MACHINE LEARNING BASED FRAME SIZE OPTIMIZATION IN WLAN DOWNLINK MU-MIMO CHANNEL WITH THE LEAST COST OF DELAY

Lemlem Kassa[1], Jianhua Deng[1], Mark Davis[2] and Jingye Cai[1]

[1]School of Information and Software Engineering,
University of Electronic Science and Technology China (UESTC),
Chengdu 610054, China
[2]Communication Network Research Institute (CNRI),
Technological University Dublin, D08 NF82 Dublin, Ireland

ABSTRACT

The key enhancement in the medium access control (MAC) layer is frame aggregation introduced by the IEEE 802.11n/ac standard to accommodate the growing traffic demand in the WLAN by allowing multiple packets aggregated per transmission. Frame aggregation efficiently reduces control overhead in the MAC layer, such as the MAC header and thus it helps to enhance transmission efficiency and throughput performance of WLAN. However, heterogeneous traffic demand among streams in the WLAN downlink MU-MIMO channel creates a challenge to efficiently utilize the benefits of frame aggregation. Transmission efficiency is also compromised during frame size setting determination because when a frame size is larger, the impact of the overhead frame can be lower, but they are also more vulnerable to transmission errors. Thus, this trade-off between maximizing frame size and minimizing overhead frames should be addressed by employing an adaptive frame aggregation technique to derive the optimal frame size that would maximize the throughput in WLAN downlink MU-MIMO channel. Moreover, when frame aggregation approach is employed, more frames must wait before transmission in a buffer which causes a delay in the performance of WLAN. Thus, analysing the trade-off between maximizing throughput and minimizing delay is a critical issue that should also be addressed to enhance the performance of WLAN. However, the majority of the existing adaptive aggregation algorithms in the WLAN downlink MU-MIMO channel are focused to maximize the throughput or minimize the delay. The main contribution of this paper is to propose a machine learning-based frame size optimization algorithm by extending our earlier approach in considering the cost of delay to maximize the system throughput of WLAN. The effectiveness of the proposed scheme is evaluated over the FIFO Baseline Approach and earlier conventional approaches under the effects of various traffic patterns, channel conditions, and the number of STAs.

KEYWORDS

Delay, downlink MU-MIMO, frame size optimization, WLAN, heterogeneous traffic, machine learning, neural network, throughput optimization.

1. INTRODUCTION

Wireless technology has been continuously evolving to support basic coverage and satisfy the advanced needs of today’s network [1]. To improve the system performance of WLAN, IEEE 802.11 introduced single-user MIMO (SU-MIMO) [2] from IEEE 802.11n [3] and multi-user multiple in multiple out (MU-MIMO) in IEEE802.11ac [2-4]. SU-MIMO transmission is the conventional approach that allows multiple transmission of data streams to a single station. The SU-MIMO WLAN is effective if the connecting stations are equipped with multiple receiving

DOI: 10.5121/ijcnc.2022.14601

antennas. However, in a WLAN that connects many stations, due to the limitation of receiving antennas at receiving stations, SU-MIMO cannot efficiently utilize the full advantage of multiple transmission antennas found at access points (APs) [2]. As a result, this situation causes degradation in the system performance in terms of system capacity and user throughput. As a consequence of this background, by adopting MU-MIMO, simultaneous data transmission to multiple stations (STAs) is achieved by forming virtual spatial transmission channels between the AP and the multiple receiving STAs [2].

The other key enhancement in the medium access control (MAC) layer is frame aggregation introduced by the IEEE 802.11n/ac to accommodate the growing traffic demand in the network. By using frame aggregation technique, multiple packets can be aggregated for a single transmission thus this reduces overhead frame consumption in the MAC layer, such as the MAC header. This approach provides a throughput enhancement and transmission efficiency in WLAN [5-8]. However, heterogeneous traffic demand among streams in the WLAN downlink MU-MIMO channel creates a challenge to efficiently utilize the benefits of frame aggregation. This is because of that, when shorter and longer streams are grouped in downlink MU-MIMO transmission, wasted space channel time will occur which is a time duration where a part of spatial streams carries a data frame whereas the others do not [9-11]. To accommodate the airtime of the shorter streams, frame pad bits, called padding bits, can be added at the tail of the shorter stream until it has the same transmission duration as the longer streams. However, this approach severely degrades transmission efficiency [11]. Transmission efficiency is also analyzed in terms of errors occurring during frame size setting because when a frame size is larger the impact of the overhead frame can be lower, but they are also more vulnerable to transmission errors. This trade-off is addressed by frame aggregation techniques to derive the optimal frame size that would maximize transmission efficiency. The IEEE 802.11n/ac standard introduces two basic aggregation methods: the aggregated MAC service data unit (A-MSDU) and the aggregated MAC protocol data unit (A-MPDU) [9,12]. Thus, to enhance the system throughput and transmission efficiency in every MU-MIMO transmission, signaling overhead and frame error rate should also be reduced. Moreover, there is always a trade between maximizing throughput and minimizing delay when frame aggregation is adopted [12-14] because more frames must wait before transmission in a buffer. In this regard, the adaptive aggregation mechanism (AAM) proposed by [14] studied that there is a trade-off between maximizing throughput and minimizing delay when frame aggregation is adopted. This study considered the varying nature of the packet size and the packet arrival time to assemble the target aggregate packet size in considering the minimum delay. In the work [13], the delay lower limit (DLL) in the IEEE 802.11 wireless network was studied using the DCF model. However, this study considered best-case scenarios such as an ideal channel condition with only one active station that always has frames to send and receive frames and send acknowledgments (ACKs). However, both [13] and [14] were considered single-user WLAN scenarios.

In the case of an MU-MIMO-enabled WLAN, the majority of the existing studies have only attempted to maximize the throughput or minimize the delay. For instance, work [16] proposed a machine learning-based frame size optimization to improve the system throughput performance of WLAN in the downlink MU-MIMO channel. Different channel conditions, traffic patterns, and number of competing stations are considered in this study. However, the effect of delay is not studied while maximizing the throughput. But the trade-off between maximizing throughput and minimizing delay needs to be considered when frame aggregation approaches are adopted [14]. The main contribution of this paper is to propose a machine learning-based frame size optimization algorithm by extending our earlier approach in considering the cost of transmission delay. In this approach, the optimal frame size setting is achieved by adopting an adaptive aggregation approach that can cope with the dynamic effects of the traffic pattern and channel conditions.

Due to the increasing complexity of wireless networks along with unsophisticated deployment, distributed management systems, and density of network systems, it is becoming a challenge to efficiently operate the IEEE 802.11 networks. The dominant approach to solving these performance-related problems is to apply machine learning (ML). ML is a type of artificial intelligence where algorithms can learn from training data without being explicitly programmed [15]. There are two types of learning strategies in machine learning such as online learning (also called incremental learning) and offline learning (or batch learning) [15 -17]. In online learning, the algorithm updates its parameters after learning from each training instance which allows the learning algorithm to keep learning on the fly [16,17]. Our proposed ML approach adopts an online learning approach to cope with the time-varying channel conditions and traffic patterns. In this approach, the Access Point (AP) collects the "frame size-system throughput patterns" which contain knowledge about the effects of transmission delay), traffic condition, channel condition, and a number of stations (STAs). Based on these patterns, the neural network is used to correctly model the system throughput as a function of the system frame size. Once the training is completed by the neural network, the gradient information obtained is used to adjust the system frame size. Finally, the effectiveness of the proposed scheme is validated under various traffic patterns, channel conditions, and the number of STAs for WLAN downlink MU-MIMO channels.

The rest of the paper is organized as follows, in Section 2, we introduce related works about the frame aggregation schemes and the performance challenges of multi-user transmissions in the WLAN downlink MU-MIMO channel. The description of the proposed approach is discussed in Section 3. In Section 4, results and discussions are presented to evaluate the performance of the proposed approach under various channel conditions, traffic models, and a number of stations. Finally, the conclusions are given in Section 5.

