Toward Fair and Efficient Congestion Control: Machine Learning Aided Congestion Control (MLACC)

Department: Distributed Data Storage and Management Lab – Network and Protocols Team
Authors: Ahmed Elbery, Yi Lian, and Geng Li
Date: June 29-30, 2023
目录

• Introduction and Motivation
  • New Application Requirements
  • Current Congestion Control
  • State-of-the-Industry Examples

• MLACC
  • High Level Overview
  • System Architecture
  • How the Rate Control Algorithm Works?
  • Bandwidth Estimation Model

• Evaluation

• Conclusion and Future Work

• Modern applications are sensitive to packet drops and latency
  • Transaction based application
  • Remote Direct Memory Access (RDMA)
• Parallel computing also requires fair sharing of BW resources
• Most of the current Congestion Control (CC) attempt to maximize the throughput, which may result in:
  • Longer queuing delay
  • Higher delay variance
  • Higher packet drops

We need a CC that can minimize drops, minimize latency and increase fairness
Current Congestion Control

Single Flow TCP CC Algorithms
- loss-based
  - CUBIC
  - Reno
  - NewReno
- delay-based
  - Vegas
  - New Vegas
- hybrid
  - BBR
  - YeAH

Multi-Flow TCP CC Algorithms
- MP-TCP
  - LIA
  - OLIA
  - BALIA

Alternative CC Algorithms
- network assisted CC
  - PFC

Alternative CC Algorithms
- YeAH
- MP-TCP
- LIA
- OLIA
- BALIA

Fast

BBR

LIA

OLIA

BALIA

PFC
State-of-the-Industry Examples

• Google TCP-BBR:
  • Maximizes throughput
  • Can cause congestion
  • High latency variation
  • Lack of fairness

• Priority Flow Congestion Control (PFC) (Microsoft):
  • Prevents congestion and achieves low latency
  • Can affect other flows (victim flow problem)
  • Lack of fairness (parking lot problem)
Machine Learning Aided CC (MLACC): High level Overview

- Idea: scarify a small portion of throughput to...
  - Minimize congestion
  - Minimize latency and latency variation
  - Have good level of fairness
- How it works
  - Define a bottleneck utilization threshold $u_{thr}$
  - Based on the current utilization $u_t$ and $u_{thr}$, different senders can coordinately change their rates to avoid congestion.
- Doesn’t pause traffic like PFC
- Doesn’t overload the bottleneck like TCP-BBR

Now, the questions became:
1- How senders can know the bottleneck link utilization
2- How senders can update their rate
How senders can know bottleneck utilization?

- Using feedback from network (In-Network Telemetry (INT))
  - More accurate
  - Not all devices are INT enabled
- Using ML models to estimate the bottleneck utilization based on QoS measurements
  - Easy deployment
  - ML estimation errors may be a problem

We use the second method because it can be deployed in different use cases without any hardware changes.
Machine Learning Aided CC (MLACC): Bandwidth Estimation Model

- We use long short-term memory (LSTM) model to predict the available ratio of the BW
- Input features:
  - Average and standard deviation values of RTT, RTT 1st and 2nd order derivative
  - Packet loss
- Time step = 0.1 second
- Data collection using NS3 simulation
  - Different BW
  - Dynamic background traffic rates

Achieve errors less than 5 % MAPE.
Machine Learning Aided CC (MLACC): System Architecture

Rate Controller

Sx. Socket

Rx. Socket

BUEM

BUPM: Bottleneck Utilization Estimation Module

Feedback (RTT, drops rates,..)
Machine Learning Aided CC (MLACC): How the Rate Control Algorithm Works

- To avoid congestion, MLACC compares the current link utilization $u_t$ to its threshold $u_{thr}$ and a tolerance factor $\gamma$
  - If **over** utilized --->>> the sender **decreases** its rate
  - If **under** utilized --->>> sender **increases** its rate
  - If **in range** --->>> increases its rate *carefully*
- Amount of increase/decrease
  - The **higher** $r_t$, the **lower the increase** and the **higher the decrease**
  - The **higher** the gradient, the **lower the increase** and the **higher the decrease**

Why the gradient?
It can represent how a traffic flow contributes to the congestion on the bottleneck link.

How can senders update their rate?
Evaluation

• Network setup:
  • MLACC vs. TCP BBR using NS3 simulation
  • N clients sending a server through the same bottleneck link
  • Bottleneck capacity [5 Mbps, 10 Mbps, 15 Mbps]
  • $u_{thr} = 0.95$, $\gamma = 0.05$
  • Metrics: Latency and Throughput

MLACC scarifies a very small ratio of throughput (< 3.2 %) to save up to 52% of delay

<table>
<thead>
<tr>
<th></th>
<th>5Mbps</th>
<th></th>
<th>10Mbps</th>
<th></th>
<th>15Mbps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Av. Rate</td>
<td>4.82</td>
<td>Av. Rate</td>
<td>9.04</td>
<td>Av. Rate</td>
<td>13.60</td>
</tr>
<tr>
<td>Av. Lat.</td>
<td>30.34</td>
<td>Av. Lat.</td>
<td>44.94</td>
<td>Av. Lat.</td>
<td>36.07</td>
</tr>
<tr>
<td>BBR</td>
<td></td>
<td>MLACC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Av. Rate</td>
<td>4.67</td>
<td>Av. Rate</td>
<td>9.14</td>
<td>Av. Rate</td>
<td>13.65</td>
</tr>
<tr>
<td>Av. Lat.</td>
<td>22.51</td>
<td>Av. Lat.</td>
<td>21.49</td>
<td>Av. Lat.</td>
<td>21.82</td>
</tr>
<tr>
<td>Saving %</td>
<td>-3.19</td>
<td>Saving %</td>
<td>-1.12</td>
<td>Saving %</td>
<td>-0.31</td>
</tr>
<tr>
<td></td>
<td>25.80</td>
<td></td>
<td>52.17</td>
<td></td>
<td>39.50</td>
</tr>
</tbody>
</table>
Evaluation

- MLACC achieves
  - lower latency
  - Lower latency variation
  - Fair BW sharing
Conclusion and Future Work

• Using the gradient of the link utilization w.r.t the rate can be a good signal to CC techniques.
• MLACC scarifies a very small ratio of throughput (< 3.2 %) to save up to 52% of delay
• The proposed CC is a first step toward designing an efficient CC, There are several extensions can be conducted in future such as:
  • Perform more in-depth analysis and comprehensive simulation in different use cases to understand any its limitation
  • Compare MLACC with the SOTA protocols such as PFC to study it suitability for emerging applications such as RDMA
  • Apply the MLACC with QoS protocols such as IEEE802.1p to study how different traffic type can affect the overall performance
Thank You!

Any questions?

For any questions please contact Ahmed Elbery on ahmeda.elbery@huawei.com aelbery@vt.edu

把数字世界带入每个人、每个家庭、每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd. All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.