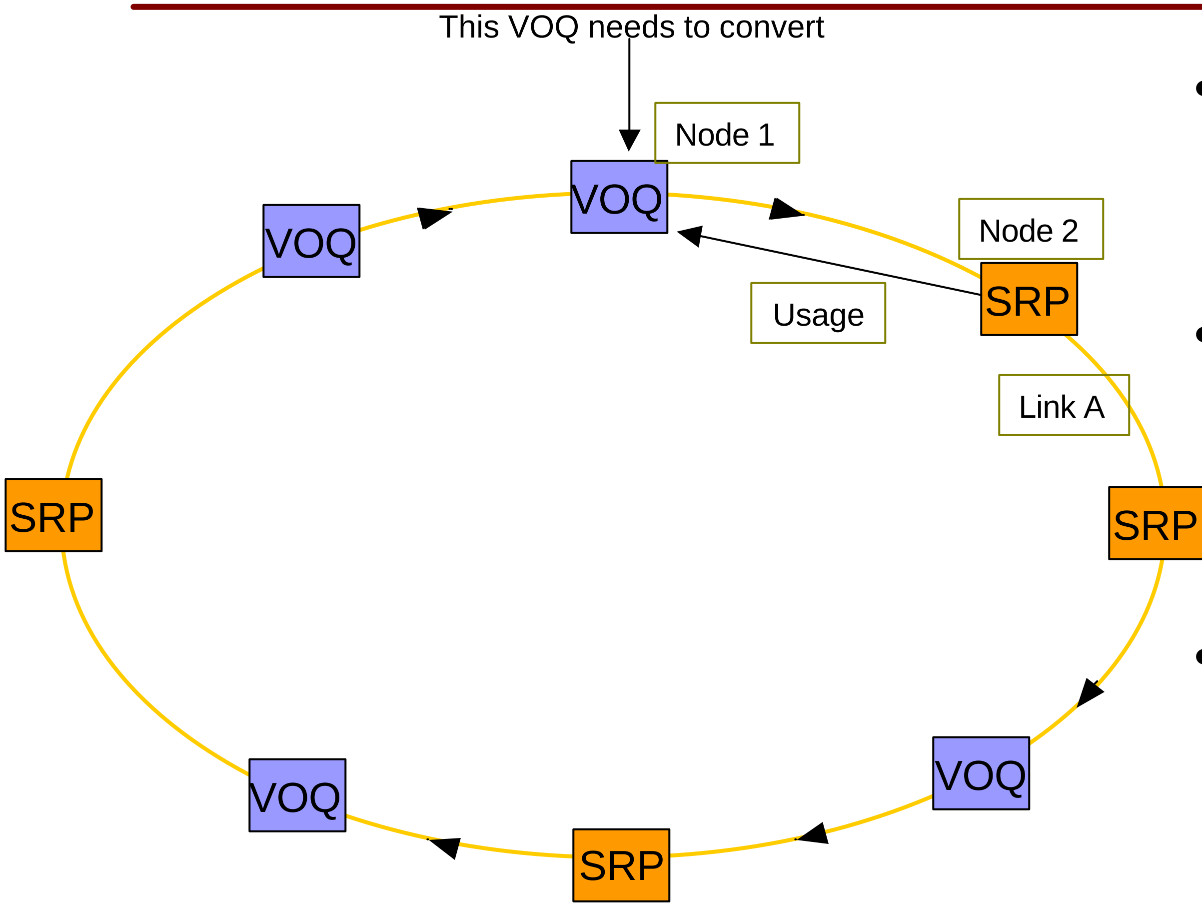
See all nodes' RCMs and fill silent SRP nodes with RCM=MAX



