CHAPTER-3: LITERATURE SURVEY

Survey of the literature for the following topics: QoS mechanisms in WLANs, Transmit Beamforming and MU-MIMO are sections 3.1 to 3.7, 3.8 and 3.9 respectively.

3.1 QOS Mechanisms for IEEE 802.11 WLANs

In this chapter, the various techniques of QoS performance improvement in WLANs have been studied, each one focusing on a specific approach. Appendix-2 contains theoretical concepts of WLANs and underlying QOS mechanisms. Each of the paper is studied considering the QoS problem experienced in a particular scenario, the approach used by the authors to enhance QoS and finally an analysis of the solution. Exhaustive survey of all the mechanisms is avoided as the idea is to demonstrate the range of techniques which impact QoS rather than the extensive work in this area, which is very vast. Some of the algorithms designed by the authors in this paper are in Table 3.1

Table 3.1 QoS Mechanisms discussed in literature

Sl.No.	**QoS Techniques**	**Algorithms**				
1	Queueing	ZCD				
2	Aggregation	SA				
3	Rate Adaptation	FD & RS	HAMM/CP	ZIG ZAG		
4	Scheduler	QDRA	ADWISER	CAT	ETA-EQN	WTN
5	Additional mechanisms	ROC	FAIR-WLAN	M-HCCA		

Appendix-4 provides background on Schedulers used in Cellular networks.

Appendix-5 contains characteristics of VOIP. Previous surveys in this area are in [9], [10] and [11].

3.2 Queueing

Drop tail, one of the most widespread passive buffer management algorithms conventionally deployed in Internet routers, is computationally uncomplicated. However, when this algorithm is employed in WLANs, it can greatly decrease the throughput and lead to priority inversion.

In [12], the authors study various solutions to resolve this problem. Manipulating of various EDCA parameters (reducing the AIFS or the CW) is not found to be very helpful. Eliminating RTS/CTS mechanism shows performance enhancement, but root cause of the issue is not isolated. Furthermore, RTS/CTS elimination is not recommended as it has multiple uses. Another simple approach would be partitioning memory. Each partition will be used exclusively by each AC. Weighted Random Early Detection (WRED), an extension to the RED mechanism is suitable for 802.11e and is considered next. WRED thresholds must be cautiously selected in order to get the optimum performance. The authors propose an AQM mechanism coined ZCD or Zero Configuration Drop. Frames transmitted to PHY are examined by ZCD and the values of the ACs distribution as also the average packet size are noted. Using these values, the egress queues are controlled by setting a maximum value on the memory which individual access categories can be granted while ensuring that packets do not get discarded. It is concluded that while the priority inversion problem is resolved by WRED, multiple memory pools and ZCD, ZCD leads to the highest throughput with minimum configuration.

3.3 Call Admission Control

In [13], the AP constantly senses the network condition and maintains records of the events occurring. When there is a request for a new session, the proposed algorithm gets activated and computes the quantity of airtime resources consumed by means of the measurements made at the AP. It also assesses the past history of transmission rates and evaluates how well the overall system performance is affected when the new session is admitted. Based on whether sufficient resources are obtainable to maintain the new call without obstructing existing sessions, the Access Point notifies the requesting STA whether the call is accepted or rejected.

3.4 Aggregation

802.11n Aggregation is a technique whereby multiple packets can be transmitted simultaneously, and not by waiting for each individual STA to complete smaller packets.

Problem: 802.11e attempts to offer QoS by providing varying amount of backoff periods for individual TCs. As an instance, a voice packet (highest priority TC) has less waiting time before being sent as compared to a BE packet. This results in traffic flows with lesser priority undergoing starvation. Thus, 802.11n performs better in the absence of 802.11e features.

Solution: Smart aggregation (SA) framework [14] makes certain changes to the standard. While the last packet is getting set, the two final blocks of the normal aggregation mechanism are swapped. The packets are prepared during the time contention for the channel is taking place, till just prior to dispatching. This permits filling the aggregated frame using additional packets, as newer ones should have reached the buffers during the contention period. The collisions are reduced as there is a reduction in the attempts in accessing the channel. **Conclusion:** Simulation results demonstrate that SA is beneficial for all the ACs, for both delay and throughput. This is true both when the network is densely crowded as well as only lightly crowded.

3.5 Rate Adaptation

The limitations of the Existing RA Schemes:

Slow response of ARF-based RA schemes, (i.e., ARF, AARF, and CARA) is problematic for VoWLANs. In RRAA, large number of ewnd frames is sent in order to estimate current PHY layer's FER. This is done so as to eliminate intermittent frame losses and to collect useful samples. This limits RRAA from being sufficiently quick in tracing the fast-fading channel.

IEEE 802.11 MAC: Continuous retransmissions in EDCA result in the packet loss getting worse as explained below. After transmission errors are caused due to a bad channel, the WLAN station attempts to retransmit the frames. However, if the status of the medium continues to be unimproved as well as PHY data rate has not changed, performance of VoWLANs using EDCA will become even worse with retransmissions.

3.5.1 Fast Decrease and Retry Scheduling

In [15], a proposal is offered to enhance the QoS of VoWLANs by resolving the problems of the current RA schemes using two mechanisms mentioned below:

(i) Fast Decrease (FD): Based on the problems mentioned in the above paragraph, it is advisable to use good PHY rates for retransmission in VoWLAN, to ensure good voice quality. FD basically reduces the PHY rate when failures occur due to retransmissions, as the reason for frame loss is either because of collisions or unfavorable channel.

(ii) Retry Scheduling (RS): Still, if the channel continues to be unfavorable as a result of strong fading and/or interference from hidden terminals , FD also may not be of use. For this reason, a scheduling policy is developed with delayed retransmissions and it is referred to as RS. That is to say, VoWLAN station waits until the channel condition improves.

The QoS of VoWLANs is evaluated with various RA schemes.

1. ARF-based families (ARF, Agile ARF with RS, Agile ARF).
2. Families based on FER-statistics (RRAA based algorithms).
3. Genie, an ideal reference RA scheme, which decides the optimum PHY rate in the absence of overhead.

Tests are conducted under 3 topologies.

(i) Static Topology: Here the WLAN has an Access Point and K static VoIP STAs, which are evenly spaced by 20 meters, from each other and the Access Point.

The schemes in 1, 2 and 3 are modified with the proposed mechanism . The R-score is improved with Fast Decrease, while Retry Scheduling provides an additional gain. The performance improvement with FD is particularly significant in RRAA as its original rate selection is relatively very static.

(ii) Random Topology with Heterogeneous Traffic Types: Here, the AP is present at the centre of a circle and VoIP and TCP data stations are arbitrarily installed within 20 m radius. Out of the five static TCP stations, two flows are uplink while the other three are downlink. All the stations employ the same RA algorithm whereas the five static VOIP stations also have the RS functionality. The average R-score in this case is to some extent more than the previous one having only VoIP stations. This is because the VoIP stations move nearer to the AP in the random topology.

Measured against the R-score, RRAA and ARF perform averagely. Modified with FD and RS, they fare far better with Agile ARF with RS able to maintain all VoIP STAs having an R-score > 80.

(iii) Mobile scenario: In this case, there is only one VoIP STA that travels to and fro in a straight line with a speed of 1 meter per second. When the distance of the station from the AP increases, ARF performs badly with RRAA faring still worse. However, with the FD and RS modifications, there is a remarkable improvement in the R-score.

3.5.2 HAMM-CP

Background: Among the IEEE 802.11 MACs providing multicast service, the ARF RA mechanism is most commonly used. ARF chooses the data rate by considering the status of the channel using the ACK packets.

The MCP or Multicast Collision Prevention algorithm has decreased remarkably the rate of multicast collision as compared to the rate which typical multicast mechanisms claim. However, packets corrupted due to the media conditions, cannot be recovered. To do this, a feedback procedure is introduced wherein STAs receiving corrupted packets can ask for retransmission, by sending NACK to the transmitter. ARF algorithm is included in the feedback mechanism. It is proposed to use hierarchical video coders based on the H.264 standard. The MRs facing the most unfavorable channel conditions is given the Base Layer (BL) of the encoded H.264 video, whereas the MRs with improved channel will attain both BL and Enhancement Layers (ELs).

The authors had earlier proposed a scheme ARSM or Auto Rate Selection mechanism for Multicast to detect MR with least SNR. This MR is designated group leader by the AP and it represents the multicast group members in acknowledging the multicast packets. The other MRs can send NACK message on detecting transmission errors. This method of determining the leader is called as MCPO or Multicast Channel Probe Operation.

The ARSM algorithm adjusts the multicast group's PHY rate by considering the receiver whose SNR is present in the lowest range. Thus, ARSM assures that all the group members obtain all packets correctly while also ensuring the packets are sent at the greatest achievable PHY data rate. Conversely, in some situations (e.g., video services), users who are able to receive data at higher rates are made to compromise. For these scenarios, the author proposes Hierarchical ARSM (HARSM) --upgrading ARSM to maintain the services which allow users having varied abilities.

As per this mechanism, video encoding happens in layers - BL and EL. Video BL packets are transmitted to every group member using ARSM rules. Chief variation for EL compared to ARSM is that HARSM chooses a member with maximum SNR. Now the EL rate is chosen based on BL's transmission rate.

Problem: Shortcomings of HARSM and ARSM - (1) incompatibility with the 802.11 standard (2) SNR thresholds utilized by HARSM and ARSM need to require a compromise between maximum data rate for transmission and signaling cost.

Solution: HAMMP/CP [16], is a unique scheme that is QoE-aware and custom-made for functions based on layered-video communications. As per the scalable encoding plan, the elements of a video sequence are handled as:

1. The more significant ones are included in BL. The BL receives preferential treatment and it gives video image of adequate quality.
2. The less significant elements are included in EL and it, therefore, receives less favorable treatment.

Due to this scheme, the encoders are capable of providing communication systems with good QoE.

When the encoders adjust the encoding attributes based on the network feedback, it leads to temporary reduction of video generation rate but preserves satisfactory video features and a graceful degradation in quality.

In [16], a video encoding protocol , H.264 aka MPEG-4 AVC, is used. An SVC bit stream consists of a BL and one or several ELs. Here, the protocol schemes which were earlier discussed and Intra-Access Category differentiation methods of 802.11e EDCA are integrated into the HAMM/CP. Hence a queue design to efficiently transport scalable video to a multicast group is proposed.

Encoded video is transmitted in 2 layers with an assigned User Priority as per HAMM/CP. Packets pertaining to the BL are provided a superior User Priority and queued into a primary video AC (AC VI) whereas the EL is queued into an video AC (AAC VI). Therefore, separate control mechanisms can be employed for the different video streams.

HAMM/CP is a QoE Aware Multicast Mechanism for Scalable Video Communications. In HAMM/CP, the procedure for transmitting the video packets of both Access Categories is initiated by the MAC. The feedback ARF mechanism only controls errors faced by packets holding the BL and the transmission speed is updated, thereby assuring a satisfactory video quality. Alternatively, packets containing the EL are delivered using an unacknowledged service and they will not be retransmitted if they get corrupted.

