



THE **GAME THEORY** BEHIND RUNNING CUBIC AND BBR ON THE INTERNET

Ayush Mishra¹, Jingzhi Zhang², Melodies Sim¹, Sean Ng¹,
Raj Joshi¹, Ben Leong¹

¹ National University of Singapore

² Wuhan University

A lot of websites have made the performance driven decision to switch to BBR.

It has been reported that switching to BBR has improved throughput and reduced delay

Google Cloud

TCP BBR congestion control comes to GCP – your Internet just got faster

Neal Cardwell
Senior Staff Software Engineer

Yuchung Cheng
Senior Staff Software Engineer

C. Stephen Gunn
Senior Staff Site Reliability Engineer

Sohell Hassas Yeganeh
Senior Software Engineer

Van Jacobson
Research Scientist

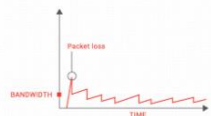
Amin Vahdat
Google Fellow

July 21, 2017

We're excited to announce that Google Cloud Platform (GCP) now features a cutting-edge new congestion control algorithm, TCP BBR, which achieves higher bandwidths and lower latencies for internet traffic. This is the same BBR that powers TCP traffic from google.com and that improved YouTube network throughput by 4 percent on average globally – and by more than 14 percent in some countries.

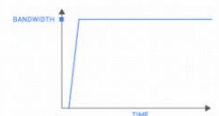
TCP before BBR

Today's Internet is not moving data as well as it should. TCP sends data at lower bandwidth because the 1980s-era algorithm assumes that packet loss means network congestion.



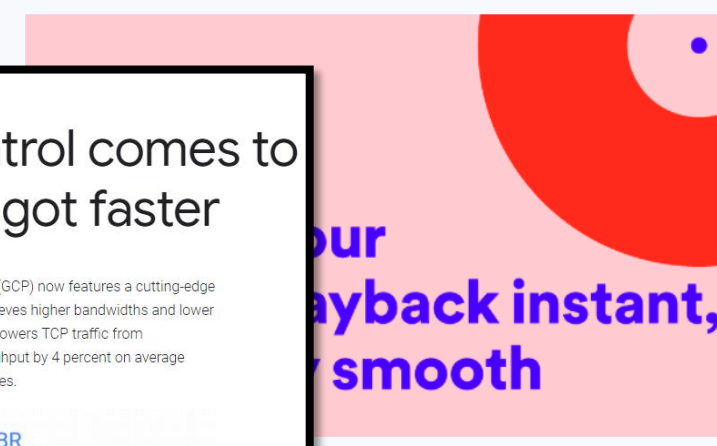
TCP BBR

BBR models the network to send as fast as the available bandwidth and is 2700x faster than previous TCPs on a 10Gb, 100ms link with 1% loss. BBR powers google.com, youtube.com, and apps using Google Cloud Platform services.



Smoother Streaming with BBR

August 31, 2018
Published by Erik Carlsson, Eirini Kakogianni

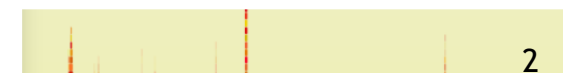


Dropbox.Tech

Topics ▾ Developers Jobs ↗

Optimizing web servers for high throughput and low latency

// By Alexey Ivanov · Sep 06, 2017



Close to **18%** of the Alexa Top 20,000 websites run BBR

This share is even larger among websites that are more popular or have a higher share of the downstream traffic

The Great Internet TCP Congestion Control Census

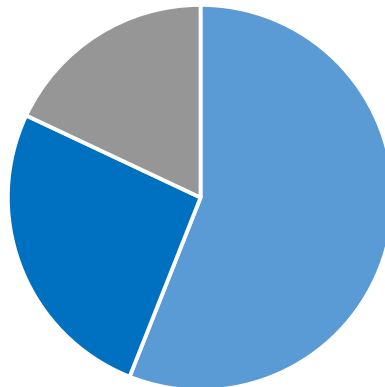
AYUSH MISHRA, National University of Singapore, Singapore
XIANGPENG SUN, National University of Singapore, Singapore
ATISHYA JAIN, Indian Institute of Technology, Delhi, India
SAMEER PANDE, Indian Institute of Technology, Delhi, India
RAJ JOSHI, National University of Singapore, Singapore
BEN LEONG, National University of Singapore, Singapore

In 2016, Google proposed and deployed a new TCP variant called BBR. BBR represents a major departure from traditional congestion-window-based congestion control. Instead of using loss as a congestion signal, BBR

Variant	Websites	Proportion
CUBIC [15]	6,139	30.70%
BBR [4]	3,550	17.75%
BBR G1.1	167	0.84%
YeAH [2]	1,162	5.81%
CTCP [34]/Illinois[22]	1,148	5.74%
Vegas [3]/Veno [13]	564	2.82%
HTCP [21]	560	2.80%
BIC [37]	181	0.90%
New Reno [28]/HSTCP [12]	160	0.80%
Scalable [20]	39	0.20%
Westwood [7]	0	0.00%
Unknown	3,535	17.67%
Short flows	1,493	7.46%
Unresponsive websites	1,302	6.51%
Total	20,000	100%

