

Research Article

Elastic QoS Scheduling with Step-by-Step Propagation in IEEE 802.11e Networks with Multimedia Traffic

Anna Lina Ruscelli , Gabriele Cecchetti , and Piero Castoldi

TeCIP Institute, Scuola Superiore Sant'Anna, Via G. Moruzzi 1, Pisa, Italy

Correspondence should be addressed to Anna Lina Ruscelli; a.ruscelli@sssup.it

Received 31 December 2018; Revised 7 March 2019; Accepted 18 March 2019; Published 18 April 2019

Academic Editor: Giovanni Stea

Copyright © 2019 Anna Lina Ruscelli et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The spreading diffusion of wireless devices and the crowded coexistence of multimedia applications greedy of bandwidth and with strict requirements stress the service provisioning offered by wireless technologies. WiFi is a reference for wireless connectivity and it requires a continuous evolution of its mechanism in order to follow increasingly demanding service needs. In particular, despite the evolution of physical layer, some critical contexts, such as industrial networks, telemedicine, telerehabilitation, and virtual training, require further refined improvements in order to ensure the respect of strict real-time service requirements. In this paper an in-depth analysis of *Dynamic TXOP HCCA* (DTH) MAC enhanced centralized scheduler is illustrated and it is further refined introducing a new improvement, *DTH with threshold*. DTH and DTH with threshold can be integrated with preexisting centralized schedulers in order to improve their performances, without any overprovisioning that can negatively impact on the admission control feasibility test. Indeed, without modifying the centralized scheduler policy, they combine together the concepts of reclaiming transmission time and statistical estimation of the traffic profile in order to provide, at each polling, an instantaneous transmission time tailored to the variable traffic requirements, increasing, when necessary, the service data rate. These mechanisms can coexist with advanced physical layer-based solutions, providing the required service differentiation. Experimental results and theoretical analysis, based on elastic scheduler theory, show that they are effective especially in the case of Variable Bit Rate traffic streams in terms of transmission queues length, packets loss, delay, and throughput.

1. Introduction and State of the Art

IEEE 802.11e *Hybrid Coordination Channel Access* (HCCA) networks [1] are stimulated to improve their provided *Quality of Service* (QoS) by the pervasive diffusion of multimedia applications over *Wireless Local Area Networks* (WLANs) [2]. On-line gaming, multimedia streaming, *Voice Over IP* (VoIP), *High Definition TV* (HDTV), mobile applications, etc., expect a service differentiation suitable to deliver these types of traffic respecting their requests and avoiding data degradation. Furthermore, the increasing numbers of users heighten the levels of required service stressing the provisioning mechanism.

Moreover, some particular domains, e.g., industrial networks [3] or telerehabilitation, telemedicine, and virtual training [4], require a particular attention in providing and ensuring the required service, showing strict service expectation about delay, packets drop, and real-time guarantees.

For instance, in the context of industrial networks, railway communications, both between trackside devices [5] than on-board control networks, have stringent requirements that cannot be unattended since they affect the safety of the system [6]. In general, industrial networks used the basic *Distributed Coordination Function* (DCF) of IEEE 802.11 but the real-time requirements pushed the investigation of alternative and advanced mechanism [7], such as that of HCCA of IEEE 802.11e suitable to ensure the respect of such strict service requirements [8, 9]. Indeed, the need of real-time guarantees forces the support of service differentiation and deterministic transmission that contention-based solutions or only physical layer schemes cannot ensure.

This implies that, despite the evolution of physical layer, where the improvement of offered service is obtained making the bandwidth and frequency assignment more efficient optimizing the spectrum assignment (see, for instance, IEEE802.11ax), in the last years a new attention is

rising also at the *Medium Access Control* (MAC) level; see [3, 10–14].

Indeed, for instance, IEEE 802.11ac [15] and IEEE802.11s [16] are based on physical techniques, such as *Multiple Input Multiple Output* (MIMO), that improve the spectrum assignment but show fluctuations that jeopardize the provided QoS. Physical tools aim to provide an overprovisioning of transmission capacity that, however, in some particular cases, e.g., the mentioned industrial networks or telemedicine, is not sufficient implying strict temporal requirements. Indeed, making the spectrum more crowded is not sufficient in the mentioned domains where the service needs are more strict with respect to multimedia applications and service differentiation is required.

Thus, in such particular critical domains, physical improvements require to be supported also by MAC-level tools suitable to improve the provided QoS making the system more deterministic [3]. Furthermore, these improvements are useful also in the case of dense networks, where the high number of users jeopardizes the provided QoS and only physical solutions are not sufficient.

At the best of our knowledge, this justifies the availability of on-the-shelf equipment based on IEEE802.11e, e.g., Ralink RT5370/RT5350/RT2860 [17], that supports the HCCA mechanism, that is tailored to these particular sectors of the market, and that demonstrates the usefulness and practical usability of IEEE 802.11e.

Finally, HCCA mechanisms can coexist with the EDCA ones, based on contention schemes, providing further differentiation mechanisms [18] and allowing the coexistence of HCCA with installed contention-based solutions.

Focusing on HCCA, despite the fact that it supports until eight different and negotiable types of traffic profile, the eight *Traffic Streams* (TSs), being designed for service differentiation, multimedia applications are more exigent requiring a dynamic scheduling of transmission time and polling period, which are, respectively, related to the two main protocol parameters, *Transmission Opportunity* (TXOP) (transmission duration) and *Service Interval* (SI) (polling period). Anyway, the standard IEEE 802.11e reference scheduler manages these parameters as fixed and computed taking into account mean value service expectations agreed during the admission control. As a consequence the reference scheduler results tailored for *Constant Bit Rate* (CBR) TSs, whereas it is underperforming in the case of *Variable Bit Rate* (VBR) TSs; see [19] and the cited works herein. However, when data rate rapidly changes its profile, providing a variable transmission time tailored for multimedia traffic streams is a challenging task also for many of the existing HCCA scheduling algorithms [19–25] introduced to improve the QoS, including those with variable TXOP.

Many algorithms have been proposed in order to improve the QoS support as provided by the reference HCCA scheduler, and different techniques have been investigated to compute variable TXOP. For instance, some of them use transmission queues models to reduce the experienced delay, such as the reference *Fair HCF* (FHCF) [23] or [26] that considering the Service Data Units ready to be transmitted. In particular, FHCF, which will be used in the performance

evaluation, increases the fairness of CBR and VBR TSs and reduces the delay assigning variable TXOPs by means of the estimation of the uplink TSs queues length. It uses a *QoS Access Point* (QAP) scheduler to estimate the queue length for each QoS *STATION* (QSTA) at the beginning of the next *Service Interval* (SI), comparing the obtained estimation with the ideal queue length corrected with its expected value. Then, it computes the additional required time for each TS of a QSTA considering the number of packets in the queue. This computation is refined by the node scheduler that redistributes the additional time among its TSs. Feedback-based algorithms try to correct the fixed TXOP as provided by the reference, such as the classical *Feedback Based Dynamic Scheduler* (FBDS) [24], or the more recent [27, 28]. Information about the size of video frames is used by [29] or the arrival time of the subsequent video frame of the uplink traffic [30] to compute dynamic TXOPs. Real-time theory is borrowed to manage the polling list considering temporal requirements in [8], or in *Wireless Capacity-Based Scheduler* (WCBS) [21], that is based on *Earliest Deadline First* (EDF) [31] algorithm and on a capacity recharging mechanism that implies a postponing of the deadline and a remix of the polling list order, and in the classical Scheduling *Estimated Transmission Time-Earliest Due Date* (SETT-EDD) [32], and *Real-Time HCCA* (RTH) scheduler [20] that regulates the polling list using EDF. In particular, RTH, which will be compared through simulation with DTH and DTH with threshold, is a periodic scheduler, based on EDF algorithm plus *Stack Resource Policy* (SRP) algorithm [33], that introduces the concept of nonpreemptability of frame transmissions that are considered as critical sections. It is designed to provide real-time support in HCCA ensuring the traffic streams a fixed amount of capacity during a fixed period. Its scheduler activity is composed by an offline task at stream lifetime timescale, which performs the more complex activity (admission control and timetable computation), and an online task (reading next polling time and transmission duration) at the frame transmission timescale. Cross-layer optimization is used to provide a dynamic assignment of TXOP for video traffic [11].

Some directions of investigations approached also the joint use of both HCCA and *Enhanced Distributed Channel Access* (EDCA) mechanisms allowed by the IEEE 802.11e *HCCA-EDCA Mixed Mode* (HEMM): for instance, [34] is focused on optimizing the relative duration of HCCA and EDCA, whereas *Overboost* [18] exploits EDCA capacity to integrate the HCCA one.

Also bandwidth reclaiming techniques have been applied to increase the assigned transmission time, such as in [35], where a Proportional-Integral-Derivative controller is used to recover the resources using queue level information. *Unused Time Shifting Scheduler* [36] proposes a greedy reclaiming algorithm to recover the unused time from the previous polled stations: the recovery capacity is simply and greedily added to the TXOP of the next polled station without any adaptation to the effective station needs. This mechanism is refined in [37] where *Immediate Dynamic TXOP* is presented. Here, in order to make more efficient the transmission time assignment, the transmission time

used during the previous polling is evaluated. However, this scheme is not suitable to perceive the effective traffic behavior, especially in the case of data rate changes in intensity and trend and, thus, a variable *TXOP* is provided but not strictly tailored to the station needs as expected to make the capacity assignment more efficient. However, if resource overprovisioning is not considered as a shortcut, making dynamic the transmission interval is a twofold problem: on one side it is required to dynamically adapt *TXOP* to the traffic expectations; on the other hand it is compulsory to respect the resource threshold assigned to the stations and preserved by the admission control feasibility test. Solving this challenge suggests refined methods, suitable to answer both of these questions, sometimes choosing the best tradeoff, and without jeopardizing the levels of service accorded to admitted QoS Stations QSTAs.

An approach alternative to proposing a further new scheduler is taking advantage of the existing schedulers and intervening on their policy by integrating a scheduling module focused on improving the offered QoS. *Dynamic TXOP HCCA* (DTH) [38] acts in this way, refining the behavior of a centralized HCCA scheduler with the ability of providing a dynamic *TXOP*. In this scope, an advanced version of DTH, *Dynamic TXOP HCCA with threshold*, is presented. It provides, as an improvement of DTH, a “step forward” enriching, as its predecessor, the preexistent schedulers with additional functionalities suitable to better satisfy QoS expectations. Both of them are focused on *TXOP* computation, making that not only dynamic but also adaptable to the applications needs. This makes possible to assign different values of transmission duration *TXOP*, providing a service more tailored to the considered station. In particular, they mix together the concepts of capacity reclaiming, overprovisioning with respect to the admission control assignment, and traffic estimation in order to provide, at each polling, a transmission duration tailored to the station changing requirements. First of all, the capacity reclaiming is the mechanism used to feed the overprovisioning one without violating the resource threshold set by the admission control; then the recovered resources are assigned taking into account the station needs that can be varying with respect to the declared ones during the admission phase. This is achieved through a statistical estimation of the traffic profile based on the use of time series forecasting suitable to highlight the deviations of the TS parameters with respect to their mean values, typical of VBR traffic. These actions are done by means of a scheduling module suitable to be integrated in a preexistent centralized scheduler, without modifying its scheduling engine.

In this paper both the mentioned algorithms, DTH and DTH with threshold, are deeply analyzed theoretically and by simulations. In particular, the analysis is focused on the real-time properties suitable for the management of VBR traffic and, specially, on feasibility of the algorithm and its features in the scheduling of multimedia application. To that end, a new analysis model is introduced, able to represent the obtained modifications of the transmission time. The analysis is derived by the elastic task model, originally used in the context of processor systems for the management

of overload due to soft real-time tasks. In that case the rate of the processor that executes the tasks is changed modifying the tasks frequency. In the case of multimedia applications over wireless networks, the approach is overturned accepting varying traffic rate as that of the different TSs and modifying their transmission time accordingly. This produces as effect a practical modification of the serving rate, whereas the reference scheduler rate remains unchanged. The mentioned analysis and the simulations shows as both DTH and its improvement are suitable to ameliorate the behavior of a centralized scheduler when multimedia applications are involved, positively impacting on transmission queues length and, consequently, traffic delay and jitter.

In particular, in Section 2 DTH is described deeply explaining the motivation of its features and its scheduling rules. In Section 3 the algorithm is analyzed from the point of view of the feasibility by means of the introduced model and its improvement is introduced. Finally, in Section 4, simulation results are described and in Section 5 conclusions are drawn.

2. DTH Scheduling Algorithm

In this section DTH algorithm will be deeply described, starting from the motivation of its use and, then, going deep inside its behavior.

