Performance Analysis of Distributed Resource Reservation in IEEE 802.11e-Based Wireless Networks

Xiaobo Yu, Pirabakaran Navaratnam, and Klaus Moessner

Centre for Communication Systems Research
University of Surrey
Guildford, GU2 7XH, UK
Email: {x.yu, p.navaratnam, k.moessner}@surrey.ac.uk

***Abstract*—Guaranteeing quality of service is one of the most critical challenges in IEEE 802.11-based wireless networks. This paper proposes an analytical framework to evaluate hybrid MAC scheduling mechanisms with distributed resource reservation, that was proposed for the IEEE 802.11e enhanced distributed channel access protocol for guaranteeing quality of service. The hybrid MAC scheduling mechanisms split the airtime into service intervals with contention-free period for quality of service guaranteed real-time sessions, and contention access period for other traffic sessions. The distributed resource reservation ensures that the resources are allocated to real-time sessions without the support of a centralised controller - this makes it suitable for ad-hoc networking applications. The proposed analytical framework models the quality of service (i.e. delay and throughput) performance of real-time sessions with dedicated resources in a distributed environment, and also estimates the overall capacity of the network. Moreover, the derived models can be used to investigate the impact of changes to individual system parameters, such as service interval or size of transmission opportunity. The simulation results show that the proposed analytical framework precisely models the quality of service performance of real-time sessions and predicts the optimum resource allocation for improved network capacity.**

***Index Terms*—quality of service, IEEE 802.11e, resource reservation;**

I. Introduction

IEEE 802.11-based wireless technology is ubiquitous. Owing to advantages such as low cost, robustness, and increased coverage, the technology can be deployed in many applications such as Wi-Fi networks, ad-hoc wireless networks, vehicular ad-hoc networks, wireless mesh networks, etc. Since multimedia applications become more and more popular, supporting real-time services such as video and voice is regarded a necessity. However, this is challenging in 802.11-based wireless networks. One of the main problems is how to satisfy increasing demand of bandwidth to Real-Time Sessions (RTSNs) in order to guarantee their performance, for instance, delay bounds and throughput.

This paper proposes an analytical framework for distributed hybrid MAC Resource Reservation (RR) schemes (i.e. hybrid MAC scheduling + distributed RR) for Enhanced Distributed Channel Access (EDCA), and models the delay and throughput of RTSNs under different traffic loads. Based on this, the network capacity (i.e. the maximum amount of RTSNs) with guaranteed Quality of Service (QoS) is investigated under different system parameters, and traffic conditions.

The IEEE 802.11e standard specifies two channel access mechanisms: EDCA and Hybrid coordination function Controlled Channel Access (HCCA). As a preliminary QoS support mechanism, EDCA has been chosen as the key channel access method for mesh coordination function in IEEE 802.11s [1] which is a new amendment in the IEEE 802.11 standard family [2] for wireless mesh networking. Different from the legacy Distributed Coordination Function (DCF), EDCA can provide a rudimentary scheduling mechanism operated by differentiated services for distinct traffic types. Using EDCA, QoS enhancement can be achieved [3]. However, this prioritisation does not guarantee the required QoS. Inter-node collision can still happen because of the hidden terminal problem and is more grievous in the presence of high traffic load. In contrast to EDCA, the centralised channel access mechanism HCCA can inherently render guaranteed QoS for RTSNs by reserving dedicated Transmission Opportunities (TXOPs). It can be further developed to support handoff latency improvement for the nodes with mobility function [4], [5]. However, complexity of HCCA makes it difficult to be implemented in a distributed network environment such as ad-hoc wireless networks [6], due to lack of a centralised controller. Compared with HCCA, EDCA is a preferable mechanism that can be easily implemented in distributed wireless networks.

Due to the inherent shortcomings of HCCA, many approaches to provide QoS for RTSNs in distributed environment use EDCA. The aim of EDCA enhancements is to minimize collisions posed by interference and to reduce uncertainty of channel access incurred by deferral and back-off algorithms. If a node can not sense the ongoing transmission of a neighbouring node, they may interfere with each other and degrade the QoS performance of the network. The deferral and back-off algorithms proposed in the legacy standard can help reduce the issue of collisions. However, they are not helpful in a situation where high density of transmitting nodes exist. Even if the amount of senders within a shared interference range is limited, high arrival rates can still degrade the ratio of successful transmissions in a contention-based environment [7]. In addition, waiting time posed by deferral and back-off themselves can be regarded as a scheduling overhead which affects the QoS of RTSNs.

Hence, to support guaranteed QoS in a distributed environment, dedicated RR in a distributed manner is indispensable. The purpose of RR is to guarantee QoS for RTSNs by reserving resources (i.e. bandwidth) at all intermediate nodes along the traffic route from source to destination [8]. The bandwidth is also deprived from the off-route nodes which are located within the interference range of the tagged route. Using explicit signalling or piggyback mechanism, the distributed RR schemes can assign dedicated and periodic bandwidth to the RTSNs in order to meet their QoS requirements. Meanwhile, by announcing the reservation information to the neighbouring nodes, these mechanisms can help to prevent the violation of QoS assurances in the transmissions of RTSNs. The explicit signalling or the piggyback mechanism can take charge of negotiating the state information of reservation among all nodes which need to be informed. These include the nodes along the traffic route and the nodes that are located within the interference range of the tagged route [9], [10].

The investigations in this paper focus on hybrid MAC RR schemes for EDCA. In general, hybrid MAC RR schemes introduce contention-free channel access for RTSNs with strict QoS requirements while maintaining a certain duration of contention-based channel access for other types of traffic sessions in order to ensure their sustainable services. Airtime is split into consecutive Service Intervals (SIs), each of which contains both Contention-Free Period (CFP) and Contention Access Period (CAP). The scheduler of the distributed RR scheme can be part of the MAC layer, with the support of explicit signalling messages that propagate among the nodes in order to determine and configure the reservation parameters.