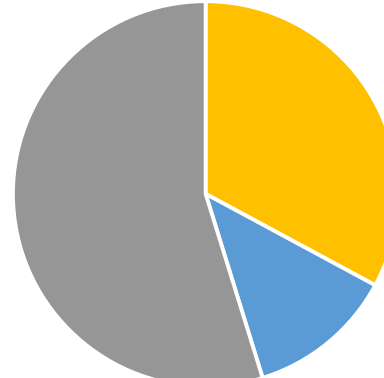
Where is this evolution headed?

What is the next paradigm shift in the Internet's congestion control landscape going to look like?



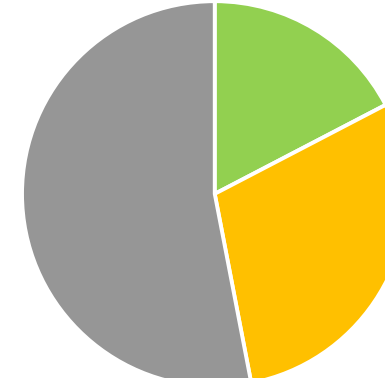
■ Reno ■ Tahoe ■ Others

[2001] Padhye et al.



■ BIC/CUBIC ■ Reno ■ Others

[2011] Yang et al.



■ BBR ■ CUBIC ■ Others

[2019] Mishra et al.

?

On Inferring TCP Behavior

Jitendra Padhye and Sally Floyd
AT&T Center for Internet Research at ICISI (ACIRI)
padhye@aciri.org, floyd@aciri.org

ABSTRACT

Most of the traffic in today's Internet is controlled by the Transmission Control Protocol (TCP). Hence, the performance of TCP has a significant impact on the performance of the overall Internet. TCP is a complex protocol with many user-configurable parameters and a range of different implementations. In addition, research continues to produce new developments in congestion control mechanisms and TCP options, and it is useful to trace the deployment of these new mechanisms in the Internet. As a final concern, the stability and fairness of the current Internet relies on the voluntary use of congestion control mechanisms by end hosts. Therefore it is important to test TCP implementations for conformant end-to-end congestion control. Since web traffic forms the majority of the TCP traffic, TCP implementations in today's web servers are of particular interest. We have developed a tool called TCP Behavior Inference Tool (TBIT) to characterize the TCP behavior of a remote web server. In this paper, we describe TBIT, and present results about the TCP behavior of major web servers, obtained using this tool. We also describe the use of TBIT to detect bugs and the consequences in TCP implementations in the Internet.

The overall congestion control behavior of the Internet is heavily influenced by the TCP implementations in web servers, since a significant fraction of the traffic in the Internet consists of TCP traffic from web servers to browsers [8]. TCP is a complex protocol with a range of user-configurable parameters. A host of variations on the basic TCP protocol [27] have been proposed and deployed. Variants on the basic congestion control mechanism continue to be developed along with new TCP options such as Selective Acknowledgment (SACK) and Explicit Congestion Notification (ECN). To obtain a comprehensive picture of TCP performance, analysis and simulations must be accompanied by a look at the Internet itself. Several factors motivated us to develop TBIT.

One motivation for TBIT is to answer questions such as "Is it appropriate to base Internet simulation and analysis on Reno TCP?" As Section 4.2 explains in some detail, Reno TCP is a older variant of TCP congestion control from 1990 that performs particularly badly when multiple packets are dropped from a window of data. TBIT shows that newer TCP variants such as NewReno and SACK are widely deployed in the Internet, and this fact should be taken

IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 21, NO. 4, AUGUST 2013
Pages: 1311–1324, DOI: 10.1109/TNET.2013.2278271

TCP Congestion Avoidance Algorithm Identification

Peng Yang, Member, IEEE, Juan Shao, Wen Luo, Liang Xu, Member, IEEE, Jitendra Dugan, Member, IEEE, and Ying Lu, Member, IEEE

Abstract—The Internet has recently been evolving from homogeneous congestion control to heterogeneous congestion control. Several years ago, Internet traffic was mainly controlled by the traditional Reno, whereas it is now controlled by multiple different TCP algorithms, such as Reno, CUBIC, and Compound TCP (CTCP). However, there is very little work on the performance and stability study of the Internet with heterogeneous congestion control. One fundamental reason is the lack of the deployment information of different TCP algorithms. In this paper, we first propose a tool called TCP Congestion Avoidance Algorithm Identification (CAAI) for actively identifying the TCP algorithm of a remote Web server. CAAI can identify all default TCP algorithms (e.g., Reno, CUBIC, and CTCP) and most non-default TCP algorithms of major operating system families. We then present the CAAI measurement result of about 20,000 Web servers. We found that only 3.31% (~14.47% of the Web servers still use Reno), 46.95% of the Web servers use BBR or CUBIC, and 14.5% (~25.95% of the Web servers use CTCP). Our measurement results show a strong sign that the majority of TCP flows are not controlled by Reno anymore, and a strong sign that the Internet congestion control has changed from homogeneous to heterogeneous.