Result Analysis: The performance results for HAMM/CP and HARSM demonstrate proper delivery of the BL, ensuring at the least the essential video quality to all MRs. In addition, HAMM/CP is superior to HARSM in reaching out EL's multicast traffic because the transmission data rate adapts better in the two layers. In HAMM/CP, reason for loss in EL's multicast packets is the media conditions, while in HARSM it is because of the growing load on the network and hence the dropping packets are more than the deadline.

3.5.3 Dual Queue Rate-Controlled AP (DRAP)

The issue of QoS provisioning for videoconferencing in hotspots is addressed [17]. Here, a binomial group of schemes has been evaluated to describe a superior solution for this kind of traffic.

Problem: DRAP works above DCF MAC layer, to assure QoS during video conferencing in hotspots with TCP traffic in the background. DRAP can be implemented without hardware modification. DRAP assures soft QoS guarantees, but it exhibits unsteadiness during rate limiting of TCP background traffic and also video traffic.

DRAP is a dual queue mechanism where the video and BE traffic are rate controlled. The RT queue for audio and video traffic is shorter and has priority over BE NRT queue. Thus, audio does not suffer much delay over video and BE traffic. Conversely, DRAP rate control is based on an AIMD or Additive Increase Multiplicative Decrease algorithm which is not appropriate for multimedia due to the multiplicative reduction of the transmission rate. Delay is increased, it also influences video packet loss and it may be possible that BE sources send uncontrolled bursts impacting audio and video.

Solution: One of the binomial rate control mechanisms [17] evaluated are implemented in DRAP to handle the issues stated above. The authors have in earlier literature, presented a category of nonlinear algorithms to manage congestion in Internet transport layer protocols. Some of the algorithms to be used in areas like Internet video and audio don't respond effectively to sudden fall in rates arising from lower user-identified quality. Their attempts are successful by generalization of the familiar algorithms like AIMD, having linear control. Inverse Increase /Additive Decrease or IIAD and SQuare RooT or SQRT are introduced.

Conclusion: Box plots are plotted for original DRAP based on AIMD, SQRT and IIAD mechanisms relating to delays introduced in the audio packet. SQRT binomial algorithm performs in a superior manner in a hotspot set-up for video conferences. SQRT allows an easier method for decrementing or increasing data rates in DRAP, permitting finer control over TCP sources, avoiding their introduction of background traffic burst, thereby offering improved protection for RT traffic.

3.5.4 Zig Zag MiRA

802.11n has channel bonding feature with 40 MHz bandwidth covering two contiguous channels and thus has double stream (DS) and single stream (SS) modes of working.

Learning Automata: (LA) The stochastic learning automata are a reinforcement learning method, which attempts to resolve a problem devoid of any major knowledge concerning the optimal solution. It has two major blocks, one being the stochastic automata consisting of a finite number of actions. This learning method takes a decision of selecting an optimal action from a list of permissible actions (i.e., rates). In every process where a decision on rate has to be taken, an action is chosen from this set, based on its probability vector, and this is implemented on a random network. Next, the chosen action is assessed and the goodput of this action is measured and dispatched to the automata. Automata use this response to make a choice about the next action. Finally, the optimal action is chosen by repeating these steps.

The second component, the algorithm for learning the optimal action, affects the performance of the LA.

Background: Zig zag algorithm (MIRA ALGORITHM), a MIMO based RA algorithm in 802.11n WLANs, alternates between SS and DS modes. It searches upwards and downwards in the current mode until it does not find any further opportunity for enhancing the throughput. After finishing the intra-mode actions, it commences inter-mode operations ,searching and altering the rate to the alternate mode.

Problem: RRAA and SampleRate are not good enough in 802.11n although they perform better in legacy WLANs.

Solution: Learning_RA algorithm [18] is proposed. Each time the algorithm is executed, current rate's goodput is entered in the goodput array, and it calculates the desired_goodput:

$$desired_{goodput} = (goodput \times a) + (goodput[previous\_rate]) \times b$$

Where a = 0.4 and b = 0.6. Next, it checks if obtained goodput of current rate meets desired goodput. If so, then it increases selecting probability of the current rate and for others rates, it is reduced. If not, then it does the opposite.

Next, it assigns a random number to each rate, which is their priority to select. Nevertheless, if there is a rate which has not yet been chosen, it has more priority. Or else it selects the rate which is given a bigger number.

Simulation: Three scenarios are defined - (1) No RA algorithm used. (2) Zigzag: MiRA (3) Proposed Learning_RA algorithm.

MiRA ZIGZAG gives better results over the scenario that doesn't use any RA algorithm. On the other hand, Learning_RA is superior to MiRA ZIGZAG (in better throughput and less delay).

3.6 Scheduling

Thirteen case studies on scheduling methods to assure QoS and guaranteeing fairness in WLAN RT flows are discussed below. Scheduling approaches are classified broadly under eleven headings based on the following layers / mechanisms:

Traffic Scheduling schemes, Polling, Token Passing, MAC layer (Aggregation and TXOP), PHY layer (scheduling using Resource Blocks), Cross platform and Out of band approaches. The last scheduler example meant for 802.11ac WLANs consolidates MU MIMO precoding, MCS and user selection and link adaptation features.

3.6.1 Traffic scheduling schemes

IMM or Interactive Multimedia provides video and audio conferencing in high speed WCN or Wireless Campus Network. Interactive Multimedia transmission is enhanced to address the issue of scheduling. The foremost challenge to schedulers of RT traffic is to grant the assured QoS. These traffic schedulers transmit the priority based traffic flows through different bandwidth divisions in the channel.

The scheduling algorithms for IMM applications [19] are discussed next.

3.6.1.1 Packet based traffic schedulers

a. **PDAS - Path Delay Adaptive Scheduling:** Here, routers are connected in a multicast tree topology. The maximum path delay from each child node to the destination parent node is maintained by the routers and these routers, in turn, exchange this information with other routers in the multicast tree. PDAS assures QoS for RT applications by consulting the delay information in the routers and selecting paths with least delay.

b. **Connection-Aware Temporally Fair Scheduler (CATFS)** algorithm provides equal channel usage times to each flow. Decentralize-CATFS (D-CATFS), an extension packet-based traffic flow algorithm to CATFS, assures fairness for each uplink transmission flow of 802.11a/b WLANs.

3.6.1.2 Priority based packet schedulers

These are used to assure QoS. They are of two types- static scheduling where packets are delivered as per the fixed priority assigned and also on dynamic scheduling. In the latter, where packet priority is used to sort them and this priority is changeable in the course of transmission.

a. **TSBS - Token and Self- Policing Based Scheduler** is used to guarantee QoS in the presence of large traffic loading over 802.11e WLANs. Voice, video and data applications are marked as first, second and third priority respectively. TSPBS works in a distributed manner without a centralized controller by means of passing tokens.

b. **VPQS - Voice Priority Queue Scheduler** splits the traffic flow into VoIP and Non-VoIP traffic flows (VF and NVF). The enqueue classifier receives the input which is divided in VF and NVF flows. Figure 3.1 explains the VPQS

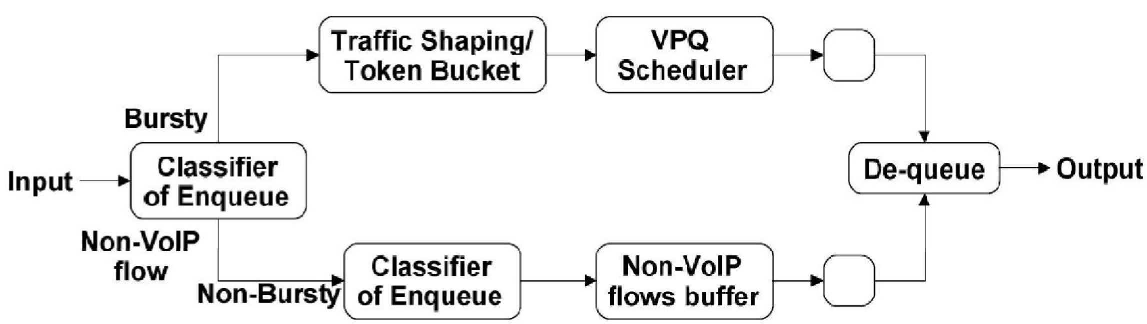

Fig 3.1 VPQ System Model

c. **CBWFQS - Class Based Weighted Fair Queue Scheduler** has classified flows in high speed networks depending on traffic class weight. CBWFQS guarantees small delay for multimedia applications, but results in increased buffer size besides requiring expensive hardware set up for RT applications.

3.6.1.3 Frame based packet schedulers

These consider a frame i.e., combination of multiple packets as a traffic flow.

a. **Deficit Transmission Time Scheduler (DTTS):**is applied at the AP in a FIFO mode. Here, the link quality is not based upon SNR but the channel quality is instead quantified to assess the time for transmitting to each STA. Thus, it is possible for different STAs to obtain perfect isolation between various traffic flows. However, although DTTS accepts TCP and UDP traffic flows, multimedia flows are not accepted.

b. **Fair Hybrid Coordination Function Scheduler (FHCFS):** In FHCFS, the time is allocated to different types of multimedia flows in mobile nodes by estimating queue lengths of the multiple flows.

Advantages: Performance is comparable with IEEE 802.11e HCF scheduler, it provides better fairness to meet delay requirements of different multimedia flows and supports bandwidth for huge networking loads.

Disadvantages: It is unsuitable for fixed BW reservation wireless media.

c. **Dual Queue Based Rate Selecting Scheduler (DQBRSS)** has two queues at the AP, meant for unicast and multicast. The data from Access Point is provided access to the channel depending on the data types, media condition and transmission rates .

d. **Wireless Access using Variable Expansion Scheduler (WAVES)** algorithm provides efficient bandwidth management to achieve QoS for multimedia applications.

Advantages: The scheduler is reliable, achieved top QoS performance and has very low schedule calculation complexity.

Disadvantages: The computational load has very small increment compared to other schedulers over WLANs.

3.6.2 Request reservation approach

Polling and request reservation scheduling schemes are non-applicable for UL.

In 802.11ac, DL MU MIMO enables SDMA, permitting multiple users to receive streams in parallel. In WiMAX and LTE mobile networks, because of TDMA, polling and Resource Reservation Request - Grant scheduling mechanism ensures that QoS guarantees are met for users. It has been arrived at in [20] that both polling and ULR-based mechanisms are unsuitable for RT applications.

3.6.3 Polling

A novel technique called Channel Access Throttling (CAT) enhances QoS [21] by enabling scheduled fine-granularity based QoS policy.

Problem: EDCA offers service differentiation with inherent limitations such as - 'soft' QoS support (not parameterized), coarse service differentiation (only 4 access categories) and inefficient channel access (long CW size). While these problems of EDCA can be overcome by HCCA which is a scheduled access method, HCCA itself is not easily supported by the existing WLAN devices.

Solution: A CAT AP performs scheduled channel access in the upper MAC layer, through CSMA/CA with adaptable CW as with EDCA. Due to the compatibility of CAT with EDCA, it can be readily ported onto EDCA hardware.

While polling (as per a pre-computed schedule) a member STA, the CAT Access point modifies the channel access attributes so as to assign a comparatively greater channel access priority to the “polled” member station.

