Is Multi-Link Operation of 802.11be TCP friendly? Analysis and Solution

Jagrati Kulshrestha
IIIT Delhi
India
jksjagrati@gmail.com

Srimant Mohanty
IIIT Delhi
India
srimant21207@iiitd.ac.in

Pasquale Imputato
University of Naples
Italy
p.imputato@gmail.com

Stefano Avallone
University of Naples
Italy
stavallo@gmail.com

Mukulika Maity
IIIT Delhi
India
mukulika@iiitd.ac.in

***Abstract*—IEEE 802.11be introduces MLO (Multi-Link Operation) where MLDs (Multi-Link Devices) can take advantage of the multiple available radios to concurrently transmit over frequency channels located in the 2.4, 5 and 6 GHz bands. As the flow is distributed across multiple links, there might be some interesting consequences of this on the performance, especially in the case of TCP traffic, as it involves the transmission of TCP ACKs in the reverse direction of TCP data packets. In this paper, we analyze the interaction of MLO with TCP using the latest 802.11be models supplied by network simulator NS-3 (version 3.40). As expected, throughput increases with multiple links; however, the latency suffers in many cases, specifically the tail latency. This is in contradiction to one of the key objectives of MLO, which is to improve latency. We then find out that the root cause is an increase in collisions with the number of links. We thus propose a solution where we map certain links to a specific traffic type. We observe that it reduces collisions and helps improve both throughput and latency.**

***Index Terms*—IEEE802.11be, WiFi7, Multi-Link Operation, Multi-Link Devices, TID-to-Link Mapping, TCP, Performance Analysis, Network Performance.**

I. Introduction

To meet the ever-increasing demands from WiFi networks, IEEE introduced 802.11be standard for the latest WiFi7 networks, also referred to as the Extremely High Throughput (EHT) networks [1], [2]. EHT networks can support applications like augmented reality, virtual reality, cloud gaming, etc., which have very high throughput and low latency requirements. One of the new and prominent features of WiFi7 is the introduction of Multi-Link Operation (MLO). Using MLO, a transmitter or receiver can utilize multiple links at the same time. These links can be on different bands like, 2.4, 5 and 6GHz, or they can also be on the same band. Devices using MLO feature are termed as Multi-Link Devices (MLDs).

MLO has three modes of operation:

1) **Simultaneous Transmit Receive (STR):** MLDs can operate on different links independently and can transmit and receive over multiple links simultaneously.
2) **Non-Simultaneous Transmit Receive (NSTR):** MLDs can either receive or transmit simultaneously but not both at the same time.
3) **Enhanced Multi-Link Single Radio (EMLSR):** MLDs can transmit or receive on only one link at a time, and the link is chosen dynamically.

ISBN 978-3-903176-63-8 © 2024 IFIP

Among all these, STR mode is expected to provide the best performance as it can utilize all the available links [3]–[6]. There are multiple prior works on the performance evaluation of MLO [3]–[14]. They show that MLO with STR mode increases the throughput with saturated traffic, and it reduces latency for unsaturated traffic in proportion to the number of links. Specifically, [7], [8] show that MLO can reduce the 95^{th} percentile latency by 78% and the average latency by 69%. However, all prior works consider only UDP (User Datagram Protocol) traffic for evaluating the performance gains of MLO compared to Single Link Operation (SLO). On the other hand, most of the Internet traffic is based on TCP (Transmission Control Protocol). TCP throws interesting challenges. Even if the TCP data is in downlink, the TCP ACK (Acknowledgement) will be in the uplink. Hence, the clients need to contend to transmit TCP ACKs. Further, as different links might have different link characteristics, they will incur different latency, which might lead to out-of-order packets. UDP remains unaffected by this. But, TCP is sensitive to out-of-order packets which can impact its performance.

In this paper, we study the interplay between TCP and MLO in WiFi 7. For this, we use 802.11be models provided by the open-source network simulator NS-3. We measure the performance in terms of throughput, latency, round trip time (RTT), etc. We evaluate with single link, double links, and triple links for both sparse (with 10 STAs) and dense (with 40 STAs) networks. We evaluate for two types of traffic (1) only Best Effort (BE) Access Category (AC) traffic and (2) three ACs - BE, VI (Video) and VO (Voice) traffic. In each traffic type, the STAs have traffic in - a) only downlink (DL), b) only uplink (UL), and c) both downlink and uplink (DL+UL) directions with both sparse and dense networks.

We observe that, as expected, the throughput increases with single to double to triple links. However, oftentimes, higher percentile TCP RTT and higher percentile application layer latency (E2E: End to End latency) increases with the increase in links. In some of the scenarios, especially where both DL and UL (DL+UL) traffic is present, and the traffic is of three ACs - BE (Best Effort), VI (Video), and VO (Voice), even the median RTT increases compared to SLO. This causes tail E2E latency to increase as well compared to SLO. In addition to RTT, the number of retransmissions and out-of-order packets also increases with MLO with the increase in the number of links. All these factors are detrimental to any network and thus restrict MLO from realizing its true potential. Providing low latency is one of the key objectives of MLO; however, observations are in complete contradiction to what MLO is expected to achieve.

Next, we perform a root cause analysis for the same. For this, we utilize relevant statistics such as transmission (Tx) failures and Physical Service Data Unit (PSDU) response timeouts generated during the simulation. We undercover the root cause behind such a latency increase is the increase in collisions as we move from single link to multiple links. This limits the performance gains with multiple links. We also observe that the PSDU response timeouts increase from single to double to triple links.

Having discovered the root cause, we design a solution to limit contention/collisions. Our solution utilizes TID (Traffic Identifier)-to-Link mapping. Here, a set of TIDs of different ACs (Access Categories) are mapped to specific links. Thus, the stations with the said categories will be operating on some specific links, and not on all the links that are available. This limits the number of active STAs at any link and hence reduces contention/collisions. We next evaluate the performance of MLO with TID-to-link mapping and compare it with default mapping where no such mapping exists. We observe that such a solution helps in reducing collisions and therefore improves throughput compared to MLO with default mapping. Further, the median and tail E2E latency values, median and tail RTT values, retransmissions and out-of-order packets are also reduced compared to the former case, and even with SLO in most of the cases. Thus, such a solution can help us realize the true potential of MLO.

