Reinforcement Learning-Based Transmission Opportunity Control in Multi-Link Networks

Ferdinando Marrone
University of Naples Federico II
Naples, Italy
ferdinando.marrone@unina.it

Nicola D'Ambra
University of Naples Federico II
Naples, Italy
ni.dambra@studenti.unina.it

Pasquale Imputato
University of Naples Federico II
Naples, Italy
pasquale.imputato@unina.it

Stefano Avallone
University of Naples Federico II
Naples, Italy
stefano.avallone@unina.it

***Abstract*—The IEEE 802.11 standard has evolved with the introduction of Wi-Fi 7 and its Extremely High Throughput (EHT) amendment, which incorporates Multi-Link Operation (MLO) to enable concurrent transmissions across multiple frequency links. While MLO can significantly improve throughput and reduce latency under favorable conditions, its performance may degrade in the presence of irregular or high traffic loads due to traffic imbalance and increased collision probability. TCP flows are particularly sensitive to these effects, as congestion on individual links triggers congestion control mechanisms that reduce sending rates and impair overall throughput. This work investigates the impact of MLO on TCP performance under varying levels of UDP congestion through ns-3 simulations. Results highlight significant performance instability under unbalanced load conditions, motivating the need for adaptive traffic management. To address these challenges, we propose a Reinforcement Learning (RL)-based Transmission Opportunity (TXOP) Random Discard strategy that dynamically adjusts discard probabilities according to network conditions. Simulation results show that the proposed method improves TCP performance while preserving UDP throughput, providing a flexible and effective solution for traffic balancing in multi-link Wi-Fi 7 networks.**

***Index Terms*—Network Testing; Performance Evaluation; TCP; RL; MLO; ns-3**

I. Introduction

Modern applications increasingly demand higher throughput and lower latency. In response, IEEE 802.11be (Wi-Fi 7) introduces Multi-Link Operation (MLO), which enables simultaneous communication over multiple wireless links. MLO can provide link aggregation, load balancing, latency reduction, and increased reliability.

However, the benefits of MLO are highly context-dependent. When links are similarly loaded and channel conditions are favorable, MLO can substantially improve throughput and latency compared to single-link operation [1]. In congested or unevenly loaded scenarios, its performance depends on how traffic is distributed across the available links. Without effective load balancing, increased contention on specific links can degrade throughput and offset the expected latency gains [2], [3]. Traffic allocation has therefore emerged as a key factor for efficient MLO operation.

The contribution of Nicola D'Ambra was strictly limited to work performed as part of their Master's thesis project, with no subsequent involvement.

The interaction between MLO and transport-layer protocols is also relevant. In particular, TCP is sensitive to packet losses, retransmissions, and round-trip time variations, and its congestion control may lead to performance degradation under certain MLO conditions [4], [5].

To the best of our knowledge, existing works have largely considered simplified or unrealistic traffic scenarios and have only pointed out the need for further investigation of asymmetric traffic conditions. In addition, most available studies rely on numerical simulations that do not capture the behavior of the full protocol stack.

This work investigates the impact of MLO on TCP traffic under heterogeneous link conditions and competing UDP flows using realistic ns-3 simulations. We show that TCP flows may experience severe performance degradation in the presence of interfering UDP traffic, in some cases performing worse than in a single-link configuration. To address this issue, we propose the TXOP Random Discard mechanism, in which stations probabilistically relinquish TXOPs for best-effort traffic, together with an RL-based strategy that dynamically adapts the discard probability to current network conditions. The proposed approach improves TCP throughput and latency, enables MLO to outperform SLO even under moderate and heavy traffic loads, and preserves UDP stability, ensuring more balanced and efficient multi-link operation.

II. Methodology

The evaluation of TCP over MLO, and then of the proposed RL-based TXOP random discard strategy, is carried out through an extensive simulation campaign, conducted using the ns-3 network simulator. This setup allows comparing MLO performance against traditional Single Link Operation (SLO), investigating how it varies with link congestion.

