Master Mini Project

Performance Evaluation of Minimum Delivery Delay Multipath TCP Scheduler Implemented in Linux Kernel

Faridul Alam

Date: 21 June, 2016

Supervisors

Amanpreet Singh, M.Sc

Dr. Andreas Könsgen

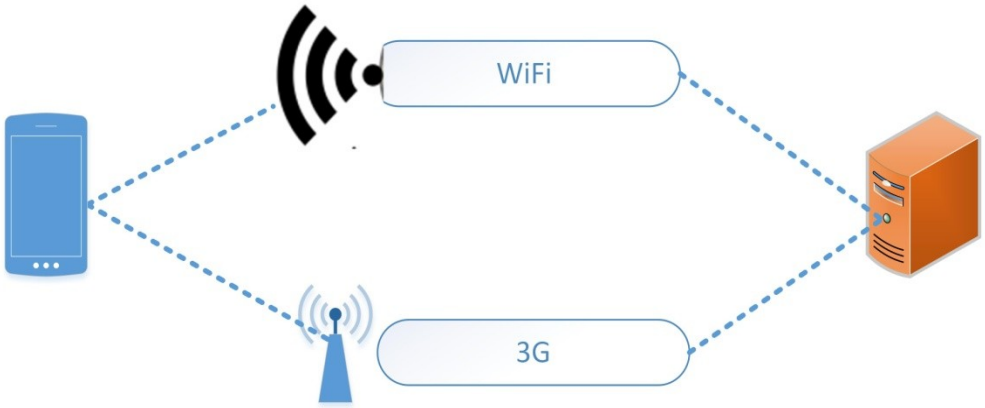
Prof. Dr. Anna Förster

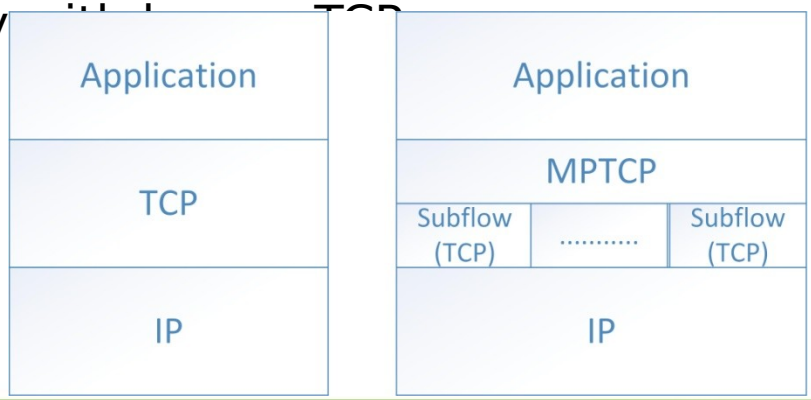
Outline

- Multipath TCP (MPTCP)
- MPTCP Scheduler
  - Existing MPTCP Schedulers on Linux
    - Round Robin Scheduler
    - Shortest Round Trip Time Scheduler
- Implementation of Delivery Delay Scheduler
- Performance Evaluation
- Conclusion & Outlook

Multipath TCP (MPTCP)

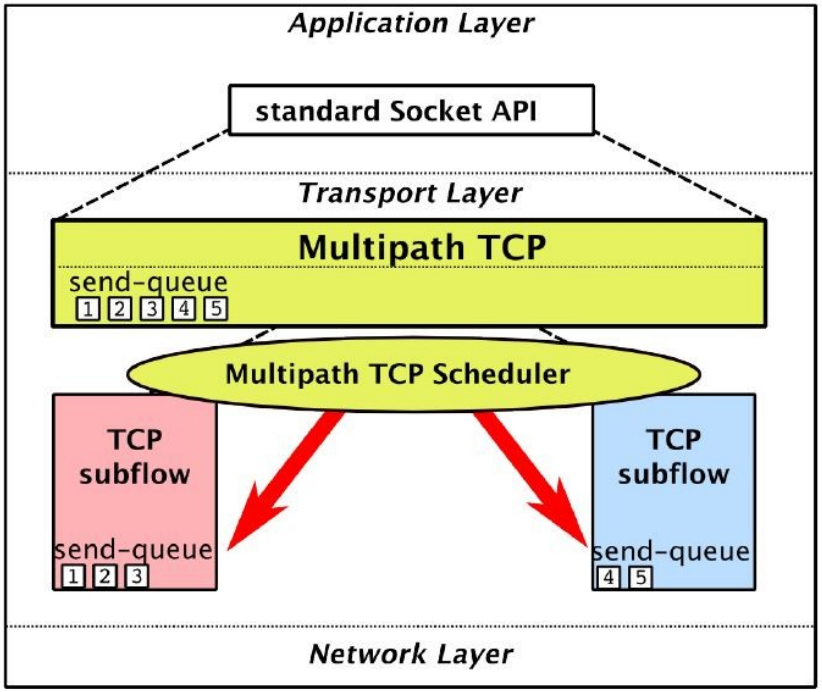
- With the emergence of multihoming capable devices, data transfer through multiple paths for a session become a viable option
- IETF has developed the tra to facilitate multipath data
- MPTCP offers– [1]
  - Increased throughput
  - Robustness against link fai
  - Balance congestion
  - Compatibility





MPTCP Scheduler

- Heterogeneous nature of delay and bandwidth asymmetry of the path can create out-of-order data arrival at the receiver [2]
  - Head-of-Line Blocking
  - Receiver-Window Limitations
- Results higher reordering delay and might give lower throughput



[2]

Round Robin (RR) Scheduler

- Schedule data packets in round robin or cyclic fashion
- Maximum number of consecutive scheduled segments for an instance is configurable
  - Fill up the congestion window of a subflow or
  - Strictly Round Robin
- Disadvantages –
  - May not be able to utilize the whole path capacity when paths have significant delay or capacity differences as the scheduler always tries to distribute equally among all available subflows [5]

Shortest Round Trip Time (sRTT) Scheduler

- Schedule data packets on a subflow with the lowest RTT amongst all available subflows
- Considers delay (RTT) of the paths, which also reflects network condition to some extent
- Disadvantages –
  - High delay assymetry in backward path might affect the scheduling which might influence out-of-order delivery

Delivery Delay (DD) based Scheduler

- Schedule data packet on a subflow with the lowest Delivery Delay (DD) amongst all available subflows
- DD scheduler was first proposed in [8] with the title `` Performance Comparison of Scheduling Algorithms for Multipath Transfer''
  - Was implemented in a Simulator and shown that the End-to-End (ENE) delay is reduced

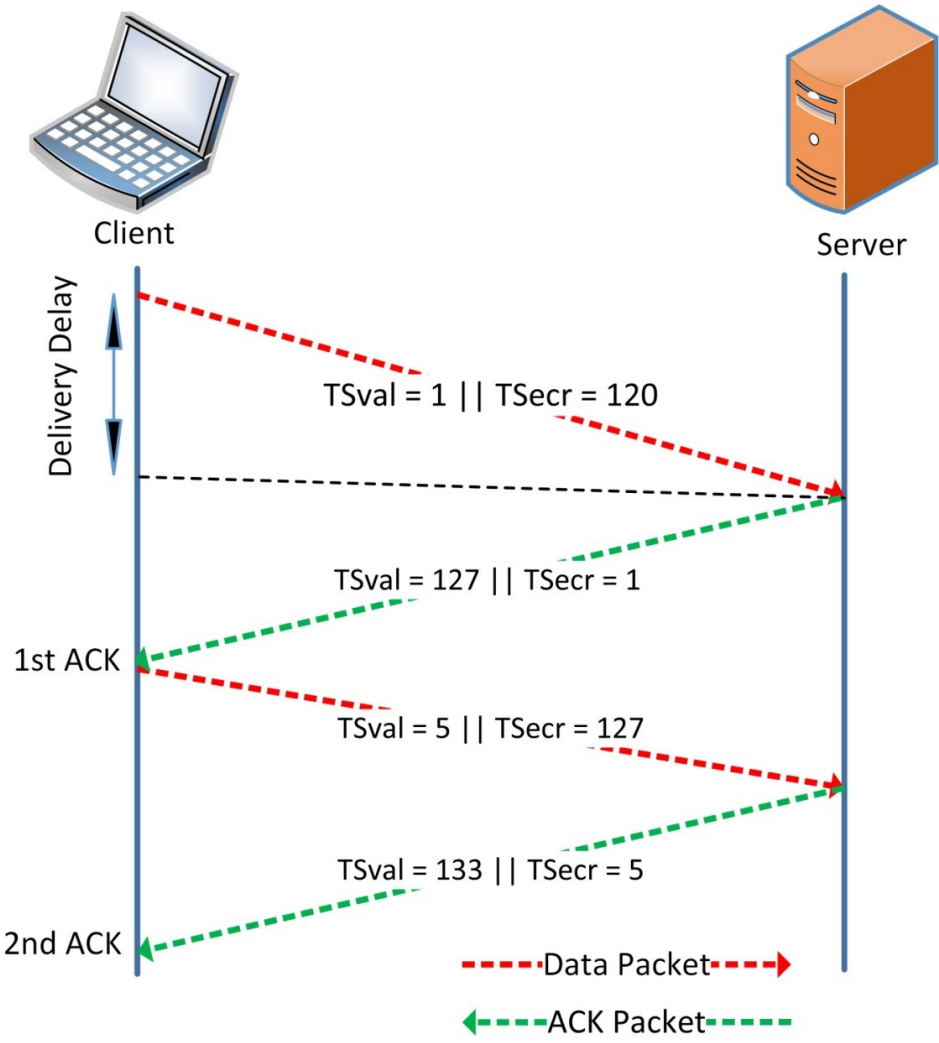
- Focus point –
  - Understand the implemented schedulers (sRTT and RR) in Linux Kernel along with their behaviour
  - Implement the Delivery Delay (DD) based scheduler in Linux Kernel
  - Performance analysis of DD scheduler along with the other two implemented schedulers for different network scenarios

Estimation of Delivery Delay

TCP Timestamps [9]			
Kind = 8	Length	TSval	TSecr
1 Byte	1 Byte	4 Bytes	4 Bytes

- TSval is the current value of the timestamp's clock
- TSecr is the TSval value contained in the received data/ack packet

- Delivery Delay – **1st ACK** = TSval – TSecr = 127 – 1 = 126
- Delivery Delay – **2st ACK** = TSval – TSecr = 133 – 5 = 128



Scheduling Algorithm – For each packet

```
min_delivery_delay = MAX_INT

for each subflow i do
    if (current packet is a retransmission) and (has been transmitted on
  the subflow i) then          continue
    end if
    if (subflow i is definitely unavailable) or (subflow i is temporary
  unavailable) then            continue
    end if
    if (delivery_delay _i < min_delivery_delay) then
        min_delivery_delay = delivery_delay_i
        selected_subflow = i
  end if
end for
```

Performance Evaluation

Test Topology - 2 Node

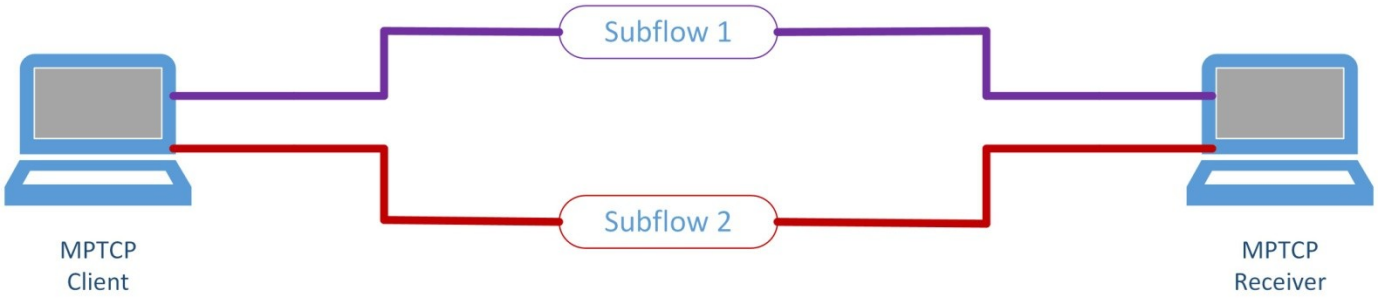


Figure : Two Node Testbed

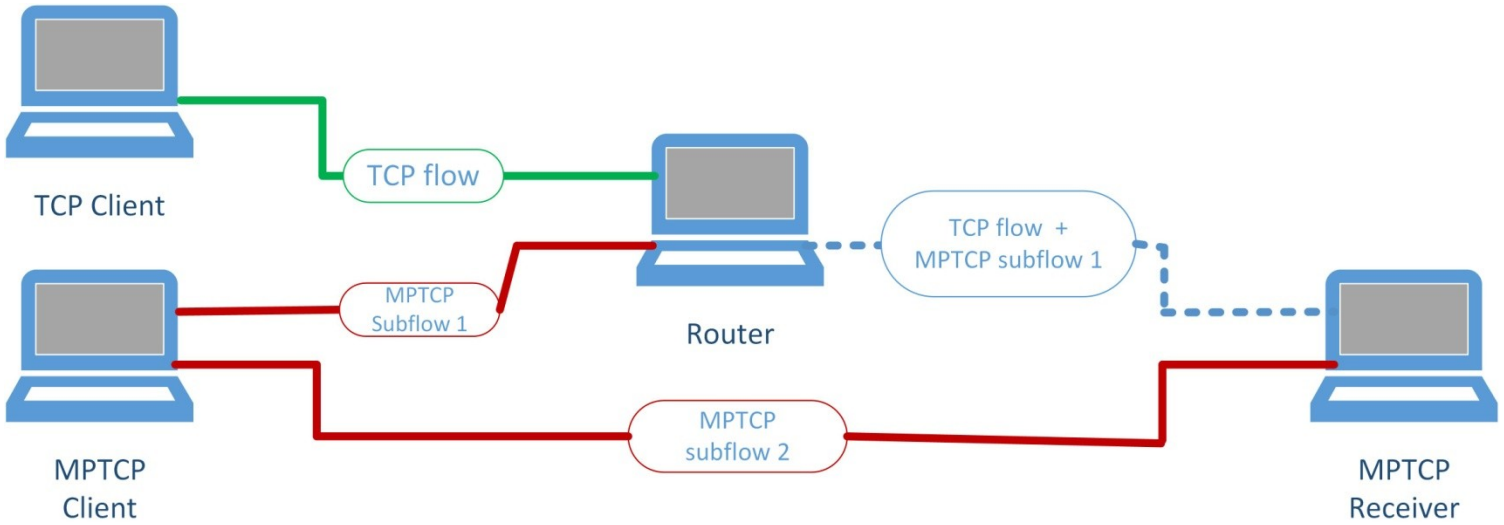


Figure : Extended Testbed

Test Topology – Two Node Testbed with Router

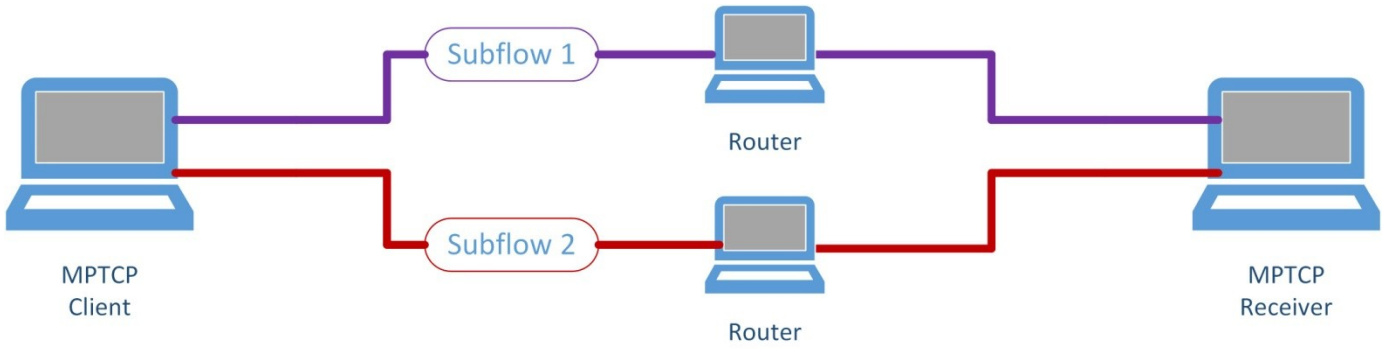
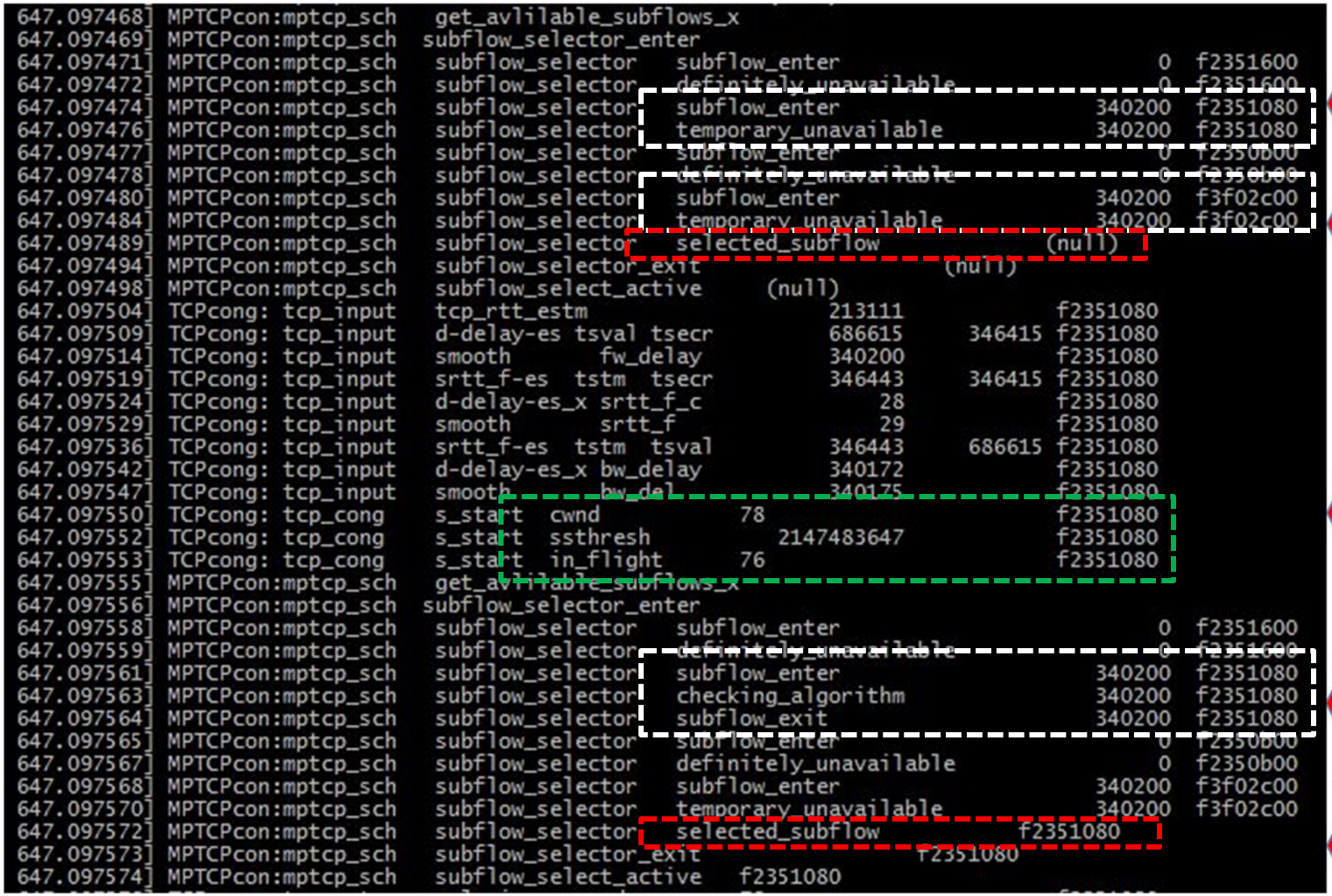


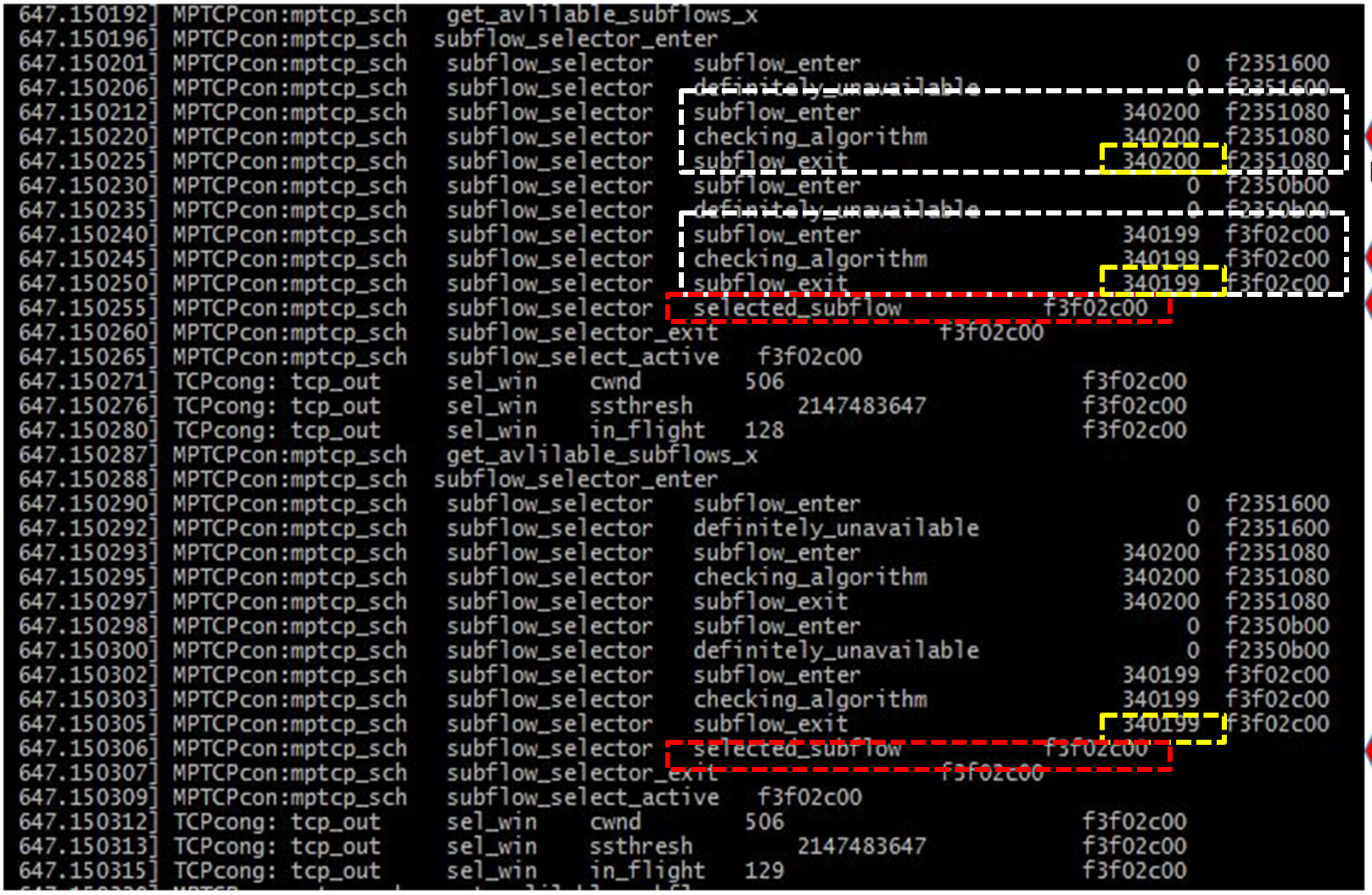
Figure : Two Node Testbed with Router

- Configuration :
  - Congestion Control = Reno
  - tc qdisc "add" dev "ethX" root netem delay “Yms“ – at router

Validation of Delivery Delay Scheduler - I



Validation of Delivery Delay Scheduler - II



Both subflows are available

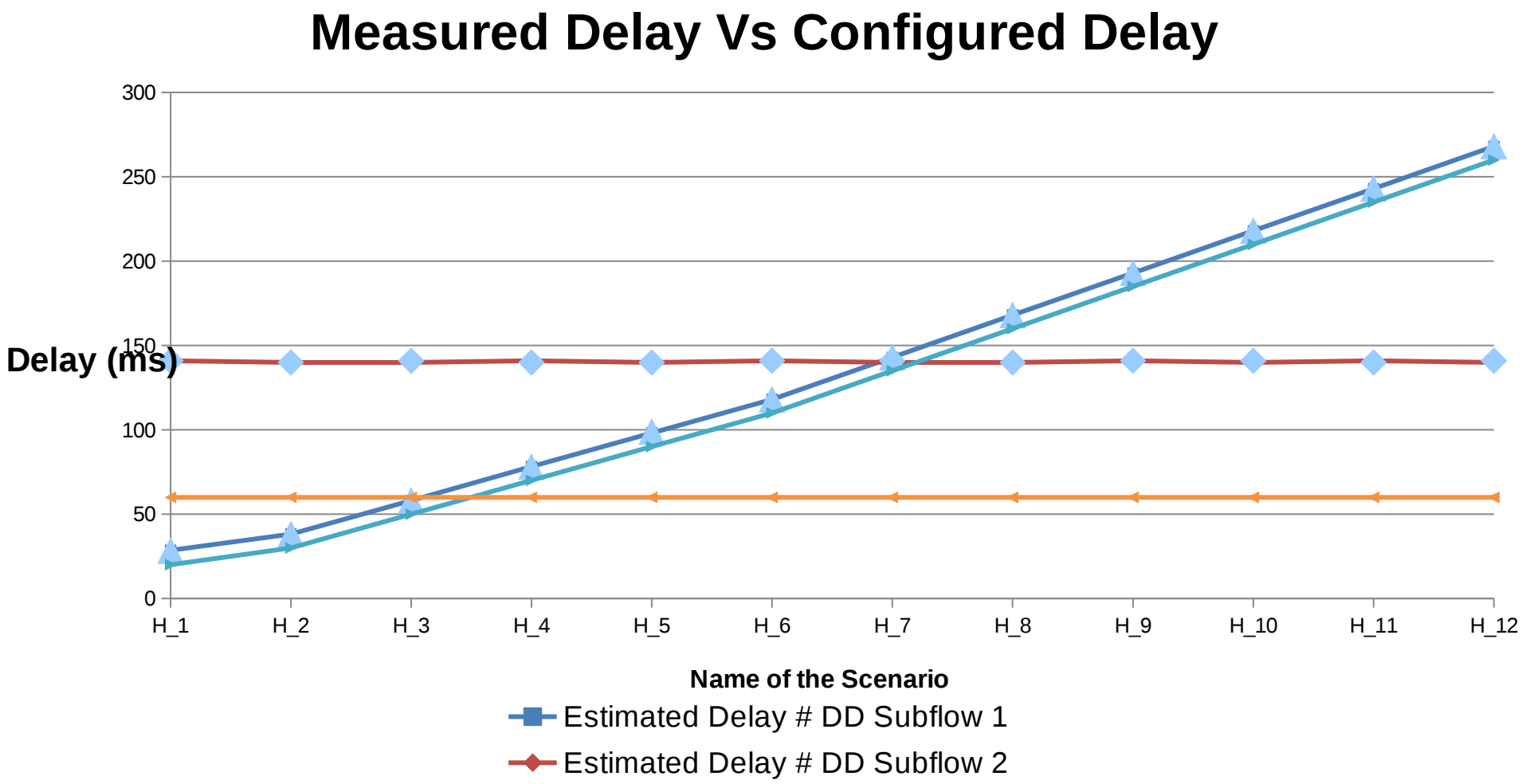
Subflow 1's delivery delay is lower than that of subflow 2, therefore subflow 1 is selected

For the next packet, subflow 1 is also selected

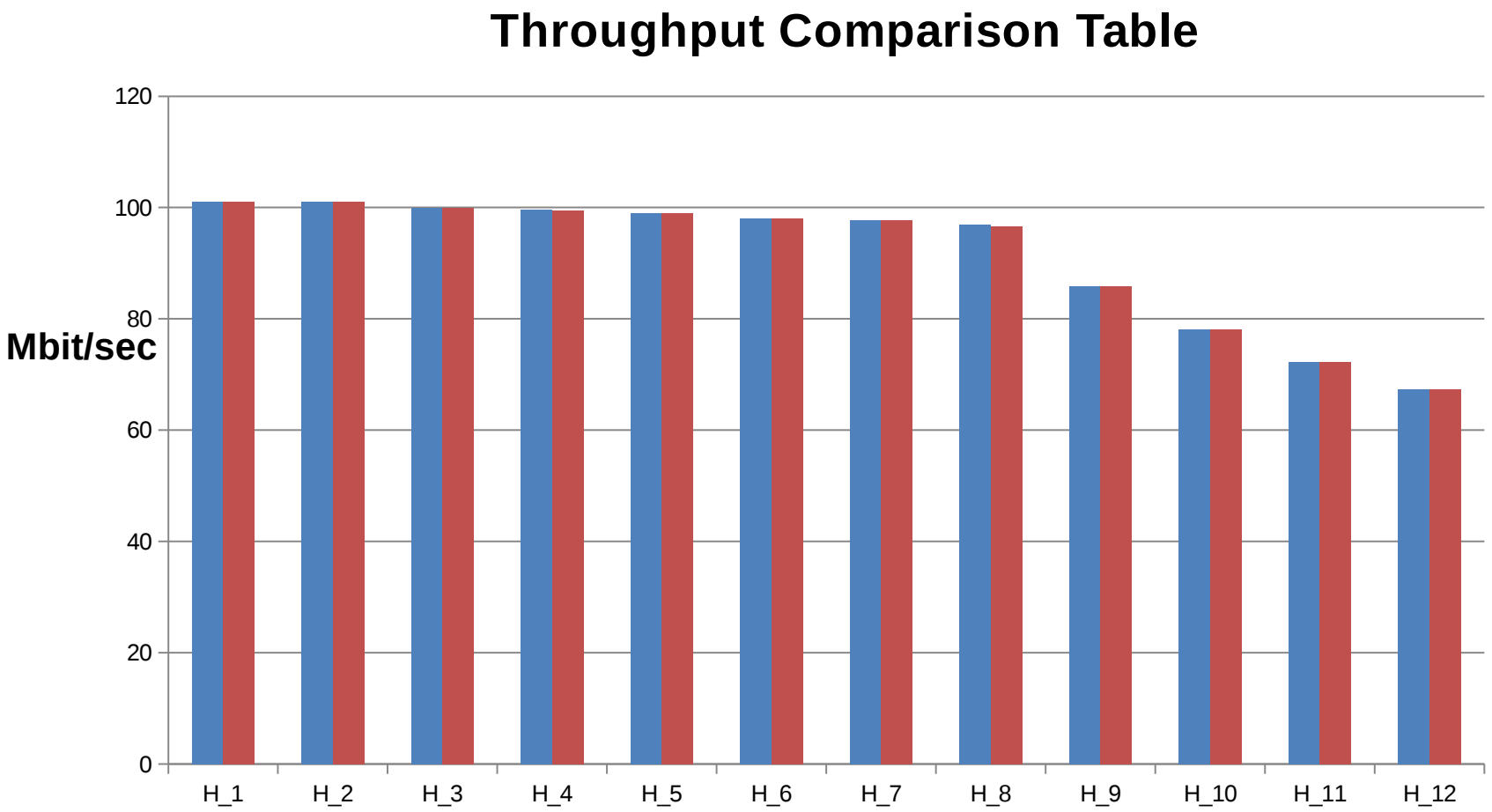
Performance Evaluation - Two Node Testbed with Router

