Review on QoS provisioning approaches for supporting video traffic in IEEE802.11e: challenges and issues

MOHAMMED A. AL-MAQRI1,2, MOHAMED A. ALRSHAH1,(Senior Member, IEEE), AND MOHAMED OTHMAN.1,3,(Senior Member, IEEE)

1Department of Communications Technology and Networks, Universiti Putra Malaysia, Serdang, 43400, Malaysia
2Azal University for Human Development, 60th Street, Sana’a, Yemen
3Computational Science and Mathematical Physics Lab, INSPEM, Universiti Putra Malaysia, Serdang, 43400, Malaysia

Corresponding author: Mohamed A. Alrshah (e-mail: mohamed.asnd@gmail.com)

This research was funded by the Malaysian Ministry of Education under UPM/700-2/1/GPB/2017/9557900 Putra with High-Impact Grant.

ABSTRACT Recently, the demand for multimedia applications is dramatically increased, which in turn increases the portion of video traffic on the Internet. The video streams, which require stringent Quality of Service (QoS), are expected to occupy more than two-thirds of web traffic by 2019. IEEE802.11e has introduced HCF Controlled Channel Access (HCCA) to provide QoS for delay-sensitive applications including highly compressed video streams. However, IEEE802.11e performance is hindered by the dynamic nature of Variable Bit Rate (VBR) video streams in which packet size and interval time are rapidly fluctuating during the traffic lifetime. In order to make IEEE802.11e able to accommodate with the irregularity of VBR video traffic, many approaches have been used in the literature. In this article, we highlight and discuss the QoS challenges in IEEE802.11e. Then, we classify the existing QoS approaches in IEEE802.11e and we also discuss the selection of recent promising and interesting enhancements of HCCA. Eventually, a set of open research issues and potential future directions is presented.

INDEX TERMS 802.11e, HCCA, MAC, Multimedia, VBR, QoS, Survey.

I. INTRODUCTION

THE optimal transport of delay-constrained multimedia services over WLANs requires adaptation to many aspects of Open Systems Interconnection (OSI) model layers starting from delay constraints and bandwidth variations of the traffic at the application layer up to accommodation to wireless channel conditions and power constraints at the physical layer. The efficiency of 802.11e HCF Controlled Channel Access (HCCA) function mainly depends on the accuracy of its scheduler in assigning network resources, such as channel bandwidth, to the traffic streams without jeopardizing the QoS constraints such as delay and throughput. Moreover, with the presence of delay-sensitive multimedia traffic with variable profile, the existing scheduling approaches become inefficient. Thus, the scheduler is required to consider the fluctuation of traffic in the scheduling process.

This article introduces an overview of the prime challenges for provisioning QoS for multimedia traffic with emphasize on Variable Bit Rate (VBR) traffic in IEEE 802.11e wireless networks. Then, it presents a taxonomy for the existing solutions, and describes the most representative properties, advantages, and design challenges. This taxonomy comprises the core approaches and techniques on IEEE802.11e protocol, with more emphasize on HCCA enhancements. Additionally, a systematic summarization and comparison for research contributions in each field are used to clearly identify the current challenges for further research. Finally, the article discusses the most critical issues which hinder the provisioning of QoS in wireless networks with a special attention to polling and Transmission Opportunity (TXOP) allocation enhancements.

This paper is a survey of QoS provisioning for video transmission in IEEE802.11e, which is organized as follows: Section II exhibits the background about IEEE802.11e standard and its functions. Section III presents the main challenges in IEEE802.11e WLANs. Section IV classifies and reviews the core approaches in IEEE802.11e WLANs, which were proposed to enhance QoS provisioning for multimedia traffic. A number of leading approaches aiming at
improving the QoS for multimedia traffic has been discussed in Section V. Section VI shows a general comparison of the IEEE802.11e approaches and their targeted features, and lists some of the strength and limitation criteria of these approaches. Section VII and Section IX identifies research trends, challenges, and potential future areas related to the article’s scope, and finally Section X concludes the article.

II. IEEE802.11E STANDARD

Several amendments have been made to the legacy IEEE802.11 WLAN standard [1], as shown in Table 1. IEEE802.11e is one of the approved versions of IEEE802.11 standard, which defines a combination of Quality of Service (QoS) improvements on the Medium Access Control (MAC) layer for WLAN applications, as shown in Fig. 1. The standard is critically important for applications that are very sensitive to delay, such as Voice over WLAN (VoWLAN) and multimedia streaming.

In IEEE802.11e, the QoS feature includes an extra coordination function called Hybrid Coordination Function (HCF). This function combines both functionalities of the well-known Point Coordination Function (PCF) and Distributed Coordination Function (DCF). In order to permit the use of a uniform assortment of frame exchange sequences for QoS data transfer during the time of both Contention Period (CP) and Contention Free Period (CFP), the HCF introduced some enhanced frame subtypes and QoS-specific mechanisms. As for contention-based transfer, HCF employs a contention-based channel access approach, namely Enhanced Distributed Channel Access (EDCA), while for contention-free transfer it uses a controlled-channel access method, so-called HCCA. Stations (STAs) might obtain TXOPs using EDCA, HCCA or both schemes together. Thus, a TXOP is defined as EDCA TXOP if it is obtained by the contention-based channel access, while it is defined as HCCA-TXOP if it is obtained by the controlled channel access.

![FIGURE 1. MAC architecture in IEEE802.11e.](image)

A. ENHANCED DISTRIBUTED CHANNEL ACCESS (EDCA)

EDCA mechanism has been designed to provide sort of differentiated distributed access to Wireless Medium (WM) for STAs by using eight uneven User Prioritiess (UPs). It determines four Access Categories (ACs) to provide support for traffic delivery at the STAs using UPs, which produces the AC, as shown in Table 2. For every AC, an enhanced variant of DCF, called Enhanced DCF (EDCF), contends for TXOPs using a set of EDCA parameters. For more details about the EDCF refer to [11]. Implementation of this mechanism is easy; however, the QoS requirement of a realtime traffic can not always be met, especially when the heavy load conditions occur. In heavy loaded scenarios, higher prioritized traffic QoS requirement may easily be broken even though it exhausts most of the available bandwidth. However, lower prioritized traffic may be starved and severely deteriorated in both efficiency and effectiveness.

B. HCF CONTROLLED CHANNEL ACCESS (HCCA)

As known in IEEE-802.11e, a synchronization signal is rhythmically sent to all of the connected stations in the Basic Service Set (BSS). The time between two subsequent signals makes a super-frame, where a service can be delivered through this super-frame over two periods of time, CP and CFP. The data of any station has to be transmitted during a period of time, namely TXOP, which is dedicated for a QoS-enabled Station (QSTA) to transfer its MAC-Service Data Units (MSDUs). Fundamentally, TXOP is acquired through the contention-based access, which is known as EDCA-TXOP. As for the controlled medium access, the Hybrid Coordinator (HC) grants the TXOP to the QSTA (known as polled TXOP). Fig. 2 shows a clear example of 802.11e super-frame which demonstrates the interchange of one controlled medium access and one contention-based period, where the later includes one QoS-enabled Access Point (QAP) and three QSTAs. In general, controlling medium access occurs either within the CP or through the CFP if the medium remains idle for at least one period of PCF Inter Frame Space (PIFS). In order to support QoS in HCCA, many researchers have proposed to improve the existing PCF by controlling the transmission only within the CFP. Therefore, the data packets of any wireless station in HCCA can be only transmitted during a declared period of time in the poll frame.

