Understanding the Performance of TCP BBRv2 using FABRIC

Jose Gomez, Elie Kfoury, Jorge Crichigno

Integrated Information Technology Department, University of South Carolina, Columbia, South Carolina

Abstract

- This work presents a performance evaluation of the Bottleneck Bandwidth and Round-trip Time version 2 (BBRv2)
- The experiments are conducted in FABRIC, a national-scale experimental network infrastructure
- Evaluations conducted on FABRIC to reproduce the network conditions of Wide Area Networks (WANs)
- The tests presented in this paper evaluate:
  - ✓ The throughput as a function of the Round-trip Time
  - ✓ The RTT unfairness of BBRv1 and BBRv2
  - ✓ The queue occupancy,
  - ✓ The packet loss rate as a function of the router's buffer size
- This work also presents and discusses the influence of Active Queue Management (AQM) algorithms

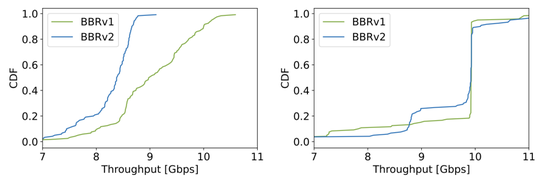
Related Work

- Kfoury et al. [1] used Mininet to conduct a performance evaluation of BBRv2
- Gomez et al. [2] executed additional evaluation tests on BBRv2, but they also used Mininet
- Tierney et al. [3] described and performed experiments to assess the suitability of BBRv2 for use on Data Transfer Nodes (DTNs)
- Scherrer et al. [4] presented a fluid model of BBRv1 and BBRv2 to complement the previous studies

[1] Kfoury et al., "An emulation-based evaluation of TCP BBRv2 alpha for wired broadband," Computer Communications, 2020.
[2] Gomez et al., "A performance evaluation of TCP BBRv2 alpha," in 2020 TSP.
[3] Tierney et al., "Exploring the BBRv2 congestion control algorithm for use on data transfer nodes," in 2021 INDIS.
[4] S. Scherrer et al., "Model-based insights on the performance, fairness, and stability of BBR," in IMC 2022

Motivation

- Understanding the behavior of BBRv2 in a testbed with real propagation delay
- Observing the dynamics of BBRv2 in a Wide Area Network (WAN)
- Analyzing the differences between an emulated environment and a real testbed
- This work leverage the distributed architecture of the FABRIC testbed to reproduce WAN conditions and test the performance of BBRv2



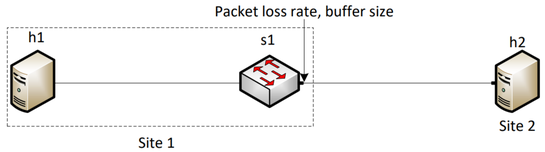
(a) (b)

- CDF of the bottleneck bandwidth estimation of BBRv1 and BBRv2. (a) with 45ms emulated delay. (b) with 45ms propagation delay.

Experimental Setup

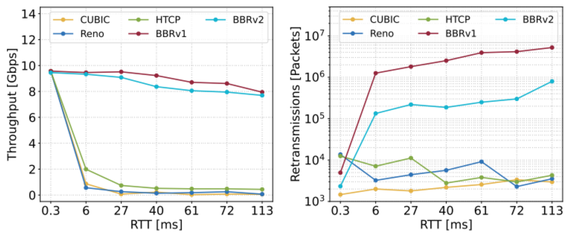
- Software switch to limit the rate and emulate packet losses
- Rate is limited using the Token Bucket Filter (TBF) in Linux
- Packet losses are emulated using NETEM
- The sites are selected based on the experiment

Site 1	Site 2	RTT
TACC (TX)	TACC (TX)	0.3ms
DALL (TX)	TACC (TX)	6ms
DALL (TX)	WASH (DC)	27ms
SALT (UT)	FIU (FL)	44ms
GPN (MO)	DALL (TX)	61ms
UTAH (UT)	WASH (DC)	72ms
GPN (MO)	FIU (FL)	113ms



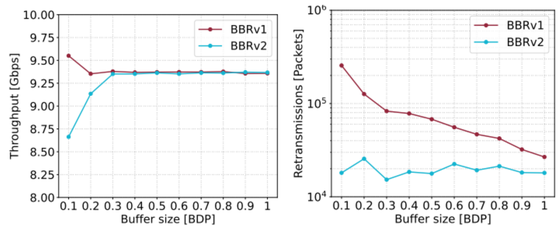
Performance in a WAN with Packet Losses

- Experiment 1: Performance in a WAN with packet losses
- The rate is limited to 10Gbps
- The emulated packet loss rate is 0.0046% (i.e., 1/22,000)
- Cubic, Reno, HTCP: Loss-based congestion control algorithms
- BBRv1, BBRv2: model-based congestion control algorithms
- Throughput with loss-based algorithms degrades when the RTT is high and there are packet losses



