Simulation Results

on

SRP Fairness Algorithm (SRP-fa)

Byoung-Joon (BJ) Lee

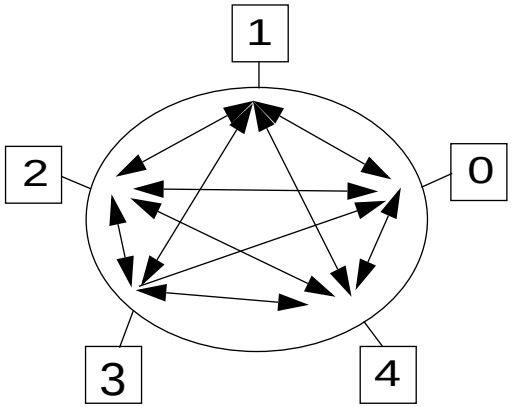
March 8, 2000

Overview

- Provide simulation results to demonstrate the behavour of SRP fairness mechanism under various traffic pattern and load scenarios.

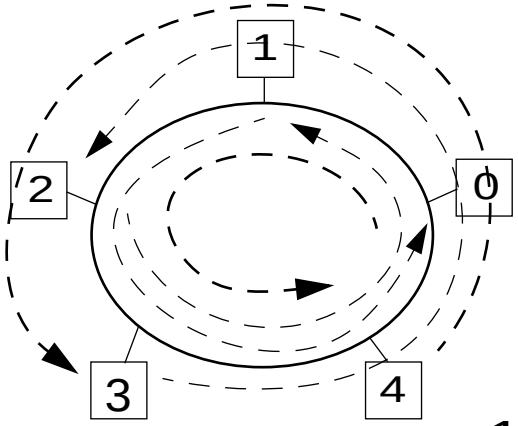
- Cases considered:
  - ❖ Homogeneous traffic
    - Mesh, Worst-case, Aggregation, Best-case
  - ❖ Heterogeneous traffic
    - Aggregation
  - ❖ Large number of nodes

Traffic Pattern Scenarios



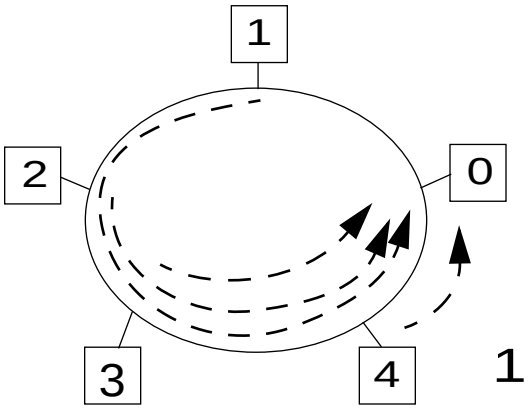
Mesh

200%



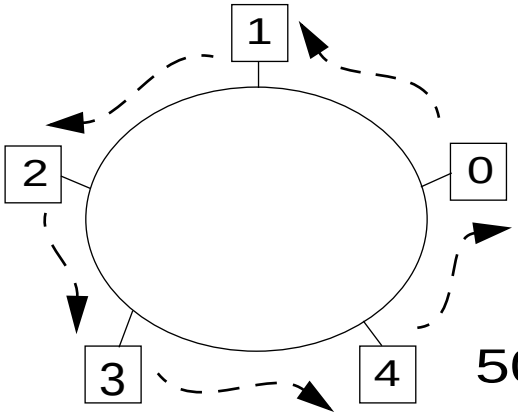
Worst-Case

120%



Aggregation

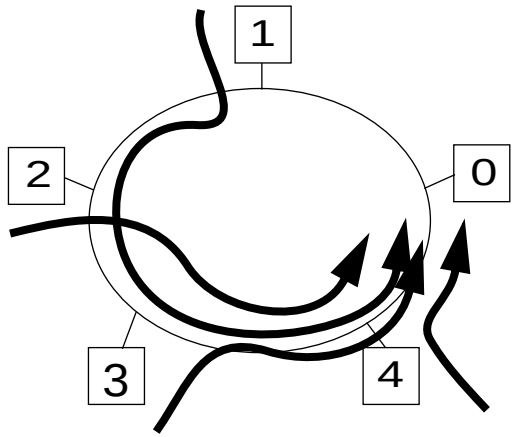
100%



Best-Case

500%

Traffic Pattern Scenarios *(cont'd)*



Aggregation (heterogeneous input)

❖ case 1

1 -> 0: 5 %
2 -> 0: 25 %
3 -> 0: 50 %
4 -> 0: 75 %

(Total: 155%)

