Driving TCP Congestion Control Algorithms on Highway

Jae Chung, Feng Li, and Xiaoxiao Jiang
Vz Labs

4/6/17

VERIZON LABS
Reimagining the Future
verizon

NetDev

Intro

The FASTEST Radio Access Network in the Market

Network Performance Challenges

- TCP is the major traffic source in the market
- Most TCP flows use AIMD-based Congestion Control Algorithm (CCA)
- AIMD-based CCA is not RAN friendly
  - AIMD does not effectively consume available bandwidth in LTE (4G) and 5G high-bandwidth high-delay RAN.
  - eNodeB vendors implement AQM to manage buffer resources.

Demand for PEP to buffer L4 packets and control TX rate on the RAN-side

- Fast small object download time
- Maximize goodput for large object transfers
- Maintain low self-inflicted RTT to avoid unnecessary drops by eNodeB AQM

Performance Enhanced Proxy (PEP)

Technical Challenges

- Fast time-to-market
- Fast adaptation to emerging technology
- Reduce software maintenance headache

Attractive Potential Solution

- Transparent PEP using
  - Open source TCP proxy
  - Linux TCP and networking stack
  - Existing / new / home-grown TCP Congestion Avoidance Module

Understanding TCP CCA Performance on LTE

No winner TCP Congestion Control Algorithm (CCA) for LTE

- Not very impressive LTE performance by TCP CUBIC, Westwood+ (low link utilization).
- Experimental TCP for wireless links implemented as UDP tunnels (e.g TCP Sprout, TCP Verus).
- New CCAs Designed for Data Centers (e.g., BBR, NV, etc).

Less Knowledge on CCAs' Performance on High Mobility

- No real measurement studies on High-Speed driving on LTE.
- No measurement studies to compare different CCAs performance.
- Difficult to model or simulate RF condition on highway.

Evaluation

Outline

- Methodology
- Radio Network Characteristics
- Compare CCAs' Performance
- Discussions
- Conclusion

Congestion Control Algorithms Compared

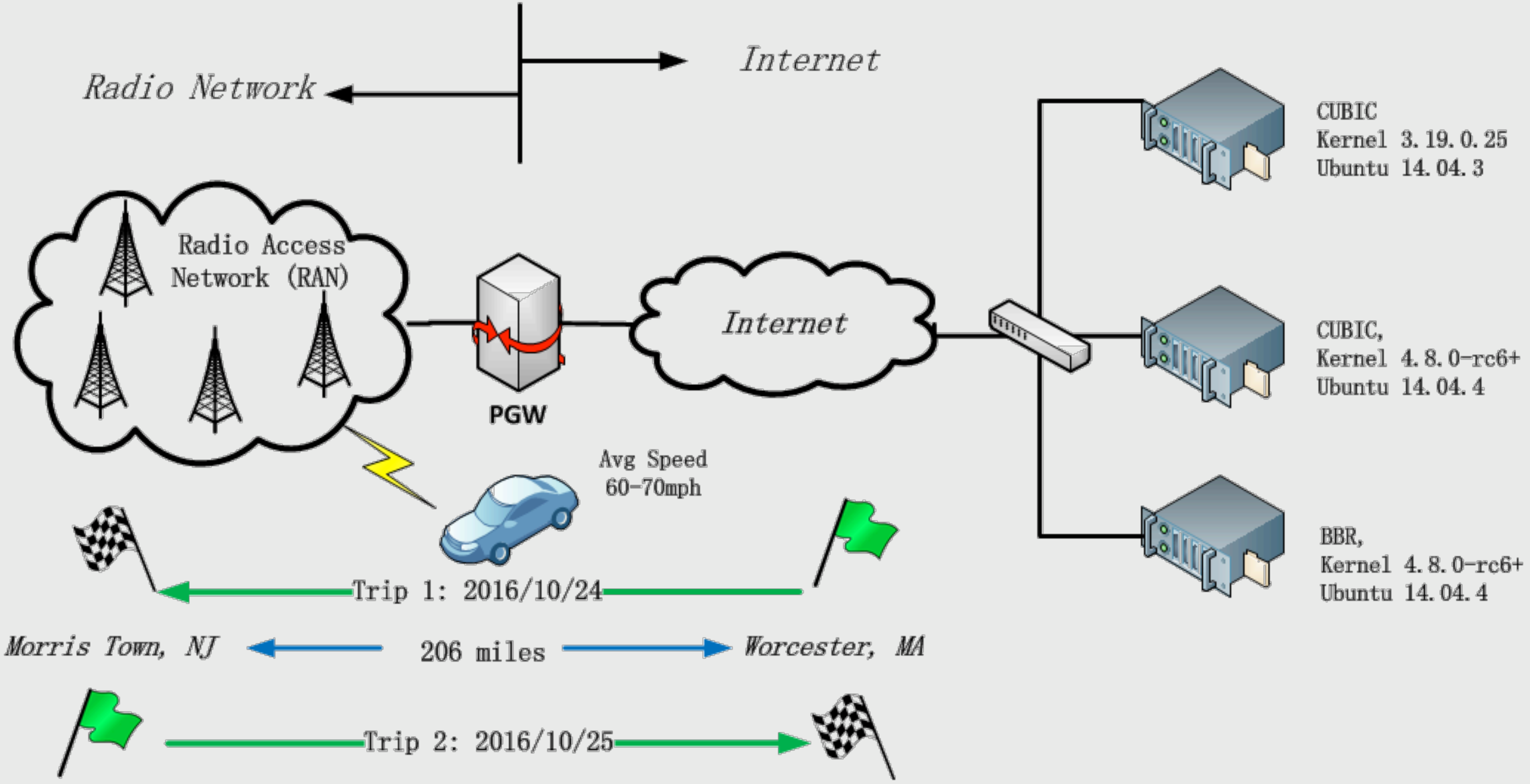
BBR (Bottleneck Bandwidth and Round trip propagation time).