Our Contributions: To the best of our knowledge, this is the first work that studies the interplay of MLO and TCP. Our key contributions are as follows:

- We perform an exhaustive evaluation of MLO with TCP traffic. We consider both sparse and dense deployments with single AC (BE), 3 ACs (BE, VI & VO), and with traffic in DL, UL, DL+UL directions. We observe that though the throughput increases with multiple links, the number of retransmissions, out-of-order packets, the tail RTT, sometimes even the 50^{th} percentile RTT, and the tail E2E latency suffers as we increase the number of links.
- We perform a root cause analysis and observe that TCP traffic encounters transmission failures due to the collision of TCP data packets in one direction and TCP ACK packets in the reverse direction. The number of such failures increases with the increase in the number of links, as with more links the data traffic increases multifold.
- We then implement a solution to reduce contention/collision by using TID-to-Link Mapping. The experiments show that TID-to-Link mapping improves throughput, RTT, and both 50^{th} percentile and tail E2E latency values compared to MLO without TID-to-Link Mapping and even with SLO, in most of the cases.

II. Related work

We categorize the related work as follows:

A. Prior works on STR mode of MLO

Here we discuss prior works that focus on STR mode of MLO and are closely related to this work. Carrascosa *et al.* [7], [8] study the impact of different modes of MLO on the latency of the network under varying occupancy levels by replaying traces collected in a Football stadium. They consider 1AP and 1 STA in the presence of interference from OBSS (Overlapping Basic Service Set) transmissions. They show that with symmetrically occupied links, MLO with STR mode can reduce the 95^{th} percentile latency by 78% and the average latency by 69%. With asymmetrically occupied links, latency with MLO can be higher than SLO due to frequent freezing of the backoff counter as the link is occupied by interfering transmissions of other OBSS. Then they proposed STR+ mode where a packet is allocated to the interface whose backoff expires first. With this, MLO can decrease the 95^{th} percentile latency by up to 60% compared to SLO for asymmetric links.

Naik *et al.* [6] study the impact of MLO on the latency for different amounts of load on MLD devices and different numbers of MLD devices in the network. They consider the presence of contending SLO devices. The authors show that at a low load with a single MLD, adding one more link reduces the 90^{th} percentile latency significantly compared to increasing the bandwidth and MCS. Increasing further links does not lead to much improvement in latency. At high loads, increasing the bandwidth and MCS has a major impact in reducing the latency compared to adding a second link. When the number of MLDs is more, then adding links gives significant improvements in 90^{th} percentile latency, and this improvement increases with an increase in the traffic load. Adhikari and Verma in [3] study the relative gain of MLO with respect to SLO in terms of throughput and latency for different channel and network configurations. They mention that both in the absence and presence of contending devices and with full buffer, MLO STR can provide double the throughput of SLO with two symmetric links. With unsaturated traffic, the latency will be reduced to half. A similar kind of evaluation is done by Chen *et al.* [9].

B. Prior works on the other modes of MLO

Here we discuss prior works that focus on other modes like NSTR, EMLSR. Song and Kim in [11] analyze the performance of NSTR operation using analytical modeling based on Markov chain. They further verify their analysis with the help of simulations in MATLAB. Naribole *et al.* in [12] present and evaluate the NSTR operation where the transmissions on the two channels are aligned. Naribole *et al.* in [13] propose the end alignment of the transmissions of STAs where the separation between the different channels is not enough. They compare the performance of the proposed scheme with single link and with unaligned overlapping transmissions. Lan *et al.* in [10] evaluate the performance of both MLSR and

EMLSR through analytical modeling and simulations in NS-3 and compare it with SLO. They show that both SLO and MLMR perform almost similar while EMLSR outperforms both of them.

All the prior studies on STR or other modes of MLO consider UDP traffic. MLO can impact the performance of TCP traffic differently as (1) it involves traffic in both directions even if the data flow is in one direction, (2) it is sensitive to out-of-order packets. We therefore, evaluate the impact of MLO on the performance of TCP traffic.

III. Simulation Setup

We utilize the latest 802.11be models in Network Simulator 3 (NS-3, version 3.40) [15] for our simulations. NS-3 uses the default STR mode, i.e., the STA contends on all the available links and the packets are transmitted on whichever link wins contention first. The parameters for the simulation

Table I: Simulation parameters [1]

Parameter	Value	Parameter	Value
Stations	**40**, 10	Links	1, 2 and 3
Radius	20m	Bandwidth	40MHz
TXOP Duration	5.440ms	Spatial Streams	1
Transport Protocol	TCP	Guard Interval	800 ns
Application Layer Payload Size	1000 B	MCS	11, **variable**
MSDU Aggregation	Disabled	MPDU Aggregation	Enabled
MAC Queue Size	5,000 packets	Simulation Duration	10s

setup are summarized in Table.I. The STAs are uniformly distributed around the AP in a radius of 20 meters. The MAC queue size is fixed at 5,000 packets to allow sufficient frames for aggregation. We consider OFDM transmissions. The MAC acknowledgments are the usual implicit Block ACKs. We set the TCP segment size to 1500 bytes. The rest of the TCP parameters are the default ones of NS-3. We evaluate under both perfect channel conditions where the highest MCS is possible (MCS 11) and with practical channel conditions with path loss and fading. In such cases, MCS is selected by MinstrelHtWifiManager module of NS-3. We use FriisPropagation as the loss model and NakagamiPropagation as the fading model.

We consider traffic in (1) only downlink (DL), (2) only uplink (UL) and (3) both downlink and uplink directions (DL+UL). Table II shows the different link configurations. For single link ***SL***, we use a 40MHz link on the 5GHz band. For double link, we consider two possible scenarios - 1) ***DL1:*** 40MHz links on 2.4GHz and 5GHz bands, and 2) ***DL2:*** 40MHz links on 5GHz and 6GHz bands. For the triple link ***TL***, we consider 40MHz links on all the 2.4GHz, 5GHz and 6GHz bands.

In each of the link configurations, we offer a total load of 280Mbps which is equal to the 40MHz link capacity for single

[1]For, more than one parameter in a cell, parameters marked in bold are used generally in the entire work. Parameters not in bold, are used only under specific sections.

Table II: Link configurations considered.

Abbreviation	Meaning	Configuration
SL	Single Link	40MHz link on 5GHz band
DL1	Double Link Scenario 1	40MHz links on 2.4GHz and 5GHz bands
DL2	Double Link Scenario 2	40MHz links on 5GHz and 6GHz bands
TL	Triple Link	40MHz links on all 2.4, 5 and 6 GHz bands

user transmission at MCS 11 and a guard interval of 800ns. So for 10 STAs, it will be 28 Mbps per STA and for 40 STAs it will be 7 Mbps per STA for the DL or UL traffic. For the DL+UL case, it will be half of the load for DL or UL direction (14 Mbps for 10 STAs and 3.5 Mbps for 40 STAs).