Index Terms—Heterogeneous congestion control, Internet measurement, TCP congestion control.

THE INTERNET has recently been evolving from homogeneous congestion control to heterogeneous congestion control. Several years ago, Internet traffic was mainly controlled by the traditional Reno, whereas it is now controlled by multiple different TCP algorithms, such as Reno, CUBIC, and Compound TCP (CTCP). However, there is very little work on the performance and stability study of the Internet with heterogeneous congestion control. One fundamental reason is the lack of the deployment information of different TCP algorithms. In this paper, we first propose a tool called TCP Congestion Avoidance Algorithm Identification (CAAI) for actively identifying the TCP algorithm of a remote Web server. CAAI can identify all default TCP algorithms (e.g., Reno, CUBIC, and CTCP) and most non-default TCP algorithms of major operating system families. We then present the CAAI measurement result of about 20,000 Web servers. We found that only 3.31% (~14.47% of the Web servers still use Reno), 46.95% of the Web servers use BBR or CUBIC, and 14.5% (~25.95% of the Web servers use CTCP). Our measurement results show a strong sign that the majority of TCP flows are not controlled by Reno anymore, and a strong sign that the Internet congestion control has changed from homogeneous to heterogeneous.

The Great Internet TCP Congestion Control Census

AYUSH MISHRA, National University of Singapore, Singapore
XIANGPENG SUN, National University of Singapore, Singapore
ATISHYA JAIN, Indian Institute of Technology, Delhi, India
SAMEER PANDE, Indian Institute of Technology, Delhi, India
RAJ JOSHI, National University of Singapore, Singapore
BEN LEONG, National University of Singapore, Singapore

In 2016, Google proposed and deployed a new TCP variant called BBR. BBR represents a major departure from traditional congestion-window-based congestion control. Instead of using loss as a congestion signal, BBR uses estimates of the bandwidth and round-trip delays to regulate its sending rate. The last major study on the distribution of TCP variants on the Internet was done in 2011, so it is timely to conduct a new census given the recent developments around BBR. To this end, we designed and implemented Gordon, a tool that allows us to measure the exact congestion window (cwnd) corresponding to each successive RTT in the TCP connection response of a congestion control algorithm. To compare a measured flow to the known variants, we created a localized bottleneck where we can introduce a variety of network changes like loss events, bandwidth change, and increased delay, and normalize all measurements by RTT. An offline classifier is used to identify the TCP variant based on the cwnd trace over time.

Our results suggest that CUBIC is currently the dominant TCP variant on the Internet, and it is deployed on about 36% of the websites in the Alexa Top 20,000 list. While BBR and its variant BBR G1.1 are currently in second place with a 22% share by website count, their present share of total Internet traffic volume is estimated to be larger than 40%. We also found that Akamai has deployed a unique loss-agnostic rate-based TCP variant on some 6% of the Alexa Top 20,000 websites and there are likely other undocumented variants. The traditional assumption that TCP variants "in the wild" will come from a small known set is not likely to

Given this performance improvement, how do we expect the Internet to evolve?

Is it reasonable to expect everyone to switch from CUBIC to BBR?

APNET 2021

Conjecture: Existence of Nash Equilibria in Modern Internet Congestion Control

AYUSH MISHRA, National University of Singapore, Singapore

JINGZHI ZHANG, Wuhan University, China

MELODIES SIM, National University of Singapore, Singapore

SEAN NG, National University of Singapore, Singapore

RAJ JOSHI, National University of Singapore, Singapore

BEN LEONG, National University of Singapore, Singapore

The Internet's congestion control landscape is currently in the midst of an unprecedented paradigm shift. A recent measurement study found that BBR, a congestion control algorithm introduced by Google in 2016, has seen rapid adoption and is deployed at more

China, China. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3469393.3469397>

Given this performance improvement, how do we expect the Internet to evolve?

Is it reasonable to expect everyone to switch from CUBIC to BBR?