BW and usage conversion

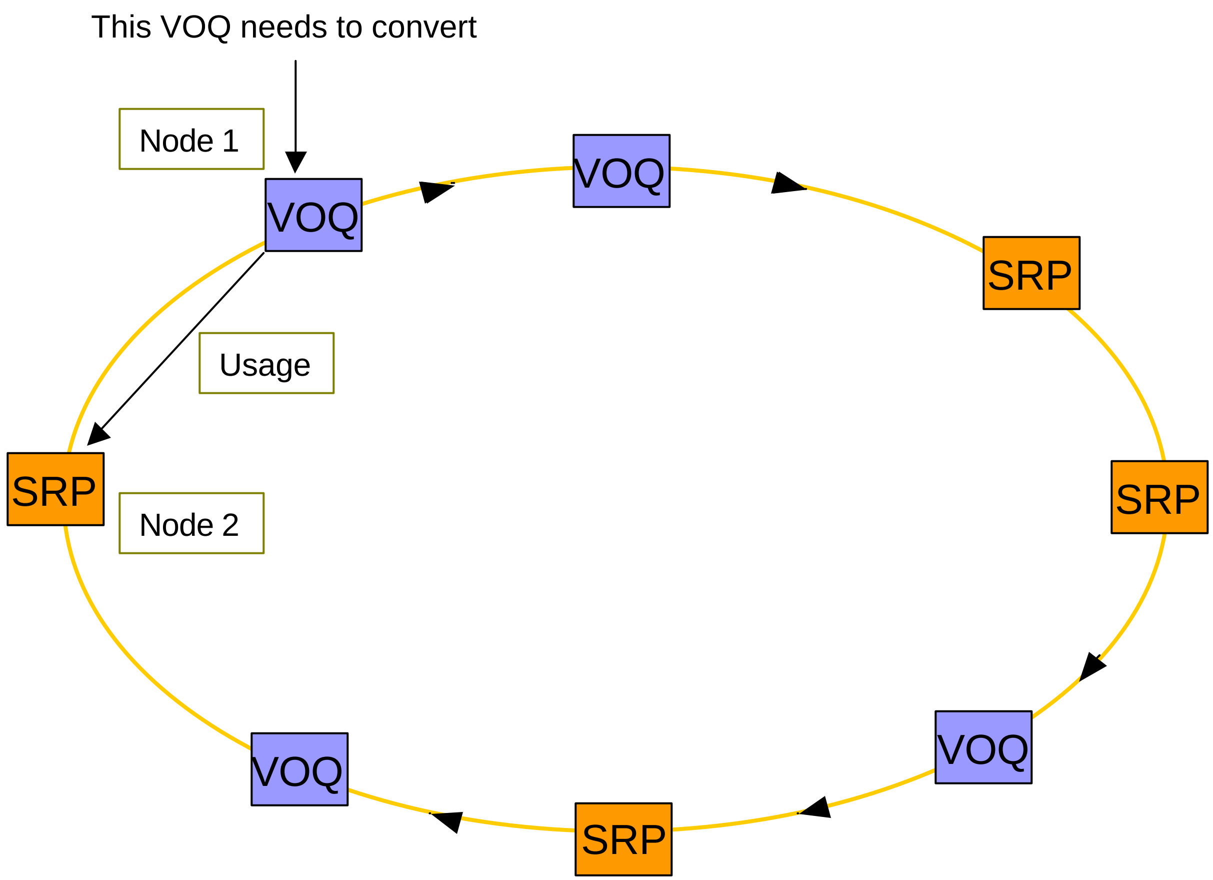
- SRP advertises a usage value instead of a rate value.
- This usage is the byte count over the decay interval.
- Usage over a fixed interval == “rate”

Receiving usage from SRP



- VOQ node 1 receives usage from downstream SRP node 2.
- This usage is each node's share on link A, and hence translates to rcm of SRP node 2.
- Convert usage to BW.
  - BW = usage/decay_interval