Metrics: We measure throughput, E2E latency, RTT, number of TCP retransmissions and out-of-order packets. Throughput is the number of bits transmitted per second. E2E latency is defined as the time interval between the packet being transmitted by the application layer at the sender and received by the application layer at the receiver side. RTT is the Round Trip Time which is the time interval when the TCP packet was transmitted and when its ACK was received. Retransmissions are the TCP packets that are retransmitted by the source transport layer. Out-of-Order packets are the ones received successfully but not in order, at the receiver.

IV. Performance Evaluation

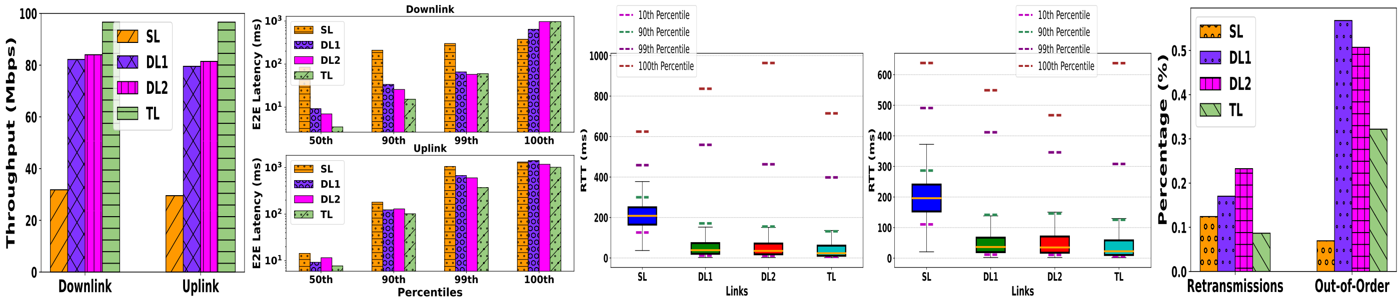
We first analyze the impact of MLO on the performance of TCP flows. For this, we compare the performance of TCP flows with single, double and triple links with sparse (10 STAs) and dense (40 STAs) deployments. We perform the comparison for two scenarios - 1) with only one Access Category (AC) Best Effort (BE), and 2) with three different ACs Best Effort (BE), Video (VI) and Voice (VO). We discuss the evaluation of only BE in subsection IV-A and of VI, VO and BE in subsection IV-B. In both the scenarios, we perform evaluation for (1) only downlink (DL), (2) only uplink (UL), and (3) both downlink and uplink directions (DL+UL). However, we do not report the results for all configurations and only present the relevant set of results in each scenario.

A. Analysis with only BE AC

Here, we consider only BE traffic at all the STAs.

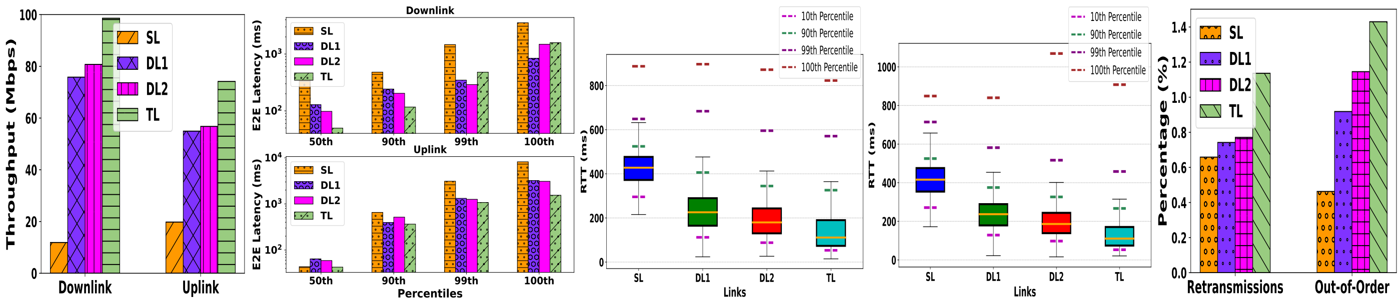
*1) **Sparse Deployment**:* Here, we consider 10 STAs uniformly distributed around the AP. We present the results of DL+UL traffic as this is the most challenging scenario. As the traffic is bidirectional, we report the metrics separately in both directions.

DL+UL Traffic: Fig. 1 shows the different metrics in this case. Fig. 1(a) shows that both the UL and DL throughput increase with the increase in the number of links which is also one of the prime objectives of 802.11be with the introduction of MLO. Throughput with two links can sometimes be more than twice that of a single link based on the occupancy characteristics of the two links, as also identified in [8]. In our case, the amount of increase with links will also depend on the MCS selected and collisions faced on the other links



(a) (b) (c) (d) (e)

Figure 1: Sparse deployment with BE AC traffic for DL+UL: (a) Throughput Downlink & Uplink (Mbps), (b) E2E latency Downlink & Uplink (ms), (c) RTT (ms) in DL, (d) RTT (ms) in UL, (e) Retransmission and out-of-order percentage.



(a) (b) (c) (d) (e)

Figure 2: Dense deployment with BE AC traffic for DL+UL: (a) Throughput Downlink & Uplink (Mbps), (b) E2E latency Downlink & Uplink (ms), (c) RTT (ms) in DL, (d) RTT (ms) in UL, (e) Retransmission and out-of-order percentage.

in addition to the occupancy. Fig. 1(b) show the different percentile values of E2E in both UL and DL directions. We see that the E2E latency in UL direction decreases with MLO as now the STAs can contend on multiple links and transmit on whichever link becomes available. However, in the DL direction, the tail ($> 99^{th}$ percentile) E2E values increase with MLO. This is due to the increase in collision of the packets transmitted by the AP with the packets transmitted by the STAs. Due to TCP traffic, the STAs will be contending to transmit both the TCP data packets and the TCP ACK packets. As the data received will be more with MLO, as can be seen from the throughput graph of Fig. 1(a), the contention and hence collisions will be more with MLO. We discuss these factors in detail in §.V. Fig. 1(c) and 1(d) shows the Round Trip Time (RTT) both in the DL and UL direction. Like E2E latency, the 99^{th} and higher percentile RTT is higher with MLO compared to SLO. This is specifically true in DL.