Fig. 1 shows the network topologies used in the simulation campaign. The SLO topology in Fig. 1a represents the Wi-Fi 6 baseline, with a single 5 GHz link and an IEEE 802.11ax AP, DUT, and nine additional Wi-Fi 6 stations. MLO 1 in Fig. 1b extends this setup by upgrading the AP and DUT to Wi-Fi 7 and adding a second 5 GHz link, which is left free for AP–DUT transmissions. MLO 2 in Fig. 1c adds nine Wi-Fi 6 stations on the second link, generating competing UDP traffic and introducing congestion. In all topologies, the

979-8-3195-0167-7/26/$31.00 ©2026 IEEE

DUT and STAs are uniformly distributed within a circular area of 10 meters radius, with the AP placed at the center. While SLO stations remain static, the DUT follows a Gauss-Markov mobility model. This choice is motivated by the need to emulate realistic human movement, where inertia prevents instantaneous changes in direction or speed. By setting $\alpha = 0.9$ and a mean velocity between 1.0 and 1.3 m/s, we ensure a smooth and continuous evolution of the wireless links, challenging MLO adaptation mechanisms to react to realistic signal fluctuations.

Traffic is generated using OnOff applications, with 0.1 s Off periods and exponentially distributed On periods with a mean of 0.1 s. The TCP traffic rate generated by the DUT is set to 30% of the PHY rate, while the total UDP load on link 1 is 30% and that on link 2 is varied between 30%, 50%, and 70%. Table I outlines the specific scenarios under test, combining the SLO, MLO 1, and MLO 2 topologies with the various traffic loads investigated on link 2.

Each scenario is simulated 10 times with different random seeds, with a duration of 10 seconds. The DUT generates TCP traffic in the best-effort access category.

			Link 1		Link 2	
	Topology	**TCP Rate**	**STAs**	**UDP Load**	**STAs**	**UDP Load**
s1	SLO	30%	9	30%	0	0%
s2	MLO 1	30%	9	30%	0	0%
s3	MLO 2	30%	9	30%	9	30%
s4	MLO 2	30%	9	30%	9	50%
s5	MLO 2	30%	9	30%	9	70%

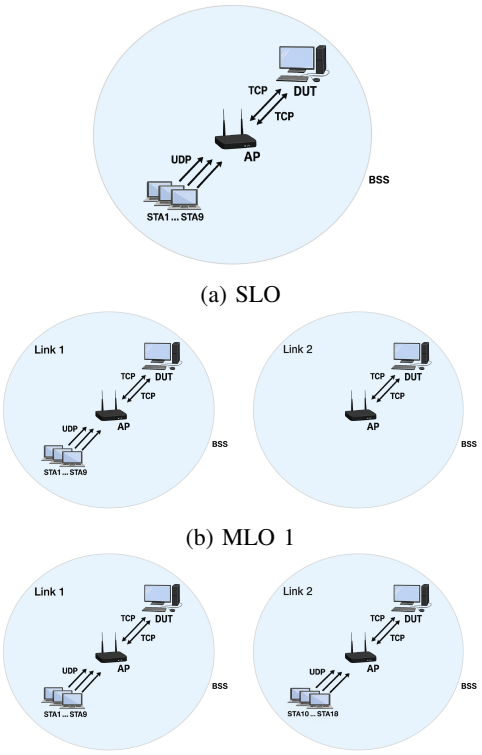
TABLE I: Simulation scenarios for TCP over MLO.

III. Analysis of TCP Performance over MLO

To understand the interaction between TCP and MLO, we first evaluate the baseline performance of TCP flows without introducing any traffic management mechanisms. This preliminary analysis is crucial, as it reveals whether MLO alone is sufficient to improve end-to-end performance or if its benefits are counterbalanced by transport-layer dynamics.

The results in Fig. 2 show that multi-link operation provides clear benefits only when the additional link is lightly loaded or completely idle. Fig. 2a (2b) shows the throughput distribution for TCP downlink (uplink) flow in the five scenarios. Results reveal that increasing the offered load on link 2 (in scenarios s3, s4, and s5) leads to performance degradation of the DUT. In particular, when the load on link 2 reaches or exceeds 50% (in s4 and s5), MLO under TCP performs worse than the single-link case. The latency characteristics further highlight this trend. The 95th percentile of latency of downlink (uplink) TCP flow is reported in Fig. 2c (2d). MLO reduces latency when link 2 is lightly loaded (in s3), but as soon as the load increases (in s4 and s5), latency grows beyond the single-link baseline.

