GRU-Based Learning for the Identification of Congestion Protocols in TCP Traffic

Paul Bergeron, Sandhya Aneja

{paul.bergeron1, sandhya.aneja}@marist.edu

MARIST UNIVERSITY

Abstract

This paper presents the identification of congestion control protocols TCP Reno, TCP Cubic, TCP Vegas, and BBR on the Marist University campus, with an accuracy of 97.04% using a GRU-based learning model. We used a faster neural network architecture on a more complex and competitive network in comparison to existing work and achieved comparably high accuracy.

1 Introduction

Identifying TCP congestion control protocols is vital for performance monitoring, optimization, and future Internet research.

Contributions:

- Achieved 97.04% accuracy using GRU.
- Used size, maximum window size, throughput, smoothed throughput, RTT as representative features.
- BBR → max throughput
- Cubic → lowest RTT.

2 Motivation

- **Reno:** halves window on loss, additive increase.
- **Cubic:** cubic growth, smooth convergence.
- **Vegas:** compares expected vs. actual throughput using RTT.
- **BBR:** probes bandwidth/RTT, aggressive startup.

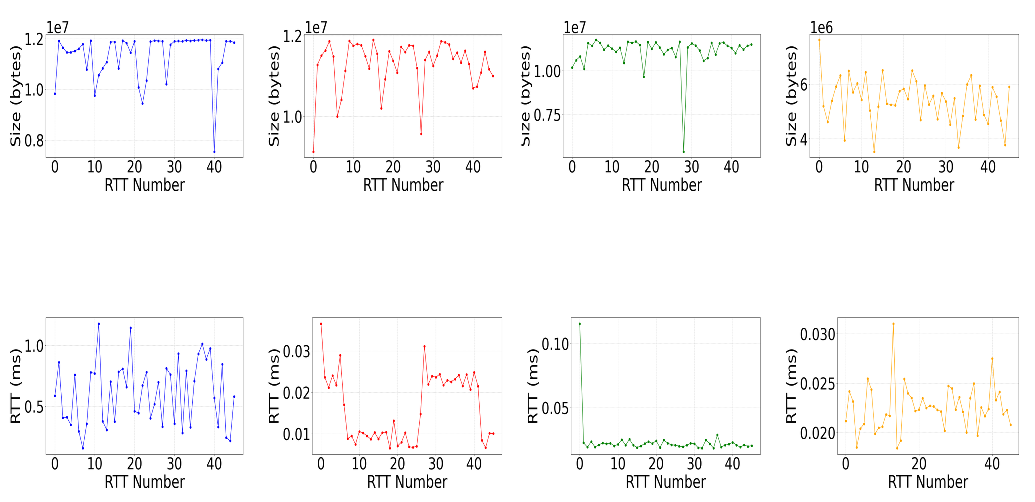

Figure 1: Size and RTT variation for BBR, Cubic, Reno, and Vegas across 47 RTTs.

3 Problem Statement

We address the problem of distinguishing among communication flows that operate under different TCP congestion control protocols (Reno, Cubic, Vegas, BBR) by modeling their temporal behavior.

To capture protocol-specific patterns in throughput, window size, and RTT, we employ a GRU with attention, which learns the characteristic dynamics (e.g., Reno's sawtooth, BBR's ramp-up) from flow sequences.

4 Related Work

- **TBIT** [1]: heuristic, rule-based, active probing.
- **CAAI** [3]: active probing, 30k servers, shows Reno→Cubic shift.
- **DeepCCI** [2]: CNN+LSTM on packet arrival histograms, 99% accuracy in lab.

Our work: GRU + attention on 1 Gbps real campus network, increasing complexlity.

5 Experimental Setup

- Client–server testbed at Marist (1 Gbps bottleneck).
- Protocols: Reno, Cubic, Vegas, BBRv1.
- 500 MB transfers, repeated 15 days, 3× daily.
- Captured via Wireshark, features extracted with `tshark`.