❖ case 2

1 -> 0: 5 %
2 -> 0: 15 %
3 -> 0: 30 %
4 -> 0: 45 %

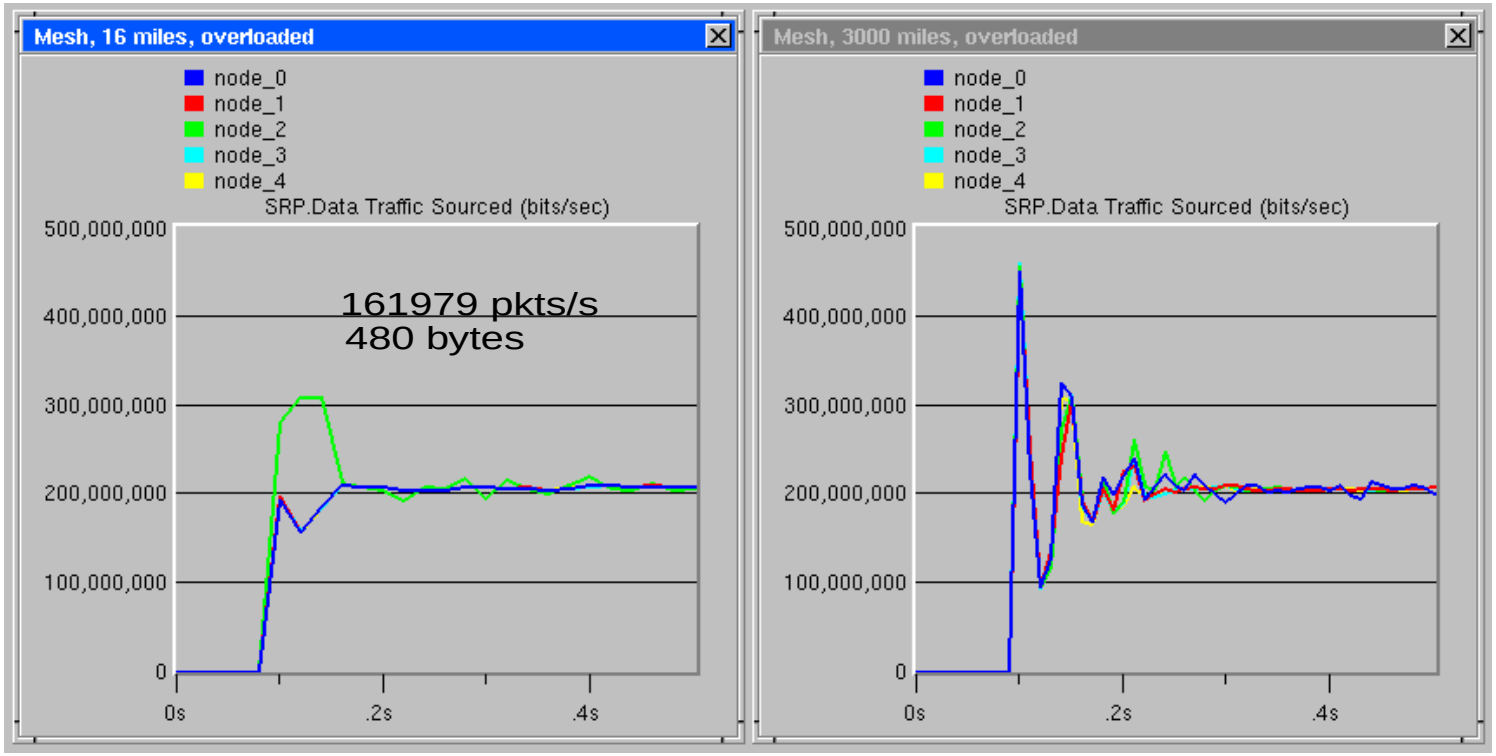
(Total: 95%)

Simulation Parameters

* 5 nodes with outer ring traffic only and equal-distanced

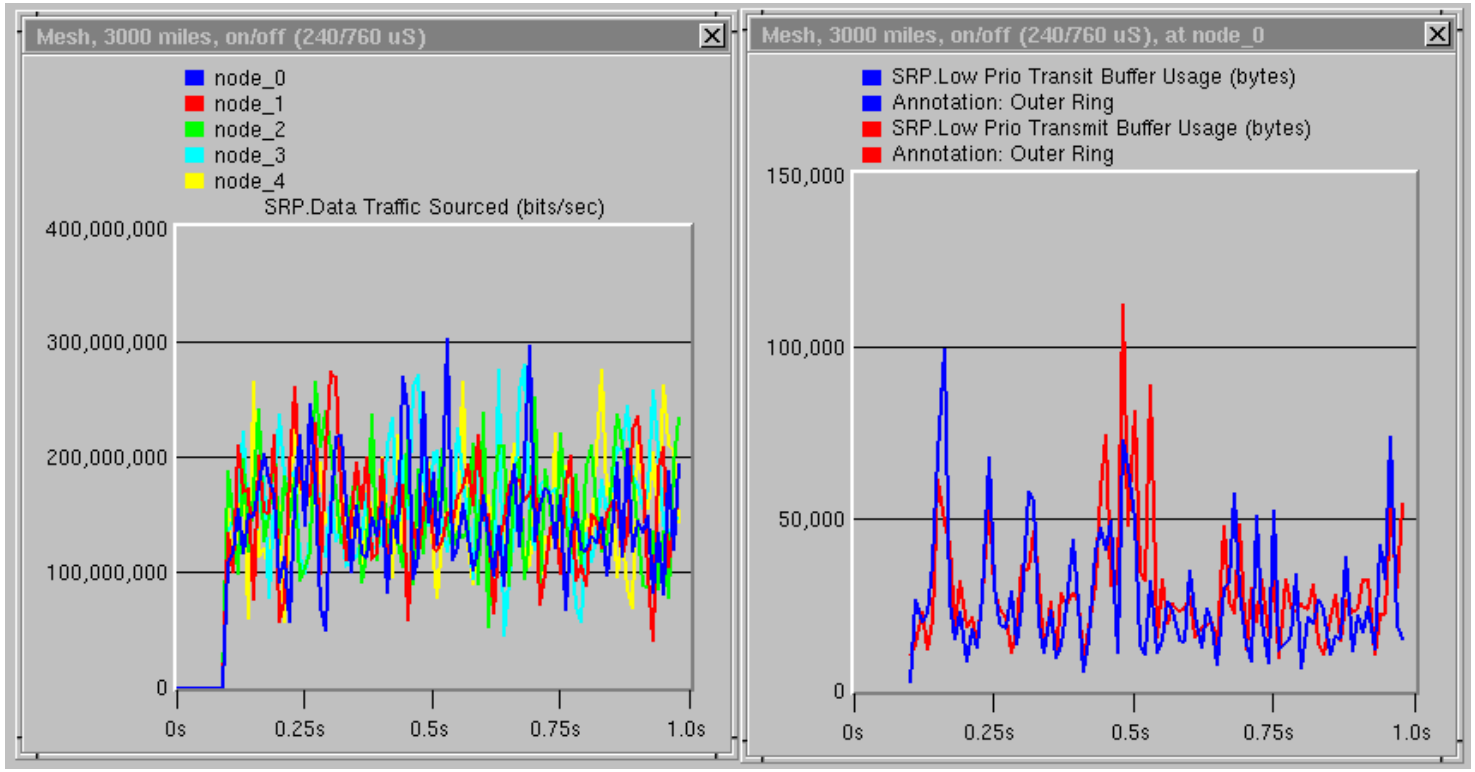
Ring circumference	:	16, 3000	(in miles)
Hi/Low priority ratio	:	varies	
Avg pkt rate	:	161979 pkts/sec	(exponential)
Avg pkt size	:	480 Bytes	(exponential)
Burst ON duration	:	varies	(exponential)
Burst OFF duration	:	varies	(exponential)
Tx high priority size	:	2.7 KB	
Tx low priority size	:	1 MB	
TB high priority size	:	2.7 KB	
TB low priority size	:	512 KB	(OC-12)
TB low threshold	:	256 KB	
TB high threshold	:	500 KB	(OC-12)
LP_FWD	:	64	(low pass filetering weight)
LP_MU	:	512	
LP_ALLOW	:	64	
AGECOEFF	:	4	(Aging Coefficient Value)
Decay Interval	:	32000	(in byte times, 100μS)
Usage Interval	:	Decay Interval + Δ	