2.1. Motivations. The assignment of resources and the related computation of the HCCA protocol parameters suitable to manage the access to the medium (the shared resource) is performed during the admission control considering the mean value of the transmission parameters. This approach is acceptable in the case of multimedia applications that have soft real-time requirements; i.e., a deadline miss causes only a deterioration of provided QoS and *Quality of Experience* (QoE) without critical effects. Instead, considering max or min values in terms of the worst case conditions ensures the respect of all requirements, including the most stringent ones. However, the consequent overprovisioning can imply the not complete depletion of the accorded resources since, in the case of VBR streams, the data rate R can vary in the interval $[R_{min}, R_{max}]$ in dependency of the traffic profile and, when $R < R_{max}$, assigned resources are not exhausted. Thus, from an efficiency point of view, choosing mean values of protocol parameters is a quite satisfactory but conservative choice, since it merely allows to meet mean requirements, whereas max values allow to satisfy all requirements with the drawback of a waste of resources. Moreover, the choice of mean values is a tradeoff since it does not allow to work on the border zone, where the data rate is less or greater than R_{mean} , in order to refine the resource assignment and made that more performing. Indeed, trying to tune the scheduling in these areas allows a more precise resource allocation tailored to the traffic profile, avoiding loosening resources when not needed and trying to provide more when necessary. An alternative way to provide a *TXOP* suitable to the traffic is its recomputation at each polling. This choice can be more accurate compared to the “single-shot” computation

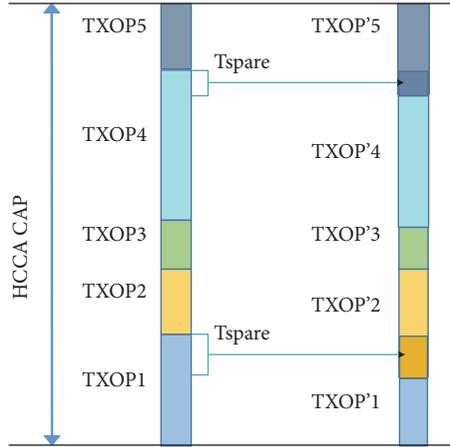


FIGURE 1: Recovery mechanism: DTH does not impact on the admission control.

of $TXOP$ during the admission control phase, but it weights the computation activity.

Starting from these considerations our aim is varying $TXOP$, i.e., providing an instantaneous $TXOP$ taking into account the stations current requirements, but without effects on resource repartition set during the admission control, which remains the reference, and avoiding a computation overload due to a recalculation of $TXOP$ at each polling.

A possible answer to this goal is adding to the accorded transmission time, at each polling, further resources taking into account the current station traffic and the consequent current service needs. An easy way to find missing resources, without perturbing the correlated assignments done by the admission control for each station, is reclaiming unused capacity, already considered in the admission control threshold and facing off their waste. Indeed, differently than in the case of general real-time systems, HCCA standard protocol provides some basic rules that help the recovery mechanism. In particular, when a stream ends its transmission before the expected time, it does not continue occupying the medium, and the unused resources are released and, in absence of reclaiming schemes, are simply lost. As set by the standard, this underutilization of $TXOP$ implies a polling in advance of the consecutive station in the polling list, which gains the control of the medium after a *Short Interframe Space* (SIFS). This mechanism makes unused resources also available for the recovery and for the assignment to the next polled station. Thus the varying network load itself allows, thanks to unexhausted resources freed by the scheduler, to provide an instantaneous overprovisioning useful in the case of overload of the subsequent stations and, as will be shown in the following, without violating the admission control feasibility test. In general, in HCCA, being a *Time Division Multiplexing* (TDM) system, the different tasks are scheduled consecutively; thus, considering the total length of all streams transmission time equal to the maximum global assigned resources, each variation in the resource utilization can impact on other streams, assuming the presence of an admission control threshold. Instead, in the case of DTH, as shown in Figure 1 related to a *Controlled Access Phase*

where HCCA takes action, the reassignment of not-exhausted resources to the stations requiring a longer transmission time does not overcome the admission control upper bound U_{lub} (least upper bound) indeed, since the unused portions, or part of them in dependency of the needs of the subsequent stations, are simply translated, and no more resources with respect to the global assigned amount are required. This approach impacts only on the transmission time and does not involve the polling mechanism; this consideration is valid in general, since the algorithm does not operate on the polling engine and, in particular, since the period is not modified, the polling list is not changed in the case, for instance, when the polling order is set taking into account deadlines. Moreover, the shift of recovered resources to the next stations performed by DTH can imply only a polling advance, coherently to the reference scheduler and as shown in [38].

2.2. DTH Algorithm: How It Works. In the introduction of DTH we assume starting the resource scheduling from the assignment as set by the reference scheduler: mean computation time $C_{mean_i} \equiv TXOP_{AC_i}$, the $TXOP$ assigned during the admission control, is adopted as a “frugal” solution, despite max values parameters that should require $C_{max_i} \equiv TXOP_{max_i}$. Then, since in dependency of the VBR traffic profile resources can be not exhausted, or can be not sufficient, a recovery mechanism is exploited updating the assigned capacity of the next station with the purpose of providing a $TXOP$ nearest to the instantaneous station needs and trying to meet the heavier resource expectations. Furthermore, in DTH the assignment of the new instantaneous $TXOP$ is refined by considering the current station expectations in function of the actual traffic profile in order to allocate the actual required capacity, as it will be explained in the following. This is the added value introduced by DTH with respect to other recovery mechanisms.

However, since, in general, in the case of multimedia applications, the traffic does not have a deterministic behavior, a deterministic descriptive function is missing and the traffic trend has to be derived through mathematical or statistical models. DTH traffic estimation is based on time

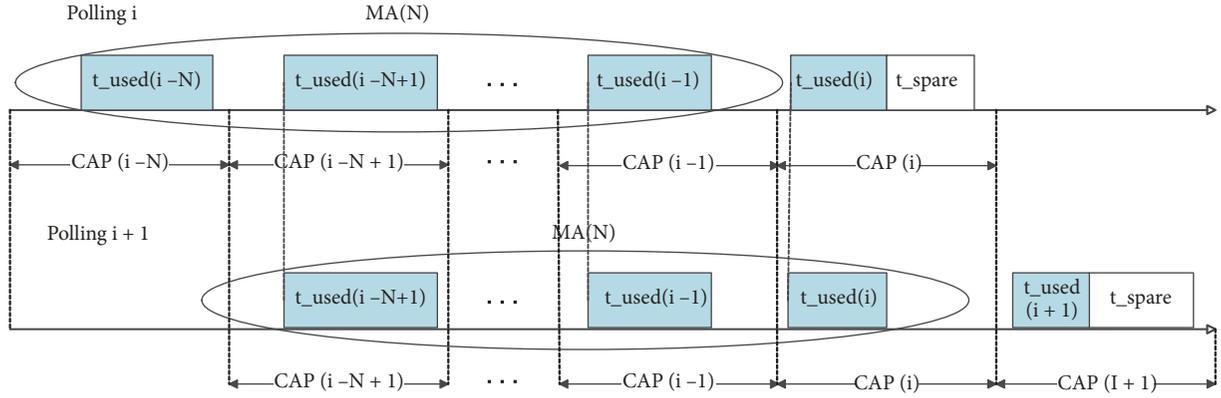


FIGURE 2: Simple Moving Average window mechanism of DTH.

series forecasting, which is generally exploited to analyze stochastic phenomenon. This method uses a series of temporally ordered random variables to represent the samples of a process, to forecast its future behavior, and to derive the future values of the process variables. In this scope in the case of multimedia applications with a stochastic behavior, where the traffic profile is unknown and the transmission time can vary, the actual values of the transmission time during the previous polling phases are used to estimate the transmission length needed for the next polling and this estimation is considered for the assignment of *TXOP*.

Following this approach the transmission time Δ actually used by QSTA_{*j*} is described by a time series as follows:

$$\Delta = \{t_{used_{j,t}}, t \in T\}, \quad (1)$$

where the time t is the state variable, $t_{used_{j,t}}$ is the observation of Δ of QSTA_{*j*} at time t , and T is the discrete *index set*, i.e., the sampling periods. In *time series forecasting* different methods of samples collection can be used. For instance, a *Moving Average (MA)* based on a mobile time window where we find the samples is suitable to evaluate the trend of the investigated phenomenon considering its more recent samples y_i with $i \in T$ and it can be used to forecast the sending interval for the subsequent transmission. Indeed, the study of the traffic behavior during the previous polling phases can suggest the trend of the current transmission time. In particular, a *Moving Average MA(t)* of order $m_1 + m_2$ at time t collects the $m_1 + m_2$ samples around time t and it is expressed as follows:

$$MA(t) = \sum_{i=-m_1}^{m_2} \theta_i \cdot y_{t+i} \quad (2)$$

where m_1 is the number of periods before t , m_2 after t , and θ_i is the weight of the element i . In the case of DTH, which directly uses the information provided by the samples without any further elaboration, $\theta_i = 1/(m_1 + m_2), \forall i$. Moreover, since all samples of a QSTA_{*j*} have the same period, SI_j , the *Simple Moving Average (SMA)* is used. With this choice a *mobile window* flows along the time axis following the traffic variations to collect the samples that are elaborated

through an arithmetic average, as illustrated in Figure 2. Moreover this method helps to diagnose the inversion in the data rate trend. Thus, $t_{est_j}(t)$, the estimated transmission time of QSTA_{*j*} at time t needed for its $(N + 1)^{th}$ polling, is computed by means of the SMA of the transmission intervals used during its previous N polling:

$$t_{est_j}(t) = \frac{\sum_{i=1}^N t_{used_{j,i}}}{N}, \quad (3)$$

where $t_{used_{j,i}}$ is the transmission time of QSTA_{*j*} actually used during its i^{th} polling. DTH chooses the index set T and the length of the sampling window taking into account traffic stream period and type in order to have a sufficient number of samples. For instance, if T is composed of 250 elements in order to have an estimation of the transmission time used during the last 250 pollings and updated by the samples moving window, in the case of a video streaming with a period of 40 s, the window length is equal to 10 s.

Knowing the estimated *TXOP* required for the next polling time, in the case of missing resources, the recovery mechanism comes into play by providing further resources with respect to the assigned ones during the admission control, avoiding assignments recomputation for all stations that could cause unfeasible schedulability or inefficient over-provisioning.

The reclaimed transmission time from the transmission of the preceding station in the polling list T_{spare} is expressed as

$$T_{spare} = t_{end} - t_{stop}, \quad (4)$$

where t_{stop} is the actual finishing transmission time, t_{end} is the ending transmission time when *TXOP*, assigned using the adopted scheduling rules, is completely exhausted and it is equal to $t_{end} = t_p + TXOP$, with t_p being the polling time. This additional capacity is wisely assigned to the stations effectively requiring that.

Combining these two mechanisms, DTH exploits time series to refine the management of recovered resources, since its use is similar to the combination of *trendline* and price trend in finance. As when the price trend, growing, intercepts

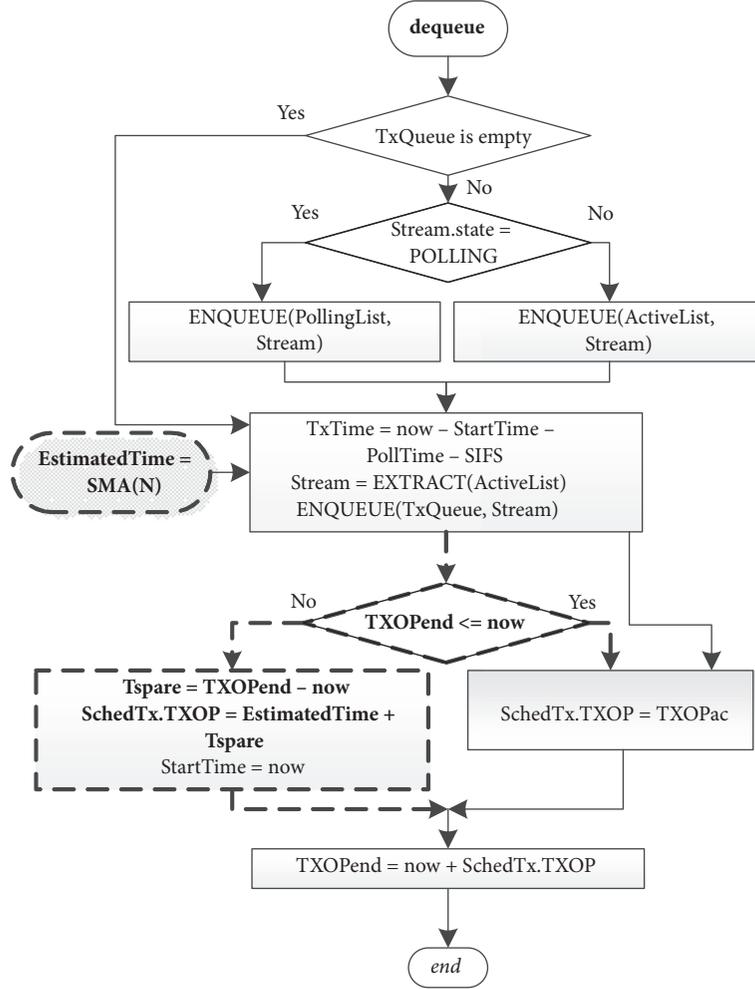


FIGURE 3: DTH dequeue flow diagram.

the trendline and a buying signal is triggered, so when the punctual value of the estimated transmission time at time t exceeds its reference value $TXOP_{AC}$, the recovery mechanism is activated and, if $T_{spare} \neq 0$, it is shifted to the next polled station. Finally, when the DTH recovery mechanism cannot be applied since there are no spare resources available, the centralized scheduler continues assigning $TXOP_{AC}$ for efficiency reasons and in order to avoid saturation effects, as theoretically demonstrated in [38]. As a result the new transmission time $TXOP'_i$ for the next polling of $QSTA_i$ is computed as follows:

$$TXOP'_i = \begin{cases} TXOP_{AC(i)} & \text{if } T_{spare} \equiv 0 \\ t_{est_i} + T_{spare} & \text{if } T_{spare} > 0 \end{cases} \quad (5)$$

DTH integrates the explained mechanism of tuning the resource assignment with the centralized scheduler, without replacing that and taking action only at the polling time, as shown by the flow graph in Figure 3: it illustrates the deque function of a centralized HCCA scheduler with DTH integrated into the scheduling engine, i.e., the function that extracts TS from the waiting queue and sends it. In this figure DTH activities are highlighted in bold, its functional

blocks have dashed contour, and it is shown where DTH inserts its activity and how it cooperates with the centralized scheduler.