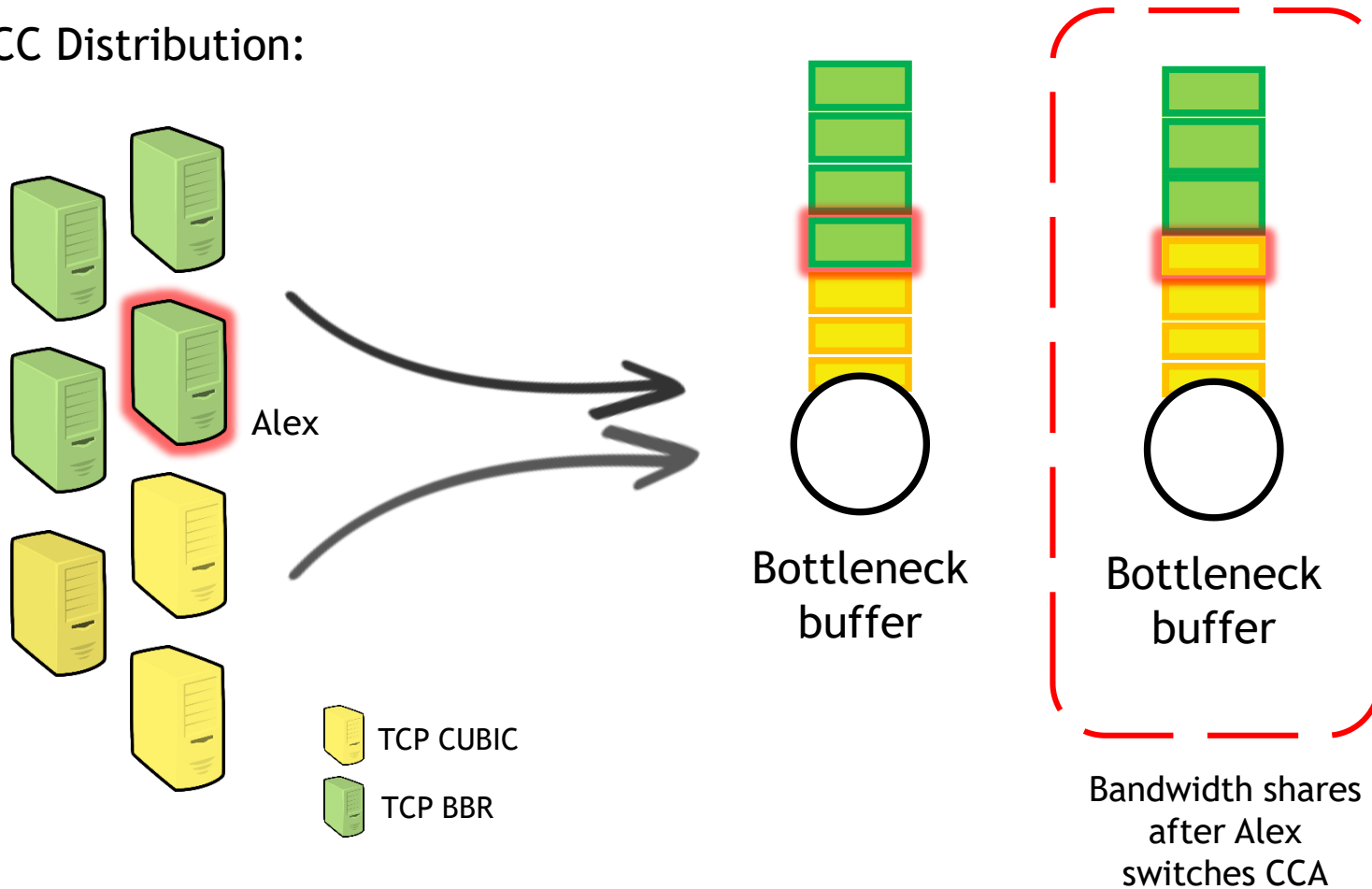
Key Insight:

We have some players that can maximize some **utility**
All the players have **strategies** (CUBIC/BBR) available to them to maximize their utility.

It's a Normal Form Game!

Given some network, we can calculate the **Nash Equilibrium** CC distribution.

CC Distribution:



If **everyone does worse after making a switch**, then the given CC Distribution is the **Nash Equilibrium**.

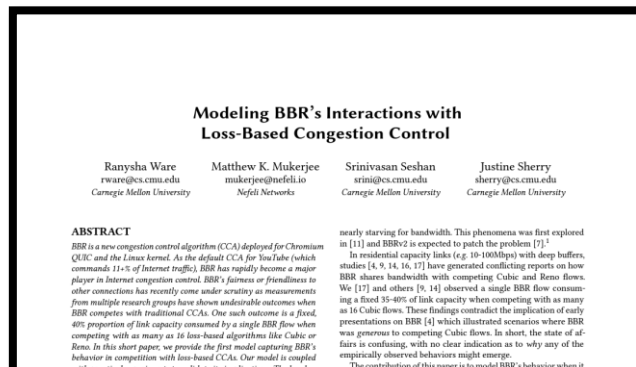
If everyone on the Internet chooses between CUBIC and BBR based on throughput, this Nash Equilibrium distribution is our best *estimate* of the future of the congestion control landscape.

We think that **there will always be a Nash Equilibrium** in a network with CUBIC and BBR flows.