- Developed by Google, originally for server to server communication.
- BBR was released with 4.8-rc6 kernel

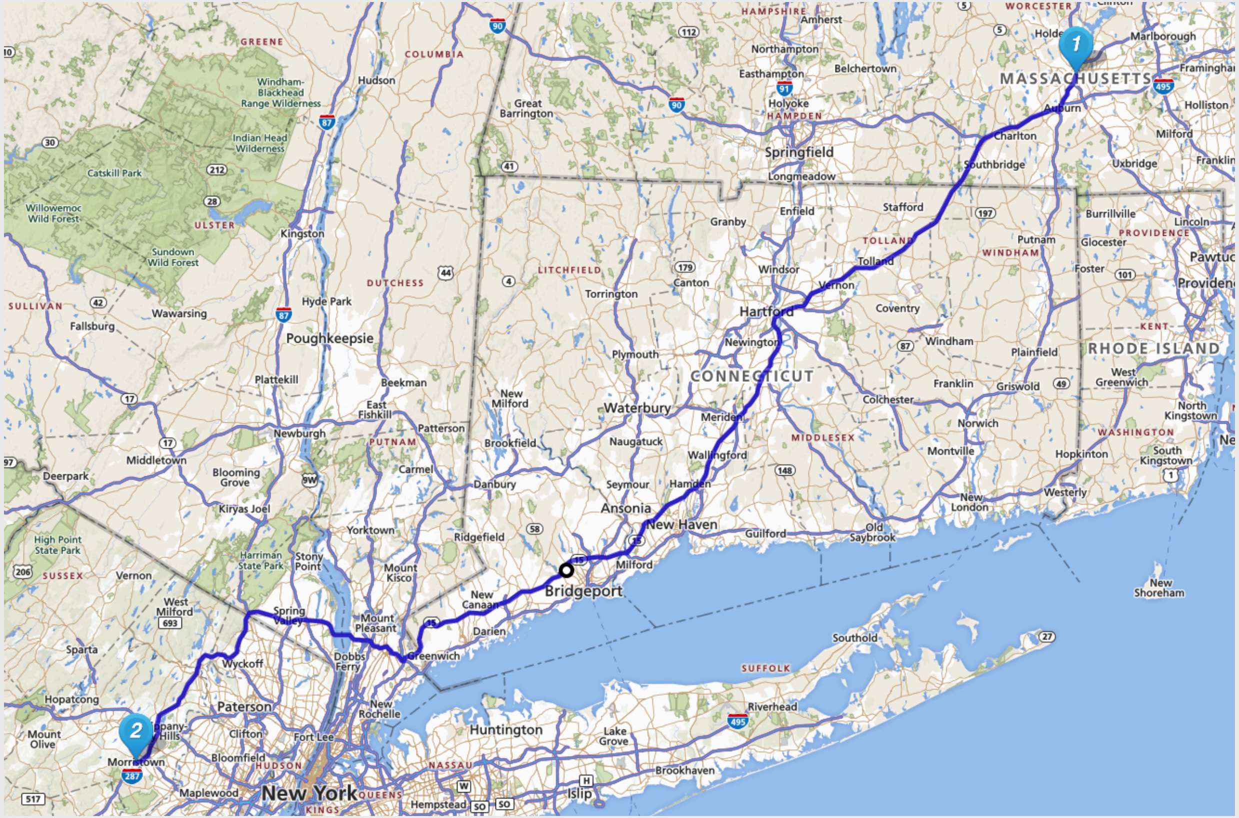
CUBICs

- The current default CCA in Linux
- Two servers running 4.8-rc6 and 3.19 kernels.
  – CUBIC in 4.8 introduces a patch to keep cwnd growth to cubic curve after "application limited" long idle time (bictcp_cwnd_event()).

Experimental Setup



Driving Route



- Date:
2016/10/24 and 2016/10/25

- End Points
Worcester, MA
Morris Town, NJ

- Distance
410 miles+ round trip,

- Data Volume
15.0+ GB traffic as 720 20MB file downloading in 6 hours.

some “large scale” research only collect 90GB traffic in 8 months.

Measurement Tools Used

Commercial Tool (Qualipoc) on smart phone (LG G2 VS980)

- Ping tool to measure propagation round trip time between server and phone.
- Throughput measurement tool.
- Physical and Link Layer statistics collected from device drivers.

Four HP Proliant 460c Gen9 blade Servers

- All run with Ubuntu 14.04: two with 4.8.0-rc6 kernel, and two with 3.19.0.25 kernel.
- Same kernel settings and Ethernet (NIC) settings, except default congestion control algorithm.
- Apache 2.4.7 Web server with PHP 5.0, dynamically generating file to avoid caching.
- Tcpdump running as a service in background,
- Dedicated performance study servers, light load (< 1% CPU usage).

700MHz Radio Spectrum

700MHz (Band XIII)

- Verizon provide 700MHz and 1700/1900MHz (AWS) radio spectrum.
- AWS only provide extra capacity in urban area.
- None of US carrier provides national wide AWS coverage.

Lock phone on 700MHz spectrum.

- Lost GPS location and velocity in test, could only estimate average speed through checkpoints.

Efforts to Reduce Random Variables

- Same route, Same Driver, Same Car
- Identical Servers, except default congestion control algorithm.

