



Title	[奨励講演]超低遅延ネットワークアーキテクチャL4Sを用いた遠隔強化学習の性能評価
Author(s)	奥村, 丈; 三橋, 力麻; 砂原, 悟 他
Citation	電子情報通信学会技術研究報告, 126(132), 1-6
Issue Date	2026-07-23
Doc URL	https://hdl.handle.net/2115/100413
Rights	Copyright ©2026 IEICE
Type	conference paper
File Information	NS2026-48.pdf



[奨励講演] 超低遅延ネットワークアーキテクチャL4Sを用いた 遠隔強化学習の性能評価

奥村 丈[†] 三橋 力麻^{††} 砂原 悟^{†††} 飯田 勝吉^{†††}

[†]北海道大学 大学院情報科学院 情報理工学コース 先端ネットワーク研究室
〒060-0811 札幌市北区北 11 条西 5 丁目

^{††}, ^{†††}北海道大学 情報基盤センター 〒060-0811 札幌市北区北 11 条西 5 丁目
E-mail: [†]okumura.jou.d2@elms.hokudai.ac.jp, ^{††}mitsuhashi@tanega.iic.hokudai.ac.jp,
^{†††}{suna,iida}@iic.hokudai.ac.jp

あらまし 近年、自動運転車の制御や自律ドローンの操作など、機械制御を学習・実行するシステムの研究開発が活発化している。機械制御に遠隔強化学習を適用したシステムでは、実際の機器とサーバの間で通信データを頻繁にやり取りするため、ネットワーク品質が制御性能や安全性に大きく影響する。最近では、ネットワーク遅延を抑制する手法として RFC9330 で標準化された L4S (Low Latency, Low Loss, and Scalable Throughput) が注目を集めているが、現時点で遠隔強化学習に L4S を適用した報告はほとんど見られていない。前回の研究会では遠隔強化学習の基礎的評価を目的とした、L4S を適用したネットワーク上での遠隔強化学習の環境構築について述べた。しかし、以前の発表では環境の構築に留まり、性能評価について紹介することが出来なかったため、本発表ではネットワーク遅延が強化学習の性能に及ぼす影響を中心とした、遠隔強化学習の評価の結果及び考察について述べる。

キーワード 遠隔強化学習, 倒立振子, L4S, ns-3, DEFIANCE

[Encouragement Talk] Performance Evaluation of Remote Reinforcement Learning Using L4S Ultra-low Latency Network Architecture

Jo OKUMURA[†], Rikima MITSUHASHI^{††}, Satoru SUNAHARA^{†††}, and Katsuyoshi IIDA^{†††}

[†] Laboratory of Advanced Information Network,
Course of Computer Science and Information Technology, Graduate School of Information Science & Engineering,
Hokkaido University, Kita 11 Nishi 5, Kita-ku, Sapporo, 060-0811, Japan.
^{††}, ^{†††} Information Initiative Center, Hokkaido University, Kita 11 Nishi 5, Kita-ku, Sapporo, 060-0811, Japan.
E-mail: [†]okumura.jou.d2@elms.hokudai.ac.jp, ^{††}mitsuhashi@tanega.iic.hokudai.ac.jp,
^{†††}{suna,iida}@iic.hokudai.ac.jp

Abstract In recent years, research and development of systems for learning and execution of machine control, such as autonomous vehicles and drones, has been gaining momentum. In machine control systems employing remote reinforcement learning, data is frequently exchanged between physical devices and remote servers. Consequently, network quality has a significant impact on control performance and operational safety. Recently, L4S (Low Latency, Low Loss, and Scalable Throughput) standardized in RFC 9330, has attracted considerable attention as a technology for reducing network latency. However, despite its potential advantages, there have been very few studies investigating the application of L4S to remote reinforcement learning environments. In our previous presentation, we described the construction of a remote reinforcement learning environment over an L4S-enabled network. However, that presentation focused solely on the construction of the experimental environment and did not include a performance evaluation. Therefore, this presentation reports the evaluation results obtained using the constructed environment, with the impact of network latency on reinforcement learning performance.

Key words Remote reinforcement learning, cartpole, L4S, ns-3, and DEFIANCE.

1. はじめに

近年、自動運転車の制御や自律ドローン操作など機械制御の高度化が進み、これらの制御アルゴリズムを学習・実行するシステムの検討が活発になっている [1], [2]。機械制御に遠隔強化学習を適用したシステムでは、機械側で取得した観測データが中央制御や学習を行うサーバ側に送信され、そこで更新された方策に基づく次の行動が機械側へ返されることになる。つまり、実際の機器とサーバとの間でデータを頻繁にやり取りするため、ネットワークの帯域、遅延、輻輳といった品質が機械制御の性能や安全性に大きく影響する。