EDCA’s channel access differentiation is only with different AC. CAT has added two more dimensions - different member STA and different time. As an example, if it is assumed that the member stations can be grouped under only two categories.

1) CAT-high group: Special stations with higher priority.

2) CAT-low group: Normal stations.

The difference in channel access priority between the groups is so large that the CAT-high stations have almost exclusive access to the channel and the CAT-low ones are in backoff condition. The memberships of the two groups is rotated to ensure that all member stations get transmit opportunity as per the QoS stream requirements. The duration for which a station is in the CAT-high group, with good bandwidth and data rate service, and also the time when a station joins this group are well planned by CAT to meet the requirement of delay for the QoS streams.

With proper planning, CAT can allocate bandwidth and duration within the schedule for the stations so that their throughput and delay requirements are satisfied.

Simulation: Simulation of CAT is performed in NS2 using the enhanced 802.11 module 'yans' which simulates a more detailed PHY layer, while EDCA and HCCA are supported by its MAC layer. The module 'yans' is BER based and simulates wireless conditions which closely resemble the actual WLAN receiver with varying transmission rates. Results of the simulation using EDCA and HCCA are studied along with CAT. HCCA used a reference scheduler in yans as per IEEE 802.11e protocol. The results proved that CAT:

- Effectively partitioned wireless channel capacity as per category
- Increased channel usage by improving its access efficiency
- Improves performance of VoIP applications.

3.6.4 Using TXOP

Explicit Traffic Aware scheduling algorithm with Explicit Queue length Notification (ETA-EQN) [22] is designed for IEEE 802.11e HCCA for multimedia services having the characteristics of VBR traffic.

Problem:

1. Although HCCA has several scheduling algorithms for multimedia services, issues still exist with polling and inefficient TXOP allocation, which restrict the HCCA scheduler from practical deployment.

2. Due to the nature of VBR video traffic - variations in bit rates and unpredictability, meeting the preferred QoS specifications is difficult.

Computation of TXOP is a critical activity in HCCA scheduling to achieve efficiency and performance. Choosing a very short TXOP duration increases the delay leading to QoS failure. It is advantageous to choose a large TXOP duration as STAs may redistribute the TXOP which has not been utilized to other flows in the same STA. But, huge value of TXOPs lead to session blocking probability i.e., admission of new streams.

Solution: ETA-EQN extends TXOP limit defined by the standard and avails the EQN or explicit queue length notification frame. It is suggested to add a TXOP Level field . The procedure how the TXOP Level is determined is described now. A STA requiring to request a new traffic stream configures the TXOP enable field to 1 and transmits ADDTS Request frame. The Access Point on receiving this frame determines TXOP of the stream and also computes the value of TXOP level .The Access Point adds the TL value into the TSPEC and sends it to the STA in a ADDTS Response frame. The STA upon receiving this frame verifies the TXOP level from TSPEC. Each time the STA computes TXOP limit from QoS CF-Poll frame, it multiplies the TXOP level :

$$TXOP = TXOP_{LIMIT} \times 32\mu S \times TL$$

Where $TXOP_{LIMIT}$ is value recovered from the QoS CFPoll frame after the STA receives it.

Uplink Scheduling: Each STA transmits a EQN frame to the Access Point which holds information about its own queue length. Next, the Access Point obtains queue length information of each STA from the EQN frames.

QAP uses queue length information for calculating TXOPs. As the EQN frames hold the STA's up-to-date queue lengths , it is possible for the Access Point to accurately compute the TXOP of each STA.

Downlink Scheduling: As the QAP knows the queue lengths of each stream to be sent, it is easy to allocate TXOP.

Conclusion: The performance of ETA-EQN is verified through simulations and the results compared with the reference NS2 scheduler. The results indicate that ETA-EQN performs better compared to the reference scheduler with respect to the throughput, delay and packet loss rate.

3.6.5 Token Passing Mechanism

A scheduling based QoS scheme called Wireless Token Network (WTN) is a centralized polling based procedure [23]. WTN reduces transmission overhead by reducing frame / sub frame size, uses TDM to minimize packet losses that occurs due to collisions and thus improves RT traffic throughput in WLAN. The wireless node transmits, only when it receives a token from the AP - which is responsible for the management functions.

The TDM activities in a WTN cycle are addressing, down-stream and up-stream traffic. The average delay at each node can be at the most the cycle time of a round-robin cyclical network, which for e.g. in voice and video traffic is 150 ms and 200 ms respectively. The data traffic does not have such delay constraint as it has low priority and hence uses the unused available bandwidth.

The cycle period is 128 / 120 ms based on whether addressing takes place or not. A cycle is sequenced as follows:

1. Downstream traffic from the AP is sent continuously until there is either an end of traffic or end of downstream duration. It may be noted that traffic pattern is symmetrical without a traffic holdup at the AP as the AP need not compete like other stations for channel access.

2. Up-stream sequence starts. The token holding data concerning the slot allocated and duration of the allocated slot is passed to each station. QoS is achieved by allocating to the AP and to all the stations, dual queue which permits differentiation between best effort and RT traffic. During the upstream sequence each station updates the modifications in its queue lengths in the transmit data frames, which is later stored in a management list.

3. Addressing is initiated to allow the wireless nodes that are new to the

network to connect with the AP. After upstream transmission is completed, the AP looks for a free address to send an Address Send Frame (ASF). This serves as an indication to the unassociated nodes about the availability of an address which the wireless nodes can associate with. The nodes in turn are expected to apply a random back off slot to avoid collision and then transmit an Address Reply Frame. Address is allotted by the AP to the nodes .

4. When there are no free addresses, the AP does not send ASF and initiates a downstream division.

3.6.6 MAC Aggregation Approach

3.6.6.1 Approach-1

A frame aggregation scheduler is proposed [24] which manages QoS for RT multimedia applications .

The approach used by this scheduler is to unite 802.11e differentiation with 802.11n aggregation. Dynamic adjustment of the aggregated frame size is done based on the frame's Access Category. The procedure is to (1) define the upper and lower limits for the aggregation sizes used with each Access Category's priority, (2) verify frame AC, (3) calculate the complete size of frames comprising the same Access Category in the queue and (4) select the equivalent aggregation method based on the Access Category value.

Aggregation is extremely useful when used with traffics with high speeds. In other cases, waiting for other packets to reach the queue, amplifies the delay highly, particularly for packets that have reached earlier. In the current proposal, packets which delay-insensitive (BK and BE ACs) are made to wait till the other packets arrive. After the aggregation size specific to each Access Category is obtained, frames are sent by means of A-MPDU aggregation. On the contrary, as the tolerable delay should not be violated for delay-sensitive applications (VO and VI Access Categories), when the MAC layer obtains AC2 or AC3 packets, all queued packets with the same Access Category are immediately transferred with / without aggregation.

3.6.6.2 Approach-2

Problem: Aggregation, present in both IEEE 802.11n as well as in IEEE 802.11ac, mixes multiple packets in one transmission. This improves throughput significantly although there is more delay with low rate applications. The delay has two components: (1) delay for packet aggregation is the time taken up for queuing by the first packet which has arrived and is waiting in the buffer and (2) the time needed to convey the aggregated frame - i.e., IFSs and the delay due to headers. Therefore, it is required to arrive at a compromise between increasing throughput and reducing delay.

Solution: Joint urgency delay scheduler and adaptive aggregation mechanism [25] will be employed in the AP. Suppose there are the following data sources: voice, video and streaming . Frames are classified into Access Categories decided by EDCF defined in IEEE 802.11e.

Scheduling metrics defined for this paper include waiting delay(WD) and urgency delay (UD). Packets with least UD are considered first for addition into the aggregated frame. Next, we shall modify the aggregated frame's payload size .

A. Scheduling based on UD

DT = delay threshold within which limit the packet should have reached its destination.

UD = time remaining to serve a packet.

WD = delay while waiting in the queue.

The proposal recommends that packets with the lowest *UD* should be scheduled on priority. Packets = -ve/ NULL *UD* shall be discarded.

B. The Access Point forecasts the size of the aggregated frame knowing the *UD* of accumulated sub-frames. As the first sub-frame had to be transmitted before *UD,* the transmission delay of the aggregated frame must not be larger than this. The AP checks if it can add one more packet based on the space of current aggregated frame. If so, it embeds the packets into the aggregated frame. Or else, they are placed into a new aggregated frame.

Figure 3.2 describes the working. Firstly, Access Point calculates the UD using UD = DT - WD. Then, the packet with the lowest *UD* is sorted and added to the aggregated frame. Then, it follows the approach outlined above and is depicted in the figure 3.2

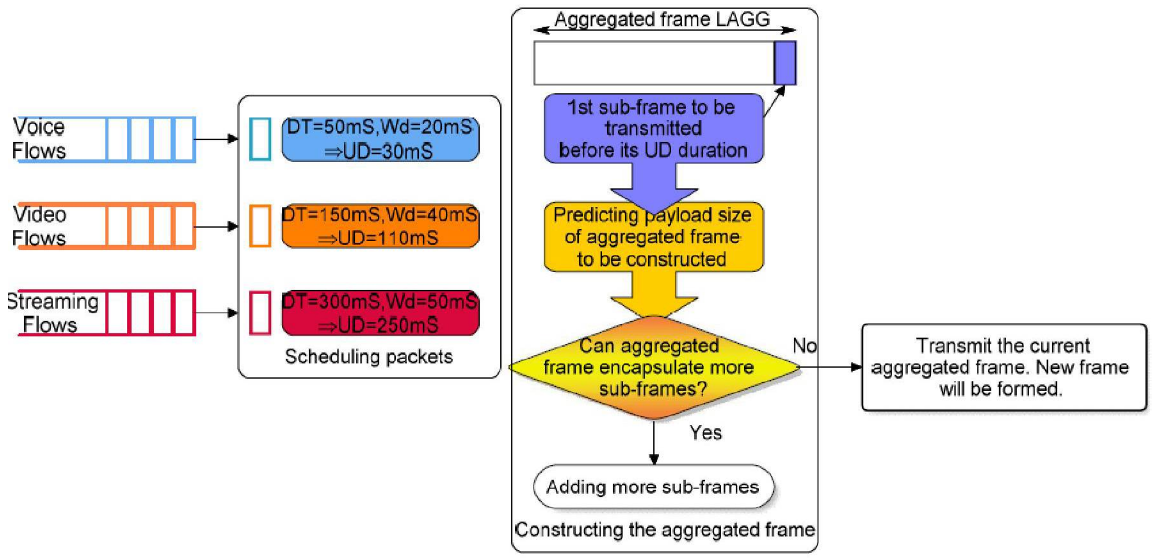

Fig 3.2 Functional Example - Adapting the Aggregated Frame

3.6.7 PHY Layer Scheduling

3.6.7.1 Approach-1

Problem: In DCF all contending stations are offered equal opportunities in a distributed way, and thus it is highly appropriate for BE traffic. Conversely, because of the same contention and unpredictable characteristic, it is inappropriate for transmitting traffic having QoS requirements. As the operational frequency range of WLANs is unlicensed, it is difficult to forecast the other existing sources of signals which can cause interference and hence QoS is not assured. However, demand for transmission with QoS guarantee is constantly rising.