Fig. 1(e) shows that the number of retransmissions and out-of-order percentages increases with MLO compared to SLO. These metrics are lower with 3 links (TL) compared to 2 links (DL1 & DL2). As there are only 10 STAs, increasing the links helps in reducing the number of retransmissions.

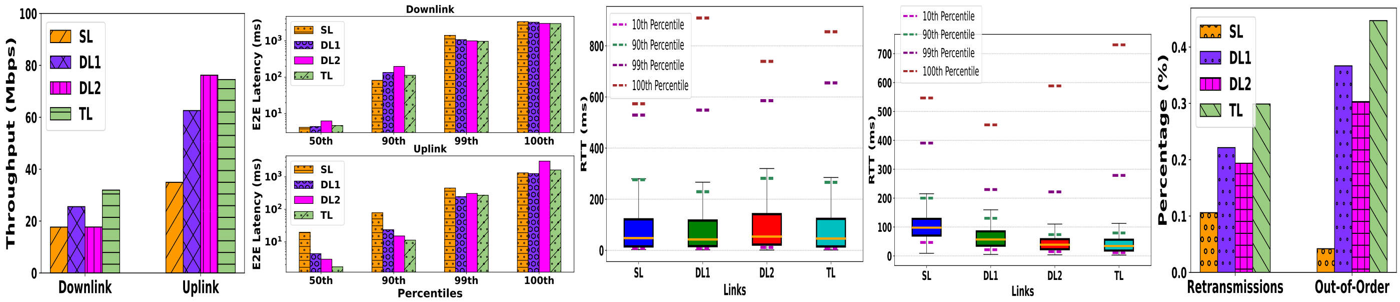
*2) **Dense Deployment**:* Here, we consider 40 STAs uniformly distributed around the AP.

DL+UL Traffic: Fig. 2 shows the different metrics of interest. Here, the throughput again increases with the increase in the number of links. However, the throughput in UL direction has decreased compared to the previous case as the collision would increase with the increase in the number of STAs. In the DL direction as well, the throughput decreases significantly with SL, but not with multiple links. This is due to the increased transmission opportunity with multiple links. In this case, the E2E latency does not suffer for MLO compared to SLO. However, going from 2 to 3 links, the E2E latency increases for DL. Fig. 2(c) shows that the 99^{th} percentile RTT in case of DL with DL1 is higher than SL. Fig. 2(d)) shows that the 100^{th} percentile RTT in case of UL increases with increasing links. Fig. 2(e) shows that retransmissions and out-of-order percentages also increase. In the dense network, such retransmissions and out-of-order percentages are higher compared to the sparse network.

B. Analysis with BE, VI and VO ACs

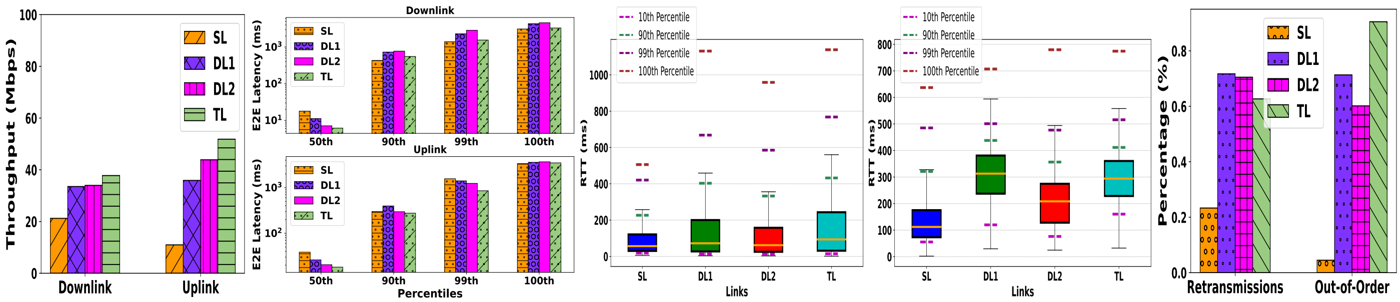
In this case, we consider traffic of BE, VI and VO ACs at the STAs. The flow of each AC will be enabled on an almost equal number of STAs.

*1) **Sparse Deployment**:* **DL+UL Traffic:** Fig. 3 shows the different metrics of interest. Here, the throughput decreases compared to the case with only BE AC (Fig. 1). This is due to the increase in contention level with different ACs as $2/3_{rd}$ of the STAs have high-priority traffic. The contention parameters for VI and VO i.e., CW_{min} (Contention Window), CW_{max}, and $AIFSN$ (Arbitration Inter Frame Space Number), are less compared to BE AC, so the contention happens more



(a) (b) (c) (d) (e)

Figure 3: Sparse deployment with BE, VI and VO ACs traffic for DL+UL: (a) Throughput Downlink & Uplink (Mbps), (b) E2E latency Downlink & Uplink (ms), (c) RTT (ms) in DL, (d) RTT (ms) in UL, (e) Retransmission and out-of-order percentage.



(a) (b) (c) (d) (e)

Figure 4: Dense deployment with BE, VI and VO ACs traffic for DL+UL: (a) Throughput Downlink & Uplink (Mbps), (b) E2E latency (ms) Downlink & Uplink, (c) RTT (ms) in DL, (d) RTT (ms) in UL, (e) Retransmission and out-of-order percentage.

frequently. Here, the UL throughput is more than the DL throughput. This is because as there are more STAs with higher priority traffic, they will contend for the channel frequently and hence will transmit more data. Also, the E2E latency, RTT, retransmissions, and out-of-order packets increase in this case compared to the previous case of sparse BE AC (Fig. 1). Compared to SLO, the throughput increases for MLO, significantly in the UL direction. However, the 50^{th} and the 90^{th} percentile E2E in the case of DL are higher with MLO compared to SLO. For UL, the 100^{th} percentile E2E suffers with MLO compared to SLO. $>= 99^{th}$ percentile RTT suffers with MLO compared to SLO in the case of DL and the 100^{th} percentile for UL. Besides, the number of retransmissions and out-of-order packets also increase with the increase in the number of links.

2) Dense Deployment: Here, we discuss all three traffic directions - DL, UL and DL+UL.