The aforementioned distributed RR method has been developed and evaluated for example in [9] and [10]. However, previous works did not consider the mathematical analysis of this concept, which is needed to prove its effectiveness and accuracy aspects. Some related analytical works focusing on hybrid MAC schemes can be found for centralised solutions such as Point Coordination Function (PCF) and HCCA. For instance, the authors of [11] propose a delay model, considering arbitrary amount of users, packet arrival rate, as well as packet size for the sessions following contention-free channel access in PCF. However, the model based on PCF assumes that in each CFP, only one frame can be transmitted by a node. The proposed model in [12] considers telephony traffic under contention-free environment in PCF. This model is devised under the assumption that the size of super-frame can be stretched under the control over the centralised device, which is not practical in distributed RR schemes. The queuing model proposed in [13] for the RTSNs in the CFP with HCCA considers the packet priority using MAP/PH/1 queue. Another model proposed in [14] also utilizes the information of queue size and other parameters to predict the maximum suffered delay. They assume that the queue size of each station can be detected by a centralised controller and thereby portion of the allocated TXOP for a corresponding station can be temporarily utilized by other stations. This can hardly be implemented in distributed RR schemes.

The analysis for the hybrid MAC RR schemes with EDCA is part of the focus of this paper. Apart from the performance analysis, the optimization of network capacity is another open issue and is addressed here. Here, the network capacity refers to the amount of RTSNs that can be served in the CFP given that each of them can obtain satisfactory QoS. The bandwidth needs to be assigned strictly according to the QoS demands of the RTSNs with different data rates, while meeting the delay bounds. However, low network capacity can be the result if the bandwidth allocation in the CFP is not conducted optimally. To study this problem, an analysis to optimize the network capacity with the guaranteed RTSNs is documented in this paper. Built on top of the satisfactory QoS, the optimization can accommodate more RTSNs in the CFP by re-tuning the resources allocated for RTSN as well as utilizing optimized system parameters such as the SI.

The remainder of this paper is organized as follows. Section II specifies the mechanism of hybrid MAC RR for EDCA. The analytical model for delay and throughput performance and the subsequent optimization study are presented in Section III. Simulation and theoretical evaluations are shown and discussed in Section IV. Finally, Section V concludes the paper.

II. Hybrid MAC Resource Reservation in IEEE 802.11e Networks

The hybrid MAC RR refers to the mechanisms that divide the airtime into SIs and allocate transmission time-slots periodically for traffic sessions. For example, as depicted in Fig. 1, the admitted RTSNs will be allocated dedicated resources (i.e. TXOPs) within the CFP periodically, and the other sessions will only be allowed to transmit during the CAP. This kind of RR can be established in a centralised or distributed manner. Because of the widespread applicability of distributed schemes in which resource allocation as well as the admission decision can be made by every node (or super node in some cases) in the network, the distributed RR mechanisms are more appealing for the next generation networks. Thus, this paper considers distributed mechanisms for demonstrating the principles of hybrid MAC RR in IEEE 802.11e networks for guaranteeing the QoS performance of RTSNs.

A. Signalling for Distributed Resource Reservation

To implement the RR in a distributed manner, information regarding the reservation parameters needs to be exchanged between the member nodes using either implicit or explicit signalling process. Parameters including service start time, mean data rate, nominal frame size, and delay bound will be included in a RR signalling message. This RR signalling message is

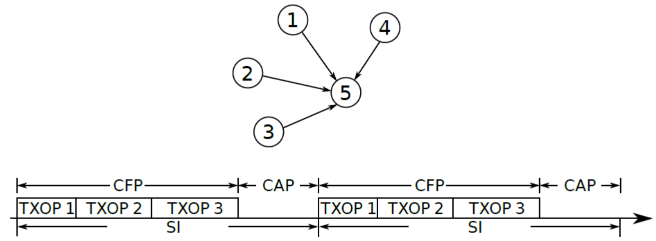


Fig. 1. Hybrid MAC resource reservation scheme

broadcasted from the source node if a new RTSN with strict QoS demand can be admitted into the network. Please note that this admission decision will be made by an Admission Control Algorithm (ACA) as explained in the next sub-section. The nodes receiving the RR signalling message will update their local state information of RR using the information available on this signalling message and align themselves for this new RR scheduling within the CFP. Moreover, this RR will be confirmed by the corresponding destination node on a feedback signalling message. Whenever a RTSN with dedicated TXOP in the CFP finishes its transmission, the allocated resources for this session will be released by the nodes with a notification from the source node. By this way, each node sharing overlapping transmission range can be well synchronized for the RR scheduling in a distributed manner, and thus the strict QoS demands of RTSNs can be satisfied in distributed environments, like ad-hoc networks.

B. Distributed Admission Control

The RR mechanism needs to be equiped with an ACA. Note that within a fixed size of SI, the transmission time-slots available within a CFP is limited. Therefore, bandwidth reservation in the CFP can not be made for excessive amount of RTSNs. In this case, admission control is a necessity to manage the permission of dedicated RR in the CFP. In centralised wireless networks, the admission decision of RR is always performed by centralised controllers such as access point or based station. A node needs to send request to the centralised controller and can only obtain dedicated bandwidth if the request is granted. In contrast to the centralised mechanisms, in distributed wireless networks, the admission decision can be made by every node (or super node). This can be achieved by knowing the information of previous RR within the CFP, obtained from previous signalling message exchanges, and thereby a node can make decision to admit a newly arrived RTSN if its QoS demand can be satisfied by the available resources in the CFP. Thus, ACA ensures that the existing RTSNs are not affected by newly arriving RTSNs. Moreover, it can administer the priority of QoS provisioning for newly arriving RTSNs within the network based on predefined resource provisioning criteria. Please refer [10] for specific details on such RR and admission control mechanisms.

This paper aims to provide an analytical framework for evaluating such RR mechanisms in IEEE 802.11e wireless networks.

III. Analytical Framework

In this section, an analytical framework that models contention-free access using hybrid MAC RR is proposed. As described in the section II, the primary objective of hybrid MAC RR is to provide guaranteed QoS for RTSNs.