Name of the Scenario	Subflow 1		Subflow 2	
	RTT (ms)	FPD+BPD (ms)	RTT (ms)	FPD+BPD (ms)
H_1	20	10 + 10	60	50 + 10
H_2	30	10 + 20	60	50 + 10
H_3	50	10 + 40	60	50 + 10
H_4	70	10 + 60	60	50 + 10
H_5	90	10 + 80	60	50 + 10
H_6	110	10 + 100	60	50 + 10
H_7	135	10 + 125	60	50 + 10
H_8	160	10 + 150	60	50 + 10
H_9	185	10 + 175	60	50 + 10
H_10	210	10 + 200	60	50 + 10
H_11	235	10 + 225	60	50 + 10
H_12	260	10 + 250	60	50 + 10

Link Capacity – 100/10 (Mbit/sec)



Link Capacity 100/10 (Mbit/sec)



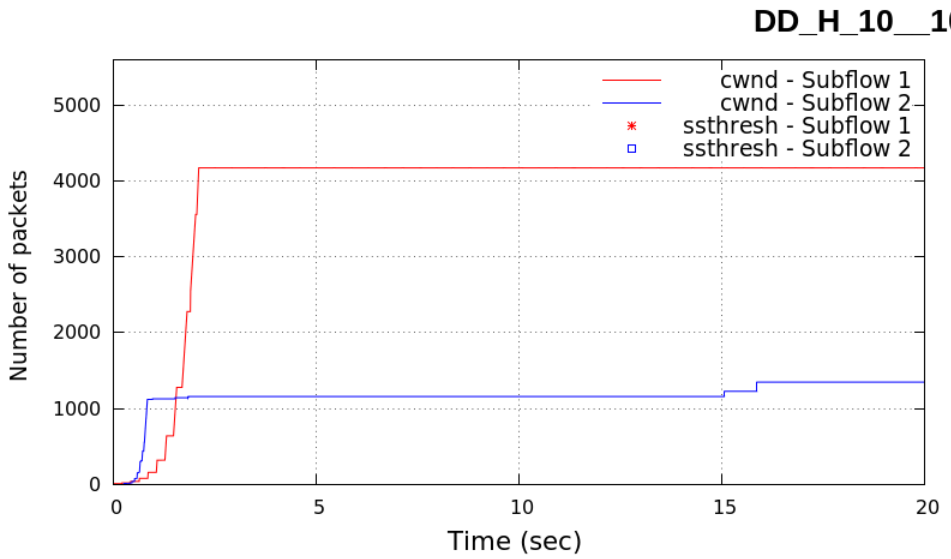
Graphical Analysis

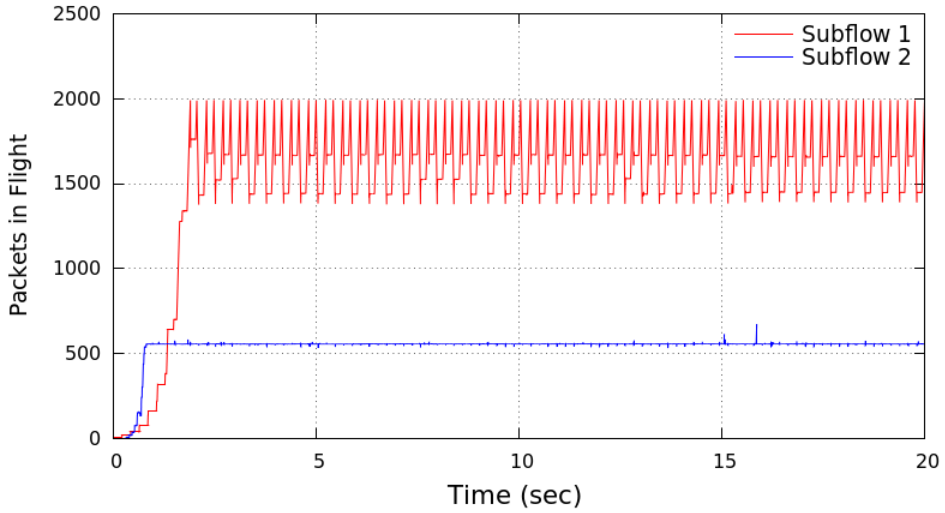
Link Capacity 100/100 (Mbit/sec)

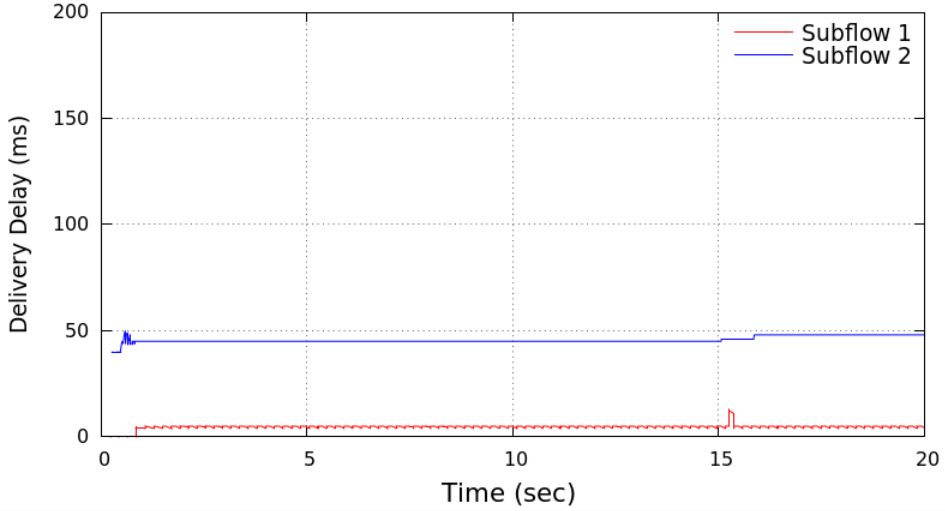
Scenario H_10

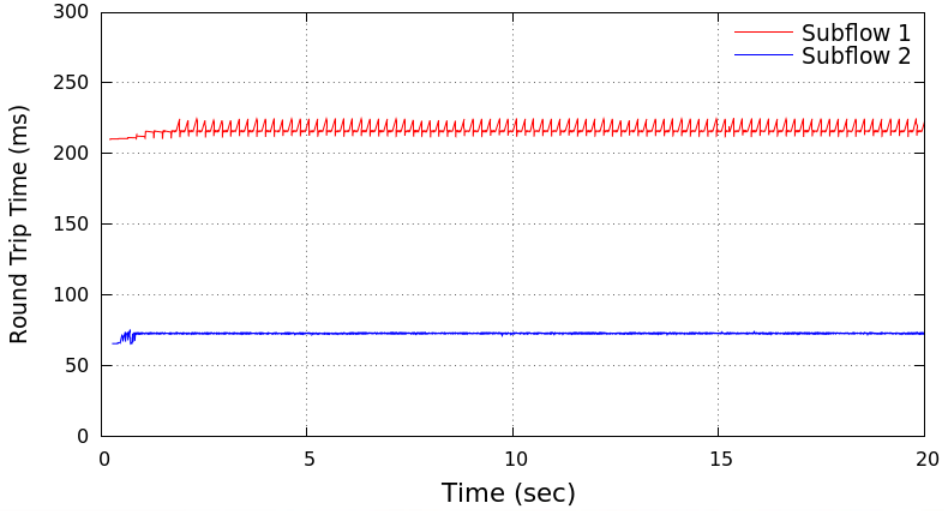
Subflow 1 RTT – 210 ms (10 ms + 200 ms)
Subflow 2 RTT – 60 ms (50 ms + 10 ms)

Delivery Delay Scheduler

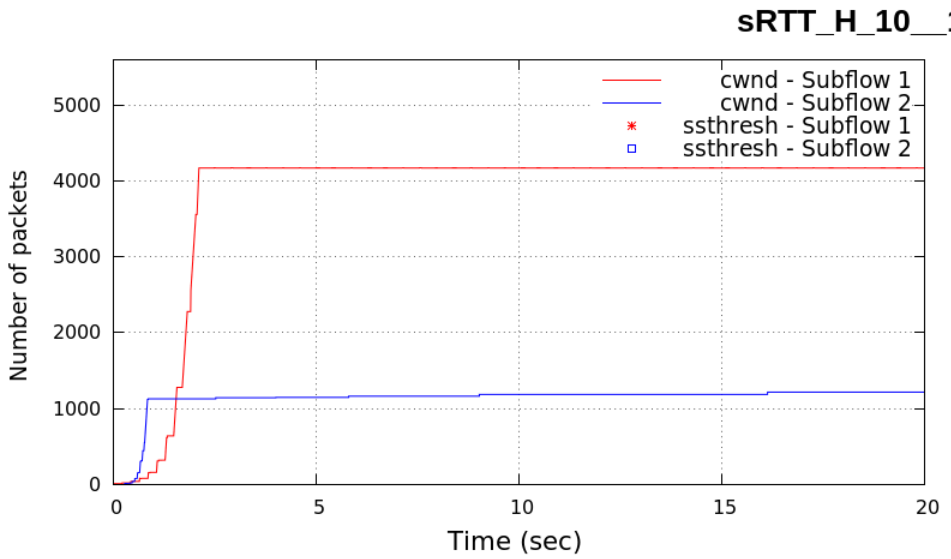


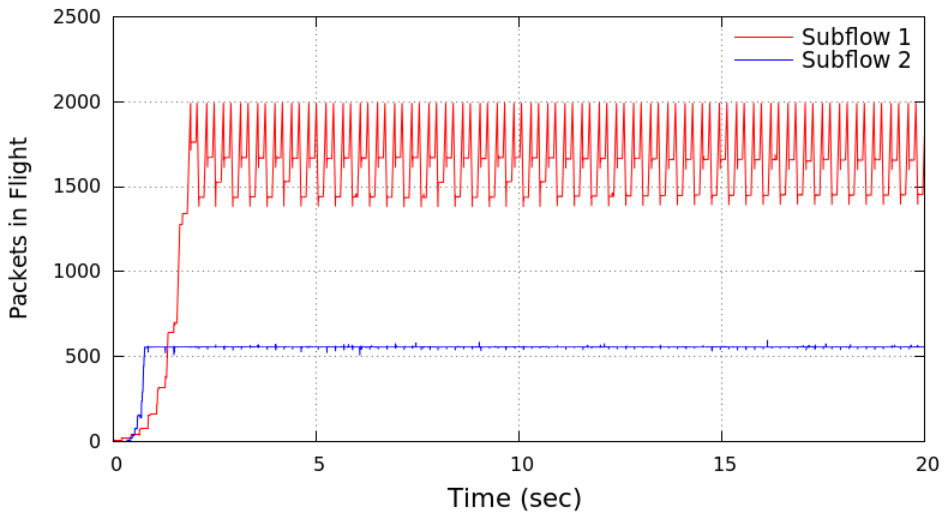


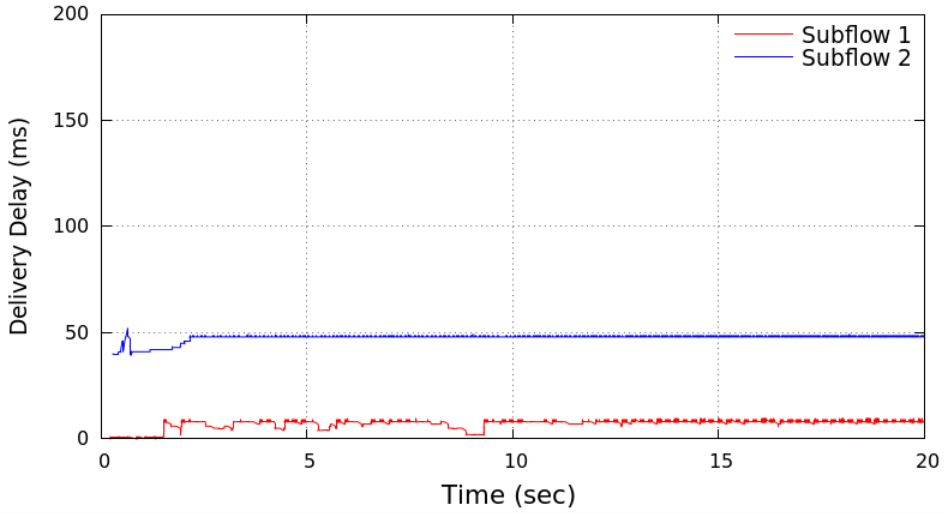


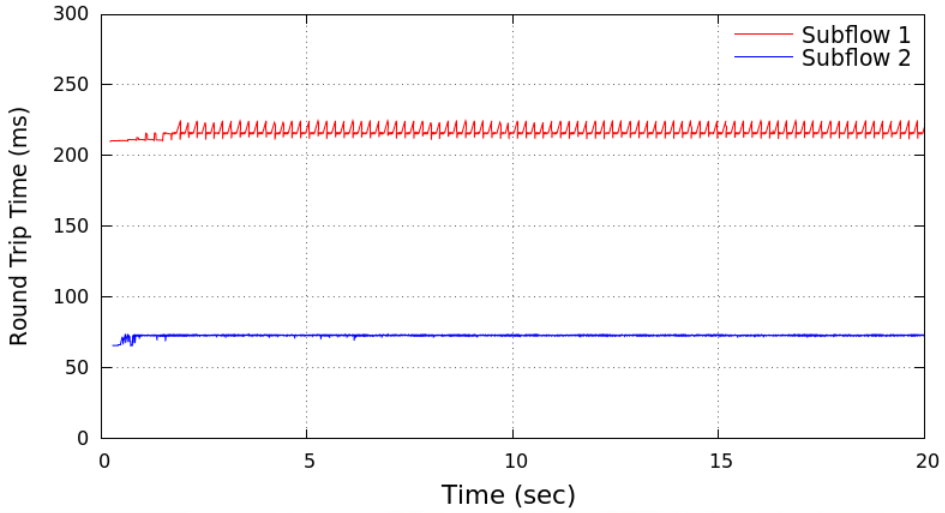


Shortest Round Trip Time Scheduler



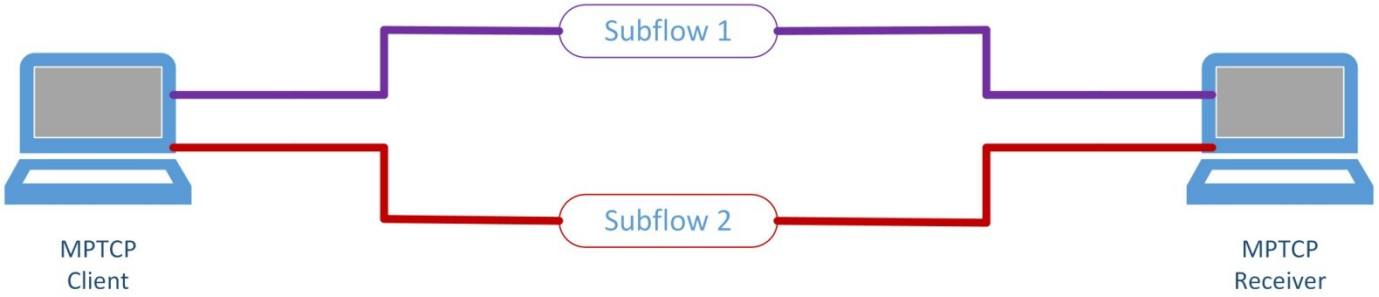






Faridul Alam
(faridul_alam@ymail.com)

Graphical Analysis- Two Node Testbed



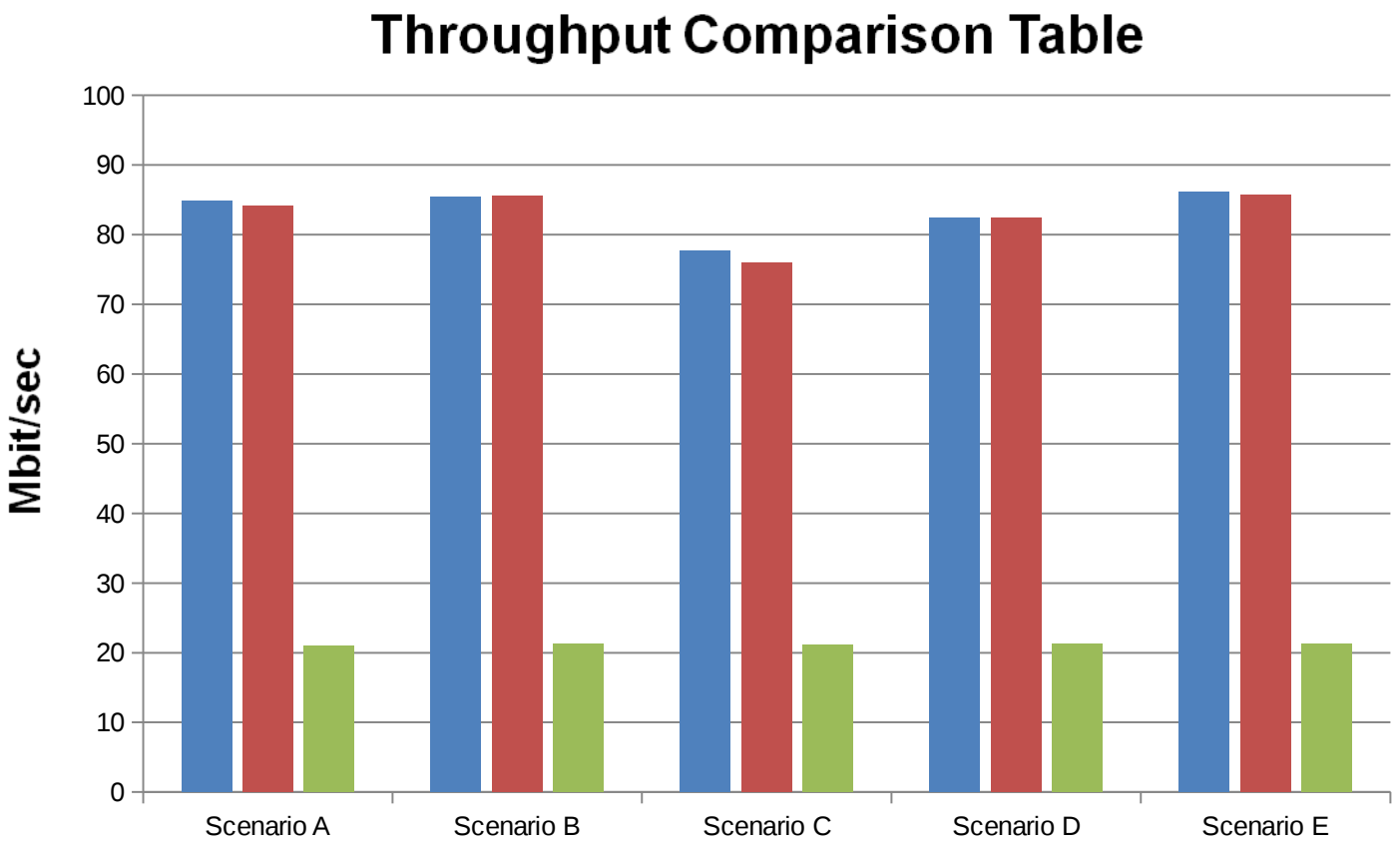
Configuration

- Congestion Control = Reno
- tc qdisc "add" dev "ethX" root netem delay “Yms“ – at Client and Server

Scenarios

Capacity (Mbit/sec)	Name of Scenario	Subflow 1			Subflow 2		
		FPD (ms)	BPD (ms)	RTT (ms)	FPD (ms)	BPD (ms)	RTT (ms)
100/100 & 100/10 & 10/10	A	10	20	30	20	10	30
	B	10	10	20	20	10	30
	C	10	40	50	20	20	40
	D	10	20	30	20	40	60
	E	10	10	20	20	20	40
	F	10	200	210	20	400	420

Link Capacity 100/10 (Mbit/sec)



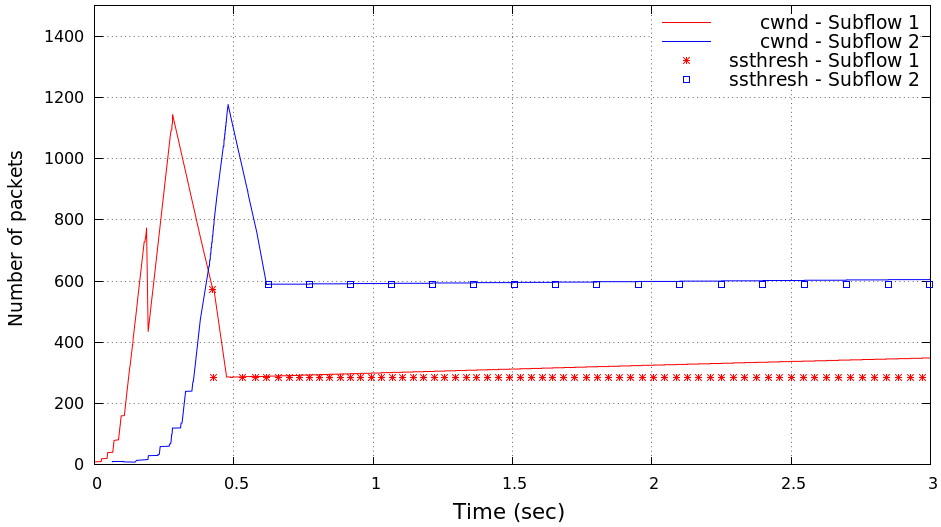
Graphical Analysis

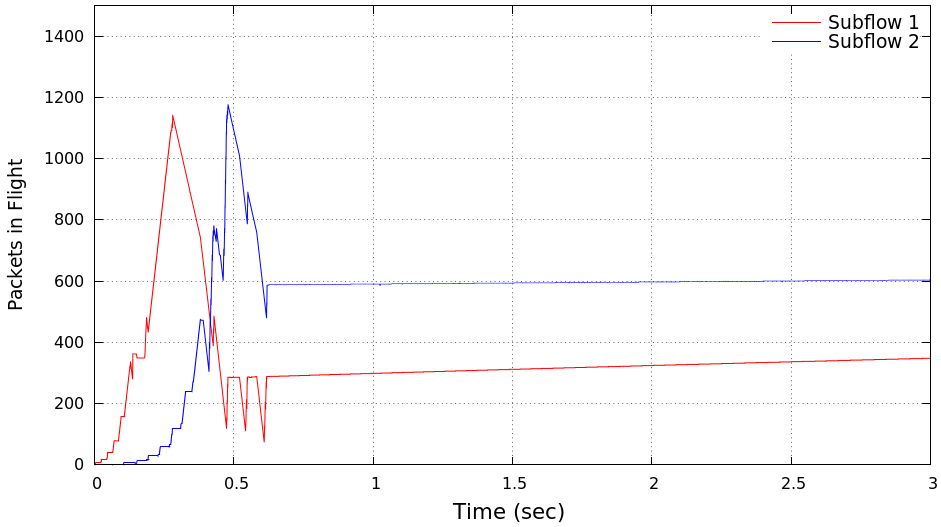
Link Capacity 100/100 (Mbit/sec)

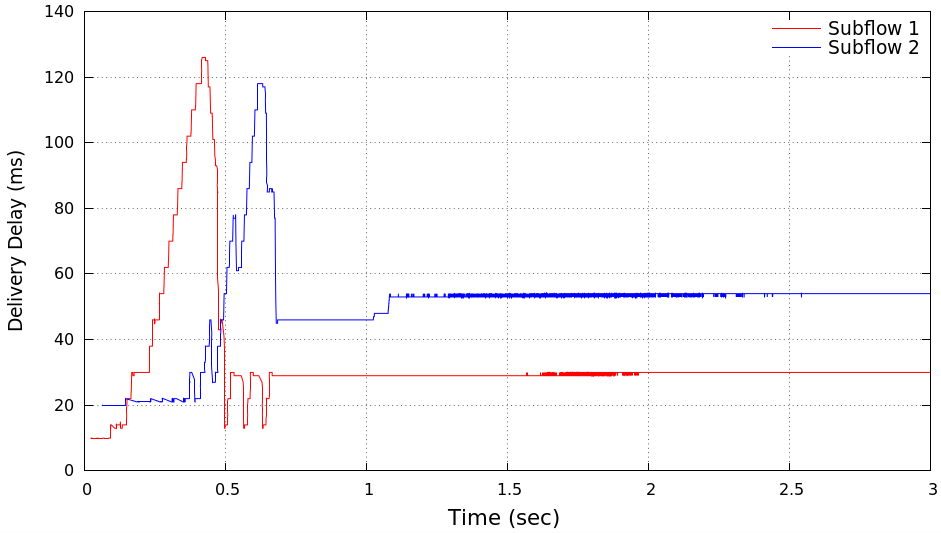
Scenario E

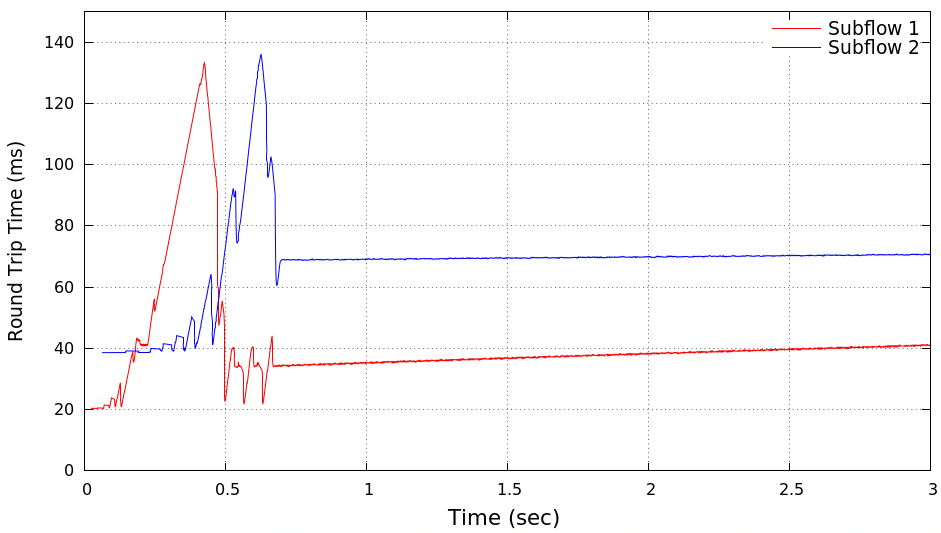
Subflow 1 RTT – 20 ms (10 ms + 20 ms)
Subflow 2 RTT – 40 ms (20 ms + 20 ms)

Delivery Delay Based Scheduler









Receiver Buffer Occupancy

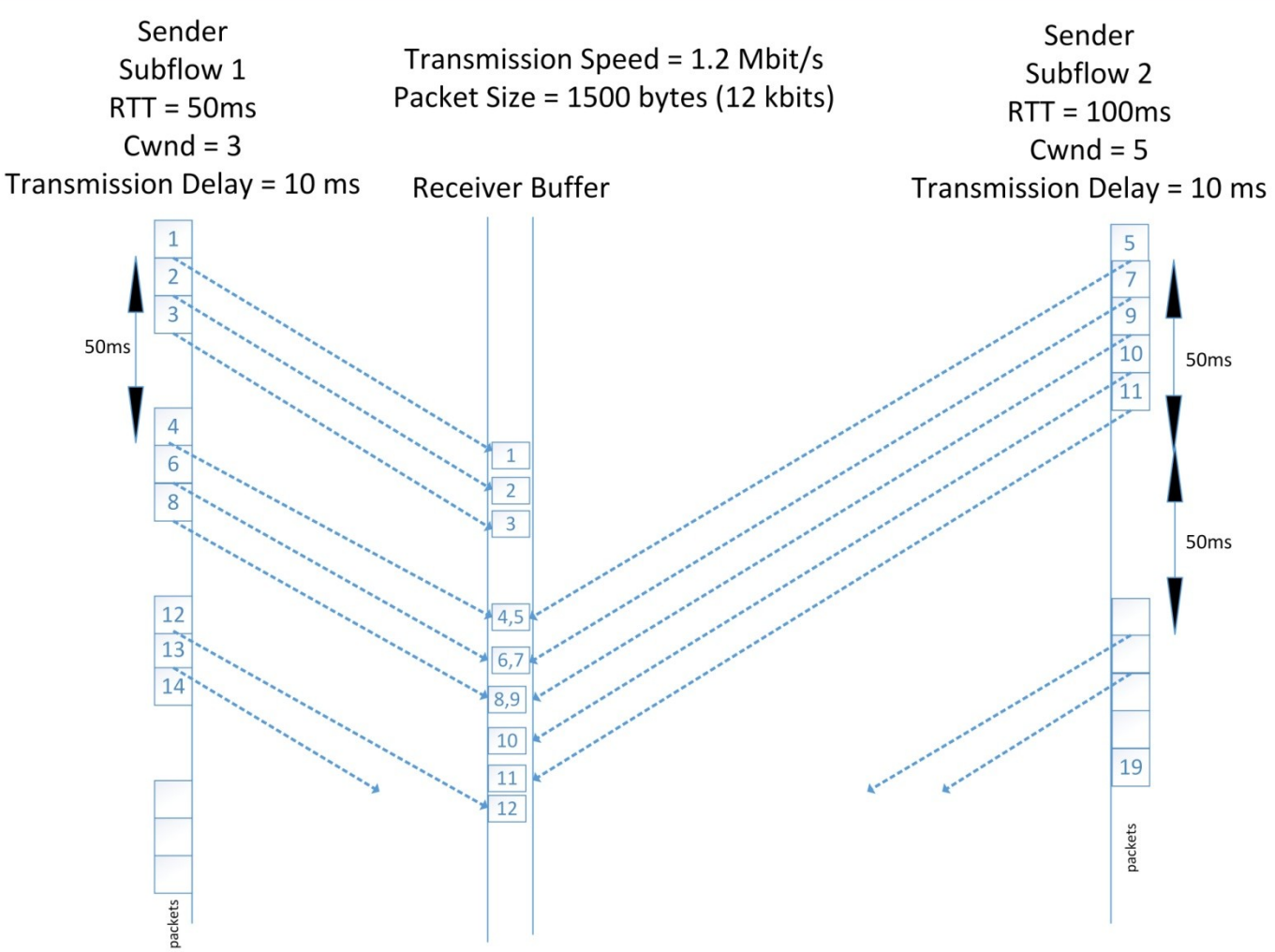
Scenario			
Link Capacity = 1.2 Mbit /sec	Transmission Delay	RTT (ms)	FPD+BPD (ms)
Subflow 1	10 ms	110	10 + 100
Subflow 2	10 ms	60	50 + 10

Receiver Buffer (sRTT Scheduler)															
4	5	6	1	2	3	7	10	8	9	12	14	11	13	15	19
20	16	21	17	18	22	23	24	28	29	30	25	26	27	31	32
33	36	38	39	34	35	37	40	41	44	42	46	48	43	45	47

Receiver Buffer (DD Scheduler)															
1	2	3	4	5	6	7	10	8	11	9	13	12	14	15	19
16	20	17	21	18	22	23	24	28	29	25	30	26	27	31	32
33	36	38	39	34	35	37	40	41	42	44	46	48	43	45	47

Outlook (Intelligent DD Scheduler)

Packet no	Reordering Delay (ms)	End-to Delay (
1	0	50
2	0	50
3	0	50
4	0	50
5	0	100
6	0	50
7	0	100
8	0	50
9	0	100
10	0	100
11	0	100
12	0	50



Conclusion

- The performance of Round Robin scheduler is poor compared to Delivery Delay and sRTT scheduler in case of high link capacity and delay asymmetry

- Throughput performance of Delivery Delay scheduler and sRTT scheduler are similar (ack-clocked)
  - Scheduling is considered only among the subflows which have an open congestion window
  - Congestion window's increase is dependent on the ACK packets and hence also controls the scheduler's performance
  - When the backward delay (delivery delay for the ACK packet) on the subflows remains constant throughout the simulation run, the sRTT scheduler almost acts as the delivery delay scheduler

- Artificial delay configuration at the sender nodes created performance issues in the experimental setup
  - Delay should be configured on the routers

Outlook

- The performance comparison of different schedulers can be evaluated in different network conditions
  - In wireless networks
  - With bidirectional data traffic
- Intelegent Delivery Delay scheduler can be implemented to minimize the out-of-order delivery
  - Advance packet scheduling by estimating the forward paths behavior

References

1. Costin Raiciu, Mark Handley, and Damon Wischik. “Coupled congestion control for multipath transport protocols “. RFC6356, October, 2011.
2. Christoph Paasch, Simone Ferlin, Ozgu Alay, and Olivier Bonaventure. “Experimental evaluation of multipath tcp schedulers”. In Proceedings of the 2014 ACM SIGCOMM workshop on Capacity sharing workshop, pages 27–32. ACM, 2014.
3. Tuan-Anh Le and Loc X Bui. ‘’Forward delay-based packet scheduling algorithm for multipath tcp’’. arXiv preprint arXiv:1501.03196, 2015.
4. Multipath TCP, Multipath TCP in the stack. http://cacm.acm.org/magazines/2014/4/173230-multipath-tcp/fulltext , Last visited: 02.03.2016.
5. J Popat Khushi, Jigar A Raval, Samuel Johnson, and Bhavesh Patel. ``An efficient scheduling scheme of multipath tcp for mpi’’.
6. Costin Raiciu, Mark Handley, and Damon Wischik. ``Coupled congestion control for multipath transport protocols’’. RFC6356, October, 2011.
7. Sébastien Barré, Christoph Paasch, and Olivier Bonaventure. ``Multipath tcp: from theory to practice’’. In NETWORKING 2011, pages 444–457. Springer, 2011.
8. Amanpreet Singh, Carmelita Goerg, Andreas Timm-Giel, Michael Scharf, and T-R Banniza. ``Performance comparison of scheduling algorithms for multipath transfer’’. In Global Communications Conference (GLOBECOM), 2012 IEEE, pages 2653–2658. IEEE, 2012.