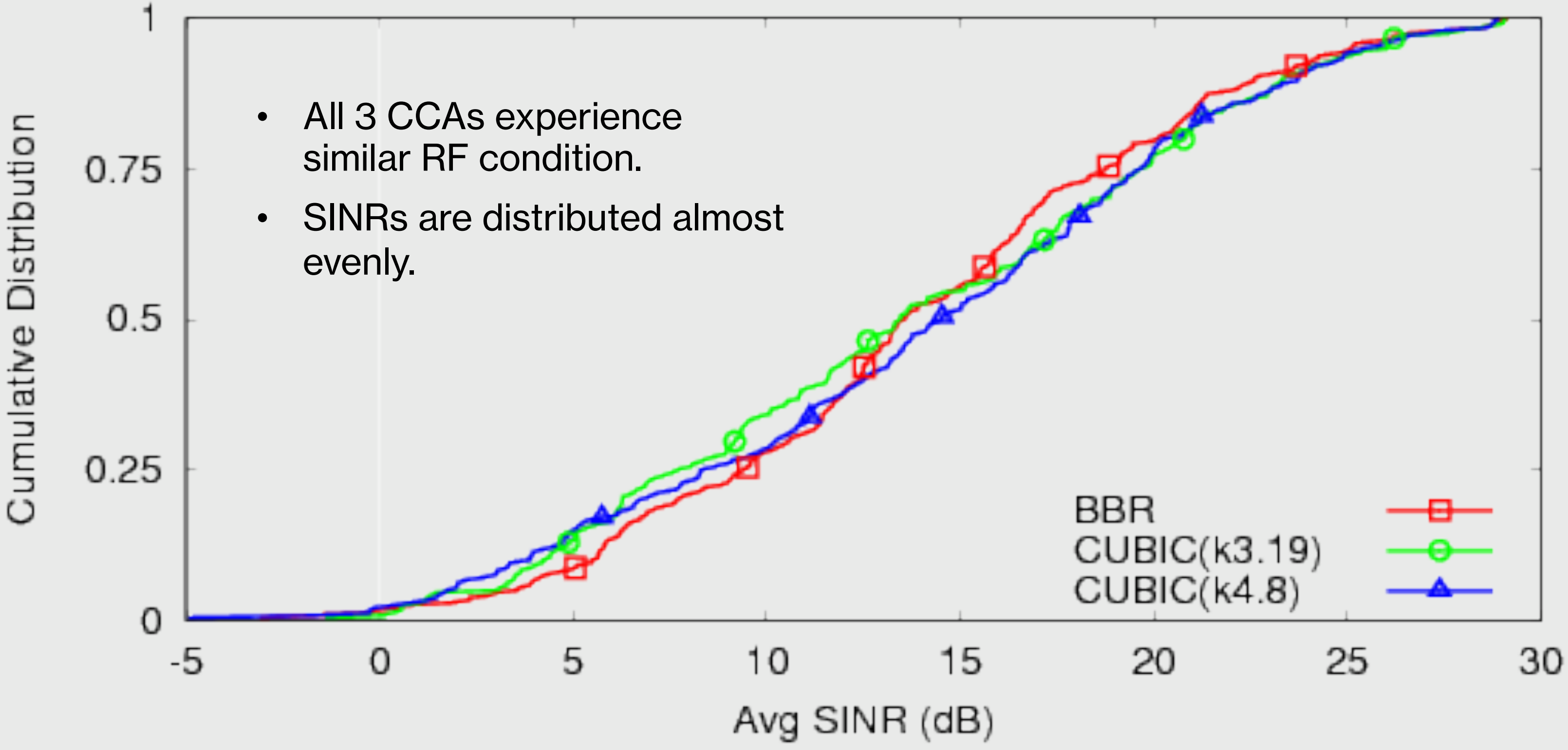
Band XIII Radio Spectrum

Metric	Value
Band Number	Band XIII (13)
UP Link Freq.	777-787 MHz
Down Link Freq.	746-750 MHz
Channel Width	10MHz
Modulation	**QPSK, 16QAM, 64QAM**
Theoretic TCP Throughput	45 – 50 Mbps (maximum)

Outline

- Methodology
- Radio Network Characteristics
- Compare CCAs’ Performance
- Discussions
- Conclusion