![FIGURE 2. An 802.11e super-frame example, CFP and CP. In the CFP, the frame exchange takes a place throughout the polling mechanism, while in CP the QSTAs have to listen to the medium transmitting data packets.](image)
TABLE 1. The family of IEEE-802.11 versions

<table>
<thead>
<tr>
<th>Standard</th>
<th>Objective</th>
<th>Frequency and Modulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE802.11 [2]</td>
<td>To provide up to 2 Mbps bit rate.</td>
<td>2.4 GHz by utilizing DSSS and FHSS.</td>
</tr>
<tr>
<td>IEEE802.11a [3]</td>
<td>To provide up to 54 Mbps bit rate.</td>
<td>5 GHz by utilizing OFDM.</td>
</tr>
<tr>
<td>IEEE802.11b [4]</td>
<td>To provide up to 11 Mbps bit rate.</td>
<td>2.4 GHz by utilizing HRDSSS.</td>
</tr>
<tr>
<td>IEEE802.11c [5]</td>
<td>To ensure proper bridging operations.</td>
<td>-</td>
</tr>
<tr>
<td>IEEE802.11d [6]</td>
<td>To cover more regulatory domains.</td>
<td>-</td>
</tr>
<tr>
<td>IEEE802.11e [7]</td>
<td>To define new QoS enhancements to 802.11a and 802.11b.</td>
<td>-</td>
</tr>
<tr>
<td>IEEE802.11f [8]</td>
<td>To provide interoperability for roaming among different APs.</td>
<td>-</td>
</tr>
<tr>
<td>IEEE802.11g [9]</td>
<td>To provide up to 54 Mbps bit rate.</td>
<td>2.4 GHz by utilizing OFDM.</td>
</tr>
<tr>
<td>IEEE802.11n [10]</td>
<td>To provide up to 600 Mbps bit rate.</td>
<td>2.4 and 5 GHz by utilizing MIMO-OFDM.</td>
</tr>
</tbody>
</table>

Table 2. Mappings of user priority to access category

<table>
<thead>
<tr>
<th>Priority</th>
<th>UP</th>
<th>AC</th>
<th>Designation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowest</td>
<td>1</td>
<td>AC_BK</td>
<td>Background</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>AC_BK</td>
<td>Background</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>AC_BE</td>
<td>Best Effort</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>AC_BE</td>
<td>Best Effort</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>AC_VI</td>
<td>Video</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>AC_VI</td>
<td>Video</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>AC_VO</td>
<td>Voice</td>
</tr>
<tr>
<td>Highest</td>
<td>7</td>
<td>AC_VO</td>
<td>Voice</td>
</tr>
</tbody>
</table>

1) Reference design of HCCA
At the point if a QSTA wants to transmit its realtime Traffic Stream (TS) within the contention-free period, it has to send an ADDTS-Request to the QAP. This ADDTS-Request declares the requirements of QoS for that specific TS within the relevant TS Specification (TSPEC) domain. Consequently, the QAP will try to fulfill the requirements while conserving the QoS of existing admitted flows. If the ADDTS-Request is accepted, the QAP will reply an ADDTS-Response back to the relevant station, then, this station will be admitted to the QAP polling list. Table 3 shows the compulsory TSPEC parameters and their symbols.

TABLE 3. Symbols used for TSPEC and scheduling parameters

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>Mean Data Rate</td>
</tr>
<tr>
<td>$L$</td>
<td>Nominal MSDU Size</td>
</tr>
<tr>
<td>$M$</td>
<td>Maximum MSDU Size</td>
</tr>
<tr>
<td>$D$</td>
<td>Delay Bound</td>
</tr>
<tr>
<td>$SI$</td>
<td>Service Interval</td>
</tr>
<tr>
<td>$\min SI$</td>
<td>Minimum Service Interval</td>
</tr>
<tr>
<td>$MSI$</td>
<td>Maximum Service Interval</td>
</tr>
<tr>
<td>$R$</td>
<td>Physical Transmission Rate</td>
</tr>
<tr>
<td>$BI$</td>
<td>Beacon Interval</td>
</tr>
<tr>
<td>$O$</td>
<td>Physical Layer mode (PHY) and MAC Overhead</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of packets</td>
</tr>
<tr>
<td>$T$</td>
<td>superframe duration</td>
</tr>
<tr>
<td>$TC_P$</td>
<td>Contention-based duration</td>
</tr>
</tbody>
</table>

After accepting new ADDTS-Request, the HCCA scheduler will go through the following steps:

1) Assigning service interval
HCCA [11] computes the Service Interval (SI) as a sub-multiple of the whole Beacon Interval $BI$, which is calculated as the minimum of the maximum SIs of all priorly accepted traffic streams including the incoming data traffic. Equation (1) is used to calculate the SI:

$$SI = \frac{BI}{\max(MSI_i)}$$

where $MSI_{min}$ is computed as in Equation (2):

$$MSI_{min} = \min(MSI_i), i \in [1, n]$$

where $MSI_i$ denotes the maximum SI of the $i^{th}$ stream and $n$ denotes the number of all previously admitted QSTAs’ traffic streams.

2) Allocating TXOP
Variant TXOP is allocated by HC to every accepted QSTA based on the declared QoS parameters in the TSPEC, which allows the QSTA to obtain the required QoS. The HC calculates TXOP for the $i^{th}$ QSTA based on the expected MSDUs, which may arrive at $\rho_i$, as calculated in Equation (3):

$$N_i = \left\lceil \frac{SI \times \rho_i}{L_i} \right\rceil$$

where $L_i$ denotes the MSDU of the $i^{th}$ station.

Thereafter, the TXOP of the $i^{th}$ station ($TXOP_i$) is calculated as the required time to transmit $N_i$ MSDU or one maximum MSDU at the relevant physical rate $R_i$, as in Equation (4) below:

$$TXOP_i = max \left( \frac{N_i \times L_i}{R_i} + O, \frac{M}{R_i} + O \right)$$

where $O$ represents the total overhead, including MAC and physical headers, poll frames overheads, inter-frame spaces (IFSs) and acknowledgments.

3) Admission control
The Admission Control Unit (ACU) regulates the admission of the TS while maintaining the QoS of the previously admitted TSs. When the ACU receives a request of admitting a new TS, the ACU calculates a new SI using Equation (1) and estimates the number of MSDUs that may arrive at this new SI based on Equation (3). Then, the ACU calculates the $TXOP_i$ for the particular TS using Equation (4). Finally, the ACU would admit the relevant TS only if the following inequality is satisfied:

$$\frac{TXOP_{n+1}}{SI} + \sum_{i=1}^{n} \frac{TXOP_i}{SI} \leq \frac{T - TC_P}{T}$$
Fig. 3 shows an example of an admitted stream from STA. The beacon interval is 100ms and the maximum SI for the stream is 60ms. The scheduler sets a scheduled SI to 50ms with complying to Equation (5), where $n$ represents the number of all admitted streams, $n+1$ denotes the index of incoming TS, $T$ indicates the beacon interval and $T_{CP}$ is the time reserved for EDCA contention-period. The HC sends an ADDTS-Response to the relevant QSTA only if Equation (5) is satisfied, and it sends a message of rejection otherwise. Then, the HC will add the accepted TS to its polling list.