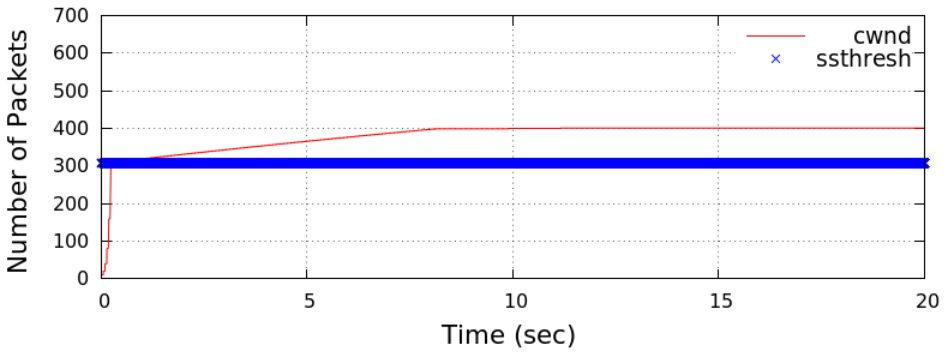
Questions??

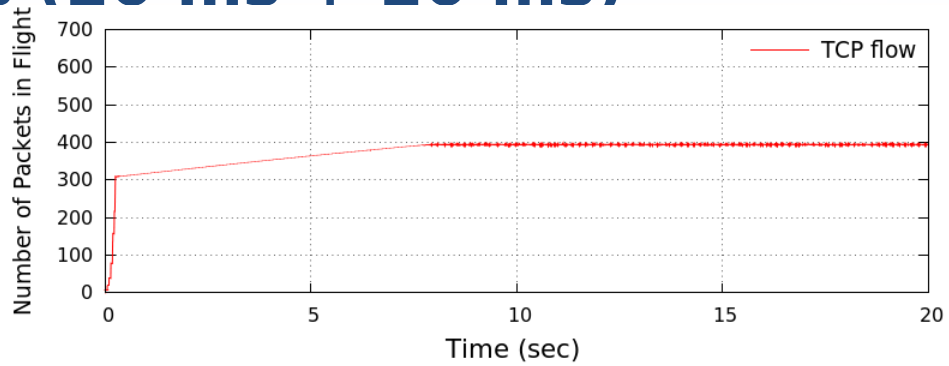
Thank you!!!

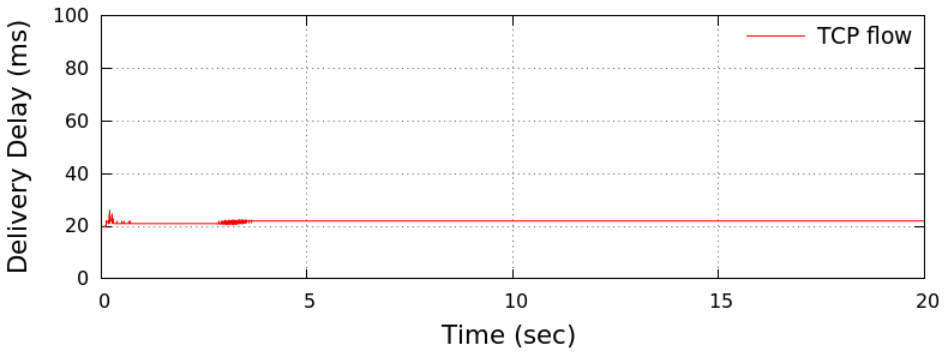
Backup Slides

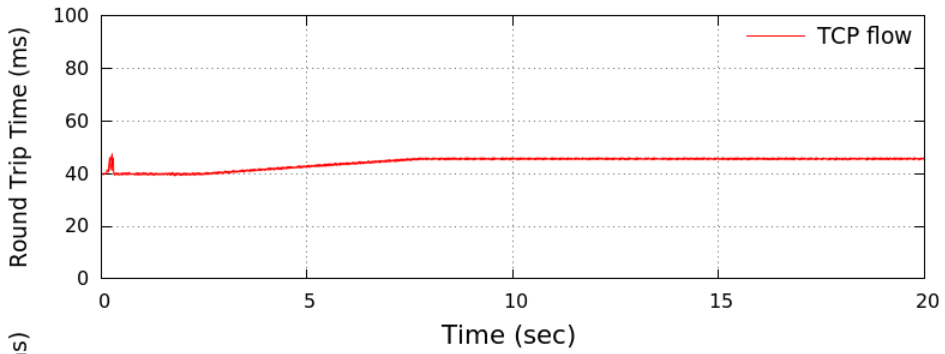
TCP with Router – Link Capacity 100 (Mbit/sec)

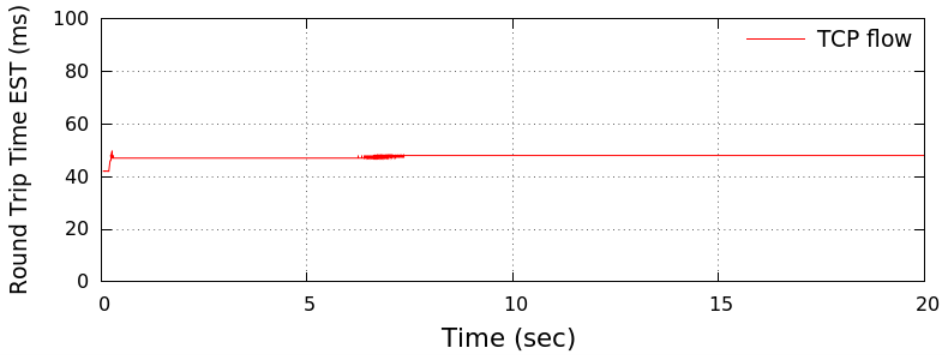
Delay - 40 ms (20 ms + 20 ms)

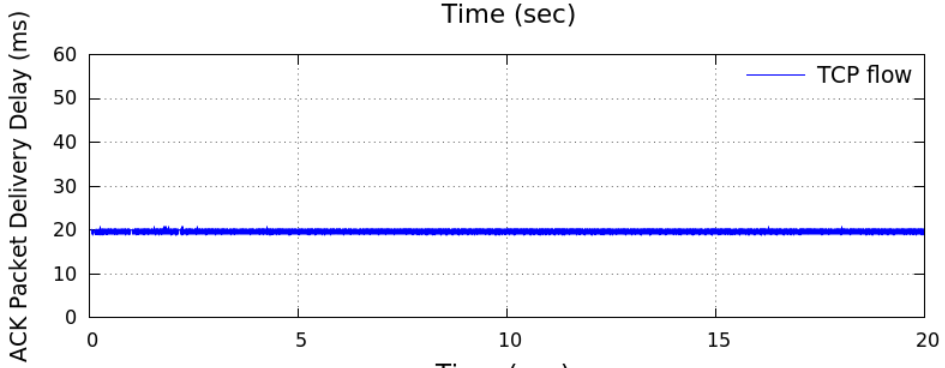






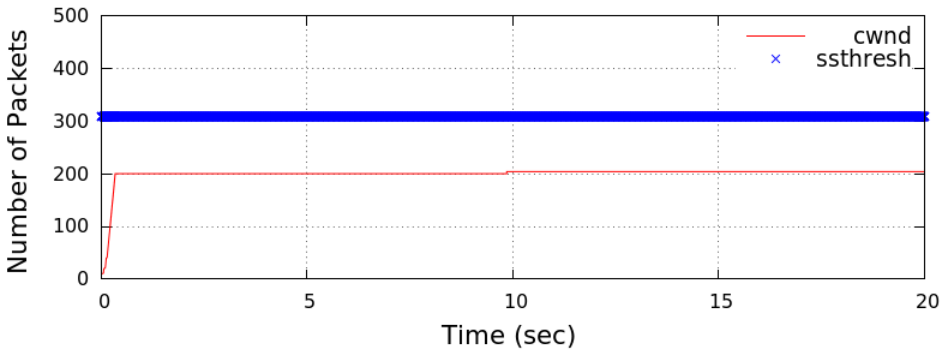


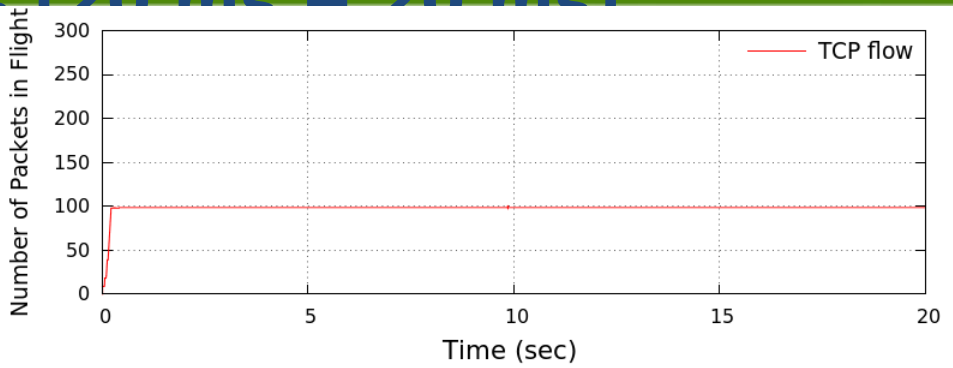


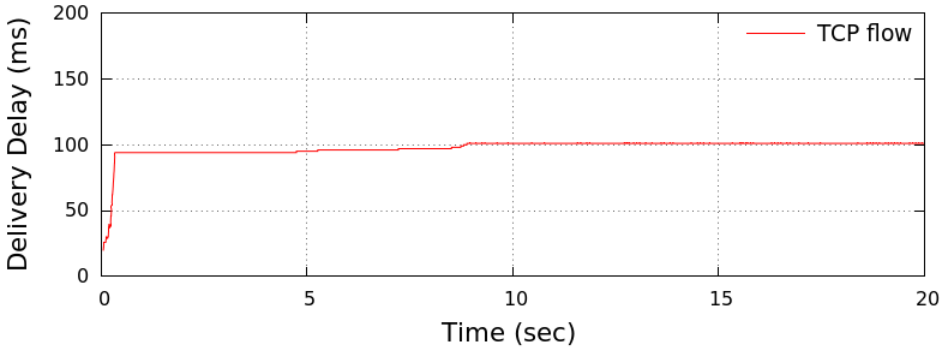


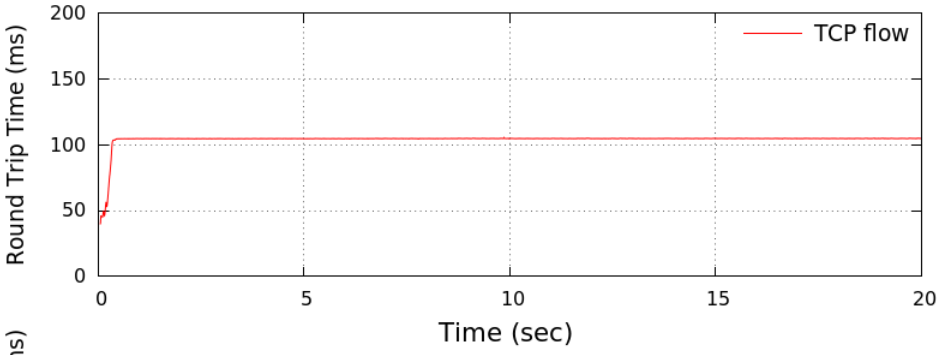
TCP with Router – Link Capacity 10 (Mbit/sec)

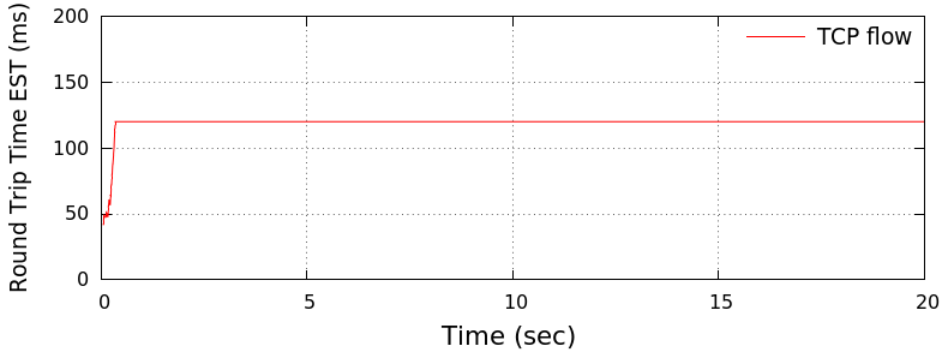
Delay - 40 ms (20 ms + 20 ms)

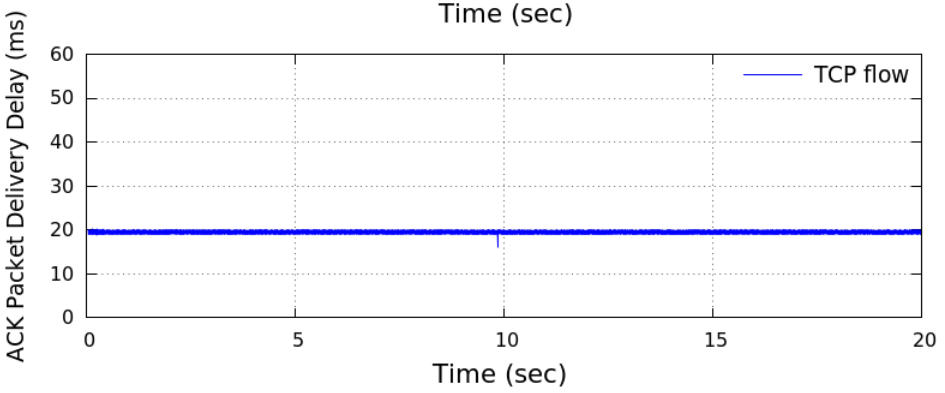






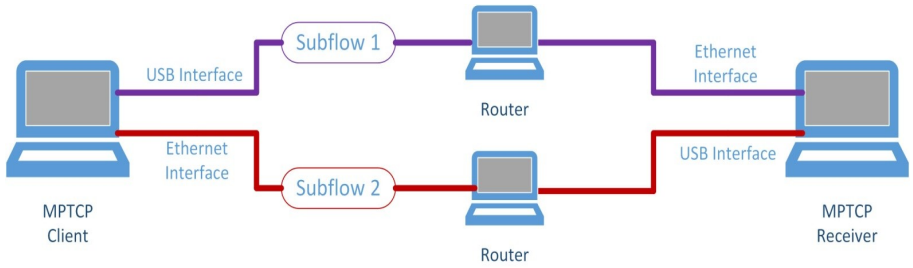






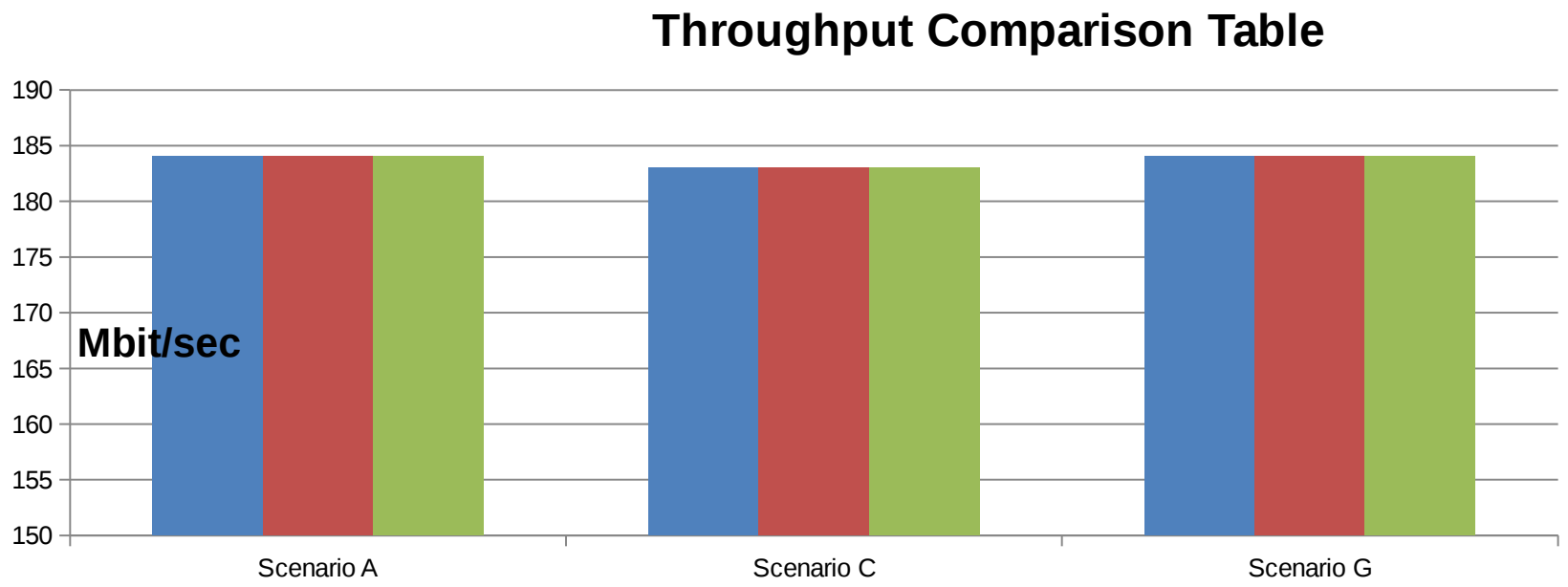
Performance Evaluation - Two Node Testbed with Router

- Configuration :
  - Congestion Control = Reno
  - tc qdisc "add" dev "ethX" root netem delay "Yms" – at router

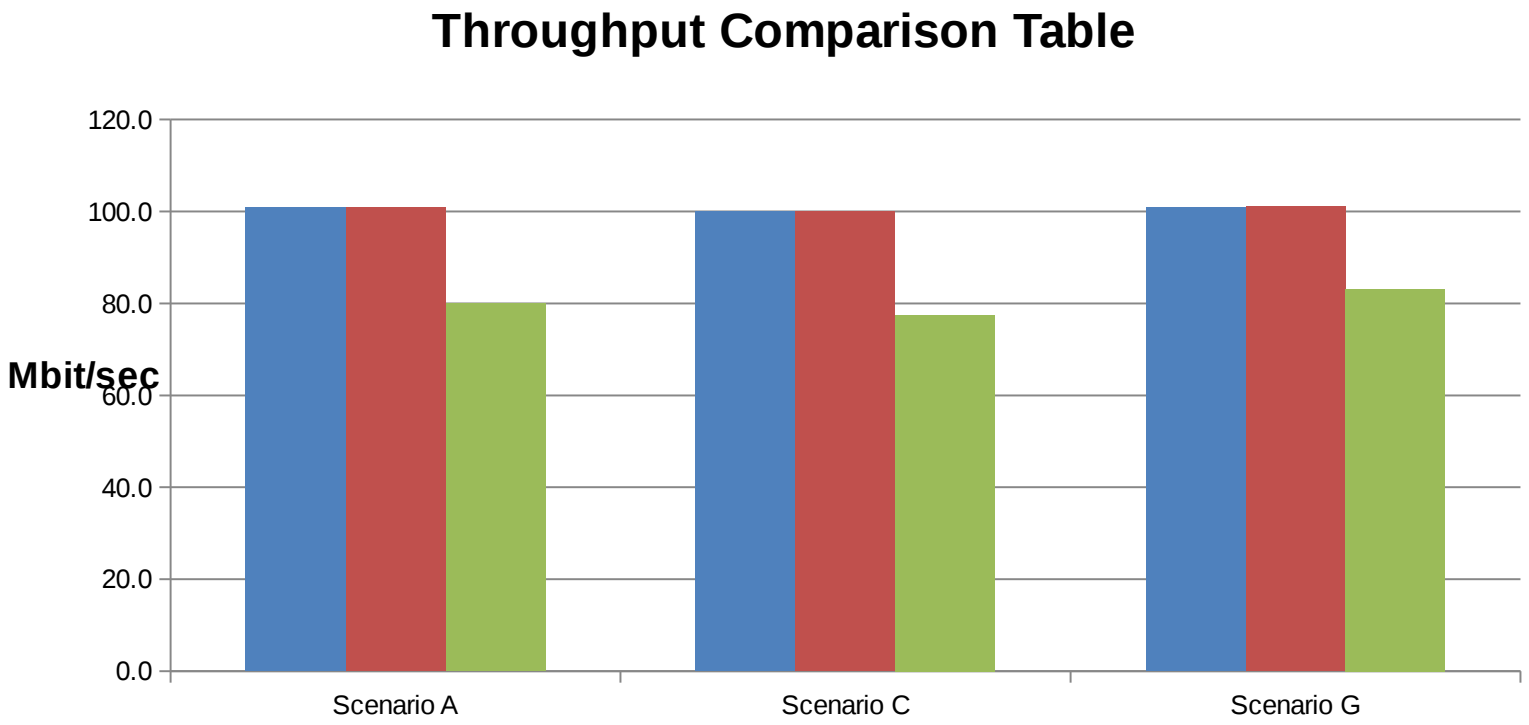


Capacity (Mbit/sec)	Name of Scenario	Subflow 1			Subflow 2		
		FPD (ms)	BPD (ms)	RTT (ms)	FPD (ms)	BPD (ms)	RTT (ms)
100/100 & 100/10 & 10/10	A	10	20	30	20	10	30
	C	10	40	50	20	20	40
	G	10	10	20	10	10	20

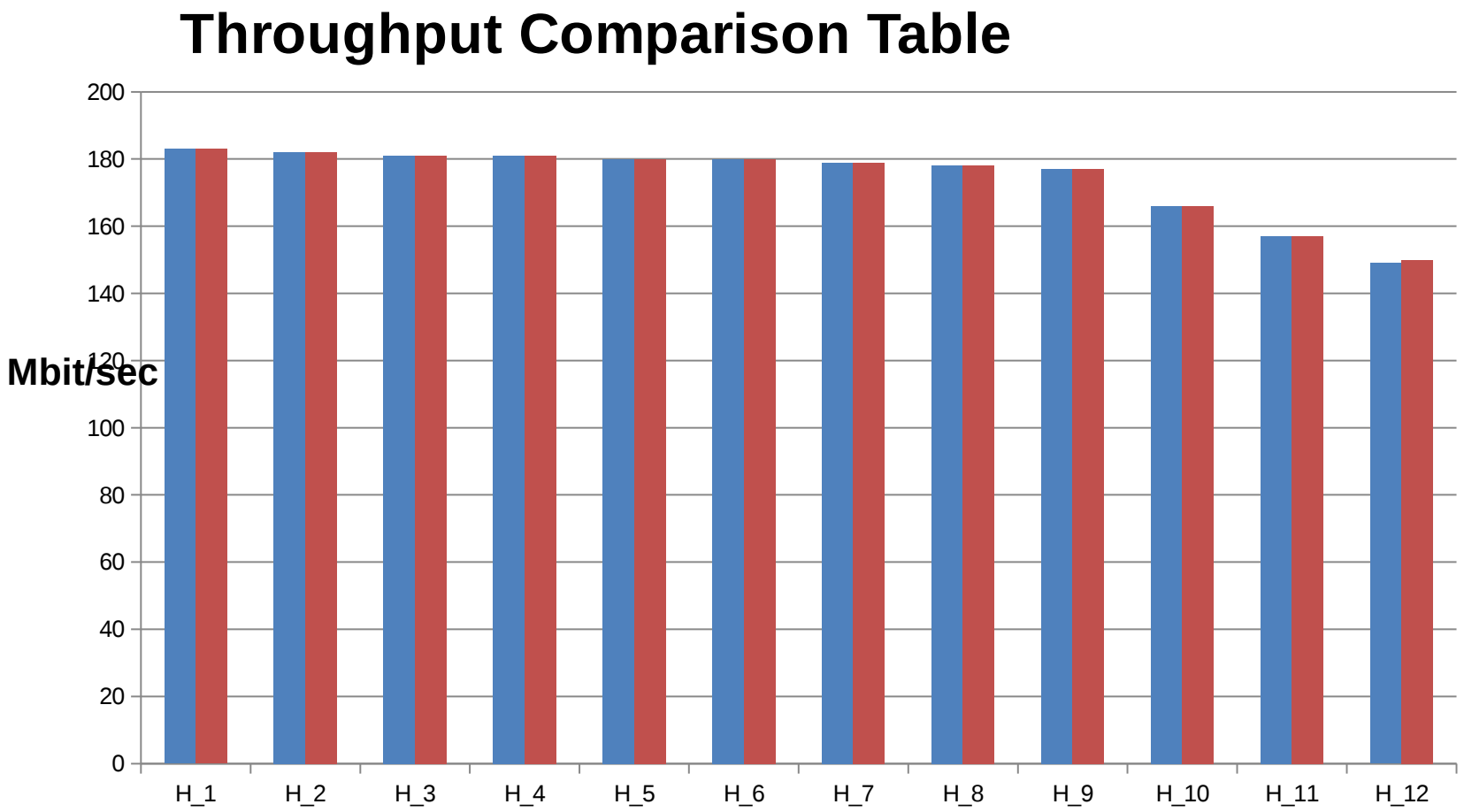
Link Capacity 100/100 (Mbit/sec)



Link Capacity 100/10 (Mbit/sec)



Link Capacity 100/100 (Mbit/sec)



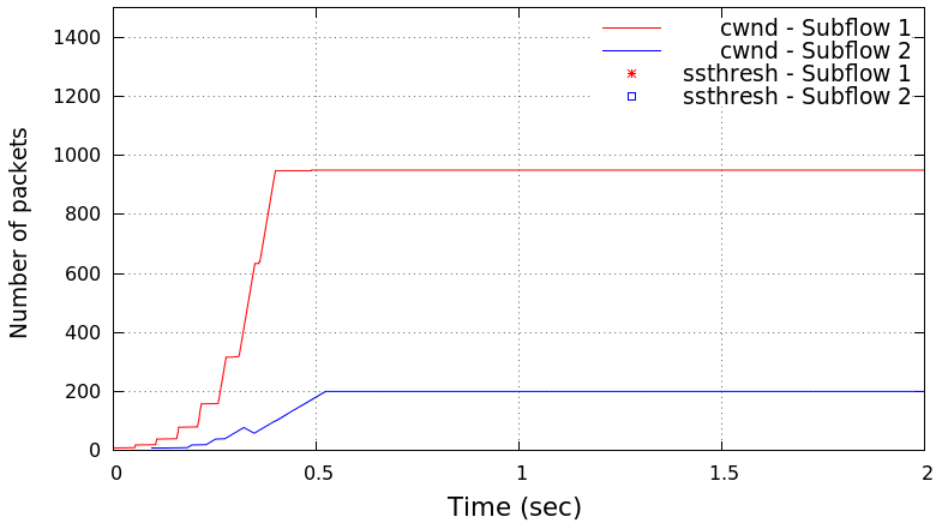
Graphical Analysis

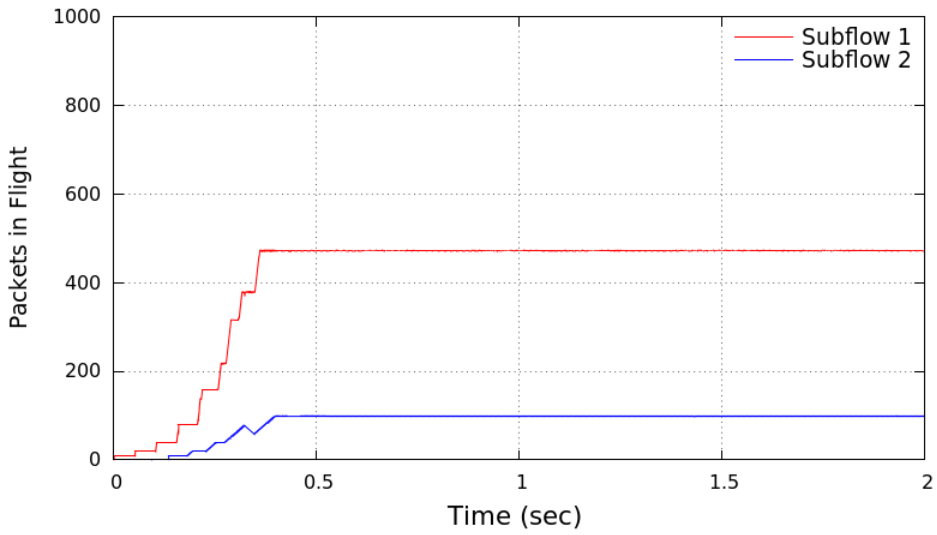
Link Capacity 100/10 (Mbit/sec)