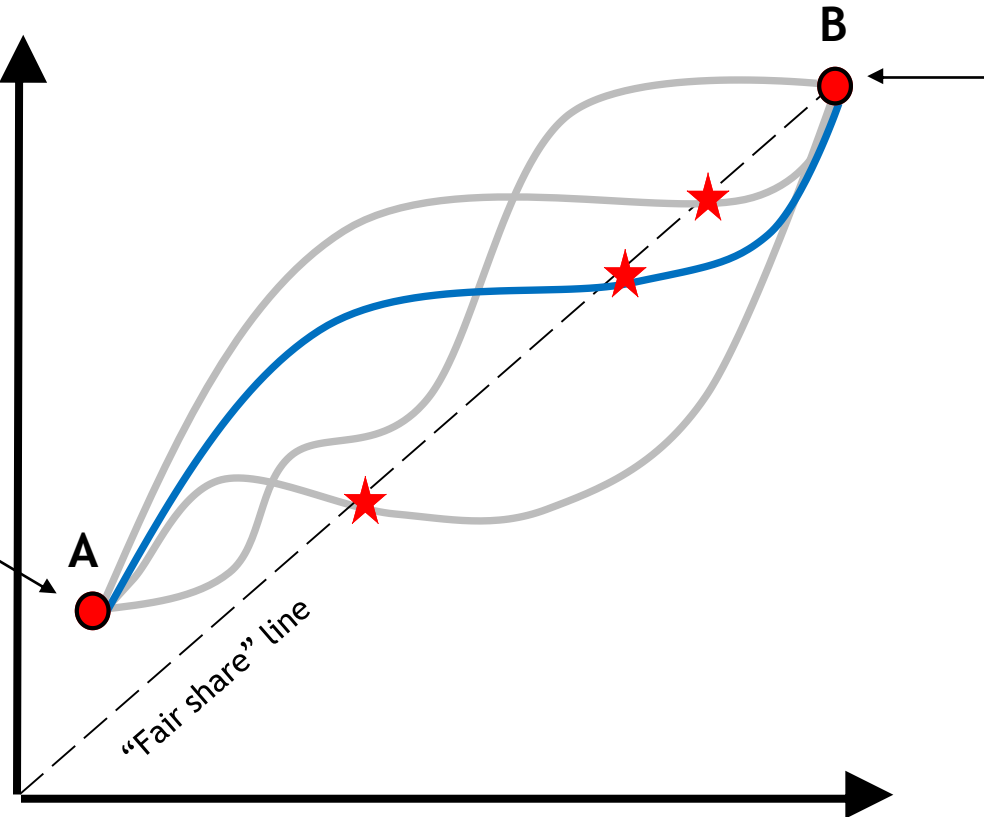
When all the flows at the bottleneck are BBR flows, they will utilize the bottleneck link

A small number of BBR flows can get a disproportionately high share of the bottleneck bandwidth

[2019] Ware et al.



Combined throughput of all the BBR flows



Percentage of flows running BBR at the bottleneck



Each point of intersection is a **Nash Equilibrium distribution!**

We think that **there will always be a Nash Equilibrium** in a network with CUBIC and BBR flows.