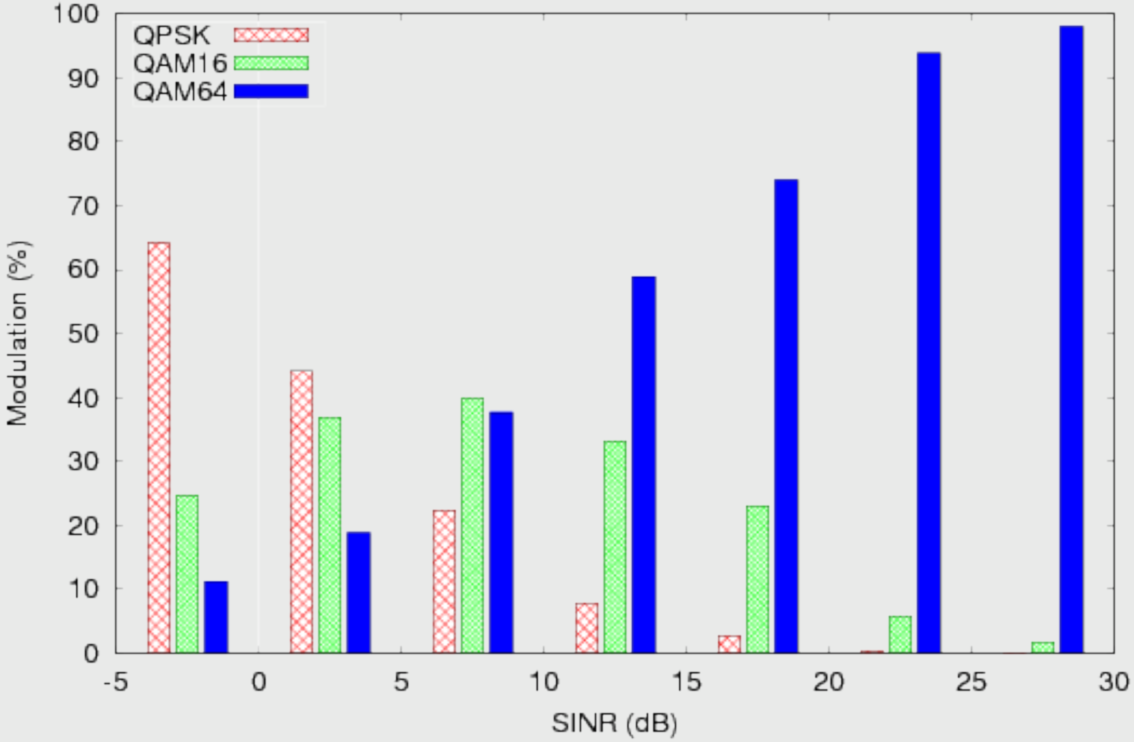
Radio Condition (SINR) on Highway



- All 3 CCAs experience similar RF condition.
- SINRs are distributed almost evenly.

Modulation / Rate Adaption

Fig. Modulation on Highway



Theoretical Max PHY Throughput	**10MHz**
QPSK	17 Mbps
16QAM	25 Mbps
64QAM	50 Mbps

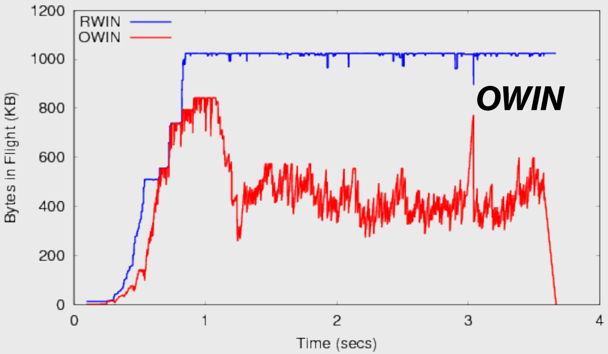
- Modulation/Rate Adaption changes would impact bandwidth estimation algorithm, for example BBR.
- Rate drop suddenly increase the RLP queuing layer delay that cause eNodeB AQM drops.

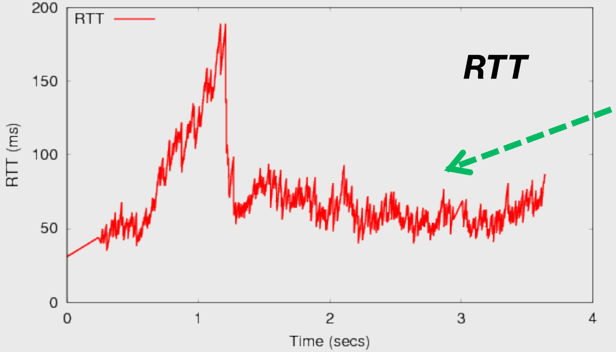
Outline

- Methodology
- Radio Network Characteristics
- Compare CCAs' Performance
- Discussions
- Conclusion

BBR Case Study

- **SINR is greater than 20dB**

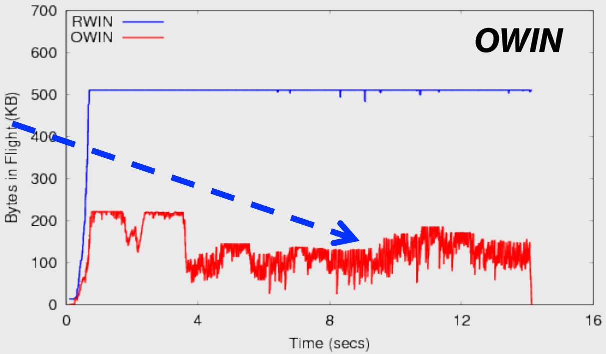


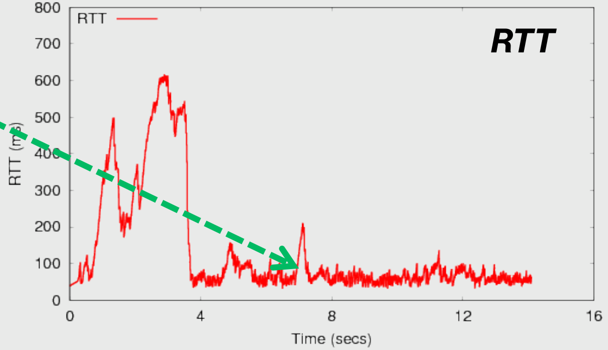


BBR chooses smaller CWND to control RTT as low.

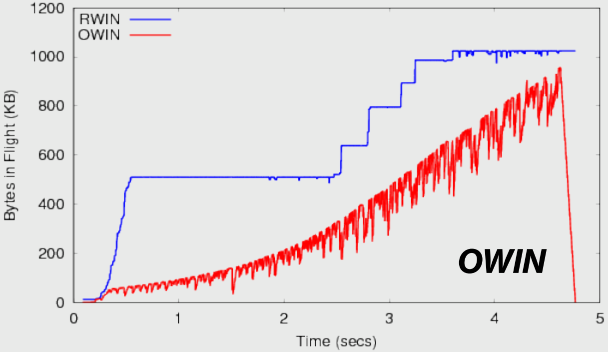
BBR attempts to keep a low RTT.