![FIGURE 3. Schedule for streams from STAs i to k. The streams are scheduled in Round-Robin fashion governed by the admission control unit](image)

III. QOS CHALLENGES IN IEEE802.11E WLANS

QoS is the overall effect of the service performance, which defines the satisfaction degree of a service user and manifests itself in a number of subjective or objective parameters [12]. There are two ways to investigate the QoS, subjective (perceptive) and objective (network) measurements. In the subjective measurement, the user involves to carry out a series of assessment tests, while in objective measurement, typical network performance throughput, packet loss, packet jitter and delay is evaluated. In order to meet the user satisfaction, the subjective QoS parameters shall be translated into a set of objective QoS parameters, e.g. throughput, delay and losses.

QoS could be supported in different ways at different protocol layers as illustrated in Fig. 4. Some applications have the capability to adapt the generated traffic to the conditions of the underlying network in order to meet user expectations. An example is the use of the Real-time Transport Protocol (RTP) and associated RTP Control Protocol (RTCP) [13] to dynamically adapt the parameters of an audio and/or video streams, minimizing the losses due to congestion in the network [14]. Nevertheless, application layer mechanisms are usually not enough, since end-to-end QoS requires support in the lower layers of the protocol stack throughout the network nodes that the traffic must traverse from sender to receiver. However, this work mainly concerns with QoS provisioning at MAC layers.

The QoS provisioning of diverse multimedia streams in a wireless environment imposes a chain of challenges due to many factors of OSI model layers [15], [16], [17] ranging from traffic characteristics in application layer down to the wireless channels nature in physical layer. In this section, a review of the major challenges that may emerge when providing QoS for delay-sensitive applications in IEEE802.11e wireless networks.

A. ADAPTATION TO FLUCTUATION OF APPLICATION PROFILE

Generally, the application profile of a traffic is defined by the alternation of the traffic over the time. The QoS provision of a VBR flow is substantially influenced by the variation of the application profile over the time. The accurate estimation of the traffic at the application layer can significantly enhance the performance of underlying functions of MAC layer to adapt its parameters according to these changes.

The VBR video source can be generally classified into three main categories [18], [19]: I) variable packet size with constant Generation Interval (GI), e.g., MPEG-4 videos; II) constant packet size with variable GI, e.g., Voice over Internet Protocol (VoIP); and III) variable packet size with variable GI, e.g., H.263.

The transmission of video streams can be significantly affected by the compression techniques used, such as MPEG-4 and H.263. The nature of the frame structure and the compression algorithm used along with the variations within video scenes can significantly influence the burstiness level of the stream [20], [21]. The burstiness of a VBR stream traffic increases the complexity of network resources management to ensure QoS support for continuous stream playback. Although, the reference design of the HCCA scheduler is simple and efficient in supporting constant application profile, yet it is not adequate since it cannot address the fast-changing imposed by the VBR bursty traffic, which hinders the performance of HCCA by causing packets to wait for a longer time in their transmission queues.

In case of downlink traffic, from QAP to QSTAs, the QAP is aware about its data queues and shall use its highest priority to seize the channel if it remains idle for a duration of PIFS without undergoing back-off procedure. However, due to the fact that QAP suffers from the lack of information about the uplink transmission queue status, an adaptive scheme is required to allow the scheduler to adjust its behavior based on the current application characteristics. Generally, adaptation to the application can be categorized according
to its variability level-based in the three well-known types mentioned in III-A.

In MAC layer, the uplink traffic profile can be determined using different ways, such as estimating the data buffer of the flow, predicting the packet generation time and/or traffic load at a specific time, or obtaining actual information through cross-layer architecture design. By having the traffic profile, the HCF can adjust one or more of its functions such as polling [22], SI assignment [23], TXOP allocation mechanisms [24] which allows it to instantaneously adapt to QoS requirement of the flow.

The QoS of VBR video transmission is unguaranteed due to the fact that those packets are queued for a duration equivalent to SI until already-queued packets in the buffer are delivered. Recall that during each SI, the reference HCCA scheduler allocates a fixed TXOP to each QSTA based on its mean rate requirements regardless the real VBR traffic changes. There are three QoS challenges relevant to Class I, II and III of VBR traffics.

1) QoS Challenges of Class I video flows

HCCA scheduler fails to accommodate to variability Class I traffic which, in turn, leaves the wireless bandwidth in underutilization status. Assume, without loss of generality, that an identical TXOP duration is allocated for every QSTA, consequently, each QSTA will waste the same amount of unused TXOP ($T_u$). Thus, Equation (5) can be rewritten as follows:

$$\frac{TXOP_{n+1}}{SI} + \frac{TXOP - T_u}{SI} \leq \frac{T - T_{CP}}{T} \quad (6)$$

According to Reference [25], using different SIs for different streams will improve the bandwidth utilization up to 50%. In other words, the $T_u$ in Equation (5) will be equal to $\frac{TXOP}{2}$. Therefore, the Equation (5) can be again rewritten as follows:

$$\frac{TXOP_{n+1}}{SI} + \frac{TXOP}{2 \times SI} \leq \frac{T - T_{CP}}{T} \quad (7)$$

which means that the number of admitted flows can be maximized to double the number of admitted flows when different SIs are used.

2) QoS Challenges of Class II video flows

In Class II, when $QSTA_i$, at any SI, exploits only portion of its allocated $TXOP_i$ at the traffic setup time, namely $T_{eff}$, leaving an unspent amount of $T_u$. Thus, the following relation can be held [26]:

$$\sum_{i=1}^{N} T_i = T_{eff} + T_{eff}^2 + \cdots + T_{eff}^N$$

$$= TD_1 - T_u + TD_2 - T_u^2 + \cdots + TD_N - T_u^N$$

$$= \sum_{i=1}^{N} TD_i - \sum_{i=1}^{N} T_u^i$$

$$\sum_{i=1}^{N} T_u^i \geq 0, \sum_{i=1}^{N} TD_i, \text{and } \sum_{i=1}^{N} T_u^i \text{is the total TXOP scheduled in any SI used in HCCA and ATXOP, respectively. It is worth noting that } TD_i \text{is the TXOP duration of the } QSTA_i \text{including the poll overhead. Thus, the delay of } QSTA_i \text{in an SI is computed as follows:}$$

$$D_{SI} = \sum_{j=1}^{i-1} (TD_i - T_u^i) + T_L + T_{poll} + 2 \times SIFS \quad (9)$$

Altogether, the real QoS challenge is to minimize packet delay by minimizing the surplus amount, namely $T_u^i$.

3) QoS Challenges of Class III video flows

In video streams like H.263, the deviation comprises not only packet size but also shows up to high variation in generation interval which makes the matter much worse. In any SI, scheduling a QSTA based on its TSPEC likely imposes allocating surplus of TXOP duration which leads to wasting of the resources. This waste of resources due to the variations in data rate influences the efficiency of the scheduler that does not implement any recovery policy. Besides, due to the variation in the packet generation interval, perhaps there are some QSTAs that are not ready to transmit which will be considered as over-polling state. This waste of resources, due to the variations in data rate, influences the efficiency of the scheduler that does not implement any recovery policy. Overall, it hinders the meet of delay bounds requirements, which leads to a degradation in QoS provisioning.

Consider the example illustrated in Figure 5 where four QSTAs are polled for transmission in both CFP and Controlled Access Phase (CAP). In this example, $TXOP_1$, $TXOP_2$, $TXOP_3$ and $TXOP_4$ to $QSTA_1$, $QSTA_2$, $QSTA_3$ and $QSTA_4$, respectively. The wasted TXOP and over-polling issues experienced using reference HCCA, inspired from the example, are as illustrated in Figure 5.