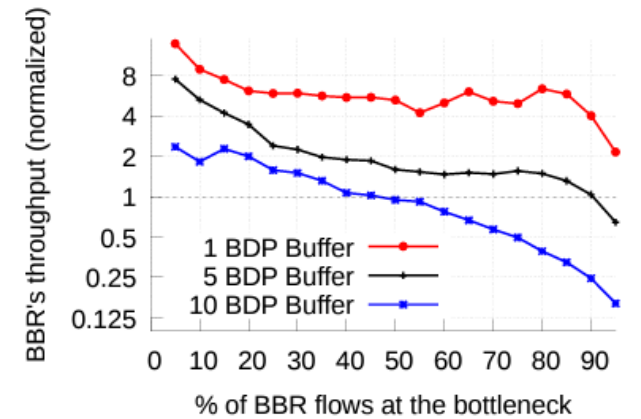
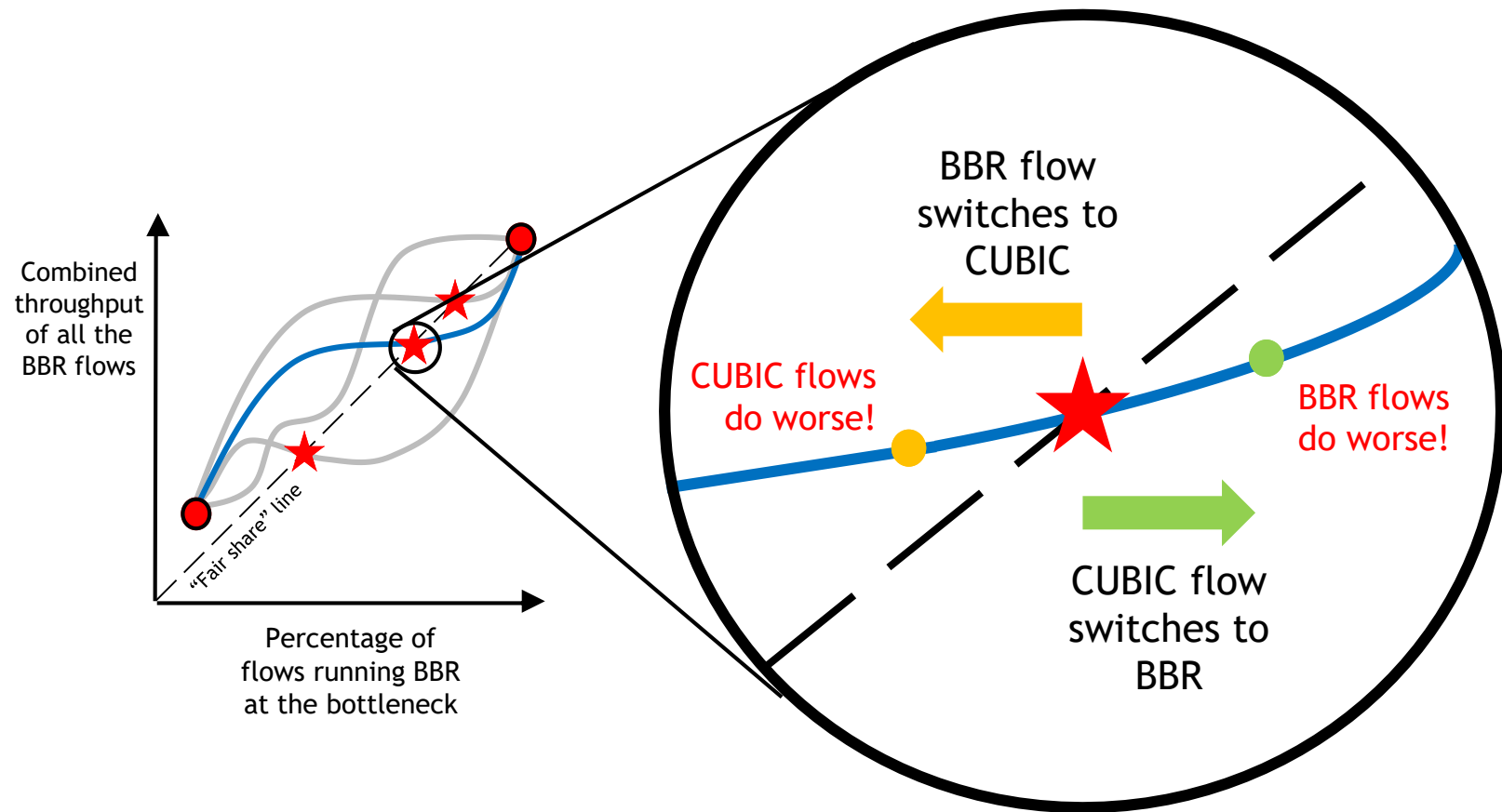


Fig. 2. BBR's throughput vs. % of BBR flows.

We can **exhaustively** prove a NE will always exist when 2 flows compete.

The proof is based on simple observations made by other measurement papers on how CUBIC and BBR compete.

Table 2. Outcomes in a two-flow game. ($RTT_1 > RTT_2$, winning strategies are **highlighted**)

	$Buff < T_2$				$T_2 < Buff < T_3$				$T_3 < Buff$			
Scenario	Strategies		Outcome		Strategies		Outcome		Strategies		Outcome	
	S_1	S_2	S_1	S_2	S_1	S_2	S_1	S_2	S_1	S_2	S_1	S_2
1	C	C	L	W	C	C	L	W	C	C	L	W
2	C	B	L	W	C	B	W	L	C	B	W	L
3	B	C	W	L	B	C	W	L	B	C	L	W
4	B	B	W	L	B	B	W	L	B	B	W	L
<u>Nash Equilibrium</u>	$(B, *)$				$(B, *)$				$(*, C)$			

Empirical validation

Checking the claims of this conjecture in a limited state space

- NE was computed in 6, 9 and 12 flow systems with each third of the flows having RTTs 20, 50 and 80 ms respectively. All flows shared a common bottleneck with a fixed link speed.
- For a given number of flows and a network configuration we ran all the possible combinations of flows running either CUBIC or BBR.