2. RELATED WORK

In this section, some previous studies on frame size optimization problems in WLAN will be discussed. According to [16], a machine learning-based frame size optimization approach is proposed aiming to improve the system throughput performance of WLAN in the downlink MU-MIMO channel. In this approach different channel conditions, traffic patterns, and a number of competing stations are considered to evaluate the performance of the proposed approach. However, this approach did not study the effect of delay while maximizing throughput thus, it provides a suboptimal solution. The work in [17] proposed a machine learning-based frame size optimization approach in considering both channel conditions and contention effects of users to maximize the throughput performance of WLAN. However, this approach similarly ignored the issue of delay, and it can not operate in the IEEE 802.11 MU-MIMO-enabled WLAN. In addressing the energy-throughput trade-off, [18] proposed an online learning-based frame aggregation, Intelligent Energy-Efficient Frame Aggregation (IE2FA), to design an energy-efficient MAC for high throughput wireless local area network (HT-WLAN). According to the simulation results, they improved network performance significantly compared to the other existing studies. Aiming to maximize goodput, [19] proposed an adaptive ML-based approach for frame size selection on a per-user basis by taking into account channel conditions and global performance indicators. The main approach of this study is to propose ML techniques in the specific case of Software-Defined Wireless Local Area Networks (SD-WLAN) particularly focusing on frame length optimization. According to the results, by analyzing a multitude scenarios, an average improvement of 18.36% is achieved in goodput over standard aggregation mechanisms. However, the issue of delay is not addressed in both [18] and [19]. [20] proposed a frame-aggregation size determination approach in the WLAN downlink MU-MIMO channel to improve channel utilization. This approach considered the issue of delay for frames waiting in transmission queues and attempted to reduce the delay by appropriately determining the aggregation size according to the traffic variation. However, the authors did not elaborate on the effects of channel errors, and different traffic situations, also the study mainly focused to enhance channel utilization. The work in [21] proposed an adaptive aggregation algorithm in the downlink MU-MIMO channel in considering the issue of minimizing the cost of increased delay. This approach considers both queue delay and transmission delay to improve the system throughput performic of WLAN. According to the results, maximum throughput performance is achieved with a minimum delay over the baseline FIFO algorithm evaluated under the effects of traffic patterns, channel conditions, and a number of stations. The main difference between our approach and [21] is that we adopted a machine learning-based frame size optimization solution to enhance the system throughput of WLAN in considering the cost of transmission delay in particular.

According to [22], queueing length and number of active nodes have significant impacts on the system throughput performance. Thus, their frame size-based frame aggregation scheme achieved the maximum system throughput performance by generating the same frame length in all spatial streams. However, the limitation of this approach is that the issue of delay is ignored, and the frame aggregation policy is not adaptive to the change in the conditions of traffic and channel conditions. By employing a novel user selection criterion that provides a high priority to the stations expecting high throughput in the next MU-MIMO transmission and having a large amount of data to reduce signaling overhead, [23] proposed a frame duration-based frame aggregation scheme. The main approach of this study is equalizing the transmission time of all spatial streams in all streams according to their Modulation and Coding (MCS) level thus, the maximum system throughput performance is achieved by minimizing the space channel time in the WLAN downlink MU-MIMO channel. However, the main focus of this approach is to improve the system throughput of WLAN without studying the issue of delay. Moreover, in focusing on the effects of adding a padding bit, [24,25] provides a strategy of replacing padding bits with data frames from other users in one stream, thus the space of frame padding bits will be filled by important frames aiming to improve the transmission efficiency. However, this approach violates the rules of MU transmissions by increasing the complexity of both the transmission and reception process to allow the transmission to multiple destinations within a special stream. According to the literature, there is little research exploring the use of ML techniques to tackle frame size optimization problems in WLAN, particularly in considering the trade-off between maximizing the throughput and minimizing delay.

3. PROPOSED APPROACH

Aiming to tackle the effects of heterogenous traffic demand among streams in WLAN downlink MU-MIMO, a machine-learning-based adaptive approach is proposed for frame-size optimization that attempts to maximize the system throughput performance of WLAN in minimizing delay. Different traffic models are adopted such as (Pareto, Weibull, or fractional Brownian Motion (fBM) [9, 10] to generate different traffic scenarios. The AP collected the data set as a pattern of "frame size–system throughput "which contains knowledge about the effects of traffic patterns, channel conditions, number of stations, and minimum transmission delay using the simulation environment [21] which considers both transmission delay and queue delay. However, the data collection in this study particularly considers transmission delay. Transmission delay is defined in this study as the time duration a station takes from sensing the channel state i.e., idle or busy to transmitting a frame until it receives the Acknowledgment (ACK) of the frame [19]. A specific aggregation algorithm cannot always contribute to the maximum throughput performance in minimizing transmission delay under the actual dynamic conditions of heterogeneous traffic patterns and channel conditions of WLAN. In addressing these challenges, different frame aggregation policies such as (FIFO FA (baseline approach), Equal Frame Size FA, Equal MPDUs Agg FA, and Avg Num MPDUs FA) are adopted [9,10, 19] in this experiment to achieve the adaptive aggregation strategy. Moreover, the time required for completing a single MU-MIMO transmission depends on the types of aggregation policies employed. Thus, the adaptive aggregation approach is significant in determining the minimum delay. In downlink MU-MIMO transmission, the longer frame determines the transmission delay in a single MU- MIMO transmission [19] which is defined as max *(TData)* in formula (1) which presents the mathematical expression used for the transmission delay considered in this study.

$$TDelay_t = TDIFS + BOTime + max(TData) + NumSTA(TSIFS + TBA) + \ldots \quad (1)$$

Therefore, in the simulation, the average system transmission delay ($AverageTransDelay$) is computed per MU-MIMO transmission as the ratio of the sum of the transmission delay and the number of stations, as illustrated in formula (2). Where n represents a number of stations in the network and *t* is the simulation time.

$$AverageTransDelay_t = \frac{\sum_{i=1}^{n} TDelay_t}{n} \quad (2)$$

The average delay of each aggregation policy is evaluated to determine the optimal aggregation policy that provides the minimum system delay. Thus, the optimal aggregation policy is always the one that provides the minimum transmission delay [21].

Under the actual dynamic effects of channel conditions and traffic patterns, it is difficult to obtain an accurate throughput function *f(frm)* in all network conditions. The main contribution of this study is to study how the throughput *Thr* can be maximized by optimizing the frame size *frm*. To solve such an optimization problem the well-known gradient ascent algorithm is adopted in this study that can build the knowledge and accurately model the throughput *Thr* as a function of the frame size *frm*. After the knowledge building, the gradient information obtained from the neural networks is used to adaptively adjust the frame size based on the gradient information. In the formation of the frame-size optimization problem, the throughput *Thr* is a complex function of the frame size *frm* under the conditions of minimum delay, channel conditions, traffic patterns, and a number of stations. Formula (3) shows the throughput function *f* varies to the frame size due to the above-specified changing conditions.

$$frm_{Opt} = \underset{frm}{\operatorname{argmax}} Thr = \underset{frm}{\operatorname{argmax}} f(frm) \quad (3)$$

Thus, the local maximum of the throughput function *Thr = f(frm)* can be found by adaptively adjusting frame size *frm* using gradient ascent, by taking steps that are proportional to the gradient. For instance, at the n^{th} time of adjustment, the frame size was *frm(n)*, and the throughput was *Thr(n)*. At the next time of adjustment, the frame size *frm* will be set as shown in formula (4).

$$frm(n+1) = frm(n) + \Delta frm(n) \quad (4)$$

Where $\Delta frm(n)$ depends on the gradient of the estimated throughput $Thr(n)$ with respect to $frm(n)$, i.e.,

$$\Delta frm(n) = \mu \frac{\partial Thr(n)}{\partial frm(n)} \quad (5)$$

The parameter μ is a variable adjustment rate heuristically selected for different network scenarios.