Results: Mesh, Homogeneous, Overloaded, OC-12



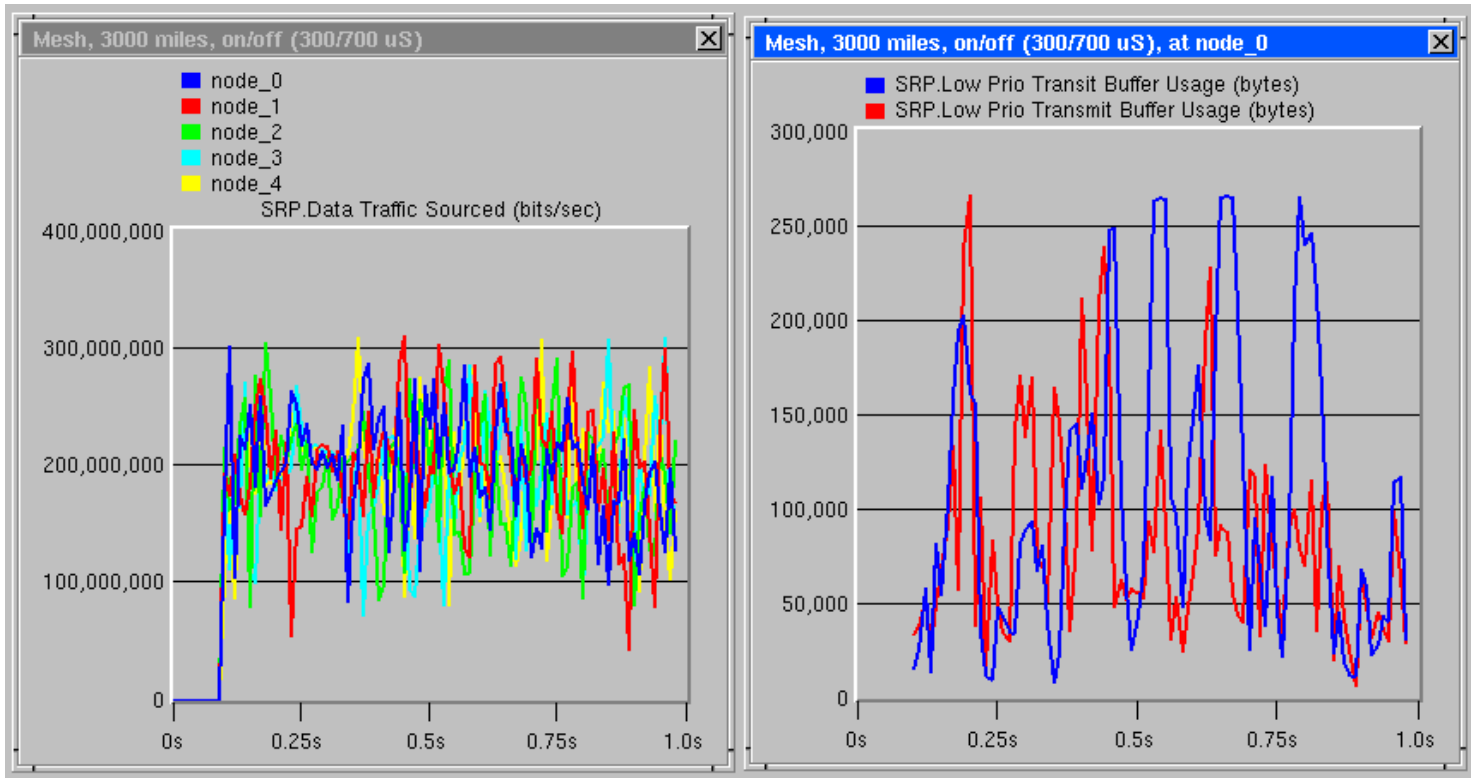
- Throughput achieves 160% of link capacity (spatial reuse)
- With small ring size, faster convergence (16 vs 3000 miles)