CBC

Is the NE if:

In BBC, flow 1 gets worse throughput

In CCC, flow 2 gets worse throughput

In CBB, flow 3 gets worse throughput

Properties of observed NE

- In each case, we observed **exactly one Nash Equilibrium**
- In each Nash Equilibrium, when the flows were sorted by the RTT, CUBIC was always picked by the smallest RTT flows

(CCC...)(...BBB)

m flows

(n-m) flows

- That is, if the NE for 6 flows is when 50% of the flows are running CUBIC, the NE is at:

C C C B B B

RTTs: 20 ms 50 ms 80 ms

Effect of buffer size and link speed on the NE

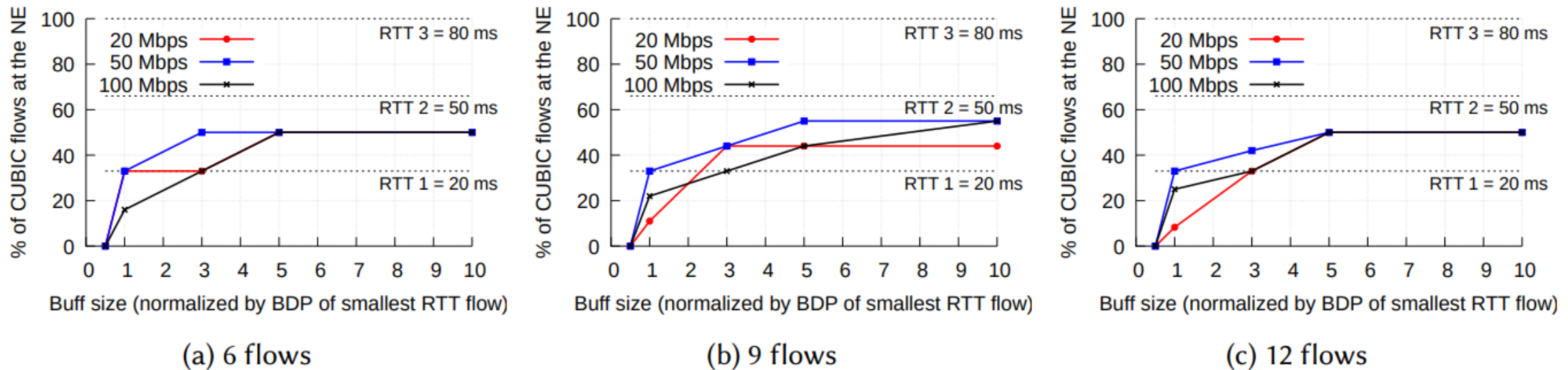


Fig. 4. The effect of link capacity and number of flows on the Nash Equilibrium. RTTs 20 ms, 50 ms and 80 ms.

Predictably, buffer size had the biggest impact on the CC distribution at the NE

Effect of RTT distribution on the NE

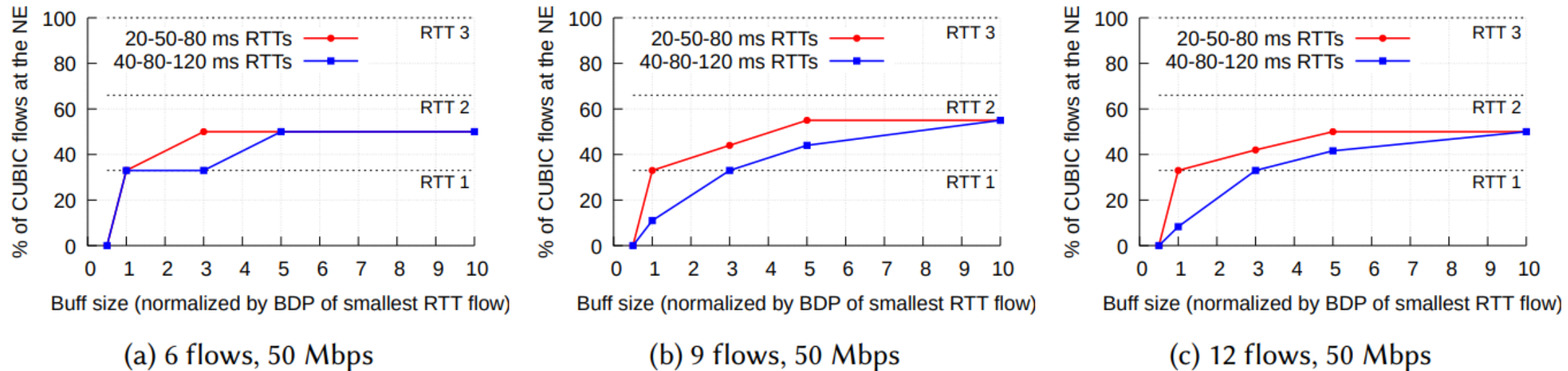


Fig. 5. The effect of the RTT distribution on the Nash Equilibrium.

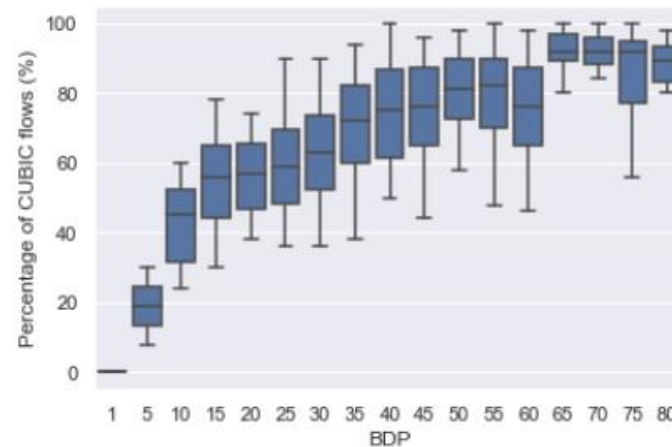
RTT distribution had little effect of the where the NE was