従来、ネットワーク遅延を抑制する手法としては、CoDel (Controlled Delay) [3]~[5] や PIE (Proportional Integral Controller Enhanced) [6] などの AQM 技術が利用されてきたが [7]~[10]、遠隔強化学習を適用するネットワーク環境では、より大きな性能改善が期待される。こうした中、最近では、RFC9330 として標準化された L4S (Low Latency, Low Loss, Scalable throughput) [11] は、ECN (Explicit Congestion Notification) [12], [13] を活用した柔軟な輻輳検知・回避により、低いキュー遅延、低い輻輳損失、スケーラブルなスループット制御を実現できる次世代通信アーキテクチャとして注目を集めている。

L4S の適用範囲としては、動画配信サービス、仮想現実 (VR) や拡張現実 (AR)、オンラインゲームなど高品質なネットワークを要求するサービスでの利用が想定されており [14]、自動運転車や産業用ロボットなどを機械制御するための遠隔強化学習と L4S の親和性は高いと考えられるが、現時点で遠隔強化学習に L4S を適用した報告はほとんど見られていない。

そこで本研究では、機械制御の基本の一つである倒立振り子を対象に、遠隔強化学習の制御ループに対する L4S 等のキュー管理方式の影響を評価する。具体的には、ネットワークシミュレータである ns-3 [15] と、ns-3 上で遠隔強化学習を実現する DEFIANCE [16] を用いて構築したシミュレーション環境において、競合する背景トラフィックを投入して実質的なネットワーク輻輳を機動的に再現する。これにより、FIFO や DualPI2 などのボトルネックキュー管理方式の違いが、輻輳下における学習および制御性能の維持にどう寄与するかを明らかにする。

2. 関連研究

前章で述べたとおり、本研究では、L4S を適用した高品質なネットワークを用意したうえで、倒立振り子を対象とした遠隔強化学習の環境構築を行う。以下では、L4S と倒立振子の遠隔制御についての関連研究を説明する。

2.1 超低遅延ネットワークアーキテクチャ L4S

L4S の中核となるアイデアは、エンドホストとネットワーク機器が協調して遅延の増加を事前に検出し、輻輳発生前に送信レートを微調整することである。その実現手段として、L4S では ECN を用いた輻輳検知シグナルが導入されている。L4S 対応のパケットは ECT(1) としてラベル付けされ、ECN マーキングが生じると、エンドホストはパケットロス待たずに輻輳を検知し、送信レートを調整する。さらに L4S 非対応の通信を Classic サービスと呼び、Classic のパケットには ECT(0) とラベ

ル付けすることで、ルータが Classic と L4S パケットを区別し、L4S パケットの優先送出を可能とする。

このアプローチにより、従来の輻輳制御が直面していたバッファ滞留によるキュー遅延を抑制することが可能となり、ネットワーク遅延を数ミリ秒以下のレベルに維持することが可能となる。また、スループットのスケーラビリティを犠牲にすることなく、安定した低遅延通信が提供される。

最近の研究では、Graff ら [17] は Cloud Gaming を対象として、従来のトラフィック制御方式と L4S 方式を比較評価した。実験では、L4S を適用したルータのキューに DualPI2 AQM [18] を用いており、classic トラフィック (大容量追求型) と「scalable トラフィック (低遅延重視型)」を異なるキューに分離することで低遅延重視型のトラフィックに対して低遅延かつ低損失な転送環境を実現した。実験結果では、Cloud Gaming を利用した環境における遅延などの QoS 指標の改善が報告されている。小倉ら [19], [20] は、AQM 機構として DualPI2, FQ-Codel, FQ-PIE, FQ-COBALT を用いた状況において、従来型トラフィック制御方式である CUBIC TCP を最大 500 本、L4S に用いる DCTCP を最大 500 本流した状況における性能評価を実施している。

2.2 倒立振子の遠隔制御

倒立振り子とは支点の上側に重心がある不安定な振り子のことである。倒立振り子は制御理論の代表的な教材として広く扱われており、英語では Inverted Pendulum, Pendulum Cart または Cart Pole と呼ばれている。制御理論の教材としての倒立振り子は次のように定義される。「直線レール上に配置された台車の上に振り子を設置し、振り子は直線レールと同じ向きだけに自由に動くように設置されている状況において、台車をレール上で小刻みに前進もしくは後進させることで棒を倒れないように制御する問題である。」PID 制御 (Proportional-Integral-Derivative Control) などの古典制御理論を用いた倒立振子の制御については [21] など研究されている。

近年、機械学習の一種である強化学習を利用した倒立振子の研究が始まっている [22]。一連の研究においては、OpenAI 社が作成した OpenAI Gym [23] という強化学習のソフトウェアプラットフォームが提供する、倒立振子の制御問題に関する Cartpole-v1 環境が幅広く利用されている。