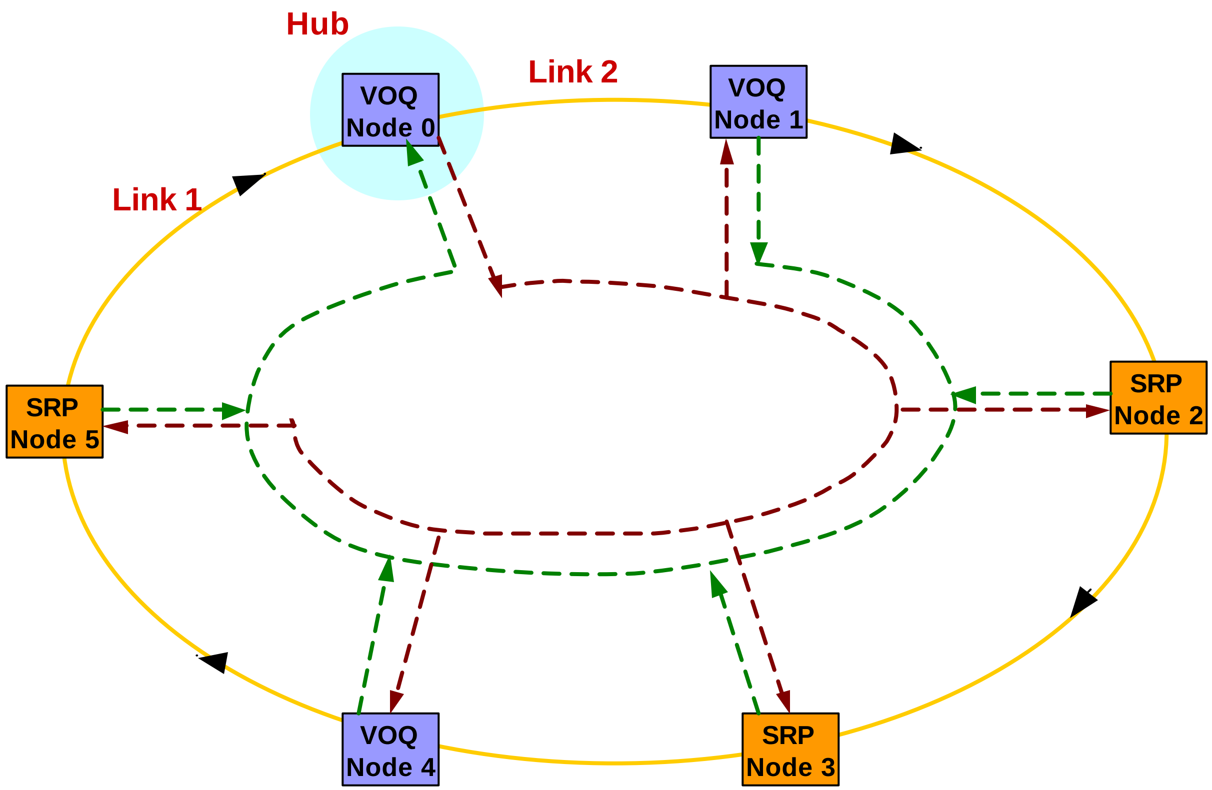
Sending usage to SRP



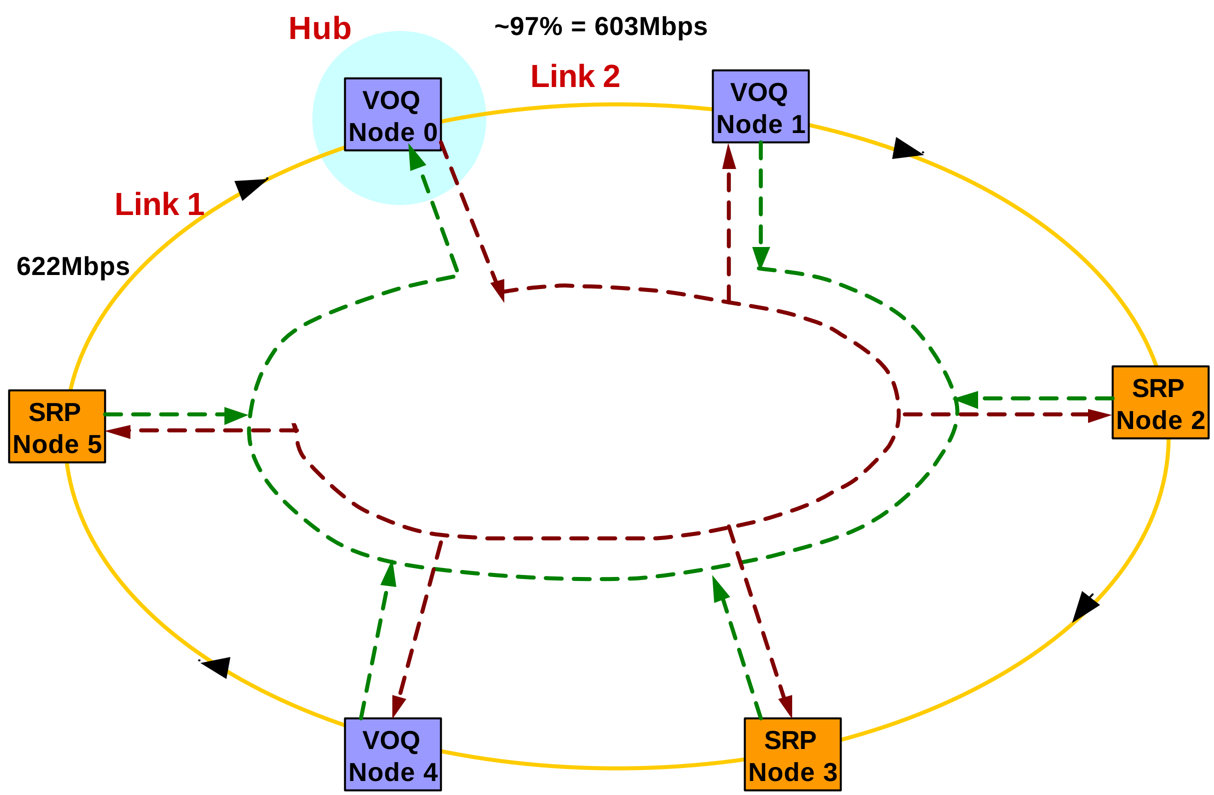
- VOQ node 1 selects rcm value to send to the upstream SRP node 2.
  - rcm = min (me, upstream nodes until rcm = link speed)
- This rcm is the equivalent to the usage for a single choke point.
- Convert BW to usage.
  - Usage = BW * decay_interval