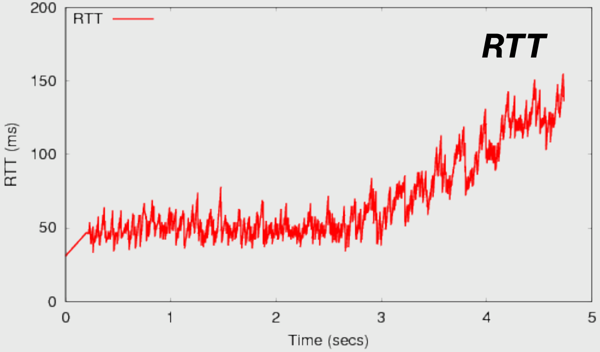
- **SINR is between 10 to 20dB**





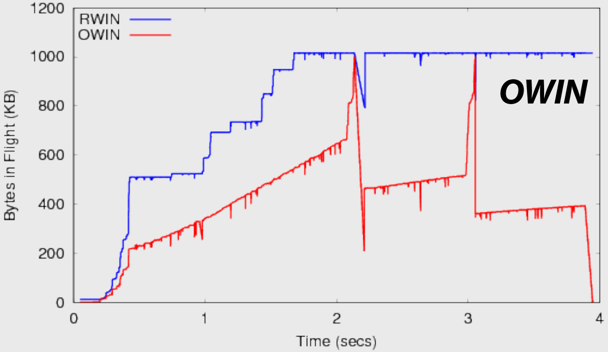
CUBIC(4.8) Case Study

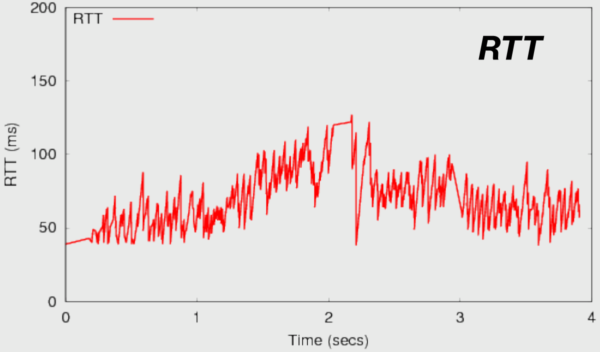




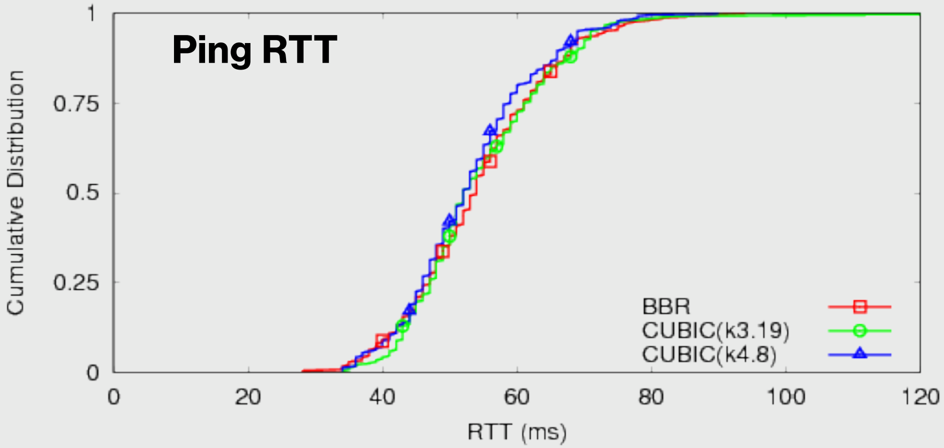
2 instances of CUBIC on Highway

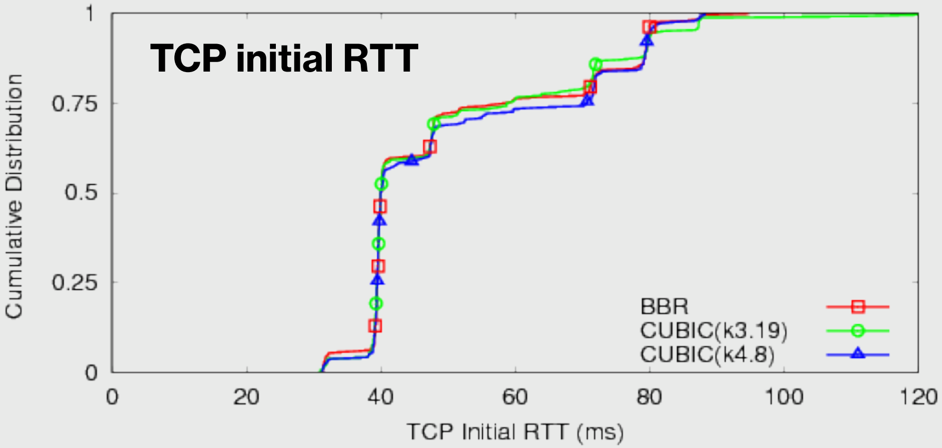
- 4 seconds to ramp up to its max owin. (left)
- Occasional loss triggers owin deduction. (right)
- Both have low link utilization b/c RTT is so small.
- No TCP loss (left) on high way.





Ping RTT vs TCP Initial RTT on Highway





- Both RTTs are in same range 40 – 100 ms
- Different Distribution, TCP packet and ICNP packet might handle differently.
- RTT based congestion control needs estimate round trip propagation delays. (e.g. BBR needs to measure the min RTT during probing phase.