To cope with the effects of time-varying channel conditions and heterogeneous traffic patterns, the online machine learning strategy is employed to achieve the data collection, knowledge building, and frame-size adjustment kept online. The proposed MLP ML approach consists of one hidden layer with four neurons and an output layer. The backpropagation algorithm only consists of two passes: 1) a forward pass and 2) a backward pass [16,17]. To obtain the gradient information $\partial Thr(n)/\partial frm(n)$ which is used to adjust the frame size, we add a third pass, i.e., the tuning pass that will be discussed in Section 3.1. In the proposed MLP approach, the backpropagation algorithm is used to adjust the network and minimize the error between the actual response and the desired (target). The description and summary of notations used in this study are presented in Table 1.

Table 1. Simulation Parameter and Notation Summary

Parameters	**Symbol**	**Value**
# Of Antenna at AP	N_{Ant}	4
# Of Stations	Num_{STA}	2–4
VoIP traffic payload size		100Byte
Video traffic payload size		1000Byte
Learning Rate	η	0.5
Mean Square Error Threshold	MES	0.00001
Epoch Threshold		1000 times
Activation Function	Sigmoid (σ)	
Number of training patterns	n	
Indices of neurons in different layers	i, j	
Frame size(input) of nth training pattern	$frm(n)$	
Target response for neuron j	$Thr(n)$	
Actual response of the nth training pattern	$\widetilde{Thr}(n)$	
Synaptic weight in layer l connecting the output neuron of i to the input neuron j at iteration n	$w_{ji}^{(l)}$	
Weight sum of all synaptic inputs plus bias of neuron j in layer l at iteration n.	$v_j^l(n)$	
Signal of output of neuron j in layer l at iteration n	$y_j^l(n)$	
Local gradient of neuron j in layer l in the tuning pass of hidden layer	$\lambda_j^l(n)$	
Local gradient of neuron j in layer l in the tuning pass of the output layer	$\lambda_1^2(n)$	
Adjustment rate	μ	

3.1. Tuning Pass Strategy

The diagram shown in Figure 1 illustrates the signal flow of the tuning pass in the machine learning model to estimate the gradient $\frac{\partial \widetilde{Thr}(n)}{\partial frm(n)}$ and the key to adjusting the frame size to maximize the throughput in minimizing delay. The initial weight is denoted as w_{ij}^i in the neural network is randomly chosen. The synaptic weights that have been well adjusted in the backward pass are set as fixed in the tuning pass. An adaptive learning rate is adopted to improve the convergence speed [16].

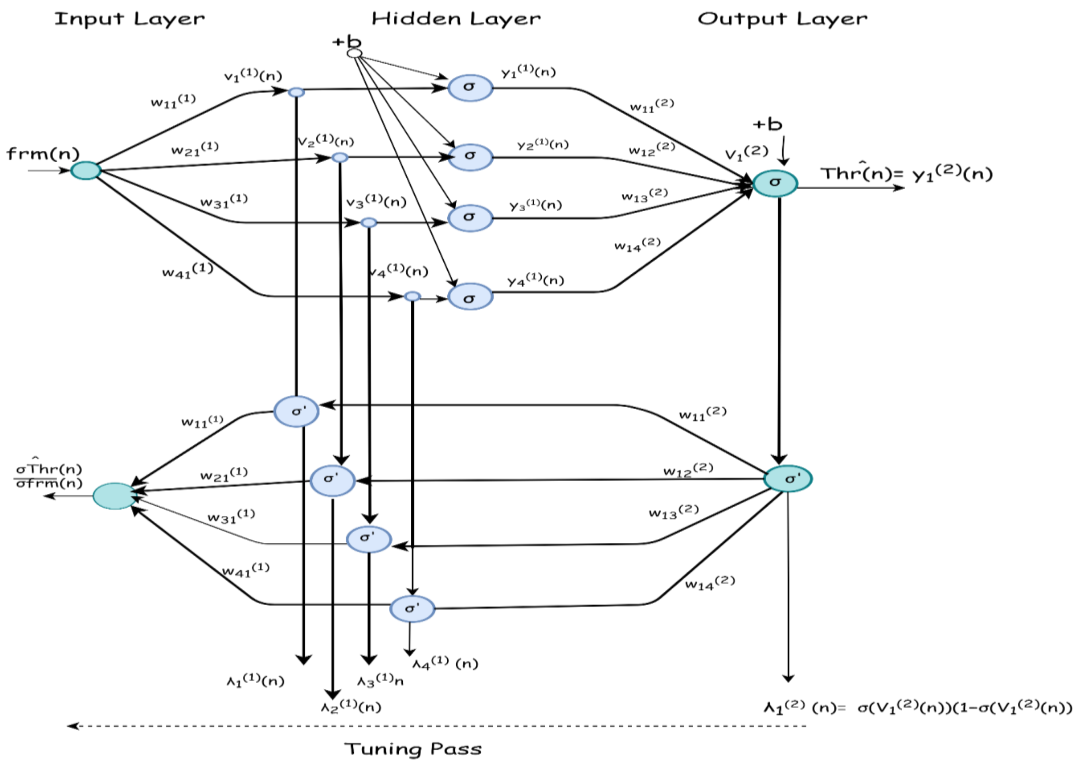


Figure 1. Flow chart of the proposed machine learning approach consisting of tuning pass, which depicts the derivation of the local gradients and the gradient for frame size adjustment.

In the following discussion, the procedure how the estimated gradient $\frac{\partial \widetilde{Thr}(n)}{\partial frm(n)}$ is computed by adopting [16]. Considering the hidden layer, the local gradient $\lambda_j^{(l)}(n)$ for the tuning pass is defined as follows:

$$\lambda_j^{(l)}(n) = \frac{\partial \widetilde{Thr}}{\partial v_j^{(l)}(n)} \tag{6}$$

Where $v_j^{(l)}$ in formula (6) is the weight sum of synaptic input plus bias of neuron j in layer l. Similarly, considering the output layer, the local gradient $\lambda_1^{(2)}(n)$ is defined as follows:

$$\lambda_1^{(2)}(n) = \frac{\partial \widetilde{Thr}(n)}{\partial v_1^{(2)}(n)} = \partial'(\,v_1^{(2)}(n)\,) = \partial(v_1^{(2)}(n))(1 - \partial(v_1^{(2)}(n)) \tag{7}$$

While considering the hidden layer, the local gradient $\lambda_j^{(l)}(n)$ can be expressed as follows using the chain rule:

$$\lambda_j^{(1)}(n) = \frac{\partial \widetilde{Thr}(n)}{\partial v_j^{(1)}(n)} = \frac{\partial \widetilde{Thr}(n}{\partial v_1^{(2)}(n)} \cdot \frac{\partial v_1^{(2)}(n)}{\partial y_j^{(1)}(n)} \cdot \frac{\partial y_j^{(1)}(n)}{\partial v_j^{(l)}(n)}$$

$$\lambda_1^{(2)}(n).\;\; w_{1j}^{(2)}(n)\;.\;\partial'(\,v_j^{(1)}(n)) \tag{8}$$

Therefore, using results (7) and (8), the gradient can be written as follows:

$$\frac{\partial \widetilde{Thr}(n}{\partial frm(n)} = \frac{\partial \widetilde{Thr}(n}{\partial v_1^{(2)}(n)} \cdot \frac{\partial v_1^{(2)}(n)}{\partial frm(n)}$$

$$= \lambda_1^{(2)}(n) \cdot \frac{\partial v_1^{(2)}(n)}{\partial frm(n)} \tag{9}$$