倒立振子の制御においては、制御を行う「エージェント」と制御対象である「環境」の間の通信遅延が大きな影響を与える。Neto [24] は強化学習を一般の制御問題に応用する際の遅延に関する研究のサーベイを行っている。Ashraf ら [25] は、エージェントと環境間を無線通信で接続した倒立振子の制御を、PID 制御および強化学習で実施するシミュレーション評価を行っている。Purves ら [26] は Cartpole-v1 による強化学習による倒立振子の制御問題において、エージェントと環境間を有線イーサネットや WiFi で接続した際の評価を行っている。しかし、倒立振子の遠隔制御に L4S を適用した報告はみられていない。

3. 提案手法：遠隔強化学習への L4S の適用

前章で述べたとおり、本研究では、倒立振り子を対象とした遠隔強化学習の環境構築を行う。以下では、遠隔強化学習へ L4S

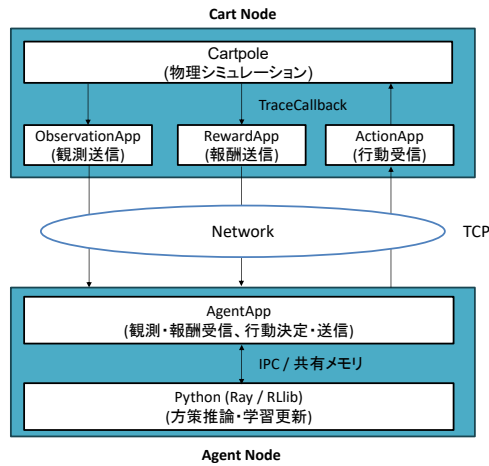


図 1: DEFIANCCE における遠隔強化学習環境

を適用するための提案手法について述べる。なお、本提案手法は、著者らがすでに [27] で発表しているが、同発表は提案のみを行ったもので、定量評価を行っていなかった。本稿では定量評価の結果等を記載するが、読者に対し十分な情報提供を行うために提案手法の概略を記す。

3.1 DEFIANCCE による遠隔強化学習の実現

強化学習 (Reinforcement Learning) は、エージェントが環境との試行錯誤を通じて、将来にわたる報酬の累積を最大化する行動方策を学習する機械学習の一分野である。教師データを用いて正解を学習する教師あり学習とは異なり、強化学習ではエージェント自身が行動の結果として得られる報酬を基に、望ましい行動を自律的に獲得する点に特徴がある。

本研究では、ネットワークを経由した遠隔強化環境を構築するために DEFIANCCE [16] を用いる。DEFIANCCE は ns-3 ネットワークシミュレーション環境上で強化学習を可能とするフレームワークである。DEFIANCCE では強化学習におけるアプリケーション間の通信を抽象化するチャンネルインタフェースが導入されており、TCP/IP を用いたネットワーク通信を扱うことができる。この設計により、通信帯域や通信遅延が強化学習の性能に与える影響を体系的に評価することが可能となる。ただし、現在の DEFIANCCE の実装では TCP/IP を用いたネットワーク通信を利用する場合、マルチエージェントおよびマルチカートはサポートされていない点には注意が必要である。

従来、ns-3 上で強化学習を評価する際には、ns3-gym [28] や ns3-ai [29] などのモジュールが用いられてきた。しかし、これらの手法では、強化学習エージェントの分散配置を伴う学習を十分に扱うことが困難であった。特に、強化学習の各構成要素 (観測、行動、報酬) がネットワーク上の異なるノードにエージェントとして配置した環境での評価が容易ではなかった。DEFIANCCE はこの課題に対し、図 1 のように、強化学習プロセスを構成する要素を観測アプリケーション、エージェントアプリケーション、行動アプリケーション、報酬アプリケーションの 4 種類の RL アプリケーションとして分離し、それぞれを ns-3 の任意のノードに配置可能とする設計を採用している。このモジュール化により、強化学習アルゴリズム自体を

変更することなく通信経路のみを変更した実験が可能となる。DEFIANCCE では Cartpole のシナリオが標準で用意されている。

また、本研究では強化学習の枠組みを実現する方法として、方策勾配法に基づくアルゴリズムである PPO (Proximal Policy Optimization) を採用する。PPO は方策の更新幅を制限することで学習の安定性を向上させる手法であり、複雑な制御問題に対しても比較的安定した学習が可能であることが知られている。この仕組みは、DEFIANCCE が内包している Ray/RLlib [30] の仕組みによって実現する。強化学習の進捗状況については、各エピソードの平均エピソード報酬 (最大 500) で評価する。

3.2 L4S によるネットワークの低遅延化

遠隔強化学習においては、各エージェントが観測した状態に基づいて行動を決定し、その結果が環境に反映されるまでの時間的遅延が学習および制御性能に大きな影響を与える。特に、遠隔制御環境では、通信遅延やジッタにより、エージェントが取得する状態情報が実際の環境状態と乖離する可能性がある。