Issue with DCF is that it inefficiently utilizes radio resources. Here, once a STA gains a TXOP, the STA occupies the radio resource which becomes unavailable to other STAs until released. This implies that the user is not efficiently utilizing the severely attenuated part. Hence, it may be a better idea to release this part for other users on a temporary basis and thus improve spectrum usage.

Solution: A transmission scheme with multiple-users using OFDMA is introduced [26]. An AP controls all STAs in the BSS. Radio resource are reserved for some time, using DCF. Fine-grained RB or resource blocks are utilized to make access control flexible. Resource block is the radio resource's minimum allocation unit in a 2D plane characterized by time and frequency domains.

AP has K back off counters, and none of the stations has a back off counter. Once any of the K back off counter's value becomes zero, Access Point sends in broadcast mode, a RTS frame. The individual STAs responds with CTS as per the sequence in RTS. The CTS from each STA holds its CSI and QoS required. Using the CTS information, the AP manages to reallocate the RB for DL transmission, and conveys this by sending DL-ARBI or Down-Link Assigned Resource Block Information frame. Refer Figure 3.3.

Next, the Access Point sends DL data frames to many STAs in the resource blocks allocated. Individual STA responds with a 'DL ACK' to indicate whether the DL-DATA frames are successfully transmitted or not. The STAs also renews their requirement for QoS and CSI using Down-Link Acknowledge (DL-ACK) frame. This information is utilized by Access Point to allocate resource for UL transmission, and notified in the UL-ARBI or UL Assigned Resource Block Information. The STAs now send UL-DATA frames to Access Point using the RB allocated to them in UL-ARBI. AP now transmits in broadcast mode, UL-ACK frame and then releases the allocated resource.

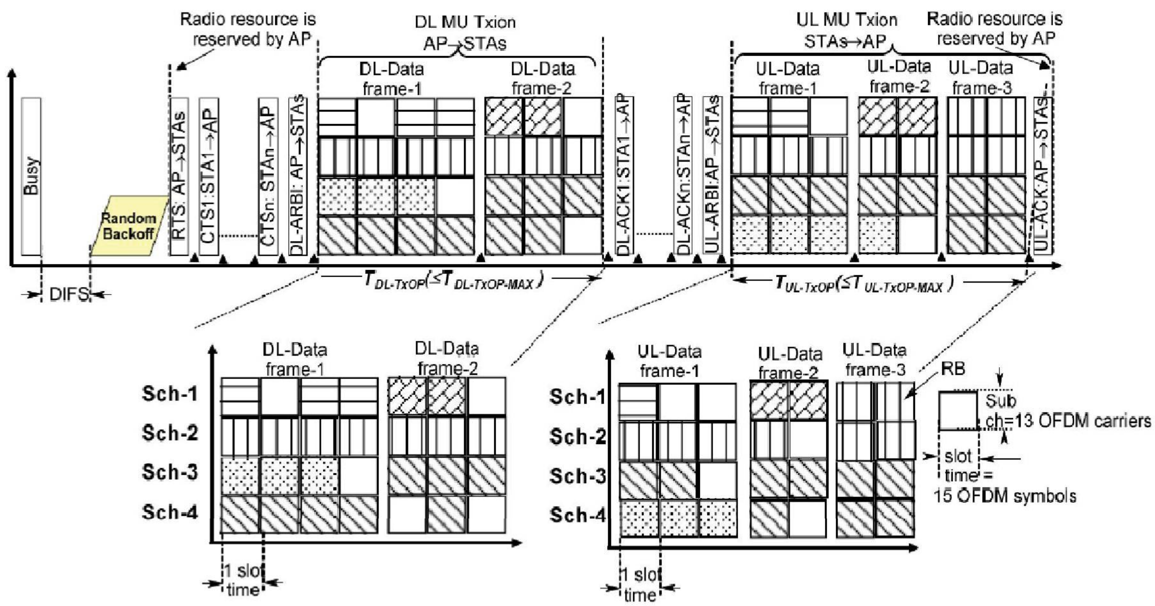

Fig 3.3 Resource Allocation using DL-ARBI

Resource Allocation Algorithm

A heuristic resource allocation scheme is proposed and implemented for WLANs comprising real-time (RT) and non-real-time (NRT) traffic. The aim is to enhance the spectrum efficiency and simultaneously meeting RT traffic's QoS requirements. A STA having RT traffic and one with NRT traffic are referred to as “RT-STA” and “NRT-STA”. These procedures are depicted in figure 3.4.

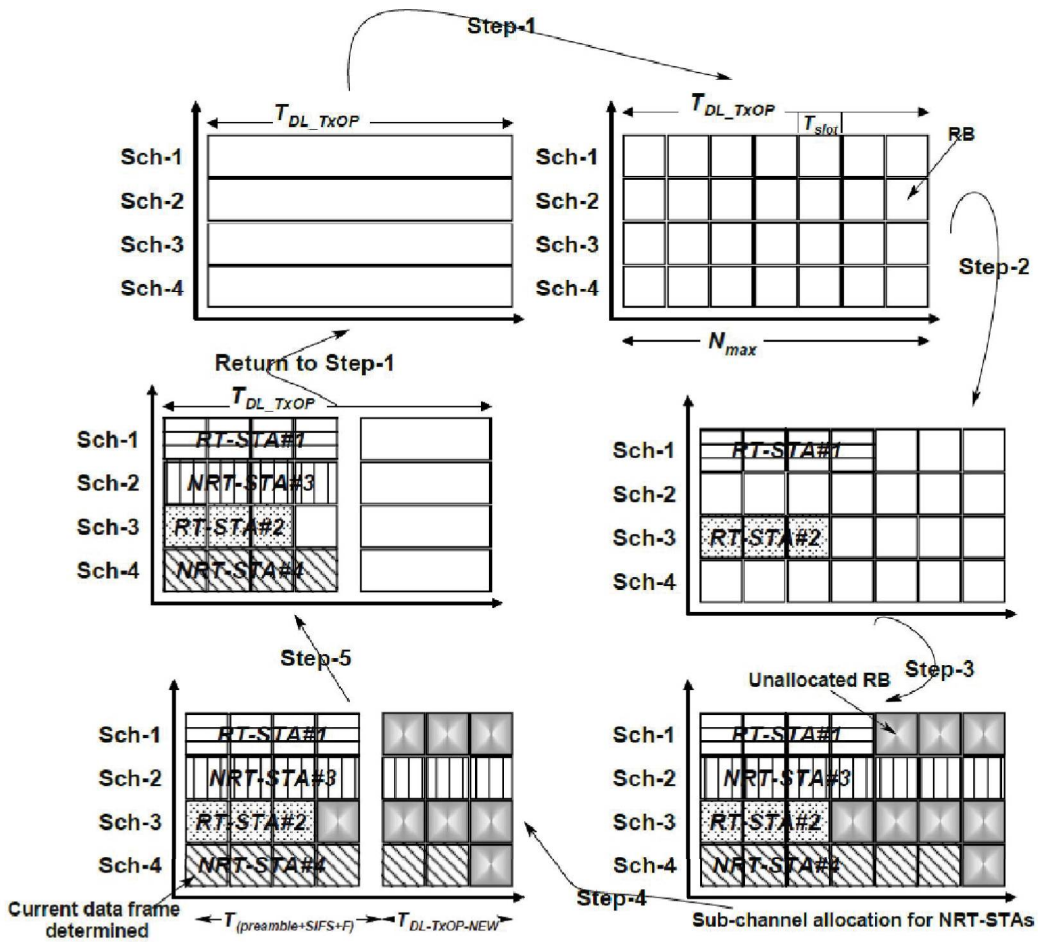

Fig 3.4 A Heuristic Resource Allocation Algorithm

STEP 1: RB generation.

STEP 2: Allocation of sub-channels for individual RT-STAs. Data from RT-STA is time-dependent and satisfies time delay criteria normally. Hence, RT-STAs are given priority during resource allocation. Using the EDF or Earlier Deadline First algorithm, the Access Point chooses the RT-STA which has a lesser TTL or time-to-live remaining and assigns to this RT STA a sub-channel having largest SNR.

STEP 3: Sub-channels not allocated in step-2, are allocated to NRT-STAs in step-3. Here also the sub-channels having largest SNR are individually assigned to the NRT-STA in order to optimize utilization.

In the WLAN comprising RT and NRT-STAs, meeting QoS requirements is tough when NRT-STAs having high load are present.

$$Qos\ satisfaction\ Ratio = \frac{Number\ of\ RT\ packets\ sent\ successfully}{Number\ of\ RT\ packets\ generated\ at\ RT - STA}$$

The results show that this ratio drastically degrades for the 802.11 DCF as the load is raised. However, this algorithm has good probability of guaranteeing QoS requirement in WLANs in addition to increasing the throughput.

3.6.7.2 Approach-2

A resource allocation algorithm termed QoS-aware dynamic resource allocation algorithm (QDRA) is recommended [27] for the DL transmission in an OFDMA and WLAN integrated system.

A channel comprising numerous sub-carriers within an allocation period is known as a sub-channel. Allocation time slot is the time period for the resource allocation process. While the process of allocation takes place, Access Point assigns sub-channels for delivering frames to STAs and verifies the amount of transmit power permitted. The decision on allocation is done when each time slot starts. It is assumed that the AP has information on the CSI, available spectrum and QoS request of the user .

The aim of the optimization is to achieve maximum data rates and also that both real time and non-real time users have their QoS requirements satisfied. Users are not allocated channel resource based on the highest data rates but according to their needs. The steps involved are :

1) QoS requirements convert

All the QoS requirements (BER, time delay and data rate) of each user are converted into various rate requests at the start of the allocation period.

2) Priority determination

Received rate requests are sorted in descending order for users. A user having the highest rate request gets the topmost priority while spectrum is being allocated. For RT user, highest rate request stands for packets which will be sent at the shortest time.

3) Sub-channel assignment

When sub-channels are available, they are assigned to RT users. For NRT users, as long as there are available sub-channels, they are assigned. When the RT user has to wait for a long while, the user can select sub-channels before the NRT users.

4) All the above steps will be repeated during every allocation period.

To offer fairness, the RT users use sub-channels which just meet the requirements, whereas other resources are given to NRT users having higher data rates.

Simulation Results: QDRA ,on comparing with three other resource allocation algorithms , not only meets QoS assurances but supports large data rates and provides fairness , as well.

3.6.8 Out of Band Approach

The TCP traffic faces unfairness in WLANs; many such issues dealt in [22] will be discussed below. A solution called ADvancedWi-fi Internet Service EnhanceR (ADWISER) is proposed to resolve these issues [28]. This module is located to ensure that all the packets that go through the Access Point go through this device. The working of ADWISER is explained (Figures 3.5 and 3.6).