Where $v_1^{(2)}(n)$ can be defined as $v_1^{(2)}(n) = \sum_{i=1}^{4} w_{1j}^{(2)}(n).\ y_j^{(1)}(n)$. Thus, the second term at the rightmost side of equation (9) can be written as:

$$\frac{\partial v_1^{(2)}(n)}{\partial frm(n)} = \sum_{i=1}^{4} w_{1j}^{(2)}(n). \frac{\partial y_j^{(1)}(n)}{\partial frm(n)}$$

$$= \sum_{i=1}^{4} w_{1j}^{(2)}(n). \frac{\partial y_j^{(1)}(n)}{\partial v_j^{(1)}(n)} \cdot \frac{\partial v_j^{(1)}(n)}{\partial frm(n)}$$

$$= \sum_{i=1}^{4} w_{1j}^{(2)}(n).\ \partial'(\ v_j^{(1)}(n)\ ).\ \ w_{j1}^{(1)}(n)$$

$$= \sum_{i=1}^{4} \lambda_1^{(2)}(n).\ w_{1j}^{(2)}(n).\ \partial'(\ v_j^{(1)}(n)\ )\ .\ \ w_{j1}^{(1)}(n)$$

$$= \sum_{i=1}^{4}\ \lambda_j^{(1)}(n)\ .\ w_{j1}^{(1)}(n) \tag{10}$$

Therefore, the gradient $\frac{\partial \widetilde{Thr}(n)}{\partial frm(n)} = \sum_{i=1}^{4} \lambda_j^{(1)}(n)\ .\ w_{j1}^{(1)}(n)$ (11)

The derivation of the local gradients at each layer and the gradient $\frac{\partial \widetilde{Thr}(n)}{\partial frm(n)}$ is depicted in Figure 1. Based on the result from equation (11), the frame size *frm* is adjusted as shown in formulas (4) and (5).

4. Results and Discussion

In this section, the experimental procedure is discussed. The performance of the proposed machine learning-based adaptive approach will be evaluated by considering the effects of channel conditions, heterogeneous traffic patterns, and a number of stations.

4.1. Experimental Procedure

This experiment is conducted to enhance the performance of our previous machine-leaning-based adaptive approach for frame size optimization to maximize the system throughput performance of WLAN in the downlink MU-MIMO channel [16]. However, this earlier work didn't consider the issue of delay. Thus, to collect the training data set as a pattern of "frame size - system throughput", we adopt the simulation environment [21]. Because, according to [21] both queue delay and transmission delay are considered in the simulation, but in this experiment, the data collection is conducted only by considering the transmission delay. The frame size in the data set represents the input data set that represents the average offered traffic load generated in the network to obtain the corresponding target system throughput. System throughput is the target data set that determines the maximum system throughput values achieved [21]. The training data set is collected once every 50 seconds in considering different network scenarios such as channel conditions, traffic patterns, number of stations, and the minimum transmission delay. Thus 50 samples will be collected for each training to train the neural network until the stopping criteria of Mean Square Error (MES) fall below 0.00001 or when the training epoch exceeds 1000 times is satisfied by using Forward and backward passes iteratively. Then the weight is updated following the procedure in the backward pass. The error threshold and the maximum number of iterations determine the accuracy of the function and the computing cost. Finally, once the training is over and the knowledge-building model is achieved, the tuning pass will be invoked to adjust the frame size *frm* by using gradient information from the neural network.

We adopted the same simulation parameters considered in [16] as we are attempting to extend the limitation of this earlier work in considering the issue of delay which is one of the significant performance evaluation metrics in networking. To evaluate the performance of the proposed approach, different comparative approaches in terms of system throughput performance are considered in the experiment such as FIFO FA (Baseline Approach) which does not employ an adaptive aggregation scheme, Adaptive FA Conv. Approach 1 [9] is an adaptive aggregation algorithm that did not consider channel error and the cost of delay aiming to achieve only maximize the throughput and Adaptive FA Conv. Approach 2 [10] is an adaptive approach that can achieve the maximum system throughput by considering the effects of channel error and the cost of increased delay. However, both [9, 10] didn't adopt a machine learning-based optimization solution. The abbreviations 'FA' and 'Conv.' in this paper refer to 'frame aggregation' and 'conventional' respectively. Thus, the performance of our proposed ML approach will be evaluated over the FIFO FA (Baseline Approach), Adaptive FA Conv. Approach 1, and Adaptive FA Conv. Approach 2.

In general, the proposed machine-learning-based adaptive approach will be evaluated under the following performance factors. In Section (4.2) the performance of the proposed ML approach is evaluated under the effects of different traffic models such as Pareto, Weibull, and fBM. In Section (4.3) the performance of the proposed approach is evaluated under the effect of channel conditions where SNR= 5, 10, and 20 dB. In Section (4.4) the performance of the proposed approach under a varying number of STAs (2,3,4) is evaluated. Finally, the performance of the proposed ML approach is evaluated in terms of system throughput versus optimal system frame size in Section (4.5). All experiments are conducted with a traffic mix of 50% VoIP and 50% [10] video with a constant frame size of 100 Byte and 1000 Byte, respectively.

4.2. Performance Under the Effect of Various Traffic Models

In this section, the performance of the proposed ML approach is evaluated under the effects of different traffic models such as Pareto, Weibull, and fBM [9], SNR = 10 dB, and Num_{STA}= 4. According to the result, the proposed approach achieved different performances in different traffic models while the same number of stations and channel conditions are considered. For instance, the proposed approach copes better with the Weibull traffic model with a maximum performance of 807Mbps than the Pareto traffic model 667Mbps which is the least even when compared with fBM 705Mbps. These results demonstrated how the optimal system throughput performance of WLAN in the downlink MU-MIMO channel is affected by the heterogeneous traffic patterns among spatial streams in different traffic conditions. The average maximum system throughput achieved by the proposed ML approach, FIFO FA (Baseline Approach), Adaptive FA Conv. Approach 1, and Adaptive FA Conv. Approach 2 under the conditions of different traffic models is illustrated in Table 2.

Table 2. Simulation result achieved by the FIFO FA (Baseline Approach), Adaptive FA Conv. Approach 1, Adaptive FA Conv. Approach 2, and Proposed ML approach for average system throughput performance in Mbps under the effects of different traffic models.

Comparative Approaches	Traffic Models		
	Pareto	Weibull	fBM
FIFO FA (Baseline Approach)	692.87875	397.50775	482.20575
Adaptive FA Conv. Approach 1	868.09675	695.20775	724.73675
Adaptive FA Conv. Approach 2	809.04125	674.4325	706.42475
Proposed ML Approach	807.872	667.94675	705.45175

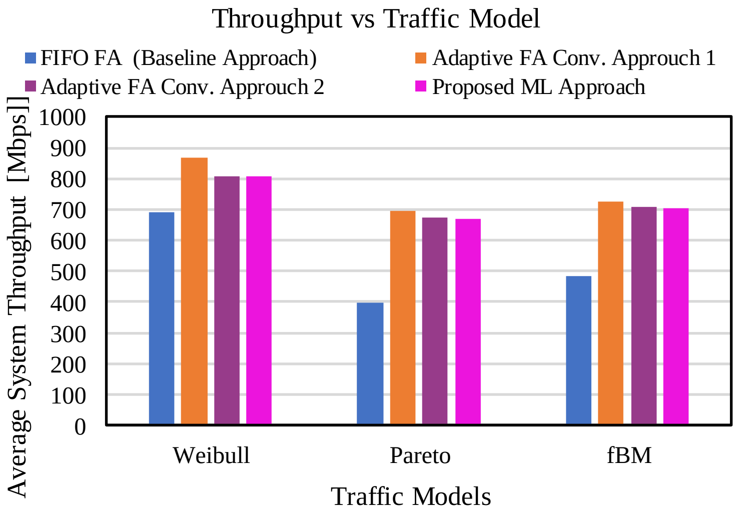


Figure 2. Performance of average system throughput under the effects of heterogenous traffic models when SNR = 10dB.