この課題に対して、本研究では低遅延通信を実現するためのネットワークアーキテクチャとして L4S を用いる。L4S を適用したネットワークでは、L4S トラフィックと従来のトラフィックを論理的に分離し、低遅延を要求するトラフィックに対して優先的かつ低遅延なキュー管理を行う。これによりキュー遅延を数ミリ秒以下に抑えることが可能となる。本研究では、遠隔強化学習における状態、報酬、行動のデータをやり取りする通信経路に L4S を適用する。これにより遠隔強化学習に用いられるデータが滞りなく送受信されることで、エージェントが効率的に行動を決定できるようになることが期待される。

低遅延通信を実現するためのボトルネックキュー管理方式として、L4S の標準的な AQM フレームワークとみなされている DualPI2 [18] を導入する (図 2)。DualPI2 では、ECT(1) フラグを持つ L4S パケットを L4S キューに、Classic パケットを Classic キューに格納する。いずれかのキューからパケットを取り出す際、PI2 AQM は ECN マーキングを行い、その後、L4S キューに高い優先順位が与えられた条件付き優先度スケジューラが、どのパケットを送信するかを決定する。また、DualPI2 と性能を比較するために FIFO を導入する。FIFO は一つのキューを持ち、キューに届いた順番にパケットを送出する仕組みであるため、キュー遅延は比較的大きくなる傾向がある。

低遅延通信を適用する遠隔制御フローには DCTCP [31] の輻輳制御アルゴリズムを適用する。DCTCP は、ECN マーキングの頻度に応じて送信ウィンドウを段階的に調整する輻輳制御アルゴリズムであり、L4S アーキテクチャにおける低遅延フローの実現に不可欠である。DCTCP のパケットを L4S による優先パケットにするために ECT(1) マークを付与してパケットを送信する。一方で、競合する背景フローの TCP 輻輳制御に CUBIC を適用し、ルータで DualPI2 などの AQM 機構を導入することでフロー分離を実現する。CUBIC は Linux に標準搭載されている輻輳制御アルゴリズムである。この設計により、遠隔制御フローのみに L4S 対応の低遅延 TCP 設定を適用し、背景フローには従来の CUBIC を使用するというフロー種別に応じた輻輳制御の使い分けを可能とする。ネットワークの稼働状

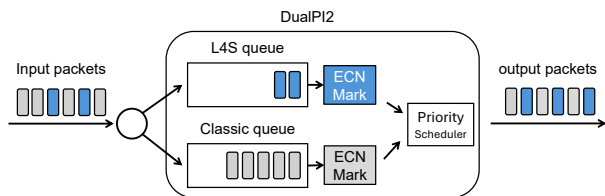


図 2: DualPI2 の概要

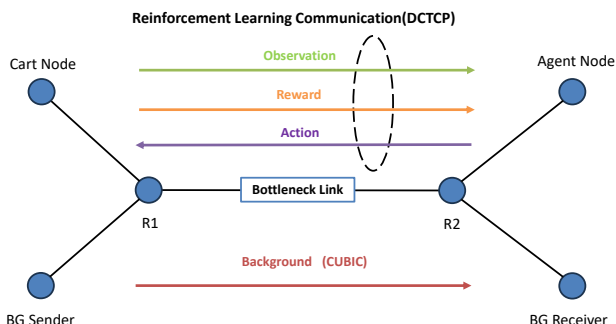


図 3: 遠隔強化学習の評価に用いるネットワークトポロジ

況については、キュー遅延およびスループットで評価する。

4. 性能評価

前章では、倒立振子を対象とした遠隔強化学習の環境を構築し、ネットワークにL4Sを適用する提案手法について述べた。本章では、前章で述べた遠隔強化学習環境を実装し、L4Sによる低遅延化が遠隔強化学習の学習効率および制御性能に与える影響の定量的な評価を行う。

4.1 評価環境

評価環境は、ns-3.47に含まれるDEFIANCEを用いて構築する。また、ネットワーク環境や強化学習環境はCartpoleのシナリオはDEFIANCEに付属するbalance2-scenario.ccをベースとする。OSはUbuntu 24.04 LTSであり、ハードウェアは、Intel Core i5-10500H CPU @ 2.50GHzである。

ネットワーク環境については、DEFIANCEに付属するbalance2-scenario.ccではLTE (Long Term Evolution)で構成されているが、本研究ではこれをPoint-to-Point (P2P)に置き換えて、図3に示すネットワークトポロジを構成する。遠隔制御の制御対象であるカートはCart Nodeに、エージェントはAgent Nodeに配置し、競合フローはBG SenderとBG Receiverに配置する。この結果、遠隔制御(観測: observation, 報酬: reward, 行動: action)に用いる3本のフローと競合フローは、ボトルネックリンクを共有する。本実験では、競合フローの通信を0秒に、遠隔制御フローの通信を0.5秒に開始している。

次に、ボトルネックリンクの帯域は10Mb/s、アクセスリンクの帯域は1Gb/s、伝搬遅延は0.5msとする。伝搬遅延により定まる最低RTTを30msまたは50msとするため、ボトルネックリンクの伝搬遅延は14msまたは24msとする。以降では、30msまたは50msのRTTのことを設定RTTと記す。ボトルネックキュー管理方式であるDualPI2またはFIFOはルータR1に設置し、そのバッファサイズは125,000Byte (125,000Byte / 1,500Byte = 約83パケット長)とする。なお、本研究では、バッ