Compare Throughputs of CCAs on Highway

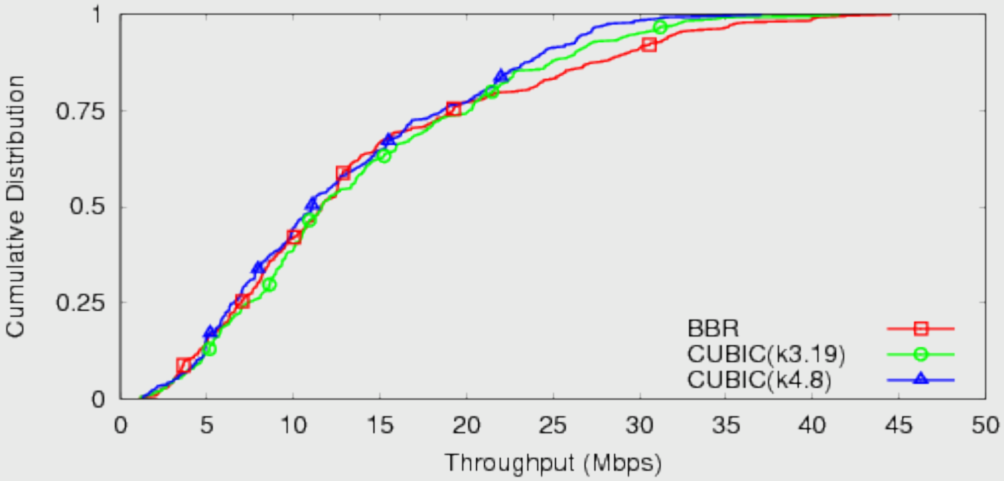


Table Overall Throughputs

CCAs	Mean	Median
BBR	14.1 ± 9.5	11.6
CUBIC(k3.19)	14.0 ± 8.4	11.6
CUBIC(k4.8)	13.0 ± 7.8	11.1

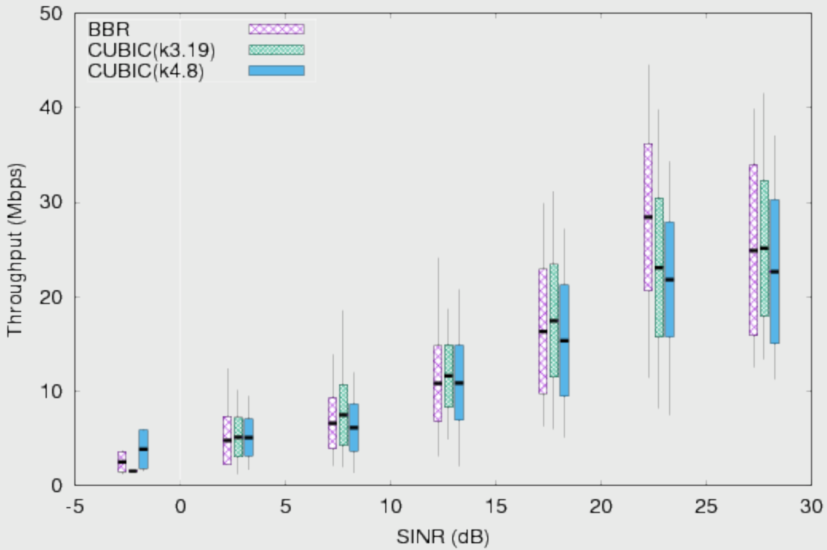


Fig. Throughputs under Different SINR

- **BBR yields comparable throughput with CUBICs on highway.**

Hand-over Between eNodeBs

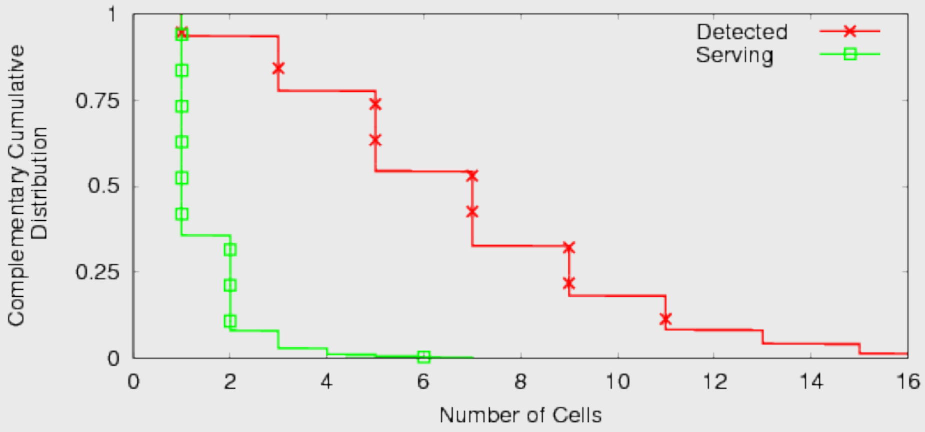


Fig. Complementary Cumulative Distribution of #Cells

- Hand-over are not as frequent as we throughput, 65%+ does not have handovers.
- Only 1 out of 720 TCP sessions experience lost connection.
- 700MHz eNode serves a large area (up to 4000 meters in radius), and car speed is only 30 m/s.
- Flows on LTE are small “mice” and “dragonflies” (short-live)