The previous considerations introduce a different approach to the recovery scheme performed by other proposed reclaiming algorithms, such as UTSS and IDTH, since in these algorithms none estimation mechanism was adopted. T_{spare} was simply shifted to the next polled station, without verifying if needed or not through a statistical estimation. From the point of view of DTH this approach can be significantly improved integrating the basic idea to assign a more appropriate transmission capacity tuned to the station current expectations.

DTH does not impact on the centralized scheduler, as mentioned before, nor on the admission control policy, as demonstrated in [38] and as will be deepened in the following.

2.3. Future Traffic Statistical Estimation Vs. Transmission Queues Length Values. The future traffic estimation mechanism can be considered an alternative to the use of transmission queues length for the $TXOP$ computation, based on an opposite approach.

Numerous works have been proposed based on the use of the queues length information in order to update the transmission time, for instance, from the first algorithms, such as [24], until the new ones, such as [27, 28]. This solution implies recovery actions that try to reduce the queues length; i.e., correct the not suitable $TXOP$ assignment, for instance, using feedback mechanisms. Indeed, most of the works based on the use of the queues length value recover the misbehavior using some feedback mechanisms; see, for instance, [35]. Indeed, the length of the transmission queues increases when the traffic is changing with respect to its mean value parameters, in particular when the rising of data rate or the presence of bursts of traffic implies that $TXOP_{AC}$ is no more sufficient to dispatch the incoming traffic.

The approach of DTH is different being based on the forecast of the future traffic behavior; thus it is a proactive method that tries to make the $TXOP$ assignment more precise taking into account the future traffic trend; before that the transmission queues length increases. Thus the approach is the opposite in the sense that it does not take action after the undesired behavior but, before, trying to prevent that.

The only advantage of the use of queues length value with DTH could be in the case of empty queues, when the polling of the station could be avoided saving $2SIFS$ time. Otherwise, in the case of queues length related to a transmission duration less than $TXOP$, the use of queues length can help in computing a transmission duration more tailored to the effective traffic to dispatch, but it shows its effects only in the polling phase subsequent to the one when the queue has been shortened, whereas the adoption of a future traffic estimation scheme, also in this situation, can act in advance. Furthermore, in this case the polling handshake cannot be, obviously avoided and the MAC scheme itself regulates the use of the transmission time freeing the medium when the transmission is finished, also before the end of the assigned $TXOP$.

Thus, in general, it is eventually preferable to add the use of queues length to DTH instead of substituting the future traffic estimation scheme, in order to take advantage from both of these mechanisms.

3. Scheduler Analysis

In this section the DTH algorithm is theoretically analyzed. First of all some basic scheduling properties are considered, and then the scheduling activity is examined through a model derived from the elastic task analysis. As it will be shown, this approach overturns the classical elastic task model used in processor systems, where the scheduling of soft real-time tasks aims at dynamically adapting their period to variable QoS requirements. This analysis deepens the investigation of the impact of DTH on resource management. Finally, some considerations about a possible improvement of DTH will be discussed describing an enhancement of DTH, DTH with threshold, and highlighting its impact on scheduling and performance.

3.1. Real-Time Analysis. In the following some properties of the scheduler that specify how it deals with temporal

requirements are analyzed focusing on starvation and respect of deadlines, on service rate, and on transmission queues length.

As explained in the previous sections, in DTH the assignment of $TXOP$ is refined taking into account the station needs. Differently than an approach agnostic of the stations current requirements, like that of UTSS, in DTH the traffic is monitored to estimate the stations expectations and, based on that, their transmission opportunities are modified. This recomputation is executed on-line, before each polling, monitoring T_{spare} and the station load and assigning a new $TXOP'$. Furthermore, the computation remains centralized, being carried out by the centralized scheduler with the “add-on” of the DTH module, and it is performed dynamically in order to faithfully reflect the traffic dynamicity. Taking advantage from the recovery mechanism the instantaneous $TXOP'$ can overcome $TXOP_{AC}$ in order to follow data rate increase and better serve bursts of traffic.

However, an upper bound of $TXOP$ can be set in order to avoid the starvation of the subsequent stations and the consequent risk of deadlines miss due to an excessive accumulation of recovered resources if these are not used by a sequence of stations or, for instance, are propagated between different CAPs, as will be illustrated in the following Proposition 1.

Proposition 1. *In order to avoid starvation of subsequent stations in the polling list and reduce the risk of deadlines miss, when DTH is used an upper bound $\Theta = d_i + t_{end_i} + \delta$ is set for the new assigned $TXOP'$.*

Proof. The demonstration follows the same steps than in [36] due to the use of a recovery mechanism in HCCA schedulers. In general, considering the propagation of T_{spare} from a station to the consecutive one, then for the new assigned $TXOP'$, $TXOP_i \leq TXOP'_i \leq TXOP_i + TXOP_{i-1}$. In worst case conditions, where none station uses its assigned transmission time travelling through consecutive CAPs, resources can accumulate and grow up. Hence, in this case, if a station uses the whole accumulated capacity, subsequent stations can risk to experience starvation and QoS degradation, in particular about temporal requirements and, at worst, deadline miss. Indeed, this abnormal growth of $TXOP$ can distort the relationship between polling time and respect of deadlines of the subsequent station, since the excessive polling advance cannot correlate anymore polling time and temporal constraints of each station. To avoid this misbehavior the following upper bound Θ of the transmission duration assigned to a station i can be set:

$$\Theta = d_i - t_{end_i} + \delta. \quad (6)$$

In the expression of Θ , d_i is the absolute deadline of QSTA_{*i*}, t_{end_i} is its actual transmission finishing time, and δ is a safety offset. Setting this upper bound for the assignable portion of T_{spare} ensures not overcoming d_i . \square

For the sake of completeness, some meaningful theoretical results about saturation effects, respect of deadlines, and transmission queues length, obtained in [38], are here simply

reported referring to the cited paper for the demonstrations. Then, as a consequence of these considerations and results, some deductions about the offered service rate will be formalized in Theorem 5.

Starting from the analysis of the basic behavior of DTH, the following proposition explains how its scheduling rules are set.

Proposition 2. *Setting $TXOP'_i = TXOP_{AC,i}$ when $T_{spare} = 0$ is suitable to avoid saturation when data rate begins growing and only t_{est} is considered, then, it is suitable to efficiently manage variations of VBR streams data rate.*

Focusing on the respect of temporal requirements, Theorem 3 demonstrates how DTH does not affect the respect of deadline as ensured by the centralized scheduler alone.

Theorem 3. *The global scheduler, composed by the centralized algorithm and DTH, continues meeting deadlines.*

Finally, the theorem below demonstrates the positive impact of DTH on the transmission queues length that is a key parameter that influences delay and packets drop performance, as it will be shown in the following.

Theorem 4. *The introduction of DTH is suitable to reduce transmission queues length.*

As a consequence of the previous theorem the shortening of transmission queues suggests that DTH can affect the transmission rate, allowing to dispatch more traffic and, precisely, as demonstrated in the following theorem, to increase the actual instantaneous service rate.

Theorem 5. *DTH is effective on the transmission rate, whose current value results to be increased by the factor $1 + T_{spare}/TXOP_{AC}$.*

Proof. Without loss of generality, assuming transmission queues empty at the beginning of the polling, the actual transmission rate offered at a station $QSTA_i$ is defined as the ration between the traffic Φ incoming during the accorded transmission time and the sending time itself. Without considering network state conditions, in the case of a scheduler that assigns a constant $TXOP = TXOP_{AC}$, the provided transmission rate R_t is

$$R_t = \frac{\Phi}{TXOP_{AC}}. \quad (7)$$

When DTH is integrated with the centralized scheduler and T_{spare} is recovered, this additional time can be used to deliver more traffic and the actual transmission rate R'_t evolves as follows:

$$R'_t = \frac{\Phi'}{TXOP_{AC}} = \frac{\Phi + (\Phi \cdot T_{spare})/TXOP_{AC}}{TXOP_{AC}}, \quad (8)$$

where Φ' is the total traffic incoming during $TXOP'$. In the ratio the denominator remains equal to $TXOP_{AC}$, that is,

the original transmission time accorded by the centralized scheduler, and it is the point of comparison in order to highlight the possible improvements with respect to the originally provided service. Thus, when $T_{spare} \neq 0$, $(\Phi \cdot T_{spare})/TXOP_{AC} > 0$ and $R'_t > R_t$. In particular, considering the ratio between R' and R , the factor of increase is equal to

$$\begin{aligned} \frac{R'}{R} &= \frac{(\Phi + \Phi \cdot T_{spare}/TXOP_{AC})/TXOP_{AC}}{\Phi/TXOP_{AC}} \\ &= \frac{\Phi + (\Phi \cdot T_{spare})/TXOP_{AC}}{\Phi} = 1 + \frac{T_{spare}}{TXOP_{AC}}. \end{aligned} \quad (9)$$

As shown by the obtained expression, the percentage of increase is strictly related to the ratio between T_{spare} , i.e., the added portion of transmission time, and the original transmission interval.

The same conclusions can be drawn considering the parameters involved in the calculation of $TXOP$. As set by the standard, the assigned transmission time is computed as

$$TXOP_i = \max\left(\frac{N_i \cdot L_i}{\Gamma_i}, \frac{M_i}{\Gamma_i}\right) + O, \quad N_i = \left\lfloor \frac{SI \cdot R_i}{L_i} \right\rfloor, \quad (10)$$

where M_i and L_i are, respectively, the maximum and the nominal size of a *MAC Service Data Unit* (MSDU) (the maximum is set by the standard equal to 2304 bytes), Γ_i is the minimum physical rate, N_i is the number of SDU that can be sent during SI at the mean data rate R_i , and O is the additional time due to the interframe spaces expected by the transmission mechanism. The standard computes $TXOP$ as the maximum between the times required to send, at the physical rate Γ_i , a max MSDU or N_i nominal size MSDUs. When DTH is used the effective transmission time is increased of T_{spare} and, since in the computation of $TXOP_i$ all the terms L_i , M_i , Γ_i are constant, except for N_i that is proportional to R_i , thus $TXOP' = TXOP_{AC} + T_{spare} > TXOP$ and $N'_i > N_i$. In particular

$$\begin{aligned} TXOP'_i &= \max\left(\frac{N'_i \cdot L_i}{\Gamma_i}, \frac{M_i}{\Gamma_i}\right) + O, \\ N'_i &= \left\lfloor \frac{SI \cdot R'_i}{L_i} \right\rfloor, \end{aligned} \quad (11)$$

and then R_i is the unique term that can reflect this increase. Hence, when $T_{spare} > 0$, then $R'_i > R_i$.

Thus, despite the fact that the physical rate remains untouched, increasing $TXOP$ by means of the bandwidth reclaiming mechanism of DTH has the effect to increase the actual current service rate. \square

3.2. Scheduler Analysis and Feasibility. In this subsection the issues related to the schedulability of the system when DTH is integrated with the centralized scheduler are investigated. Since adding T_{spare} makes the obtained global scheduler as alternative to overprovisioning techniques that can jeopardize the respect of the admission control resource threshold,

here the question is deepened with respect to [38], where it has been demonstrated that the admission control feasibility is satisfied also in the presence of DTH. The presented analysis, focused on the resource management, relies on the introduction of a model derived from the elastic task one [39] that operates in the processors system field. The derived model is applied to stations with multimedia traffic streams with the aim to ensure QoS expectations by adapting the transmission time to data rate variations, as it will be explained below.

3.2.1. Elastic Task Model. The elastic task model is focused on rate control of periodic tasks in the case of processor overload. In particular, considering soft real-time tasks, it assumes that it is possible to modify the utilization factor of the tasks acting on the service period, thus on their activation frequency, in order to satisfy their service expectations avoiding to discard some requests at the cost of a reduction of the offered service due to the consequent decrease of the processor rate. In particular, since most applications can show data rate fluctuations and can tolerate slightly data rate variations imposed by transmission conditions, the authors proposed an adaptive rate control in order to speed up applications with strict QoS requirements, taking advantage from a speed down of those with less strict needs. The action was on the processor rate where the tasks were running. The main idea, borrowed from the mechanical systems, was to consider the task flexible as a spring, with given elasticity coefficient and length, and to iteratively modify the tasks period and, consequently, the utilization factor, such as a spring: compress the period when the requirements are more stringent and decompress when they are less strict. As a consequence of these modifications, thinking the tasks system as a set of consecutive springs whose length influences that of each other, when a task period is changed in order to modify its data rate, all other tasks are subject to correlated modifications of the execution rate with the aim to avoid system overload or manage that in a flexible way. In this model each task τ_i is represented by the following parameters:

$$\tau_i = (C_i, T_i, T_{i_{min}}, T_{i_{max}}, e_i) \quad (12)$$

where C_i is its execution time, T_i , $T_{i_{min}}$, and $T_{i_{max}}$ are, respectively, its period and the corresponding min and max values, and $e_i \geq 0$ is the elastic coefficient that expresses how and how much the task period can be changed. In particular, the intensity of the possible modifications is dependent from the level of priority of the task: a task with more strict QoS requirements is assigned a higher priority, and thus a greater elasticity coefficient that translates the possibility to modify its period.