Simulations

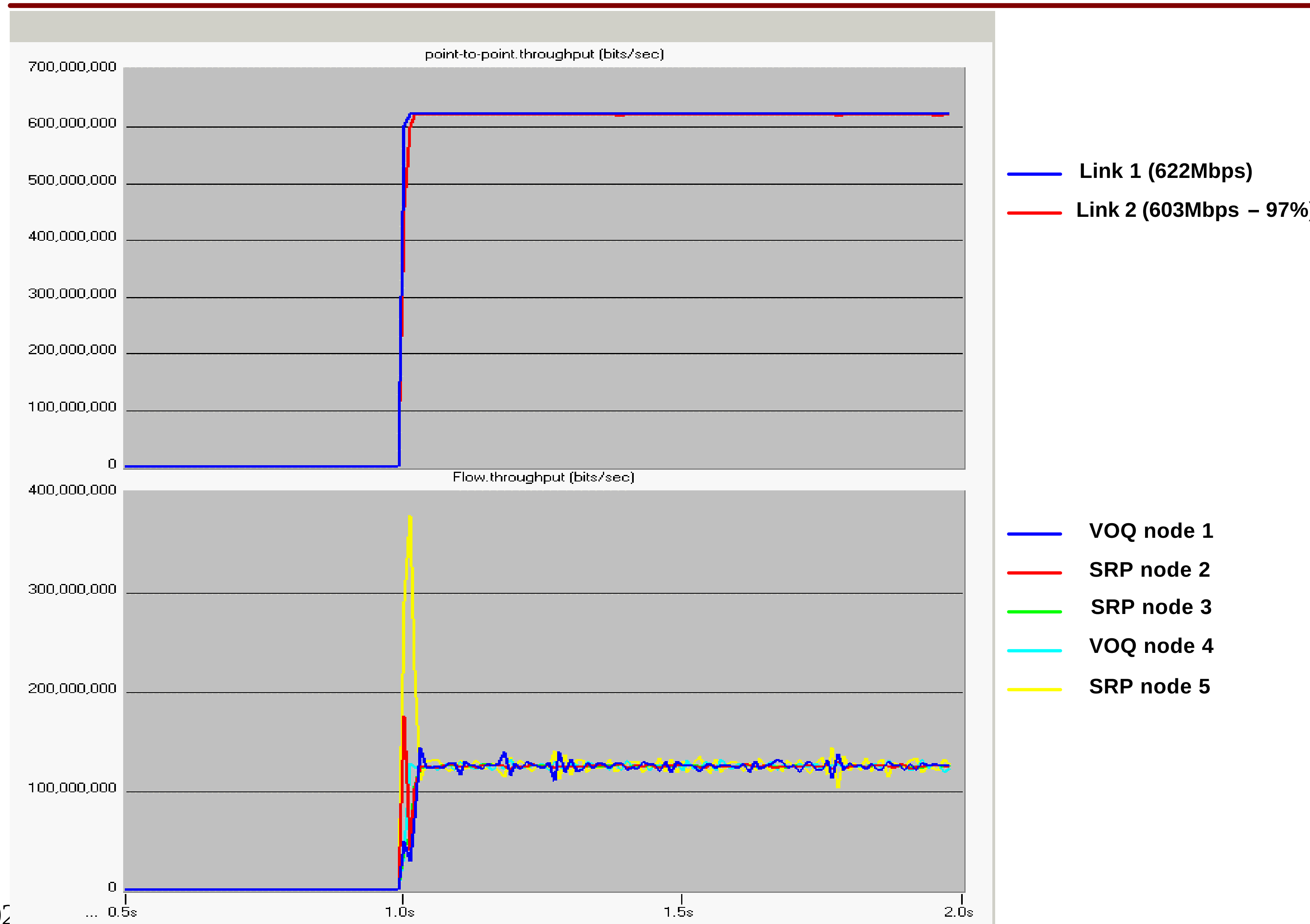
Scenario 1: BW fairness



Scenario 1: Expected results

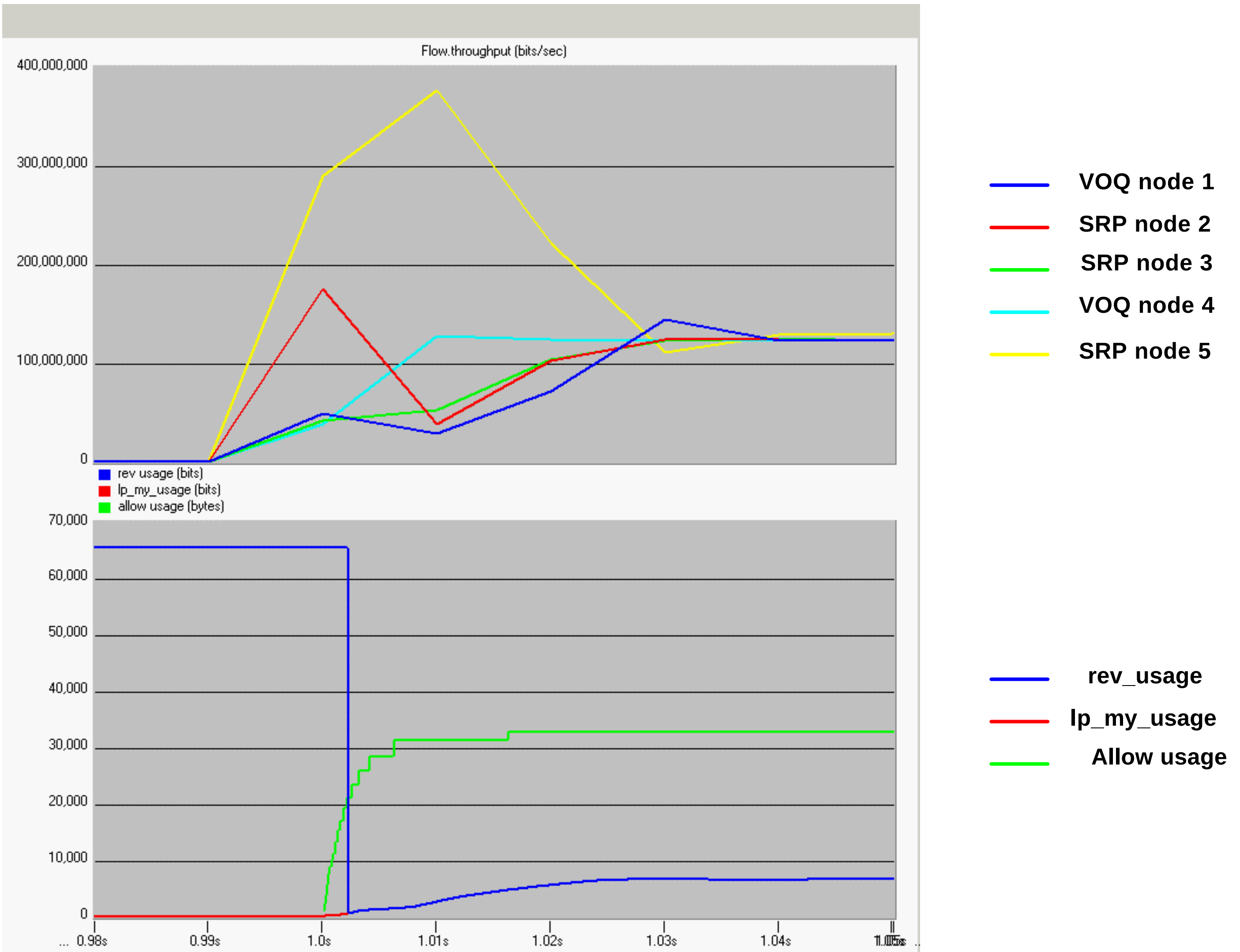


Results



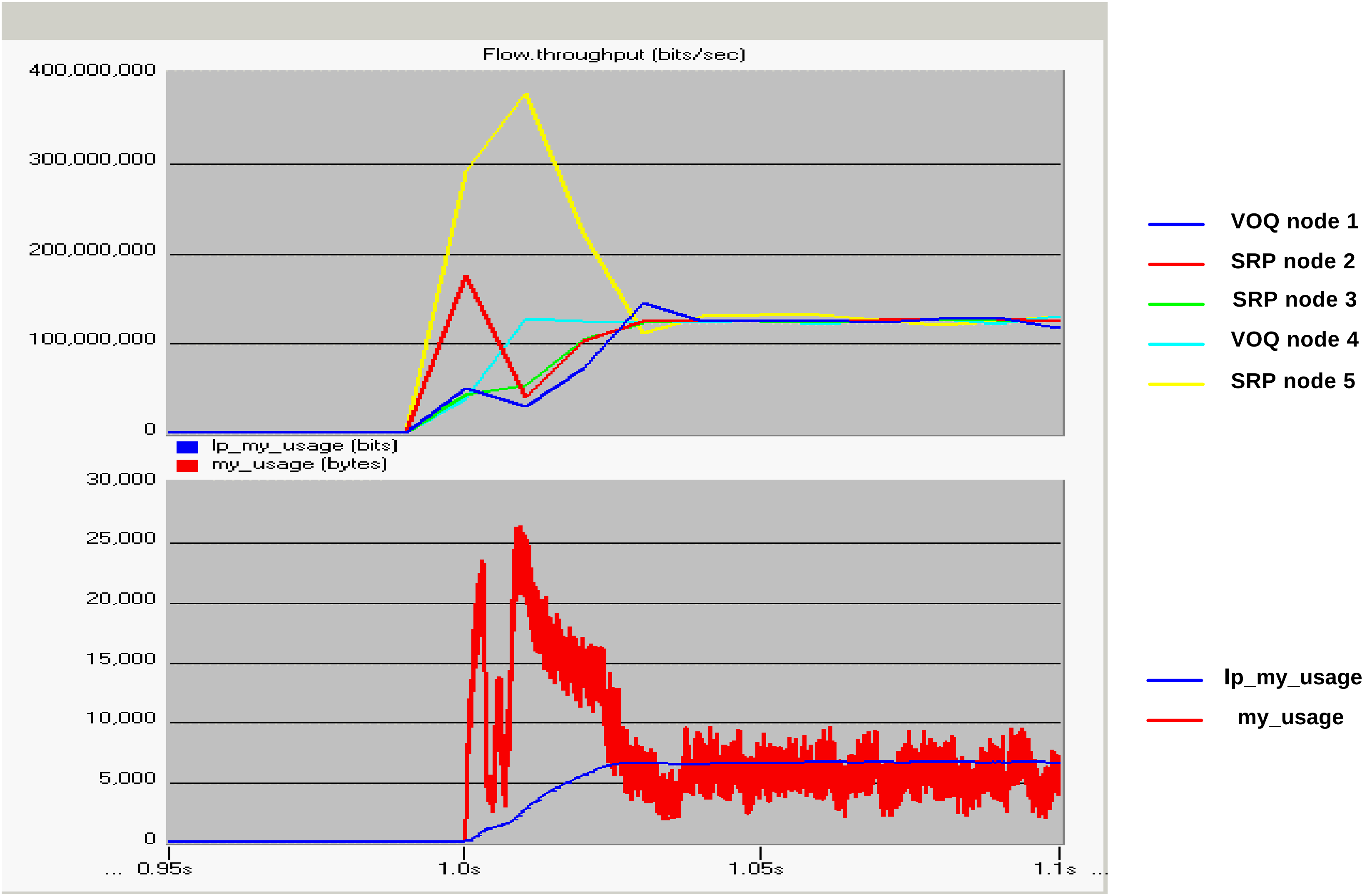
point-to-point.throughput (bits/sec)
700,000,000
600,000,000
500,000,000
400,000,000
300,000,000
200,000,000
100,000,000
0
Link 1 (622Mbps)
Link 2 (603Mbps – 97%)
Flow.throughput (bits/sec)
400,000,000
300,000,000
200,000,000
100,000,000
0
... 0.5s
1.0s
1.5s
2.0s
VOQ node 1
SRP node 2
SRP node 3
VOQ node 4
SRP node 5

Results



Flow.throughput (bits/sec)
400,000,000
300,000,000
200,000,000
100,000,000
0
rev usage (bits)
lp_my_usage (bits)
allow usage (bytes)
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
0.98s
0.99s
1.0s
1.01s
1.02s
1.03s
1.04s
VOQ node 1
SRP node 2
SRP node 3
VOQ node 4
SRP node 5
rev_usage
lp_my_usage
Allow usage