Results: Mesh, on/off (240/760 uS), 3000 miles, OC-12



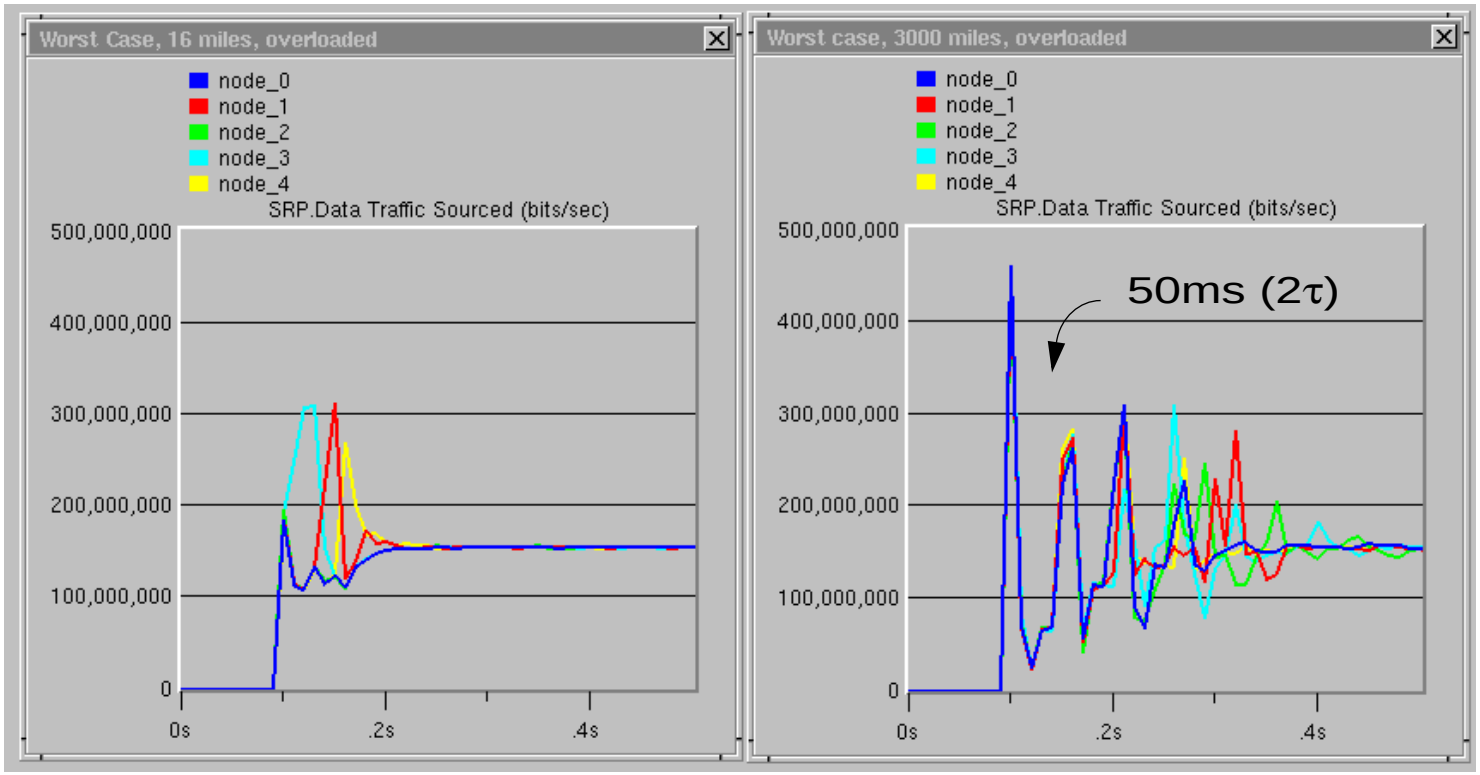
- 24% x 5 = 120% of link capacity (spatial reuse)
- TB and Tx occupancy indicate underloaded condition

Results: Mesh, on/off (300/700 uS), 3000 miles, OC-12



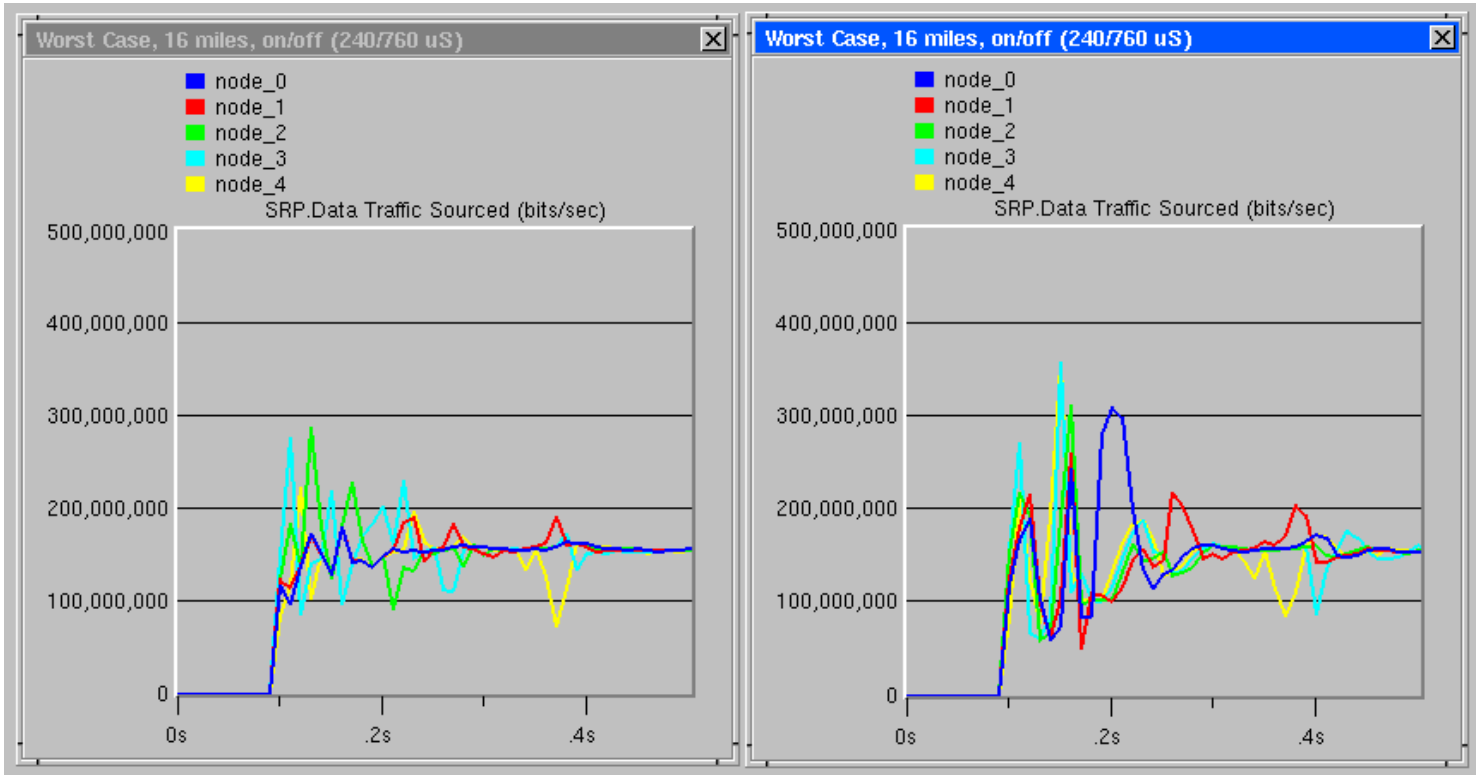
- 30% x 5 = 150% of link capacity (spatial reuse)
- TB and Tx occupancy indicate heavier utilization