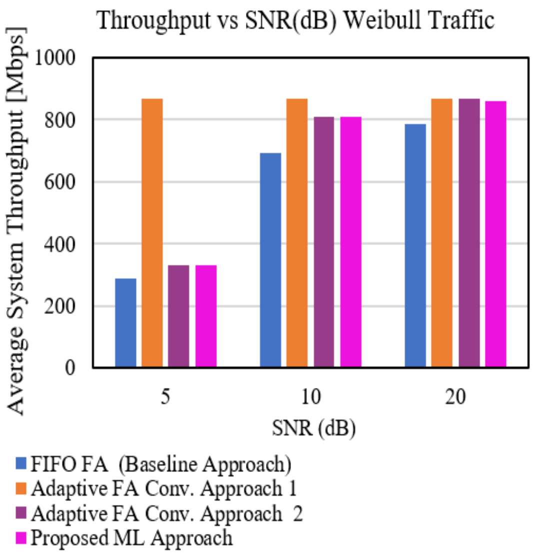
As the result shows in Figure 2, Adaptive FA Conv. Approach 1 achieved the maximum performance in all traffic models as it was only concerned with the maximum throughput with the expense of delay and ideal channel conditions. Similarly, Adaptive FA Conv. Approach 2 achieved a better performance than the proposed approach. But this approach doesn't consider the issue of delay even though transmission error is considered. However, the proposed ML approach achieved the maximum performance of 807Mbps by the Weibull traffic model which is closer to the performance 809Mbps with Adaptive FA Conv. Approach 2. The FIFO (Baseline Approach) achieved the worst performance of all traffic models particularly the least performance of 397Mbps achieved using the Pareto traffic model due to its non-adaptive aggregation policy employed in it. Thus, these results indicate that traffic patterns in the network determine the system's performance.

4.3. Performance Under the Effects of Channel Conditions

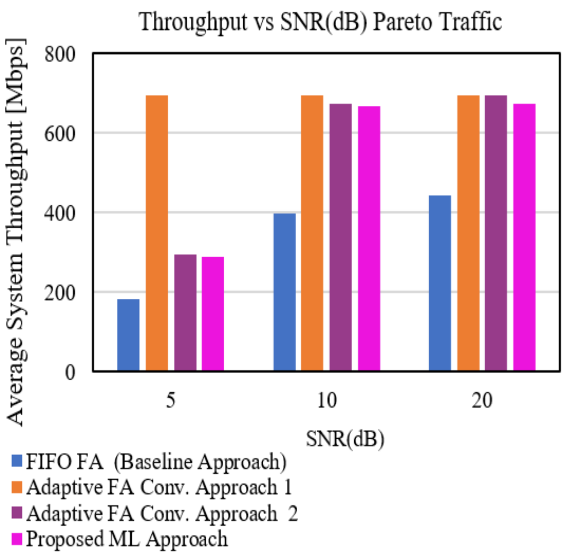
The performance of the proposed approach under different channel conditions when SNR = 5, 10, and 20dB, and Num_{STA} =4 is evaluated in this section. When the SNR value increases the channel quality improves thus the throughput performance is enhanced as the occurrence of transmission frame error rate will be less. As shown in Figure 3 (a), (b), and (c), for the case of different traffic models such as Pareto, Weibull, and fBM, the system throughput performance of all approaches increases when the channel quality improved from 5dB to 20dB. However, the proposed ML approach achieved better performance than that of the FIFO (Baseline Approach) due to the adaptive aggregation approach adopted in the Proposed ML Approach. On the contrary, the Proposed ML approach achieved the lowest performance of 673Mbps in the Pareto traffic model compared with 696Mbps achieved by Adaptive FA Conv. Approach 1 and Adaptive FA Conv. Approach 2 when SNR=20dB. This is due to the reason that both Adaptive FA Conv. Approach 1 and Adaptive FA Conv. Approach 2 approach only focuses on maximizing throughput with a cost of maximum delay which provides a suboptimal solution. On the contrary, under the worst channel condition, e.g. SNR = 5dB in the figure, a better performance of 330Mbps is achieved by the Weibull traffic, and the worst 266Mbps is achieved by fBM traffic. This result illustrates that the system throughput performance is affected by traffic patterns even under the same channel condition. However, the proposed approach achieved maximum performance better than FIFO FA (Baseline Approach) as it employed an adaptive aggregation approach to realize the optimal system throughput. Table 3 illustrates quantitative performance results of the average system throughput performances achieved by the FIFO FA (Baseline Approach), Adaptive FA Conv. Approach 1, and Adaptive FA Conv. Approach 2 and Proposed ML Approach, under the effects of different channel conditions and traffic models.

Table 3. Simulation result achieved by the Proposed ML Approach, Maximum Throughput, and FIFO (Baseline Approach) for average system throughput performance in Mbps under the effects of different traffic models and channel conditions.

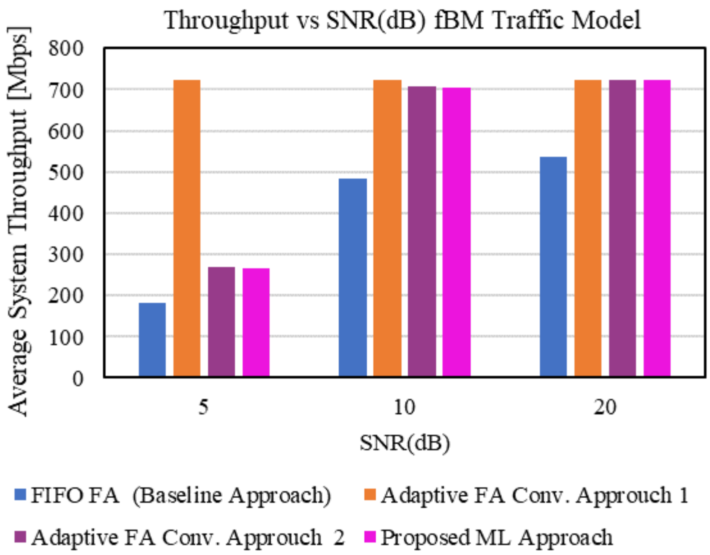
Comparative Approaches	**Traffic Models**	**SNR (dB)**		
		5(dB)	10(dB)	20(dB)
FIFO FA (Baseline Approach)	Weibull	288.020	692.878	784.021
Adaptive FA Conv. Approach 1		868.096	868.096	868.096
Adaptive FA Conv. Approach 2		331.562	809.041	868.096
Proposed ML Approach		330.067	807.872	857.716
FIFO FA (Baseline Approach)	Pareto	181.847	397.507	442.630
Adaptive FA Conv. Approach 1		695.207	695.207	695.207
Adaptive FA Conv. Approach 2		295.262	674.432	695.207
Proposed ML Approach		287.630	667.9467	673.619
FIFO FA (Baseline Approach)	fBM	180.726	482.20575	535.097
Adaptive FA Conv. Approach 1		724.736	724.7367	724.7367
Adaptive FA Conv. Approach 2		267.664	706.4247	724.7367
Proposed ML Approach		266.499	705.4517	723.764



a) Weibull Traffic Model



b) Pareto Traffic Model



c) fBM Traffic Model

Figure 3. System throughput versus SNR for different traffic models such as Pareto, Weibull, and fBM when Num_{STAs}= 4.