TABLE I
Parameters used in the analysis

Parameter	Definition
$\overline{s_{DATA_i}}$	Mean MSDU size for RTSN$_i$
ΔSI	Duration of a service interval
λ_i	Required scheduling rate for RTSN$_i$
n_i	Number of MSDUs allowed to be transmitted with in a TXOP for RTSN$_i$
t_{TXOP_i}	TXOP duration allocated for RTSN$_i$
t_{ACK}	Transmission time of an ACK frame
t_{SIFS}	Time interval of SIFS
t_{DATA_i}	Duration of an MSDU transmission for RTSN$_i$
$E[t_{DATA_i}]$	Average transmission time of an MSDU for RTSN$_i$
$d_{ave,i}$	Average delay for RTSN$_i$
$d_{ca,i}$	Channel access delay for RTSN$_i$
$d_{q,i}$	Queuing delay for RTSN$_i$
$d_{tr,i}$	Transmission delay for RTSN$_i$
μ_i	Sending interval for RTSN$_i$
$\zeta(j)$	Normalized offset of the arrival time for the j^{th} MSDU
$\eta(j)$	Offset duration for the j^{th} MSDU
t_{DA_i}	Transmission duration of an MSDU of RTSN$_i$ including its consequent ACK frame
σ	Duration of a slot time
R	Physical transmission rate
$\lambda_{r,i}$	Reserved scheduling rate for RTSN$_i$
Φ_i	Interval $[0, t_{TXOP_i} - t_{DA_i}]$ in a TXOP for RTSN$_i$
$dc_{j,i}$	Channel access delay for the j^{th} MSDU for RTSN$_i$
p_i	Minimum period for RTSN$_i$
$dq_{j,i}$	Queuing delay of j^{th} MSDU for RTSN$_i$
$\overline{S_i}$	Average throughput of RTSN$_i$ with TXOP
$\overline{nd_i}$	Average amount of MSDUs for RTSN$_i$ that have been transmitted in each TXOP
O_i	Transmission overhead caused by MAC header
$d_{rmax,i}$	Delay bound of RTSN$_i$

In such RR mechanisms, TXOPs will be reserved for the corresponding RTSNs within the CFP. Instead of transmitting each data frame individually within a reserved TXOP, several data frames can be encapsulated into MAC Service Data Units (MSDUs) and transmitted together if the corresponding TXOP allows. This is useful for improving the throughput and bandwidth utilization. Note that the default TXOP specification is included in IEEE 802.11e standard. It suggests that the number of MSDUs allowed to be transmitted within a TXOP for RTSN$_i$ depends on the mean MSDU size $\overline{s_{DATA_i}}$, duration of SI represented by ΔSI and required scheduling rate λ_i[1]. Let n_i be the number of MSDUs that are allowed to be transmitted in a single TXOP for a RTSN$_i$. It can be given by:

$$n_i = \lceil \frac{\Delta SI \times \lambda_i}{\overline{s_{DATA_i}}} \rceil \tag{1}$$

As mentioned in section II, the allocated TXOP for each RTSN subjects to the QoS requirement embedded in the signalling

[1] In this paper, it is assumed that the required scheduling rate is equivalent to the application rate.

message. Thus, sessions will secure TXOPs with distinct durations according to their QoS demands. In order to study and evaluate the performance of RTSNs with different rates, we formulate delay and throughput performance of a QoS guaranteed RTSN in the following sub-sections. Please refer to Tab. I for the parameters used in the subsequent sections.

A. Delay Model for RTSNs with TXOPs

In this sub-section, we model the delay performance of RTSNs with TXOPs. Note that, here the average delay experienced by a QoS guaranteed RTSN$_i$ is defined as the mean duration from the instant that its MSDUs arrive the MAC interface queue of the source node to the instant that they are successfully transmitted. Thus, the average delay $d_{ave,i}$ experienced by session RTSN$_i$ can be determined by [15]:

$$d_{ave,i} = \overline{d_{ca,i} + d_{q,i} + d_{tr,i}} \tag{2}$$

where $d_{ca,i}$ is the channel access delay, $d_{q,i}$ is the queuing delay, and $d_{tr,i}$ is the transmission delay. For contention-free channel access with pre-fixed SI, each of the delay can be assumed to be independent of each other. Therefore, the above expression can be transformed into:

$$d_{ave,i} = \overline{d_{ca,i}} + \overline{d_{q,i}} + \overline{d_{tr,i}} \tag{3}$$

1) Channel access delay: The channel access delay is measured from the time when an MSDU arrives at the head of the interface queue to the moment that it begins accessing the channel. Since each MSDU will wait for its dedicated TXOP to be transmitted, the channel access delay for each MSDU depends on the instant when it arrives the head of the queue.

As mentioned before, each admitted RTSN will have one TXOP in each CFP. Therefore from a RTSN point of view, the airtime can be perceived as periodical TXOPs and non-TXOPs. Note that during non-TXOP durations, the RTSN will wait for its next TXOP for data transmissions. Based on the traffic load of a RTSN and the size of its reserved TXOP duration, there can be three transmission states: (i) unsaturated transmission state indicates that the traffic load of the RTSN is less than its reserved TXOP capacity and results in bandwidth wastage, (ii) saturated transmission state implies that the traffic load can exactly fit its reserved TXOP capacity without any bandwidth wastage, and (iii) over-saturated transmission state indicates that the traffic load exceeds the reserved TXOP capacity. In hybrid MAC RR schemes, the resources (i.e. TXOP duration) reserved for each RTSN are based on its QoS demand. Since n_i is set as the upper bound integer value calculated by (1), the resources allocated for each RTSN will not be less than its QoS demand. Please note that a RTSN will only be allowed to transmit in the CFP if there are enough resources to be reserved for it. Further, the reserved resources can exactly meet the requirement of the RTSN in some cases in which $\frac{\Delta SI \times \lambda_i}{s_{DATA_i}}$ is an integer value. In other cases, the reserved resources will be slightly higher than the demand. Hence, only the saturated and the unsaturated transmission states will be the focus here in this paper.

As a key parameter, the instant that an MSDU reaches the head of queue is dependent on the MSDU arrival rate. Let μ_i be the sending interval and it can be computed as:

$$\mu_i = \frac{\overline{s_{DATA_i}}}{\lambda_i} \tag{4}$$

As shown in Fig. 2, sending interval can map the arrival time of each MSDU to an instant within a SI.

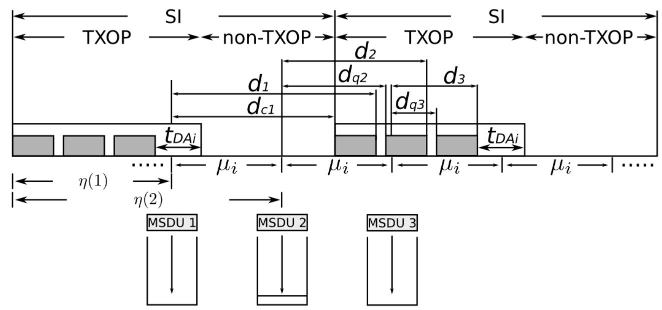


Fig. 2. Channel access delay and queuing delay for different MSDU arrival instances

Note that the arrival time of each MSDU determines its channel access delay. Let $f = \delta(x)$ be a function and is defined as:

$$\delta(x) = x - [x] \tag{5}$$

where [x] represents the integer part of variable x. Using $\delta(x)$, we can figure out the normalized offset $\zeta(j)$ of the arrival time for the j^{th} MSDU based on the start time of its SI.

$$\zeta(j) = \delta(\frac{j \cdot \mu_i}{\Delta SI}) \tag{6}$$

The arrival offset duration of the j^{th} MSDU $\eta(j)$ within its SI can be denoted by:

$$\eta(j) = \delta(\frac{j \cdot \mu_i}{\Delta SI}) \cdot \Delta SI \tag{7}$$

An example of arrival offset is shown in Fig. 2. It illustrates the arrival offset durations of the MSDU 1 and MSDU 2. By obtaining the offset duration of each MSDU's arrival time, the channel access delay can be precisely computed as follows.