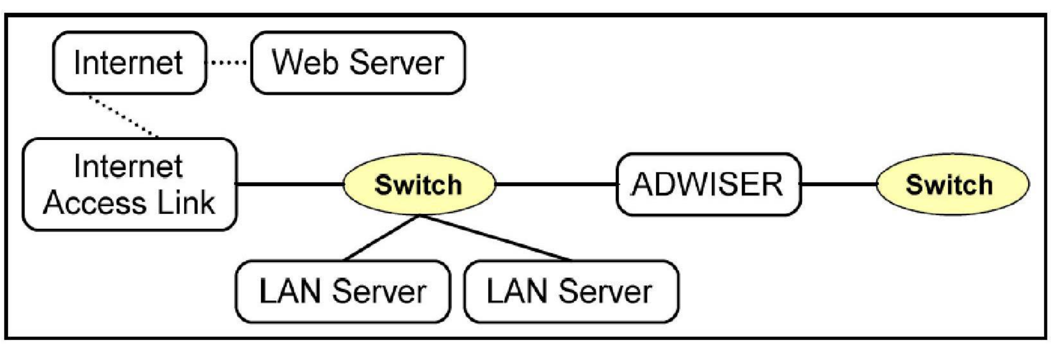

Fig 3.5 ADWISER

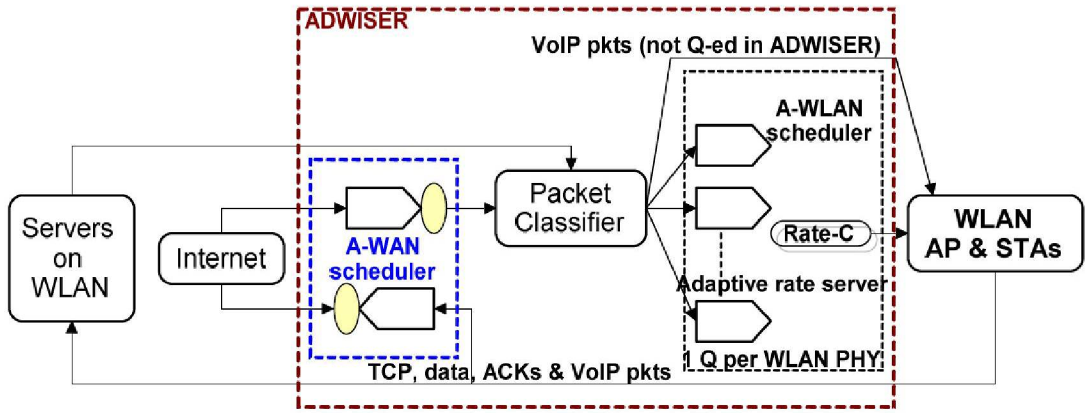

Fig 3.6 Data Flow in ADWISER

There is one virtual server - ADWISER WLAN (A-WLAN) scheduler. Two virtual servers relate to the inbound and outbound directions of the Internet WAN access link - ADWISER WAN (A-WAN) scheduler. Thus, queuing occurs in ADWISER (instead of at the Access Point, for the WLAN channel) and the access routers (in the Internet link). Coordination between these two schedulers is key for handling the unfairness issues.

For TCP-controlled traffic, all packets flowing toward the WLAN (TCP DATA /ACKs incoming from Internet servers or on the wired LAN) are classified and added to queues behind the A-WLAN server, which manages the queues using a fair queuing algorithm.

Control of TCP download traffic (from the servers to the STAs) and of TCP upload traffic (from the STAs to the servers) is realized by appropriately serving the TCP DATA packets and the TCP ACKs.

VoIP packets flowing in the direction of the WLAN are positioned in a strict priority queue in the A-WLAN server and freely flow into the Access Point. Every packet directed to the servers on the wired LAN pass through ADWISER without queuing. All packets arriving into ADWISER from the WAN access link /WLAN are queued in the A-WAN scheduler at the virtual server corresponding to the inbound / outbound direction of the Internet link.

Unfairness issues and their resolution using ADWISER are discussed in [22].

3.6.9 Cross platform Approach

A multi-traffic scheduler with genetic algorithm (GA) is proposed, which seeks to improve QoS, taking into account both traffic and status of the channel [29].

The APP layer of AP generates different traffic types which is categorized in MAC and buffered in memory queues in FIFO manner. When the packet wait time in the queue gets larger than the delay limit of the packet, it will be dropped. When scheduling begins, AP initially detects queue length for every type of traffic. It then offers an opportunity for scheduling the queues that are not empty. If a DL channel has K time slots in every frame for transmitting, they are fragmented for the present M traffic. The K time slots are equally divide for the M traffic.

Next the AP obtains from STA, feedback of CSI and CQI i.e. channel state and quality information. Prior to scheduling a frame, the scheduler estimates traffic QoS using queue and channel information, traffic QoS requirements and priority. Considering delay satisfaction, transmission rate of traffic can be obtained from CQI. The problem is of selecting the optimal traffic slot in the frame (MAC layer) where delay satisfaction, packet loss and data rate are considered. GA is used to reduce the search time for obtaining the optimal solution.

The simulation results are compared with other existing similar algorithms - PF, M-LDWF, Random and it is shown that GA can provide better quality with respect to delay, data rate and packet loss.

3.6.10 QoE based scheduling

Providing QoE during streaming of video in WLANs is mandatory for good reception at user end. Normally, the scheduler fixes the priority in scheduling of the queued packets before transmission. **Earliest Deadline First (EDF)** algorithm schedules the packets by giving highest priority to the packets closest to deadline. The **gradient-based scheduling algorithm (U'R)** gives priority to applications that are sensitive to delay. The **Modified Largest Weighted Delay First (M-LWDF)**mechanism which is similar to U'R has quadratic utility function and is sensitive to both channel and delay. It is also capable of improving throughput. All the above schemes support throughput but not QoE. Hence, they are not the best choice for video streaming.

QoEAS or QoE-aware scheduling mechanism [30] derived from U'R is a less complex and more accurate and hence can be implemented. This framework is shown in Figure 3.7.A factor derived from the following is used for scheduling:

1. **Packet importance index** derived from the slope of video QoE function, and it indicates the importance of the packet in improving quality of video. It is derived at the APP layer and inserted into the IP header's ToS field
2. **Packet delay** is a utility parameter defined for streaming of video which is delay-sensitive.
3. **Current status of channel** is done using a link model

At the MAC layer, the above three factors are all taken into consideration by the scheduler and then a packet is selected from one of the queues followed by transmission.

The proposed QoEAS scheme is validated by QualNet which simulates streaming of video streaming over dense networks with 11n/ac protocols in a lecture hall wherein videos are recorded and streamed simultaneously from an AP to students. QoEAS is compared with MLWDF and round-robin (RR) schemes. QoEAS considerably improves quality of the video streamed while ensuring fairness amongst the users. It has almost 5 dB better Peak SNR ever for the 10 percentile of users from the bottom.

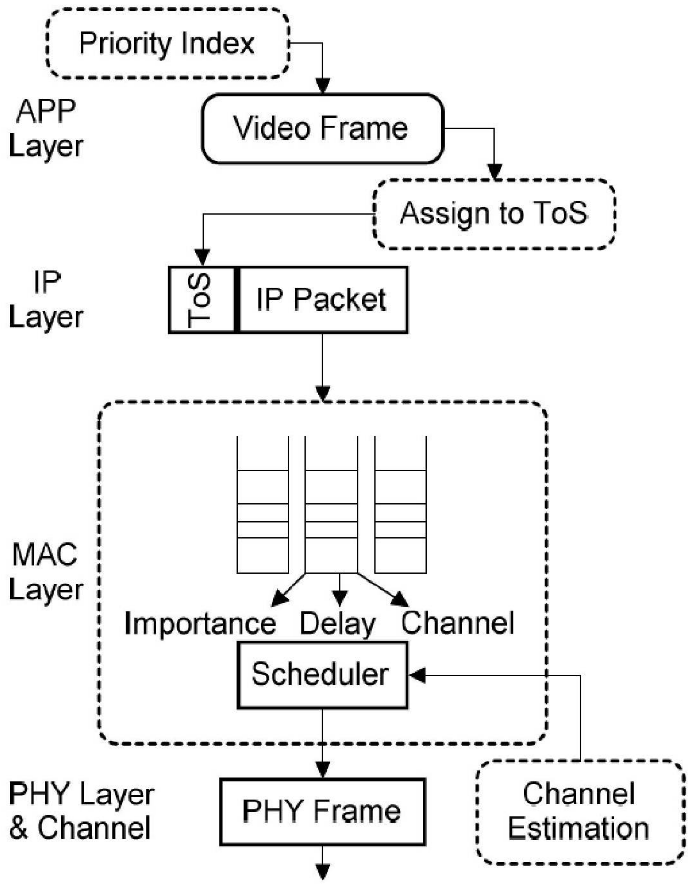

Fig 3.7 Proposed QoE-Aware Scheduling Framework

3.6.11 Scheduler for 802.11ac

By far, the most complete Scheduler for 802.11ac is explained in [31]. Here, a framework for link adaptation in multiuser MIMO-OFDM networks with limited feedback information is discussed. The link adaptation problem is formulated as a maximization of the sum rate subject to a FER constraint.

The steps involved in the approach are as follows:

I. CSI Acquisition

This mechanism is described in the 802.11ac standard (briefly explained below).

1) Sounding: Here, the AP sends a VHT NDP announcement control frame (NDPA), which consists of the set of users going to be polled. The frame, in addition, also includes information on the requested feedback (single user or multiuser—SU/MU). The AP next sends an NDP.

2) Quantization and Feedback: The CSI obtained by each user is first quantized. The main pertinent parameters for the MU-MIMO operation are the preferred beamforming matrices which are represented using Givens decomposition.

The first user in the NDPA list estimates the channel and transmits quantized feedback information , and the remaining users (if any) transmit their CSI by replying to future beamforming (BF) report polls.

Based on the limited feedback obtained, the AP performs link adaptation.

II. Link Adaptation

PRECODING AND EQUALIZATION WITH INTERFERENCE ESTIMATION

A. Block Diagonalization (BD) Precoding

The precoders are based on BD. This precoding removes the interference between the different users but not the interference between streams associated with the same user.

B. Quantization

The presence of limited feedback caused by the quantization procedure generates unknown interference leakage (IL) among the multiple users. As the post-processing SNR is dependent on IL, it is required to estimate this parameter. It is preferable to do this at the transmitter, without the receiver sending feedback .

A closed form approximation of the inter-user IL is derived from BD precoding with zero forcing (ZF) receivers. This is simply calculated at the transmitter by means of statistical characterization of the quantization error and is found to be very accurate for different feedback rates.

MCS SELECTION

Features derived from the channel are classified into the highest MCS that meets the target FER constraint. A supervised machine learning approach is used which includes the following tasks: Feature Extraction and Classification .

User and Mode Selection is done using a greedy algorithm which makes use of information on the SNRs, the feedback rate and the number of users.

3.7 Additional Mechanisms to Handle QoS

3.7.1 Synchronized Time

Synchronized time facilitates exact delay measurements in both directions. It is now possible to get this easily as NTP is deployed in a large scale and precise time sources such as GPS receivers are available. As per ITU-T G.114 recommendation, delay in one-way is maximum 150milliseconds.