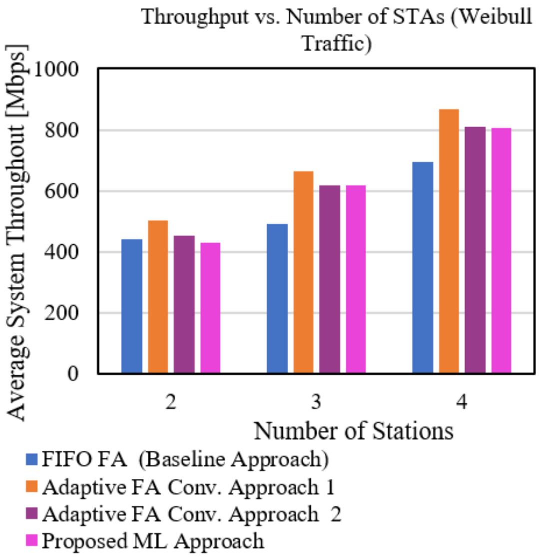
FIFO FA (Baseline Approach) achieved the worst performance 181Mbps and 180Mbps in the worst channel condition of 5dB by the Pareto and fBM traffic models respectively. In general, the FIFO FA (Baseline approach) aggregation policy is the worst compared to the proposed approach in all scenarios because of the non-adaptive aggregation strategy it employs. However, the Proposed ML Approach always achieved the maximum performance close to Adaptive FA Conv. Approach 1, and Adaptive FA Conv. Approach 2 only considers the achievement of maximum throughput with the cost of maximum delay [9,10].

4.4. Performance Under the Effects of Number of Stations

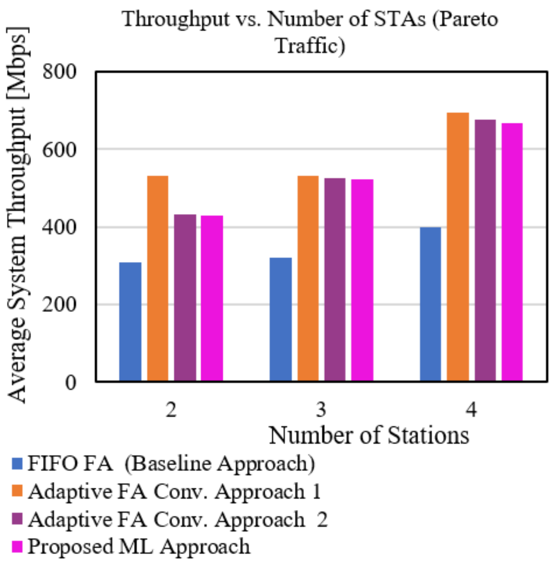
Figure 4 (a), (b), and (c), demonstrate the performance of the proposed approach evaluated under the effect of a different number of stations (Num_{STA} =2, 3, and 4), and when the channel condition is SNR=10dB for the case of Weibull, Pareto, and fBM traffic models. As the results show, the system throughput performance significantly increases in all traffic models as the traffic rate increase when the number of stations ranges from 2 to 4. However, the performance of the proposed ML approach achieved different performances in different traffic models due to the effect of heterogeneous traffic patterns even under the same number of stations. The quantitative comparative results of the average optimal system throughput achieved by the FIFO FA (Baseline Approach), Adaptive FA Conv. Approach 1, Adaptive FA Conv. Approach 2, and the Proposed ML Approach under the effects of a variable number of STAs are illustrated in Table 4.

Table 4. Simulation results achieved by comparative approaches such as FIFO FA (Baseline Approach), Adaptive FA Conv. Approach 1, Adaptive FA Conv. Approach 2, and Proposed ML approach for average system throughput in Mbps under the effects of variable number of stations in Weibull, Pareto, and fBM traffic models.

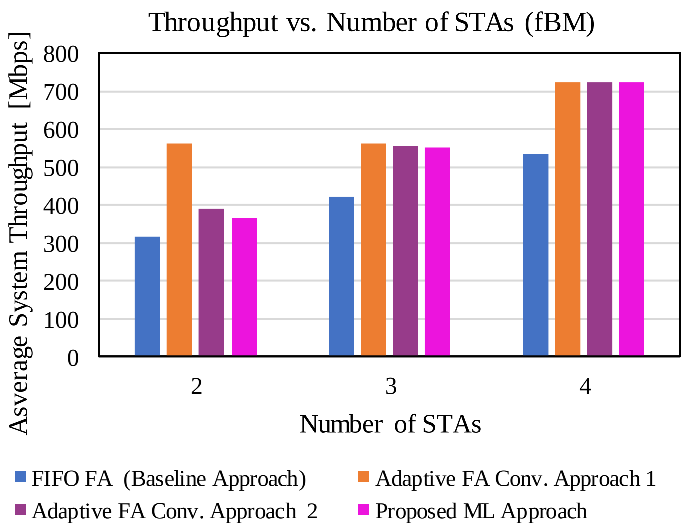
Comparative Approaches	**Traffic Models**	**Number of Stations**		
		2	3	4
FIFO FA (Baseline Approach)	Weibull	440.9687	489.502	692.8787
Adaptive FA Conv. Approach 1		501.84	662.1969	868.0967
Adaptive FA Conv. Approach 2		450.9877	618.7347	809.0412
Proposed ML Approach		428.3877	617.7457	807.872
FIFO FA (Baseline Approach)	Pareto	307.5327	320.9385	397.5077
Adaptive FA Conv. Approach 1		529.9385	529.9385	695.2077
Adaptive FA Conv. Approach 2		432.1387	524.8727	674.432
Proposed ML Approach		428.1055	523.189	667.9467
FIFO FA (Baseline Approach)	fBM	316.3205	421.6932	482.2057
Adaptive FA Conv. Approach 1		564.47	564.47	724.7367
Adaptive FA Conv. Approach 2		389.5777	555.8375	706.4247
Proposed ML Approach		365.8767	553.0777	705.4517



a) Weibull Traffic Model



b) Pareto Traffic Model



c) fBM Traffic Model

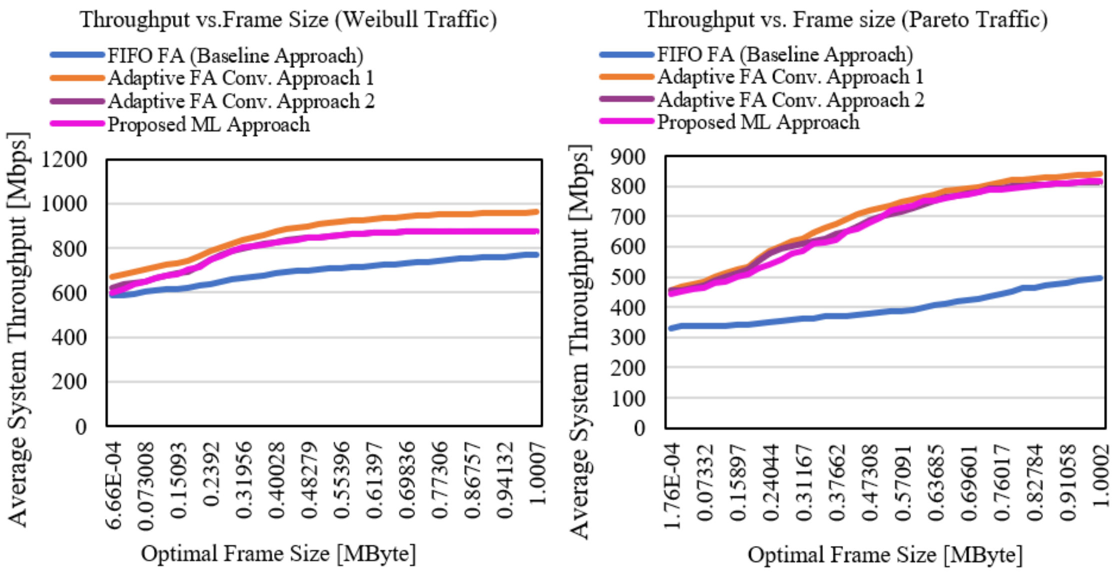
Figure 4. System throughput versus number of stations when the channel condition is SNR =10dB for the Weibull, Pareto, and fBM traffic models.