- **Over-polling of QSTAs** As illustrated in this example, due to the lack of awareness about the change in the traffic profile, some QSTAs may receive unwanted poll messages as their transmission queues are empty. $QSTA_2$ and $QSTA_3$ in CP, and $QSTA_2$ in CFP will respond with a null-frame causing unwanted delay to all QSTAs that may come after them in the same SI.

- **Wasted TXOP duration** Since some $QSTA$s, such as $QSTA_1$, experience a high instant drops-down in data rate, only a short amount of the given TXOP duration is utilized. In this case, the channel might remain idle for a period of time greater than the Short Inter Frame Space (SIFS) and the control of the medium conveyed to Access Point (AP) to poll the next station in the list.

Even though the effect of wasted TXOP duration in the packet delay is not as high as that caused by over-polling case, however, it is considerably can go high as the number of stations in the network increases.
B. ADAPTATION TO VARYING NETWORK CONDITIONS

Due to the phenomena of path loss, multipath fading, shadowing, and interference, wireless networks likely suffer from Signal-to-Interference Plus Noise Ratio (SINR) [23]. The fluctuation of the underlying channel capacity will hinder the QoS provisioning for time-sensitive applications. Consequently, two possible ways to be applied on the QoS algorithms in order to encounter this challenge and meet the required QoS needs. The first one is by computing the transmission time for the packets based on the minimum physical bit rate announced. By doing so, the QoS is guaranteed, however, this technique gives rise to degradation in bandwidth efficiency as the bandwidth might get higher anytime while only the minimum link rate is considered. The second one is to encourage the QoS algorithm to take into account the link adaptation mechanism of WLANs over the time.

Although, piggybacking feature of HCCA is basically designed to improve the channel capacity, it may inversely behave when a station experience successive retransmission or channel noise. This issue has been referred to in [27] as "the piggyback problem as the low physical transmission rate". If any QSTA was transmitting at a low physical rate due to channel error, QAP will accordingly decrease the transmission rate of the piggybacked Contention Free (CF-Poll) frame. This, in turn, will result in channel efficiency degradation and will increase the TSs’ frame delay of other stations involving in the Network Allocation Vector (NAV) process.

In case of VBR traffic transmission over WLANs and apart from the issues and challenges of HCCA reported in [18], [28], the major issue of the reference scheduler is the unawareness about the inherent wireless time-varying channel condition[29]. Keeping aware about the channel status has a major impact on the scheduler performance as it can potentially degrade the service differentiation process, even though, HCCA has been observed to perform well in heavy loaded network [30], [31], especially with the emergence of several physical layer technologies such as Adaptive Modulation and Coding (AMC) schemes.

C. BANDWIDTH UTILIZATION

In HCCA, after receiving an ADDTS-request from the station, the scheduler needs first to calculate the required TXOP duration taking into account its TSPEC parameters. Thereafter, the used admission control mechanism will check the ability to accept the new TS. If the new TS is accepted, the SI will be computed as the minimum among all delay bounds of admitted streams, which is enough to meet the most urgent delay requirement to guarantee the required QoS service for the admitted streams. Finally, the round robin approach is utilized to allocate TXOPs to the involved station. Even though the use of this design is very simple and straightforward, it still suffers from some challenging issues related to the efficient use of the bandwidth. Indeed, the use of round robin approach in HCCA scheduler to serve all TSs in one SI might lead to over-allocating the bandwidth, which in turn leads to under-utilizing the channel bandwidth. Moreover, the waste of the wireless bandwidth may reach up to more than 50% in some cases [32].

In fact, based on the minimum physical rate and the characteristics of the incoming TSs, the ACU decides the number of admitted TSs to which the wireless resource will be allocated. This approach leads to allocating a constant amount of resource to every TS using the mean of single physical transmission rate, which is not compatible with the condition of current wireless bandwidth, especially VBR traffic. In other words, the ACU should consider both the physical layer and the service specific QoS parameters in order to be able to achieve effective bandwidth utilization [33]. With noticeably VBR flows, one of two scenarios likely occurs at some specific SIs. In the first scenario, the data rate becomes lower than the average value determined in the TSPEC, thus, the allocated TXOP will not be completely consumed which is considered as a wasting of resources. As for the second scenario, the data rate becomes greater than the average value determined in the TSPEC, thus, the assigned TXOP will not be enough to transmit the relevant data which increases the end-to-end delay of the flow. The possible solutions to solve these two problems as explained in these references [34], [35] are: (1) By increasing the TXOP duration to the average TXOP of traffic for the first...
case, knowing that it will reduce the bandwidth utilization, especially if the data rate is dropped down. (2) By applying the bandwidth reclaiming approach [36], [37], [38], [39].

D. NETWORK RESOURCES MANAGEMENT

Indeed the HCF of IEEE802.11e protocol is targeted to the provisioning of QoS throughout the service differentiation, yet the proper network resource management, such as coordinating between distributed (CP) and controlled (CAP) periods and link layer resources still in request [40]. In addition, a feasible ACU scheme is also required, which in such way can ensure that the QoS requirements are satisfied. The HCCA scheduler operates based on the static configuration of its traffic TSPEC parameters where they are constantly served for their lifetime to enforce resource sharing with ensuring that the desired QoS constraints are met. To this aim, a good resource utilization is often left to the heuristic network administrator know-how. However, this constant resource sharing policy might highly cause a scarce bandwidth utilization since it cannot adapt to the transformation of the traffic profile and the lifetime due to dynamic VBR traffic evolution.

As a resolution to this issue, a bandwidth sharing strategy is suggested to rely on a criteria which is driven by the performance [41], in which a common performance metric is recommended to be defined to differentiate between the traffic streams based on their performance requirements.

IV. CLASSIFICATION OF QoS SUPPORT FOR MULTIMEDIA TRAFFIC APPROACHES IN IEEE802.11E WLAN

In general, the enhancement approaches in IEEE802.11e protocol can be classified based on the access medium control fashion into distributed control and centralized control enhancements. In IEEE802.11e, EDCA operates based on the distributed access control while the HCCA represents the centralized access control. In [42], many QoS enhancements for 802.11 WLAN have been proposed and classified along with their advantages and disadvantages. Another survey in [43] has focused on the QoS provisioning in both EDCA and HCCA over IEEE802.11e networks. The HCCA enhancement approaches can be themselves classified into different categories according to several aspects such as the functional, structural, environmental and location aspects. In [44] and [45], the authors presented a survey of various admission control in IEEE802.11e and they classify schemes based on several aspects such as Measurement-Based, Model-Based and Hybrid schemes. In [46], the delay-EDD based scheduler has been compared to the feedback control based scheduler in order to provide a better comprehension about the so-called packet scheduling in 802.11 WLANs. Below is a short description of the different possible ways of IEEE802.11e approaches classification.

- Traffic flow direction: In infrastructure mode of IEEE802.11 WLANs, the traffic directions would be either downlink or uplink. "Downlink" refers to a traffic flow transmitted from AP to a mobile device, while "uplink" refers to a flow with a reverse direction. IEEE802.11e enhancements can be tailored to enhance the performance for either downlink or uplink traffic or in some cases for both directions.

- Targeted environment: Although IEEE802.11e MAC was originally designed for wireless infrastructure networks and widely used in WLANs, there have been some enhancements for adapting IEEE802.11e to work with other networks such as the improvement of polling and scheduling scheme over IEEE802.11a/e [47], IEEE802.11p networks [48], Ad-hoc Wireless Networks in [49] or in Integrated model of IEEE802.11e and IEEE802.16 [50], [51].