The corresponding utilization factor $U_i = C_i/T_i$ summarizes the previous parameters and the state of resource utilization of the considered task. The drawback of this model was that any change in the period of a task impacts on the feasibility of the resource allocation, implying a recomputation of all periods in order to respect the following feasibility condition related to the resource utilization bound:

$$\sum_{i=1}^N U_i = \sum_{i=1}^N \frac{C_i}{T_i} \leq 1. \quad (13)$$

In particular, the change of a period of a station and, consequently, of its utilization factor causes an iterative recomputation of all the corresponding parameters of the other stations, until a feasible solution for the resource utilization threshold test is found. The recomputations induced by the change of a period take into account the QoS requirements of each task and the acceptable levels of service degradation. In particular, when a task requires a greater QoS, if possible, its utilization factor U is increased, reducing its period and, consequently, the utilization factors of tasks with less strict service expectation are compressed increasing their periods using as regulatory element the feasibility test upper bound previously shown. From this point of view the periods and the utilization factors are considered as springs where they act to modify, when needed, the provided QoS and the maximum amplitude of the changes of period and utilization factor of each task is specified by the elasticity coefficient. In particular, the recomputation is performed off-line in an iterative manner until some combinations of the values of the U_i and T_i , $\forall i$ of the tasks set that meet the admission control feasibility test (i.e., until $\sum_{i=1}^N U_i \leq 1$) are found.

3.2.2. Modelling Variable Transmission Time Scheduling

(1) Motivations. In this study an approach similar to the elastic task model is adopted to investigate the correctness and the sustainability of the resource management performed by the new scheduler composed by the centralized scheduler and DTH, i.e., when DTH is added to a centralized scheduler. We do not use the elastic task model to prove the feasibility of the tasks set, as it was proposed, but to model the scheduled situation. How can the spring model this scenario? How is the spring used to express how many variations are possible for the TSs? As it will be explained in the following, two different situations are possible that impact on the setting of the model parameter: (a) consider the elasticity coefficient related to the maximum variation that can be expected and referred to the maximum data rate and (b) consider the elasticity coefficient related to the performed modification of $TXOP$ due to the use of T_{spare} .

In the presented study not only periodic traffic such as VoIP that is characterized by a period, but also the aperiodic one, such as bursty or MPEG4 traffic, is considered. Indeed, from the QoS guarantees point of view, also for aperiodic traffic worst case conditions and related upper bound of traffic parameters are considered, in order to respect the more stringent requirements. In particular, these types of traffic are managed by the scheduler that is periodic assigning a period and a transmission time.

Furthermore, in our case the overload is due to data rate variations of the multimedia application, in particular, due to data rate increase with respect to its mean value considered for the scheduling parameters computation during the admission control phase.

Finally, the features of this global scheduler and of the IEEE 802.11e Medium Access Control mechanism determine

how to modify the elastic task model adapting to the new scenario. Indeed, since the aim is to improve the efficiency of the channel utilization by recovering unused transmission capacity, already considered in the admission control threshold and, consequently, acting on the transmission time of a stream in order to reflect traffic variations and, then, to provide a dynamic differentiated QoS, thus the attention is focused on the transmission time. Consequently the periods are not affected. Furthermore, it is not of interest modifying the scheduling period, i.e., the polling period that is related to the periodicity of the traffic profile. Hence the approach of the derived model is the opposite of the elastic one, acting on a different parameter (the transmission time instead of the period) that differently affects the service rate, as demonstrated in Theorem 5. Thus no explicit data rate control is performed by the scheduler as in the original model, whereas data rate variations due to VBR TSs are accepted and the scheduling system tries to follow them modifying the transmission parameters in order to meet QoS expectations. From the point of view of the model, instead of modifying the resource utilization factors acting on the periods, in general, of all the traffic streams in order to face off load changes and avoid scheduling infeasibility, now the transmission time is tuned. In particular, differently than in the original model where the scheduling system acts directly on the periods of the tasks, here the approach is “passive” in the sense that only when there are more resources, the recovered ones, DTH takes action and assigns that to the stations with high service levels. Furthermore, the system accepts the rate fluctuations of the stations with multimedia applications and takes advantage from that to ameliorate the provided QoS.

A further difference is that, in general for the elastic task model, as mentioned, the system schedulability implies an iterative recomputation of all the periods until a combination of values suitable to respect the feasibility test is found. Instead, in the case of DTH, the recomputation of the chosen scheduling parameter, $TXOP$, is performed limited to a station at a time, the subsequent station in the polling list, that receives T_{spare} in dependency of its needs as monitored by the estimation of the future required transmission time. Hence it is like a *step-by-step* propagation of the recalculation of the transmission time involving at each step only one station at time going across the polling list.

Furthermore, the variation of $TXOP$ is due to the recovery of T_{spare} , which is reassigned in dependency of the effective current station needs as derived by the statistical analysis of its traffic trend. Thus also in this case, the approach is different in the sense that $TXOP$ is not simply recomputed taking into account only the station needs, but this recalculation is triggered by the availability of spare resources and, then, it is influenced by the station expectations related to its traffic. Moreover, as demonstrated in Theorem 5, despite the fact that the physical rate remains untouched, increasing $TXOP$ by means of a bandwidth reclaiming mechanism has the same effect on the packets delivery than increasing the service rate maintaining a fixed transmission time.

In the following the derived model will be deeply described specifying the used notation, the resulting analysis

method, and highlighting its differences with respect to the original one.

(2) *Notation.* In the presented model each traffic stream TS_i that is equivalent to a task in the real-time systems terminology is assigned the following:

- (i) A capacity ($TXOP_i$)
- (ii) A period (SI_i)
- (iii) The stream resource utilization factor $U_i = TXOP_i/SI_i$, function of these parameters
- (iv) The elasticity coefficient e_i
- (v) The length of the corresponding spring

This notation reflects the parameters nomenclature as stated in the reference scheduler. In HCCA schedulers the period is equivalent to SI and it is correlated to the Delay Bound scheduler parameter. At its turn the Delay Bound is related to the stream deadline that sets the time until the execution/transmission has to finish; thus it expresses temporal requirements. As in the original model, in the reference scheduler both $TXOP$ and SI are set as fixed value during the admission control phase when available resources are divided between stations in dependency of their QoS expectations.

Now the elasticity coefficient e_i is related to the possible variations of the transmission time, intended as the maximum allowable variations in dependency of characteristics and priority of the TS. Indeed, the presence of $TXOP_{min}$, $TXOP_{max}$, and $TXOP_{AC}$ parameters set the upper and lower bound of the elasticity coefficient, i.e., the maximum and the minimum increase and decrease. In particular, $TXOP_{min}$ and $TXOP_{max}$ are related to the characteristics of the traffic streams, whereas $TXOP_{AC}$ represents the reference value, corresponding to the mean value parameters and related to the nominal length of the spring. Finally, the spring now is related to the behavior of the transmission time.

(3) *Differences with the Elastic Tasks Model.* Summarizing, the main differences between the original elastic task model and its derivation introduced in this study are the following:

- (i) The scheduler modifies the capacity, i.e., the transmission time, instead of the period
- (ii) The period is not modified; thus the polling list is not changed, in the case, for instance, the polling order is set taking into account the deadlines
- (iii) The original centralized scheduler assigns resources considering the mean value of the service parameters and the additional module increases the utilization factor
- (iv) Due to the recovery mechanism there is not the risk to overcome the resource threshold set by the admission control, as shown in [38]
- (v) Only the transmission duration of a station at time is recomputed when the scheduler manages the stations and goes through the polling list, whereas in the

original model a change of task parameters impacts on the values of all the other tasks

- (vi) The aim is not the rate control, but filling in the gaps of the reference scheduler and, in general, of schedulers with a fixed transmission time, trying to provide a variable $TXOP$ and to better satisfy QoS requirements of the stations
- (vii) The system tries to follow data rate variations and not to control them
- (viii) The service rate, as provided by the scheduler, is fixed and relative to the physical rate set by the IEEE 802.11e standard that serves different TSs whose applications have different data rate. Instead, in the original model each task can have a different processing rate and the model acts on that in order to better follow traffic variations and avoid processor overload. In the presented model, the obtained service rate results to be modified with respect to the one provided by the scheduler, as a consequence of the changes of transmission time, whereas the period remains unchanged
- (ix) The recomputation of $TXOP$, differently than in [39] where it is performed iteratively off-line, is performed on-line, before each polling

3.2.3. Elastic QoS Scheduling for VBR Traffic. As set by the HCCA standard protocol an underutilization of $TXOP$ produces a polling in advance of the consecutive traffic streams, since the station frees the medium and the control returns to the *QoS Access Point* (QAP) that polls the next station in the list after a *Polling InterFrame Space* (PIFS). But this behavior can be used also to reclaim and assign unspent time to the next polled station. This is the main difference between the original model and the derived one. Indeed, in the usual application of the original elastic task model the action is on the utilization factor and, consequently, on the service rate, by varying the period in dependency of the system load. In the new model the changing system load, due to multimedia applications, allows, by means of unused resources made available by the scheduler, to provide an instantaneous overprovisioning useful in the case of overload. Since resources are directly provided, the variable that is naturally changed is the transmission time instead of the period. Consequently, when an additional time is assigned to a station, its global transmission time $TXOP'$ and its utilization factor increase. In the subsequent section it will be shown, from a different point of view, that this modification impacts on the service rate. But, differently than the original model, since recovered resources are simply “shifted” from one station, which does not use that, to the subsequent one, the feasibility test is not violated and the resource threshold is not exceeded, as demonstrated in [38].

We assume starting the scheduling from the assignment as set by the reference scheduler:

$$C_{mean_i} \equiv TXOP_{AC_i} \quad (14)$$

is adopted as a “frugal” solution, despite max values parameters that should require $C_{max_i} \equiv TXOP_{max_i}$.

Since HCCA is a *Time Division Multiplexing* (TDM) protocol, the different tasks are scheduled consecutively; thus, considering the total length of all streams equal to the maximum assigned resources, each variation in the resource utilization can impact on other streams. Hence the reclaiming action consequent to an underutilization allows the “decompression” as stated by the elastic task model and, since only already assigned resources are considered, it permits a sort of overprovisioning but, differently than the elastic model, without the need of a new check of the feasibility of the admission control.

As previously stated, in the presented approach $TXOP$ s of all stations are not iteratively recomputed as a consequence of network overload but only the $TXOP$ of the next polled station is impacted, receiving the recovered resources resulting from the transmission of the previous stations. Consequently the recomputation of $TXOP$ is propagated step-by-step along the polling chain involving at each stage only two stations, the one that finished its transmission not exhausting its assigned transmission duration and the subsequent one that receives the exceeding resources. Figure 4 illustrates this situation, when at polling time $\delta TXOP_{AC}$ of the different stations are considered and it is assumed that the station $QSTA_i$ is polled, whereas at polling time $(\delta + 1)$, for instance, the station i has not exhausted its transmission time and the resulting t_{spare} is assigned to the next polled station $QSTA_{(i+1)}$. Hence, in this model the transmission time of each station can be considered as a spring and the total *Controlled Access Phase* (CAP) can perform like an accordion, compressing if, in the propagation of resources along the $QSTAs$ chain, resources are not globally exhausted and, then, passed to the next CAP that, at its turn, results in a greater length.

Each stream is represented by the following set of parameters:

$$TS_i = (TXOP_i, TXOP_{min_i}, TXOP_{max_i}, SI_i, e_i) \quad (15)$$

where the transmission duration varies in the interval $[TXOP_{min_i}, TXOP_{max_i}]$. In Table 1 a synoptic of the two versions of the model, one for variable transmission time and one for variable period, is provided.

$TXOP_{max}$ is set considering the $TXOP_{Limit}$ protocol parameter that set the maximum length of the transmission time that can be assigned to the considered station. Instead $T_i \equiv TXOP_{AC_i}$ set during the admission control as the maximum time to transmit the bits enqueued during SI at the minimum physical rate Γ_i :

$$TXOP_{AC_i} = \max\left(\frac{N_i \cdot L_i}{\Gamma_i}, \frac{M_i}{\Gamma_i}\right) + O, \quad (16)$$

where $N_i = \lceil (SI \cdot R_i) / L_i \rceil$ and M_i is the maximum *MAC Service Data Unit* (MSDU) size, i.e., 2304 bytes, L_i is the nominal MSDU size, R_i is the mean data rate, and O is the transmission overhead due to interframe spaces, ACK, CF-POLL.

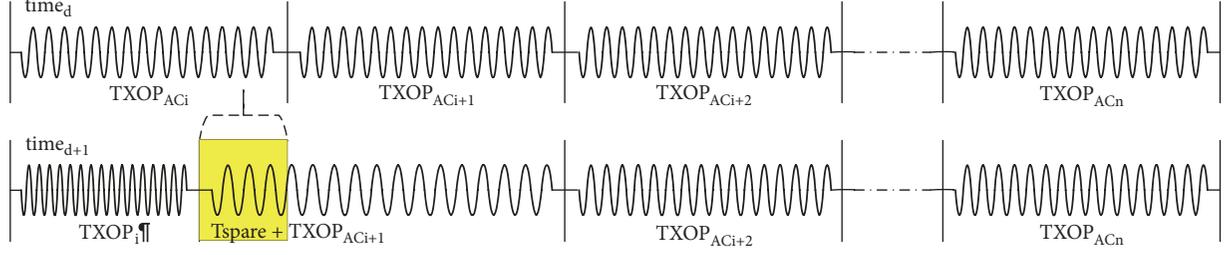


FIGURE 4: Elastic QoS scheduling.

TABLE 1: Synoptic of notations of the proposed model and of elastic task one.