DL Traffic: As the traffic is only in DL, the throughput in this case is higher for all the link configurations compared to the previous case where we considered both UL+DL traffic. Here, the 100^{th} percentile E2E latency with SL is 3063.922ms, with DL1 is 3055.377ms, with DL2 is 3147.008ms and with TL is 3225.002ms (Figure skipped due to space limitation). So even with the simple case of only DL traffic, the E2E latency with MLO is a little higher compared to that of SLO. Further, we note that the RTT increases with the increase in the number of links from 90^{th} percentile only. The retransmission percentages increase with the increase in the number of links.

UL Traffic: In this case, the throughput decreases compared to the previous case of only DL traffic since the STAs also have uplink traffic, hence more contention. The E2E latency values are higher compared to the previous case. However, in this case E2E latency for SLO is higher compared to MLO. Here, even the 50^{th} percentile RTT for MLO cases is significantly higher than SLO. The retransmissions and out-of-order percentages increase with increasing links.

DL+UL Traffic: Fig. 4 shows the different metrics of interest. Here, the throughput decreases even further and the gain in throughput that the MLO offers compared to SLO is not that significant as compared to all the previous cases. This is due to the increase in the access delay to win the channel contention. The tail E2E latency values are also higher. The RTT in the downlink direction also starts increasing with MLO from the 50^{th} percentile. The increase becomes more significant at higher percentile values. In the uplink case, the 50^{th} percentile RTT is significantly higher for MLO cases compared to SLO. The out-of-order and retransmission percentages with MLO are also higher than SLO.

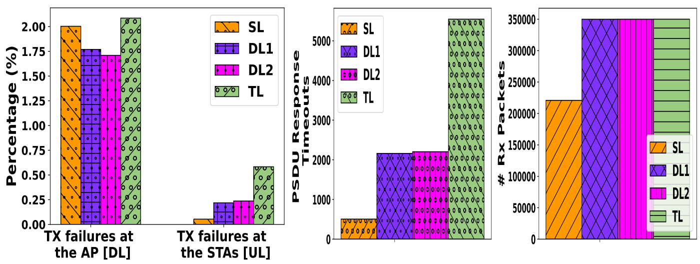
V. Root Cause Analysis and Solution Design

In this section, we perform a root cause analysis to uncover the reason behind the non-preferred increase in some of the metrics with MLO compared to SLO.

A. Root Cause Analysis

In the previous section, we considered lossy channels with path loss and fading to simulate realistic network scenarios. We saw that the 99^{th} percentile RTT and E2E latency increases with the increase in the number of links in most of the scenarios. In order to find the root cause for this, we performed experiments with an ideal channel where there are no channel-induced losses. Fig. 5 presents results for 40 STAs where only DL BE traffic was considered. Fig. 5(a) shows the number of transmission failures at the AP and at the STA. These transmission failures are collected using the *NAckedMpdu* trace source in NS-3, which reports the MPDUs that were negatively acknowledged via a Block Ack. We observe that the number of negatively acknowledged MPDUs increases with the increase in the number of links, especially at the STA side. In this case, there are no channel losses, and the traffic is also in only DL direction. Yet there are failed MPDUs both at the AP and at the STAs side. Next, we connect *PsduResponseTimeout* trace source to verify whether the MSDUs that were lost were due to collisions. *PsduResponseTimeout* reports the PSDUs whose response was not received before the timeout i.e., timeout while waiting for the MAC acknowledgment. Fig. 5(b) shows that the PSDU response timeouts increase with the increase in the number of links. This is due to collisions of TCP data packets transmitted in the DL direction with the TCP ACK packets in the UL direction.

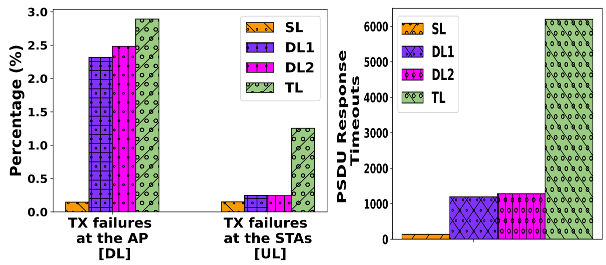
Next, we compute more statistics to understand why exactly collisions increase with the number of links. The total number of transmitted packets is same (350000) across single, double, and triple links. However, Fig. 5(c) shows that the total number of received packets increases with the increase in the number of links. As the total number of received packets increases with the increase in the number of links (SL: 220837; DL1, DL2, TL: 350000), the STAs will be contending more to transmit the TCP ACKs of the higher number of received TCP data packets. This will increase the channel contention among the AP and the STAs, thus leading to a higher number of collisions and MPDU failures.



(a) (b) (c)

Figure 5: Dense deployment with BE traffic in DL direction with ideal channel: (a) transmission failures at the AP and at the STAs, (b) number of PSDU response timeouts, (c) number of received packets.

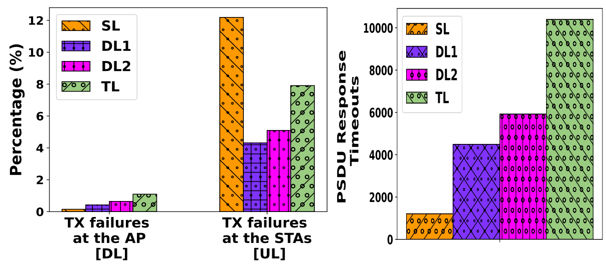
To validate the root cause, we repeated this experiment with a lossy channel and noted down the number of Rx packets, Tx failures both at the AP and STAs, and PSDU response timeouts. Fig. 6 shows that the number of failed MSDUs and the PSDU response timeouts increase with the increase in the number of links due to increased collisions with multiple links.



(a) (b)

Figure 6: Dense deployment with BE traffic in DL direction with lossy channel: (a) transmission failures at the AP and at the STAs (b) number of PSDU response timeouts.

Fig. 7 reports these metrics for the case where the traffic belongs to BE, VI and VO ACs which again confirms that the collisions increase with the increase in the number of links. Unlike in Fig. 6, the number of transmission failures at the STAs and the PSDU response timeouts increase significantly due to the more number of STAs with higher priority traffic contending frequently. The AIFSN value for VO and VI is 2 while for BE is 3. The CW_{min} for VO, VI and BE is 3, 7 and 15, respectively, while the CW_{max} for VO, VI, and BE is 7, 15 and 1023, respectively [16]. As the contention parameters' values are lesser for VO and VI, STAs with these flows will contend more frequently compared to the ones with only BE traffic. This reduces the number of received packets as can be seen from the decrease in throughput compared to when only BE traffic was present.