Conclusion

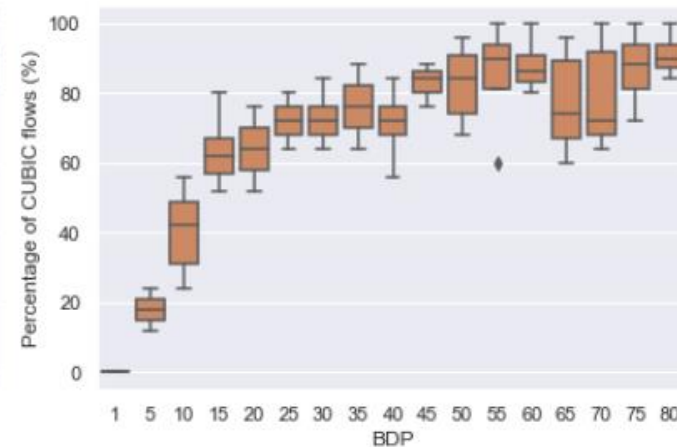
- Despite BBR's current throughput benefits, CUBIC is unlikely to disappear soon from the Internet
- The Internet is likely to remain a heterogeneous mix of congestion control algorithms
- TCP performance is highly contextual
- However, the Internet does not follow economic game theory - it is not a given that the Internet will move towards the Nash Equilibrium.

Future work

- Formal proof for NE is a **general n-flow game**
- The effect of more complex network utility functions (**delay, jitter**)
- Effect on the NE in the presence of **BBRv2, multi-hop paths, and AQMs**
- NE in **very deep buffers** and a **large number of flows**:



(a) 100Mbps, 50 flows



(b) 50Mbps, 25 flows

Thank you for your time!

Please feel free to get in touch with me at

`ayush@comp.nus.edu.sg`

Extra slides

Observation 1. When competing at the bottleneck where the buffer is deep, CUBIC tends to have higher throughput than BBR; the converse is true when the buffer is shallow.

Observation 2. When a single BBR flow competes with a single CUBIC flow at a bottleneck, there must exist some threshold bottleneck buffer size T_{fair} such that when the bottleneck buffer size $Buff < T_{fair}$, the BBR flow gets higher throughput than the CUBIC flow and when $Buff > T_{fair}$, the CUBIC flow gets higher throughput than BBR.

Observation 3. As the percentage of BBR flows at the bottleneck increases, the per-flow average throughput of BBR flows at that bottleneck decreases.

Observation 4. When two BBR flows compete at a bottleneck, the flow with a longer RTT will get higher bandwidth than the flow with a shorter RTT. When two CUBIC flows compete at a bottleneck, the flow with a shorter RTT will get higher bandwidth than the flow with a longer RTT.

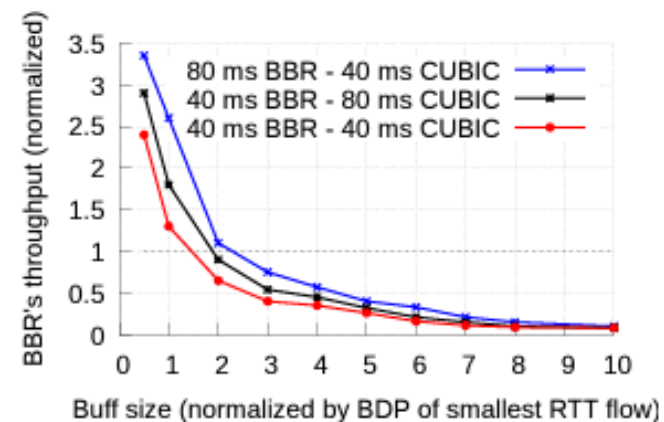


Fig. 1. BBR throughput vs. bottleneck buffer size.

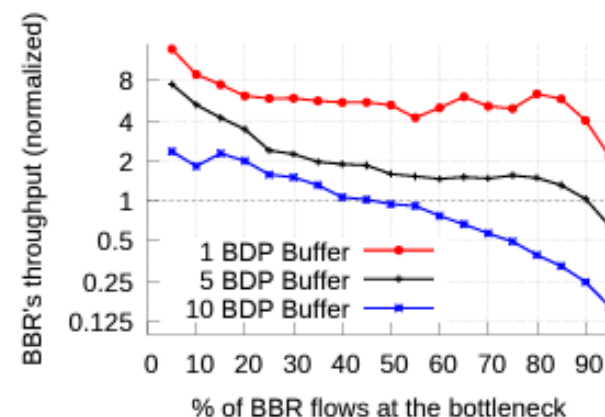


Fig. 2. BBR's throughput vs. % of BBR flows.