Derived model variable $TXOP$	Elastic task model variable period
Traffic Stream TS_i	Task τ_i
Transmission time $TXOP$	Execution time C_i
Service Interval SI_i	Period T_i
$TS_i = (TXOP_{0i}, TXOP_{min_i}, TXOP_{max_i}, SI_i, e_i)$	$\tau_i = (C_i, T_i, T_{min_i}, T_{max_i}, e_i)$

The recomputation of $TXOP$, differently than in [39] where it is performed iteratively off-line, is performed on-line, before each polling, monitoring T_{spare} and the station load and assigning a new $TXOP^l$.

In order to model the $TXOP$ variation the following expression of the length of a spring is used:

$$F = e(L_o - L_i) = e\Delta_l \quad (17)$$

where F is the strength applied to compress the spring until the length L_i , L_o is the static length, without application of any strength, e is the elasticity constant that expresses the elasticity properties of the spring (function of its material, its shape), and Δ_l is the length variation.

When DTH is used, this expression can be applied to the transmission time modeled as a spring length. In particular, the model is applied extending the concept to the case of decompression, i.e., when the transmission time is increased, whereas in this scenario the case of compression is due to the “spontaneous” reduction of transmission time due to traffic profile as managed by the MAC scheduler. In this situation the maximum value of the transmission time, obtainable assigning T_{spare} , is

$$F = e(TXOP_{max_i} - TXOP_{AC}) \quad (18)$$

In this expression the static length is $TXOP_{AC}$ assigned considering mean value parameters, and $TXOP_{max}$ is the maximum transmission time duration as requested by the traffic profile and related to the maximum values (or minimum) of the parameters. Without loss of generality, choosing for $TXOP_{AC}$ the first expression in Equation (16), i.e., attributing the variation to data rate changes, coherently with the focus of this work, this choice does not jeopardize the resource threshold since if it is the max, it is as expected; otherwise it is a conservative option. In this case $TXOP_{AC}$ results to be equal to

$$TXOP_{AC} = \frac{S_i R_{mean_i} L_i}{\Gamma_i L_i} = \frac{S_i R_{mean_i}}{\Gamma_i}, \quad (19)$$

whereas

$$TXOP_{max} = \frac{S_i R_{max_i} L_i}{\Gamma_i L_i} = \frac{S_i R_{max_i}}{\Gamma_i}, \quad (20)$$

where the variations are due to the data rate changes, as expected. Thus

$$\begin{aligned} F &= e_i \cdot \Delta_i = e_i \cdot \left(\frac{S_i R_{max_i}}{\Gamma_i} - \frac{S_i R_{mean_i}}{\Gamma_i} \right) \\ &= \frac{e S_i}{\Gamma_i} (R_{max_i} - R_{mean_i}) \end{aligned} \quad (21)$$

where Δ_i is the transmission duration variations. As in the original model, e_i is function of the priority assigned to TS_i about DTH action but, since in our case all the TSs have the same priority, there is not differentiation in the action of DTH and all TSs are treated in the same manner; thus it is possible to assign $e_i \equiv 1$. Finally, using Equation (9)

$$\begin{aligned} F &= e_i \left(1 + \frac{T_{spare}}{TXOP_{AC_i}} - 1 \right) R_{mean_i} \frac{S_i}{\Gamma_i} \\ &= e \frac{T_{spare}}{TXOP_{AC}} R_{mean_i} \frac{S_i}{\Gamma_i}. \end{aligned} \quad (22)$$

Finally, in order to model this new scenario, when the elasticity action is due to the recovery mechanism, whereas all the TSs have the same elasticity properties, it is possible to introduce an instantaneous e'_i that is function of T_{spare} collecting the priority and the recovery aspects. Hence

$$e_i \Delta_l i = e_i \frac{T_{spare}}{TXOP_{AC_i}} R_{mean_i} \frac{S_i}{\Gamma_i} = e'_i R_{mean_i} \frac{S_i}{\Gamma_i} \quad (23)$$

where

$$e'_i = \frac{T_{spare}}{TXOP_{AC_i}} \quad (24)$$

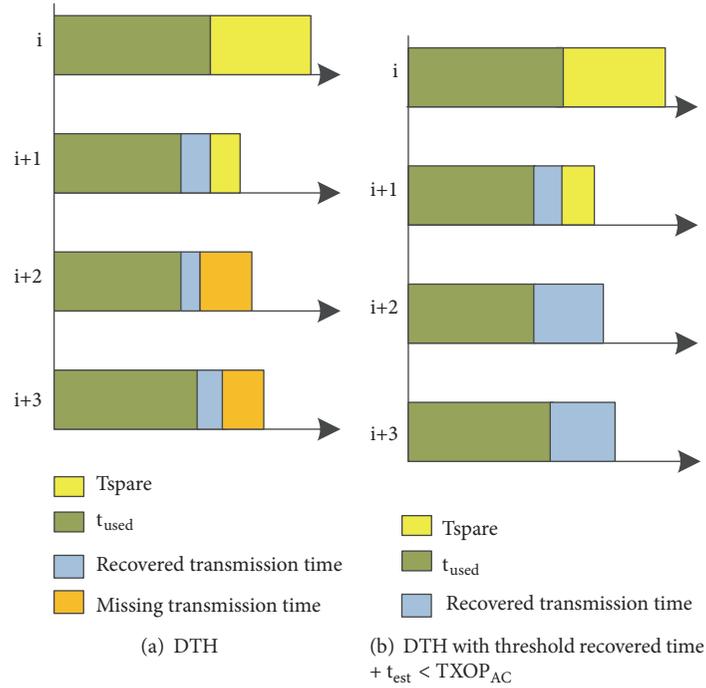


FIGURE 5: Improving DTH responsiveness.

that expresses how the instantaneous elasticity property of TS_i , i.e., the possibility to modify its $TXOP$, is function of its $TXOP_{AC_i}$ and of the instantaneous T_{spare} . This consideration highlights a further difference between the original model and the current one: the elasticity properties of a traffic stream are function of its original transmission time assigned during the admission control that collects its traffic features and of the instantaneous action of DTH that recovers T_{spare} .

3.3. Improving DTH. DTH has been shown to be apt to provide a $TXOP$ more suitable to actual stations needs with respect to the reference scheduler. From the point of view of a flexible and dynamic service provisioning an aspect that can be further improved is the responsiveness of the system to traffic profile changes in order to continuously meet service expectation and, in particular, to respect the negotiated service guarantees. Indeed, despite the fact that the moving window monitoring system is a good solution to take into account traffic fluctuations in the transmission time assignment overcoming the drawbacks of the reference scheduler, this method makes the updating process slow. Indeed, the average operation, on one side, has the advantage to permit considering traffic temporal evolution and, on the other, makes the mechanism of $TXOP$ updating slow since an increment has to be a consequence of previous samples when the effectively used transmission time is, itself, increased. In particular, it is the progression of the samples selected by the window that allows to introduce in the computation t_{used} little by little greater than the previous ones.

This is better explained in Figure 5(a), where it is shown, as in the case of DTH, the recovery mechanism mitigated by the transmission time estimation makes the additional

resource assignment slow since the added time is computed taking into account the station behavior inside the moving window, thus not considering the punctual values of the used transmission time. Despite the fact that this mechanism allows to perceive the traffic trend, it introduces this hysteresis in the speed of adaptation of the algorithm. This consideration is at the basis of the introduced improvement of DTH that from now is named *DTH with threshold*.

Deepening the concept, the DTH method of T_{spare} assignment is a stochastic process depending on the traffic profile of the previous station/stations in the polling list, its/their service expectations, its/their effectively used transmission time, and its/their position in the polling list, where the recovered capacity is propagated along the stations chain and its use in each station that generates the assigned $TXOP$ to the currently polled station. This implies that it is not possible to foresee if the station will receive the required $TXOP$ tailored to its traffic variations and suitable to respect the service contract negotiated during the admission control and set taking into account mean value parameters, such as mean data rate. If, as illustrated in the example, data rate increases after a long decrease that has conducted to assigning $TXOP' < TXOP_{AC}$, if now a $TXOP' \geq TXOP_{AC}$ should be necessary, neither the negotiated service will be provided. In order to understand the problem, for instance, the case of a chain of monitored transmission of a station is considered in three different situations where the amplitude of the future increments is (a) equal, (b) greater, and (c) less than that of the past decrements.

Case (a) In this case,

$$\forall i \ 1 \leq i \leq N : t_{used_i} < TXOP_{AC} \quad (25)$$

If $t_{used_i} < TXOP_{AC}$, $t_{used_i} = TXOP_{AC} - T'_{spare}$, where T'_{spare} is the residual time of the previous station $QSTA_{i-1}$, $TXOP'_i = t_{est_i} + T_{spare}$, and T'_{spare} is the residual time of the current station $QSTA_i$.

Assigning $TXOP_{AC_i} - t_{used_i} = \Delta_i$, thus $t_{used_i} = TXOP_{AC_i} - \Delta_i$. In order to have a new assigned $TXOP'_i = TXOP_{AC_i}$, a j has to exist, with j being a new sample, i.e., $N + 1 \leq j \leq 2N$, such that $t_{used_j} = TXOP_{AC_j} + \Delta_j$. Consequently, proceeding couple by couple in order to have that $\forall i, TXOP'_i = TXOP_{AC_i}$, each new increment has to compensate a past decrement, i.e.,

$$\forall i, \text{ with } 1 \leq i \leq N, \exists j, \text{ with } N + 1 \leq j \leq 2N \quad (26)$$

s.t.

$$\begin{aligned} t_{used_i} &= TXOP_{AC_i} - \Delta_i, \\ t_{used_j} &= TXOP_{AC_j} + \Delta_j > TXOP_{AC_j}, \end{aligned} \quad (27)$$

where $\Delta_i \equiv \Delta_j$.

Thus it is possible to state that, in this case, $TXOP_{AC}$ will be reached after $N/2$ steps.

Case (b) When the amplitude of the increments is greater than that of the previous decrements, the transmission rate will increase faster. For instance consider the case that $QSTA_{i-1}$ leaves to $QSTA_i$ a big T_{spare} equal to

$$T_{spare} = \Delta \geq \sum_{j=k}^{k+l} t_{used_j}, \quad (28)$$

where $1 \leq k \leq N, 0 \leq l \leq N - 1$; thus T_{spare} will absorb some decrements and less steps will be required with respect to the previous case. The same result is obviously true in the case $\Delta_j > D_j, \forall i, j$. In this case the steps required to reach $TXOP_{AC}$ will be less than $N/2$.

Case (c) In the opposite situation, where the data rate slowly increases and the amplitude of increment is less than that of decrement, i.e., if $\forall i$, such that $1 \leq i \leq N$ when the window moves $\exists j$, with $N + 1 \leq j \leq 2N$ s.t. $\Delta_j < \Delta_i$, thus the steps required to reach $TXOP_{AC}$ will be greater than $N/2$.

Thus two situations can occur: the current polled station could immediately receive the required transmission interval dispatching all the arriving packets or it could wait for a greater amount of capacity (in Figure 5 the missing time is in orange); thus it experiences a slow increase of sending time with respect to its requirements. This randomness, due to traffic profile and due to the position of the station in the polling list, makes the attempt to follow traffic variations less reactive.

For the illustrated motivation in *DTH with threshold* the assignment of $TXOP$ is modified as follows:

$$TXOP'_i = \begin{cases} TXOP_{AC_i}, & \text{if } T_{spare} = 0 \vee \\ & t_{est_i} + T_{spare} < TXOP_{AC_i} \\ t_{est_i} + T_{spare}, & \text{if } T_{spare} > 0 \wedge \\ & t_{est_i} + T_{spare} > TXOP_{AC_i} \end{cases} \quad (29)$$

TABLE 2: MAC/PHY simulation parameters.

Parameters	Value	Parameters	Value
SIFS (μs)	10	PLCP header (b)	24
DIFS (μs)	28	Preamble (b)	72
PIFS (μs)	19	Data Rate ($Mbit/s$)	54
Slot Time (μs)	9	Basic Rate ($Mbit/s$)	1

Assigning $TXOP_{AC}$ in the case of $T_{spare} = 0$ avoids to turn away $TXOP$ to its mean value. In the following section both DTH and its improved version will be deeply analyzed through simulation in order to illustrate their behavior.

4. Performance Evaluation

In the following the performance evaluations about the behavior of DTH and its improvement are illustrated.

First of all the simulation scenarios (settings and traffic model) are described. Then the simulation results are shown where DTH and DTH with threshold are compared with the reference scheduler, FHCF, RTH, WCBS, WCBS + UTSS (referred to as UTSS), and WCBS + IDTH (IDTH). On this scope WCBS is chosen as scheduler to add DTH and its enhancement since, being based on EDF for the management of deadline and, then, for the polling list, it shows poor delay performances, as highlighted in [21]. This makes WCBS a meaningful use case of application. FHCF has been chosen being a classical queues length-based algorithm, whereas RTH is a real-time scheduler for VBR traffic.

In order to investigate the effect of the studied algorithms on the centralized scheduler the performances about the efficient stations management expressed in terms of null rate and polling interval are illustrated. Then, real-time performances are investigated considering transmission queues length, mean access delay, and packet drop rate. Finally, the analysis about the throughput of the system and the evaluation of the reclaiming of EDCA resources complete the study on the schedulers performances along with some considerations about the fairness of the system.