ファサイズの算出には、遠隔制御に大きな遅延を与えるために大きめの値として2BDP (10Mb/s × 50ms × 2 = 125,000Byte)を用いているが、遠隔強化学習に最適なバッファサイズに関する今後の追加調査が必要である。なお、本論文のネットワークでは伝搬遅延により定まる最低RTTを30ms, 50msとしたが、これは一般的なインターネットの遅延時間を想定したものである。

また、強化学習ではエージェントと環境間で観測 (observation), 報酬 (reward), 行動 (action) の3 TCP フローを利用するが、このうち行動のTCPフローのみ情報の伝達の向きが逆になっている(図3参照)。本研究では、観測と報酬の2つのTCPフローを妨害する方向にのみ背景フローを挿入している。3フローのどれが遅延の影響を受けやすいかなど、今後の検討が必要となる。

強化学習環境については、DEFIANCEに付属するbalance2-scenario.ccを用いるが、標準的なCartpole-v1とは次の点に違いがあることが確認されている。初期状態は、DEFIANCEでは位置が(-0.1~0.1)の間でランダム、角速度が(-0.02~0.02)の間でランダム、位置が0、速度が0であるが、Cartpole-v1では全て(-0.05~0.05)の間でランダムである。Poleの長さはDEFIANCEでは1であるが、Cartpole-v1では0.5である。Terminatedの条件は、DEFIANCEでは角度が±0.418 rad、位置が±4.8のときであるが、Cartpole-v1では角度が±0.209 rad、位置が±2.4のときである。全体的にDEFIANCEは条件が緩和されているが、これはネットワーク経路で遠隔強化学習をすることの影響を抑制するための措置と考えられる。こうした差異を考慮したうえで、PPOのパラメータについては、学習率(lr)が0.0003、割引率(γ)が0.99、バッチサイズが4000、学習イテレーション数が100に設定している。

4.2 ネットワーク性能評価結果

各指標の平均値は1秒から5秒までのものを取得し、計算している。実行が6秒間しかないのは、DEFIANCEでは1回の強化学習エピソードが1回のns-3シミュレーション実行に対応し、学習はこの短いエピソードを多数反復することで進行するためである。本評価の1エピソードは、まず背景トラフィック(CUBIC)が0秒から流れ始め、ボトルネックが定常的な輻輳状態に達した1秒時点で制御対象の強化学習フローが参入し、以降10ms間隔で観測・行動をやり取りする500ステップ(5秒間)の制御区間で構成される。したがって1エピソードの総長は約6秒(背景トラフィックの立ち上げ1秒+制御区間5秒)となる。平均値の集計区間を1秒から5秒としたのは、背景トラフィックの立ち上がり過渡とエピソード終端の影響を除いた定常区間に相当するためである。なお、キュー遅延およびスループットの時系列図は、学習が十分に進行した段階における代表的な1エピソードのものを示している。

図4にボトルネックにおけるキュー遅延の評価結果を示す。(a),(c)のFIFOでは、観測(observation, 緑線)・報酬(reward, オレンジ線)、背景(background, 赤線)の3フローの遅延がほぼ一致し(約11~12ms)、強化学習フローは背景のキュー遅延増大の影響を強く受けていることがわかる。また設定RTTを30から50msへ増やすと、FIFOでは平均遅延が約11から12msへ増加した。一方、(b),(d)のDualPI2では、強化学習フロー(観