- Delay-EDD based and feedback control based: The pre-knowledge of packets arrival time is only possible for the downlink. While in the uplink, neither the delays of the head of line packets nor the quota of bandwidth needed by each flow are possible to be known by the access point. For this reason, IEEE802.11e schedulers have been categorized into the earliest due date and the feedback control class. A thorough comparison between these types has been presented in [46].

- Layered vs. cross layer: IEEE802.11e enhancements can be introduced in two structures, cross-layer and layered approaches. The cross-layer approaches rely on interactions between two layers of the OSI architecture. These approaches were motivated by the fact that providing lower or higher layer information to MAC layer to perform better. The layered approaches rely on adapting OSI layers independently of the other layers. Cross-layer is a promising direction to improve the overall performance of WLAN since it takes into account the interactions among layers [52]. Thus, several enhancements [53], [54], [55] prefer to use cross-layer design for obtaining accurate information for scheduling purposes.

- Technique or mechanism used: The HCCA scheduling approaches can be classified based on the techniques and/or mechanisms used in the design. In the literature, a diverse techniques were developed for HCCA scheduling to boost its performance for multimedia transmission over error-prone WLANs such as estimation based approaches [23], [56], predicting traffic profile [57], [58]. Moreover, some of these approaches modified one or more of HCCA mechanisms such as TXOP assignments [59], [60], [61], [62], polling mechanism [63], [64], [65] or ACU [54], [66], [67], [33].

- Analysis method used: The approach might be analyzed and/or evaluated using one of three methods, namely analytical model [68], [69], [70], [71]; simulation experiment [72], [73], [74] and test bed [75]. It is worth noting that the analytical model usually is done to capture the characteristics and the shortcoming
of the approach and prior to the proposal of a solution. Although, in simulation and test bed methods used for evaluating the proposed scheme, they might be carried out to provide a preliminary study to investigate a particular issue in the existing scheme for possible remedy.

V. QOS ENHANCEMENTS IN IEEE802.11E CONTROLLED ACCESS MODE

This section presents some of the leading approaches proposed in the literature to improve the QoS provisioning for multimedia traffic. More emphasis has been put on the transmission of VBR video streams in IEEE802.11e WLAN. The approaches are classified into six sets based on the strategy used to improve HCCA performance. However, some approaches can be matched to several types of strategies, but are only classified to their main strategy. Moreover, in the layered approaches the focus was only on the enhancements of HCCA at the MAC layer. The representative approaches are defined and their mechanisms are described along with a discussion concerning their strengths and weaknesses in improving QoS performance in IEEE802.11e WLAN. In addition, a comparison of the main characteristics of various HCCA approaches is provided for each category. Besides some mathematical models that study and provide insights to improve the HCF functions have been presented which can provide a promising avenue for further research and investigation.

A. HCCA POLLING ENHANCEMENTS

The polling mechanism of the legacy HCCA is responsible for the scheduling and the allocation of TS based on their fixed reservations. Thus, the efficiency of this mechanism highly depends on the accuracy of the flow specification declared to the HC. Yet, as the flow profile of VBR might highly vary over the time, the allocation based on fixed reservations will cause degradation in quality of multimedia flows even when the channel resource is surplus. More particularly, several issues may affect the efficiency of the HCCA such as the inefficient Round-Robin scheduling algorithm, the overhead induced by the poll frames, and the lack of coordination between the APs of the neighboring BSS. The representative approaches that address these issues and even more are summarized as follows.

CP-Multipoll is a robust multipolling mechanism aims to increase the channel utilization and to minimize the corresponding implementation overhead, which can be robust in error-prone environments like WLANs [76]. Moreover, the proposed scheme provides a polling schedule to ensure the bounded delay requirements of real-time traffic and it also provides an admission control mechanism. The main aim of this scheme is to design an efficient polling mechanism, due to its high impact on the performance of HCCA, which is able to serve both CBR and VBR real-time traffic. Unlike SinglePoll schemes where every STA receives one poll frame when polled, CP-Multipoll aggregates many polls in a single multipolling frame incorporating the DCF into the polling mechanism. The frame format of CP-Multipoll scheme is as shown in Fig. 6.

FIGURE 6. CP-Multipoll frame format

Basically, CP-Multipoll conveys the polling order into the contending order. This can be achieved by assigning different back-off values to the streams in the polling list according to their ascending order in the polling list and allow the back-off to execute as soon as they receive the CP-Multipolling frame. Besides minimizing the polling overhead by transmitting one polling message for all QSTAs in the polling list instead of sending polls as many STAs, the proposed scheme has other advantages over other multipoll schemes. The bursty traffic is better supported since the STA holds the channel only for a period needed to transmit its local buffered data. Moreover, in DCF access mode, if the STA does not use the poll frame due to empty data buffer, the other STAs in this polling group will immediately detect that the channel is idle and it will advance the starting of channel contention. However, the proposed mechanism is prone to hidden terminal problems since each STA will decrement back-off counter when it senses that the channel is idle. Thus, if hidden terminals exist in the network, different STAs will complete their back-off simultaneously and collision will happen. Due to the inherent hidden node issue of infrastructure wireless networks, CP-Multipoll cannot guarantee that all STAs in the BSS can sense the transmission of other STAs. In this case, the station will transmit its data immediately upon the expiration of its back-off timer leading to a collision.

CF-Poll piggyback scheme is presented by Lee and Kim [77] to optimize the usage rule of the CF-Poll piggyback scheme as defined in the IEEE802.11 standard [1] according to the TS load and the minimum physical transmission rate of a QSTA which suffer the deep channel fading. Consider the case of piggybacking, the CF-Poll in the QoS-ACK frame from QAP to $QSTA_3$, illustrated in Fig. 7, must be listened by all QSTA in the BSS. If any of the QSTAs experience low physical rate, which implies that $QSTA_3$ requires more time to receive the frame, the delay for all other QSTAs will be increased and the channel efficiency will be decreased.

Motivated from the aforementioned issue, the proposed work provides a guideline for the optimal usage of the CF-Poll piggyback scheme in IEEE802.11e and IEEE802.11n protocols. Simulation-based results reveal that the frame transmission delay is majorly affected by the minimum physical rate when CF-Poll is piggybacked in the QoS data frame while it is slightly influenced by the traffic load. The results show an inverse relationship between the CF-Poll piggyback scheme and the traffic load. Despite the
Since the VBR traffic exhibits variability in packet generation to provide an enhancement over HCCA mechanism, wasted.

consideration of the readiness of the station, STA with no merits of HCCA will be untapped. Moreover, there is no clear in both CFP and CP, for this reason one of the significant only functioning in CFP whereas HCCA is supposed to work be polled without colliding with others. However, DEB is back-off. When the back-off timer expires, the station can station to guarantee that the polled stations will have different occur between the nodes in the overlapped area. DEB uses a technique highlighted the issue of the collision incurred due polling the nodes in the overlapping area of two adjacent BSSs at the same time. This actually occurs due to the lack of coordination in HCCA between the adjacent APs. Consider the nodes 5, 6, 7, 8 in the overlapping area illustrated in Fig. 8. Since AP A cannot hear AP B, therefore the collision occurs between the nodes in the overlapped area. DEB uses a similar idea of sensing the carrier of EDCA since it manifests high robustness and flexibility controlling the medium at the overlapping BSSs. A virtual polling has been achieved in a distributed manner. The DEB arranges the back-off timer of node to guarantee that the polled stations will have different back-off. When the back-off timer expires, the station can be polled without colliding with others. However, DEB is only functioning in CFP whereas HCCA is supposed to work in both CFP and CP, for this reason one of the significant merits of HCCA will be untapped. Moreover, there is no clear consideration of the readiness of the station, STA with no data ready to send will be given a TXOP which, in turn, be wasted.