Results: Worst Case, Overloaded, OC-12



- Each node can source 622/4=155Mbps (25% of link cap.)
- Convergence is worse when compared to mesh pattern

Results: Worst Case, on/off (240/760 uS), OC-12

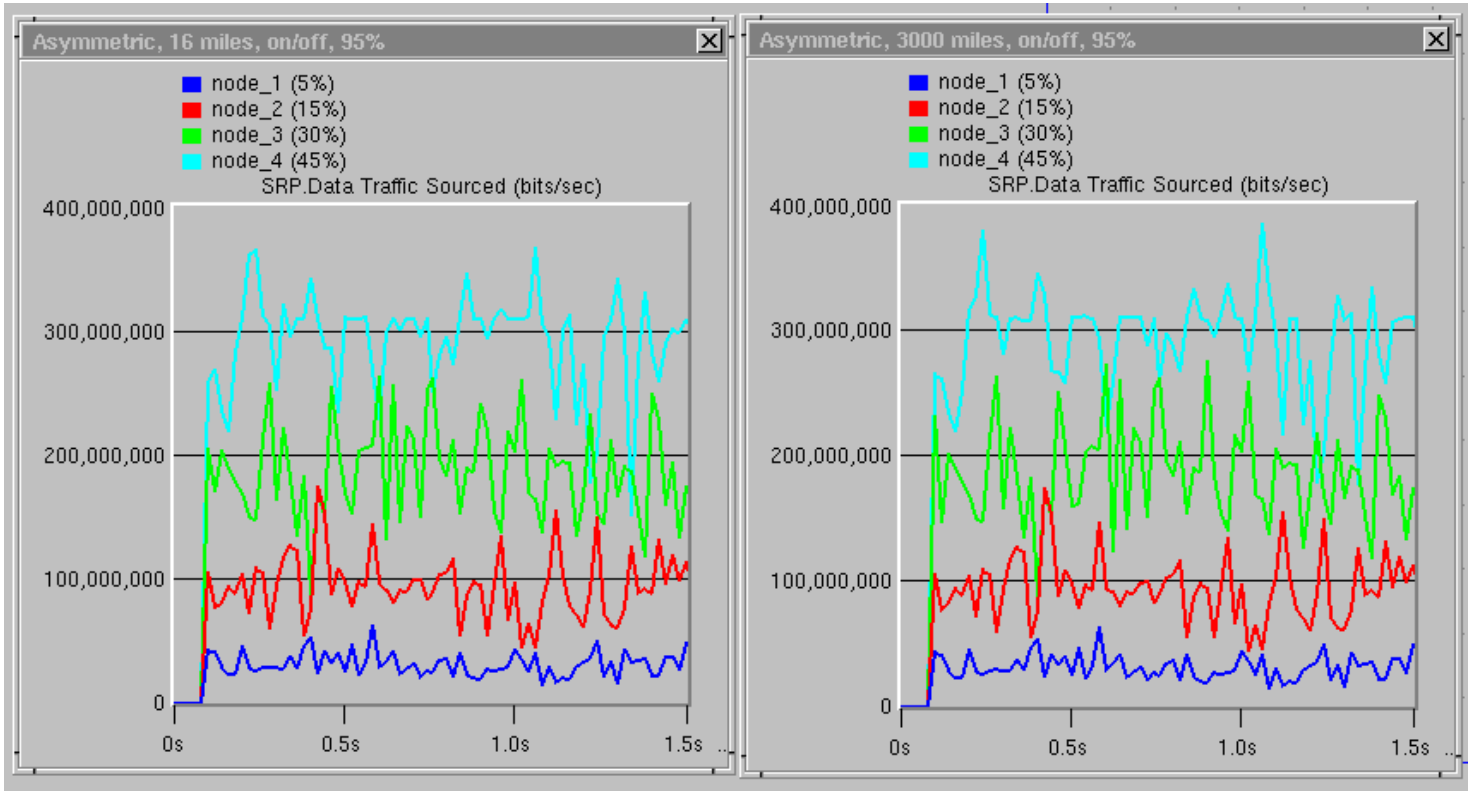


- Each node transmit 24% of link cap. on the average
- Heavy traffic condition with bursty input