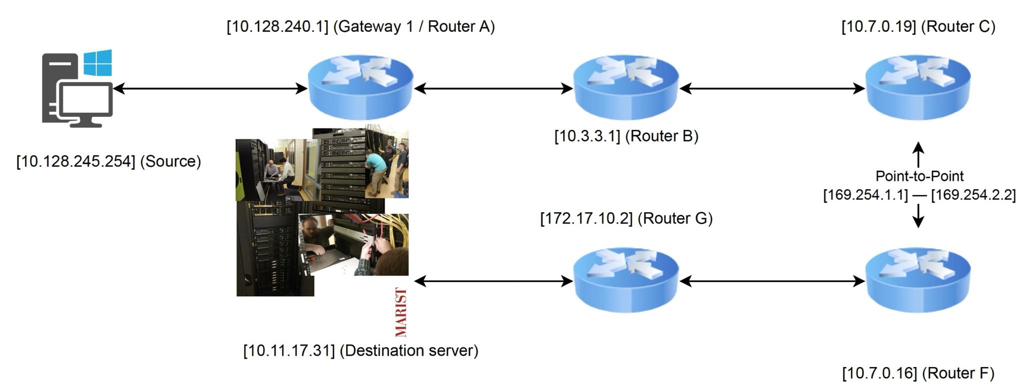

Figure 2: Experimental testbed.

5.1 Features

- Extracted every 100 ms: size, maximum window size, throughput, smoothed throughput, RTT.
- Smoothed throughput via rolling mean.
- Time excluded to avoid bias.

5.2 GRU Model

- 3-layer bidirectional GRU
- hidden size 512
- Dropout = 0.4.
- Optimizer: Adam, LR = 0.000075.
- Scheduler: ReduceLROnPlateau.
- Training: 30 epochs, batch size = 8.
- Train/val/test split: 70/10/20.

6 Results

Table 1: Data distribution

Protocol	Vegas	Reno	Cubic	BBR
Samples	3221	1802	1777	1629
% total	38.6	21.6	21.3	19.5

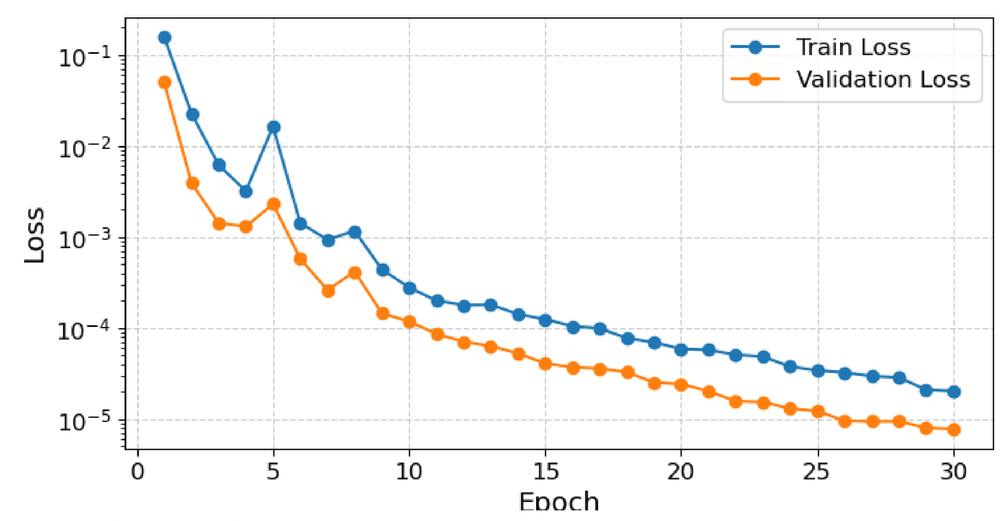

Figure 3: Training and validation loss (log scale).

Key Findings:

- Test accuracy = 97.04%.
- BBR → max throughput; Cubic → min RTT.
- Dataset imbalance corrected via standardization.

7 Conclusion & Future Work

- GRU + attention effectively identifies TCP CC protocols.
- Results hold even for encrypted traffic.
- Accuracy varies across network types.
- Future: wireless, mobile, heterogeneous networks.

References

[1] Jitendra Pahdye and Sally Floyd. On inferring TCP behavior. *ACM SIGCOMM Computer Communication Review*, 31, 2001.

[2] Constantin Sander, Jan Rüth, Oliver Hohlfeld, and Klaus Wehrle. Deepcci: Deep learning-based passive congestion control identification. In *Proceedings of the 2019 workshop on network meets AI & ML*, 2019.

[3] Peng Yang, Juan Shao, Wen Luo, Lisong Xu, Jitender Deogun, and Ying Lu. Tcp congestion avoidance algorithm identification. *IEEE/ACM Transactions on Networking*, 22(4):1311–1324, 2014.