Non-Polling based HCCA (NPHCCA) is presented in [31] to provide an enhancement over HCCA mechanism. Since the VBR traffic exhibits variability in packet generation presented analysis and guidelines, the recommendations reckon on a number of assumptions that are: the traffic is Constant Bit Rate (CBR) and each QSTA has only one TS calculated based on the Equation (3) which cannot be suitable for supporting the transmission of multimedia applications with variable profile.

**Deterministic Back-off (DEB) method for HCCA** is an enhancement of HCCA which performs virtual polling through sensing the carrier of the wireless channel [78]. This technique highlighted the issue of the collision incurred due to polling the nodes in the overlapping area of two adjacent BSSs at the same time. This actually occurs due to the lack of coordination in HCCA between the adjacent APs. Consider the nodes 5, 6, 7, 8 in the overlapping area illustrated in Fig. 8. Since AP A cannot hear AP B, therefore the collision occurs between the nodes in the overlapped area. DEB uses a similar idea of sensing the carrier of EDCA since it manifests high robustness and flexibility controlling the medium at the overlapping BSSs. A virtual polling has been achieved in a distributed manner. The DEB arranges the back-off timer of station to guarantee that the polled stations will have different back-off. When the back-off timer expires, the station can be polled without colliding with others. However, DEB is only functioning in CFP whereas HCCA is supposed to work in both CFP and CP, for this reason one of the significant merits of HCCA will be untapped. Moreover, there is no clear consideration of the readiness of the station, STA with no data ready to send will be given a TXOP which, in turn, be wasted.

The mechanism of the NPHCCA is carried out throughout a sequence of messages exchanging. First, a station with data will send a transmission frame request to the AP in order to update it about its transmission queue status, including information such as required Priority, Queue status, etc. A station only sends this frame after it receives the beacon message from the AP and senses whether the medium is idle for SIFS. Accordingly, the AP maintains this information in its scheduling table. Finally, the AP determines a transmission sequence and notifies stations to transmit data according to this transmission sequence broadcast in the beacon messages. Fig. 9 demonstrates The components of the NPHCCA. Although, NPHCCA has shown improvement in the transmission delay when the network is light-loaded, the performance was similar to that of HCCA when the network is heavy-loaded. Besides, the messaging exchange of the beacon and transmission request frames added extra overhead to the network, especially when the number of the nodes increases.

**F-Poll** In Feasible Polling Scheme (F-Poll) [79], the application layer gives the accurate arrival-time of the upcoming data frame over the uplink connection to the MAC layer, where this approach is known as a cross-layering approach. F-Poll is suitable for both type II and III of video types categorized in Subsection III-A, where the exact information of the next inter-arrival time is sent to the QAP in order to enhance the scheduling of the TSs. In order to avoid polling a station that have no ready data to transmit, a decision is made of whether to poll the relevant station in the upcoming SI or not directly after receiving a data frame. As a result, the packet access delay is minimized and a great amount of unused TXOP duration is conserved which efficiently enhances the channel utilization. Fig. 10 elaborates the F-Poll Mechanism.

**AMTXOP** [26] like D-TXOP [80], the Adaptive Multipolling TXOP Scheme (AMTXOP) calculates the TXOP for a certain data stream based on the actual frame size. Since the polling messages can increase the overhead among all QSTAs, the BSS broadcasts one multi-polling time, the station will not always have pending data to transmit, thus, it will waste time for the AP to send polling messages to the stations that have no data to transmit. For this reason, the proposed solution modifies the HCCA scheme in such way it allows stations that have pending frames to report their readiness status to AP through exchanging messages. Then, the AP schedules the only ready stations in appropriate transmission sequence.
message to the QSTAs in a single SI instead of sending one polling message for each. This approach minimizes the polling overhead and also reduces the packet delay which significantly improves the bandwidth utilization. Due to this integration, the AMTXOP outperforms both HCCA and its ancestor, D-TXOP, in terms of channel utilization and packet delay.

B. TXOP ALLOCATION ENHANCEMENTS

Usually, if a QSTA’s buffer queue is empty during its TXOP because of a non-uniform data flow from the upper layer, the media will be unutilized for the whole TXOP of the station. However, according to the 802.11e standard, the QSTA should send a QoS-NULL frame to the QAP to enforce it to start polling other sessions immediately [71]. On the other hand, if the allocated TXOP is not enough to send the backlogged packets, these data will be served in the next SI causing more delay and might impair the designated QoS requirements [38]. Several techniques have been presented in the literature to address the limitation of the TXOP assignment mechanism of IEEE802.11e, we overview here some representative approaches.

Scheduling Based on Estimated Transmission Times-Earliest Due Date (SETT-EDD) [23] has proposed a novel scheduling technique for the so-called IEEE802.11e HCF. A simple mechanism similar to the CAP timer has been employed to limit the polling-based transmission in HCF which so-called TXOP timer. This TXOP timer increases at a constant rate equal to the proportion of that TXOP duration to the minimum service interval \((T\ D/mSTI)\), which reflects the fraction of time consumed by the station in polled TXOPs. The maximum value of this timer is equal to the maximum TXOP duration \((MTD)\). The consumed time by a station in a polled TXOP is subtracted from the TXOP timer by the end of the TXOP. Thus, the station can be polled only if the TXOP timer value is greater than or equal to the minimum TXOP duration \((mTD)\), which guarantees the transmission of at least one data frame at the minimum PHY rate.

Since the TXOP is allocated in SETT-EDD based on earliest deadlines, the transmission delay and data loss have been reduced. That is why SETT-EDD shows flexibility and considered a representative dynamic scheduler, as well as it provides compatibility to the link adaptation implemented in the commercial WLANs. However, it still lacks an efficient technique to be able to calculate the accurate required TXOP for every QSTA transmission instead of estimating TXOP based on the average data rate of each TS and the packet time interval between two consecutive transmissions.

Adaptive Resource Reservation Over WLANs (ARROW) is another algorithm where the TXOP assignment is calculated dynamically based on the queued data size of the QSTAs [81], [82]. In ARROW, the SETT-EDD [23] has been extended, where the available bandwidth is allocated based on the existing amount of data which is ready for transmission in every STA. In contrast to SETT-EDD, which allocates the channel bandwidth based on the expected arriving data in every STA. In this mechanism, QSTA advertises the size of the total queued packets waiting for transmission with every poll. This information is piggybacked with the data frame prior the sending back to QAP. So, the next TXOP allocation for any particular stream will be calculated based on the advertised queue size. In this algorithm, the channel is allocated based on the exact transmission requirements for each QSTA, which is expressed by the Queue Size (QS) field indicated during the previous TXOP. By doing so, the TXOP is assigned to meet the transmission requirement at the time when the previous TXOP assignment is made and consequently the data buffered in the QSTA is taken into account at any SI leading to efficient adaptation of bandwidth allocation to actual requirements. Specifically, as illustrated in Fig. 11, data arrive during \([t_i(x), t_i(x + 1)]\) can only be transmitted after the elapsing of \([t_i(x + 2)]\), which results in a delay of packets for at least one SI.