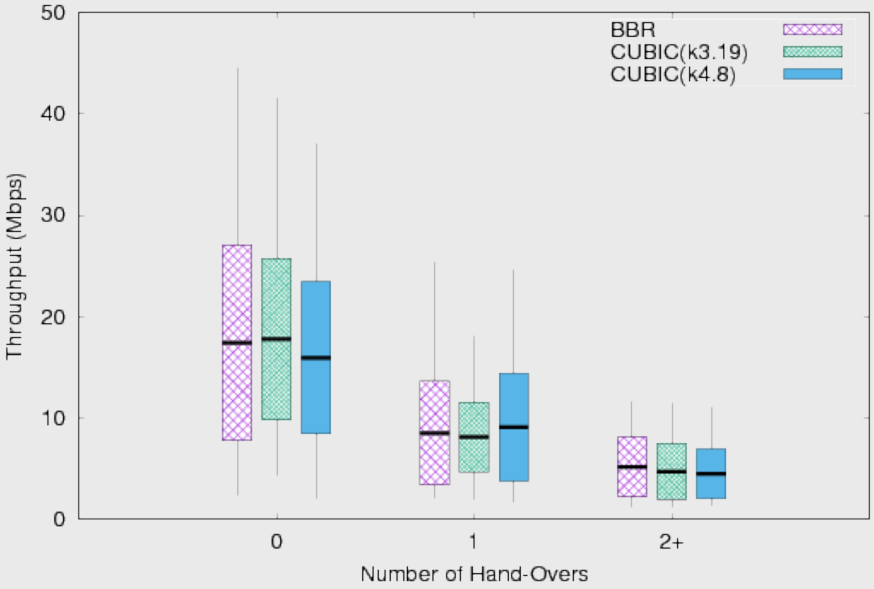
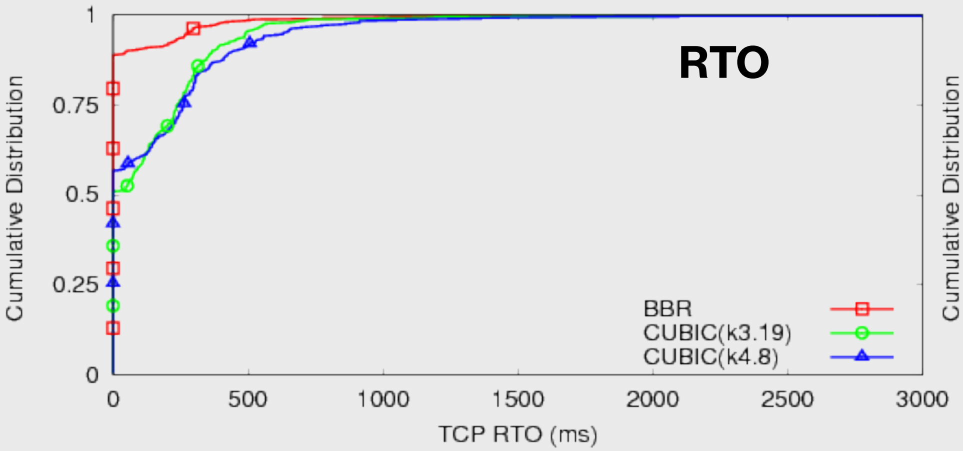
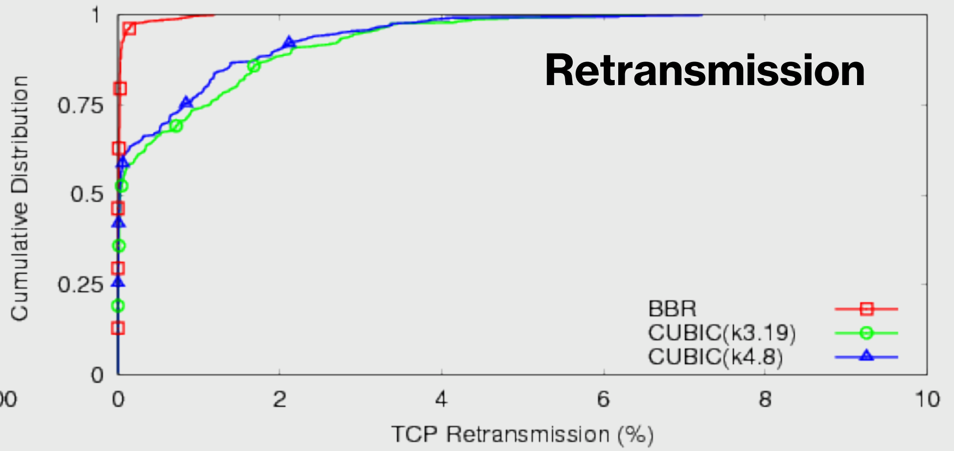


Fig. Throughput Comparison under Hand-over

- On average, multiple hand-over would lower the throughput.
- Long Live video flows would be victim of Hand-over

RTO and Retransmissions





BBR attempts to have a low RTT with smaller CWND, and its benefits are:

- Low Retransmission Rate
- Smaller RTO (lower spurious RTO rate).