Results: Asymmetric (case 1), 155%, OC-12

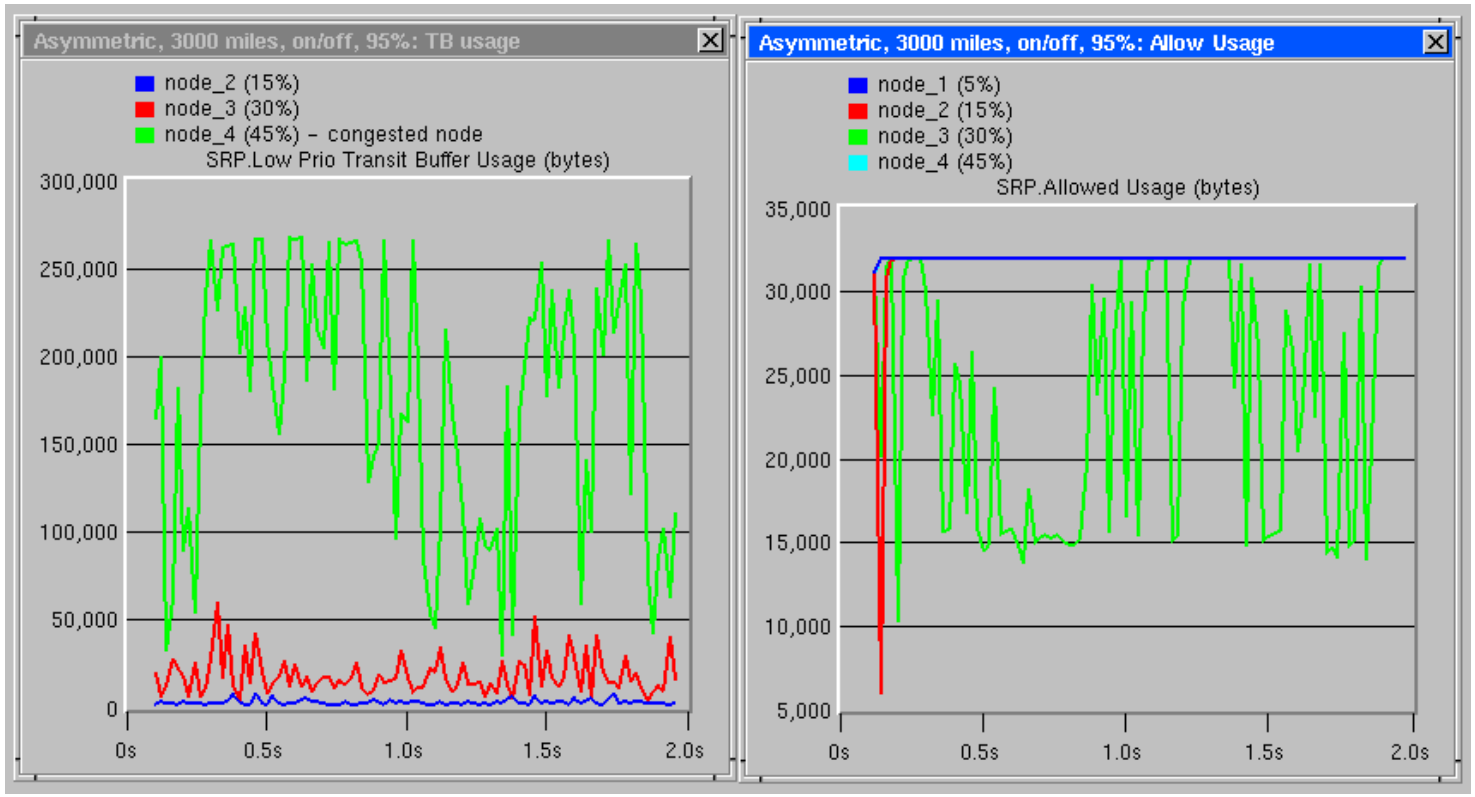
- Under overload, 75% and 50% nodes share 30% each, while small input nodes are able to send all their inputs. SRP achieves total of 90% link utilization.