4.1. Simulation Scenario. The settings of the simulations are as follows:

- (i) First is Network Simulator 2 (*ns-2*) [40], with its extension for HCCA [41]
- (ii) IEEE 802.11g physical layer: it is based on OFDM (*Orthogonal Frequency-Division Multiplexing*) and its parameters are summarized in Table 2
- (iii) MAC level fragmentation, multirate support, and RTS/CTS protection mechanism are disabled
- (iv) All nodes can directly communicate with each other, so the hidden node problem is not considered
- (v) Simulations are done running independent replications of 700 s with a warm-up time of 100 s until the 95% confidence interval. Nonmeaningful confidence intervals are omitted

TABLE 3: G.729A VoIP traffic stream parameters.

Parameter	Value	Parameter	Value
Frame size (B)	10	Payload size (B)	20
Frame per packet	2	IP/UDP/RTP	
Period (s)	0.02	Header size (B)	40
Data rate (kb/s)	24	SDU size (B)	60

TABLE 4: Video conference and video streaming traffic parameters.

Parameter	VC	VS1	VS2
Mean frame size (B)	660	3800	2900
Maximum frame size (B)	11386	16745	22239
Period (s)	0.033	0.040	0.040
Mean data rate (kb/s)	158	770	580
Maximum data rate (kb/s)	2733	3300	4400
Peak/Mean data rate	17.32	4.29	7.59
Parameter	VS3	VS4	VS5
Mean frame size (B)	2900	3500	4600
Maximum frame size (B)	15251	16960	16550
Period (s)	0.040	0.040	0.040
Mean data rate (kb/s)	580	700	910
Maximum data rate (kb/s)	3100	3400	3300
Peak/Mean data rate	5.35	4.85	3.62

The simulation scenario is composed of one QAP, of seven QSTAs that send one uplink TS to the QAP, and of a further station, not shown in the simulations, that is used to saturate the channel. This station is backlogged and sends data traffic with 1500-byte *Service Data Unit* (SDU) using the legacy *Distributed Coordination Function* (DCF).

In particular, a QSTA, named *VoIP*, sends VoIP traffic encoded with G.729A codec whose *Traffic Specification* (TSPEC) parameters are summarized in Table 3.

A QSTA, named *VC*, and five QSTAs named *VS[1-5]* transmit preencoded *high quality* MPEG4 trace files of 60 minutes each, taken from the Video Trace Library for Network Performance Evaluation [42]. These video traces are often used in literature and they are useful for performance comparison. Specifically, such trace files are as follows: *LectureHQ-Reisslein* (VC), *Jurassic Park* (VS1), *Silence of the lambs* (VS2), and *Mr. Bean* (VS3), *Die hard III* (VS4), and *Robin Hood (Cartoon)* (VS5).

The TSPECs of video conference and video streaming parameters are summarized in Table 4.

4.2. Simulation Results

4.2.1. Scheduler Efficiency. The scheduler efficiency related to the efficient management of TSs is here investigated considering the number of the admitted TSs, the *null rate*, and the *polling interval*.

Looking at the number of admitted TSs, since DTH and DTH with threshold are simply added to a centralized scheduler and act on the scheduling engine about the computation of instantaneous *TXOP*, the obtained global scheduler uses the admission control test of the original scheduler (in

the case of these simulations WCBS); thus the number of admitted TSs is unchanged and results to be greater than that of the reference scheduler. This consideration is valid in general: DTH and its enhancement not only do not impact on the admission control, as explained in Section 2, but also use the same feasibility test.

The efficiency of the scheduler intended as proper choice of the parameters is deepened considering the *Null rate* that is the CF-NuLL packets sending rate sent by a polled station without data to transmit to the QAP. Thus a value greater than zero highlights a not optimal assignment of the polling period that results to be not correlated to the period of the traffic. As a result the QAP polls the station also when it has no data to send, increasing the network overhead.

In order to understand the simulations results, first of all, the polling period, illustrated in Figure 6, is considered.

As expected, the reference scheduler assigns the same polling period to all the TSs, producing a not efficient polling without any differentiation. This is confirmed by the null rate analysis, illustrated in Figure 7.

The joint analysis of these two Figures highlights that (a) the reference and FHCF use a unique value of the polling interval for all TSs, which is less than the minimum *Maximum Service interval* (MSI) of all admitted TSs; (b) RTH, WCBS, and UTSS, IDTH, DTH assign a different polling period to the different TSs, producing a reduced null rate; i.e., they poll the stations with more attention to the traffic profile. In general, this implies that the stations are polled when they have enqueued traffic. In particular, at the beginning DTH and DTH with threshold have the same polling period than the centralized scheduler who they are added to, in this case WCBS. Then, during the

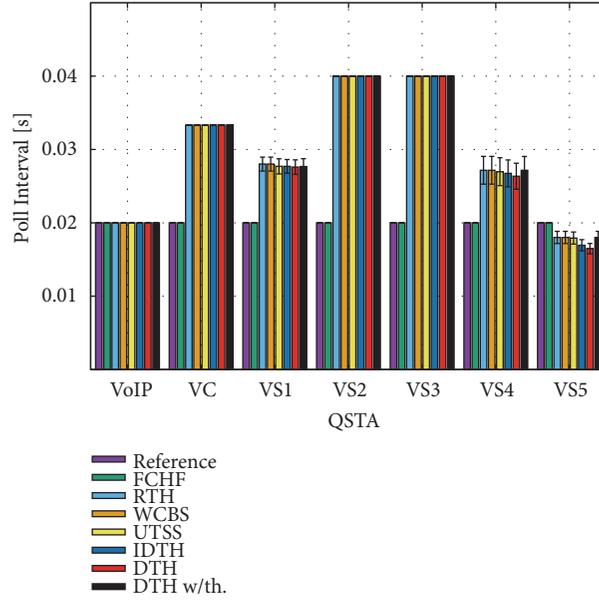


FIGURE 6: Polling period.

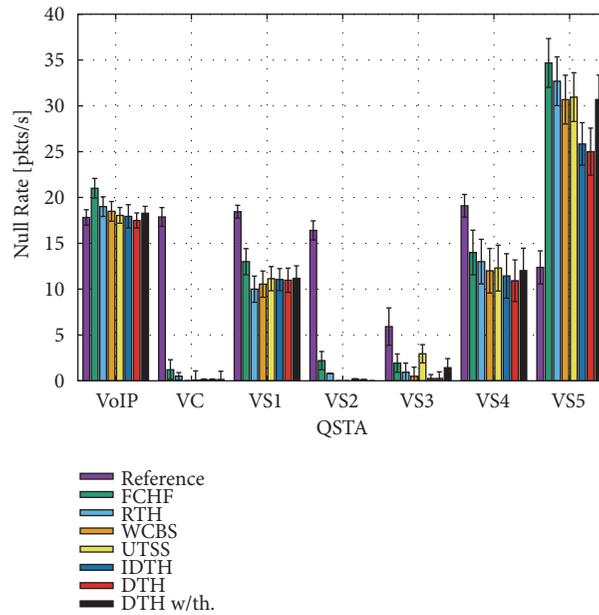


FIGURE 7: Null rate.

scheduling activity the improved assignment of $TXOP$ can generate a different polling interval, as shown in Figure 6, that reports the average polling interval experienced by the different stations during the simulations. This aspect and their refined scheduling activity that improves the centralized scheduler efficiency imply a lower null rate in the case of video-streaming stations, whereas in the case of the VoIP one the results are the same being a periodic traffic. The small differences between WCBS alone and with DTH and DTH with threshold are due to recomputation of $TXOP$, to the reclaiming mechanism and to the consequence polling shift.

Furthermore, Figure 6 highlights that reference scheduler and FCHF keep the same polling period, while the other

schedulers, except in the case of CBR traffic of VoIP, compute different polling periods. These periods, for the same TS, are quite similar between these schedulers because they use an EDF-based mechanism to compute them.

Figure 7 shows that, while for CBR traffic the null rate performance is comparable among the schedulers, with VBR TSs, in most of the cases the non-reference schedulers show a better behaviour except with VS5.

4.2.2. Delay Performance. The temporal behavior of the different schedulers is analyzed considering, first of all, the transmission queues length where the packets are enqueued waiting to be transmitted. This analysis is meaningful being

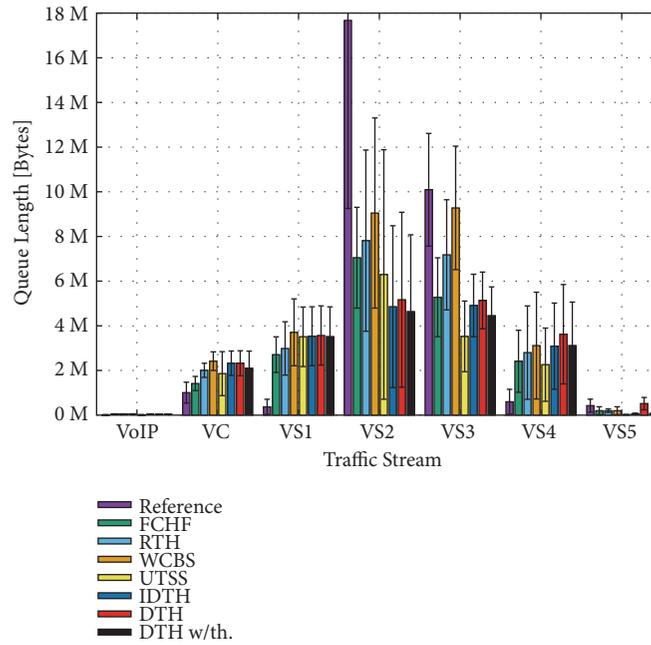


FIGURE 8: Traffic Stream queue length: 99th percentile.

related to the efficient choice of the transmission duration suitable to dispatch all the waiting TSs.

In Figure 8 the 99th percentile of queue length of the different TSs is illustrated. This statistical estimation highlights the overall behavior of the packets dispatching mechanism. As expected and demonstrated by the theoretical analysis, DTH is suitable to reduce the queues length providing additional transmission time by adding the recovered T_{spare} , reducing the queues length as generated by the centralized scheduler alone. In general, the performances of DTH and DTH with threshold are comparable with other EDF-based schedulers, but TSs with highly variable data rate, like VS2 and VS3, show the greatest benefits from these mechanisms. In the case of DTH the reduction can reach around 50%. DTH with threshold shows even better performance due to its improved responsiveness that lets to quickly recover the missing transmission time where the best results are obtained.

In Figure 9 the analysis about the transmission process is deepened by the evaluation of the *Cumulative Distribution Function* (CDF) of the queue length for VS3 station: coherently with the previous results, DTH and DTH with threshold are able to shorten the queue length with respect to different analysed schedulers for every percentage of the distribution. Among the tested schedulers, UTSS, despite its simplicity, offers the best performance until around 90% of the CDF, followed by DTH with threshold that requires less than 5.8 MB with the 95% probability. Indeed, its responsiveness in following the traffic variations allows a more efficient emptying of the transmission queues, since the increased transmission time is suitable to dispatch the packets during the same polling time of their arrival, without any further enqueueing delay.

Then the delay performances are investigated, being directly influenced by the transmission queues length. In

Figure 10 the mean access delay is evaluated. As announced by the above analysis, DTH and DTH with threshold are able to improve the performance of highly VBR TSs like VS2 and VS3, while they are penalized with other TSs. This performance is confirmed even looking at the CDF of the access delay of VS3, as shown in Figure 11, which is justified by the behavior of the transmission queue.

The delay performance is responsible, at its turn, for the number of discarded packets per second by the QAP that does not consider packets that arrived with a delay longer than the allowable maximum. In this case, Figure 12 shows that DTH and DTH with threshold perform better than the other schedulers when they deal with highly VBR TSs, where the need of further resources is more strict.

4.2.3. Throughput Performance. Finally, the throughput of the system is analyzed in order to complete the study on the schedulers performances. Figure 13 shows that DTH and DTH with threshold slightly improve the overall throughput only with high VBR TSs, because they are able to efficiently recover the unused time from other stations, losing the lowest number of packets, as shown above.

4.2.4. Performance of Reclaiming EDCA Resources. In this paper we want also to show how some of these schedulers can benefit from reclaiming EDCA resources. In particular, we consider Overboost [18] local node scheduler that is suitable for improving performance of HCCA schedulers of IEEE 802.11e networks by exploiting the EDCA function. Overboost switches the data traffic exceeding the HCCA transmission time limit to the queue of the highest priority EDCA access category. We have added Overboost local node scheduler to WCBS, UTSS, and DTH, and we have compared them with the same schedulers without this extension and with the reference scheduler.