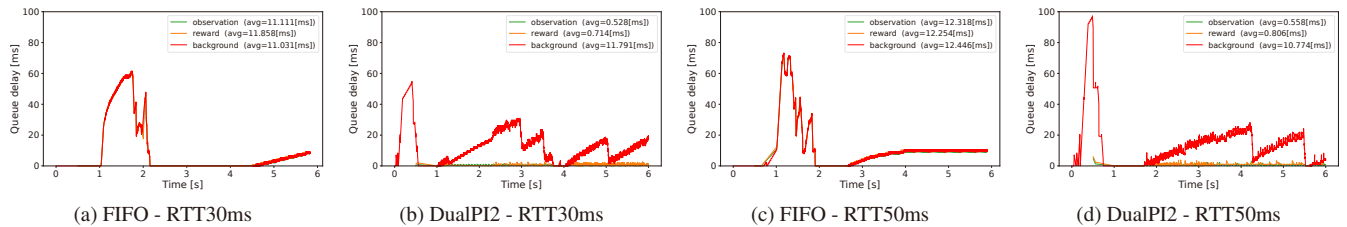


図 4: キュー遅延の時間変化

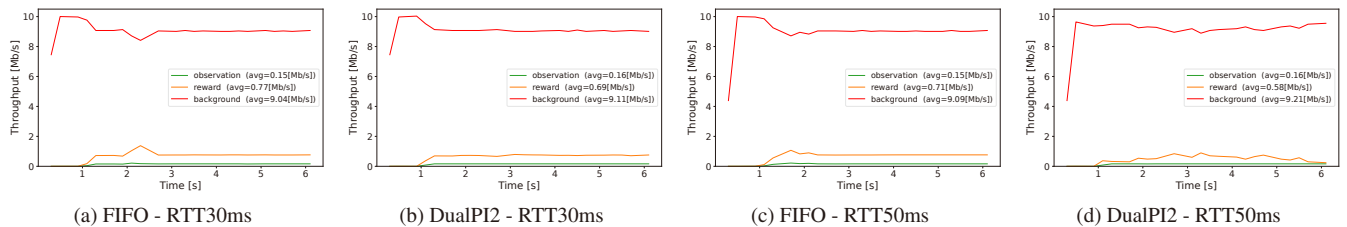


図 5: スループットの時間変化

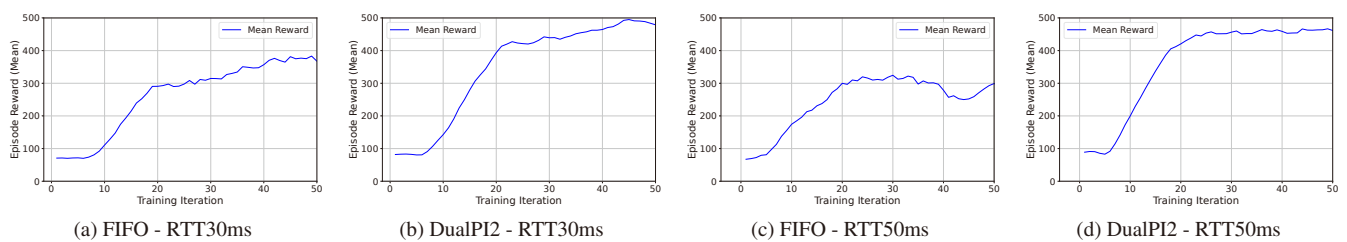


図 6: 遠隔強化学習における報酬の推移

測・報酬)のキュー遅延が設定 RTT にらず 1ms 未満に保たれ、同一ボトルネックを共有する背景トラフィック (約 11ms) より 10ms 以上小さかった。

図 5 にスループットの評価結果を示す。スループットには方式・設定 RTT にらずほとんど差が現れなかった。背景 CUBIC フローのスループットは、FIFO・DualPI2 のいずれでも、また設定 RTT 30/50ms のいずれでも約 9Mb/s でほぼ一定であった。遠隔制御フロー (観測・報酬) のスループットは合計で約 0.7~0.9Mb/s (観測約 0.15Mb/s、報酬約 0.6~0.8Mb/s) であった。つまり、全条件でリンク利用率は 97% を超えていた。

4.3 強化学習性能評価結果

図 6 に各設定 RTT の学習曲線を示す。いずれの条件でも報酬は初期の約 70~90 から上昇し、学習が進行した。(b),(d) の DualPI2 では設定 RTT30/50ms のいずれでも報酬が高い水準に収束し (約 485/460)、設定 RTT30ms では上限 500 に近い値に達した。一方 (a),(c) の FIFO では収束値が低く (約 375/300)、特に設定 RTT50ms では学習曲線が非単調で、途中で報酬が一時的に約 250 まで低下するなど不安定であった。DualPI2 は設定 RTT 増加に対する収束値の低下が小さかったのに対し、FIFO では設定 RTT を増加させることで収束値が明確に低下した。

4.4 考察

まず、遅延 (図 4) に最も顕著な差が現れた。(a),(c) の FIFO では制御フロー (観測・報酬) が背景 CUBIC フローと単一のキューを共有するため、CUBIC が形成するバッファブローからの影響を直接受けてしまい、制御フローの遅延は背景とほぼ同一の約 11/12ms (設定 RTT30/50ms) に達した。さらに設定

RTT の増加に伴い CUBIC の定常キューが深くなり、制御フローの遅延も微増 (約 11 → 12ms) した。すなわち FIFO では、遅延に敏感な遠隔制御フローが背景トラフィックの輻輳からまったく保護されない。これに対し DualPI2 では、(b),(d) にあるように、同じ背景トラフィック下でも制御フローの遅延が設定 RTT にらず 1ms 未満 (約 0.5~0.8ms) に保たれた。DualPI2 の二重キュー構造により、ECT(1) を付与した制御フローを浅いマーケティング閾値 (1ms) の L4S キューへ隔離し、CUBIC が形成する Classic キューの滞留から切り離す。さらに PI2 AQM が Classic キューを目標遅延付近に制御するため、CUBIC 自身の遅延 (約 11ms) も設定 RTT にらず概ね一定に抑えられる。結果として制御フローと背景トラフィックの遅延の間に 1 桁以上の差が生じ、ECT(1) を付与した制御フローのみが選択的に低遅延に保たれた。