Note that the head-of-line MSDU can only be transmitted in a TXOP if the residual duration left in the TXOP is sufficient for its transmission. Let t_{DA_i} be the total transmission time of an MSDU including its consequent ACK frame, and it can be determined by:

$$t_{DA_i} = E[t_{DATA_i}] + t_{SIFS} + t_{ACK} + t_{SIFS} \tag{8}$$

where t_{ACK} and t_{SIFS} stand for the duration of ACK transmission and cost of SIFS respectively. Since the size of each MSDU may vary, $E[t_{DATA_i}]$ is utilized for denoting the average transmission duration of an MSDU of RTSN$_i$.

Under the unsaturated and saturated transmission states, the MSDUs arriving inside the interval $[0, t_{TXOP_i} - t_{DA_i}]$ of a

TXOP can be transmitted within the current TXOP. This is because these states will align to the condition that there is no MSDU buffered in the queue at the instant $t_{TXOP_i} - t_{DA_i}$ within each TXOP. This condition can be justified as follows.

Let's assume that there still have MSDUs buffered in the queue at the instant of $t_{TXOP_i} - t_{DA_i}$ in a TXOP under saturated and unsaturated transmission states. Then the buffered MSDUs have to wait for the next TXOP in order to get transmitted. This means that the MSDUs generated within each SI can not be completely served by the reserved resources and therefore this case will represent the over-saturated transmission state, neither the saturated nor the unsaturated states.

Fig. 2 shows an example of channel access delay for the MSDUs at different arrival instances with respect to their TXOP durations. Let $dc1$ indicate the channel access delay of MSDU 1 and the total delay for these MSDUs are denoted by $d1$, $d2$, $d3$, respectively. Let Φ_i be the duration of $t_{TXOP_i} - t_{DA_i}$. The MSDU 1 first arrives after the instant Φ_i of its TXOP and directly reaches the head of queue. Therefore, it only has channel access delay which is equal to its channel waiting time before the start of its next TXOP. The MSDU 2 arrives inside the non-TXOP duration and it has to buffer in the queue before being transmitted. Therefore, its total delay does not have channel access delay. Likewise, the delay of MSDU 3 is also irrelevant to channel access delay. Note that the transmission duration represented by the grey block in Fig. 2 includes an MSDU transmission duration, a SIFS as well as an ACK transmission time.

Based on the above argument, the tagged MSDU within a SI can be classified into one of three categories. First, the MSDU arrives inside the interval $[0, \Phi_i]$ of its TXOP duration. Second, the MSDU arrives within the interval $(\Phi_i, \Phi_i + \mu_i]$ of the SI. Third, the MSDU arrives within the interval $(\Phi_i + \mu_i, \Delta SI)$ of the SI. The MSDU falling under the first category can be transmitted in the current TXOP after being queued until its prior MSDUs have been sent out. Consequently, it has no channel access delay.

Moreover, the MSDU within the second category will be the head-of-line MSDU and wait for the next TXOP to be transmitted. Its channel access delay is equal to the duration from the instant that it becomes the head-of-line MSDU to the start time of its next TXOP. Finally, the MSDU falling within the third category will buffer in the interface queue and wait for its reserved TXOP in the next SI. Thus, it has no channel access delay.

Consequently, in general the channel access delay of an MSDU can be determined as:

$$dc_{j,i} = \begin{cases} 0, & if \;\; 0 \le \eta(j) \le \Phi_i \\ (1 - \frac{\eta(j)}{\Delta SI}) \cdot \Delta SI, & if \;\; \Phi_i < \eta(j) \le \mu_i + \Phi_i \\ 0, & if \;\; \mu_i + \Phi_i < \eta(j) < \Delta SI \end{cases} \tag{9}$$

To compute the average channel access delay for a RTSN, the channel access delay of all the MSDUs have to be taken into account. This will dramatically increase the computational complexity. In order to simplify the calculation of the average channel access delay, we can prove that the $\eta(j)$ is a periodic function which implies that the arrival offset duration of an arbitrary MSDU will cyclically reappear. This can be proved as follows.

Proof: Consider that there exists an integer P, which represents the subsequent P^{th} MSDU from the j^{th} MSDU. Using (7), the arrival time of the P^{th} MSDU can be denoted by:

$$\begin{aligned} \eta(j+P) &= \delta(\frac{(j+P)\cdot\mu_i}{\Delta SI})\cdot\Delta SI \\ &= \delta(\frac{j\cdot\mu_i}{\Delta SI} + \frac{P\cdot\mu_i}{\Delta SI})\cdot\Delta SI \end{aligned} \tag{10}$$

The size of SI can be denoted in terms of slot-time σ as:

$$\Delta SI = K\cdot\sigma, \quad where \;\; K \in \mathbf{N}^+ \tag{11}$$

Note that a node is only permitted to access the channel at the beginning of a slot-time σ [16]. In addition, sending interval can also be represented in terms of σ as:

$$\mu_i = K'\cdot\sigma, \quad where \;\; K' \in \mathbf{N}^+ \tag{12}$$

By substituting (11) and (12) into (10), we can obtain:

$$\eta(j+P) = \delta(\frac{j\cdot\mu_i}{\Delta SI} + \frac{P\cdot K'}{K})\cdot\Delta SI \tag{13}$$

Since K' and K are integers, there exists a minimum P that can make $\frac{P\cdot K'}{K}$ as a positive integer. This implies that there exists a relationship:

$$\eta(j) = \eta(j+P), \quad where \;\; \frac{P\cdot K'}{K} \in \mathbf{N}^+ \tag{14}$$

Thus, the value of arrival offset duration for an arbitrary MSDU will appear periodically after certain subsequent number of MSDUs. Since the MSDUs sharing the same arrival offset duration have the identical channel access delay, the average channel access delay can be calculated by averaging all the channel access delay of MSDUs arrived within a certain repeating period. Consequently, the average channel access delay can be given by:

$$\overline{d_{ca,i}} = \frac{\sum_{j=1}^{p_i} dc_{j,i}}{p_i} \tag{15}$$

where p_i represents the minimum repeating period for the MSDUs of a RTSN$_i$.