To further characterize transport-layer behavior, the TCP congestion window (cwnd) histograms shown in Fig. 3 reveal that high-load scenarios (s4 and s5) are associated with a higher probability of low cwnd values. This indicates frequent

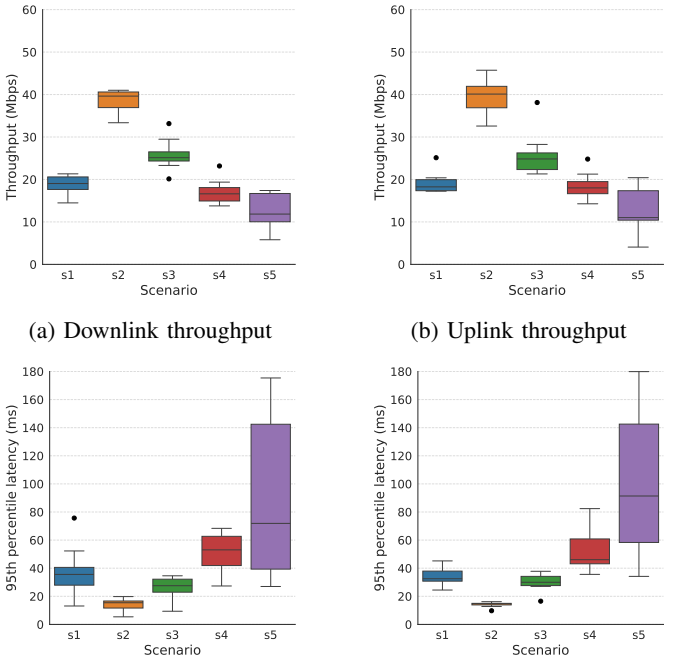


(a) SLO

(b) MLO 1

(c) MLO 2

Fig. 1: Network topologies used in the simulation campaign.



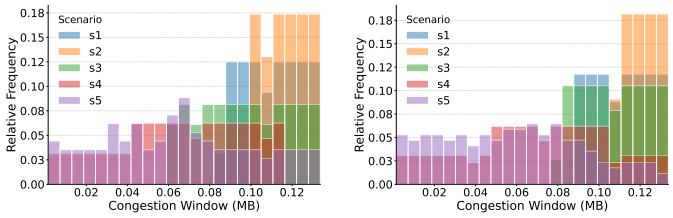
(a) Downlink throughput

(b) Uplink throughput

(c) Downlink 95th percentile latency

(d) Uplink 95th percentile latency

Fig. 2: Performance of TCP traffic.



(a) Downlink (b) Uplink

Fig. 3: Relative frequency of TCP congestion window values in downlink (a) and uplink (b).

window reductions, likely caused by TCP retransmission timeout (RTO) expiration, and explains the throughput degradation observed under unbalanced link loads.

IV. TXOP Random Discard with RL driven probability

The analysis of TCP performance under MLO has shown that medium-to-high load on one of the links can degrade performance compared to the single-link case, mainly due to increased latency caused by interfering UDP uplink flows. To mitigate this issue, we introduce *TXOP Random Discard*, a probabilistic transmission limiting mechanism in which a station may relinquish a TXOP obtained for best-effort traffic according to a configurable discard probability. The probability is the same for all STAs on a link, except for the DUT, and can differ across links. The mechanism can selectively limit interfering traffic without requiring higher-layer information.

The mechanism is implemented in ns-3 by adding the *TxopDiscProb* attribute to *WifiMac*. During *StartTransmission* in *QosFrameExchangeManager*, each BE TXOP is discarded according to the configured link-specific probability. The implementation and simulation scripts are publicly available[1].

We further propose an RL-based strategy, denoted as RL TXOP Random Discard, in which a Proximal Policy Optimization (PPO) agent dynamically adjusts the discard probability of each link according to observed network conditions. PPO was selected for its stable policy updates and suitability for dynamic wireless environments. The agent was trained with a learning rate of 0.0001, a batch size of 950, and a minibatch size of 95, while the remaining parameters followed the default RLlib PPO configuration [6]. Training consisted of 30 independent simulations of 8 seconds for each MLO scenario (Scenarios from s2 to s5 in Table I).

At each time step t, the environment state is defined as

$$s_t = \left[C_1(t), C_2(t), T_{\text{TCP-DL}}(t), T_{\text{UDP-UL}}(t)\right]$$

where $C_1(t)$ and $C_2(t)$ are the collision percentages on the two links, and $T_{\text{TCP-DL}}(t)$ and $T_{\text{UDP-UL}}(t)$ are the TCP downlink and mean UDP uplink throughput, respectively. Collision percentages are obtained from MAC-layer transmission failures inferred from missed ACKs and Block ACKs over a 100 ms observation window.