Enhanced Earliest Due Date (EDD) QoS scheduler: presented by [83] and it is an EDD-based algorithm mainly aims at addressing the above-mentioned weakness of the ARROW scheduler. Similar to ARROW scheduler, the Enhanced EDD also uses the queue length information like ARROW. However, the Enhanced EDD estimates the number of arriving packets immediately after the end of the previous transmission, as shown in Fig. 12. Thereafter, it calculates just the enough TXOP to clear up the buffer queue by the end of current transmission. To reduce the average delay, when the buffer is not empty after the current transmission completes, the next SI begins earlier, which can be achieved.
by changing the value of mSI and Maximum Service Interval (MSI). The TXOP allocation in Enhanced EDD is calculated for each station \( STA_i \), as the summation TXOP calculated exactly as in ARROW and a duration enough to transmit the packets generated during the current SI as below

\[
TXOP_i = TXOP^i_{avg} + TDR_i
\]

where \( TXOP_{avg} \) is calculated as follows and \( N_{curSI} \) is the expected number of packets generated from time \( t_{pre} \) until \( t' \).

**Dynamic TXOP HCCA**

**Dynamic TXOP HCCA (DTH)** involves a bandwidth reclaiming mechanism into a centralized HCCA scheduler in order to improve the transmission capacity and to provide additional resources to VBR TSs [34]. The main concept of DTH is to prevent wasting the underutilized portion of transmission time in order to allocate it to the next polled station that needs longer transmission period. This approach relies on the unused amount of the TXOP from the previous poll time of a QSTA, as follows

\[
TXOP_i = \begin{cases} 
TXOP^AC(i) & \text{if } T_{sare} \equiv 0 \\
T_{est(i)} + T_{sare} & \text{if } T_{sare} > 0 
\end{cases}
\]

If there is no surplus TXOP duration from previous poll time, which implies that the station exhausts the whole TXOP duration. The next TXOP duration will be the same as the one calculated in Equation (4). Otherwise, it will be calculated as the summation of the unused TXOP duration and the estimated transmission time, computed through the Simple Moving Average (SMA) of the effectively utilized duration in the previous polling intervals. Simulation results show that this approach can improve the performance, especially in terms of transmission queue size, data loss and delay, and the approach can absorb and follow the variation of VBR. Additionally, another analytical study confirms that the DTH approach has no effect on the policy of the centralized scheduler. However, the estimation of transmission time using Moving Average needs more investigation as the VBR traffic tends to high variability during the time, thus it might be not efficient to find the best setting of the mobile sampling windows.

**The Dynamic TXOP (D-TXOP) scheduling algorithm** [80] analyzes the video of the prerecorded streams before the call setup, which has been previously highlighted in [84]. The D-TXOP is suitable for transmitting type (I) of VBR video source categorized in Subsection III-A, which shows variability in packet size. Indeed, this approach assigns the TXOP for a stream based on the real frame size rather than the estimated average of frame size. It uses the unused QS field of IEEE802.11e MAC header to send the actual size of the upcoming frame to the HC. Thus, the wasted TXOPs have been minimized by this approach, which reflects lower delay compared to the previous solutions. Moreover, the EDCA benefits from the surplus TXOP duration from unused TXOP of the preceding STAs as illustrated in Fig. 13.

**C. HCCA ADMISSION CONTROL ENHANCEMENTS**

The main purpose of HCF admission control is to administer policy or regulate the available bandwidth resources which is used by the HC. The admission control is used to limit the amount of traffic admitted under a certain service category in order to guarantee the highest possible QoS level, while maximizing the utilization of the medium resources. Fig. 14 depicts a common frame format for carrying TSPEC parameters. Since the admission control relies on a fixed TSPEC element, it cannot efficiently cope with the high variability of VBR streams. To solve this problem, numerous enhancements and optimizations have been proposed to tackle this deficiency in the legacy ACU mechanism.

**Rate-Variance-envelop-based Admission Control (RVAC) mechanism** uses the Dual Token Bucket (DTB) shaper to guarantee the desired QoS specification [85]. The authors of these two references [86], [87] have derived the delay probability based on the aggregate traffic statistics rather than considering each flow individually to accept a new flow for admission [88], [89]. The effective TXOP duration of a recently arrived VBR stream can be inversely derived based on a given packet loss rate as in Equation (5).

Indeed, the RVAC takes the multiplexing gain of VBR traffic into account unlike the guarantee-rate-based scheme. More specifically, if the arrival rate of data streams extends over a wide range, where the RVAC can fully utilize the multiplexing gain among the VBR streams, the performance
gain can be noticeable. Additionally, the RVAC considers both uplink and downlink traffic streams. Simulation results have shown that the admission capacity of the RVAC approach is more than the double of its equivalent in the GRAC approach. In addition, the RVAC scheme will not violate the 0.1 second delay requirement as long as the starting time of the streams are spread over a wide time range of not less than 2 seconds. However, the performance of the RVAC in the wireless channel errors environment is not studied.

**Equal-SP** [90] has been designed of HCCA scheduling, in which the spacing of a particular stream is determined as the period of time between two consecutively scheduled streams. It has been called equal-SP scheduling because a particular stream will always get an equal spacing for its scheduling slices in the schedule. Indeed, the equal-SP scheduling relies on the well-known Rate Monotonic (RM) algorithm to achieve the QoS requirements. Despite that the equal-SP approach is similar to the SETT-EDD in terms of the general scheduling concept, however, the former assigns equal spacing for each particular stream, which is proven to violate the delay requirement in some cases.

In the example as shown in Fig. 15, the scheduler assigned 25 ms, 50 ms, and 150 ms time spacing for the flows 1, 2, and 3, respectively, which makes $T_{11} = T_{12} = T_1 = 25$ ms, $T_{22}=50$ ms, and $T_3 = 150$ ms. The equal-SP approach is easy to be implemented and it can guarantee the delay requirements and efficiently utilize the bandwidth while maintaining the compatibility to the standard since it uses the same TSPEC parameters. However, the equal-SP approach encounters the same issues faced by the standard; if a newly admitted stream has a smaller delay bound than the current $T_1$, the current $T_1$ will be set to less than or equal to the new delay bound. Therefore, the TXOP durations for the previously admitted flows are required to be recalculated with the $T_i$s. Additionally, the scheduler needs to reassign index for the admitted flows in order to maintain the condition of $T_1 \leq T_2 \leq \cdots \leq T_n$.

**PHCCA**, as described in Fig. 16, is a priority based QoS and admission control used for queue management mechanism [91]. In this approach, a mechanism for borrowing and returning bandwidth among queues has been studied. The higher priority queue, called class, has permission to borrow bandwidth from lower priority queues with the awareness of starvation protection for each priority queue. PHCCA modifies the HCCA by classifying the traffic into 3 classes, which has not been divided by the standard. Class 1 is the highest priority class suitable for voice and conference traffic implementation. Class 2 is the second highest priority class suitable for broadcast video traffic. Class 3 is the lowest priority traffic suitable for FTP and HTTP traffic.

![FIGURE 16. PHCCA admission control mechanism](image)

Experimental results reveal that the proposed PHCCA can accept more requests for Class 1 (70% better) compared to the regular HCCA. Meanwhile, the second highest priority (Class 2 in this case) is still served quite similar to the regular HCCA. Despite that the PHCCA significantly outperforms the regular HCCA, it is still not able to guarantee the required QoS for every flow; since it assumes that flows of similar types (e.g. VoIP) will have exact QoS requirements. Besides, the performance of NPHCCA could merely achieve the performance level of Best-Effort for VoIP and CBR-video applications. Moreover, parameters and environment factors for bandwidth borrowing mechanism, such as distance from the access point or mobility, should be investigated.