2) Queuing delay: The queuing delay is defined as the period from the instant an MSDU arrives the interface queue until the moment it reaches the head-of-line of the queue for transmission.

For the saturated and the unsaturated transmission states, queuing delay can be determined from the required scheduling rate, size of SI, and the size of TXOP duration allocated for the corresponding RTSN. Note that as mentioned in the previous sub-section, there is no MSDU buffered in the queue at the instant of Φ_i in each TXOP for the saturated and unsaturated transmission states.

Fig. 2 shows an example of queuing delay for the MSDUs at different arrival instances with respect to their TXOP durations. The MSDU 1 arrives within the last t_{DA_i} of its TXOP and therefore buffers at the head of the queue. After the duration of μ_i, MSDU 2 arrives during non-TXOP duration and it buffers behind MSDU 1. The total delay of MSDU 2 is shown as d_2 in Fig. 2 and its queuing delay is the duration d_{q2} which is measured from the moment that it buffers in the queue to the instant that its prior MSDU 1 finishes transmission. Likewise, the queuing delay of MSDU 3 which arrives during the transmission of MSDU 2 is illustrated as d_{q3}.

In general, the queuing delay of an MSDU can be formulated into three cases based on the MSDU's arrival interval with respect to its TXOP duration as depicted in Fig. 3.

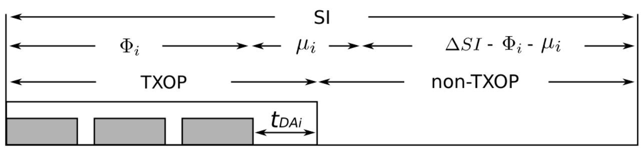


Fig. 3. Queuing delay on different conditions

An MSDU that arrives within the interval $[\Phi_i, \Phi_i + \mu_i]$ will directly reach the head-of-line and will be transmitted first at the next TXOP. Thus, it will experience no queuing delay.

If the MSDU arrives the interface queue during the interval $[0, \Phi_i)$, the queuing delay will be the total transmission time of its prior MSDUs buffered in the queue plus the residual transmission time of the current transmitting MSDU.

If the MSDU arrives within the interval $(\Phi_i + \mu_i, \Delta SI)$, the queuing delay is the transmission time of the prior MSDUs buffered in the queue plus the waiting time during the non-TXOP period.

Consequently, the queuing delay of an MSDU can be expressed by (16).

Therefore, the average queuing delay $\overline{d_{q,i}}$ of a RTSN will be:

$$\overline{d_{q,i}} = \frac{\sum_{i=1}^{p_i} dq_{j,i}}{p_i} \tag{17}$$

3) Transmission delay: Transmission delay is equivalent to the duration from the instant that an MSDU begins accessing the channel to the moment it is successfully transmitted. It can be denoted by:

$$d_{tr,i} = E[t_{DATA_i}] + t_{SIFS} + t_{ACK} + t_{SIFS} = t_{DA_i} \tag{18}$$

B. Throughput Model for RTSNs with Reserved Resources

The throughput of a RTSN with reserved resources depends on the amount of MSDUs that can be transmitted in each of its TXOP. Note that, under saturated and unsaturated transmission states, the MSDUs arriving within a SI duration will all be transmitted within the same duration. Thus, the average throughput $\overline{S_i}$ of RTSN_i with reserved resources can be expressed by:

$$\overline{S_i} = \begin{cases} \frac{\overline{nd_i} \times \overline{S_{data_i}}}{\Delta SI} \approx \lambda_i, & if \quad \lambda_i \leq \lambda_{r,i} \\ \frac{n_i \times \overline{s_{DATA_i}}}{\Delta SI}, & if \quad \lambda_i > \lambda_{r,i} \end{cases} \tag{19}$$

If the required scheduling rate exceeds the maximum transmission capacity of the reserved resources, throughput will be bounded and the value is equal to the maximum transmission capacity of the allotted TXOPs.

C. System Parameters Optimization Model

In this sub-section, we investigate the impact of changes of system parameters such as TXOP duration as well as the size of SI in order to maximize the transmission capacity of the network for RTSNs. Since the size of SI is constraint by the delay bounds to be met for RTSNs (note that RTSNs are mostly time-intolerant), the reservation for high data rate RTSNs can dramatically degrade the network capacity for accommodating more RTSNs. Furthermore, excessive bandwidth reservation beyond the QoS requirement (i.e. both throughput and delay bound) for a RTSN can also waste the network capacity. How to balance the trade-off between maximum number of RTSNs and guaranteeing QoS for each RTSN is an open issue. Smaller SI enhances the scheduling rate of each allocated TXOP and thus can accommodate RTSNs with higher required scheduling rates. Moreover, with small SI, delay bounds can be met for RTSNs because of short non-TXOP duration. However, due to limited CFP of small SI, maximum number of RTSNs capable of reserving bandwidth is limited. This issue can be alleviated by extending the size of SI to a larger value. But for meeting the delay bound, each admitted RTSN has to enlarge its reservation bandwidth. This will again bound the network capacity for more RTSNs. To deal with these problems, an optimization model is indispensable. The model can be used to study and configure the system parameters for RTSNs and thereby optimally utilize the network capacity. Thus, the objective of the optimization problem is to allocate maximum number of TXOPs in a SI while guaranteeing the QoS requirement (i.e. throughput and delay bound) of each admitted RTSN within the network.