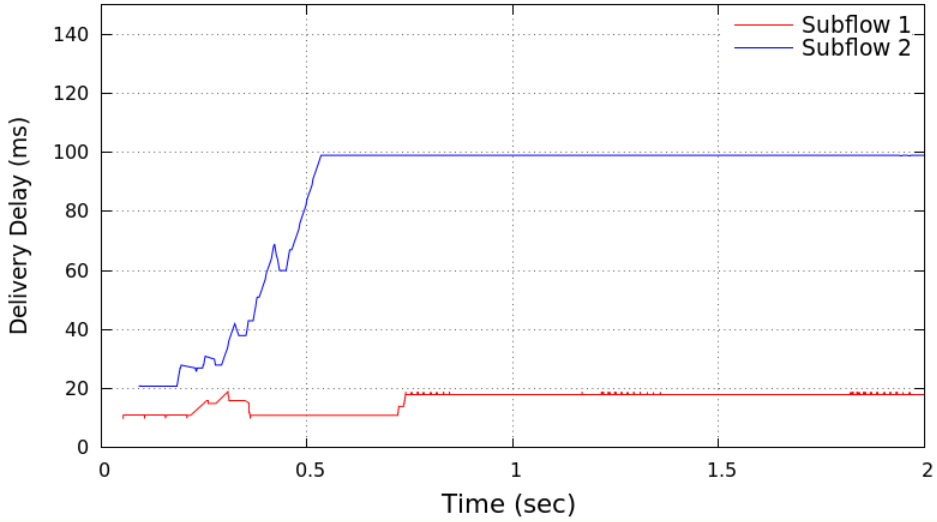
Scenario C

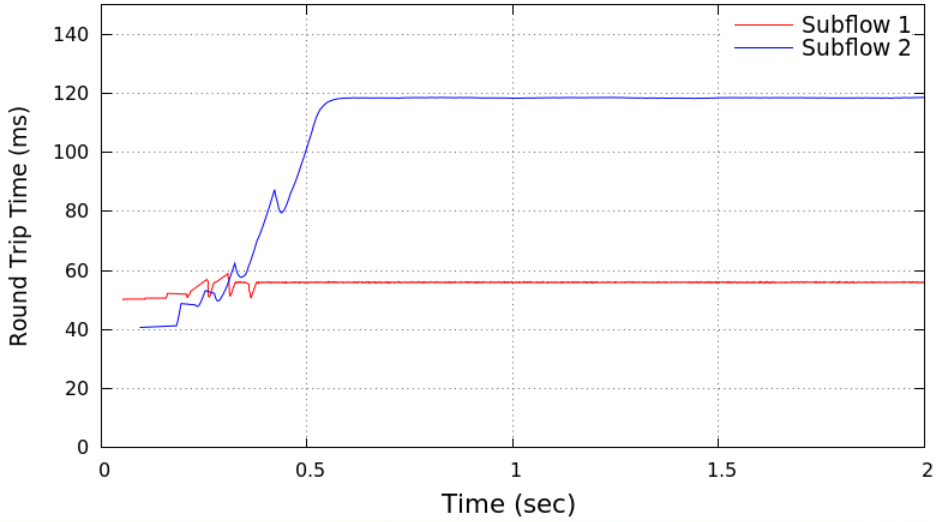
Subflow 1 RTT – 50 ms (10 ms + 40 ms)
Subflow 2 RTT – 40 ms (20 ms + 20 ms)

Delivery Delay Scheduler

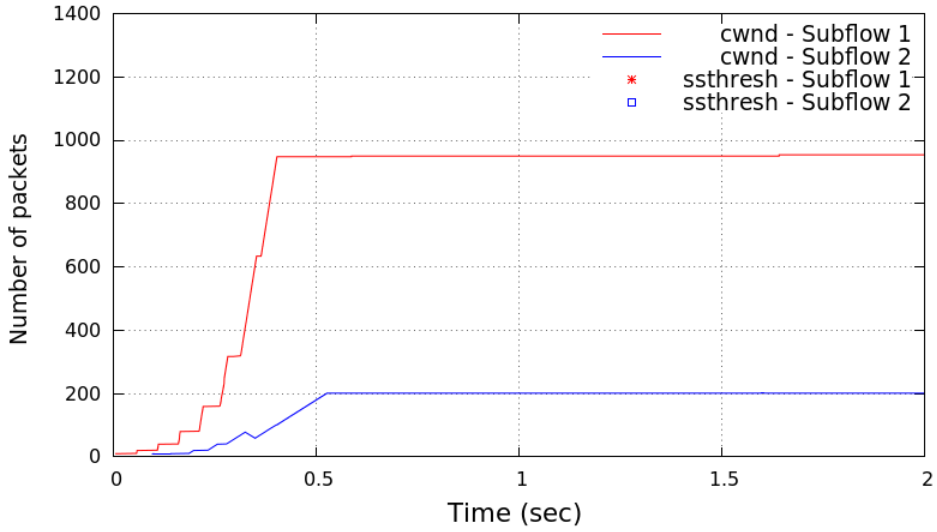


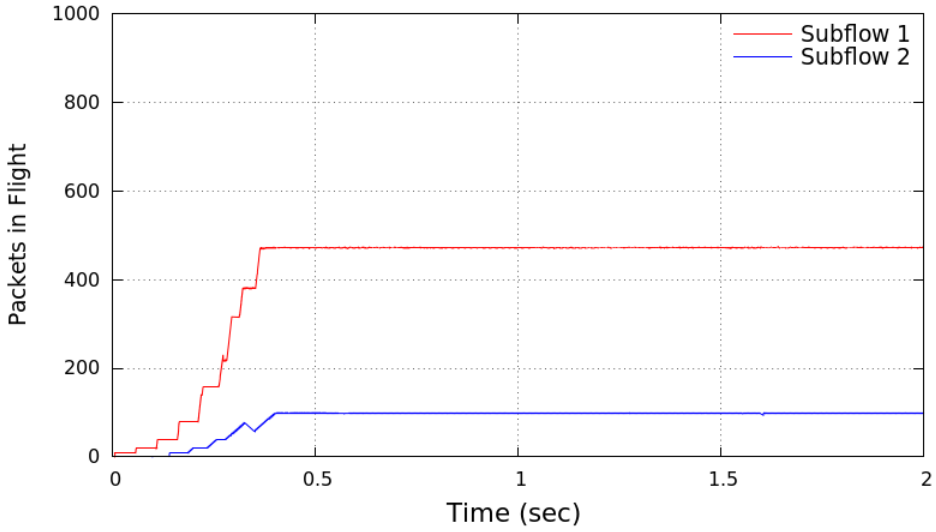


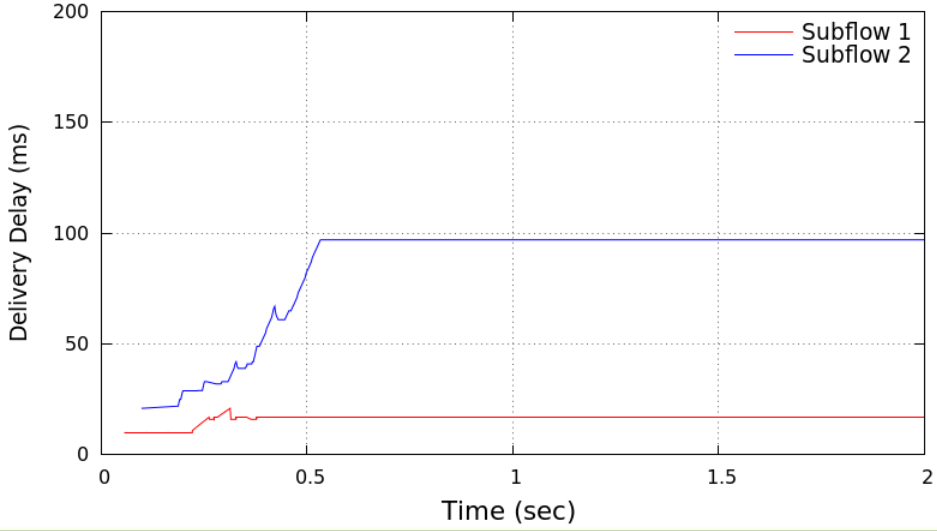


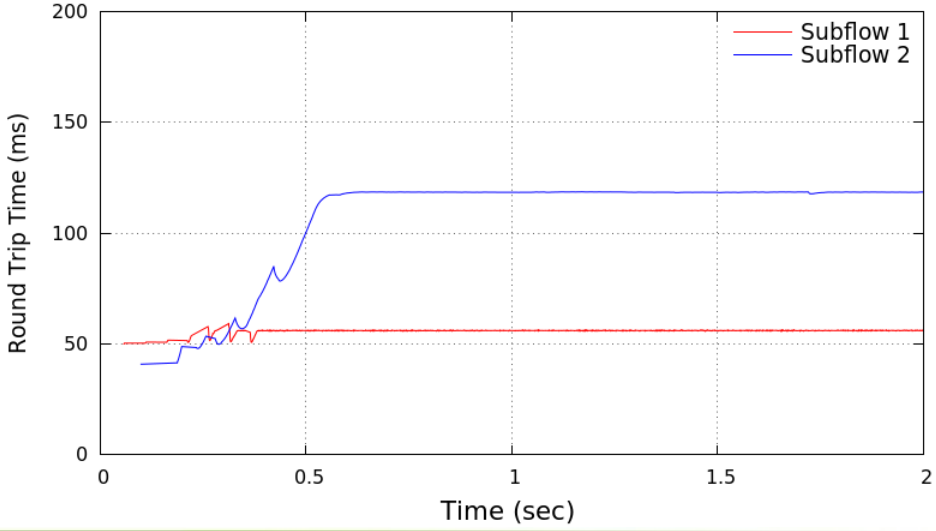


sRTT Scheduler



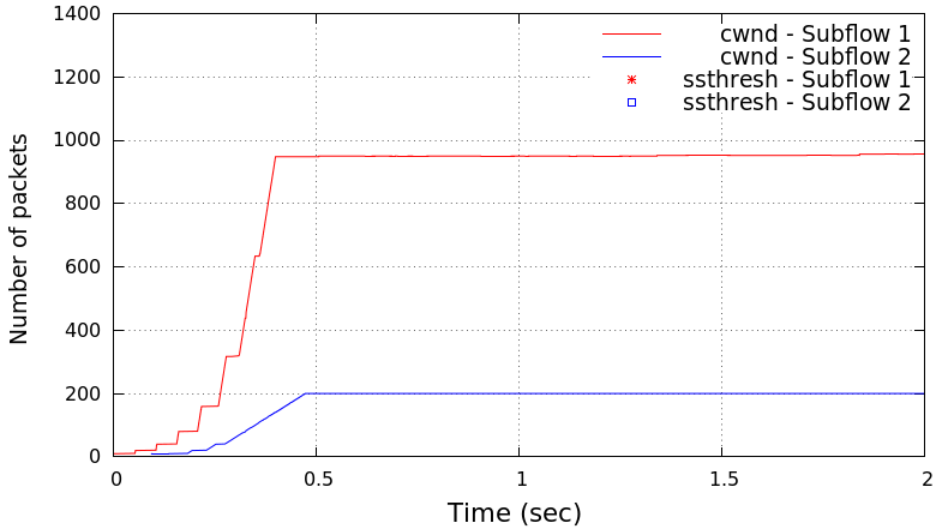


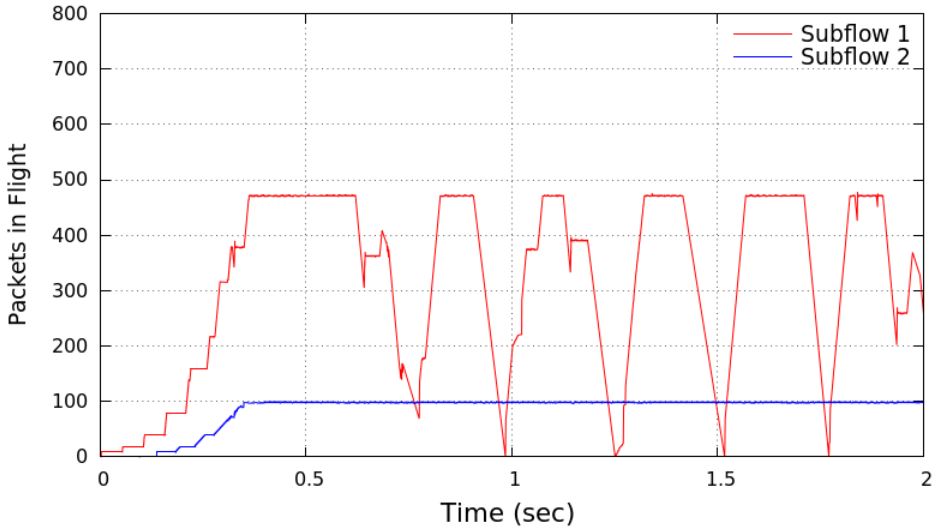


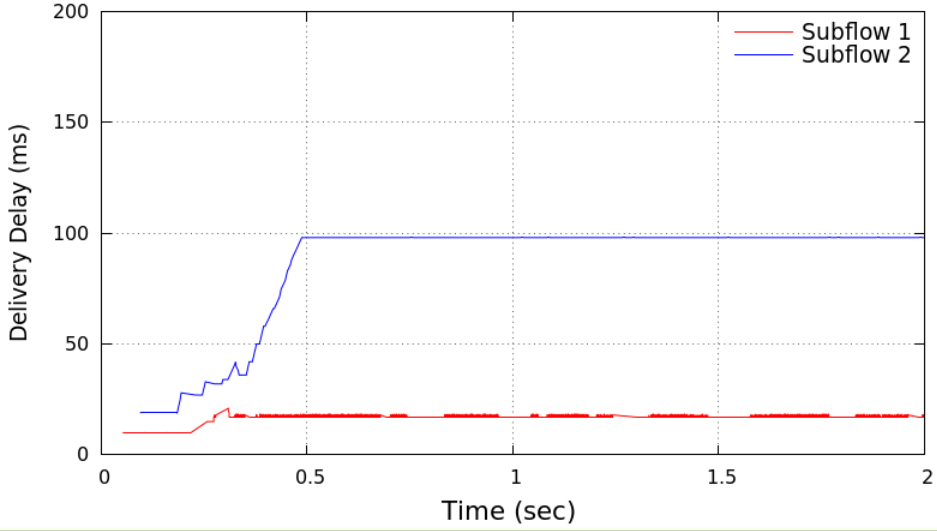


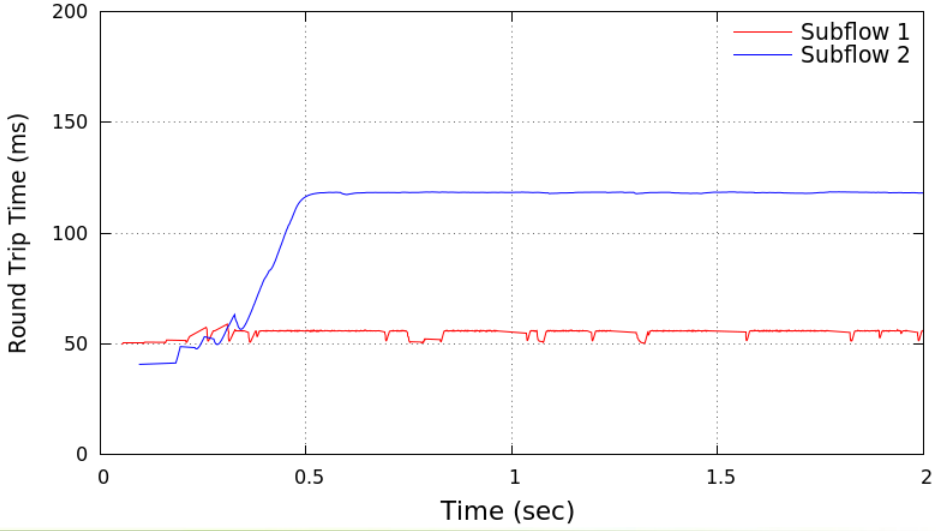
Faridul Alam
(faridul_alam@ymail.com)

RR Scheduler

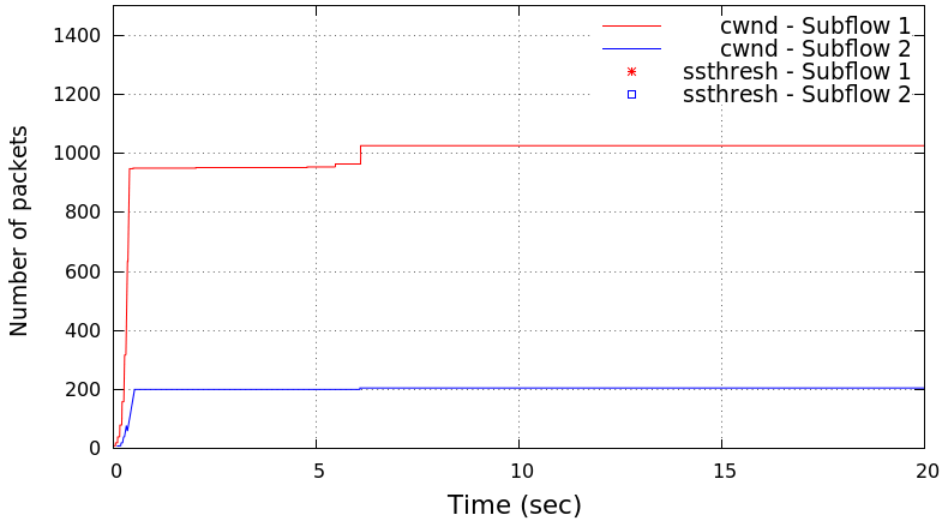


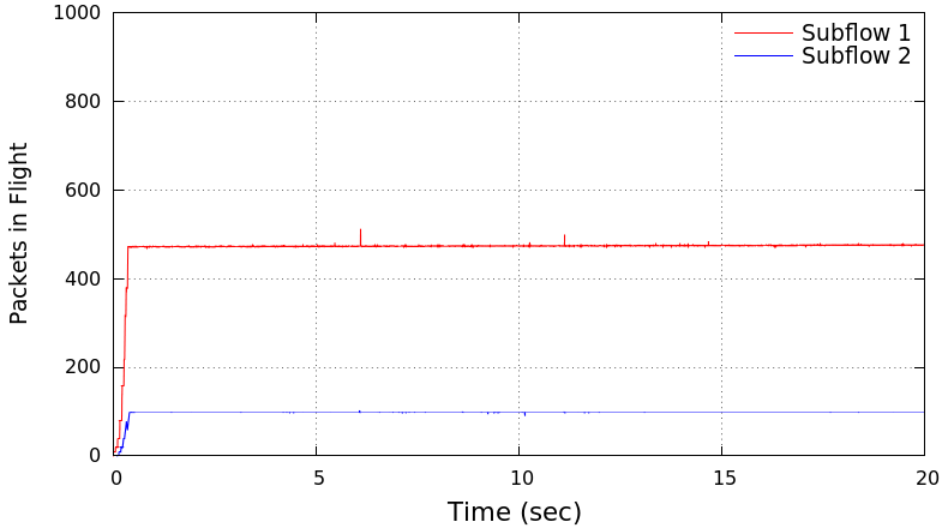


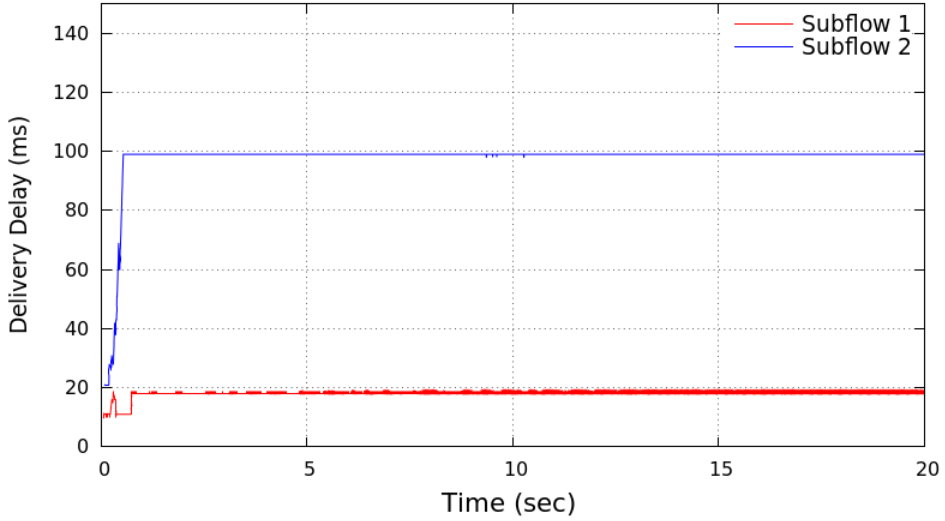


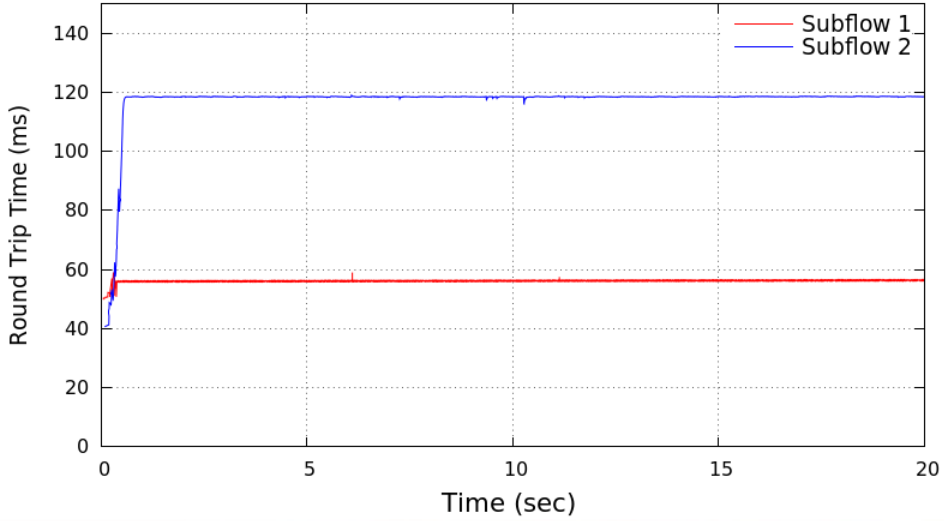


Delivery Delay Scheduler (20 sec)

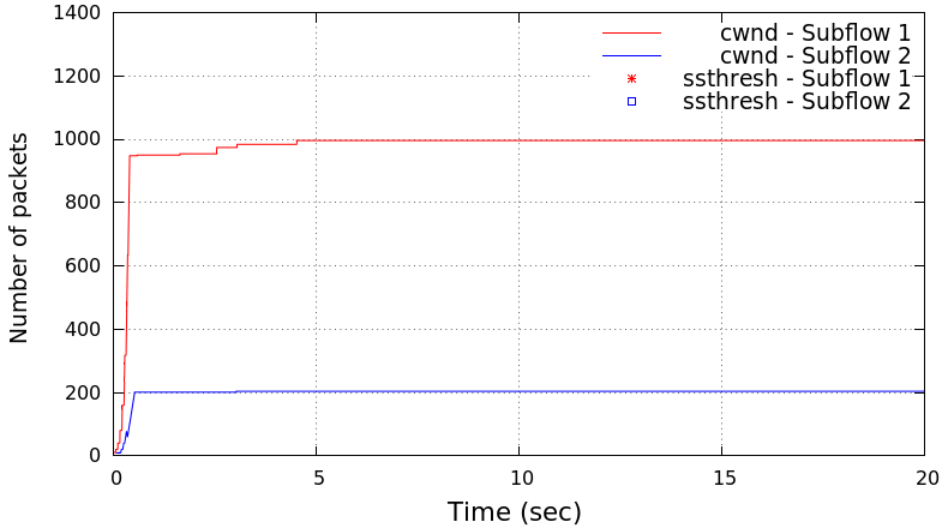


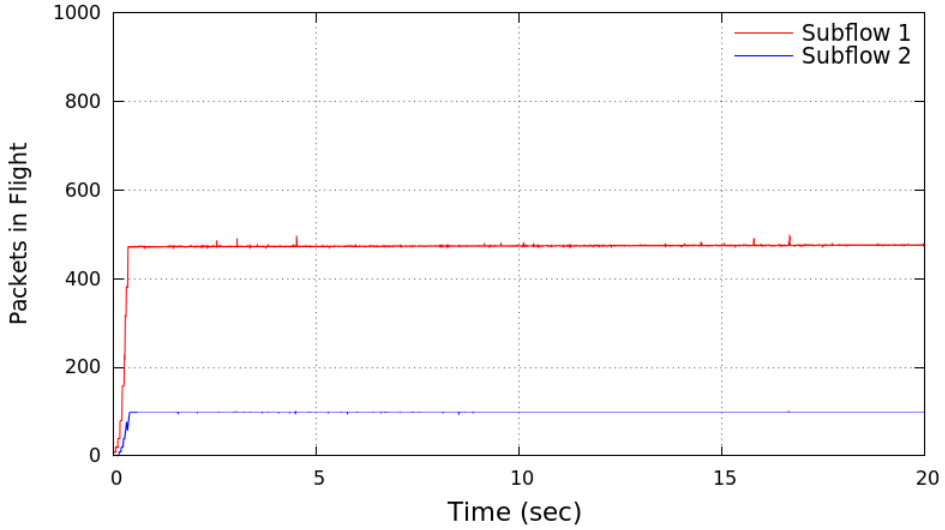


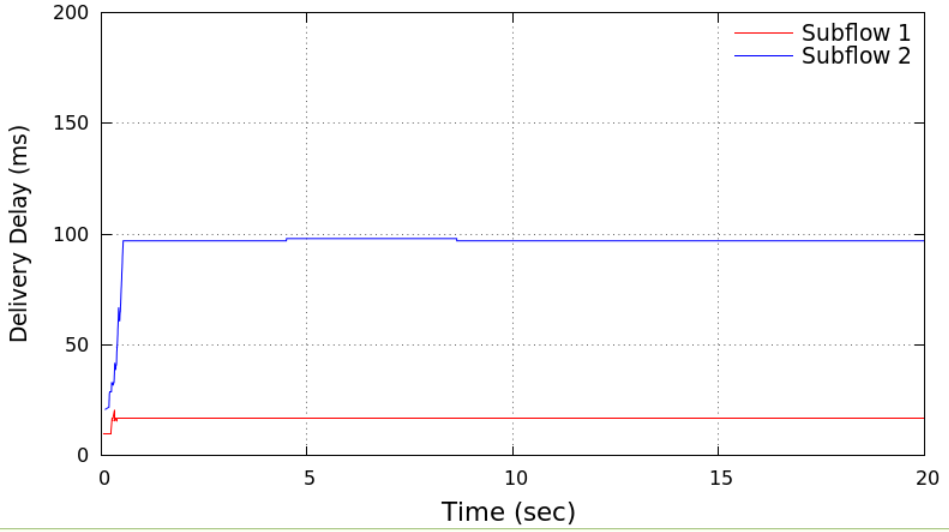


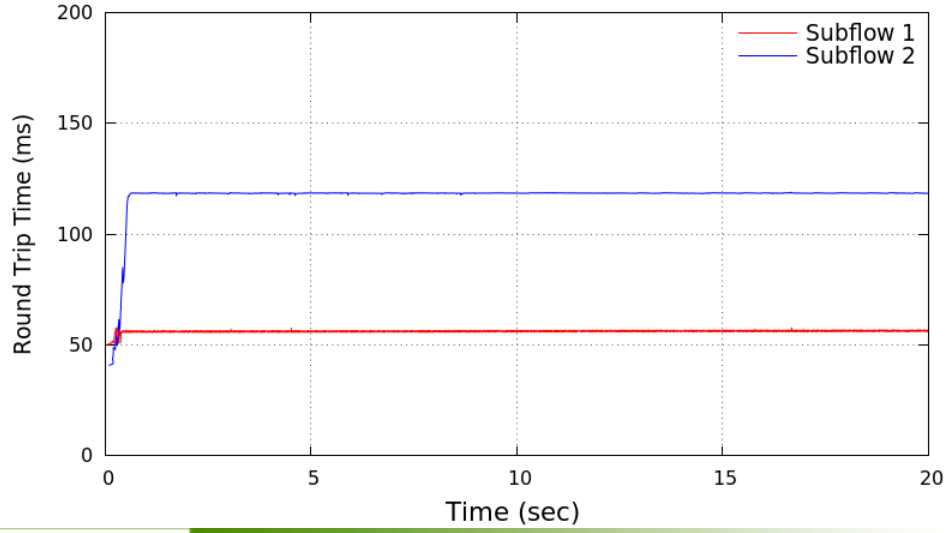


Shortest Round Trip Time Scheduler (20 sec)

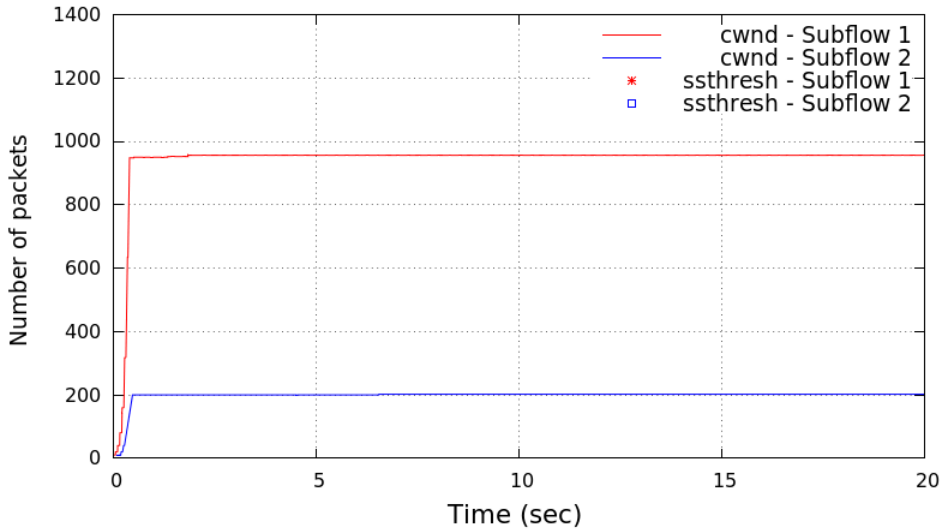


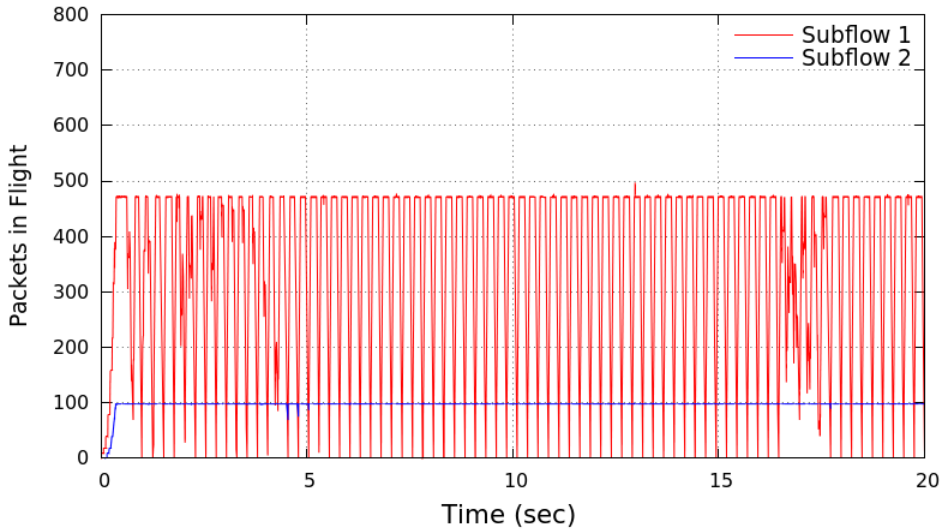


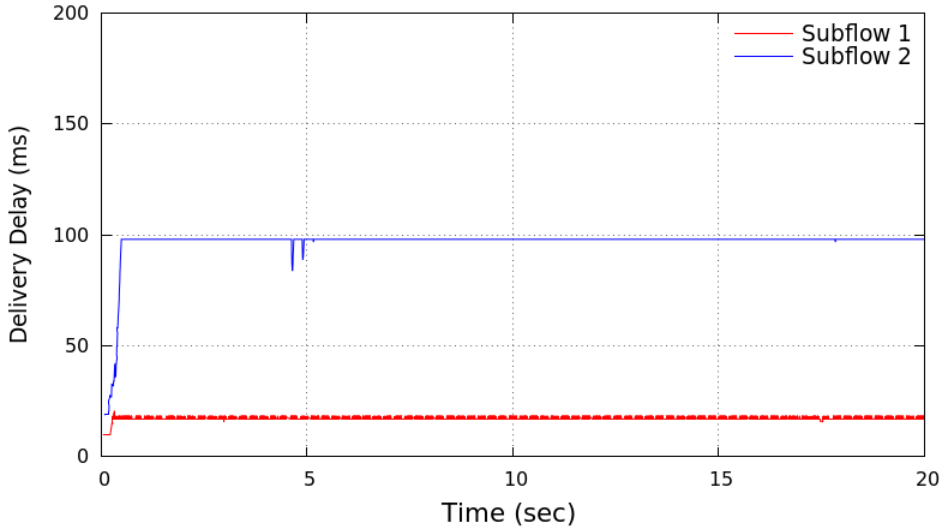


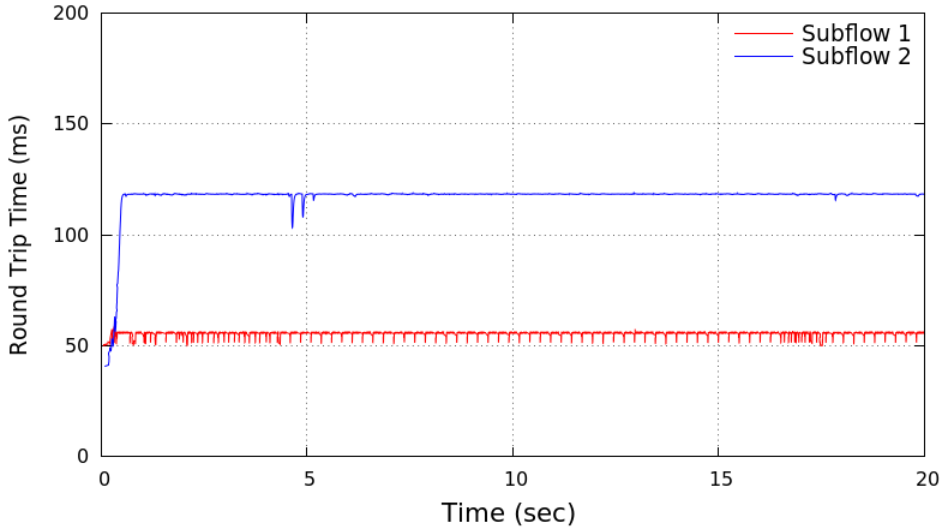


Round Robin Scheduler (20 sec)