Results

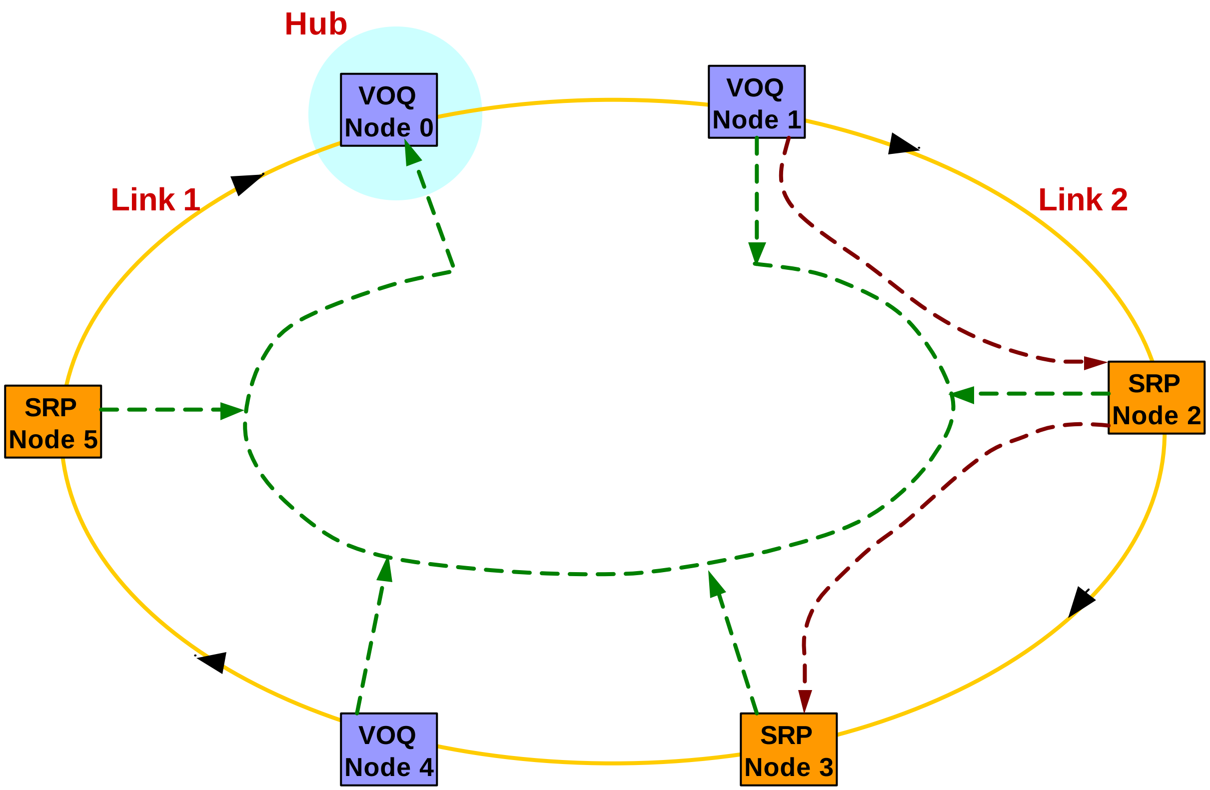


Flow.throughput (bits/sec)
400,000,000
300,000,000
200,000,000
100,000,000
0
lp_my_usage (bits)
my_usage (bytes)
30,000
25,000
20,000
15,000
10,000
5,000
0
... 0.95s
1.0s
1.05s
1.1s ...
VOQ node 1
SRP node 2
SRP node 3
VOQ node 4
SRP node 5
lp_my_usage
my_usage