**AF-HCCA** [92] enhances the experienced delay of video traffic by utilizing the surplus bandwidth and mitigates the over-polling issue. It computes the TXOP for a traffic stream based on the knowledge about the actual upcoming frame size instead of assigning TXOP according to the mean characteristics of the traffic which is unable to reflect the actual traffic. This scheduler exploits the queue size field of IEEE802.11e MAC header to transfer this information to the HC.

In AF-HCCA, the QSTAs will be prevented from receiving unnecessary large TXOP which produces a remarkable increase in the packet delay. Furthermore, the surplus time of the wireless channel conserved by reducing the number of poll frames throughout the feedback is another benefit of this research. The integrated scheme of AF-HCCA shows superior performance compared to IEEE802.11e HCCA, Enhanced EDD [83] and F-Poll [79] schedulers in terms of delay and channel utilization without affecting the system throughput. However, preserved TXOP time is not efficiently utilized to enhance the flow capacity.

**Feedback-based Admission Control Unit (FACU)** [93] aims at maximizing the utilization of the surplus bandwidth
which has never been tested in previous schemes. The FACU exploits piggybacked information containing the size of the subsequent video frames to increase the number of admitted flows.

The FACU introduces an enhancement to admission control mechanism of Adaptive-TXOP. Analytical results reveal the efficiency of FACU over the examined schemes. The results show that the conserved channel bandwidth of Adaptive-TXOP can be utilized to increase the number of admitted QoS flows and enhance the overall QoS provisioning in IEEE802.11e WLANs.

VI. HCCA SCHEDULING APPROACHES COMPARISON

Table 4 presents a summary and comparison for the HCCA enhancements in IEEE802.11e along with their targeted features classified based on the place of the enhancement. The solution column briefly describes the used technique. The complexity of an approach can be high, medium or simple estimated based on Likert-type rating scale. The complexity here represents the volume of the operations of that particular approach. The method that involves more operations is considered high-complex and vice versa. The main targeted traffic of the enhancement is stated. The targeted flow direction, which is considered by the approach, is also presented.

VII. OPEN RESEARCH ISSUES

Although the existing approaches provide several possible solutions to alleviate the deficiency of scheduling for VBR multimedia traffic in IEEE802.11e WLANs, many issues have been thoroughly discussed in the literature review section, which are potential research topics. This section highlights the most important issues in order to determine the directions for potential future research. One of the problems with HCCA is the coexistence. Several mechanisms claim to be able to coordinate different HCs that operate on the same frequency channel. Since HCCA’s QoS guarantee depends on the exclusive usage of the frequency channel, multiple HCCA can hardly coexist. On the other hand, additional delay may occur by the polling STAs with scalable video that exhibits constant quality yet introduce high variation in the traffic profile. From the cross-layer perspective and to the best of our knowledge, there is no proactive scheme that provides a good solution to the adequate interaction between the fluctuation of the uplink VBR traffic profile at the application layer and the flexible scheduling policy at MAC layer which exhibit low-complexity design. In summary, some issues are needed to be considered to provide optimal enhancement for the transmission of VBR traffic in IEEE802.11e WLANs. We believe the following suggestions are desirable for designing a good HCCA scheme in IEEE802.11e wireless networks.

- Providing efficient estimation of the bandwidth in order to achieve high connection throughput.
- Designing a scheme coupled with link adaptation mechanisms in order to provide efficient adaptation to dynamic network behavior.

FIGURE 17. Number of publication of the investigated research areas

- Exploiting the distributed feature of EDCA in designing a hybrid HCCA scheme in order to yield high integration and interoperability without jeopardizing the system simplicity.
- Achieving low algorithm complexity.

VIII. RESEARCH TREND ON QoS SUPPORT IN IEEE802.11E

Many researches have been conducted in the Literature since the first advent of the HCCA protocol draft in IEEE802.11e standard [94]. These researches can be classified into five research areas as in Table 5 aims at demonstrating the trend of the research since the first presence of the HCF functions till 2009 and from 2010 to present. The collection includes over 89 journal and conference papers. These scientific documents have been collected using IEEE Explorer Digital Library, Springer Link, ScienceDirect and Google Scholar. One can notice that the polling and TXOP allocation mechanisms have greatly received the researchers’ attention since the evolution of the HCCA till now, while admission control mechanisms have less interest. It is worth noting that the design of the hybrid EDCA-HCCA scheme has scarcely studied. The HCCA performance and mathematical analysis have been fairly covered. Yet, only few efforts have focused on designing a comprehensive analytical model for HCCA protocol. The aggregated number of papers published in three periods, namely 2004 to 2007; 2008 to 2011 and 2012 to present are depicted in Fig. 17. The figure shows that the polling and TXOP mechanism have received a great amount of attention compared to ACU and hybrid scheme. On the other hand, recently there has been a few analysis of HCCA protocol, in contrast to the period from 2004 up to 2011.

IX. FUTURE DIRECTIONS

Although all proposed schemes in their current states improve the transmission of prerecorded video, there still some issues need to be addressed and investigated. Below,
we highlight some future works that need to be carried out for further enhancement to:

- Enhance the HCCA to cope with more complicated wireless scenarios, where the hidden node problem exists and QSTAs communicate using RTS/CTS mechanism with MAC level fragmentation and multi-rate support enabled.
- Study the scalable HCCA MAC for video over wireless mesh networks that are also scalable to a wider range of MAC settings to support more robust time-bounded media applications.
- Design a new admission control algorithm to utilize the excess bandwidth and to manage the available resources among the HCCA, HCF and EDCA in order to maximize the number of served streams or applications in the network.
- Examine the performance of HCCA with the presence of collision occurred in the overlapping area when polling stations among two neighboring BSSs simultaneously.

**X. CONCLUSION**

IEEE 802.11e is aimed at providing stringent QoS support to multimedia applications such as video streaming. The controlled based function of IEEE802.11e standard, HCCA scheduler, consider a fixed TSPEC for scheduling the traffic while in fact the VBR traffic tends to change their characteristics such as data rate and packet size over the time. The inability of the IEEE802.11e MAC protocol to accommodate to the high fluctuation of VBR video profile motivates many researches to be conducted. Several enhancements have been made to alleviate these shortcomings. These enhancements tend to address particular issues or applications by improving, in most cases, one of the HCF functions. In general, designing a robust HCF solution that provides an integrated solution for all traffic classes is still a challenging task for future research.

**REFERENCES**


This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2872770, IEEE Access.


Mohamed Othman (M’06–SM’18) received his Ph.D from the Universiti Kebangsaan Malaysia (UKM) with distinction (Best Ph.D Thesis in 2000 awarded by Sime Darby Malaysia and Malaysian Mathematical Science Society). Now, he is a Professor in the Department of Communication Technology and Networks, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM). He is also an associate researcher at the Lab of Computational Science and Mathematical Physics, Institute of Mathematical Research (INSPEM), UPM. He published more than 160 International journals and 230 proceeding papers. His main research interests are in the fields of high-speed network, parallel and distributed algorithms, software defined networking, network design and management, wireless network (MPDU- and MSDU-Frame aggregation, MAC layer, resource management, and traffic monitoring) and scientific telegraph equation and modelling.