NTP is deployed at each end and RTCP sender reports link the RTP time stamps to absolute time. RTCP packets also help senders in determining the RTD time. In a round trip, if the environment is time synchronized, then the sender will be able to estimate the received packet delay from the RTCP sender and receiver reports. Hence, with the information of RTD also, it is possible to estimate delays for the to and fro trips. When many VoIP calls are routed through the same AP at the same time, one-way delays normally vary to a great extent. For instance, with calls over long distances with Wired Network Delay (WND) >100ms,the total one-way delay may go beyond150 mS as recommended by ITU-T G.114.

Problem: 802.11e supports to a certain extent functions like VOIP where QoS is critical. Still, it is prone to congestion and packet losses, resulting in delays. All VoIP sessions have dissimilar one-way Mouth-to-Ear (M2E) delays which may also vary. In spite of this, they compete with each other using the equal prioritization AC, as per the 802.11e standard.

Solution: It is proposed in [32] that if delay in one leg, of VoIP session data can be determined for both directions, we can set EDCA parameters in a different way between two VoIP sessions. Basically, priority assignment is done to voice sessions in the VO AC. In sessions where R-factor is high and one leg delay is small, packets waiting time at MAC layer may be more, subject to the fact that they are conveyed within a period that will not reduce their QoS remarkably. Conversely, a VoIP session having long network delay will have a reduced MAC layer contention delay due to prioritization. Hence the voice session will have a reduced one-way delay.

Conclusion: Compared to the existing 802.11e EDCA mechanism, an improvement has been shown and one can discriminate between VoIP sessions.

3.7.2 Receiving-Opportunity Control

Problem: Although IEEE 802.11e EDCA provides priority to data which requires a guaranteed QoS, the performance of EDCA's high-priority stations may be reduced when the EDCA and DCF STAs are linked to an Access Point, because priority of every DCF STA is greater compared to the Best Effort and Background types in EDCA.

Solution: An innovative mechanism guarantees the required QoS in such a mixed environment using Receiving-Opportunity Control (ROC) [33] in the MAC Frame. ROCS works by reducing traffic with lower priority by imposing a lengthier back-off time and permitting traffic with higher priority and hence enhance their transmission rate. To accomplish this, ROC utilizes ACK and back-off mechanisms in CSMA/CA procedure of 802.11 DCF. Say, in ROC, an Access Point deliberately refrains from returning ACK frames for traffic with low priority. Now, sender has to hold for the lengthier time for back- . Hence flows that have higher priority achieve guaranteed throughput. ROC can be applied in stations without any modifications.

Receiving Refusal Probability (RRP) is calculated periodically and also immediately when a new flow joins. When the Access Point is unable to obtain a frame from a STA during a specified period of time, it understands that the STA has completed sending data frames. The Access Point straight away computes the

RRP when STA number declines. Hence, the suitable RRP (based on alterations in STA number) can be determined.

Conclusion: As per simulation, when STAs number increases, EDCA without ROC does not necessarily result in desired throughput whereas for a small number of STAs, performance is satisfactory. Conversely, the scheme suggested by the author, assures required throughput for any number of STAs.

3.7.3 Cognitive Solution

Problem: There is a great demand for the wireless medium. Hence, there is a need to utilize the available bandwidth in a useful manner without compromising QoS. The demand for the frequency spectra has also resulted in congestion and interference within networks in unlicensed bands in which WLANs operate. A cognitive solution is proposed [34] using a 'central decision engine'.

With the aim of guaranteeing a certain level of QoS, it is essential to have knowledge of the effect of interference level at the receiver on QoS. An experimental setup is created to establish two links. While the first link operates on channel 6, the second one is configured to act as interfering link. The WLAN radio interfaces are configured to send data at the 'Tx rate' bit rate.

Two scenarios for interference are studied.

The first, attempts to assess the effect on QoS of interfering transmitter's power received at LUT. Interference was produced by varying the power from 0 to 20 dBm, and the level of interference power received at the node / LUT observed. Also the impact on the throughput, due to channel overlap at LUT was assessed.

The second study is about Channel Occupancy Degree (COD).

$$COD = \frac{Data\ Generation\ Rate}{Tx\ Rate}$$

The impact of COD on LUT throughput is assessed. The buffer for transmission is packed with data at 'data generation rate', which is then transmitted by the interface through the transmit buffer at 'Tx rate' rate which is more than data generation rate.

Both the above scenarios execute some experiments, each consisting of two phases: QoS of the LUT recorded (1) without interference, and (2)with interference generated. During the experiment, we first parse the output stream holding the receiver's periodic throughput reports and finally store the results in the controller's server database (DB).

Simulation: It was noticed that the throughput is not much affected by higher interference powers. And its impact is further less when channel separation is more than or equal to 3 channels. The throughput is impacted more for higher CODs.

Application: Introduction into a cognitive networking system. The above experiment is used to determine interference and the results obtained will help in estimating the connection link's performance and optimize the network / throughput by reconfiguring the network. As explained it the two experiments, the system architecture comprises of (1) a database, (2) an environment map, (3) a network and (4) the decision engine.

While the traffic is created, the network data (e.g., Receiver Signal Strength Indicator (RSSI), delay,…) is examined and stored in a DB. Using the data in the DB, an environmental map is constructed, envisaging QoS parameters like throughput or interference metrics like COD, RSSI. The QoS and interference metrics are measured at the node locations marked on the map while at other locations they are interpolated. As the map gets updated dynamically, it is possible to rapidly evaluate the channel occupancy. Making use of the results, it is possible to prepare models to predict the changed QoS as and when the network is reconfigured.

3.7.4 FairWLAN

Consider a situation in which one user makes extensive use of the shared medium while another uses it occasionally and hence is at a disadvantage. The available capacity is shared equally between the connected STAs by the APs, thereby making a feeble attempt to overcome the issue. But, as information of earlier interactions is not recorded, this method may not be totally fruitful.

Solution: The concept of FairWLAN [35] for centralized QoS enhancement is discussed. Details of performance and traffic history for all stations associated with an AP are obtained from the AP by the centralized server. This knowledge is utilized to decide the QoS configurations in upstream router. These actions are performed considering all devices that belong to a single user, instead of handling each device separately.

The three main components in FairWLAN are:

1) Collector on a regular basis queries (using SNMP) each Access Point, collecting information for each of the associated stations - PHY bit rate, transmitted and received bytes, MSDU retries, IP and MAC addresses.

2) Recorder processes the data gathered and stores the usage history for each user. The association between a user and his connected devices is extracted from the RADIUS server with which the STA authenticates.

3) Controller uses the data stored for configuring the QoS parameters on the router. It assigns each STA, once every five seconds, to one of the predefined traffic classes. The amount of bandwidth provided to these STAs is decided considering the user's traffic history, the PHY bit rate of the station and the MSDU retries.

FairWLAN distributes the available BW to all the STAs on each Access Point. STAs which did not send data during the previous interval are also given the BW they are entitled to.

3.7.5 MAC-HCCA

[36] designs a MAC protocol, M-HCCA or MAC HCCA which exploits multi-beam smart antennas at AP to increase WLAN capacity and to support QoS.

The Access Point's antenna system comprises of M sectors, each having its own transceiver. Hence several narrow-beam antennas in individual sectors can be treated as a logical antenna. Switched multi-beam antennas are used.

In M-HCCA, a STA having RT traffic can only after (re)association link to the polling list (PL). This mechanism makes CF-Pollable and CF-Poll Request subfields inoperative. The structure of M-HCCA has a CFP (for active M-HCCA) and CP (when DCF is active), jointly referred to as a super-frame. During CFP, the Access Point usually functions in the multi-beam antenna mode. At the beginning of each CFP, the TBTT, each STA shall wake up and starts listening for the polling list (PL) frame. Meanwhile, the medium is constantly checked by the Access Point which then takes control by broadcasting post-PIFS idle time, the beacon. In M-HCCA, the three parts of CFP are created to attach a priority, for collision resolution and for polling. Prioritization and collision resolution periods are jointly considered as 'reservation period'. During 'prioritization', the Access Point exchanges a sequence of handshake signals to ascertain that STAs with higher-priority are always included in the PL before the STAs with low-priority. As part of 'collision resolution', the Access Point implements the deterministic tree-splitting mechanism to find in the time for prioritization, the STAs which wish for admission to the PL.

After reservation process is completed, the Access Point broadcasts the PL frames to declare the beginning of period for polling. On evaluating PL frame, an STA which does not come under receiver / transmitter category in polling period can reassume 'doze' state. On completion of polling period, the Access Point transmits in broadcast mode, CF-End frames. This allows every STA to go into the CP, when the Access Point runs DCF while operating in omnidirectional-antenna mode. Hence 802.11- STAs which have not implemented M-HCCA may continue communication with the Access Point in the CP.

3.7.6 Enhanced AEDCA

Adaptive EDCA (AEDCA) is an algorithm that adaptively controls the CW size based on rate of collision of the data frames sent. Here, every station calculates its own rate of frame collision using number of successful /unsuccessful data frames sent. When the collision rate exceeds a predefined limit, AEDCA sets a bigger CW size as compared to that for EDCA. This lessens the collision rate. Hence, in comparison with EDCA, AEDCA provides an enhanced capability to discriminate QoS provisioning amongst Access Categories, even while preserving total efficiency of bandwidth.

In [37], a QoS control mechanism is proposed where throughput of STAs with assigned priority is measured and IFS length of stations is altered to match the circumstance and assure QoS. Here, the AP observes the data of every station and insists that stations with lesser priority should rate at which frames are transmitted if AP detects traffic overcrowding. The access point has an observer that estimates the traffic of every station. On sensing a congestion, the Access Point conveys this information to the stations through introducing the congestion information known as back pressure flag into the ACK frame and sending it.

To identify congestion on the media accurately, the Access Point uses leaky bucket algorithm.

3.8 Review of Transmit Beamforming Mechanisms

Here, the various approaches which are mentioned in literature, with respect to Transmit Beamforming are discussed.

3.8.1 Time Domain Quantization

In [38], a technique "time domain quantization" (TD-Q) is introduced wherein there is a feedback of time domain parameters. These are required, as in Sounding, to form the beamforming matrix at the transmitting end. It is proved that TD-Q reaches the same sum rate capacity of the conventional Givens rotation quantization GR-Q and requires less amount of feedback.

3.8.2 Pre-coding and Temporal CSI

Reference [39] proposes a method of transmitting accurate channel estimation to the transmitter which requires this to enable the calculation of precoding vectors. Since there is a huge amount of channel coefficients that is required for feedback, particularly when there are more OFDM subcarriers and antennas, limited CSI feedback works out to be an advantage. Thus , with [39] ,both accuracy of the feedback is enhanced as well as the feedback amount is decreased.

3.8.3 Explicit Feedback and Preamble Structure

The authors in [40] propose methods to reduce overhead for explicit feedback based on Givens decomposition. It is shown that the smoothing gain due to channel estimation can be more than the gain due to beamforming, specifically for low values of SNR and in channels due to low delay spread. Therefore it is proposed to utilize a preamble that permits channel smoothing to be done in transmission with beamforming. Also non-precoded channel estimation for channels with low delay spread is proposed.