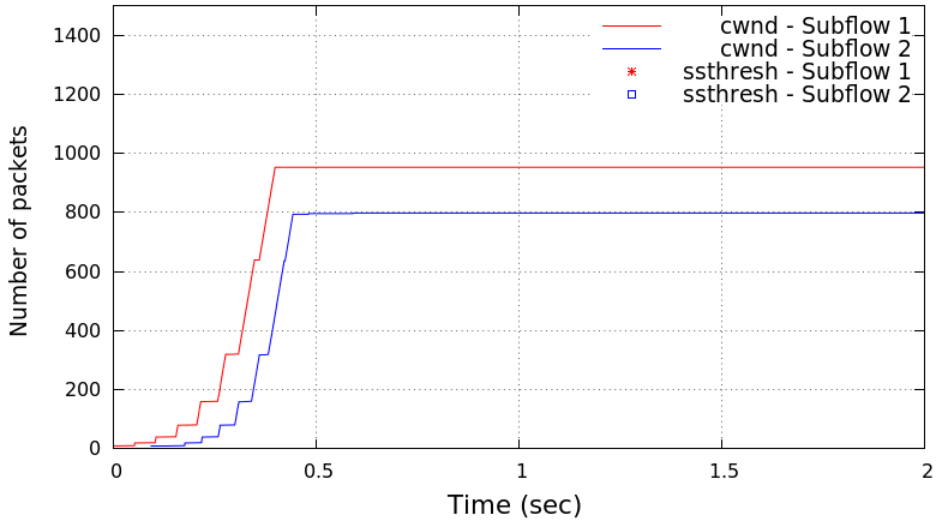
Graphical Analysis

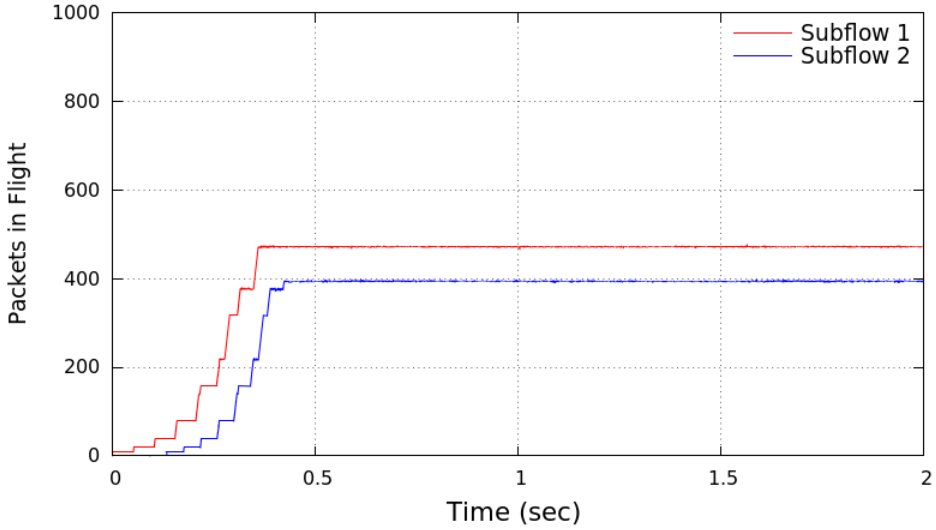
Link Capacity 100/100 (Mbit/sec)

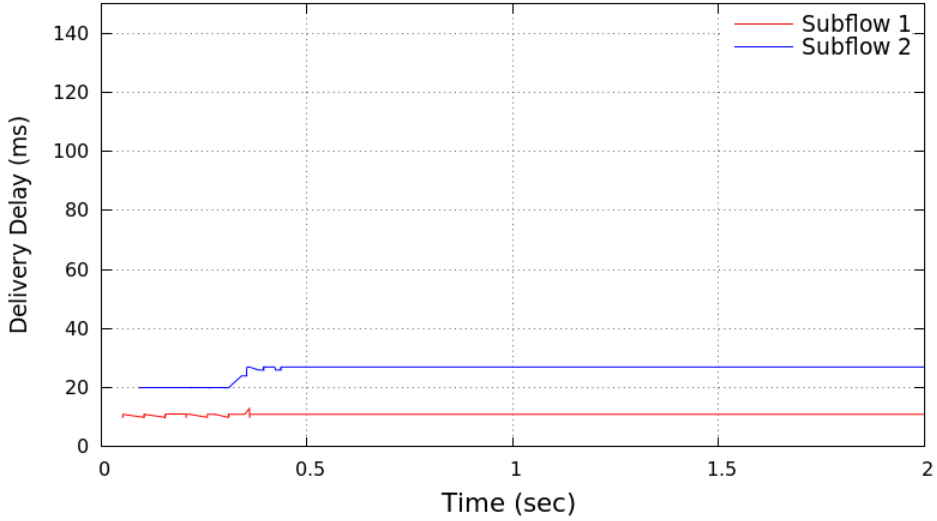
Scenario C

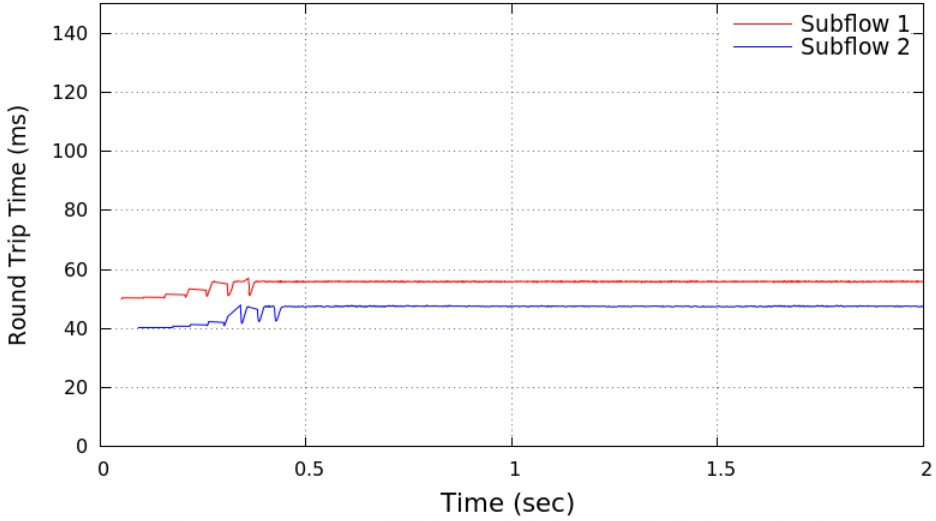
Subflow 1 RTT – 50 ms (10 ms + 40 ms)
Subflow 2 RTT – 40 ms (20 ms + 20 ms)

Delivery Delay Scheduler

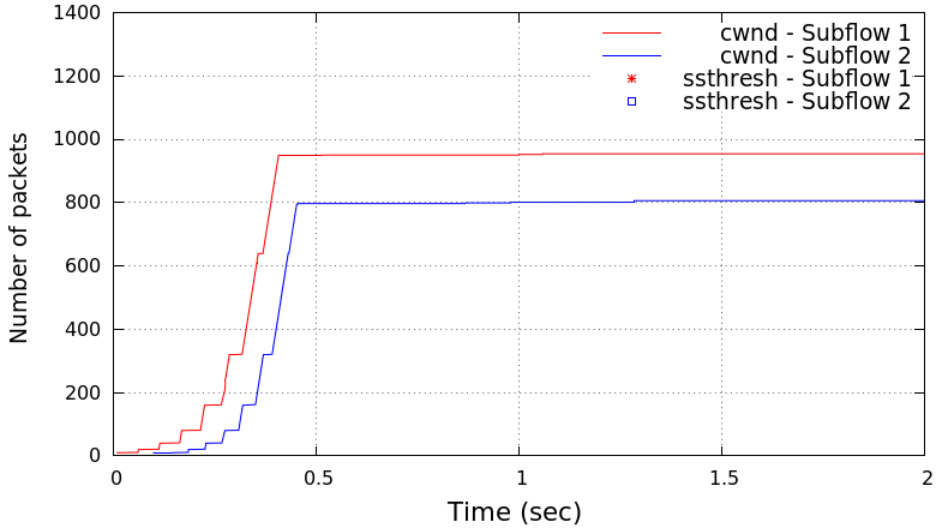


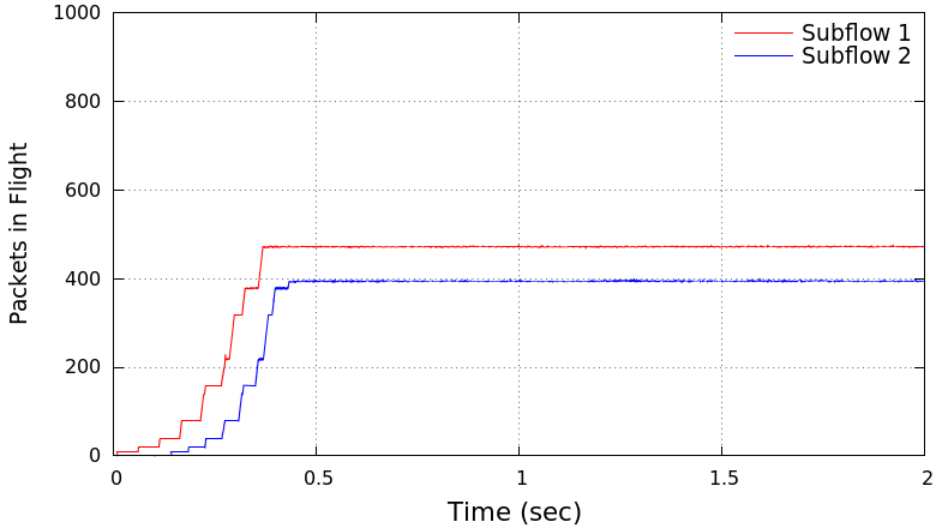


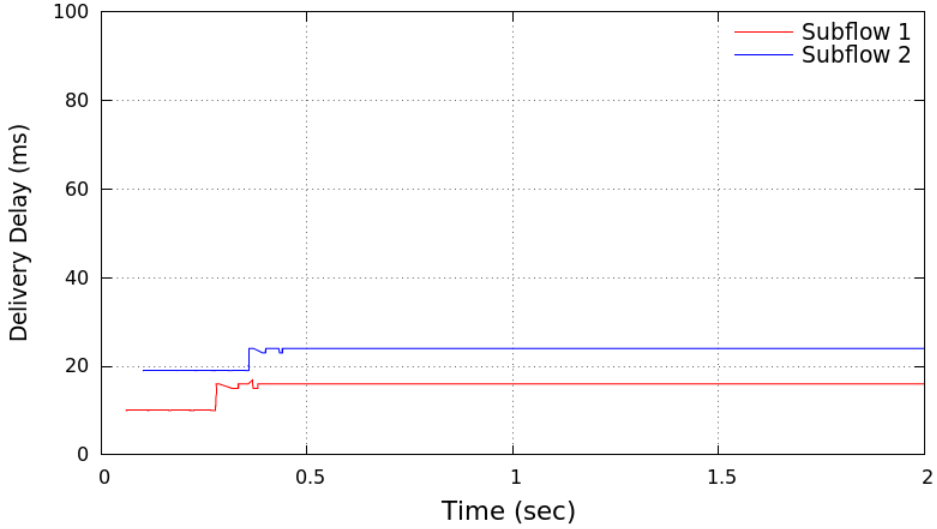


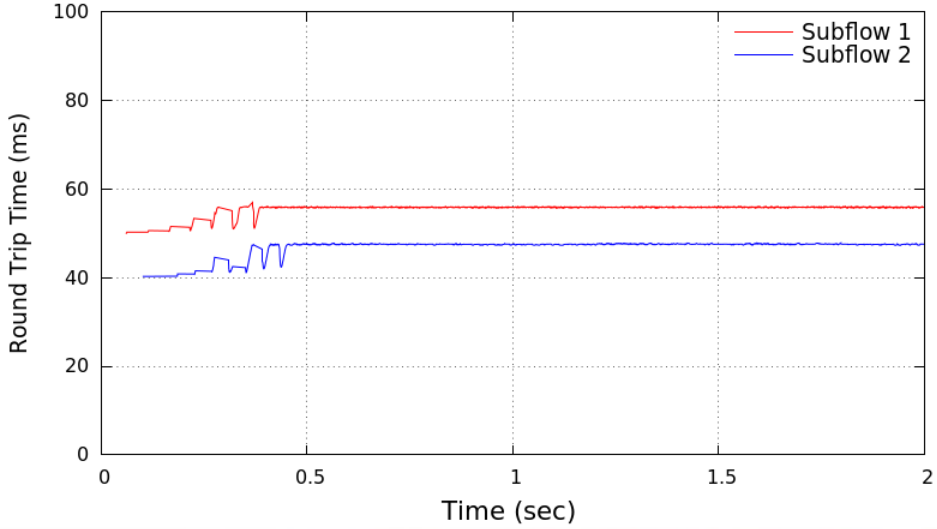


Shortest Round Trip Time Scheduler

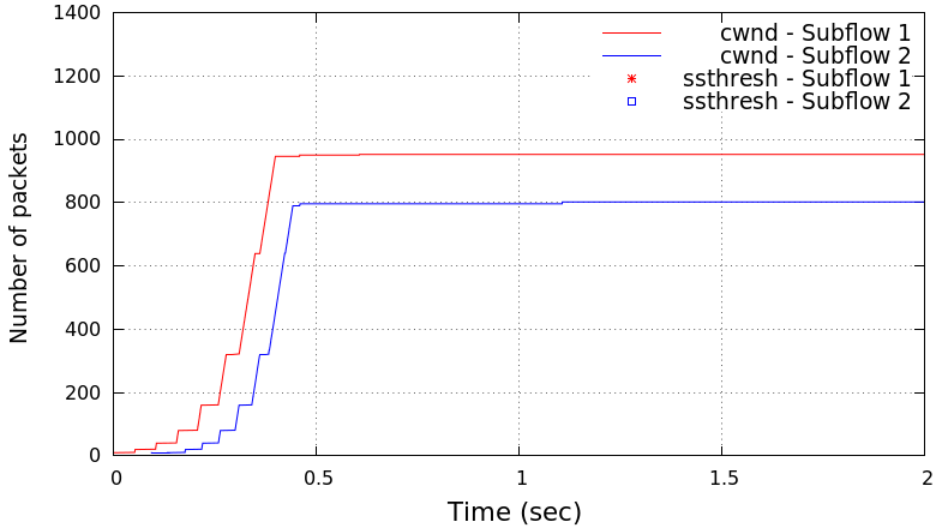


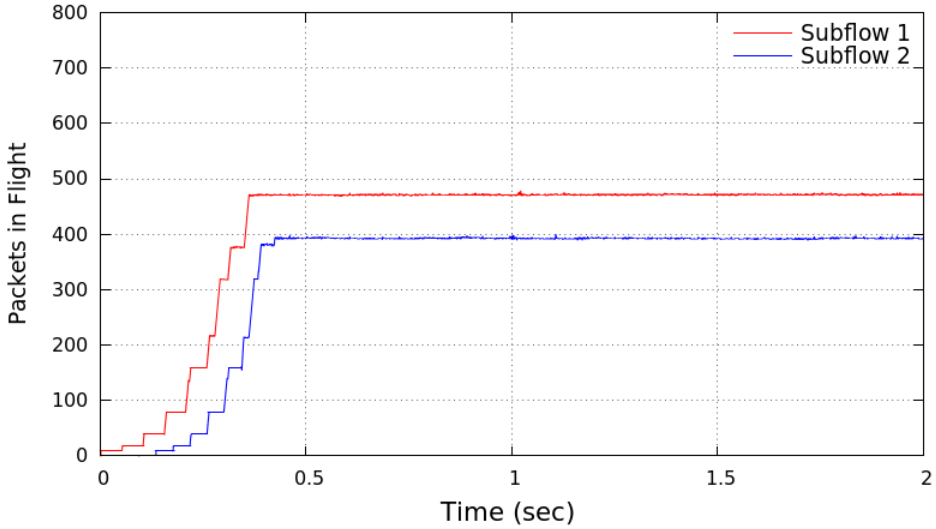


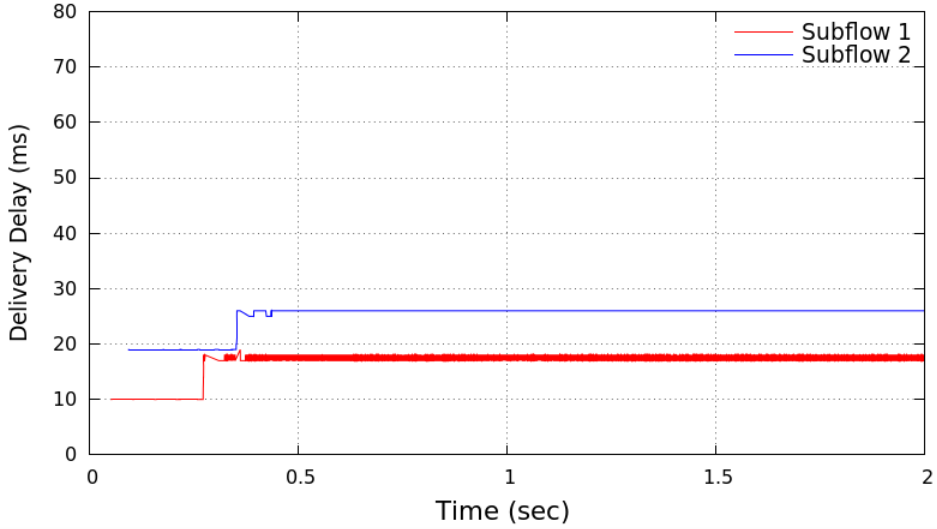


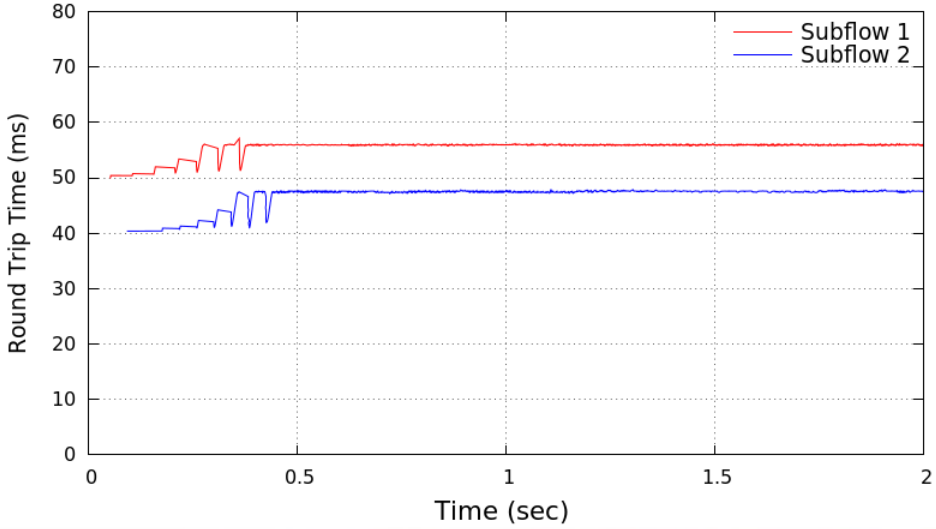


Round Robin Scheduler

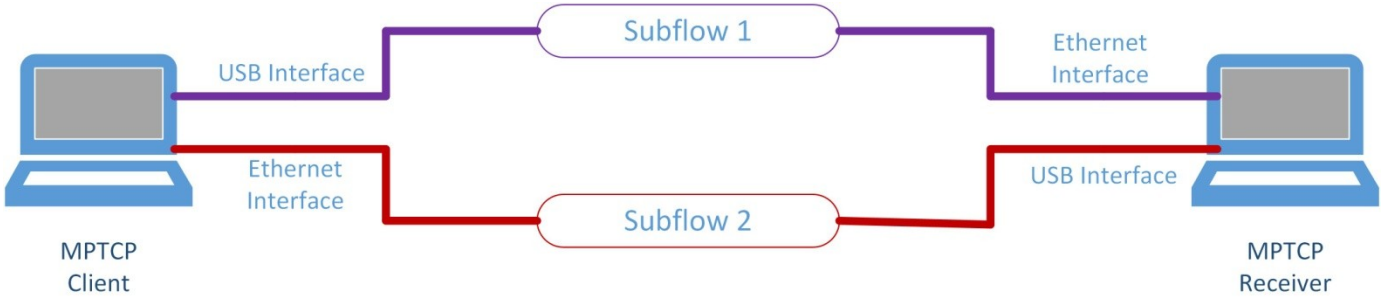








Graphical Analysis – 2 Node without Router



Link Capacity 100/10 (Mbit/sec)

Scenario B

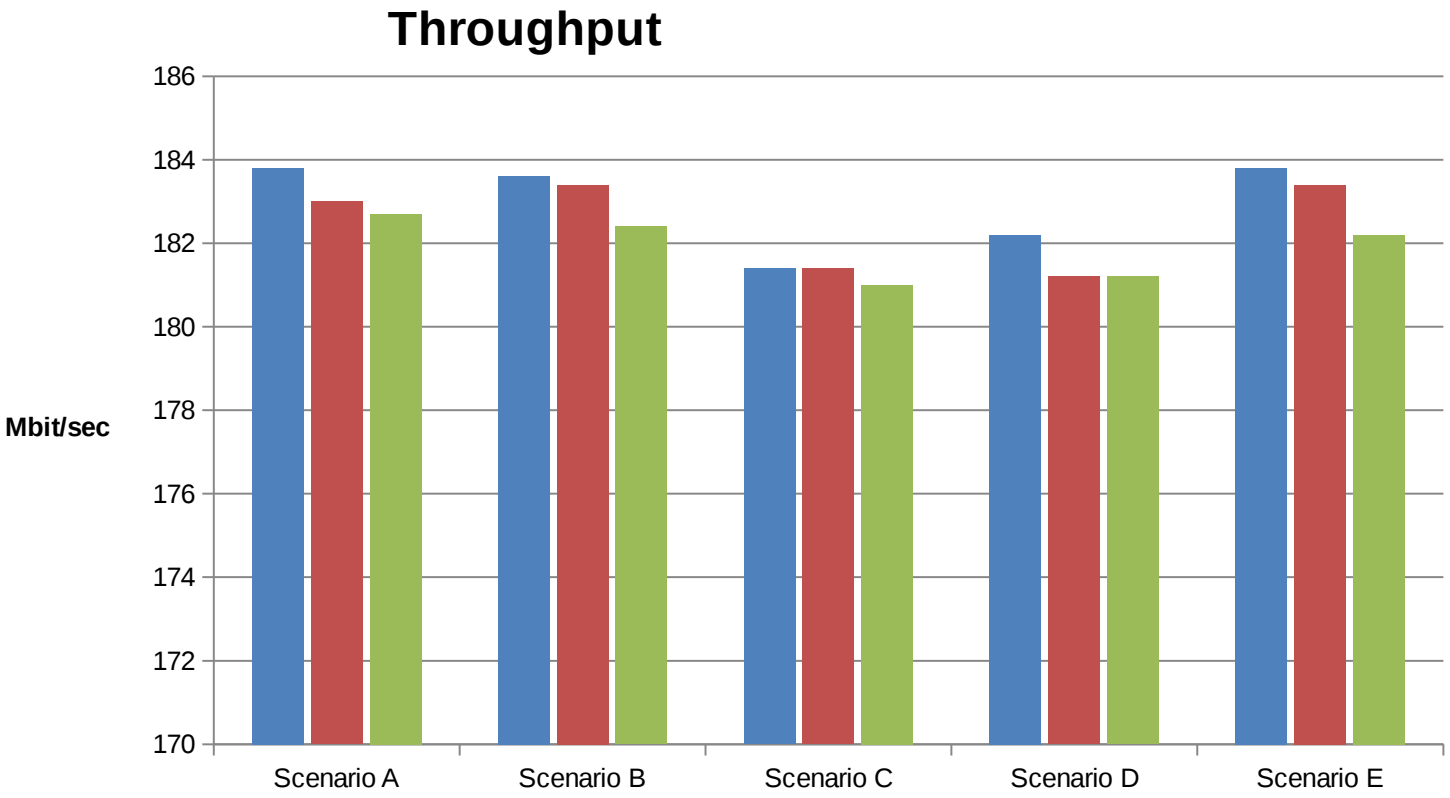
Subflow 1 RTT – 20 ms (10 ms + 10 ms)

Subflow 2 RTT – 30 ms (20 ms + 10 ms)

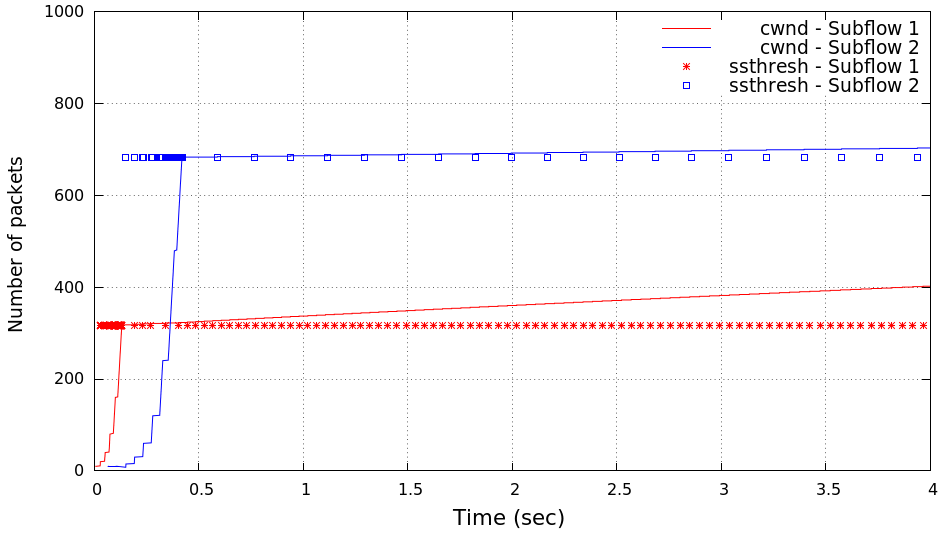
Scenarios

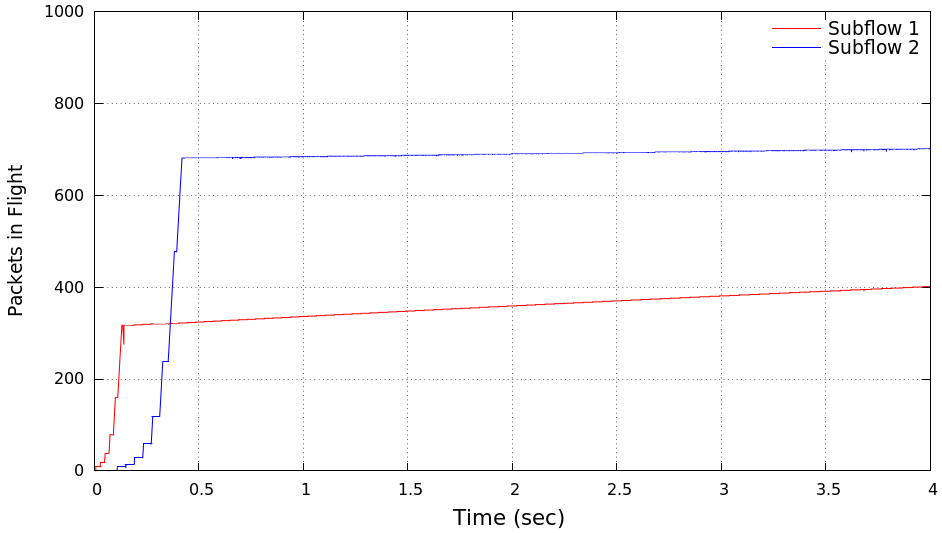
Capacity	Name Of scenarios	Subflow 1			Subflow 2			TCP Flow		
		FPD (ms)	BPD (ms)	RTT (ms)	FPD (ms)	BPD (ms)	RTT (ms)	FPD (ms)	BPD (ms)	RTT (ms)
100/100 & 100/10 & 10/10	A	10	20	30	20	10	30	10	20	30
	B	10	10	20	20	10	30	10	10	20
	C	10	40	50	20	20	40	10	40	50
	D	10	20	30	20	40	60	10	20	30
	E	10	10	20	20	20	40	10	10	20