[1]https://gitlab.com/ferdinando.marrone/txop-random-discard-v2

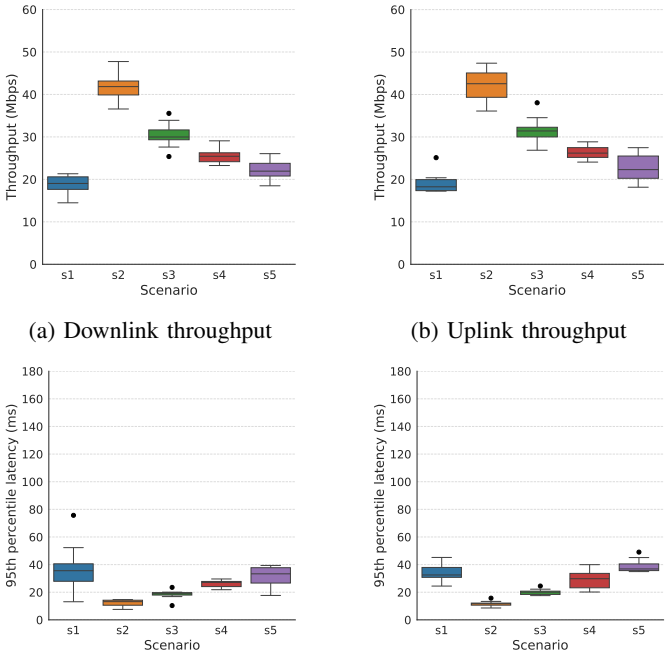
The action is defined as

$$a_t = \left[d_1(t), d_2(t)\right]$$

where $d_1(t)$ and $d_2(t)$ are the discard probabilities applied to the two links. The action space is discretized into values from 0% to 80% with 10% increments, avoiding complete blocking of UDP traffic.

The reward function balances TCP and UDP throughput with collision reduction:

$$R_t = w_T \cdot T_{\text{TCP-DL}}(t) + w_T \cdot T_{\text{UDP-UL}}(t) - w_C \cdot C_1(t) - w_C \cdot C_2(t)$$



(a) Downlink throughput (b) Uplink throughput

(c) Downlink 95th percentile latency (d) Uplink 95th percentile latency

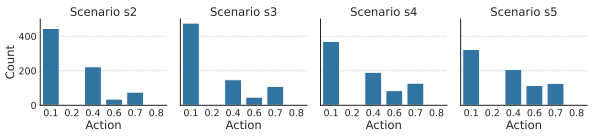
Fig. 4: Performance of TCP traffic under the RL TXOP Random Discard approach

The weights were selected through a preliminary sensitivity analysis, resulting in $w_T = 0.5$ and $w_C = 15$ which provided the best overall trade-off between throughput, collision reduction, and fairness. All metrics refer to the most recent observation window. Uplink TCP throughput is not explicitly included in the reward, but is indirectly affected through the collision-related terms.

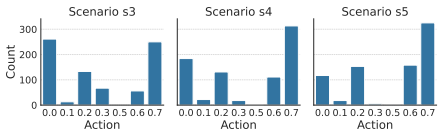
V. Performance evaluation

Using the selected configuration ($w_T = 0.5$, $w_C = 15$), we evaluate the RL TXOP Random Discard mechanism across all scenarios. As shown in Figs. 4a and 4b, the proposed approach substantially improves both TCP downlink and uplink throughput compared to the baseline configuration. In particular, MLO scenarios with the RL TXOP Random Discard mechanism never perform worse than the SLO case, confirming the effectiveness of dynamically adjusting the discard probability to mitigate UDP aggressiveness. In terms of latency, the

improvement is even more evident: as shown in Figs. 4c and 4d, all MLO configurations achieve 95th percentile latency values lower than those of a single-link setup. These results demonstrate that the reinforcement learning approach successfully adapts to varying network conditions, effectively balancing throughput and delay while consistently improving performance in MLO scenarios.



(a) Link 1



(b) Link 2

Fig. 5: Distribution of selected actions for each scenario under RL TXOP Random Discard. Each bar represents the count of occurrences of a given discard probability applied to Link 1/Link 2

The histograms in Fig. 5 show the frequency of each TXOP Random Discard value applied by the RL agent to the two links across the different traffic scenarios. For link 1, the agent maintains a predominantly conservative policy across all scenarios (s2 through s5), with a consistent peak frequency at a discard probability of 0.1. In contrast, the behavior on link 2 reveals a high degree of adaptability to network load: while scenario s3 shows a more distributed selection, the transition to high-load conditions in s4 and s5 triggers a decisive shift toward higher discard probabilities, specifically centering around 0.7. This trend demonstrates the agent's ability to dynamically mitigate increased contention on the second link. By selectively increasing the TXOP Random Discard value under heavy load, the RL agent effectively prioritizes TCP traffic, explaining the improved TCP performance observed in these scenarios.