(a) (b)

Figure 7: Dense deployment with BE, VI and VO traffic in DL direction with lossy channel: (a) transmission failures at the AP and at the STAs (b) number of PSDU response timeouts.

B. Solution Design

As collision is the root cause for such performance drop, a solution should target reducing collisions. We thus focus on designing a solution that can restrict the number of STAs per link. For example, if there are 40 STAs, then with STR mode all 40 STAs will be contending on each of the three links. If we could reduce the number of STAs contending on any link, then the amount of contention/collision would reduce. Suppose, there are N STAs in the network and 1 AP. Let the probability of the channel access at the AP and at STAs be denoted by τ_{ap} and τ_{sta}, respectively. These τ_{ap} and τ_{sta} can be computed based on the value of CW_{min}, the number of backoff stages, and the probability of collision p^c_{ap}/p^c_{sta} using either Markov chain model or using a fixed-point model. The probability of collision at AP and STAs can be computed as given in Eqs. 1 and 2.

$$p^c_{ap} = 1 - (1 - \tau_{sta})^N \quad (1)$$

$$p^c_{sta} = 1 - (1 - \tau_{ap})(1 - \tau_{sta})^{N-1} \quad (2)$$

If N reduces from 40 to say 13, the collision probability will also decrease.

A solution would be to allocate flows to distinct links; this could be achieved by implementing a link scheduler that selects packets from flows allocated to a link when channel access is obtained on that link. In this paper, we evaluate the benefits of allocating flows to links by exploiting the TID-to-Link mapping mechanism.

TID-to-Link Mapping: In MLO, as there are multiple available links, over which the data could be transmitted, flows belonging to different ACs can be mapped to different links. There are two TIDs (Traffic IDentifiers) belonging to each AC which indicate the priority of the flow. TIDs 0, 3 correspond to BE traffic; 1, 2 correspond to BK traffic; 4, 5 correspond to VI traffic; and 6, 7 correspond to VO traffic.

If TID-to-Link mapping is not specified, default TID-to-Link mapping is used which maps all TIDs to all available links. This means that the flow corresponding to any of the TID can be transmitted over any of the available links. Using TID-to-Link mapping, specific TIDs can be mapped to specific links so that data packets belonging to that TID are only transmitted over the link(s) to which that TID is mapped onto. A TID can be mapped to either one or more links using TID-to-Link mapping. Moreover, TID-to-Link mapping can be performed either in the DL direction or the UL direction or in both the DL and UL directions. In order to reduce the contention on each link we map an equal number of flows on each link using TID-to-Link mapping. Hence, the packets belonging to that flow will contend only on the link to which that TID is mapped onto.

VI. TID-to-Link Mapping, Implementation and Evaluation

A. Configuration

We configured TID-to-Link mapping using the TIDs belonging to the three ACs. For three links, we map the TID(s) belonging to 1 AC to one link. For three links, we used TID-to-Link mapping of [0,3 0; 4,5 1; 6,7 2], which means that the TID 0&3 of BE traffic is mapped to the first link (link 0), TID 4&5 of VI traffic is mapped to the second link (link 1) and TID 6&7 of VO traffic is mapped to the third link (link 2). For two links, we used TID-to-Link mapping of [0,3,5 0; 4,6,7 1], which means that TIDs - $0, 3, 5$ are assigned to link 0 and TIDs - $4, 6, 7$ are assigned to link 1. To show the potential of TID-to-Link mapping, we present a comparison with default mapping.

There are two TIDs corresponding to each AC; similarly, BE also has two TIDs - 0 and 3. Now TID-to-Link Mapping solution would be straightforward to implement with two links. However, implementing that with three links is not straightforward as there are only two TIDs that can be mapped. Hence, to provide a fair comparison of BE traffic with two links (DL1, DL2) and three links (TL), we emulate it using BE, VI, and VO traffic by modifying their access parameters. We consider three ACs, but make the access parameters like the CW_{min}, CW_{max}, $AIFSN$, $TXOP$, etc. same for all the three ACs with the value of BE so that they will behave like BE traffic. Then, the TID-to-Link mapping solution can be implemented easily by mapping the TIDs of the three ACs to the three different links, as discussed above.

For the TCP traffic in only DL direction, we use TID-to-Link mapping in the UL direction only so as to reduce the contention at the STA side and thus prevent collisions. With both UL+DL traffic, TID-to-Link mapping is used in both UL and DL directions. We discuss only the results with 3 ACs which is most challenging and omit the description of the results of 1 AC. The improvements are more for 1AC.

B. TID-to-Link mapping with BE, VI and VO ACs

1) Dense Deployment: Here, we perform the evaluation with BE, VO, and VI traffic with their default access parameters and report the results first for DL case and then for DL+UL case.

DL traffic: Here, we use TID-to-Link mapping in UL direction. Fig. 8 shows the evaluation results in this case. Fig. 8(a) shows that TID-to-Link mapping improves the throughput, especially for DL2 and TL and marginally for DL1 also. Fig. 8(b) and 8(c) show that it also improves both median and max E2E latency. There is a slight increase in median E2E for DL1 (it increases from 2ms to 5ms but for DL2 and TL cases, it decreases compared to default mapping). With TID-to-Link mapping, the max E2E latency improves significantly (from 3225ms to 3016ms for the TL case). It even improves compared to the SL case which shows that with TID-to-Link mapping can help to realize the true benefits of MLO and make it more suitable for real time applications like AR/VR applications by reducing the max E2E latency significantly.