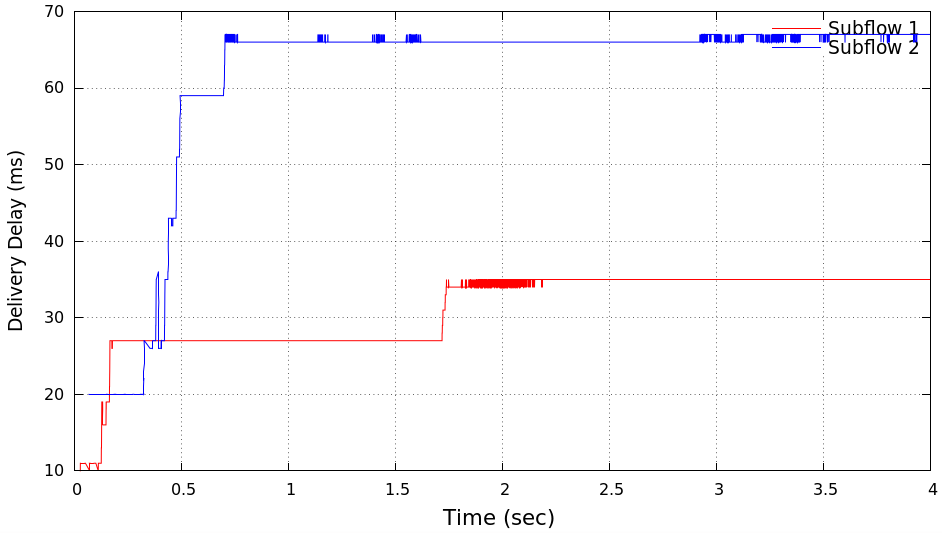
Link Capacity 100/100 (Mbit/sec)

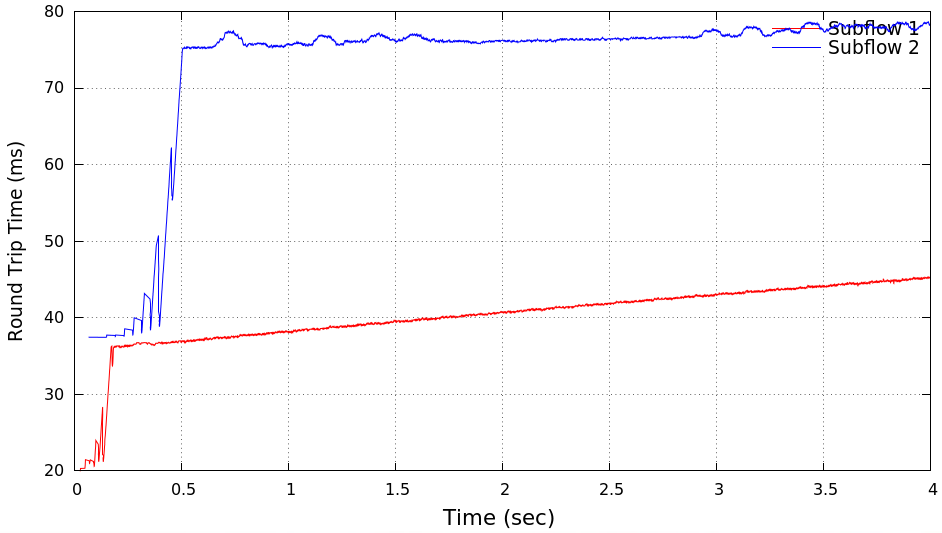


Shortest Round Trip Time Scheduler – E (100/100)

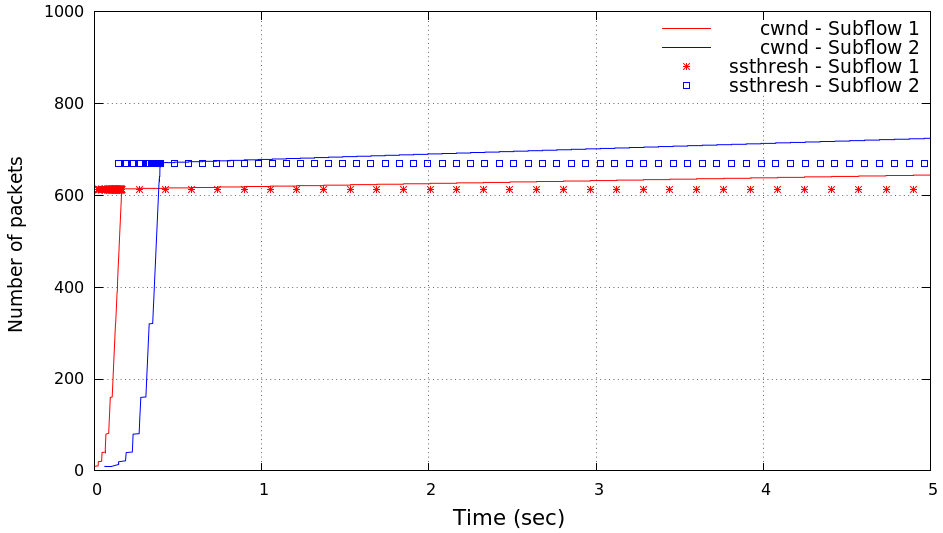


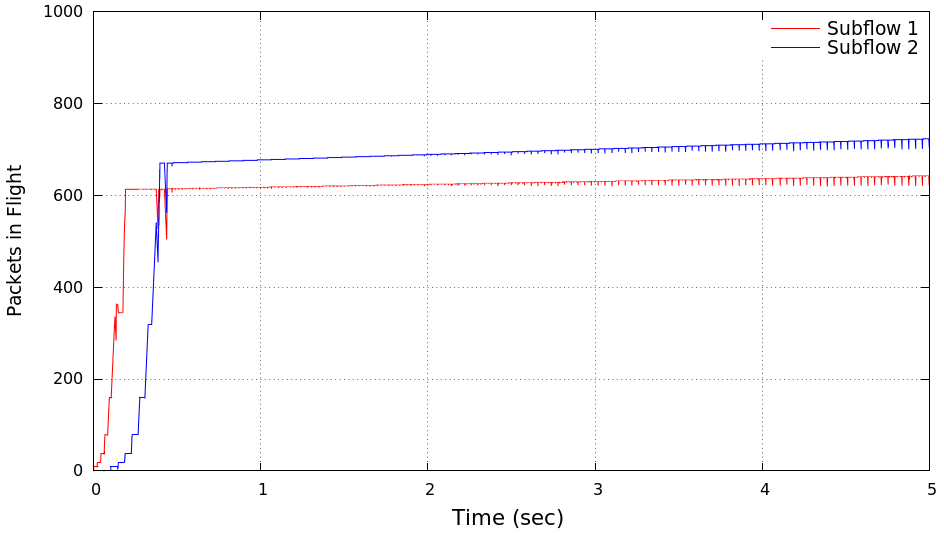


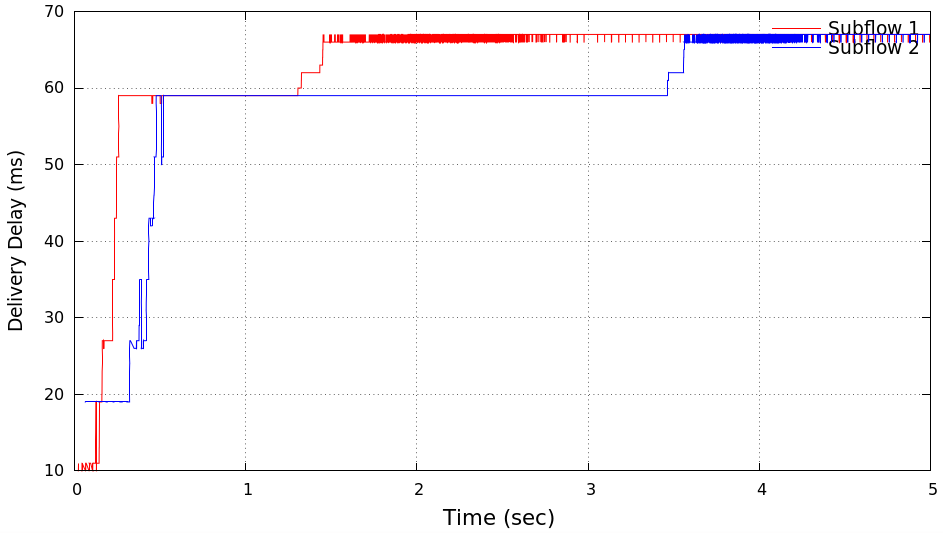


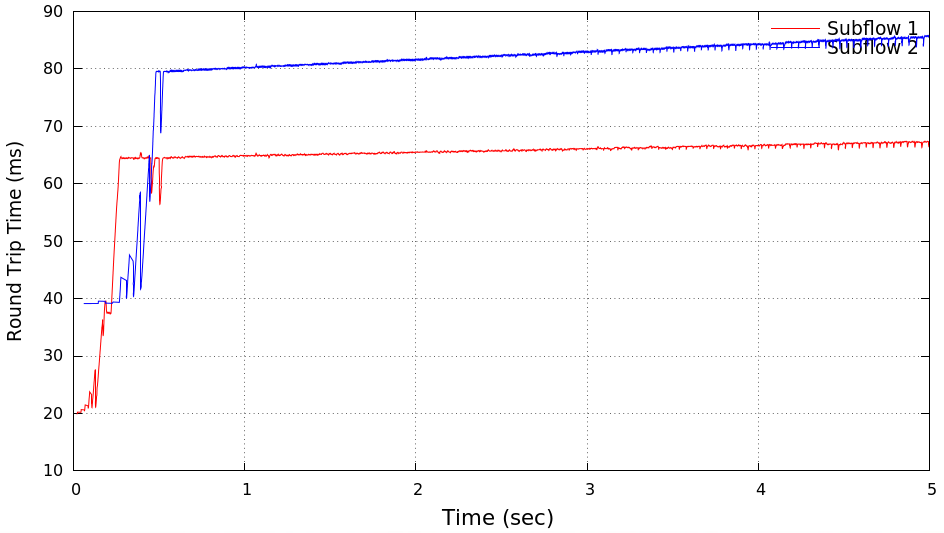


Round Robin Scheduler - E (100/100)

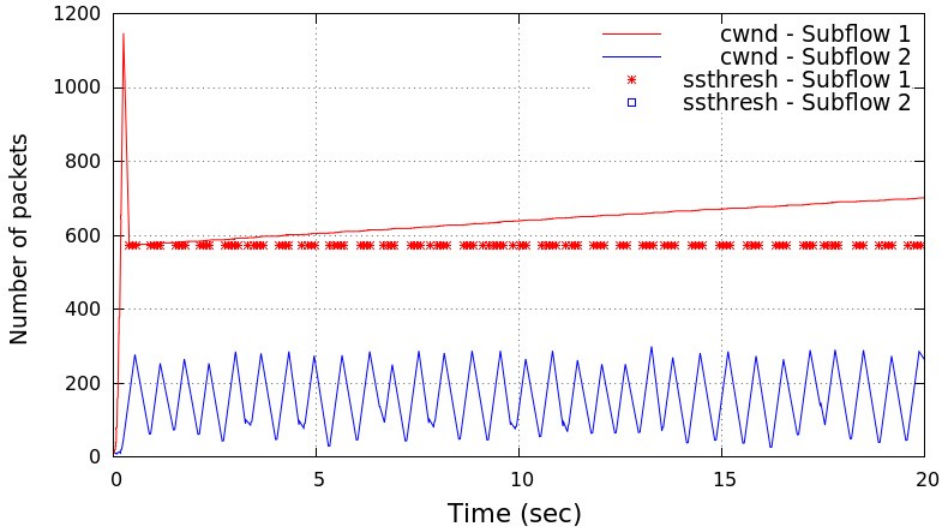


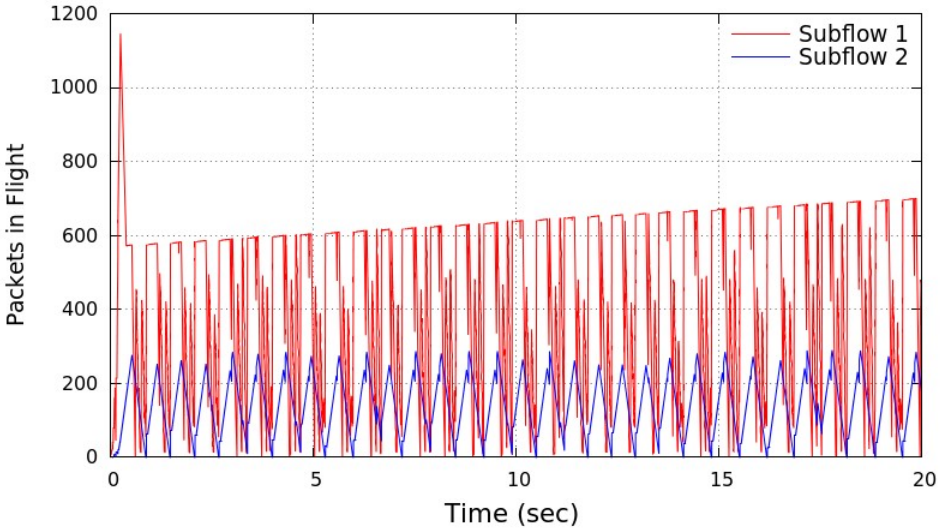


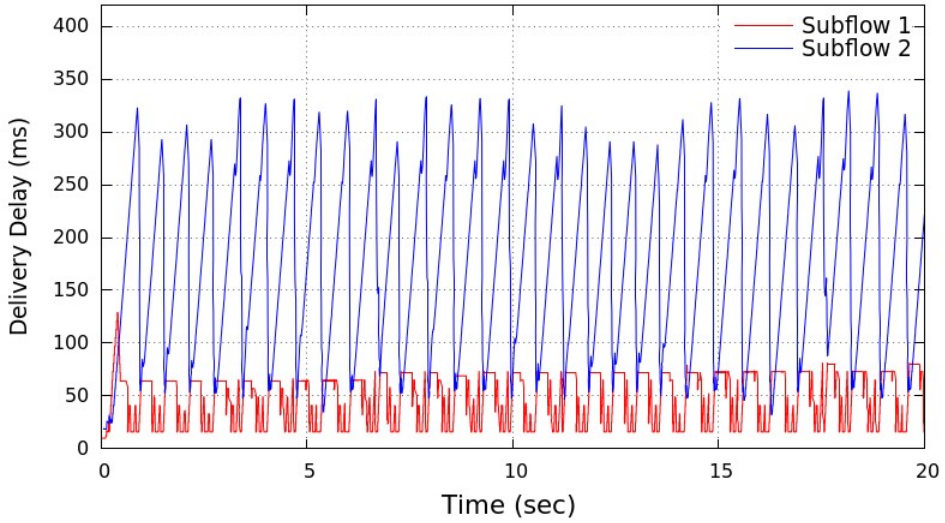


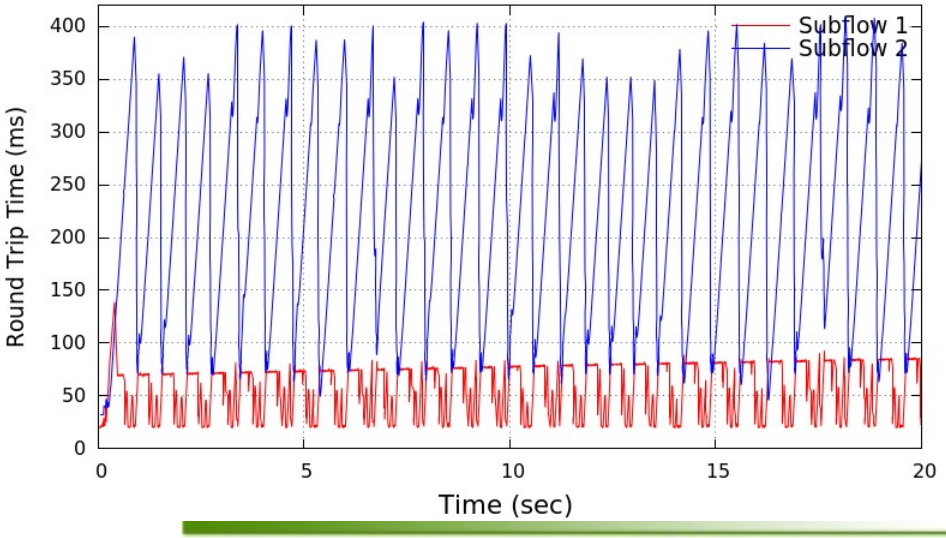


Delivery Delay Scheduler

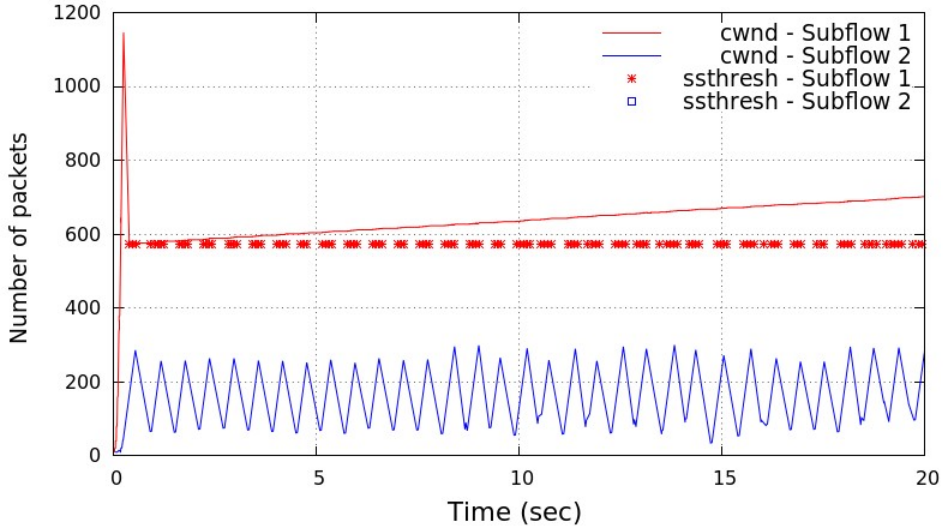


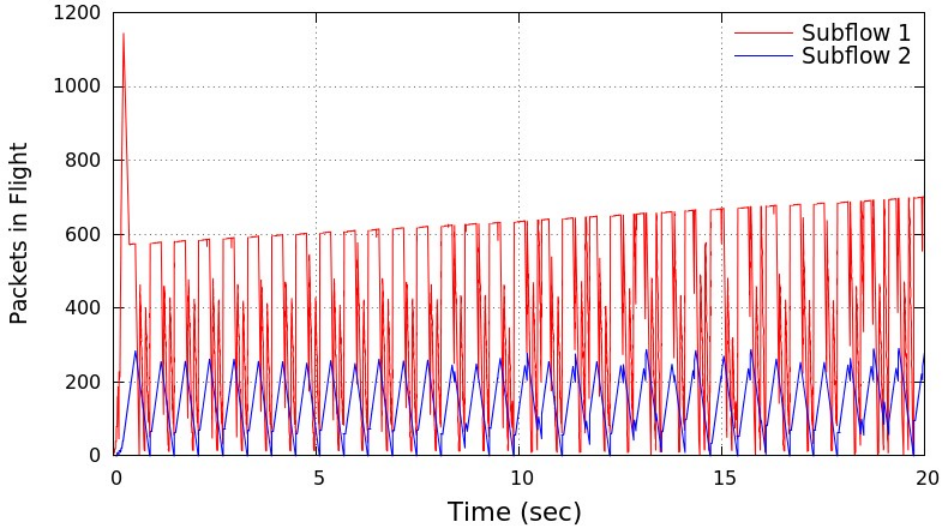


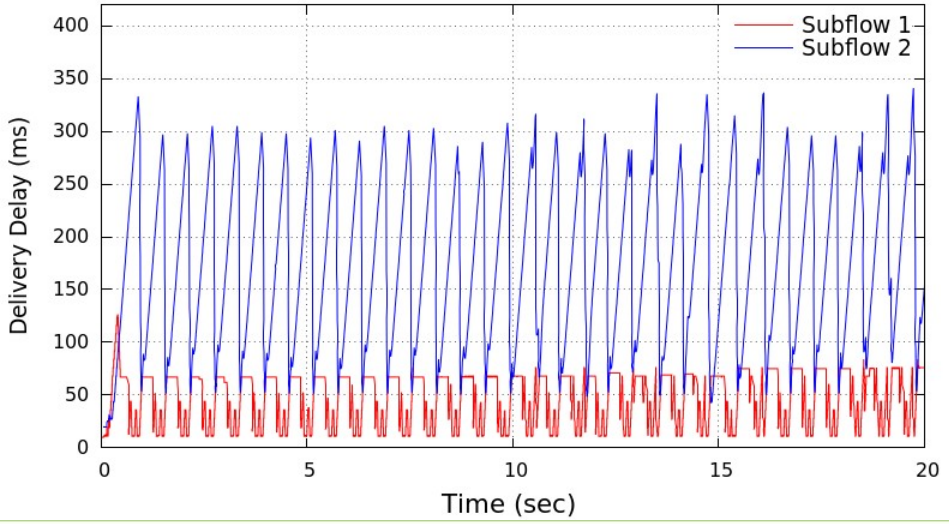


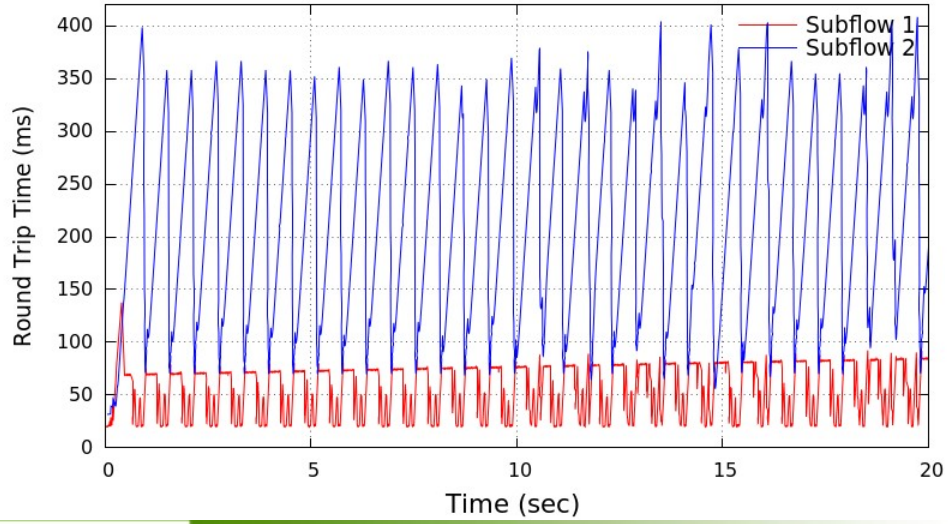


Shortest Round Trip Scheduler Scheduler

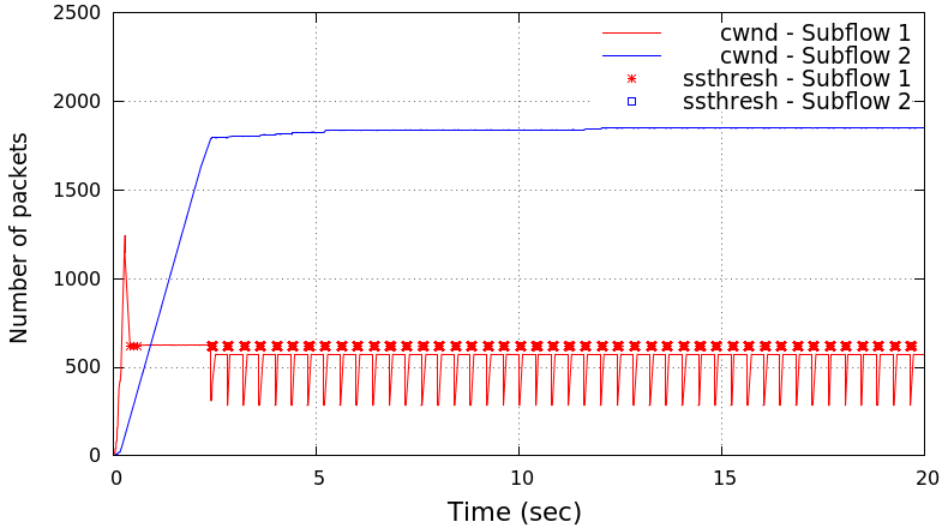


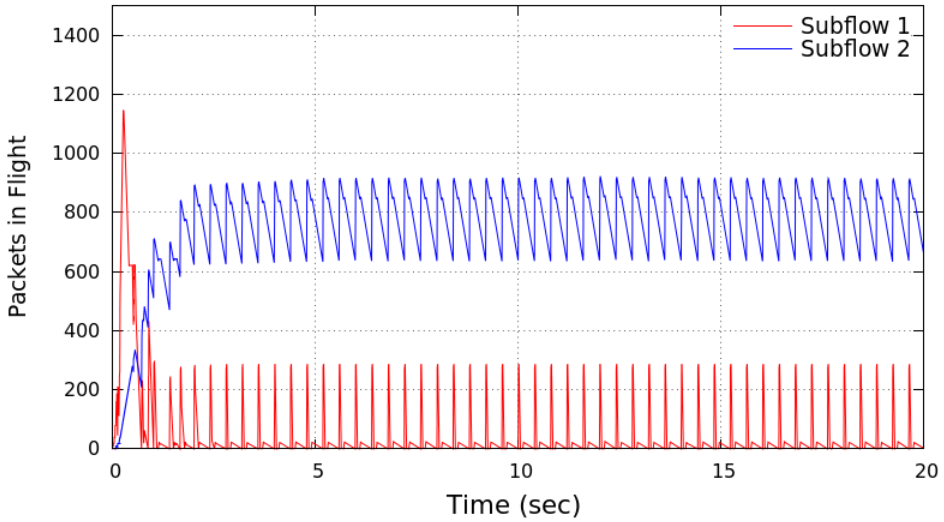


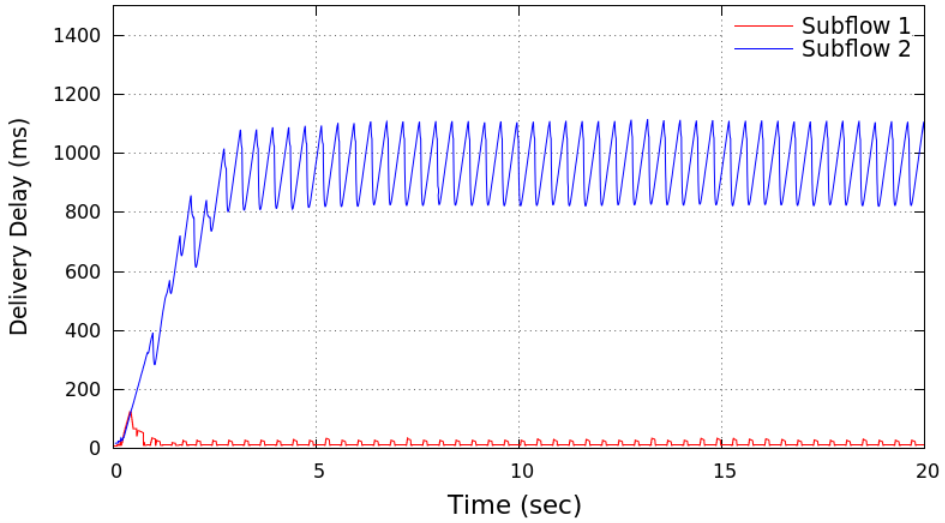


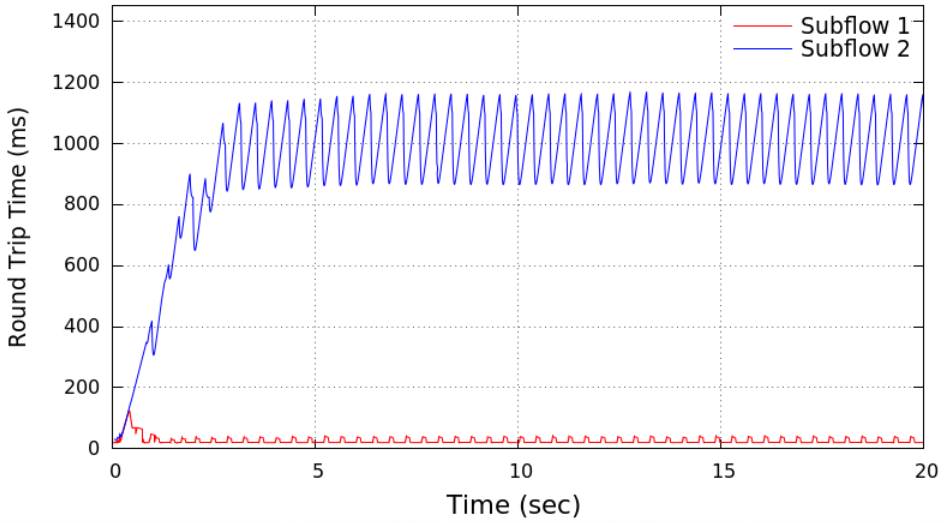


Round Robin Scheduler









Test Topology – Extended Testbed

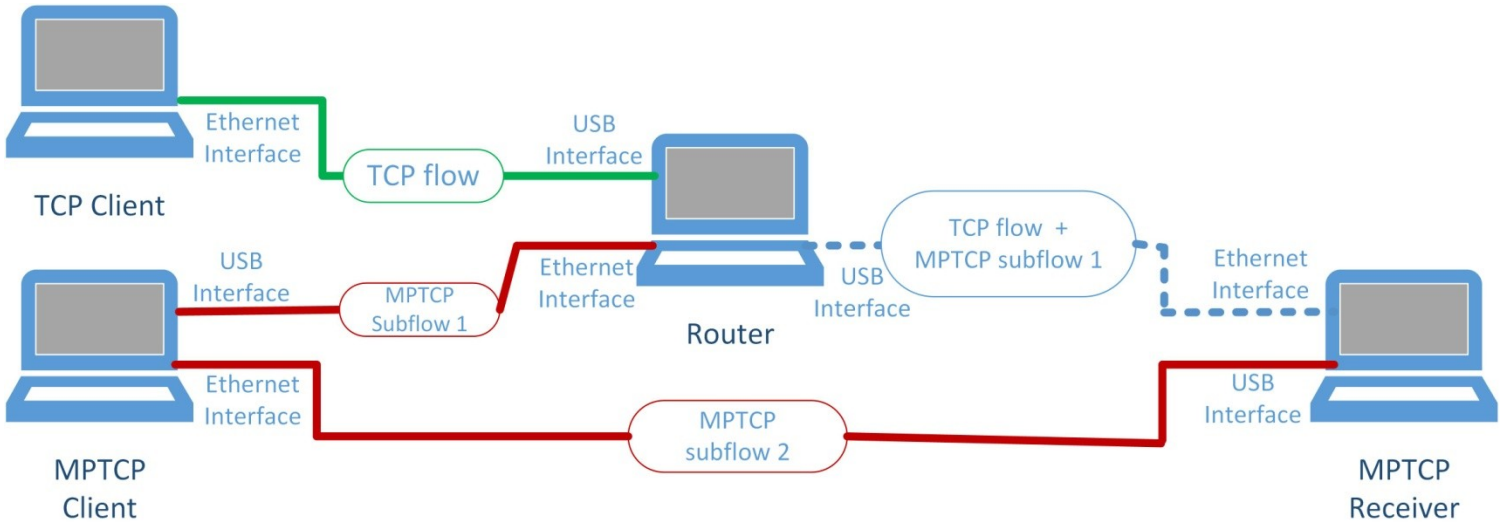
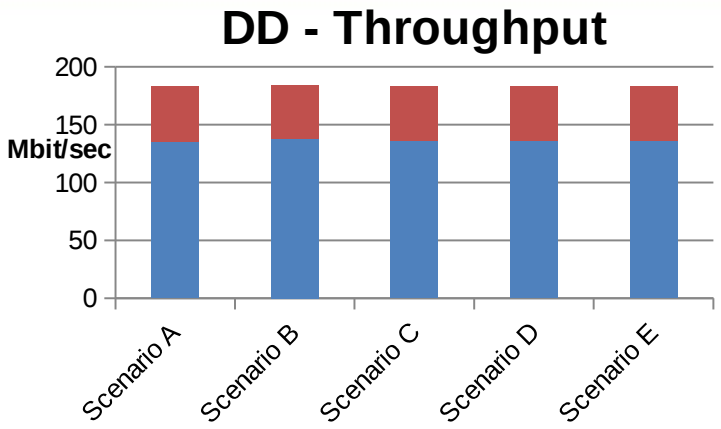
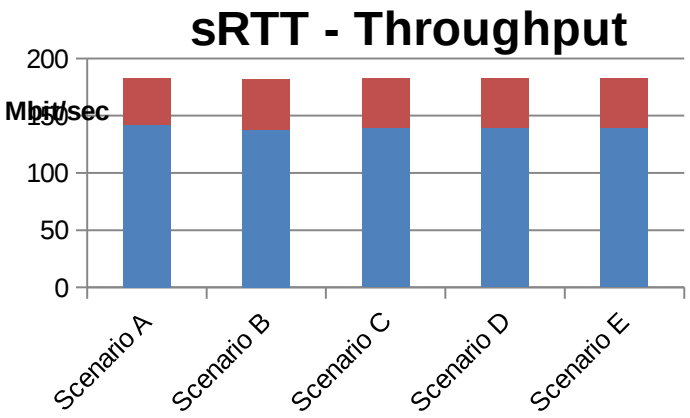
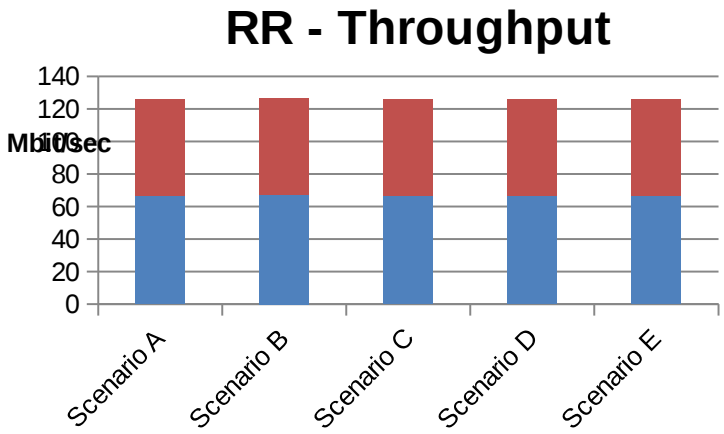