RTT and Throughput

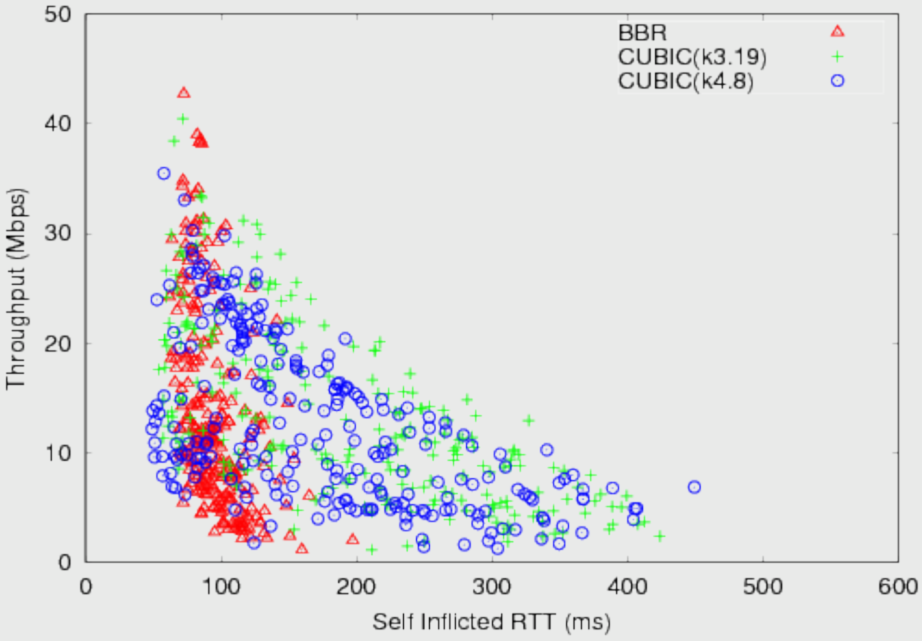


Fig. Throughput vs Self-Inflicted RTT

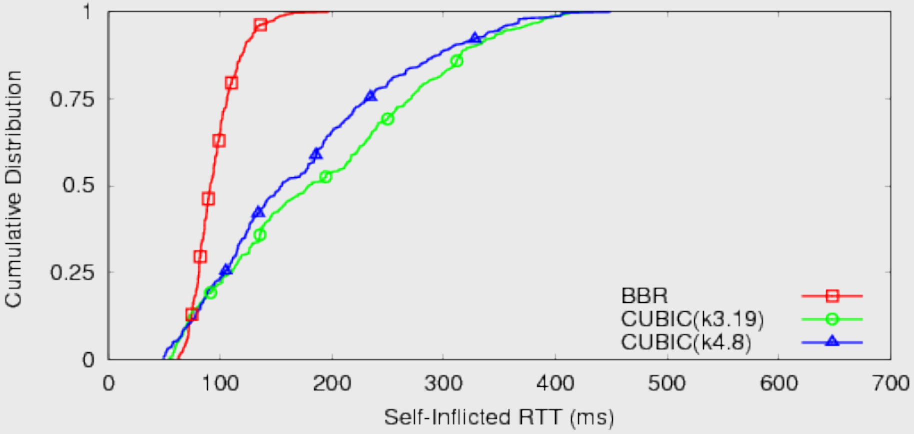
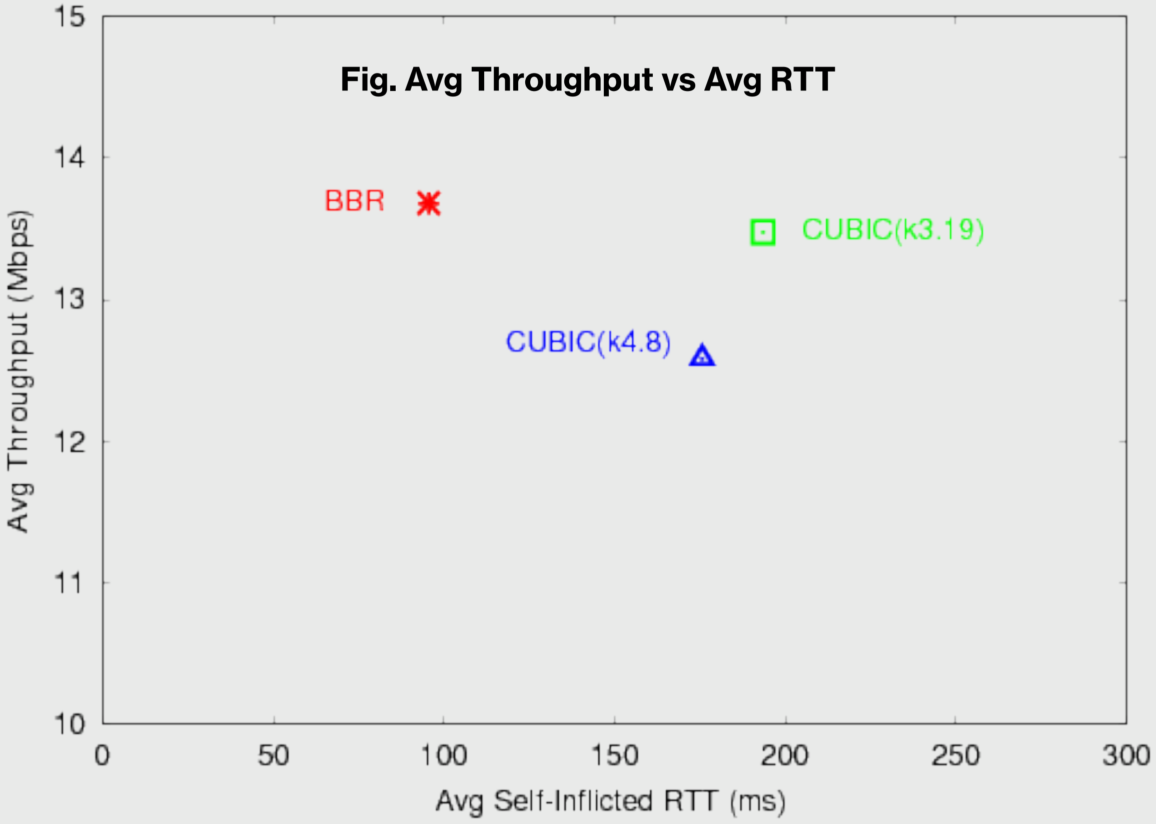


Fig. Distribution of Self-Inflicted RTT

- BBR has much less Self-inflicted RTT than CUBICs with similar throughputs.

Summary



- BBR balances the RTT and Throughput, (winner on Highway.)
- Different design principle of BBR and CUBIC

Congestion Control Algorithm over Mobile Network

- eNodeB's are bottle-neck devices over mobile network, and "buffer bloat" is the main reason for TCP performance degradation.
- Reducing maximum RWIN on UEs to avoid "buffer bloat" is not practical.
- Large buffer inside eNodeB is a double-edged sword to performance, and large buffer may increase RTT.
- Fairness may not be an important metric for CCA over LTE, because eNodes containing per-device queue.

Conclusions

Cross Layer and Comprehensive Measurement Study on Highway.

- Results as input to model and simulation in future.

CUBIC with hystart may not preform well on LTE.

- Long ramp up time to its maximum CWND, and
- Low link utilization

BBR balances RTT and Throughput.

- BBR can achieve a high throughput with low self-inflicted RTT.
- BBR would be a good CCA of choice for PEP for wireless operators.
- A good starting point to future CCA design over mobile networks.

Questions?