スループット (図 5) は両方式・全設定 RTT でほぼ同一であった。CUBIC が約 9Mb/s でボトルネックの過半を占め、2 つの制御フロー (観測、報酬) は強化学習のために必要なデータ量が少ないため低レートとなっている。すなわち帯域配分の観点では DualPI2 と FIFO に実質的な差はない。両者の本質的な差異は帯域ではなく遅延にあり、同一のスループットを維持したまま DualPI2 は制御フローの遅延のみを大きく低減するという点が、L4S の価値を端的に示している。

最終的な制御性能 (図 6) は、(b),(d) の DualPI2 では遠隔制御フローのキュー遅延がほとんど 0ms に抑えられるため、観測から行動までのフィードバックが遅延なく伝わり、制御が安定して高い報酬に収束したと考えられる。一方 (a),(c) の FIFO では

10 数 ms のキュー遅延が設定 RTT に上乗せされ、さらにキュー遅延の変動が大きいこともあり、強化学習による制御が不安定化している（特に設定 RTT50ms）。また、スループットは全条件でほぼ同一であったことから、この報酬差は帯域の差ではなく遅延の差に起因すると考えられる。すなわち、L4S/DualPI2 による遅延の分離が、遠隔強化学習の制御品質と学習の安定性を有効に改善することを示している。この効果は設定 RTT が大きいほど顕著であり、L4S の恩恵が大きいといえる。

なお本評価は各条件で単一のランダムシードによるものである。一般に強化学習は乱数の値による性能差が大きいため、ランダムシードを変えて複数回実行し、統計的な分析を行うことが通例であるため、今後の追加調査が必要と言える。特に DualPI2 の設定 RTT(30/50ms) の性能差は小さく、乱数の影響による結果の可能性がある。一方、DualPI2 と FIFO の差、および設定 RTT50ms での FIFO の劣化は乱数の影響のみとは考えにくく、本質的な傾向であると考えられる。

5. おわりに

本研究では、通信品質が遠隔強化学習の制御性能に与える影響を明らかにすることを目的として、DEFIANCE を用いた遠隔強化学習の環境を構築した。評価の結果、ネットワーク遅延が増大すると遠隔強化学習の性能は低下することが判明したが、DualPI2 を用いた高品質なネットワークを利用することで、その影響は最小限に抑制できることを示した。

今後は、より複雑な強化学習シナリオを用いた検証や、多数の背景フローが存在する厳しいネットワーク条件での評価を通じて、遠隔強化学習に適したネットワーク条件を調査する。

文 献

- [1] A. Mahmood, L. Beltramelli, S. Fakhru Abedin, S. Zeb, N.I. Mowla, S.A. Hassan, E. Sisinni, and M. Gidlund, "Industrial IoT in 5G-and-beyond networks: Vision, architecture, and design trends," *IEEE Transactions on Industrial Informatics*, vol.18, no.6, pp.4122–4137, Sept. 2022.
- [2] I. Behnke and H. Austad, "Real-time performance of industrial IoT communication technologies: A review," *IEEE Internet of Things Journal*, vol.11, no.5, pp.7399–7410, March 2024.
- [3] K. Nichols, V. Jacobson, A. McGregor, and J. Iyengar, "Controlled Delay Active Queue Management," IETF RFC 8289, Jan. 2018. <https://www.rfc-editor.org/info/rfc8289>
- [4] K. Nichols and V. Jacobson, "Controlling queue delay," *ACM Queue*, vol.10, no.5, pp.1–15, May 2012.
- [5] 妹尾龍馬, 平澤崇佳, 森弘樹, 中務諭士, 高橋賢, "超低遅延サービスの提供を見据えた遅延抑制技術 CoDel の適用性評価," *信学技報*, vol.119, no.460, NS2019-199, pp.129–134, March 2020.
- [6] R. Pan, P. Natarajan, F. Baker, and G. White, "Proportional Integral Controller Enhanced (PIE): A Lightweight Control Scheme to Address the Bufferbloat Problem," IETF RFC 8033, Feb. 2017.
- [7] A.A. Mahawish and H.J. Hassan, "Survey on: A variety of AQM algorithm schemas and intelligent techniques developed for congestion control," *Indonesian Journal of Electrical Engineering and Computer Science*, vol.23, no.3, pp.1419–1431, Sept. 2021.
- [8] J. Luo, J. Jin, and F. Shan, "Standardization of low-latency TCP with explicit congestion notification: A survey," *IEEE Internet Computing*, vol.21, no.1, pp.48–55, Jan. -Feb. 2017.
- [9] F. Baker and G. Fairhurst, "IETF Recommendations Regarding Active Queue Management," IETF RFC 7567, July 2015.
- [10] R. Al-Saadi, G. Armitage, J. But, and P. Branch, "A survey of delay-based and hybrid TCP congestion control algorithms," *IEEE Com-*