Results: Asymmetric (case 2), 95%, OC-12



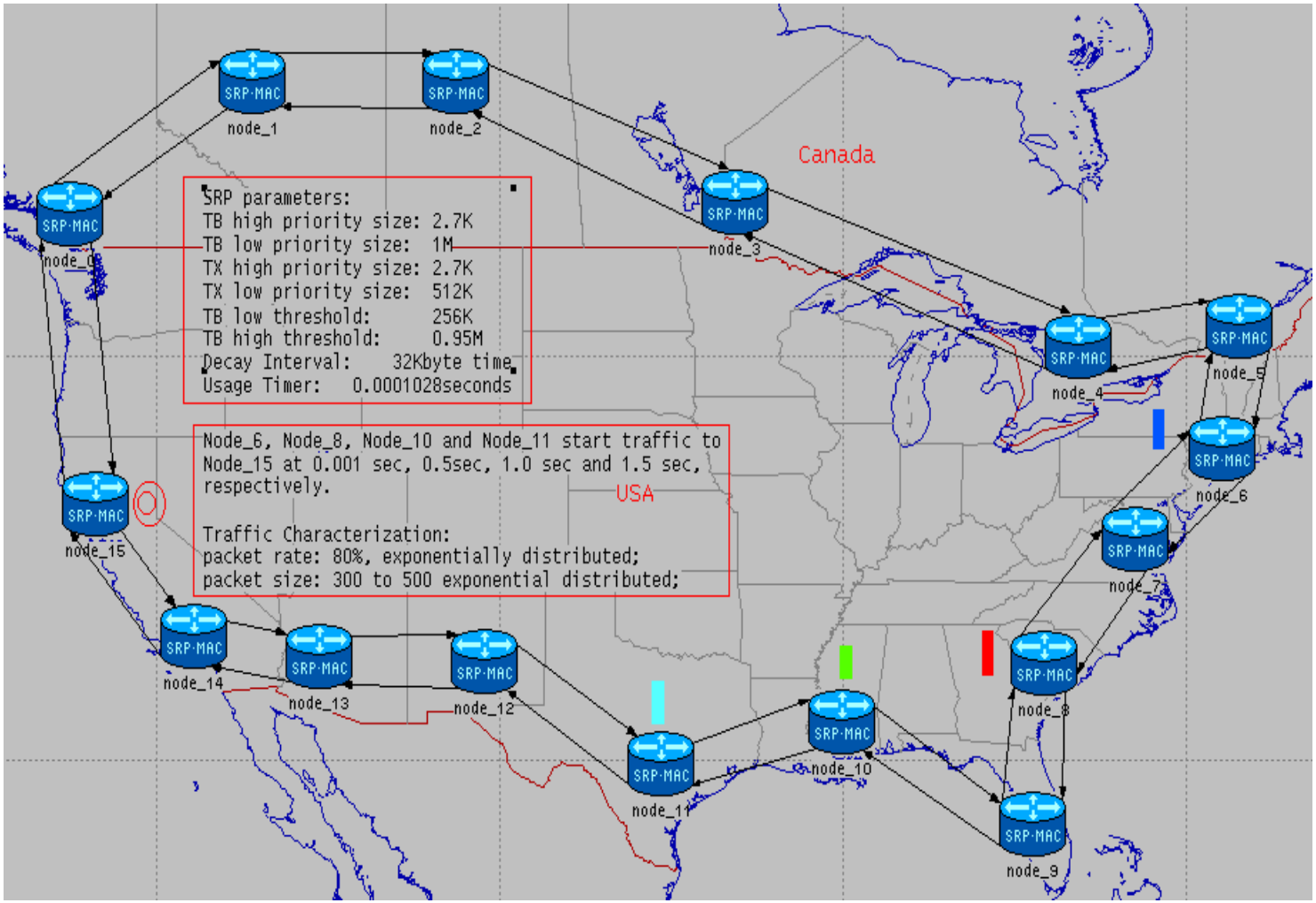
- Under 95% input load, all the nodes are able to send their full input trraffic. Similar performance both with 16 and 3000 miles.

Results: Asymmetric (case 2), 95%, OC-12

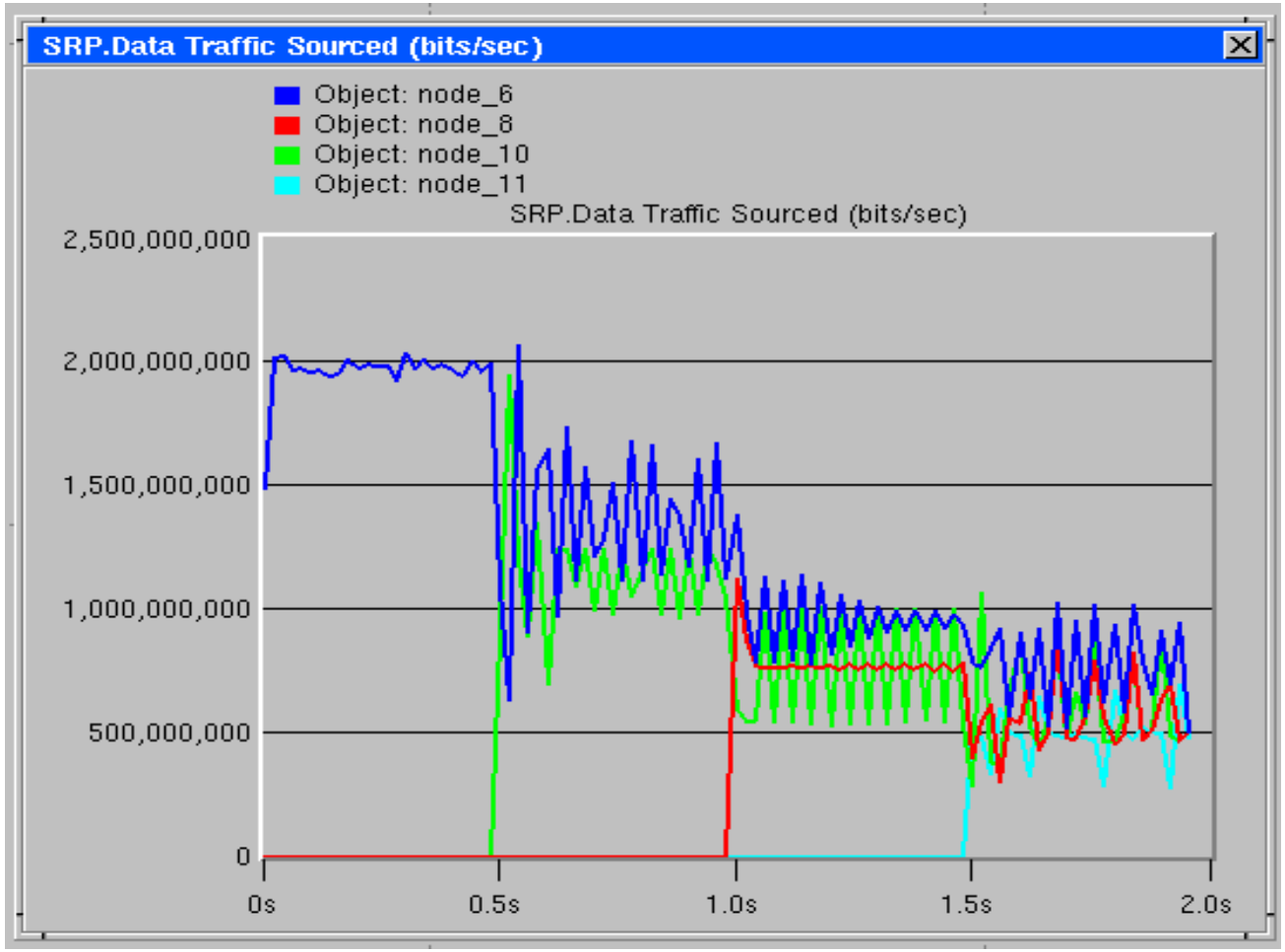


- TB Congestion only occurs at node 4 (aggregation point), and as a result, the allow usage at node 3 (upstream node) is often throttled.