Note that the reserved scheduling rate $\lambda_{r,i}$ represents the maximum transmission capacity of a TXOP of RTSN_i and it can be expressed by:

$$\lambda_{r,i} = \frac{n_i \times \overline{s_{DATA_i}}}{\Delta SI} \tag{20}$$

While satisfying the throughput requirement, the delay bound $d_{rmax,i}$ of the admitted RTSN_i also needs to be met. i.e.

$$d_{ave,i} \leq d_{rmax,i} \tag{21}$$

Using (3), the above expression can be transformed into:

$$\overline{d_{ca,i}} + \overline{d_{tr,i}} + \overline{d_{q,i}} \leq d_{rmax,i} \tag{22}$$

$$dq_{j,i} = \begin{cases} (\lfloor \frac{\eta(j)+\Delta SI-\Phi_i}{\mu_i} \rfloor - \lfloor \frac{\eta(j)}{t_{DA_i}} \rfloor - 1) \cdot t_{DA_i} + (\lfloor \frac{\eta(j)}{t_{DA_i}} + 1 \rfloor \cdot t_{DA_i} - \eta(j)), & if \ \ 0 \leq \eta(j) < \Phi_i \\ 0, & if \ \ \Phi_i \leq \eta(j) \leq \mu_i + \Phi_i \\ \lfloor \frac{\eta(j)-\Phi_i}{\mu_i} \rfloor \cdot t_{DA_i} + \Delta SI - \eta(j), & if \ \ \mu_i + \Phi_i < \eta(j) < \Delta SI \end{cases} \tag{16}$$

The aforementioned condition in (22) can limit the range of reserved scheduling rates given in (20). As discussed before, the TXOP allocated for a RTSN has to suffice the throughput requirement, reflected by the parameter required scheduling rate. Otherwise, interface queue will overflow, causing unacceptable queuing delay as well as high packet loss ratio. In order to avoid this over-saturated transmission state, the reserved scheduling rate should at least be equal to the required scheduling rate. i.e.

$$\lambda_{r,i} \geq \lambda_i \tag{23}$$

Consequently, the conditions in (22) and (23) should be satisfied for guaranteeing the QoS demand of the RTSN. This can be achieved and optimized with proper settings of system parameters. Apparently, the reserved scheduling rate is influenced by the size of TXOP. In order to study the optimization of network capacity for more RTSNs, the relationship between the required scheduling rate and the size of TXOP needs to be formulated. The formulation can be started with the TXOP duration which is denoted by:

$$t_{TXOP_i} = \frac{\overline{s_{DATA_i}} \cdot n_i + O_i}{R} \tag{24}$$

where O_i indicates the transmission overhead caused by MAC header, ACK frame, and SIFS. Thus, it can be computed by:

$$O_i = n_i(s_{ACK} + O_{mac}) + 2n_i \cdot t_{SIFS} \cdot R \tag{25}$$

By substituting (20) into (24), the relationship between the size of TXOP and the reserved scheduling rate can be given as:

$$t_{TXOP_i} = \frac{\lambda_{r,i} \cdot \Delta SI + O_i}{R} \tag{26}$$

Using (23), the derivative of (26) can be represented by:

$$t_{TXOP_i} \geq \frac{\lambda_i \cdot \Delta SI + O_i}{R} \tag{27}$$

It can be seen from (20) and (27) that n_i has to change according to the size of SI and the required scheduling rate so as to satisfy the delay bound for RTSN$_i$. Note that multiple RTSNs can reserve bandwidth in a CFP. In order to figure out the optimum allocated resources for each of the admitted RTSN within a CFP, the average TXOP duration $\overline{t_{TXOP}}$ can be introduced and given by:

$$\overline{t_{TXOP}} = \frac{1}{N}\sum_{i=1}^{N} t_{TXOP_i} \tag{28}$$

where N is the number of admitted RTSNs within a CFP.

As mentioned before, the goal of optimization is to allocate resources for maximum number of RTSNs while the QoS requirement of each admitted RTSN is satisfied (i.e. optimize the network capacity). It can be concluded from (28) that the network capacity can be maximized if all the reserved TXOPs are set as the minimum values given that the delay bound requirement is met. The above argument can be expressed in another way by using the average TXOP. The maximum amount of TXOPs and the optimum scheduling rate can be obtained when the $\overline{t_{TXOP}}$ is the minimum value given that the delay bound is satisfied. Thus, this optimization model can be represented by the following function.

$$\begin{aligned} f(\lambda_{r,i}) &= min\{\overline{t_{TXOP}}\}, \ \ if \ \ d_{ave,i} \leq d_{rmax,i} \\ &= min\{\frac{\sum_{i=1}^{N} t_{TXOP_i}}{N}\}, \ \ if \ \ d_{ave,i} \leq d_{rmax,i} \end{aligned} \tag{29}$$

By substituting (26) into (29), we can get a variant of the optimization function which is denoted by:

$$f(\lambda_{r,i}) = min\{\frac{\sum_{i=1}^{N}(\lambda_{r,i} \cdot \Delta SI + O_i)}{N \cdot R}\}, if \ d_{ave,i} \leq d_{rmax,i} \tag{30}$$

The equation (30) indicates that given the delay bound, the optimum resource allocation can be achieved when each reserved scheduling rate is taken as the minimum value. If the delay bound does not pose extra demand to the reserved bandwidth, the minimum value of (30) can be $\frac{\sum_{i=1}^{N}(\lambda_i \cdot \Delta SI + O_i)}{N \cdot R}$, as indicated in (27).

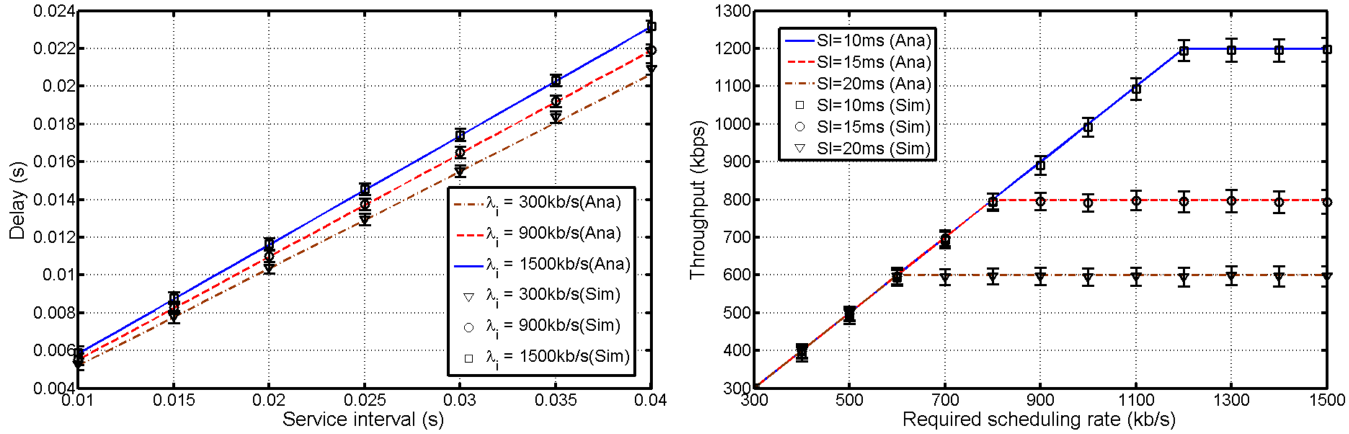
Under the optimum scheduling rate and the guaranteed delay bound, the optimum SI can be achieved when the occupied resources reserved by the existing RTSNs get the minimum proportion of the CFP. Thus the optimum SI can be formulated by the following function.

$$g(\Delta SI) = min\{\frac{\sum_{i=1}^{N} t_{TXOP_i}}{\Delta SI}\}, \quad if \ d_{ave,i} \leq d_{rmax,i} \tag{31}$$

It implies that given the amount of TXOPs, the optimum SI has the minimum proportion of occupied CFP resources so that the residual bandwidth can be maximized in order to accommodate additional RTSNs, i.e. supporting more RTSNs and thus optimize the network capacity utilization for more RTSNs.