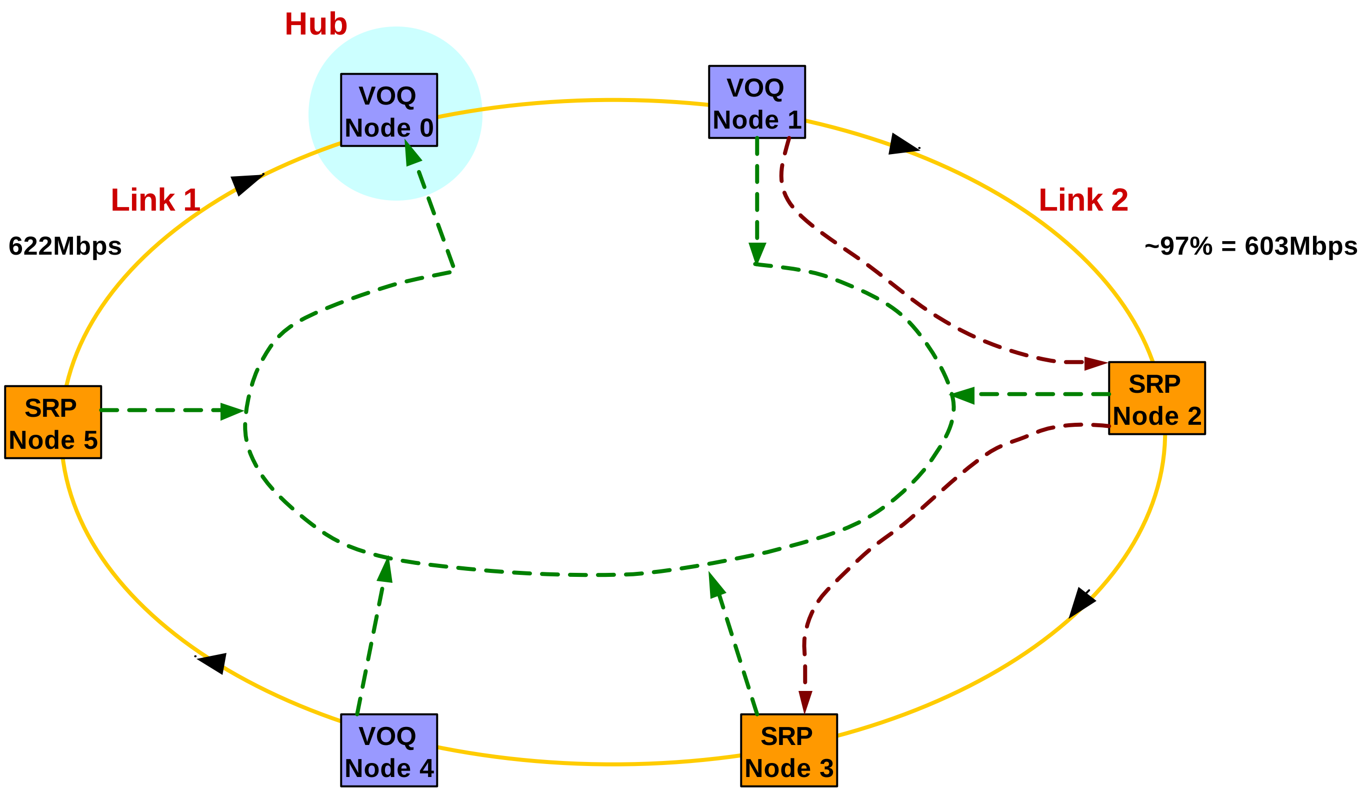
Observations

- Link into the hub is fully utilized.
- All nodes (1-5) get an equal share of link 1. Therefore, bandwidth allocation is fair between VOQ and SRP nodes.
- The VOQ hub is able to send maximum traffic.