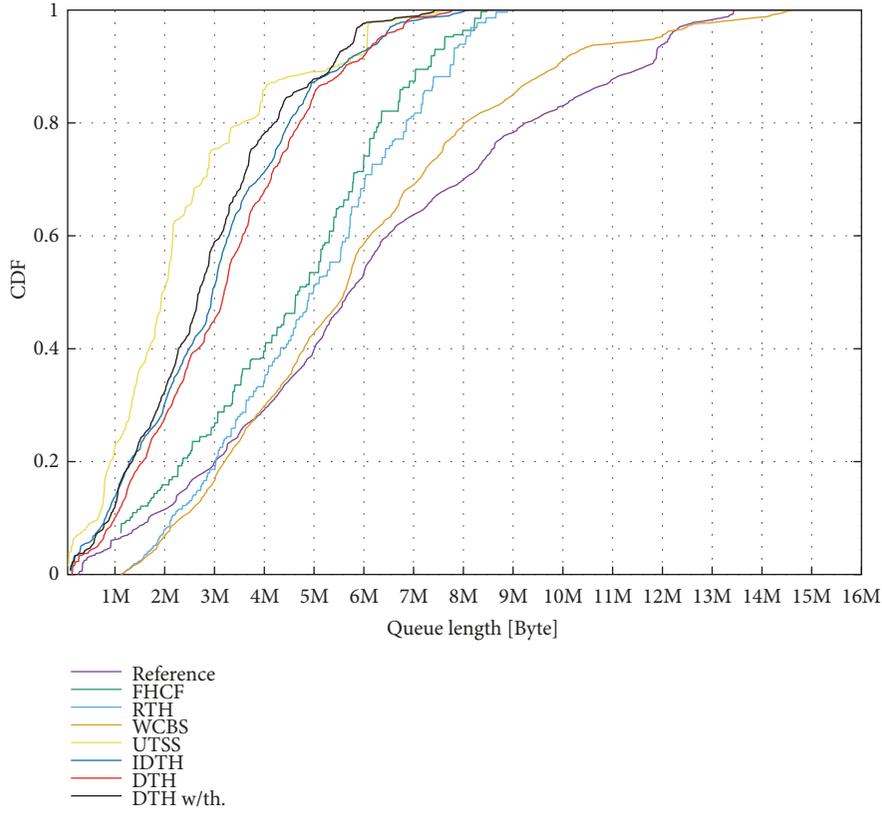


FIGURE 9: Queue length CDF of VS3 station.

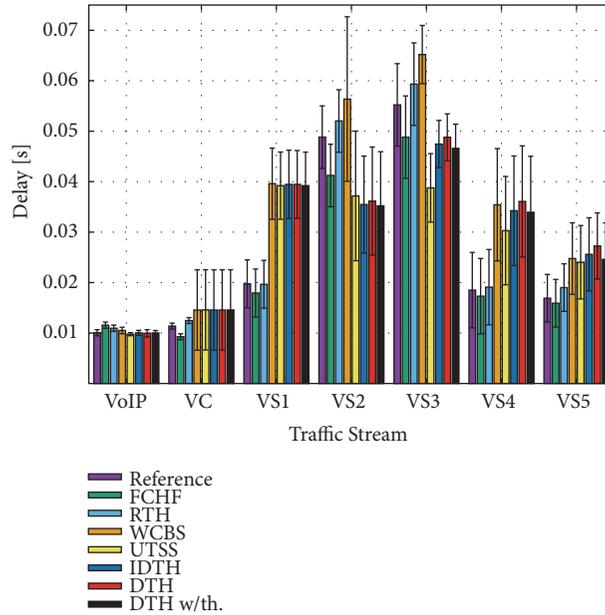


FIGURE 10: Mean Access Delay.

Looking at the mean access delay, in Figure 14 the improved schedulers reduce drastically their delays, obtaining the best values for each TS, as expected.

Regarding the throughput, in Figure 15 it is highlighted that Overboost offers a small improvement for every scheduler, because their performance was already satisfactory.

4.2.5. *Fairness Analysis.* Not only is the approach of DTH useful to improve the transmission capacity efficiency, but, since it tries to accord a *TXOP* more tailored to the station expectations, it encompasses also fairness issues.

First of all, as explained, the context is that of QoS provisioning, where asymmetries in the resources assignment

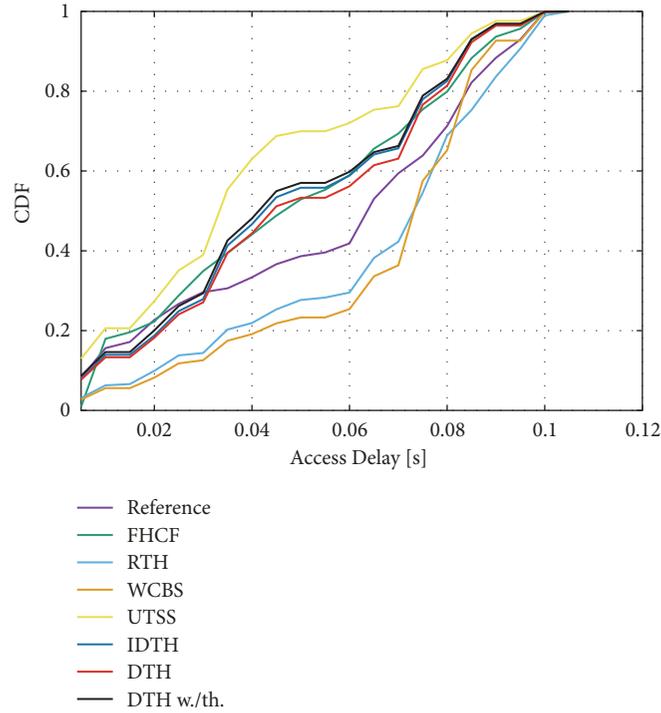


FIGURE 11: CDF of VS3 Access Delay.

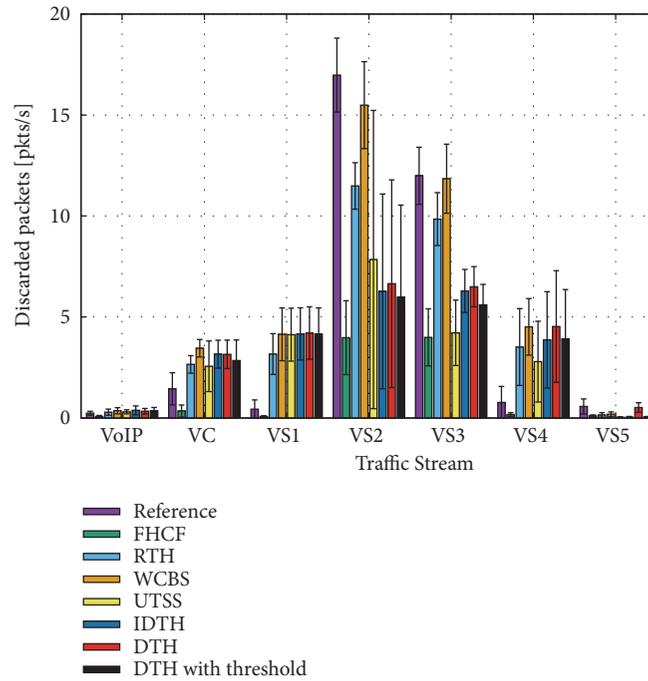


FIGURE 12: Discard packets rates.

are required in order to differentiate the service. Then, focusing on VBR applications, the goal is to calibrate the accorded $TXOP$ to the effective and current station needs: time series forecasting is used to evaluate the future traffic behavior, overcoming the assignment performed during the admission control based on mean values parameters. This is

a refinement that makes the system similar to a multirate system where the data rates of the different entities are changed in order to improve the overall fairness. Indeed, as explained above in Section 3.1, when T_{spare} is added, the effective transmission rate is increased of the ratio expressed by Eq. (8)

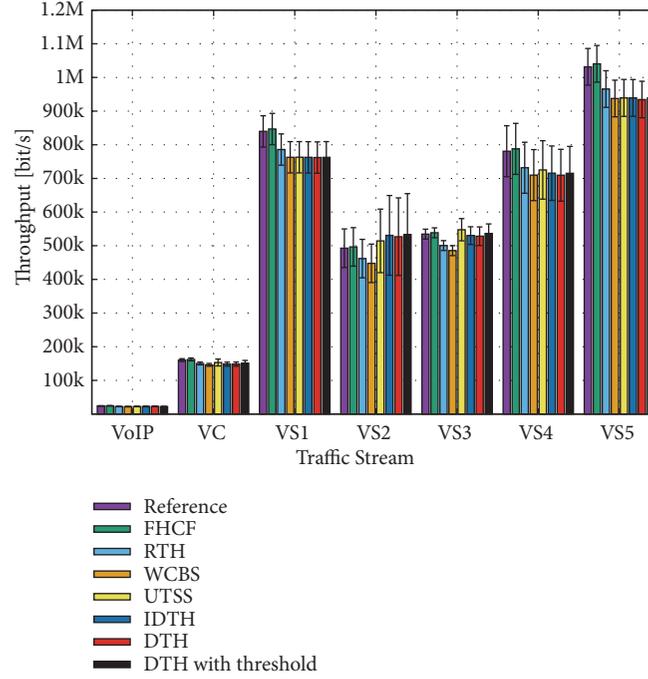


FIGURE 13: Throughput.

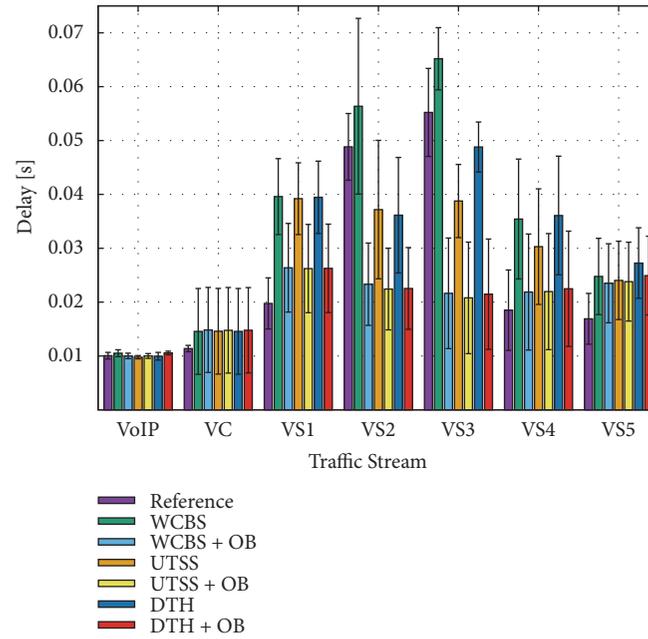


FIGURE 14: Mean Access Delay with Overboost EDCA local scheduler.

$$R'_t = \frac{\Phi + (\Phi \cdot T_{spare}) / TXOP_{AC}}{TXOP_{AC}}. \quad (30)$$

In particular, in the case of decreasing of data rate, it is the MAC protocol itself and the admission control QoS negotiation that ensure the negotiated service to the QSTA. In this case the station uses the needed portion of $TXOP$ and makes the remaining part available for other stations, just leaving more time for the next CAP, or for the EDCA part, or,

in the case of advanced schemes, such that of DTH and DTH with threshold, leaving this transmission time, for instance, by means of reclaiming mechanisms to other stations in the HCCA portion of the CAP.

In the case of growing rate, T_{spare} allows to increase the service rate; thus the system is more fair in the sense that it takes some resources from stations that do not use that and gives that to stations that instantly require more capacity. T_{spare} is assigned to the next polled station without

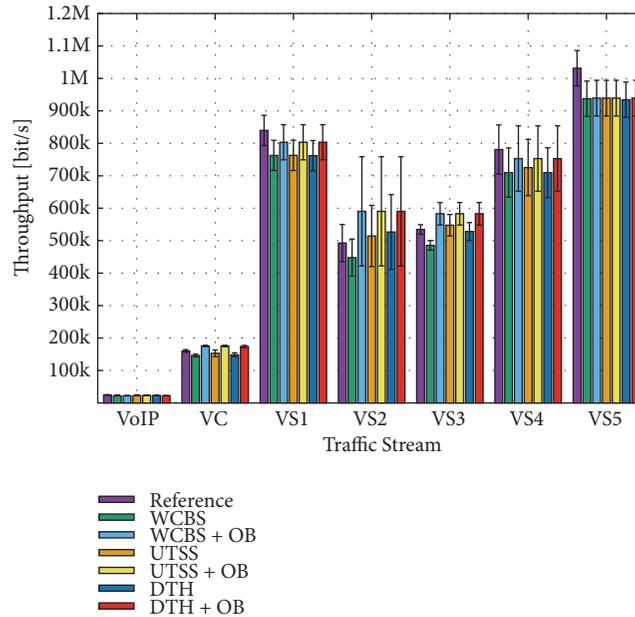


FIGURE 15: Throughput with Overboost EDCA local scheduler.

any distinction in dependency of the stations needs, being the negotiated QoS requirements ensured by the protocol.

Focusing on short-term fairness, the performances of the different QSTAs are dependent, in general, from their position in the polling list that is managed in order to meet the service expectation of the different stations. Different policies can be adopted, and this modifies the order of the stations. For instance, the reference scheduler uses Round Robin, and some algorithms are based on *Earliest Due Date* (EDD) [43] that schedules the stations in increasing deadlines order, or EDF that is a derivation of EDD allowing dynamic activation of tasks (here QSTA), the reordering of the polling list at each new insertion with the increasing deadlines order. WCBS is an EDF-based algorithm plus a recharging capacity scheme that implies a dynamic change of the polling list at each polling.

However, as mentioned the scheduler meets QoS requirements as negotiated during the Admission Control phase where the traffic is classified in the different Traffic Streams, the traffic specifications are declared, and, consequently, the accorded service is assigned by the QAP by means of the tailored computation of $TXOP$ and SI .

In this context, IDTH, which assigns T_{spare} considering the previously used $TXOP$, has the effect to improve the instantaneous fairness between stations, since it does not directly assign T_{spare} to $TXOP$ but avoids allocating resources when not necessary. Instead, DTH, using a statistical estimation of future transmission to provide an even more precise $TXOP$ considering the traffic trend, is more tailored on that and acts on the long-term fairness. Indeed, since DTH takes action in the case of VBR traffic, where data rate can vary (increases but also decreases with respect to the mean data rate); sometimes a station can use completely T_{spare} , sometimes it can partially release the remaining part, or sometimes it can produce a T_{spare} ; thus, in general, it is not

possible to predict the behavior of the different schedulers in the brief-term. However, due to random distribution of the mentioned events, it is possible to consider the system as a long-term fairness system that tries to recover resources from stations that do not use that and reassign that to stations requiring more capacity.