IV. Performance Evaluation

In this section, the proposed analytical model is investigated and verified using *ns-2* simulations. The optimization



(a) Delay under different service intervals

(b) Throughput under different required scheduling rates

Fig. 4. Simulation and analytical result of analytical model

results are also shown and discussed in this section. Tab. II summarizes the parameters used in the evaluation. The network topology consists of several nodes with one RTSN per-node. All the senders are randomly deployed within each other's transmission range. Note that MAC and physical layer parameters are selected in accordance with IEEE 802.11b standard. Nodes within the networks utilize fixed transmission power of 281mW. This results in 250m transmission range using *ns-2*'s standard channel model. The interface queue size is set to 50 by default using *ns-2* model. In order to obtain statistically meaningful results, all the results are taken over 20 simulation runs and the mean values are computed with 98% confidence interval. Each simulation is run for over 500s simulated time. This applies to all the simulation results in this paper.

TABLE II
SIMULATION PARAMETERS

Parameter(units)	Value
SIFS(μs)	10
Slot time(μs)	20
ACK size(bytes)	28
MAC header(bytes)	36
Channel capacity(Mbps)	11
Interface queue size(packets)	50
Transmission range(m)	250
Traffic application	CBR over UDP

In order to prove the accuracy of the analytical model, a set of simulations has been conducted. Fig. 4a demonstrates the delay performance of QoS guaranteed RTSNs for different SIs. Fig. 4b presents the throughput of RTSNs with different required scheduling rates. It can be noted from Fig. 4 that the analytical results have a good agreement with the simulation outcomes. Fig. 4a indicates that RTSNs with higher required scheduling rates receive higher delay. This is because the queuing delay increases along with the traffic load.

Beside that, the results from Fig. 4a also show the tendency that the delay increases with the increment of the size of SI. The reason is that non-TXOP duration augments the waiting time of each MSDU. Fig. 4b depicts that the throughput of a RTSN with TXOP increases with the increment of its required scheduling rate, and the throughput saturates when the traffic load can not be sufficiently served by the allocated bandwidth.

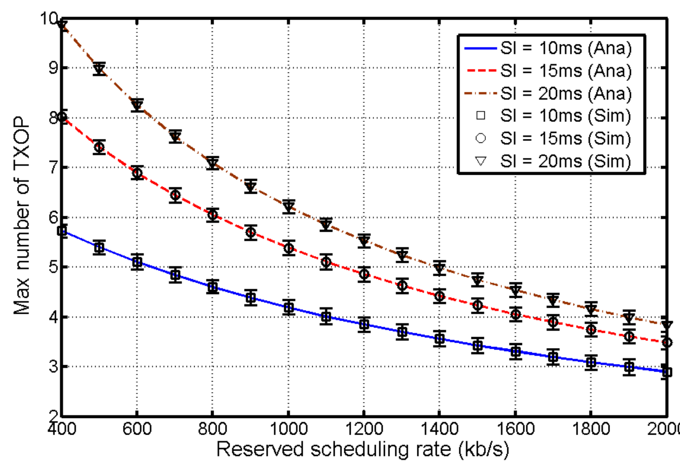


Fig. 5. Maximum TXOP allocation under different reserved scheduling rates

As discussed in the previous section, reserved scheduling rate has to be at least the required scheduling rate of a RTSN in order to satisfy its QoS demand. However, it does not indicate that the excessive bandwidth can be allocated to it. Resource allocation for each RTSN needs to be optimized in order to maximize the network capacity utilization. According to the conclusion in Section III-C, the optimal usage of network capacity can be obtained when the optimum scheduling rate is achieved by taking the minimum value which can just meet the delay bound. As shown in Fig. 6, the optimum scheduling rates are 400kb/s which is equal to the required scheduling rates of RTSNs if delay bounds are satisfied. It can also be indicated that as the reserved scheduling rates for the existing RTSNs increase, the network capacity is getting worse. Fig. 5 shows the theoretical and the simulation outcomes of the maximum number of TXOP allocation under distinct reserved scheduling rates. Both of the results indicate that the maximum network capacity is fulfilled when the optimum scheduling rates are chosen.