3.8.4 DTRA - Directional Transmit and Receive Algorithm

Reference [41] proposes a DTRA algorithm a TDMA-based MAC algorithm. This utilizes smart antenna's beamforming capabilities of to adapt resource access based on the requirements of each traffic flow. At the same time, likelihood of detection, interference and network jamming are also considered. In DTRA slots are reserved based on the traffic load. It is demonstrated from simulation results that DTRA performance is superior to IEEE 802.11.

3.8.5 Beamforming with Multiple Spatial Streams

Reference [42] evaluates the theoretical throughput required and coverage expected for a domestic WLAN. Results are provided for 802.11n MIMO deployments. Beamforming is mainly influenced by regulatory constraints in transmit power. For long-range links, throughput of beamformed data decreases greater than half. With regulatory constraints imposed on beamforming in

WLAN, the average data rate decreases by 14.3% whereas for a system with antenna selection, it results in an 8.9% reduction. While antenna selection uses single stream MIMO more commonly as compared to beamforming, beamforming is superior in performance with multiple spatial streams. It is demonstrated that Beamforming is most effective in the medium-range channels.

3.8.6 Single-User and Multi-User Beamforming

In [43], the achievable rate formulas of STBC, SU-/MU-BF modes of transmission in 802.11ac with time-varying channels, are derived. The parameters - feedback delay and operating SNR are estimated by the mode of transmission, rate of fading and frame length.

While STBC feature is built to operate in open-loop mode, TBF operates in closed-loop mode with AP supporting one STA in the SU-BF mode and many STAs simultaneously in MU-BF mode.

If a station cannot perform closed-loop beamforming, data transmission from AP should be with STBC. With SU / MU-BF, variation of channel parameters with time leads to degradation. Sometimes, STBC can yield higher throughput compared to BF even in the absence of feedback in the channel. For example, when there is not much data to be delivered, feedback turns out to be an overhead. In such cases, open-loop BF is a better choice compared to the close-loop BF. Rate loss due to variation in channel parameters is derived for throughput calculation, by considering the structure of the frame. A guideline is provided for system design of 802.11ac following these investigations.

3.8.7 Smart Antennas

Prior to the use of smart antenna techniques in 802.11n, a study was undertaken in [44] to study the benefits of these techniques for general wireless networks. [44] Discusses the circuit techniques involved in combining and splitting signals to interface with an array of elements in receive and transmit mode respectively. Also the methods employed to generated equi-phase and equal amplitude signals from/to all the array elements and thus achieve superior data rates and cancel interference are explored.

3.9 Review of MU-MIMO Mechanisms

Here, the various approaches which are mentioned in literature, with respect to MU MIMO are discussed. They are classified under General, MAC and PHY headings.

3.9.1 General

3.9.1.1 Group ID

Reference [45] focuses on utilizing the concept of group membership for DL MU-MIMO. The AP uses the Group ID to indicate to STAs which are group members, about their location and also the SS on which DL transmission will take place.

In [45], depending on the Group IDs available, the authors are concerned with allocating membership to the group and STAs positions. An empirical algorithm implements the activities listed below.

- Associated STAs are allotted alphabet letters and transmission sets are arranged alphabetically.
- The Access Point calculates the conditional probability (P) of a STA (if it is a member of the given transmission set) being in position x of the transmission set.
- Every STA is given position x in a number of Group IDs based on the value P.
- From position x, the Access Point is eligible to choose the corresponding Group ID.

3.9.1.2 TxOP for DL MU-MIMO

In [46], TXOP sharing to enhance TXOP for DL MUMIMO transmission is discussed in addition to a concept regarding revised backoff procedures for secondary Access Categories.

Conclusions: From the simulation results it is seen that the suggested backoff procedure is superior in performance, regarding fairness, than the traditional backoff mechanisms. The TXOP Sharing mechanism introduced in this paper is accepted in the 802.11ac specification.

3.9.2 MAC Layer

3.9.2.1 MAC Mechanism

a) Backoff Mechanism: Reference [47] reports the work done to improve EDCA backoff schemes, its drawbacks and approaches proposed to improve the limitations of the backoff schemes and simulated with OPNET. It may be noted that this suggestion has been implemented in the 802.11ac (draft).

Simulation results indicate that these approaches (1) decrease back offs (2) increase voice traffic channel time without overly reducing video traffic channel time.

b) Network Allocation Vector - When there are more than one BSS, network allocation vector (NAV) may be set erroneously because of MU-MIMO and overlapping BSS. The throughput may decrease considerably due to the useless or redundant NAV settings. To resolve this issue, in [48], a two-level NAV mechanism is proposed with minor changes to the standard. It is demonstrated that the proposed mechanism achieves better throughput when evaluated against the usual NAV method.

3.9.2.2 MAC TXOP

a) Modified Backoff Procedure: In [49], it is proposed to enhance the TXOP Sharing mechanism, to obtain improved DL-MU-MIMO transmission. Next, a modified backoff procedure is designed for the primary AC. Simulations show the advantages of the scheme with respect to throughput and channel utilization.

b) Markov Chain Model-1: In [50], an analytical Markov chain based model is introduced to assess performance of 802.11ac Access Point. While this model is an extension to the popular Bianchi's Model, its chief advantage is in the integration of the TXOP sharing mode to the 802.11ac. Using the Markov chain output i.e., transmission probability of a given Access Category, a mathematical model is derived to estimate the achievable throughput of a given Access Category. Using this model, it is possible to examine how TXOP sharing could enhance the inadequate wireless bandwidth utilization while obtaining fairness among the multiple ACs in accessing the channel. Note that this paper only analyses the standard for TXOP and no enhancement is provided.

3.9.2.3 Aggregation

a) **Comparison of MU-MIMO and Frame Aggregation Multiplexing Schemes:** [52] compares the performance of two DL user multiplexing schemes: MU-MIMO and frame aggregation in IEEE 802.11ac. If each user's encoded data stream has a similar length, the MU_MIMO achieves better throughput compared to aggregation. Conversely, if different lengths are present, the reverse is true. In a fast-varying channel, because of the overhead of channel feedback, throughput of MU-MIMO is lesser than frame aggregation.

Problem: The MU-MIMO combined with aggregation can improve performance considerably. If the frames in different SS have differing transmission times, the space channel time occurs .This is a period wherein data is carried by certain SS whereas the others do not have data. Due to space channel time, DL MU-MIMO channels' transmission efficiency degrades.

Recent WLANs implement multi-rate transmissions using Adaptive Modulation and Coding (AMC) to realize transmission channel adaptation. The paper, in addition to addressing the space channel time also considers AMC and VoIP transmission in aggregation mechanisms.

b) **Aggregation for VoIP**: In the proposed scheme in [53], the Access Points uniform the transmission times of frames to the optimum extent on all the SS to the different Mobile Terminals considering their MCS level. Consequently, the scheme can reduce the space channel time for the VoIP packets transmission within the allowable delay limitation.

Conclusion: The simulation results demonstrate that the new scheme augments the throughput, the space channel time ratio and maximum delay for the VoIP packets with multi-rate transmissions.

c) **Fragmented MPDU-Problem**: In EDCA TXOP sharing mode, A-MPDU pads must be added to the end of the A-MPDU for each user in order to fit within A-MPDU boundaries. These pads correspond to non-meaningful content that waste medium resources.

Solution: In [54], A-MPDU with Compressed Block ACK mechanism is modified to be able to contain a fragmented MPDU that is not even a VHT single MPDU. In the place of A-MPDU pads, fragmented MPDU may be included to fill the length of A-MPDU boundary in EDCA TXOP sharing mode. The VHT Compressed Block ACK mechanism is also re-designed to acknowledge the fragmented MPDUs in the A-MPDU frame.

Conclusion: With the proposed scheme, more frames can be sent within the A-MPDU boundaries, resulting in enhanced throughput.

d) **Combination of Aggregation and MU-MIMO**: [55] proposes a combination of packet aggregation and MU-MIMO transmission to improve performance of the system. The technique adopted is (a) RTS/CTS handshake is used to signal the selected Stations as well as for channel sounding, (b) quantify the gain possible due to aggregation and the influence of the size of the buffer on the attainable throughput (c) decide on buffer size to optimize performance.

If the number of stations is very large, due to the heterogeneity of destinations, complete benefit of packet aggregation is not realized. With a larger queue size, the throughput is enhanced - although it is at the cost of resulting in an increased delay.

e) **Efficient Aggregation Scheme - Problem**: To achieve efficient frame aggregation in downlink MU-MIMO transmission, we have to consider transmission efficiency in each SS channel. Different transmission time between frames on different SS causes space channel time for reducing which parameter, the data size based frame aggregation scheme has been proposed. This scheme uniforms the amount of data size carried in the wireless frame on each SS channel. Under multi-rate transmissions, however, wireless frame duration (time length) generally differs between SS, and then, the space channel time arises in this scheme.

Moreover, due to first-in first-out (FIFO) based MT selection, it cannot always construct longer wireless frame, leading to large signaling overheads and then low transmission efficiency.

Against these problems, we have previously proposed the wireless frame duration based frame aggregation scheme considering multi-rate transmission for downlink MU-MIMO transmissions. This scheme achieves a reduction of space channel time by uniforming the wireless frame duration considering transmission rate of each receiving MT even under multi-rate transmissions. Signaling overhead reduction is also achieved by adopting receiving MT selection with descending order selection based on wireless frame duration. However, this receiving MT selection tends to prioritize MTs with lower transmission rate over higher one, leading to degradation in throughput performance. In addition, this scheme does not consider frame error occurring in wireless channel, so it also deteriorates transmission efficiency due to excessive retransmissions of errant frames under actual environments with transmission errors.

Solution: [56] has proposed the efficient frame aggregation scheme for DL MU-MIMO transmissions which enhances system throughput and decreases frame error rate. Here, the receiving MT selection gives higher priority to MTs achieving higher throughput in the next MU-MIMO transmission while reducing signaling overhead, resulting in an improvement in throughput. The wireless frame setting in the proposed scheme, introducing hybrid frame aggregation method, provides lower frame error rate than acceptable level by using frame size adaptation.

This is an efficient frame aggregation scheme and is an enhanced version of the previously proposed wireless frame length based frame aggregation scheme.

Conclusion: The transmission performance and its fairness between MTs have been evaluated though system-level simulation. From the results, the proposed scheme greatly improves both performances, compared with the conventional frame aggregation schemes.

f) **Sub-Channel Scheduling** Details of sub-channel scheduling in cellular networks are in Appendix-4.

In [57], a new solution with 802.11 PHY and MAC layers is designed with multiuser channel access and a dynamic sub-channel assignment method based on traffic priority.

Here, selection happens, of the best set of users where they are allocated subcarriers according to their CSI. In [57], the authors have offered a distributed solution using which optimal scheduling can be determined in a practical system. The objective of the RA scheme is to choose the optimum set of transmission parameters for a given user bearing in mind the channel quality.

In particular, the designed approach can support QoS traffic over an OFDM-based network. Assuming information of the channel gains for all users, an adaptive multiuser OFDM sub-channel allocation and modulation solution is proposed.

A smart solution is designed for high-density wireless environments, such as airports, campuses, sports stadiums, etc. In addition, efficient resources are allocated for multiuser OFDMs over frequency selective channels with data and user priorities.