Fig. 8(d) shows that both median and max RTT are also reduced on using TID-to-Link mapping. **nM** denotes that TID-to-Link mapping is not used and **TM** denotes TID-to-Link Mapping is used. Especially, $> 90^{th}$ percentile RTT for DL1, DL2 and TL reduces significantly and even gets lower than that of SL. Fig. 8(e) shows that TID-to-Link mapping also reduces retransmission percentage. Fig. 8(f) shows the percentage of out-of-order packets. With DL1, there is a slight increase in out-of-order packets. This will also depend on the TID-to-Link

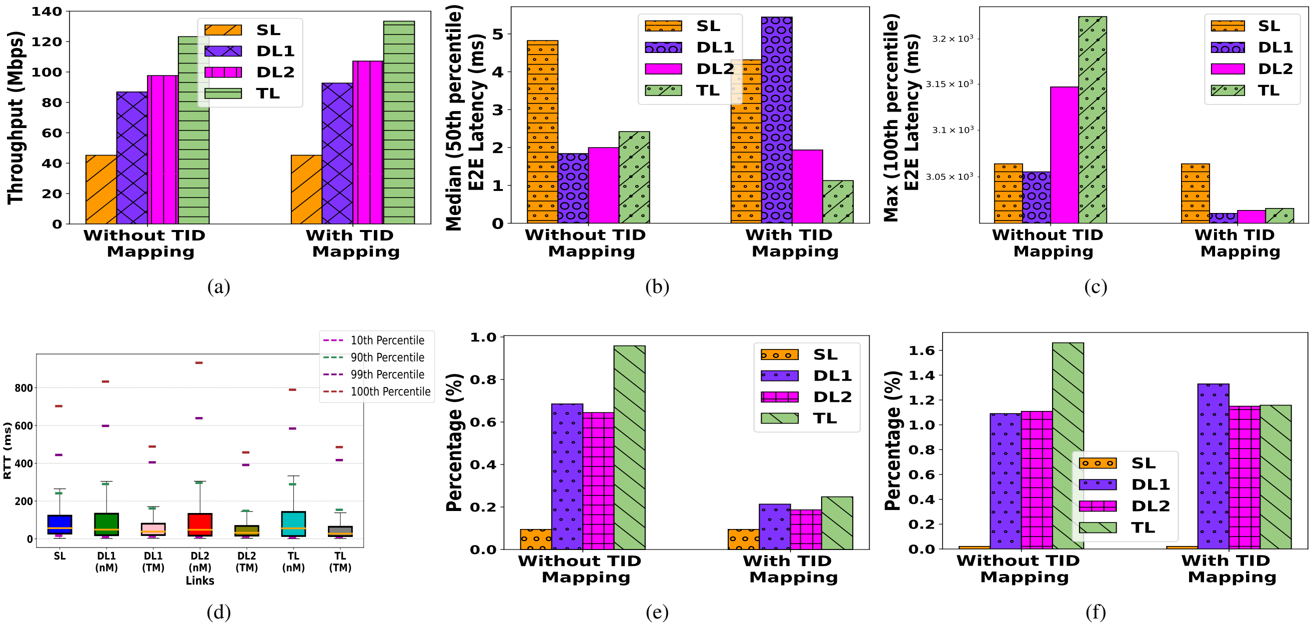


Figure 8: TID-to-Link Mapping with 3 ACs with only DL traffic: (a) throughput, (b) median (50^{th}) E2E (ms), (c) max (100^{th}) E2E (ms), (d) RTT (ms) (nM denotes that TID-to-Link mapping is not used and TM denotes TID-to-Link Mapping is used), (e) retransmissions, (f) out-of-order packets.

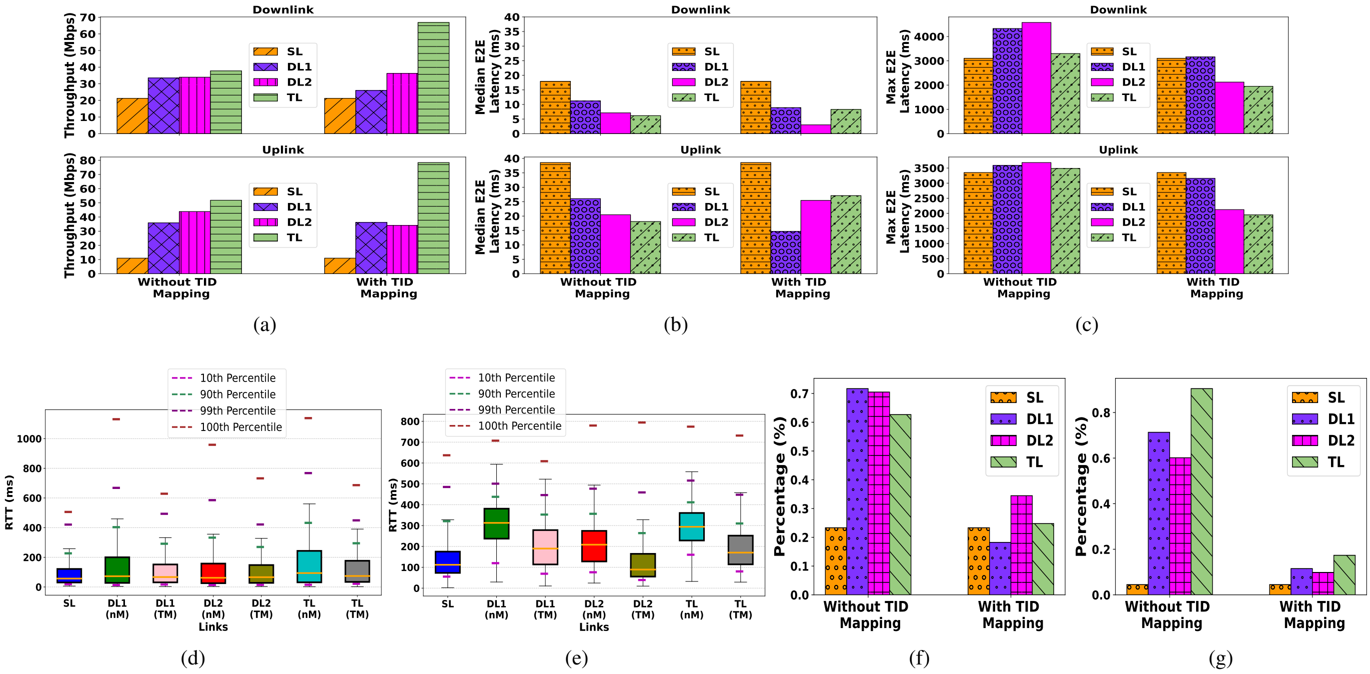


Figure 9: TID-to-Link Mapping with 3 ACs with DL + UL traffic: (a) throughput, (b) median (50^{th}) E2E (ms), (c) max (100^{th}) E2E (ms), (d) DL RTT (ms) (nM denotes that TID-to-Link mapping is not used and TM denotes TID-to-Link Mapping is used), (e) UL RTT (ms), (f) retransmissions, (g) out-of-order packets.

mapping used. With some TID-to-Link mapping configurations, this number decreased compared to those without TID-to-Link mapping. However, with TL, the out-of-order packets decrease with all the TID-to-Link mapping configurations with which we tested. This is because with TID-to-Link mapping there will be less traffic on each link in the case of triple links compared to the two links. We saw that the number of transmission failures, both at the AP and at the STAs, decreases significantly with TID-to-Link mapping compared to default TID-to-Link mapping. This validates that the collisions can be reduced using TID-to-Link mapping which leads to an improvement in throughput and E2E latency values. We also tested with utilizing TID-to-Link mapping both in the DL and UL directions, i.e. TID-to-Link mapping for both TCP data in the DL and TCP ACKs in the UL. Though it lowers the E2E latency significantly compared to the default TID-to-Link mapping, the throughput reduces marginally as the AP is not able to fully utilize the potential of MLO in that case, as the TCP flow is in only DL.