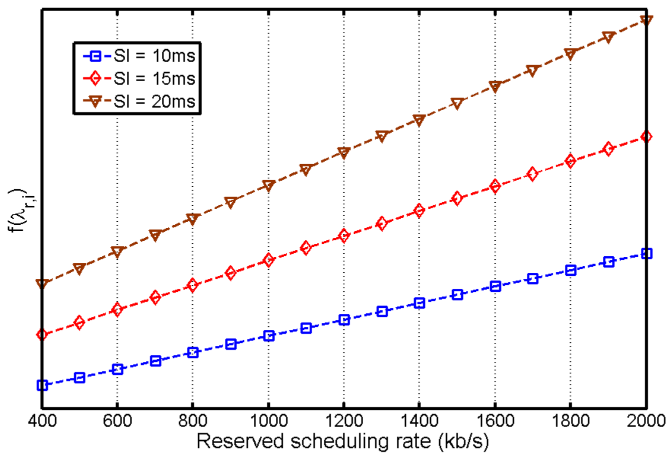


Fig. 6. Optimized TXOP

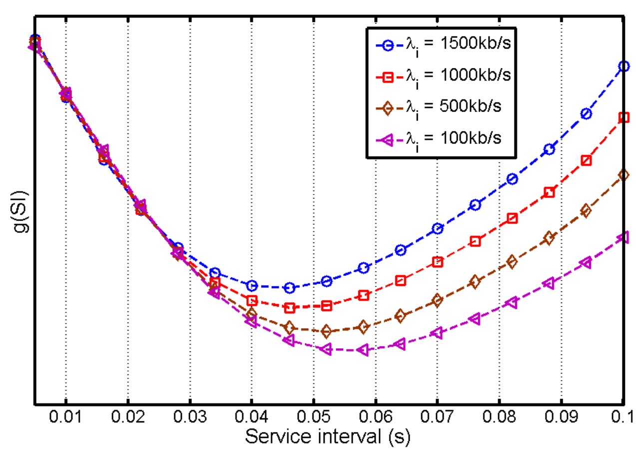


Fig. 7. Optimum SI under different scheduling rates

In order to find out the impact of required scheduling rate to the optimum SI, the delay bound is set fixed as 25ms. Several different required scheduling rates (i.e. 100kb/s, 500kb/s, 1000kb/s, and 1500kb/s) are taken and the optimum bandwidth allocation is applied. Fig. 7 indicates that the optimum SI reduces along with increment of the required scheduling rates of RTSNs. Since the optimum bandwidth allocation largely depends on the required scheduling rate, the desired bandwidth for RTSNs with higher required scheduling rates will grow faster than the RTSNs with lower required scheduling rates when the size of SI increases. This is because delay of high rate RTSN increases more significantly, which leads to more compensation of bandwidth for satisfying delay bound. Fig. 7 implies that the optimum SI can be found for the optimum scheduling rates of RTSNs.

V. Conclusion

The proposed modelling approach for QoS provision in IEEE 802.11-based networks has been evaluated throughly and the simulation results confirm the analytical approach. The model accurately evaluates the delay and throughput performance of RTSNs with reserved TXOPs under different traffic conditions. Based on this analysis, an optimization study has been performed in order to make the hybrid MAC reservation protocol accommodate maximum amount of RTSNs while satisfying the QoS demand of each RTSN. Outcomes of the analysis and the simulation results have validated the accuracy of the analytical model. They can also be used to determine the optimum network capacity as well as the optimum values for system parameters such as SI under different requirements for example traffic condition and delay bound. The evaluation method has shown to be sufficiently generic, an application to other QoS mechanisms and protocols can be anticipated.

VI. Acknowledgement

This work was performed in the project QoSMOS which has received research funding from the EU FP7 framework.

References

[1] IEEE P802.11s/D8.0: 'IEEE Draft Standard for Information Technology-Telecommunications and Information Exchange between Systems-Local and Metropolitan Area Networks-Specific Requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications-Amendment 10: Mesh Networking', 2010

[2] 'IEEE 802.11-2007: IEEE Standard for Wireless LAN Medium Acess Control (MAC) and Physical Layer (PHY) Specification', 2007 (revised version)

[3] Li, C.S. and Tsuei, T.G. and Chao, H.C.: 'Evaluation of Contention-Based EDCA for IEEE 802.11e Wireless LAN', *Journal of Internet Technology*, 2004, **5**, (4), pp. 429-434

[4] Li, C.S. and Tseng, Y.C. and Chao, H.C. and Huang, Y.M.: 'A Neighbor Caching Mechanism for Handoff in IEEE 802.11 Wireless Networks'. *Journal of Supercomputing*, 2008, **45**, (1), pp. 1-14

[5] Chang, C.Y. and Wang, H.J. and Chao, H.C. and Park, J.H.: 'IEEE 802.11 Handoff Latency Improvement Using Fuzzy Logic', *Wirel. Commun. Mob. Comput.*, 2008, **8**, (13), pp. 1201-1213

[6] He, Y. and Sun, J. and Yuan, R. and Gong, W.: 'A Reservation Based Backoff Method for Video Streaming in 802.11 Home Networks', *IEEE J. Sel. Commun.*, 2010, **28**, (3), pp. 332-343

[7] Rivero-Angeles, M.E. and Lara-Rodriguez, D. and Cruz-Perez, F.A.: 'Random-Access Control Mechanisms Using Adaptive Traffic Load in ALOHA and CSMA Strategies for EDGE', *IEEE Trans. Vehic. Tech.*, 2005, **54**, (3), pp. 1160-1186

[8] Siva Ram Murthy, C. and Manoj, B.S.: 'Ad Hoc Wireless Networks Architectures and Protocols', (Pearson Education. Inc, 5th edn. 2004)

[9] Hamidian, A. and Körner, U.: 'Extending EDCA with Distributed Resource Reservation for QoS Guarantees', *J. Sel. Telecommun Syst.*, 2008, **39**, (3-4), pp. 187-194

[10] Yu, X. and Navaratnam, P. and Moessner, K.: 'Distributed Resource Reservation Mechanism for IEEE 802.11e-Based Networks', *Proc. IEEE Vehic. Tech. Conf. 2010-Fall*, Ottawa, Canada, Sep. 2010, pp. 1-5

[11] Pangun Park and Fischione, C. and Johansson, K.H.: 'Performance Analysis of GTS Allocation in Beacon Enabled IEEE 802.15.4', *Proc. IEEE Conf. Sensor, Mesh and Ad Hoc Commun. and Netw.*, Rome, Italy, Jun. 2009, pp. 1-9

[12] Sikdar, B.: 'An Analytic Model for the Delay in IEEE 802.11 PCF MAC-Based Wireless Networks', *IEEE Trans. Wirel. Commun.*, 2007, **6**, (4), pp. 1542-1550

[13] Kim, M. and Chung, J.M.: 'Performance Analysis of IEEE 802.11e HCCA for V2I Communications in WAVE Networks', *Proc. IEEE International Midwest Symposium on Circuits and systems*, Seattle, Washington, Aug. 2010, pp. 328-331

[14] Deng, D.J. and Ke, C.H. and Chao, H.C. and Y.M. Huang.: 'On Delay Constrained CAC Scheme and Scheduling Policy for CBR Traffic in IEEE 802.11e Wireless LANs', *Wirel. Commun. and Mob. Comput.*, 2010, **10**, (11), pp. 1509-1520

[15] Reddy, T. and John, P. and Murthy, C.: 'Providing MAC QoS for Multimedia Traffic in 802.11e Based Multi-Hop Ad Hoc Wireless Networks', *J. Sel. Comput. Netw.*, 2007, **51**, (1), pp. 153-176

[16] Bianchi, G.: 'Performance Analysis of the IEEE 802.11 Distributed Coordination Function', *IEEE J. Sel. Commun.*, 2000, **18**, (3), pp. 535-547