As shown in the results in Figure 4, when the number of stations increases in all traffic models, the system throughput performance increases as the traffic load increases with an increasing number of stations from 2 to 4. However, different performance is achieved depending on the traffic condition. For instance, 807Mbps is the maximum performance achieved by the Weibull traffic model while 667Mbps is achieved by the Pareto traffic model. The Adaptive FA Conv. Approach 1 outperforms all approaches as it always considers maximizing throughput in ideal channel conditions and the expense of delay. However, the proposed approach always outperforms the FIFO (Baseline Approaches) in all scenarios due to the adaptive aggregation strategy it adopts. In this regard, the maximum performance of 807Mbps in the case of the Weibull traffic model is achieved by the proposed approach whereas the lower performance of 667Mbps is achieved in the Pareto traffic with the same number of STAs equals 4. Likewise, when the number of stations equals 2, the maximum performance of 428Mbps is achieved by the

Weibull and Pareto traffic models and the least 365Mbps by the fBM traffic model. These results show that number of stations affects the performance of the system throughput under the conditions of heterogeneous traffic patterns among streams in the downlink MU-MIMO channel. However, the proposed approach always achieved the maximum system throughput performance better than the FIFO FA (Baseline Approach) but closest to the Adaptive FA Conv. Approach 1, Adaptive FA Conv. Approach 2 [9,10] which are only considered the maximum throughput with the expense of delay.

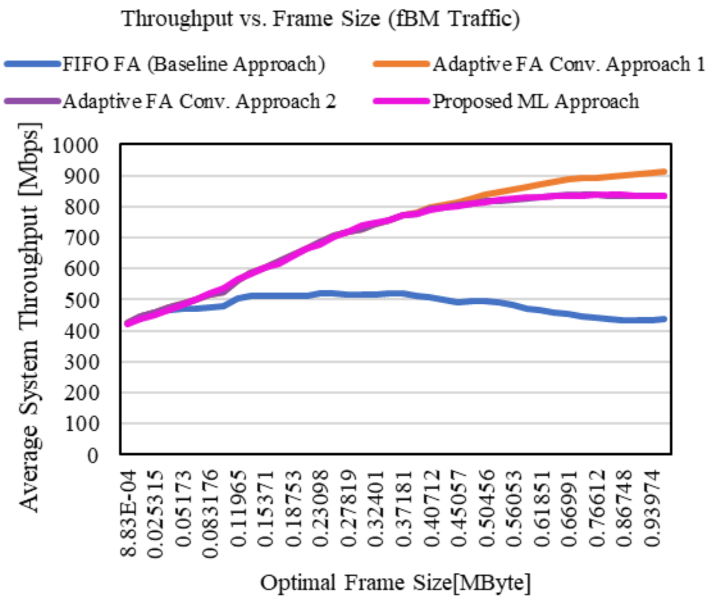
4.5. Performance of System Throughput Vs. Optimal Frame Size

Figure 5 (a), (b), and (c) show the performance of system throughput with increasing System frame size considering SNR= 10 dB, Num_{STA} = 4. To examine the effect of traffic conditions on the performance of the proposed ML approach, traffic models such as Weibull, Pareto, and fBM are considered. As the results show, the throughput performance of the proposed ML approach increased when the size of the frame increased. However different traffic models achieved different performances due to the effects of heterogeneous traffic patterns. For instance, the maximum performance of 874Mbps is achieved when the optimal frame size is 1Mbyte by the Weibull traffic model whereas the 816Mbps is the maximum performance by the Pareto when the frame size is 1Mbyte. Adaptive FA Conv. Approach 1 achieved the maximum performance in all traffic models because this approach only focuses on maximizing the throughput with the expense of maximum delay and free from transmission error. However, the FIFO FA (Baseline Approach) achieved the worst performance among all traffic models as this algorithm doesn't employ an adaptive aggreging strategy to maximize the throughput. For instance, 437Mbps is the least performance achieved by fBM traffic model when the frame size is 0.93Mbyte.



a) Weibull Traffic Model

b) Pareto Traffic Model



c) fBM Traffic Model

Figure 5. Performance of system throughput versus optimal System frame size when Num_{STAs} = 4 and SNR =10dB for the Weibull, Pareto, and fBM traffic models.

In general, the proposed ML approach provides a significant performance in optimizing the system frame size of WLAN to maximize the system throughput performance with a minimum cost of delay. This is achieved by employing an adaptive aggregation approach by considering channel conditions, traffic patterns, and a number of stations than the baseline FIFO FA (Baseline Approach). Moreover, the proposed approach achieved the closest performance to that of the Adaptive FA Conv. Approach 1 and Adaptive FA Conv. Approach 2 only considers maximizing the throughput without considering the issue of delay.

5. CONCLUSIONS

The one of key enhancements is MAC layer frame aggregation introduced by the IEEE 802.11n/ac to accommodate the growing traffic demand in the network by allowing multiple packets aggregated at once in a single transmission. By adopting frame aggregation, the overhead in the MAC layer is controlled and a tremendous throughput enhancement is achieved. However, due to the heterogeneous traffic demand among streams in the WLAN downlink MU-MIMO channel, it is challenging to efficiently utilize the benefits of frame aggregation. Because when shorter and longer streams are grouped in downlink MU-MIMO transmission, wasted space channel time will occur which is a time duration where a part of spatial streams carries data frames whereas the others do not thus it would degrade the transmission efficacy. Thus, the trade-off between maximizing frame size and minimizing overhead frames should be addressed by adopting an adaptive frame aggregation technique to derive the optimal frame size that would maximize the throughput performance of WLAN. Delay is another critical issue that needs to be taken into account when frame aggregation is employed because frame aggregation experiences more frames waiting in a buffer before transmission which degraded the performance of WLAN. The main contribution of this paper is to propose a machine learning-based frame size optimization algorithm to improve the throughput performance of WLAN in downlink Mu-MIMO channel by extending our earlier approach in considering the cost of delay. In this approach, the optimal frame size setting in the WLAN downlink MU-MIMO channel is achieved by adopting an adaptive aggregation approach that considers the dynamic effects of traffic patterns and channel conditions. The effectiveness of the proposed approach is evaluated over the FIFO Baseline Approach and earlier conventional approaches under various traffic patterns, channel conditions, and a number of STAs for WLAN downlink MU-MIMO channels. Future work will extend our approach by considering real traffic data. Study the effects of other delay factors such as processing delay and queuing delay in considering both uplink and downlink transmission channels of WLAN.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We would like to thank all authors for their contributions and the success of this manuscript and all editors and anonymous reviewers of this manuscript.

REFERENCES

[1] Mugen Peng, Wenbo Wang, and Hsiao-Hwa Chen, (2010) “TD-SCDMA evolution,” IEEE Veh. Technol. Mag., vol. 5, No. 2, pp. 28–41.

[2] IEEE Std 802.11ac-2013; Part11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Amendment 4: Enhancements for Very High Throughput for Operation in Bands below 6 GHz, 2013. DOI:10.1109/IEEESTD.2013.6687187

[3] Caire Giuseppe, and Shlomo Shamai, (2000) "On achievable rates in a multi-antenna broadcast downlink." In Proc. 38th Annu. Allerton Conf. Communication, Control, and Computing, pp. 1188-1193.

[4] Jiang Ming, and Lajos Hanzo,(2007) "Multiuser MIMO-OFDM for next-generation wireless systems." Proceedings of the IEEE 95, No. 7 pp: 1430-1469.