Conclusion: The algorithm was evaluated using NS-3 in a multiuser frequency selective fading environment with different time-delay spread values. It is seen from the results that the recommended solution leads to a RT aggregation model with a stable throughput. It surpasses traditional multiuser OFDM systems with static TDMA or FDMA methods, which use fixed and predetermined time-slots.

3.9.3 PHY Layer

1) **Precoding and Detection**

a) **Regularized Block Diagonalization Precoding - Problem**: MIMO system can increase the capacity and spectral efficiency without consuming bandwidth greatly.MU-MIMO system can enhance the performance much more. But interference of multiple users must be considered in MU-MIMO system. Interference cancellation is the main goal of precoding algorithm. BD precoding is widely employed for the sake of its low computational complexity and ability of eliminating the interference of other users. However, BD precoding doesn't consider the addictive noise and it has a problem of limited dimension antennas.

The MU-MIMO-OFDM based on bit-interleaved coded modulation (BICM) scheme is one of the enhanced air interface key technologies. In MU-MIMO, the transmitted information of different users will cause interference to each other, which can be removed by precoding. Due to its good performance and low complexity, Block Diagonalization (BD) precoding algorithm is widely used in MU-MIMO schemes. However, it also has obvious defects: the ignoring of addictive Gaussian noise and the restrictions of antenna dimensions (i.e. number of receiving antennas < number of transmitting antennas).

Solution: In [58], we employ Regularized Block Diagonalization (RBD) with signal space diversity (SSD) to solve these problems. The proposed scheme exploits time, frequency and space diversity by using constellation rotation and Q-component interleavers to optimize the MIMO-OFDM, channel coding and modulation together.

Conclusion: Simulation results clearly show that the RBD-precoded mechanism surpasses the conventional bit interleaved coded modulation (BICM) scheme without SSD by 3.0 dB SNR gain. Also, the RBD precoding obtains 1.8dB SNR gain more than the block Diagonalization (BD) precoding. Compared with BD precoding, RBD precoding not only reduces the interference of other users, but decreases the impact of addictive Gaussian noise as well. Besides, it eliminates the restrictions of antenna dimensions.

b) **Study of Precoding Mechanisms**: In [59], the performance of channel inversion (CI) and BD with limited CSI and compressed feedback are analyzed. The results evaluate the achievable sum-rate between the user selection metrics and present the PER performance.

Conclusion: Using the simulations, it is seen that superior data rates are achieved by greedy user selection algorithm and low complexity precoding schemes. BD performs better compared to CI with respect to packet error rate on using compressed feedback. Note: This paper only analyses the standard for Precoding and no enhancement is provided.

c) **MU-MIMO and SU-MIMO precoding**: [60] the precoding mechanisms in MU-MIMO and SU-MIMO are compared assuming perfect CSI at the Transmitter. Here, optimization of the BD, zero forcing and singular value decomposition (SVD) mechanisms are proposed using an algorithm that reduces the computational calculations.

Mode switching SNR, a critical metric, can be identified by verifying the impact of different precoding mechanisms in SU-MIMO and MU-MIMO. This metric aids in selecting between SU-MIMO and MU-MIMO and to meet the demands of all users to an optimum extent as well. [60] Discusses the maximum capacity that can be achieved with BD and zero forcing (ZF) for MU-MIMO and SVD precoding in SU-MIMO.

Conclusion: The analysis demonstrates that the proposed algorithm not only reduces the computations significantly but has a lesser amount of performance loss as well. Once the mode switching point is estimated, the MAC layer can opt between SU-MIMO or MU-MIMO based on the SINR values.

d) **Precoding with Detection Techniques**: The performance improvements under realistic channel conditions is measured when BD and CI are paired with various MIMO detection techniques.

In [61], BER and PER performance are examined. It is observed that high performance MIMO decoders such as LRA MMSE decoder improve the uncoded BER performance of the BD precoding but not for the CI precoding. The effective channel matrix of the BD precoding is a block diagonal matrix. Hence MIMO decoders can exploit the diversity present in the inter stream interference within a single user.

When the channel models B and D are compared and the PER performance is measured, it is seen that the channel model D is superior to channel model B. The channel model B is high correlated, the wide distribution of the singular values of the inverse increases the precoded signal power at the transmitter, and noise enhancement at the receiver side.

We considered the PER performance of each MU-MIMO transmission methods when MIMO decoder is changed. Between the Linear MMSE and the V-BLAST MMSE decoders, the PER performance measured in the Linear MMSE is better than in the V-BLAST MMSE. While the uncoded BER performance of V-BLAST MMSE is superior to Linear MMSE, performance degradation caused by error propagation limits the performance of coded V-BLAST MMSE. Also, we observe the substantial improvement of the PER performance of the BD precoding when using the LRA MMSE decoder.

Conclusion: It is seen that (1) PER performance of the BD precoding can be improved by the MIMO decoder and (2) the BD precoding is more effective compared to CI precoding.

e) **SIC MIMO detectors with LR techniques**: The challenges and concerns related to the application of lattice reduction (LR) techniques in IEEE 802.11ac systems in successive interference cancellation (SIC) receivers in SU and MU OFDM MIMO transceivers are explored in [62].

A comparative evaluation of the following OFDM MIMO detectors is undertaken: (1) MMSE; (2) ordered successive interference cancellation MMSE

(OSIC/MMSE); (3) LR Zero-Forcing (LR ZF); (4) LR MMSE; (5) LR Zero-Forcing SIC (LR ZF-SIC); (6) LR MMSE-SIC.

Conclusion: The simulation results, assume realistic operational conditions faced in IEEE 802.11ac WLANs, such as, temporal autocorrelation synchronization algorithm, least square (LS) MIMO channel estimation scheme, spatial-correlated and frequency selective TGac channel models. It is demonstrated that the application of LR MMSE-SIC MIMO detectors with hard-decision Viterbi decoding in IEEE 802.11ac WLANs allows substantial power gains in relation to linear LR MMSE MIMO detectors over 8x8 MIMO channels in a SU scenario. On the other hand, these detectors allow a power gain of only 1 dB for 3x3 MIMO channels for the SU environment and no expressive power gains when a MUMIMO system with 4 transmit antennas, 2 receive antennas per user and 2 users is considered. It is verified that the OSIC MMSE detectors offer severe performance degradation due to error propagation with realistic TGac MIMO channels models.

f) **Precoding Simulation Framework:** In [63], a framework to analyze and design the PHY layer of IEEE 802.11ac wireless LANs with focus on MU operation is explained. The performance of CI, regularized CI and BD precoding mechanisms with ZF and MMSE MIMO detectors is evaluated over different configurations of 802.11 Task Group n (TGn) and TGac channel models.

Conclusions: It is seen that the performance of MU-MIMO 802.11ac systems displays a similar performance for both TGn and TGac channel models. Finally, a unified performance assessment of five transceiver MU-MIMO architectures adapted for IEEE 802.11ac systems was undertaken: BD +ZF, BD+MMSE, CI+ZF, CI+MMSE and RI+MMSE. It is seen that for low SNR region the best performance is achieved with RI-MMSE schemes, whereas there is no noteworthy difference on the performance in high SNR region.

g) **Detection Mechanisms**: [64] proposes to detect the SS in 802.11ac device using (2×2), (4×4), (8×8) MU-MIMO systems using some of the MIMO detection algorithms. Here, the objective is to evaluate the effect of Detectors/Interference Canceller (IC) like ZF, MMSE and the proposed ZF-Successive IC (SIC) with Optimal Peak Power Ordering (OPPO), ZF-SIC-OPPO on the performance of 802.11ac in a Rayleigh fading channel.

Conclusions: The result of the proposed ZF-SIC-OPPO was weighed against the existing mechanism. The OPPO method for ZF -SIC showed enhanced performance in both the SNR and BER. ZF-SIC-OPPO detector outperformed ZF and MMSE with the lowest BER for a MIMO (8 × 8) 802.11ac system.

2) PHY Sounding

a) **Multi-User Transmission enhancer**: In [65], the evaluation of a downlink MU_MIMO sounding protocol called MUTE is presented. In a practical system, the beamformer should compromise between the frequency of sounding and CSI accuracy. MUTE decouples the sounding set selection (used to collect CSI) from the transmission set selection, thereby helping in reducing the sounding overhead and simultaneously selecting the best users. The decoupling also offers the choice of sounding a particular user or not, separately from the set of users in the list meant for the next transmission. This, consequently, reduces overhead associated with sounding by exploiting the presence of users with stable channels. At the same time, adequately accurate data about channel statistics of associated users is given to the AP. Thus, the AP can pick the user group that maximizes an objective function such as achievable rate or fairness criteria.

Conclusion: Using test bed experiments, it is shown that MUTE can decrease the sounding overhead appreciably while minimizing data rate losses resulting from inaccurate channel estimation.

b) **Comprehensive Sounding Control Scheme**: The sounding overhead may severely degrade the performance of a WLAN. In [66], a sounding control scheme for 802.11ac MU-MIMO is proposed which considers broadly the sounding control needs of the network environment, including channel coherence times, nodes' DL traffic loads, DL SNRs, etc.,. The sounding node set and sounding interval to maximize the long-term expected MU-MIMO throughput gain are jointly determined.

Although all the mobile nodes are in the same space, different wireless links experience different propagation environment variation due to mobility, behavior of nearby people, and so on. For example, some wireless links experience stationary propagation environments, such that their channel coherence times are several hundreds of milliseconds. On the other hand, some of the other wireless links are moderately time varying, such that their channel coherence times are several tens of milliseconds. Finally, the remaining wireless links vary quite fast, such that their channel coherence times are several milliseconds. Thus, MIMO channel measurements are conducted in practical WLAN environments and the performance of the mechanism is evaluated by employing the real channel data traces.

Conclusion: The proposed scheme is shown to achieve remarkable performance improvement over the existing schemes which consider channel coherence times only.

c) **Problem**: In MU-MIMO, the AP selects both the user set and the mode (number of transmit and receive antennas) and after completing sounding and before commencing transmission. Thus, optimal user groups are selected given full CSI of a set of potential receivers (which is a major overhead) or by depending on intermittent probing or stale CSI to estimate the full CSI. The measured CSI can easily become stale between packets, making data collected from previous transmissions not useful in prediction of future environments.

Pre-sounding User and Mode selection Algorithm: In [67], Pre-sounding User and Mode selection Algorithm (PUMA), an algorithm for selecting mode and user prior to sounding is designed, implemented and evaluated. PUMA, even without CSI, (i) exploits theoretical properties of MUMIMO system scaling with respect to mode, (ii) characterizes the relative cost of each potential mode, and (iii) estimates per-stream transmission rate and aggregate throughput in each mode for a potential user set. After PUMA chooses the appropriate mode and user group, the selected protocol's sounding mechanism is applied on the intended user subset to carry out the transmission.

The PUMA algorithm is executed before the start of any MU-MIMO transmission with only a priori information i.e., without any information gathered from previous multi-stream communication or channel sounding.

Conclusion: It is seen that, on average, PUMA selects the mode and group that achieves an aggregate rate within 3% of the throughput of what would have been obtained by sounding all users.