This consideration is proved by the fact that UTSS outperforms the refined DTH: in the steady state the system is fully self-sustaining, the station receives the reclaimed resources and releases the excess of resources, and the correction actions to dispatch the traffic are minimal.

5. Conclusions

In this paper *Dynamic TXOP HCCA* (DTH) and its improved version *DTH with threshold* are presented and analysed. Both of these algorithms are conceived as scheduling modules suitable to be added to centralized HCCA schedulers in order to improve their performances in the case of Variable Bit Rate Traffic, such as multimedia applications. The main idea, alternative to overprovisioning solutions, is combining together the concepts of reclaiming transmission time and statistical estimation of the traffic profile in order to provide, at each polling, an instantaneous $TXOP$ tailored to the variable traffic requirements. This allows to overcome the drawback due to the assignment of fixed transmission time during the admission control based on mean values traffic parameters. Indeed, the recovery of not exhausted transmission time is suitable to provide the required additional resources without using overprovisioning that, in general, jeopardises the admission control feasibility test. Furthermore, the use of time series forecasting based on the Moving Average Window for the statistical estimation of the traffic profile allows to assign at each time a transmission duration tailored to the actual service expectation. In particular, DTH with threshold is

even more responsive by recovering faster when *TXOP* show significant variations.

A new theoretical analysis derived by the elastic tasks model, originally used in the processors systems field, highlights how these algorithms, acting on the transmission time modulated as a spring, positively impact on the service rate provided by the scheduler following the data rate variations typical of VBR traffic streams, without modifying the scheduling policy of the centralized scheduler but improving its behavior.

A simulation evaluation has been illustrated, which confirms the analytical results and investigates the schedulers performance. It is focused on the efficiency of the schedulers, delay, and throughput performances. Simulations highlight that the algorithms positively affect the scheduling reducing the experienced null rate but without impacting on the polling time, i.e., on the centralized scheduling policy. Furthermore, the transmission queues length analysis shows that DTH and DTH with threshold are able to reduce the length of the queues, especially in the case of highly variable data rate typical of multimedia applications. As a direct consequence, also the experienced delay results to be reduced as well as the packets drop rate, whereas the system throughput is increased.

Data Availability

The software used for the simulations, NS2, is publicly available at the following URL, cited in the paper, <http://www.isi.edu/nsnam/ns/>. The software extension for HCCA, cited in the paper, is publicly available at the following URL <http://cngl.iet.unipi.it/wiki/index.php/Ns2hcca>. The data set used for the simulations are available at the following URL <http://trace.eas.asu.edu>. In particular, the traces are: VC, Video-conference traffic (LectureHQ-Reisslein trace file), Video streaming MPEG4 trace files of 60 minutes each (VS1: Jurassic Park, VS2: Silence of the lambs, VS3: Mr. Bean).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] IEEE802.11, *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE, Piscataway, NJ, USA, 2016.
- [2] CISCO, "Cisco visual networking index: forecast and methodology. 2014-2019 white paper," Tech. Rep., 2015.
- [3] Y. Cheng, D. Yang, and H. Zhou, "Det-WiFi: a multihop TDMA MAC implementation for industrial deterministic applications based on commodity 802.11 hardware," *Wireless Communication and Mobile Computing*, vol. 2017, Article ID 4943691, 10 pages, 2017.
- [4] Y. Feng, C. Jayasundara, A. Nirmalathas, and E. Wong, "IEEE 802.11 HCCA for tactile applications," in *Proceedings of the 27th International Telecommunication Networks and Applications Conference, ITNAC 2017*, pp. 1-3, Australia, November 2017.
- [5] G. Cecchetti, A. L. Ruscelli, and P. Castoldi, "Wireless spread spectrum for trackside railways communications systems," in *Proceedings of the World Congress of Railway Research (WCRR)*, Milan, Italy, May 2016.
- [6] J. Moreno, J. Riera, L. De Haro, and C. Rodriguez, "A survey on future railway radio communications services: challenges and opportunities," *IEEE Communications Magazine*, vol. 53, no. 10, pp. 62-68, 2015.
- [7] X. Li, D. Li, J. Wan, A. V. Vasilakos, C.-F. Lai, and S. Wang, "A review of industrial wireless networks in the context of Industry 4.0," *Wireless Networks*, vol. 23, no. 1, pp. 23-41, 2017.
- [8] R. Viegas, J. A. Guedes, F. Vasques, P. Portugal, and R. Moraes, "A new MAC scheme specifically suited for real-time industrial communication based on IEEE 802.11e," *Computers & Electrical Engineering, Special Issue on Wireless Systems: Modeling, Monitoring, Transmission, Performance Evaluation and Optimization*, vol. 39, no. 6, pp. 1684-1704, 2013.
- [9] Q. Zhou and P. Ma, "An improved HCCA mechanism for safety critical real time system," in *Proceedings of the 8th IEEE International Conference on Communication Software and Networks (ICCSN)*, Beijing, China, 2016.
- [10] C. Lin, C. Shieh, W. Hwang, and W. Huang, "Proportional bandwidth allocation with consideration of delay constraint over IEEE 802.11e-based wireless mesh networks," *Wireless Networks*, vol. 24, pp. 1575-1592, 2018.
- [11] Z. Wan, X. Yuan, Z. Zhou, and N. Xiong, "Cross-layer optimization of dynamic packet assignment for video transmission over IEEE 802.11e networks," *Wireless Personal Communications*, vol. 102, no. 2, pp. 2417-2428, 2018.
- [12] E. Ku, C. Chung, B. Kang, and J. Kim, "Throughput enhancement with optimal fragmented MSDU size for fragmentation and aggregation scheme in WLANs," in *Proceedings of the 13th International SoC Design Conference, ISOCC 2016*, October 2016.
- [13] S. Seytnazarov and Y.-T. Kim, "QoS-aware adaptive A-MPDU aggregation scheduler for enhanced VoIP capacity over aggregation-enabled WLANs," in *Proceedings of the 2018 IEEE/IFIP Network Operations and Management Symposium, NOMS 2018*, pp. 1-7, Taiwan, April 2018.
- [14] H. Ferng and A. Leonovich, "Periods scheduling under HCCA mode of IEEE 802.11e," *IEEE Transaction on Wireless Communications*, vol. 13, no. 12, pp. 7037-7049, 2014.
- [15] IEEE802.11ac, *802.11ac-2013 - IEEE Standard for Information Technology- Telecommunications and Information Exchange between Systemslocal and Metropolitan Area Networks- Specific Requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications-Amendment 4: Enhancements for Very High Throughput for Operation in Bands below 6 GHz*, IEEE, Piscataway, NJ, USA, 2013.
- [16] IEEE802.11s, *802.11s-2011 - IEEE Standard for Information Technology- Telecommunications and Information Exchange between Systems-Local and Metropolitan Area Networks- Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 10: Mesh Networking*, IEEE, Piscataway, NJ, USA, 2011.
- [17] L. Yuchun, "Ralink rt5370n datasheet," 2010, https://www.acmesystems.it/www/WIFI-2_documentation/RT5370N_datash-eet.pdf.
- [18] A. L. Ruscelli, G. Cecchetti, A. Alifano, and G. Lipari, "Enhancement of QoS support of HCCA schedulers using EDCA function in IEEE 802.11e networks," *Ad Hoc Networks*, vol. 10, no. 2, pp. 147-161, 2012, Special issue on recent advances in analysis

- and deployment of IEEE 802.11e and IEEE 802.11p protocol families.
- [19] M. A. Al-Maqri, M. A. Alrshah, and M. Othman, "Review on QoS provisioning approaches for supporting video traffic in IEEE802.11e: challenges and issues," *IEEE Access*, vol. 6, pp. 55202–55219, 2018.
- [20] C. Cicconetti, L. Lenzini, E. Mingozzi, and G. Stea, "Design and performance analysis of the Real-Time HCCA scheduler for IEEE 802.11e WLANs," *Computer Networks*, vol. 51, no. 9, pp. 2311–2325, 2007.
- [21] G. Cecchetti and A. L. Ruscelli, "Real-time support for HCCA function in IEEE 802.11e networks: a performance evaluation," *Security and Communication Networks*, vol. 4, no. 3, pp. 299–315, 2011, Special issue on security for QoS assured wireless networks.
- [22] A. Grilo, M. Macedo, and M. Nunes, "A scheduling algorithm for QoS support in IEEE802.11 networks," *IEEE Wireless Communications Magazine*, vol. 10, no. 3, pp. 36–43, 2003.
- [23] P. Ansel, Q. Ni, and T. Turletti, "FHCF: a simple and efficient scheduling scheme for IEEE 802.11e wireless LAN," *Mobile Networks and Applications*, vol. 11, no. 3, pp. 391–403, 2006.
- [24] G. Boggia, P. Camarda, L. G. Grieco, and S. Mascolo, "Feedback-based control for providing real-time services with the IEEE 802.11e MAC," *IEEE/ACM Trans. Net*, vol. 2, no. 15, pp. 323–333, 2007.
- [25] H. Sounni, N. Elkamoun, and F. Lakrami, "Performance evaluation of 802.11 access methods with qos enhancement under mobility impact," *International Journal of Computer Applications*, vol. 179, no. 2, pp. 36–41, 2017.
- [26] K. Ju and K. Chung, "Dynamic TXOP allocation for multimedia QoS providing over wireless networks," in *Proceedings of the 27th International Conference on Information Networking, ICOIN 2013*, pp. 397–401, Thailand, January 2013.
- [27] M. A. Al-Maqri, M. Othman, B. M. Ali, and Z. M. Hanapi, "Adaptive TXOP assignment for QoS support of video traffic in IEEE 802.11e networks," in *Proceedings of the 2013 5th IEEE International RF and Microwave Conference, RFM 2013*, pp. 144–149, Malaysia, December 2013.
- [28] A. M. Mansoor, M. A. Al-Maqri, A. Q. Sabri, H. Jalab, A. W. A. Wahab, and W. K. Al-Kopati, "A feedback-based admission control unit for QoS provision of video transmission over WLANs," in *Proceedings of the 7th IEEE Annual Computing and Communication Workshop and Conference, CCWC 2017*, USA, January 2017.
- [29] M. A. Al-Maqri, M. Othman, B. M. Ali, and Z. M. Hanapi, "Adaptive multi-polling scheduler for QoS support of video transmission in IEEE 802.11e WLANs," *Telecommunication Systems*, vol. 61, no. 4, pp. 773–791, 2016.
- [30] M. A. Al-Maqri, M. Othman, B. M. Ali, and Z. B. Hanapi, "Feasible HCCA polling mechanism for video transmission in IEEE 802.11e WLANs," *Wireless Personal Communications*, vol. 85, no. 3, pp. 773–791, 2015.
- [31] C. L. Liu and J. W. Layland, "Scheduling algorithms for multiprogramming in a hard-real-time environment," *Journal of the ACM*, vol. 20, pp. 46–61, 1973.
- [32] A. Grilo, M. Macedo, and M. Nunes, "A scheduling algorithm for QoS support in IEEE 802.11e networks," *IEEE Wireless Communications*, vol. 10, no. 3, pp. 36–43, 2003.
- [33] T. P. Baker, "Stack-based scheduling for real-time processes," *Real-Time Operating System*, vol. 3, no. 1, pp. 67–99, 1991.
- [34] W. K. Lai, C. Shien, and C. Jiang, "Adaptation of HCCA/EDCA ratio in IEEE 802.11 for improved system performance," *International Journal of Innovative Computing*, vol. 5, no. 11(B), pp. 4177–4188, 2009.
- [35] S. V. Azhari, Ö. Gürbüz, O. Ercetin, M. H. Daei, H. Barghi, and M. Nassiri, "Delay sensitive resource allocation over high speed IEEE802.11 wireless LANs," *Wireless Networks*, 2018.
- [36] A. L. Ruscelli and G. Cecchetti, "A IEEE 802.11e HCCA scheduler with a reclaiming mechanism for multimedia applications," *Advances in Multimedia*, vol. 2014, Article ID 372693, 2014.
- [37] A. L. Ruscelli and G. Cecchetti, "Improving the QoS of IEEE 802.11e networks through imprecise computation," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 23, no. 3-4, pp. 152–167, 2016.
- [38] G. Cecchetti, A. L. Ruscelli, A. Mastropaolo, and G. Lipari, "Dynamic TXOP HCCA reclaiming scheduler with transmission time estimation for IEEE 802.11e real-time networks," in *Proceedings of the 15th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, MSWiM 2012*, pp. 239–246, Cyprus, October 2012.
- [39] G. C. Buttazzo, G. Lipari, and L. Abeni, "Elastic task model for adaptive rate control," in *Proceedings of the 1998 19th IEEE Real-Time Systems Symposium*, pp. 286–295, December 1998.
- [40] "Network Simulator 2," <http://www.isi.edu/nsnam/ns/>.
- [41] C. Cicconetti, L. Lenzini, E. Mingozzi, and G. Stea, "A software architecture for simulating IEEE 802.11e HCCA," in *Proceedings of the IPS MoMe*, pp. 97–104, Warsaw, Poland, 2005.
- [42] M. e. a. Reisslein, "Mpeg4 movies traces files," 2000, <http://trace.eas.asu.edu/TRACE/ltvt.html>.
- [43] J. R. Jackson, "Scheduling a production line to minimize maximum tardiness," MAGNT Research Report 43, Management Science Research Project, University of California, Los Angeles, Calif, USA, 1955.