DL+UL traffic: Fig 9 shows the evaluation results for the case where DL+UL traffic is used. In this case, both the DL and UL throughput increase significantly for the TL case (Fig. 9(a)). For the two links, there is a marginal decrease in either the UL or the DL throughput for either DL1 or DL2. Fig. 9(b) and Fig. 9(c) shows the median and max E2E latency. We see that the max E2E latency decreases significantly with TID-to-Link mapping with a slight increase in median value in some of the UL cases. From Fig. 9(d) and 9(e), we can observe that both the median and max RTT also reduces on using TID-to-Link mapping. In this case also, the retransmission percentage is reduced (Fig. 9(f)). Besides, the out-of-order packets also reduce significantly with both two and three links (Fig. 9(g)). The transmission failures both at the AP and STAs, also reduce significantly with TID-to-Link mapping for DL1, DL2 and TL.

Summary: We conclude that TID-to-Link mapping can reduce the number of collisions. It improves throughput and reduces E2E latency and RTT values, especially the tail ones. Even in the challenging case of DL+UL traffic, TID-to-Link mapping improves the performance significantly with respect to most of the metrics. Such improvements are more significant for three links than two links. This is due to the fact that as there are only two links and the traffic is both ways, TID-to-Link mapping is not sufficient to reduce the collisions significantly. One possible solution could be to transmit TCP ACKs utilizing OFDMA. Our analysis with TID-to-Link mapping shows how the promise of MLO can be retained in realistic scenarios by reducing collisions. This opens a new direction of link scheduling that intelligently allocates traffic to links while optimizing the application performance.

VII. Conclusion and Future Work

The prime objective of MLO is to improve throughput and reduce latency. Prior works have performed an evaluation with UDP traffic and measured the performance gains of MLO. Such studies show that MLO provides better performance compared to SLO. We analyze the performance of MLO with TCP traffic and observe that MLO increases some of the undesired metrics like the RTT, failed MSDUs, etc, due to an increase in collisions with MLO. This restricts MLO from realizing its true potential. We then propose a solution utilizing TID-to-Link mapping to address the root cause of this, which is an increased number of collisions. Such a solution helps in reducing the number of collisions. This in turn, helps in improving the throughput and lowering both the E2E and RTT.

This work is an important initial step to show how reducing contention can help improve MLO performance with TCP traffic. This opens a whole new path of research in this domain of optimal traffic allocation strategy of allocating packets to links. Some other alternative solutions can also be explored that could reduce the amount of collisions with TCP traffic, say by serving only a set of STAs at a time and thus limiting the contention; or transmitting TCP ACKs through OFDMA.

References

[1] E. Khorov, I. Levitsky, and I. F. Akyildiz, "Current status and directions of ieee 802.11 be, the future wi-fi 7," *IEEE access*, vol. 8, pp. 88 664–88 688, 2020.

[2] "Ieee draft standard for information technology–telecommunications and information exchange between systems local and metropolitan area networks–specific requirements - part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications amendment: Enhancements for extremely high throughput (eht)." [Online]. Available: https://standards.ieee.org/ieee/802.11be/7516/

[3] S. Adhikari and S. Verma, "Analysis of multilink in ieee 802.11 be," *IEEE Communications Standards Magazine*, vol. 6, no. 3, pp. 52–58, 2022.

[4] 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," *arXiv preprint arXiv:2210.07695*, 2022.

[5] B. Bellalta, M. Carrascosa, L. Galati-Giordano, and G. Geraci, "Delay analysis of ieee 802.11 be multi-link operation under finite load," *IEEE Wireless Communications Letters*, vol. 12, no. 4, pp. 595–599, 2023.

[6] G. Naik, D. Ogbe, and J.-M. J. Park, "Can wi-fi 7 support real-time applications? on the impact of multi link aggregation on latency," in *ICC 2021-IEEE International Conference on Communications*. IEEE, 2021, pp. 1–6.

[7] M. Carrascosa, G. Geraci, E. Knightly, and B. Bellalta, "An experimental study of latency for ieee 802.11 be multi-link operation," in *ICC 2022-IEEE International Conference on Communications*. IEEE, 2022, pp. 2507–2512.

[8] M. Carrascosa-Zamacois, G. Geraci, E. Knightly, and B. Bellalta, "Wi-fi multi-link operation: An experimental study of latency and throughput," *IEEE/ACM Transactions on Networking*, 2023.

[9] 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.

[10] X. Lan, X. Zu, J. Yang *et al.*, "Enhanced multilink single-radio operation for the next-generation ieee 802.11 be wi-fi systems," *Security and Communication Networks*, vol. 2022, 2022.

[11] T. Song and T. Kim, "Performance analysis of synchronous multi-radio multi-link mac protocols in ieee 802.11 be extremely high throughput wlans," *Applied Sciences*, vol. 11, no. 1, p. 317, 2020.

[12] S. Naribole, S. Kandala, and A. Ranganath, "Multi-channel mobile access point in next-generation ieee 802.11 be wlans," in *ICC 2021-IEEE International Conference on Communications*. IEEE, 2021, pp. 1–7.

[13] S. Naribole, S. Kandala, W. B. Lee, and A. Ranganath, "Simultaneous multi-channel downlink operation in next generation wlans," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*. IEEE, 2020, pp. 1–7.

[14] I. Levitsky, Y. Okatev, and E. Khorov, "Study on simultaneous transmission and reception on multiple links in ieee 802.11 be networks," in *2020 International Conference Engineering and Telecommunication (En&T)*. IEEE, 2020, pp. 1–4.

[15] ns-3.40 released | ns-3. [Online]. Available: https://www.nsnam.org/news/2023/09/27/ns-3-40-released.html

[16] Quality of service in wlans. [Online]. Available: https://wifihelp.arista.com/post/qualtity-of-service-in-wlans