[5] Li Tianji, Qiang Ni, David Malone, Douglas Leith, Yang Xiao, and Thierry Turletti. (2009) "Aggregation with fragment retransmission for very high-speed WLANs." IEEE/ACM transactions on networking, No. 2, pp: 591-604.

[6] Lin Yuxia, and Vincent WS Wong. (2006) "WSN01-1: Frame aggregation and optimal frame size adaptation for IEEE 802.11 n WLANs." In IEEE Globecom, pp. 1-6. IEEE.

[7] Skordoulis Dionysios, Qiang Ni, Hsiao-Hwa Chen, Adrian P. Stephens, Changwen Liu, and Abbas Jamalipour. (2008) "IEEE 802.11 n MAC frame aggregation mechanisms for next-generation high-throughput WLANs." IEEE Wireless Communications, Vol. 15, no. 1, pp: 40-47.

[8] Ong Eng Hwee, Jarkko Kneckt, Olli Alanen, Zheng Chang, Toni Huovinen, and Timo Nihtilä. (2011) "IEEE 802.11 ac: Enhancements for very high throughput WLANs." In 2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 849-853. IEEE, 2011.

[9] Lemlem Kassa, Mark Davis, Jingye Cai, and Jianhua Deng, (2021) "A New Adaptive Frame Aggregation Method for Downlink WLAN MU-MIMO Channels." Journal of Communication, Vol. 16, No. 8, pp 311-322.

[10] Lemlem Kassa, Mark Davis, Jianhua Deng, and Jingye Cai, (2022) "Performance of an Adaptive Aggregation Mechanism in a Noisy WLAN Downlink MU-MIMO Channel." Electronics, Vol. 11, No. 5, pp 754.

[11] Kim Sanghyun, and Ji-Hoon Yun, (2019) "Efficient frame construction for multi-user transmission in IEEE 802.11 WLANs." IEEE Transactions on Vehicular Technology, Vol. 68, No. 6, pp 5859-5870.

[12] Teymoori Peyman, Nasser Yazdani, Seyyed Alireza Hoseini, and Mohammad Reza Effatparvar, (2010) "Analyzing delay limits of high-speed wireless ad hoc networks based on IEEE 802.11n" In 2010 5th International Symposium on Telecommunications, pp. 489-495. IEEE.

[13] Xiao Yang, and Jon Rosdahl, (2022) "Throughput and delay limits of IEEE 802.11." IEEE Communications letters, Vol. 6, No. 8, pp: 355-357.

[14] Jianhua Deng, Mark Davis, (2013) "An adaptive packet aggregation algorithm for wireless networks." In Proceedings of the 2013 International Conference on Wireless Communications and Signal Processing, Hangzhou, China, 24–26 October 2013; pp. 1–6.

[15] Szott Szymon, Katarzyna Kosek-Szott, Piotr Gawłowicz, Jorge Torres Gómez, Boris Bellalta, Anatolij Zubow, and Falko Dressler. (2022) "Wi-Fi Meets ML: A Survey on Improving IEEE 802.11 Performance With Machine Learning." IEEE Communications Surveys & Tutorials, Vol.24, No. 3, pp : 1843-1893.

[16] Lemlem Kassa, Jianhua Deng, Mark Davis, and Jingye Cai. (2022) "Frame Size Optimization Using a Machine Learning Approach in WLAN Downlink MU-MIMO Channel."In the 8th International Conference of Networks, Communications, Wireless and Mobile Computing (NCWC). arXiv preprint arXiv:2209.09687, 2022, Copenhagen, Denmark.

[17] Lin Pochiang, and Tsungnan Lin, (2009) "Machine-learning-based adaptive approach for frame-size optimization in wireless LAN environments." IEEE transactions on vehicular technology, Vol. 58, No. 9 pp 5060-5073.

[18] Karmakar Raja, (2019) "Online learning-based energy-efficient frame aggregation in high throughput WLANs." IEEE Communications Letters, Vol. 23, No. 4, pp: 712-715.

[19] Coronado Estefanía, Abin Thomas, and Roberto Riggio. (2020) "Adaptive ml-based frame length optimization in enterprise SD-WLANs." Journal of Network and Systems Management, Vol. 28, No. 4, pp: 850-881.

[20] Moriyama Tomokazu, Ryo Yamamoto, Satoshi Ohzahata, and Toshihiko Kato, (2017) "Frame aggregation size determination for IEEE 802.11 ac WLAN considering channel utilization and transfer delay." ICETE Proc. 14th Int Jt Conf E-Bus Telecommun. Vol. 6, pp 89–94.

[21] Lemlem Kassa, Jianhua Deng, Mark Davis and Jingye Cai, (2022). Performance of WLAN in Downlink MU-MIMO Channel with the Least Cost in Terms of Increased Delay. Electronics, , 11(18), 2851.

[22] Bellalta Boris, Jaume Barcelo, Dirk Staehle, Alexey Vinel, and Miquel Oliver, (2012) "On the performance of packet aggregation in IEEE 802.11 ac MU-MIMO WLANs." IEEE Communications Letters, Vol. 16, No. 10, pp 1588-1591.

[23] Nomura Yoshihide, Kazuo Mori, and Hideo Kobayashi, (2016) "High-Efficient Frame Aggregation with Frame Size Adaptation for Downlink MU-MIMO Wireless LANs." IEICE Transactions on Communications, Vol. 99, No. 7, pp 1584-1592.

[24] Lin Chi-Han, Yi-Ting Chen, Kate Ching-Ju Lin, and Wen-Tsuen Chen, (2018) "Fdof: Enhancing channel utilization for 802.11 ac." IEEE/ACM Transactions on Networking, Vol. 26, No. 1, pp 465-477.

[25] Lin Chi-Han, Yi-Ting Chen, Kate Ching-Ju Lin, and Wen-Tsuen Chen, (2017) "acPad: Enhancing channel utilization for 802.11 ac using packet padding." In IEEE INFOCOM 2017-IEEE Conference on Computer Communications, pp. 1-9.

AUTHORS

Lemlem Kassa has been a senior lecturer in Addis Ababa Science and Technology university in Addis Ababa, Ethiopia since February 2013. She is currently pursuing her PhD study in School of Information and Software Engineering in University of Electronic Science and Technology of China (UESTC) in China. She received the B.S. degree from Micro link Information Technology college in Addis Ababa Ethiopia in 2006 in software engineering and M.S. degree from University Putra Malaysia (UPM), Malaysia in 2009. Her research interests are in the area of wireless communications, artificial intelligence, mobile computing, and software designing.

Dr. Jianhua Deng graduated in information security from the University of Electronic Science and Technology of China (UESTC), China, in 2006. After graduated, he joined the School of Computer Science and Engineering at UESTC as a staff. From 2009 to 2014, he was a Ph.D. student in Dublin Institute of Technology (DIT) in Ireland and received Ph.D. degree in electrical engineering from DIT in 2014. Now, he is a vice professor in the School of Information and Software Engineering at UESTC. His research interests are in the area of wireless communication, statistical machine learning, artificial intelligence, deep learning. He is the reviewer of some SCI journals (e.g. wireless personal communications).

Prof. Mark Davis received his BE, MEngSc, and PhD degrees from University College Dublin in 1986, 1989 and 1992 respectively. He is currently the director of the Communications Network Research Institute at Technological University Dublin (TU Dublin). His research interests are in the area of radio resource management techniques for wireless networks, specifically IEEE 802.11 WLANs.

Prof. Jingye Cai is Professor and Associate Dean at University of Electronic Science and Technology of China (UESTC). He received BS from Sichuan University in 1983 with a major in radio electronics and Ph.D. from the University of Electronic Science and Technology of China in 1990, with major in signal and information processing. His research interests are in the area of intelligent computing, information engineering, digital Information Processing, and signal processing.