Figure : Extended Testbed

Throughput Comparison







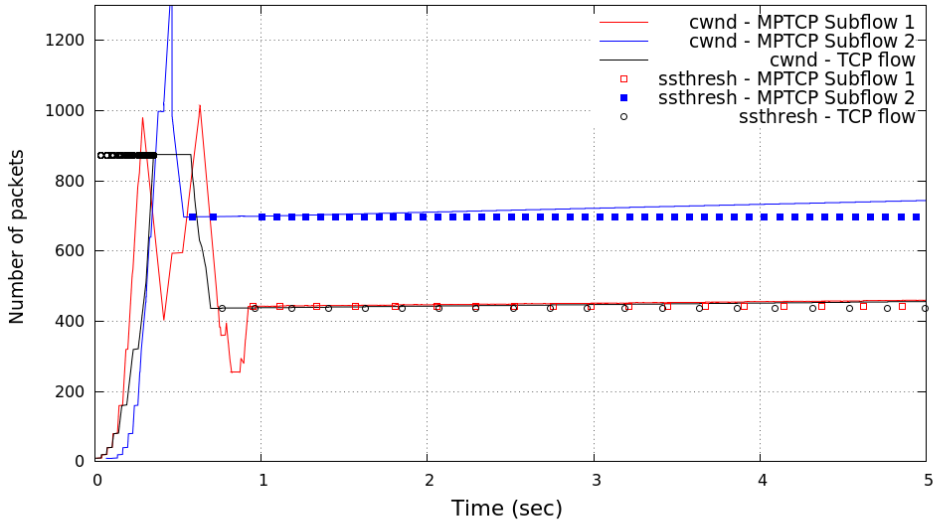
Graphical Analysis – Extended Testbed

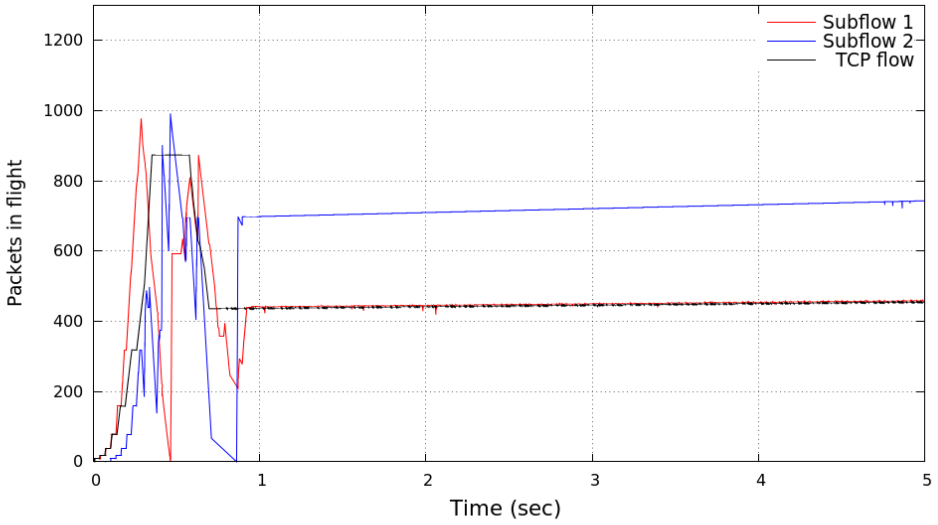
Link Capacity 100/100 (Mbit/sec)

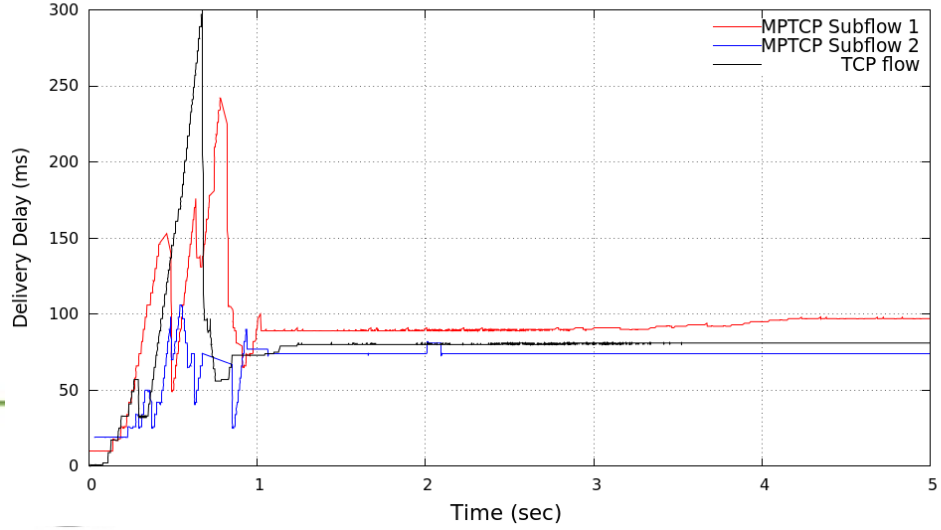
Scenario A

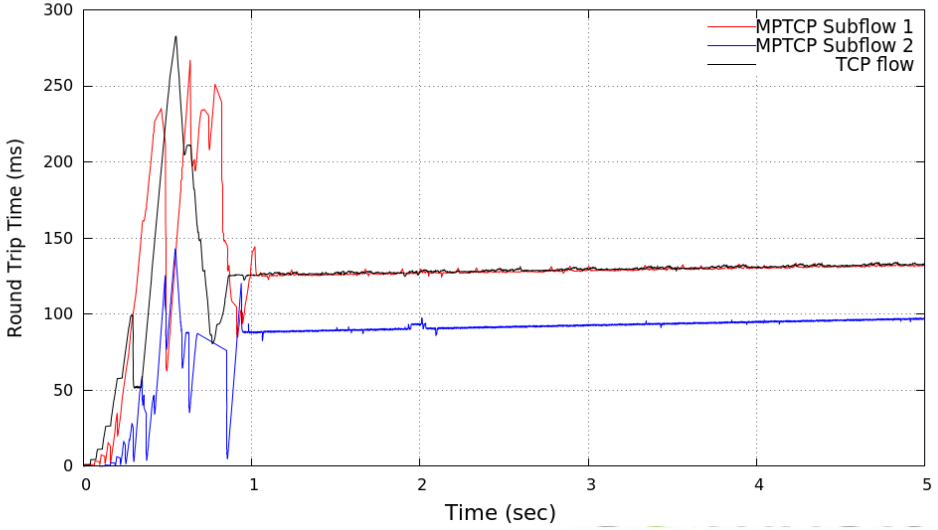
Subflow 1 RTT – 30 ms (10 ms +20 ms)
Subflow 2 RTT – 30 ms (20 ms + 10 ms)

Delivery Delay Scheduler

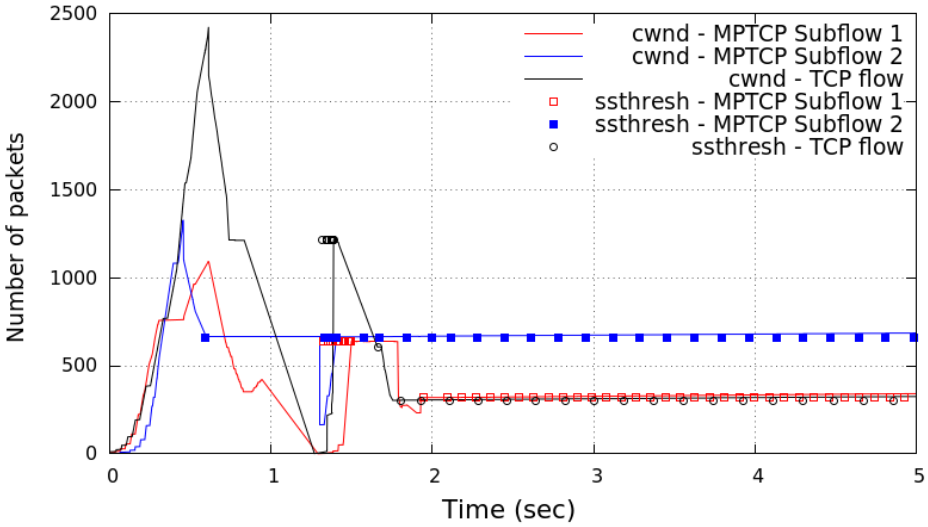


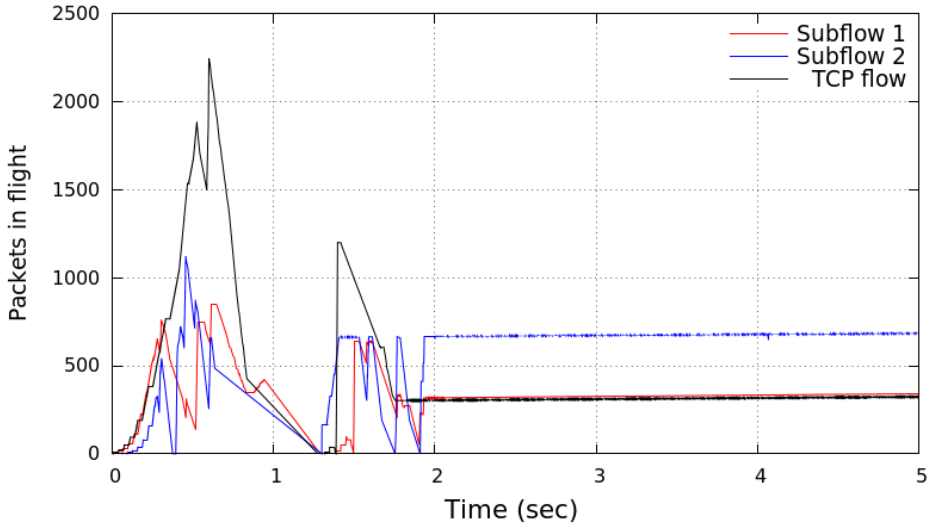


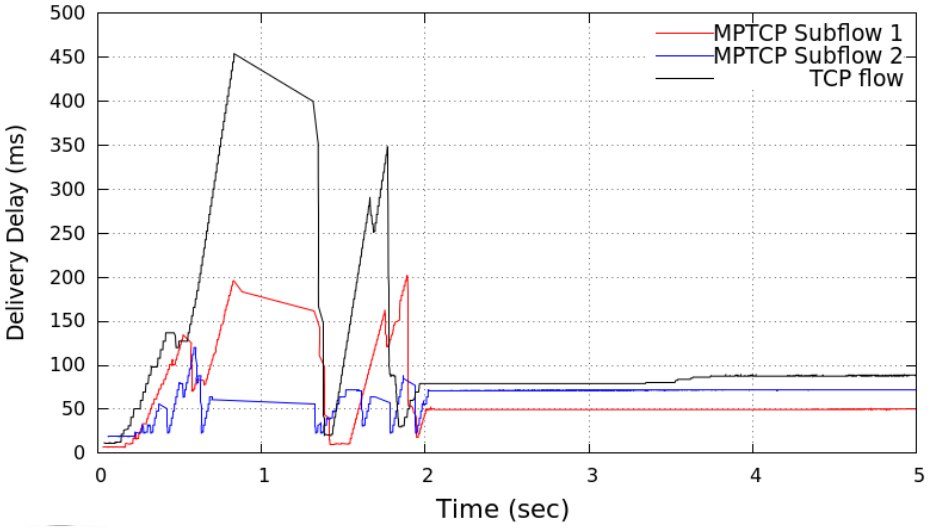


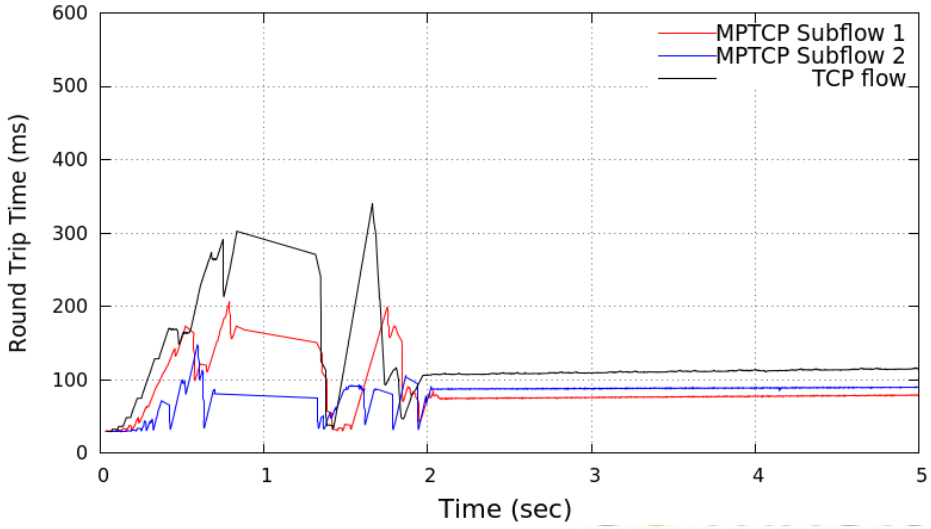


Round Trip Time Scheduler

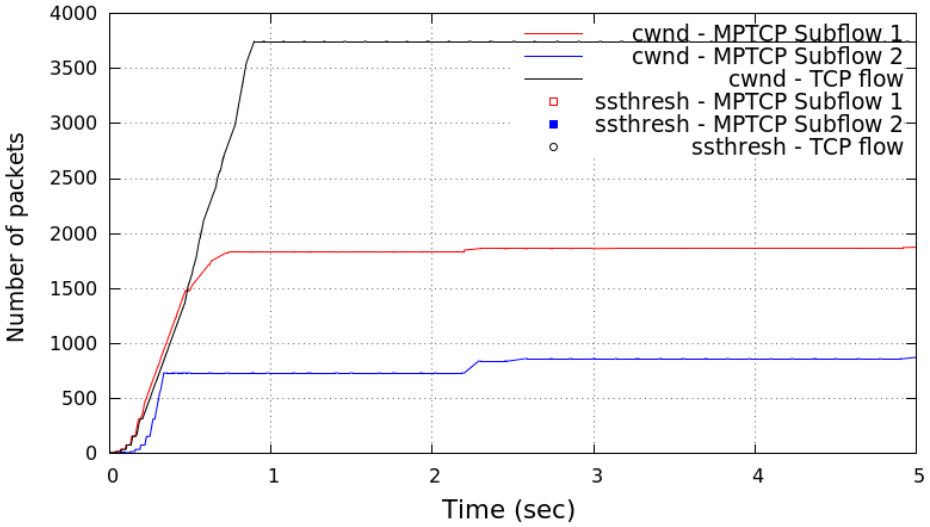


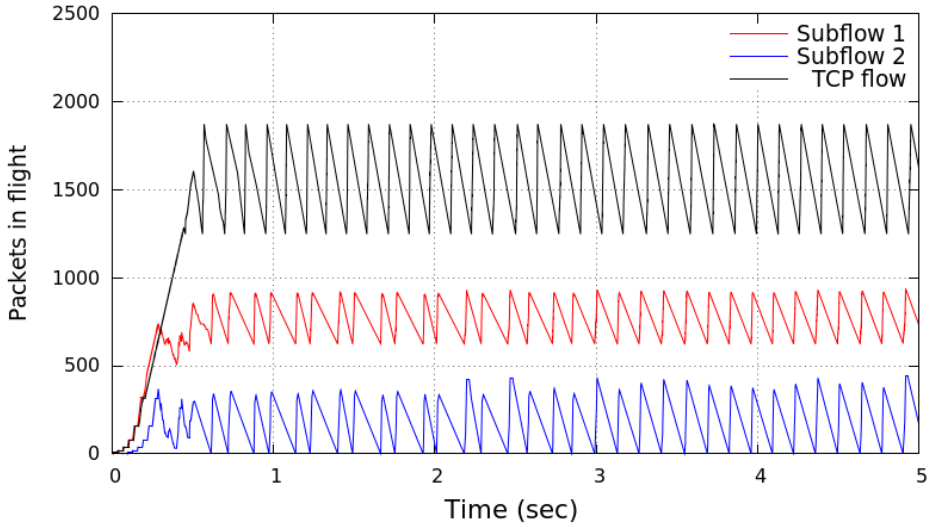


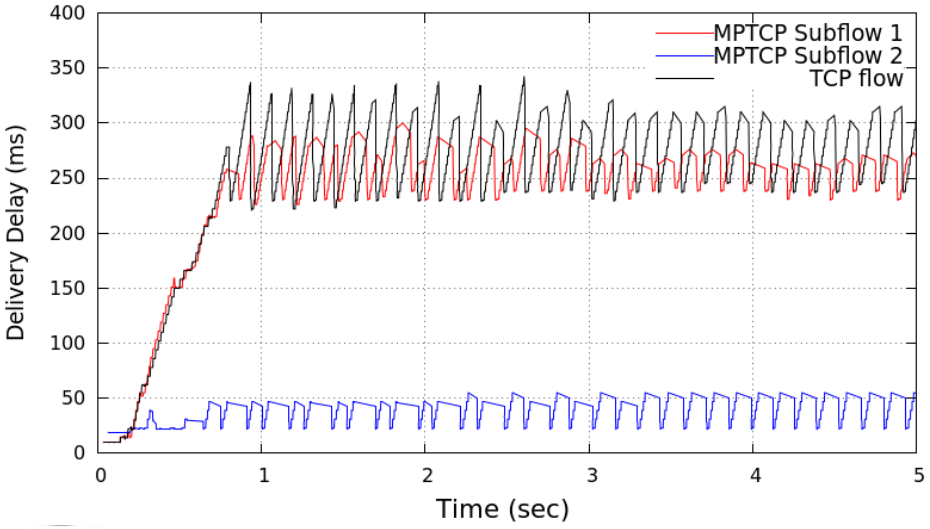


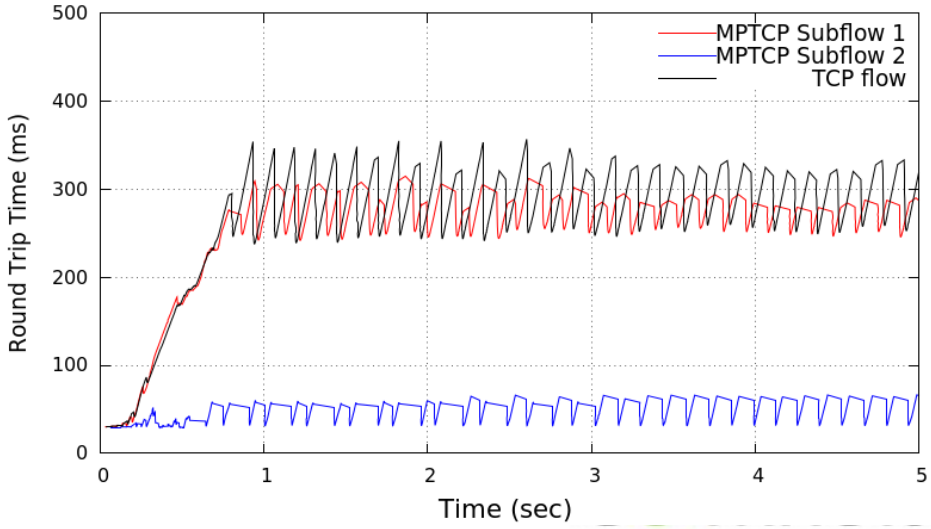


Round Robin Scheduler

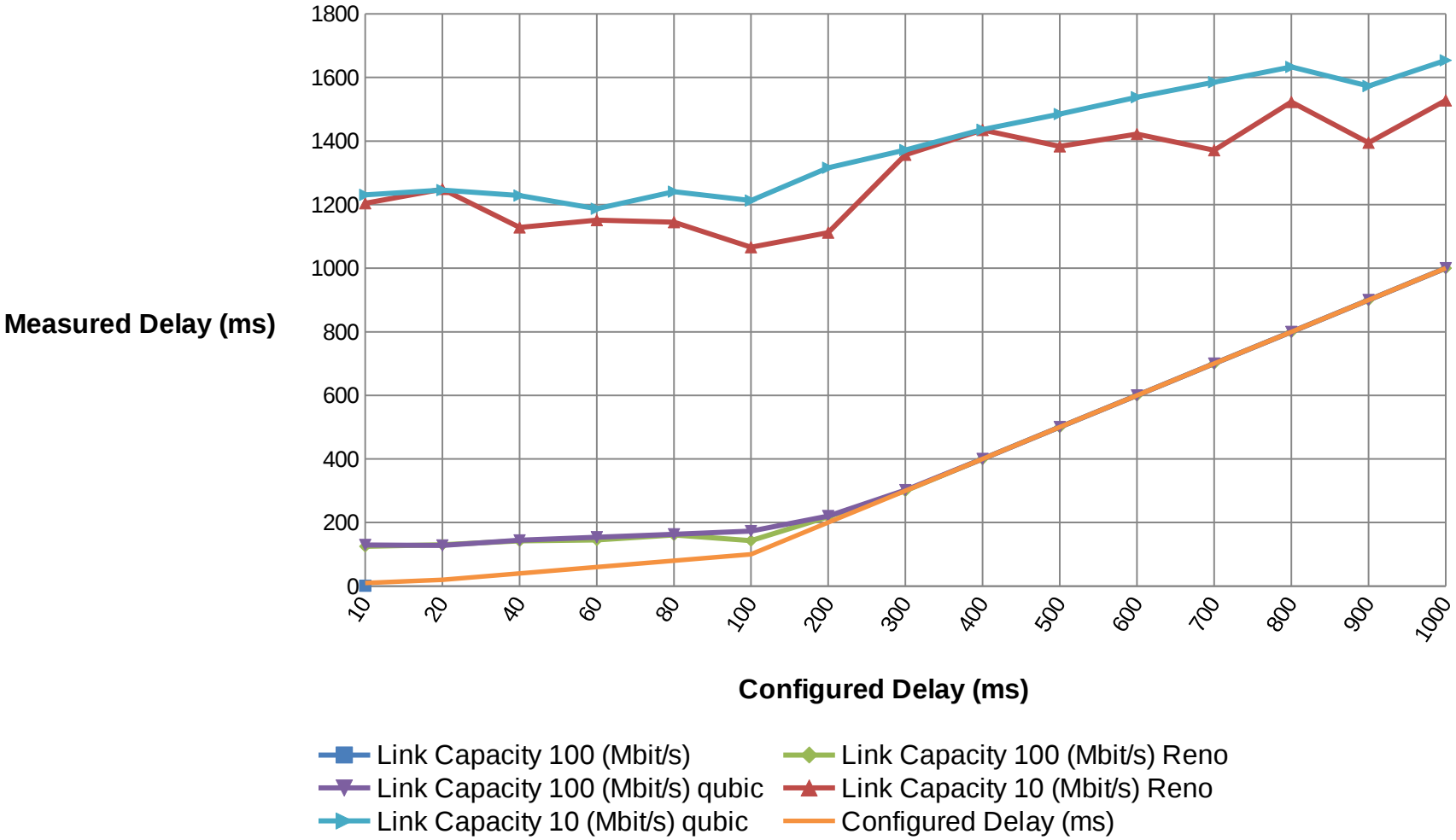




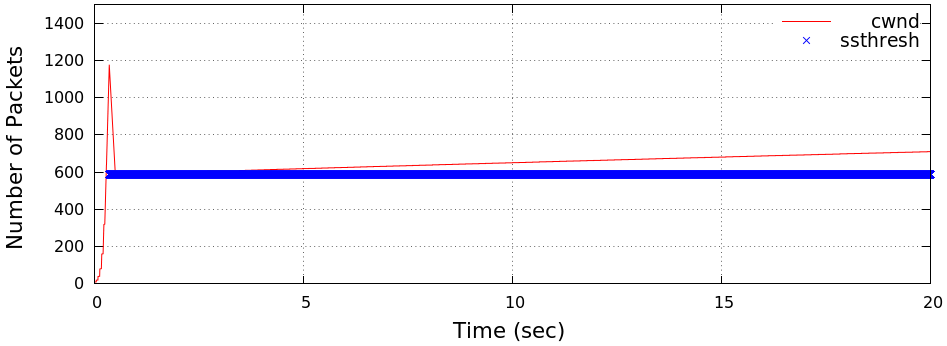


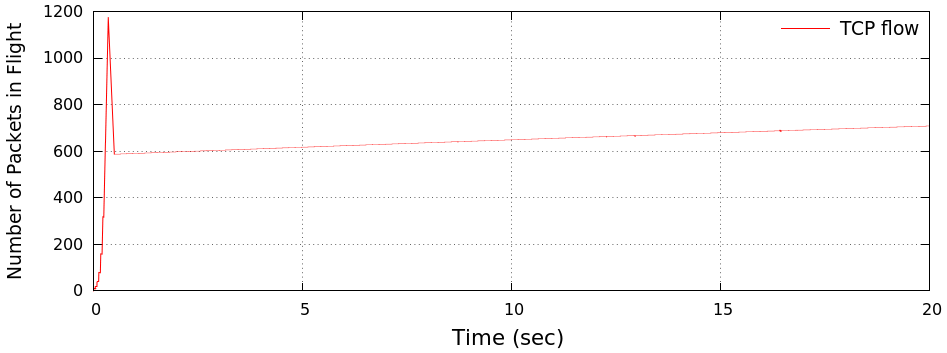


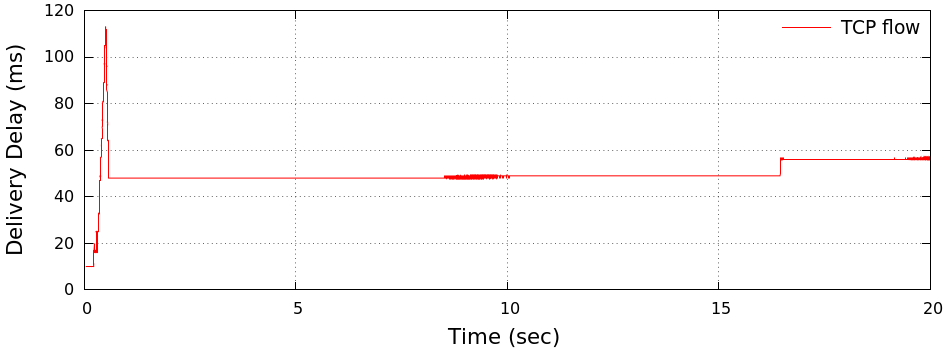
TCP – Measured Delay Vs Configured Delay

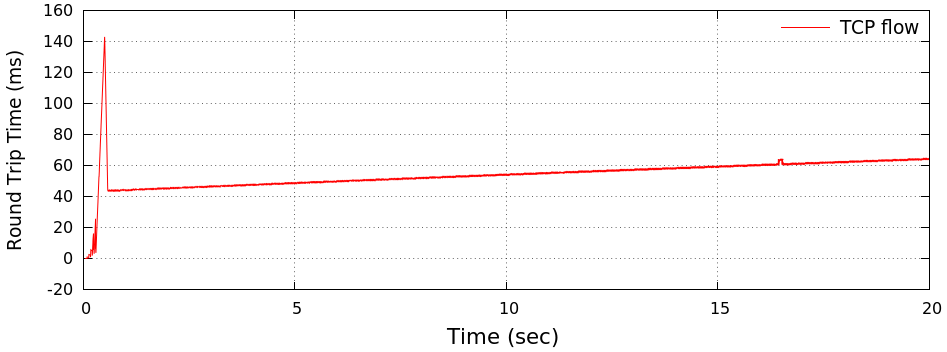


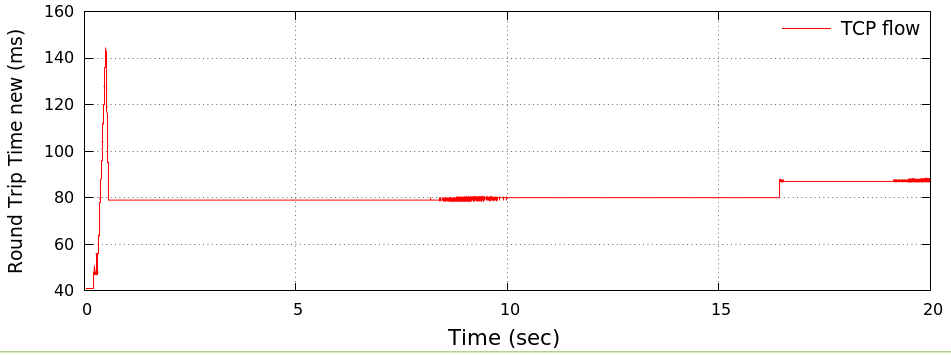
TCP – Link Capacity 100 (Mbit/sec), Delay - 40ms (20ms + 20ms)