Retransmissions with Various Buffers

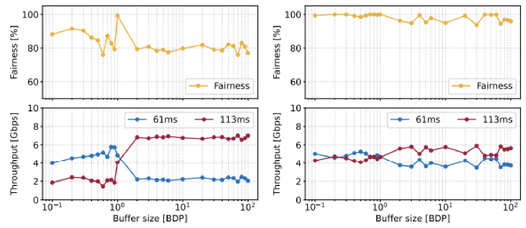
- Experiment 2: Retransmissions as a function of the buffer size
- The RTT between the hosts is 45 milliseconds (SALT, FIU)
- BBRv2 presents lower packet loss rates than BBRv1



(a) (b)

RTT Unfairness

- Experiment 3: RTT unfairness
- RTT unfairness occurs when flows with smaller RTTs obtain a higher throughput
- BBRv1 flows present the opposite behavior
- BBRv2 reduces the RTT unfairness of competing flows
- DALL – FIU: 61ms, GPN – FIU: 113ms



(a) BBRv1 with Tail Drop (b) BBRv2 with Tail Drop

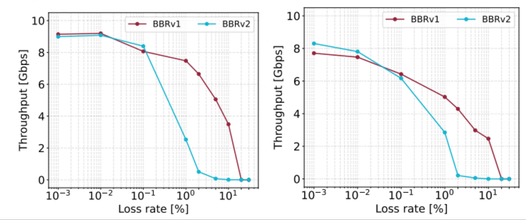
Parallel Streams and Different MTUs

- Experiment 4: Parallel streams and different MTUs
- The rate is not limited (i.e., 100Gbps)
- The RTT between the sites is 26 milliseconds (DALL, SALT)
- BBRv1 and BBRv2 achieve throughputs over 70Gbps with eight streams

Streams	CUBIC		Reno		BBRv1		BBRv2	
	1500	9000	1500	9000	1500	9000	1500	9000
1	0.624	2.5	0.542	2.09	10.2	15.4	9.35	17.7
2	1.36	3.9	1.42	3.35	21.9	33.3	18.7	31.7
4	3.58	7.84	3.13	5.31	35.5	37.2	28.9	42.1
8	9.77	9.93	8.05	9.21	32.7	70.1	40.6	73.8
16	11	14.5	8.88	13.5	41.2	86	47.6	77.8
32	11.9	18.8	11.8	17.5	47	71.3	49.5	78.3
64	12.8	66.9	12.6	22.8	44.8	79.4	44.9	80.3
120	17.2	75.7	16.1	68.2	44	67.9	43.1	77.2

Throughput with Various Packet Loss Rates

- Experiment 5: Throughput as a function of packet losses
- The performance of BBRv2 is close to that of BBRv1 for loss rates less than 1%
- (left) RTT = 26ms (DALL, SALT)
- (right) RTT = 57ms (UCSD, UMASS)

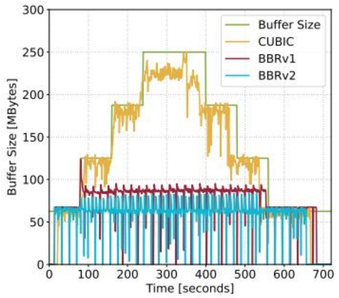


Acknowledgement

- This work was supported by the U.S. National Science Foundation, Office of Advanced Cyberinfrastructure, award #2118311

Results: Queue Occupancy

- Experiment 6: Queue occupancy
- Link is limited to 10Gbps, the RTT is 50ms
- Bandwidth-delay Product = 10Gbps * 50ms = 62.5MB
- BBRv2 keeps the queue occupancy around BDP, even with bloated buffers



Limitations

- Configuration of the intermediary devices (e.g., routers and switches)
  - ✓ Link capacity
  - ✓ Router buffer size
  - ✓ Queue allocation
- The experiments modified the buffer size of a software switch
- Shared Network Interface Cards (NICs)
- Performance isolation

Lessons Learned

- FABRIC can be used to test protocols and applications under WAN conditions
- The testbed can support a wide variety of experiments
- Its programmable infrastructure allows defining customized network environments
- BBRv2 provides improved fairness compared to BBRv1, particularly when dealing with flows that have different RTTs
- BBRv2 can achieve comparable throughput to BBRv1, while also exhibiting a lower retransmission rate

Conclusion and Future Work

- This project evaluated the performance of BBRv2 using FABRIC.
- Results show that BBRv2 performs similarly to its predecessor, BBRv1, presenting a lower retransmission rate.
- BBRv1 and BBRv2 achieve higher throughput with various RTTs than CUBIC, Reno, and HTCP.
- Future works can evaluate the performance metrics using a P4-programmable switch
- P4 switches perform fine-grained measurements, providing better visibility on the dynamics of the congestion control algorithm