At the same time, this behavior leads to a slight increase in UDP latency, as reported in Table II, since higher discard probabilities on the interfering UDP link occasionally delay the transmission opportunities of UDP packets. Specifically, while percentage increases reach up to 27.50% on Link 1 (scenario s5), the absolute 95th percentile latency remains below 51 ms. For reference, ITU-T G.114 [7] recommends a maximum end-to-end delay of 150 ms for conversational applications, indicating that the additional latency introduced by the proposed mechanism remains within a limited portion of the overall latency budget. Conversely, on Link 2, the mechanism reduces latency by up to 10.27% in scenario s4. UDP throughput remains stable across all scenarios, confirming that the proposed approach improves TCP performance without significantly affecting the interfering UDP traffic.

Scenario	No-discard [ms]	RL-Discard [ms]	Δ (%)
Link 1			
s2	41.08	42.98	+4.63%
s3	42.75	50.55	+18.23%
s4	41.90	46.54	+11.09%
s5	38.41	48.97	+27.50%
Link 2			
s3	41.46	55.45	+33.74%
s4	194.90	174.88	-10.27%
s5	429.26	455.04	+6.01%

TABLE II: Comparison of the 95th percentile of latency values and percentage variations between baseline no-discard and RL-discard mechanism for UDP BE interfering traffic.

VI. Conclusions

This study shows that MLO with two links can adversely affect TCP traffic when one link experiences moderate to high UDP interference. The resulting latency increase can trigger TCP congestion control, reducing the sending rate and making MLO perform worse than a single-link configuration.

To address this issue, we proposed the TXOP Random Discard mechanism and an RL-based strategy that dynamically adjusts the discard probability according to network conditions. The proposed approach enables MLO to outperform SLO under moderate and heavy loads while preserving UDP performance, effectively mitigating contention between heterogeneous traffic flows.

Future work will explore multi-objective reward functions and alternative RL algorithms to improve the robustness and adaptability of the proposed approach under diverse traffic and network conditions.

References

[1] C. Chen, X. Chen, D. Das, D. Akhmetov, and C. Cordeiro, "Overview and Performance Evaluation of Wi-Fi 7," *IEEE Communications Standards Magazine*, vol. 6, no. 2, pp. 12–18, 2022, https://doi.org/10.1109/MCOMSTD.0001.2100082.

[2] M. Carrascosa-Zamacois, G. Geraci, L. Galati-Giordano, A. Jonsson, and B. Bellalta, "Understanding Multi-link Operation in Wi-Fi 7: Performance, Anomalies, and Solutions," in *2023 IEEE PIMRC*, 2023, pp. 1–6, https://doi.org/10.1109/PIMRC56721.2023.10293865.

[3] M. Alsakati, C. Pettersson, S. Max, V. N. Moothedath, and J. Gross, "Performance of 802.11be Wi-Fi 7 with Multi-Link Operation on AR Applications," in *2023 IEEE WCNC*, 2023, pp. 1–6, https://doi.org/10.1109/WCNC55385.2023.10118866.

[4] P. Imputato, S. Avallone, J. Kulshrestha, S. Mohanty, and M. Maity, "TCP Latency-Sensitive Applications Over MLO in Wi-Fi 7," in *2024 IEEE Symposium on Wireless Technology & Applications (ISWTA)*, 2024, pp. 136–139, https://doi.org/10.1109/ISWTA62130.2024.10652099.

[5] J. Kulshrestha, S. Mohanty, P. Imputato, S. Avallone, and M. Maity, "Is Multi-Link Operation of 802.11be TCP Friendly? Analysis and Solution," in *2024 IFIP Networking Conference*, 2024, pp. 168–176, https://doi.org/10.23919/IFIPNetworking62109.2024.10619851.

[6] Ray Team, "Rllib: Scalable reinforcement learning," https://docs.ray.io/en/latest/rllib/index.html, 2026, accessed: 2026-08-20.

[7] International Telecommunication Union, Telecommunication Standardization Sector (ITU-T), "Recommendation G.114: One-way transmission time," ITU-T, Recommendation G.114, 2003. [Online]. Available: https://www.itu.int/rec/T-REC-G.114