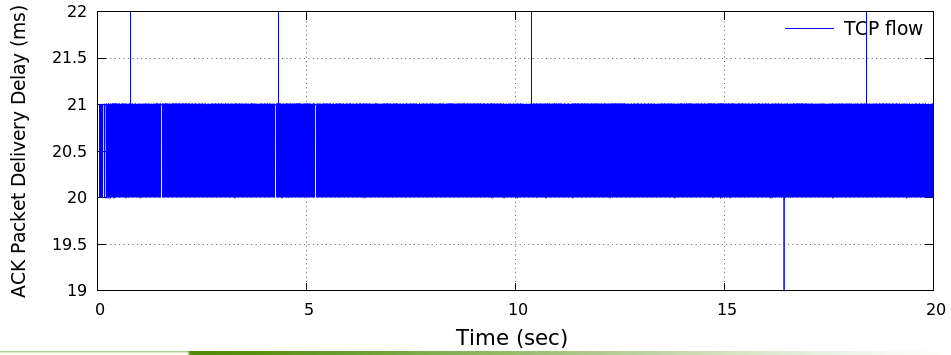




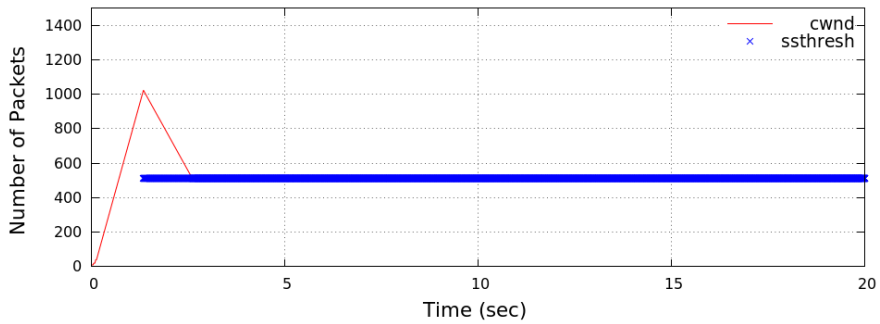


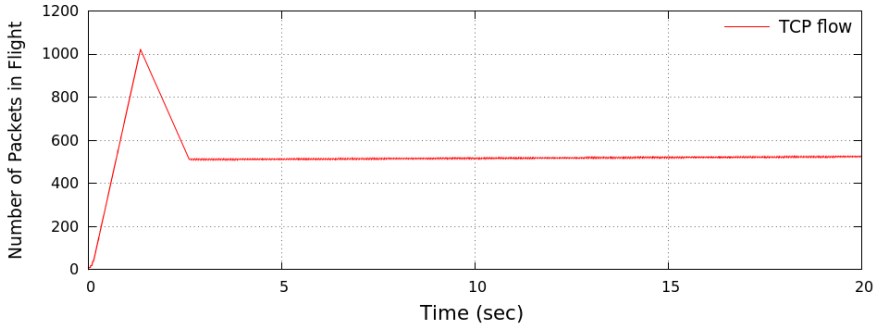


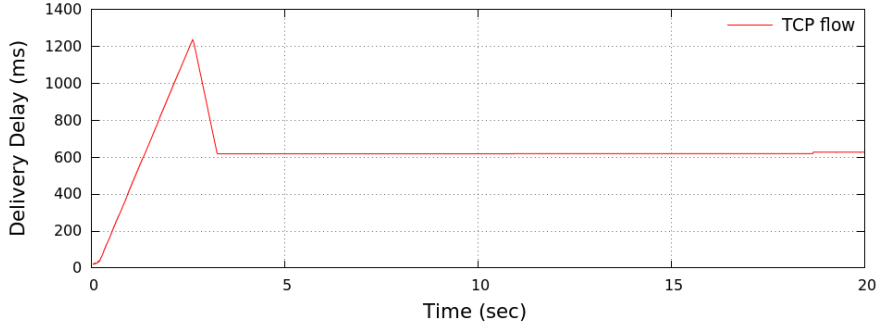


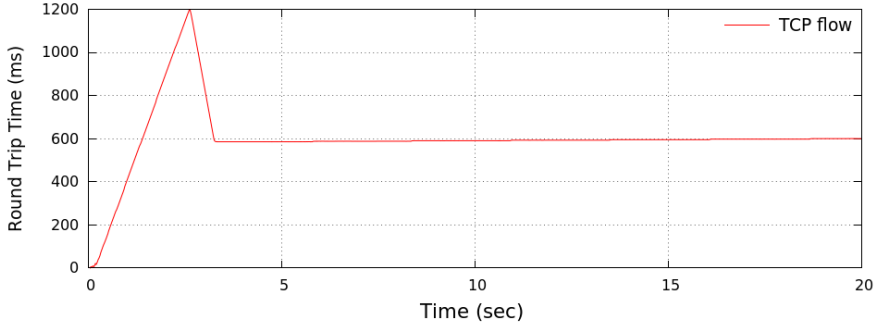


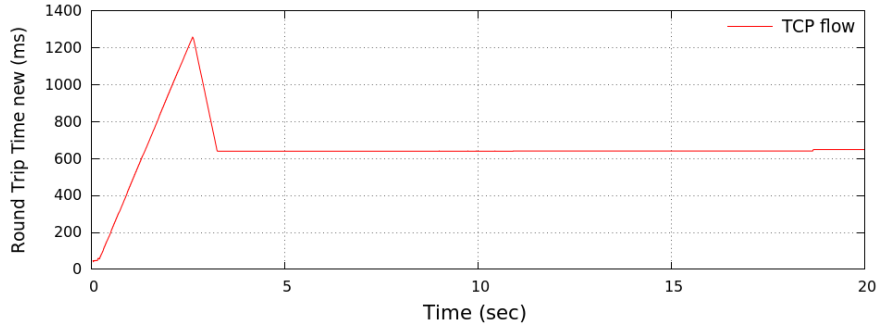
TCP – Link Capacity 10 (Mbit/sec), Delay - 40ms (20ms + 20ms)

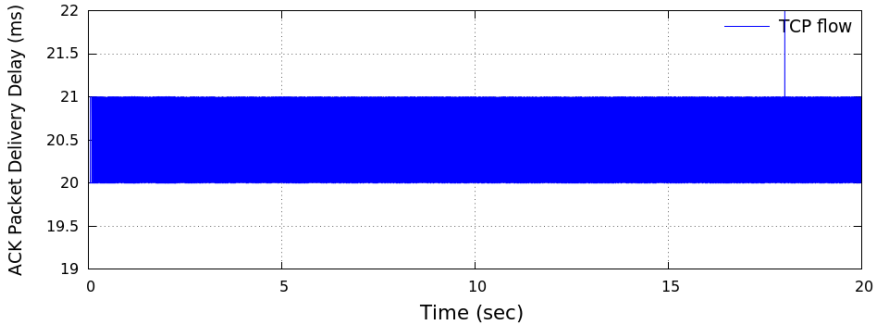






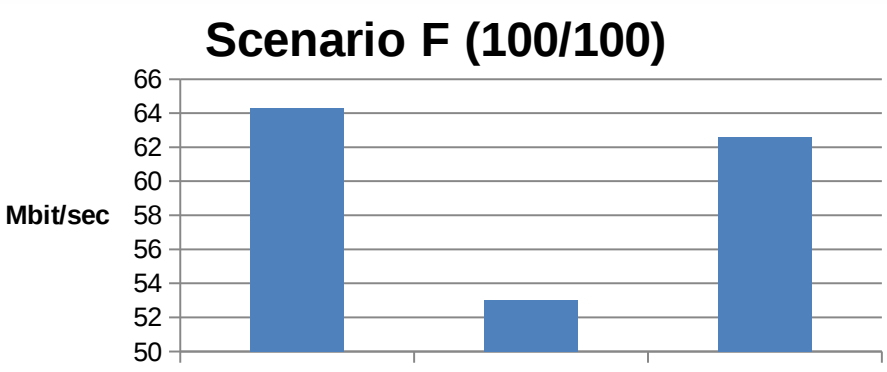


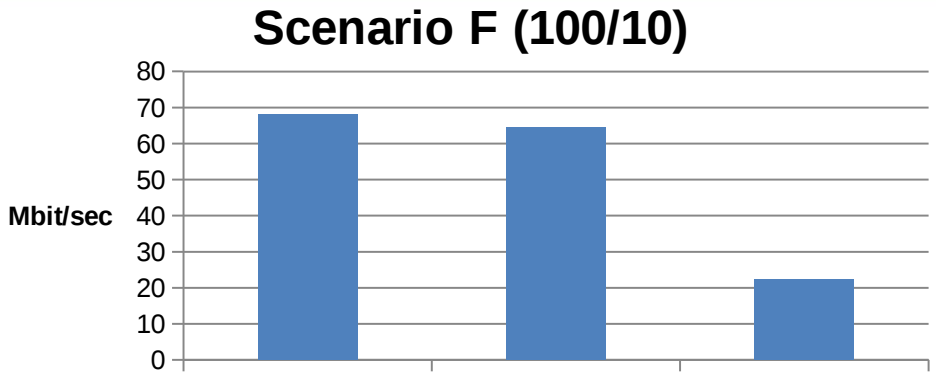


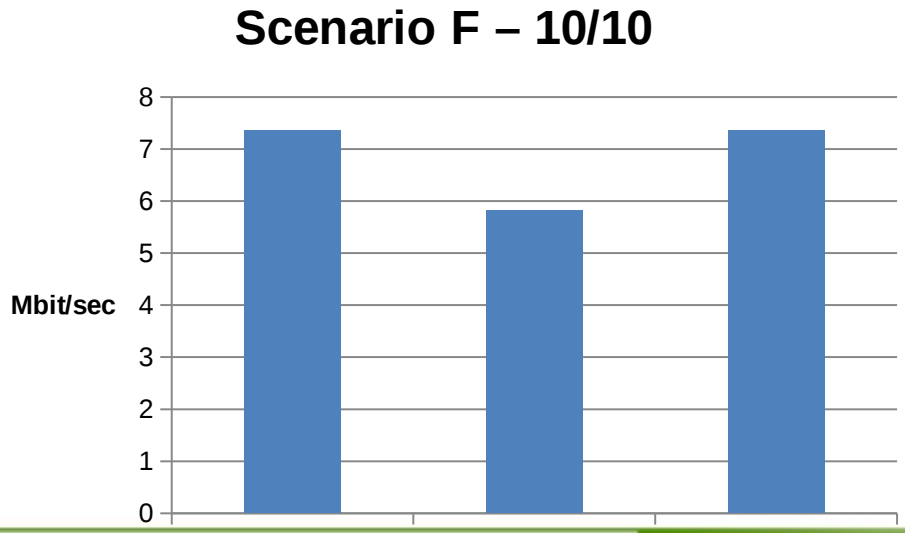


Simple Testbed
Link Capacity – 100 / 100 (Mbit/sec)
Scenario F

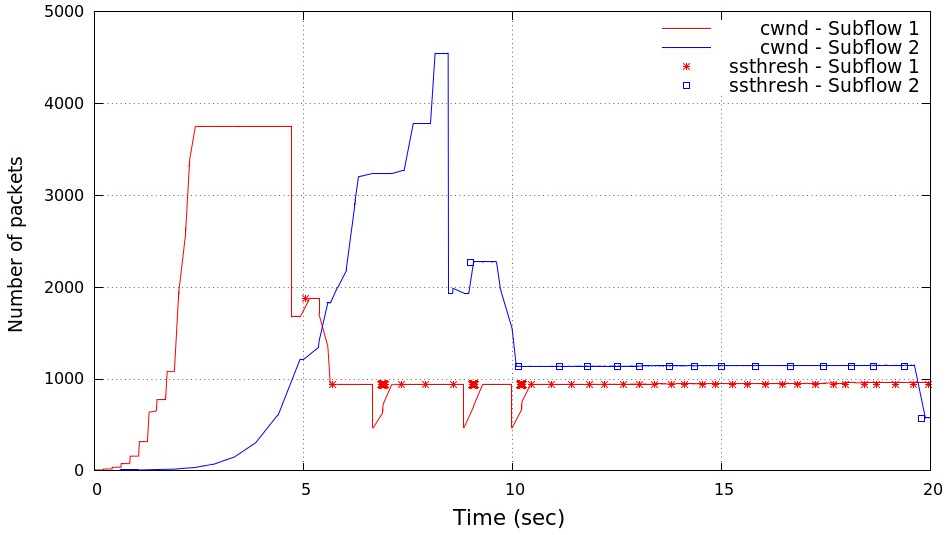
Hypothetical Scenario F

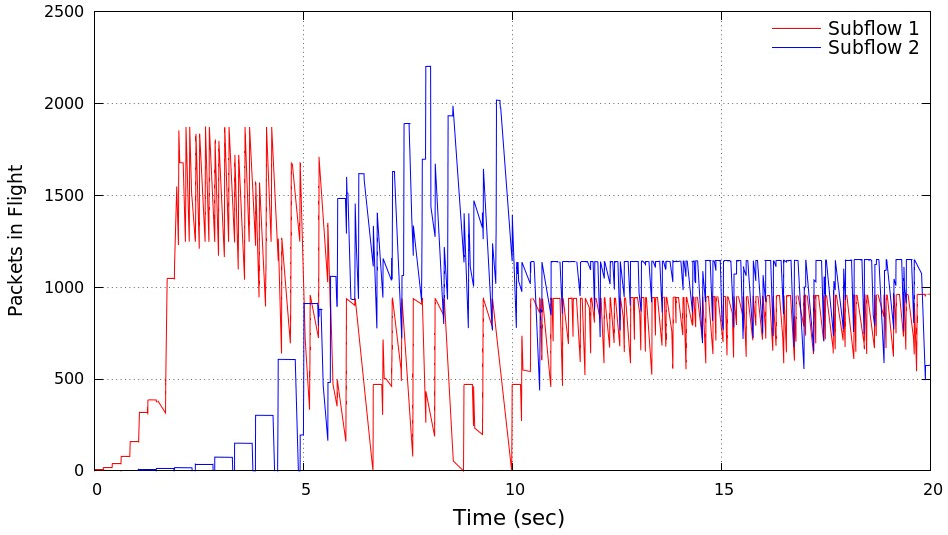


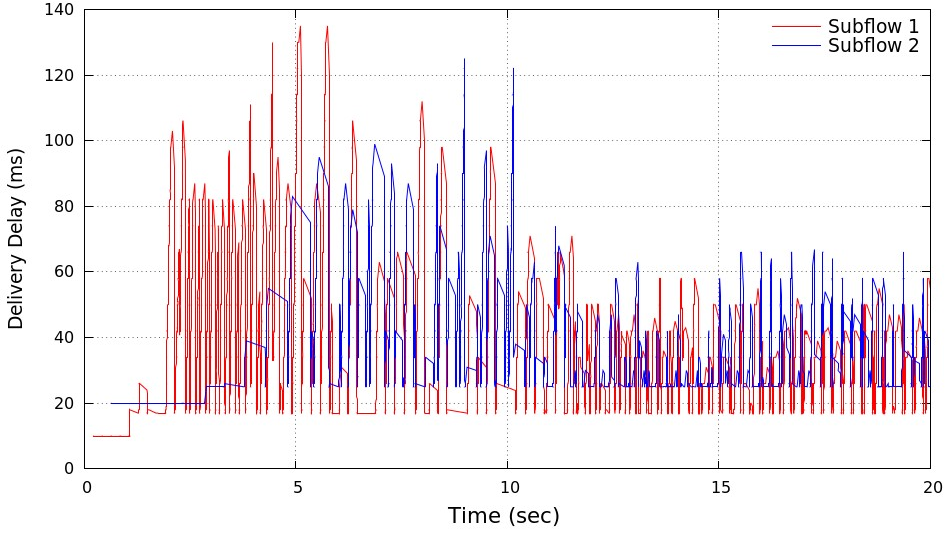


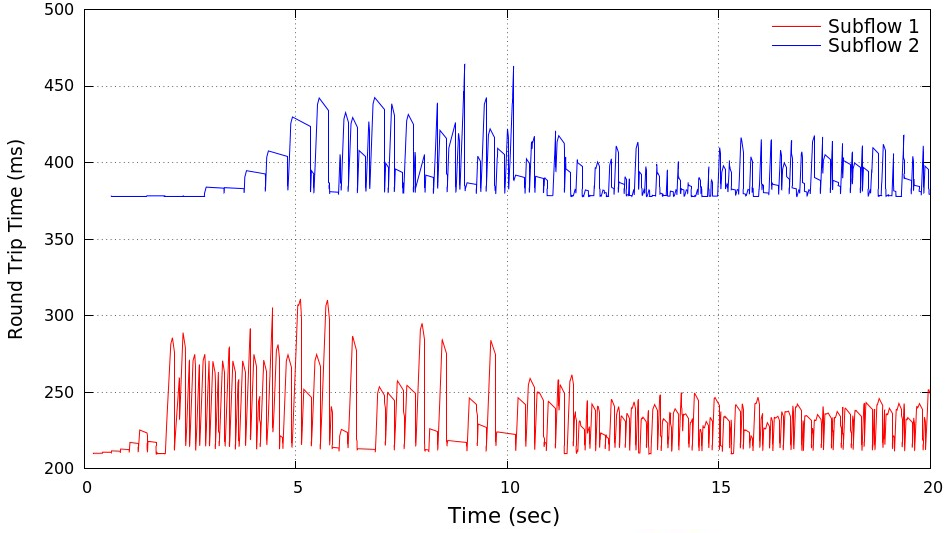


DD

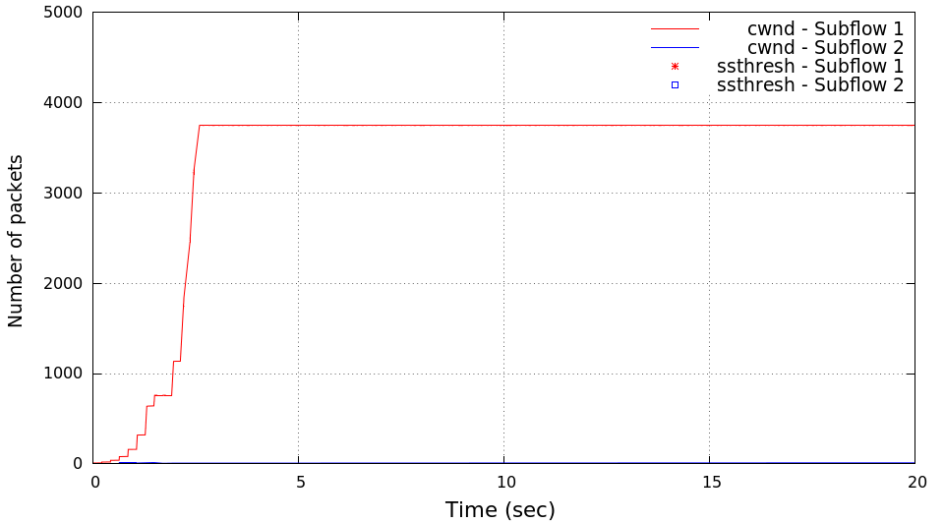

cwnd - Subflow 1
cwnd - Subflow 2
ssthresh - Subflow 1
ssthresh - Subflow 2
Number of packets
Time (sec)
0
1000
2000
3000
4000
5000
0
5
10
15
20

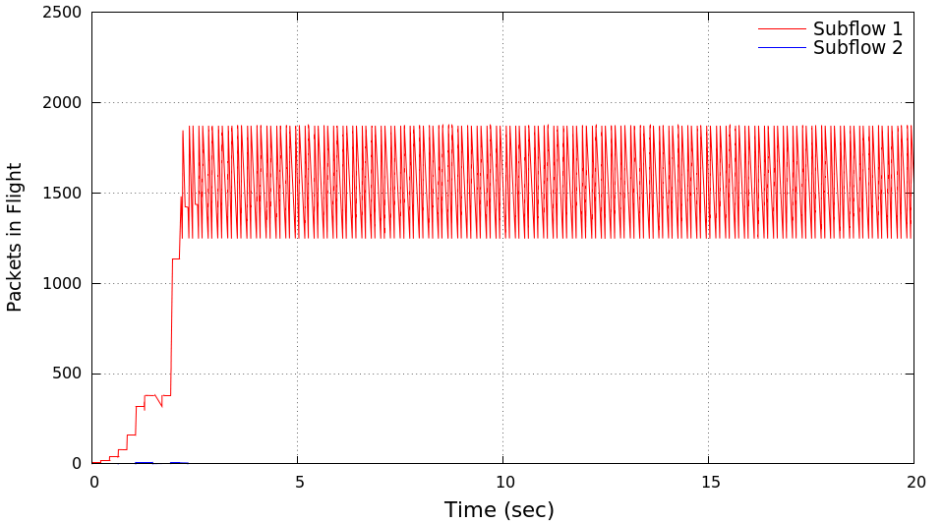

Subflow 1
Subflow 2
Packets in Flight
Time (sec)
0
500
1000
1500
2000
2500
0
5
10
15
20

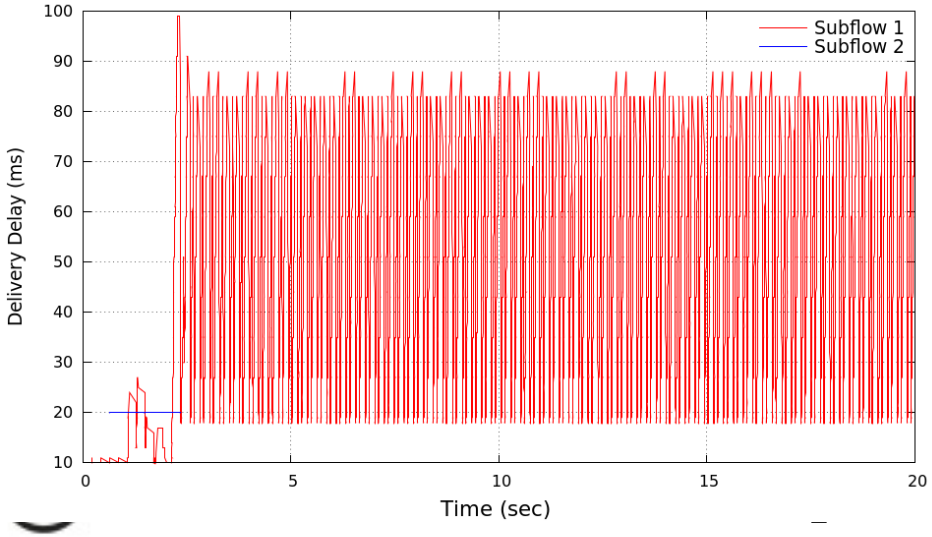

Subflow 1
Subflow 2
Delivery Delay (ms)
Time (sec)
0
20
40
60
80
100
120
140
0
5
10
15
20

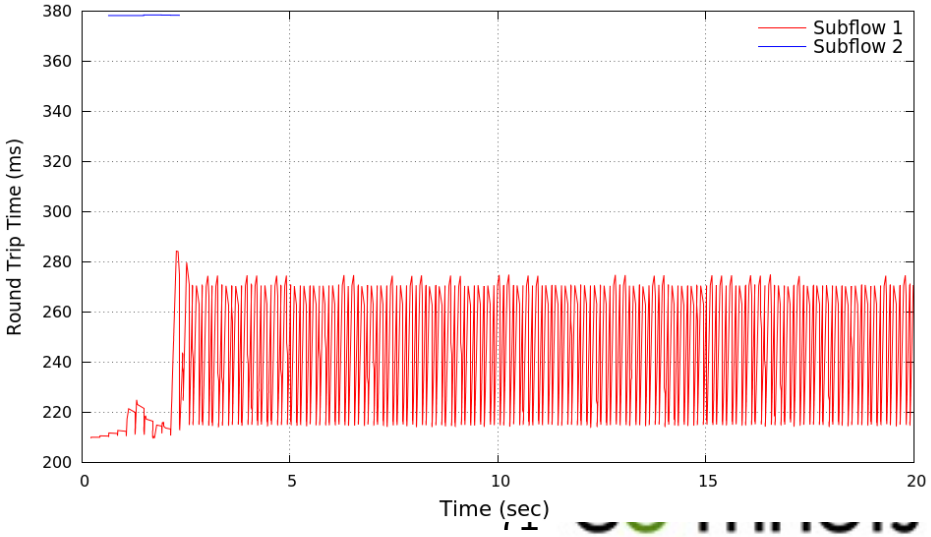

Subflow 1
Subflow 2
Round Trip Time (ms)
Time (sec)
200
250
300
350
400
450
500
0
5
10
15
20

sRTT

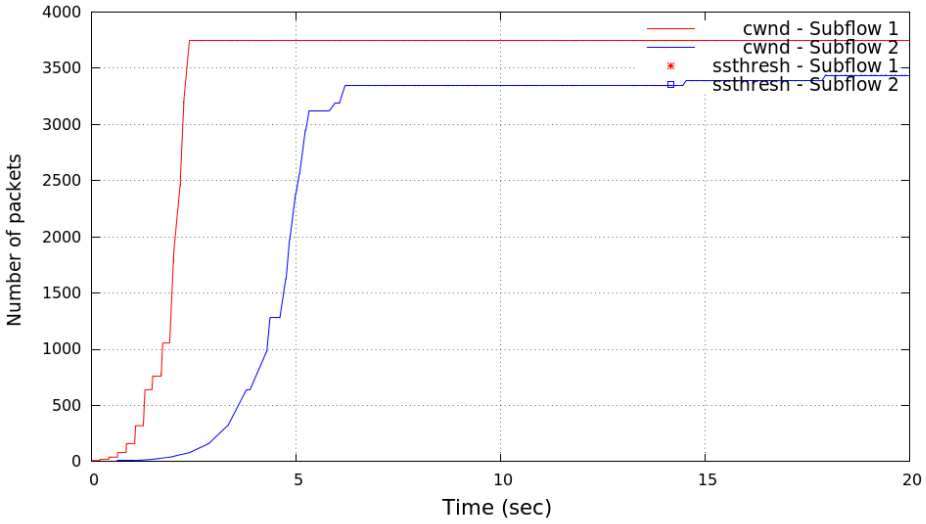

cwnd - Subflow 1
cwnd - Subflow 2
ssthresh - Subflow 1
ssthresh - Subflow 2
Number of packets
Time (sec)
0
1000
2000
3000
4000
5000
0
5
10
15
20

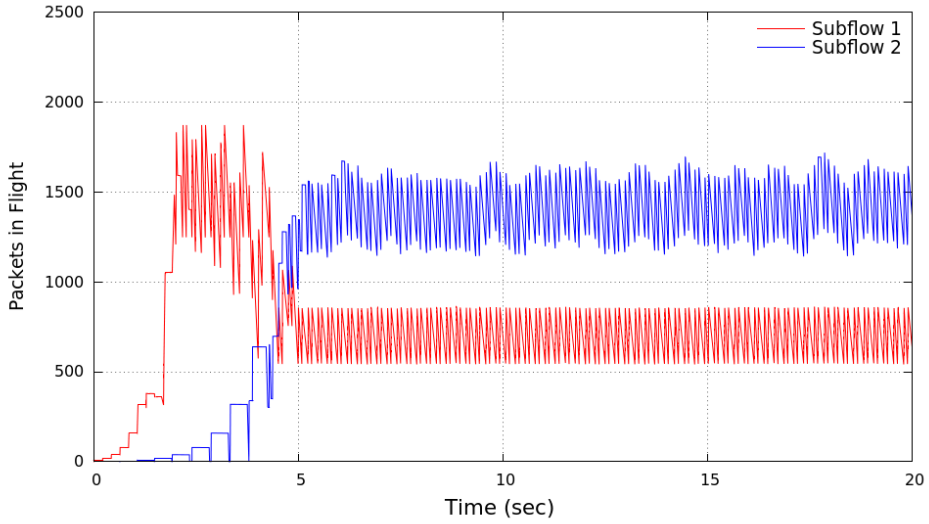

Subflow 1
Subflow 2
Packets in Flight
Time (sec)
0
500
1000
1500
2000
2500
0
5
10
15
20

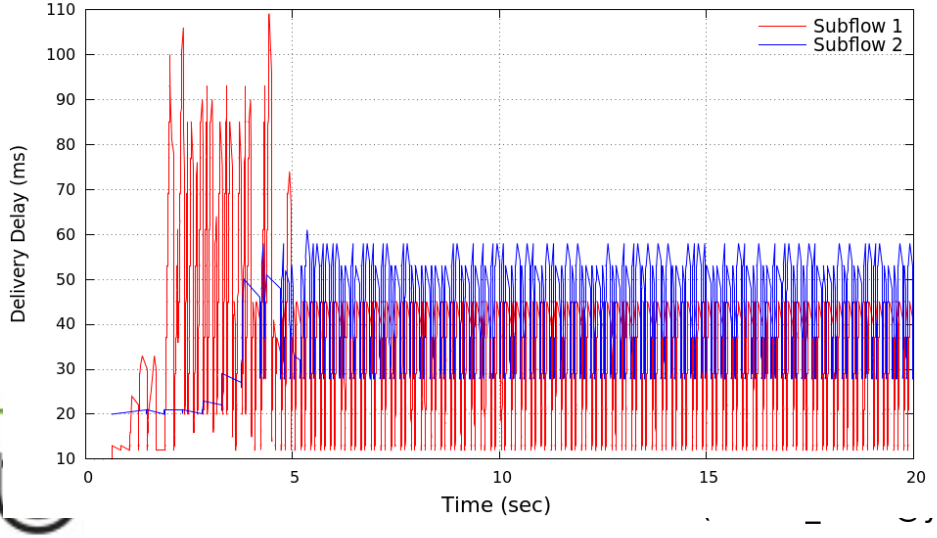

Subflow 1
Subflow 2
Delivery Delay (ms)
Time (sec)
10
20
30
40
50
60
70
80
90
100
0
5
10
15
20

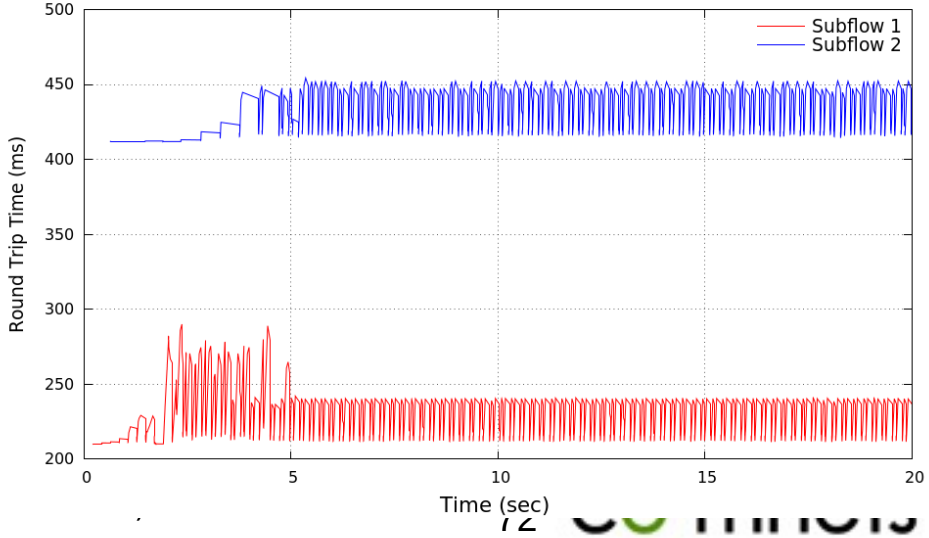

Subflow 1
Subflow 2
Round Trip Time (ms)
Time (sec)
200
220
240
260
280
300
320
340
360
380
0
5
10
15
20

RR


cwnd - Subflow 1
cwnd - Subflow 2
ssthresh - Subflow 1
ssthresh - Subflow 2
Number of packets
Time (sec)
0
500
1000
1500
2000
2500
3000
3500
4000
0
5
10
15
20


Subflow 1
Subflow 2
Packets in Flight
Time (sec)
0
500
1000
1500
2000
2500
0
5
10
15
20


Subflow 1
Subflow 2
Delivery Delay (ms)
Time (sec)
10
20
30
40
50
60
70
80
90
100
110
0
5
10
15
20


Subflow 1
Subflow 2
Round Trip Time (ms)
Time (sec)
200
250
300
350
400
450
500
0
5
10
15
20

Outlook (DD Scheduler)

Packet no	Reordering Delay (ms)	End-to-end Delay (ms)
1	0	50
2	0	50
3	0	50
4	0	100
5	0	100
6	0	100
7	0	100
8	0	100
9	40	50 + 40
10	30	50 + 30
11	20	50 + 20
12	0	50