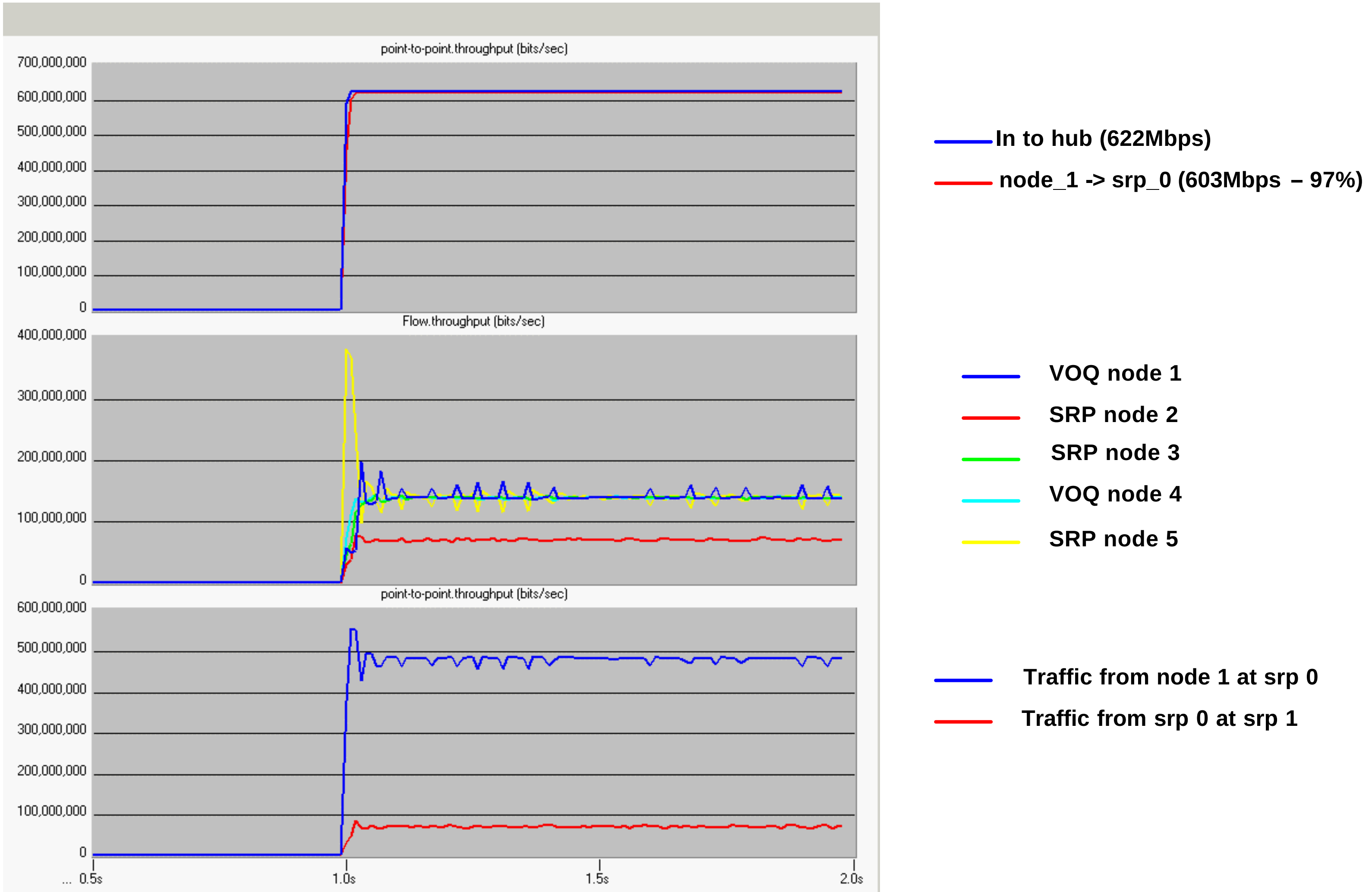
Scenario 2: No HOL Blocking



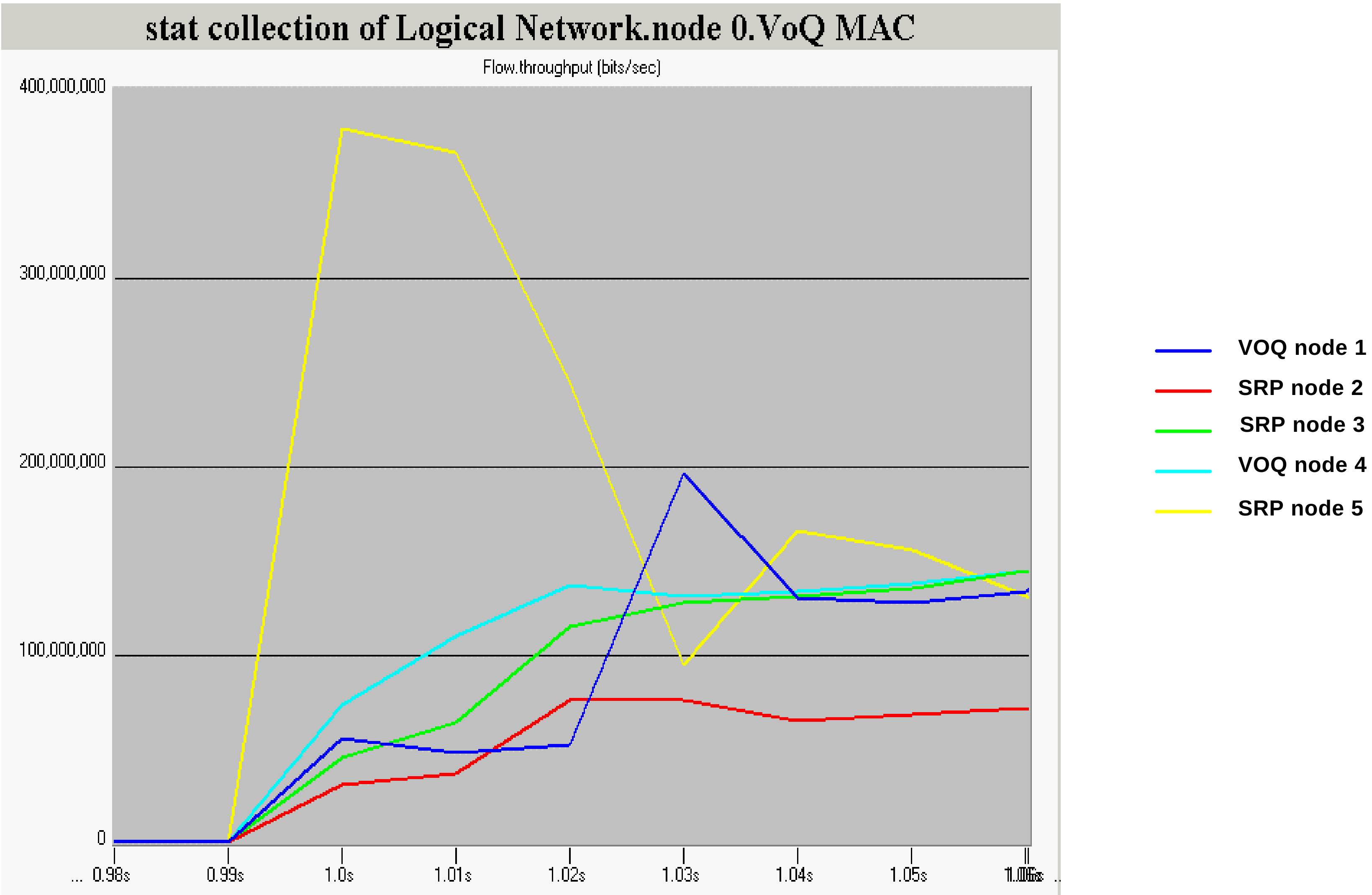
Scenario 2: Expected results



Results


point-to-point.throughput (bits/sec)
700,000,000
600,000,000
500,000,000
400,000,000
300,000,000
200,000,000
100,000,000
0
Flow.throughput (bits/sec)
400,000,000
300,000,000
200,000,000
100,000,000
0
point-to-point.throughput (bits/sec)
600,000,000
500,000,000
400,000,000
300,000,000
200,000,000
100,000,000
0
... 0.5s
1.0s
1.5s
2.0s
In to hub (622Mbps)
node_1 -> srp_0 (603Mbps – 97%)
VOQ node 1
SRP node 2
SRP node 3
VOQ node 4
SRP node 5
Traffic from node 1 at srp 0
Traffic from srp 0 at srp 1

Results (backup)



Observations

- All nodes, both VoQ and SRP, receive equal bandwidth
- VoQ nodes experience no HoL blocking while SRP nodes are unable to take advantage of spatial reuse as expected.

Summary

- VoQ and SRP can coexist on the same ring without adversely affecting each other's performance.
- Multiple choke point and single choke point fairness scheme can inter-work.
- Minimum change is required to emulate SRP fairness using VoQ, allowing ASIC to work in both modes.