Results: WAN, Triggered Input, OC-48



Results: WAN, Triggered Input, OC-48



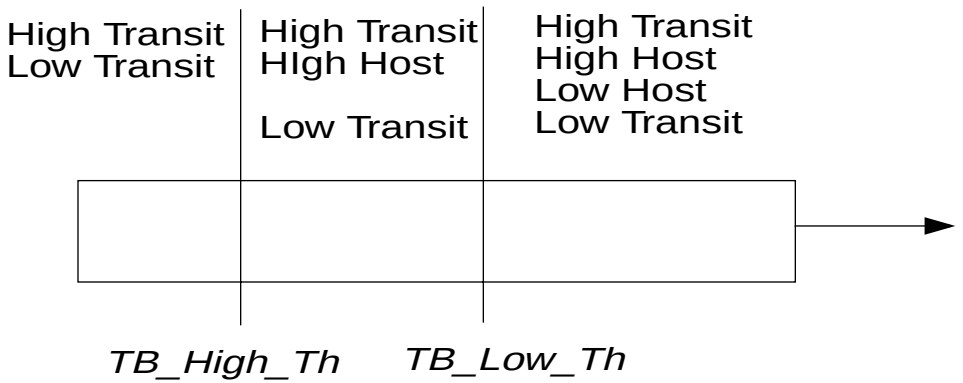
Summary

- SRP-fa achieves fair share of ring capacity under various traffic pattern and load scenarios.
- SRP-fa's fair share is based on the maximum throughput achievable with spatial reuse property.
- SRP-fa is effective in both small and large geographical coverage.

Appendix I: SRP-fa Nutshell (One)

- Layer 2 MAC Transmission Scheduling

❖ Transmission order:



❖ 4 Conditions when Low priority Host traffic is blocked:

- TB occupancy >= TB_Low_Threshold
- My_usage > Allow_usage
- My _usage > Max_allow
- TB low priority is non-empty && My_usage > Fwd_rate

Appendix I: SRP-fa Nutshell (Two)

- Generating Usage Packets
  (once every decay interval)

```
Congested = (TB_low_depth > TB_low_TH /2)

if (Congested) {
  if ( (Lp_my_usage < Rcvd_usage) ||
                        (Rcvd_usage == Null) )
1         Usage_upstream = Lp_my_usage;
  else
          Usage_upstream = Rcvd_usage;
} else if ( (Rcvd_usage != Null) &&
                      (Lp_fwd_rate > allow_usage) )
2     Usage_upstream = Received_usage;
else
3     Usage_upstream = Null;

if (Usage_upstream > Max_Lrate)
     Usage_upstream = Null;
```

(1) when congested, send usage control (lesser of the two) to throttle upstream nodes

(2) Not congested here, but it is somewhere downstream.

Throttle upstream when the forward rate is more than my allowed usage

(3) Not congested, and my allowed usage is greater than the forward rate, send NULL

Appendix I: SRP-fa Nutshell (Three)

- Updating Allow Usage

  (once every decay interval)

```
Max_Lrate = AGECOEFF * Decay_Interval

if ( Rcvd_usage != NULL)
     Allow_usage = Rcvd_usage;
else
     Allow_usage + (Max_Lrate - Allow_usage)/ LP_ALLOW
```

Reference:

D. Tsiang and G. Suwala, “draft-tsiang-srp-01.txt,” IETF Internet Draft, March 1, 2000

Appendix II: E2E Delay of High Priority Traffic

- **E2E delay** is measured from the time a high priority packet is queued at the source to the time that it is received at the destination. There are three delay componets[1]:

  ❖ Access Delay (A_d)

  WC (Worst Case) = N x B_f * HP_MTU * byte_delay + Q_d

  BC (Best Case) = HP_MTU * byte_delay + Q_d

  ❖ Nodal Delay (N_d)

  WC = N * LP_MTU * byte_delay
  + N * TB_High * byte_delay
  + N * HP_MTU * byte_delay

  BC = N x HP_MTU x byte_delay

  ❖ Fiber Delay (F_d)

  Propagation delay through N fiber link segments: 8 μS/mile.

1. N is the number of nodes traversed, while B_f is the token bucket burst size in # of packets, and Q_d is the access queueing delay. HP_MTU and LP_MTU are high and low priority Max Transmit Unit in bytes.

Appendix II: E2E Delay of High Priority Traffic *(cont'd)*

Numerical ExampleS (SRP delay at OC-12):

N	= 8 nodes traversed	(16 nodes actual)
B_f	= 2	(in HP_MTU)
Q_d	= 50 μS	(estimated)

(1) HP_MTU = 128, TB_High = 2700, LP_MTU=1500 (bytes)

WC= 77.3 + 460.8 = 538.1 μS
BC = 51.7 + 13.6 = 65.3 μS

(2) HP_MTU = 128, TB_High = 2700, LP_MTU=9000 (bytes)

WC= 77.3 + 1261.3= 1338.6 μS
BC = 51.7 + 13.6 = 65.3 μS

(2a) Same as (2) but at OC-48

WC= 56.8 + 315.4 = 372.2 μS
BC = 50.4 + 3.4 = 53.8 μS

- Simulation results provide similar range of values.