- mun. Surveys & Tutorials*, vol.21, no.4, pp.3609–3638, Nov. 2019.
- [11] B. Briscoe, K.D. Schepper, M. Bagnulo, and G. White, "Low Latency, Low Loss, and Scalable Throughput (L4S) Internet Service: Architecture," IETF RFC 9330, Jan. 2023.
- [12] S. Floyd, D.K.K. Ramakrishnan, and D.L. Black, "The Addition of Explicit Congestion Notification (ECN) to IP," IETF RFC 3168, Sept. 2001.
- [13] K.D. Schepper and B. Briscoe, "The Explicit Congestion Notification (ECN) Protocol for Low Latency, Low Loss, and Scalable Throughput (L4S)," IETF RFC 9331, Jan. 2023.
- [14] G. Pan, S. Xu, and P. Jiang, "Optimizing 5G-advanced networks for time-critical applications: The role of L4S," *IEEE Wireless Communications*, vol.32, no.2, pp.230–237, April 2025.
- [15] "The network simulator ns-3," <https://www.nsnam.org/>. Accessed: 2025/12/19.
- [16] A.-B. Burmeister, J. Dittrich, A.M. Fennig, T. Rothe, and R. Singh, "DEFIANCE: Simulate reinforcement learning deployments in realistic network scenarios," *Proc. ACM Int'l Conf. on ns-3 (ICNS3'25)*, pp.54–61, Aug. 2025.
- [17] P. Graff, et al., "Improving cloud gaming traffic QoS: A comparison between class-based queuing policy and L4S," *Proc. IEEE Network Traffic Measurement and Analysis Conference (TMA)*, pp.1–10, May 2024.
- [18] K.D. Schepper, B. Briscoe, and G. White, "Dual-Queue Coupled Active Queue Management (AQM) for Low Latency, Low Loss, and Scalable Throughput (L4S)," IETF RFC 9332, Jan. 2023.
- [19] T. Ogura, R. Mitsuhashi, S. Sunahara, K. Iida, and Y. Takai, "Comparative study of AQM frameworks in L4S networks under multi-flow traffic," submitted for publication, May 2026.
- [20] 小倉拓海, 三橋力麻, 砂原 悟, 飯田勝吉, 高井昌彰, "FQ-Codel を用いた超低遅延ネットワーク L4S の性能評価," *信学技報*, vol.125, no.385, NS2025-280, pp.351–356, March 2026.
- [21] M. Landry, S.A. Campbell, K. Morris, and C.O. Aguilar, "Dynamics of an inverted pendulum with delayed feedback control," *SIAM J. Applied Dynamical Systems*, vol.4, no.2, pp.333–351, April 2005.
- [22] R. Özalp, N.K. Varol, B. Taşci, and A. Uçar, "A review of deep reinforcement learning algorithms and comparative results on inverted pendulum system," *Machine Learning Paradigms: Advances in Deep Learning-based Technological Applications*, eds. by G.A. Tshirintzis and L.C. Jain, pp.237–256, Springer, July 2020.
- [23] G. Brockman, et al., "OpenAI gym," *arXiv preprint arXiv:1606.01540*, pp.1–4, June 2016.
- [24] A.A. Neto, "Reinforcement learning for control systems with time delays: A comprehensive survey," *arXiv preprint arXiv:2602.00399*, pp.1–30, Jan. 2026.
- [25] K. Ashraf, et al., "Co-design of a wireless networked control system for reliability and resource-efficiency," *Proc. IEEE Biennial Baltic Electronics Conf. (BEC)*, pp.1–7, Oct. 2024.
- [26] C. Purves and P. Lio', "CALF: Communication-aware learning framework for distributed reinforcement learning," *arXiv preprint arXiv:2603.12543*, pp.1–19, March 2026.
- [27] 奥村 丈, 三橋力麻, 飯田勝吉, 高井昌彰, "超低遅延ネットワークアーキテクチャ L4S を用いた遠隔強化学習の環境構築," *信学技報*, vol.125, no.320, NS2025-198, pp.37–41, Jan. 2026.
- [28] P. Gawlowicz and A. Zubow, "ns-3 meets OpenAI Gym: The playground for machine learning in networking research," *Proc. Int'l ACM Conf. Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWIM'19)*, pp.113–120, Nov. 2019.
- [29] H. Yin, P. Liu, K. Liu, L. Cao, L. Zhang, Y. Gao, and X. Hei, "ns3-ai: Fostering artificial intelligence algorithms for networking research," *Proc. ACM Workshop on ns-3*, pp.57–64, June 2020.
- [30] The Ray team, "RLlib: Industry-grade, scalable reinforcement learning," <https://docs.ray.io/en/master/rllib/index.html>. Accessed: 2025/12/23.
- [31] S. Bensley, D. Thaler, P. Balasubramanian, L. Eggert, and G. Judd, "Data Center TCP (DCTCP): TCP Congestion Control for Data Centers," IETF RFC 8257, Oct. 2017.